А. Я. ХИНЧИН

РАБОТЫ ПО МАТЕМАТИЧЕСКОЙ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

Под редакцией Б. В. ГНЕДЕНКО

ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1963

АННОТАЦИЯ

Теория массового обслуживания — важная ветвь современной теории вероятностей, развившаяся в последние годы. Эта теория может быть использована для наиболее экономного проектирования любых систем, предназначенных для удовлетворения массового потока каких-либо заявок случайного характера (например, телефонных станций, различных устройств для сбора и обработки информации и т. д.). В то же время проблемы, требующие применения тех же математических методов, возникают при автоматизации производства, организации транспорта, связи и снабжения, в военном деле.

Настоящая книга составлена из работ выдающегося советского математика, которые в своей совокупности представляют прекрасно написанное введение в изучение теории массового обслуживания. Для понимания книги необходимо владеть курсом математического анализа в объеме втузовской программы и основными понятиями теории вероятностей.

Книга рассчитана на научных работников, аспирантов и студентов старших курсов в области математики, различных отраслей техники и экономики.

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ РЕДАКТОРА

В настоящей книге собраны исследования выдающегося советского математика Александра Яковлевича Хинчина (1894—1959), относящиеся к той области теории вероятностей, которую он назвал теорией массового обслуживания.

Интерес к задачам теории массового обслуживания возник у Александра Яковлевича еще в начале тридцатых годов, когда он в качестве депутата Моссовета принял участие в работе секции связи. Для Москвы тогда был период перехода к автоматическим телефонным станциям. Многочисленные научные проблемы, возникавшие при этом, увлекли А. Я. Хинчина; в результате общественное поручение наполнилось и значительным научным содержанием. Помимо работ, возникших в результате личного контакта с крупнейшими советскими представителями науки о телефонной связи, а также изучения существовавшей в ту пору литературы, Александр Яковлевич выполнил также ряд оригинальных исследований по просьбе московского телефонного узла. Связи со специалистами телефонного дела практически не прерывались буквально до последних лет жизни Хинчина. Однако творческая работа в этом направлении разбилась на два периода, отделенных друг от друга почти двадцатью годами.

В первый период активной творческой деятельности в области задач теории массового обслуживания А. Я. Хинчин опубликовал лишь две статьи: «Математическая теория стационарной очереди» (Матем. сб. **39**, 1932, стр. 73—84) и «О среднем времени простоя станков» (там же **40**, 1933, стр. 119—123). В рукописях А. Я. Хинчина мне удалось найти литературно оформленную статью «Теория спаренных аппаратов», которая, как мне известно, нигде не была опубликована. С научной точки зрения особый интерес представляла первая из названных работ, так как в ней, помимо решения остро поставленной задачи вычисления распределения длительности ожидания для случая простейшего потока вызовов, поступающих на один прибор, обслуживаемых с очередью для произвольного распределения времени обслуживания, был впервые использован весьма перспективный метод. Этот метод впоследствии получил наименование «метода вложенных цепей Маркова» (см., например, работу Д. Кендалла «Стохастические процессы, встречающиеся в теории очередей, и их анализ методом вложенных цепей Маркова». Сборник переводов «Математика», 3:6, 1959, стр. 97—111).

Интерес к проблемам теории массового обслуживания пробудился у Александра Яковлевича вновь в 1953 г., вскоре после завершения цикла работ по использованию предельных теорем теории вероятностей для целей квантовой статистики. Этот интерес был вызван оживившимися связями с работниками телефонного дела. Консультации и курс лекций для инженеров, которые он тогда прочел, заставили его ознакомиться со значительной новой журнальной литературой. В результате наметились задачи, которые были восприняты А. Я. Хинчиным как центральные для всей теории массового обслуживания. Кратко эти задачи можно охарактеризовать как стремление изучить строение потока требований, поступающих на обслуживание. Собственно, именно этой задаче посвящена значительная часть его монографии «Математические методы теории массового обслуживания» (Труды математического института им. В. А. Стеклова, т. 49, Изд. АН СССР, М., 1955), а также две его последующие статьи: «Потоки случайных событий без последействия» (Теория вероятностей и ее применение, т. 1, вып. 1, 1956, стр. 3—18) и

«О пуассоновских потоках случайных событий» (там же, т. 1, вып. 3, 1956, стр. 320—327). В другом направлении было выполнено лишь исследование «О формулах Эрланга в теории массового обслуживания», где дано распространение известных формул Эрланга на случай простейшего входящего потока требований и произвольного распределения длительности обслуживания. Эта работа, будучи подготовлена к печати, не была передана автором, как мне известно, для опубликования.

Сейчас, когда интерес к теории массового обслуживания значительно вырос, издание собрания оригинальных исследований А. Я. Хинчина в этом направлении мне представляется своевременным. Я счел невозможным для себя изменять текст публикуемых работ и ограничился лишь исправлением заведомых опечаток, а также библиографическими замечаниями, которые помогут читателю при разыскании более поздних публикаций. Примечания редактора помечены буквами *Б. Г.*

Для того чтобы читатель составил себе представление о некоторых новых направлениях, в значительной мере связанных с прикладными исследованиями, в конце книги помещен небольшой обзор, написанный мною. В обзоре, в частности, указана основная монографическая литература в этой области, выпущенная в последние годы.

Б. В. Гнеденко

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

ПРЕДИСЛОВИЕ

В науке, практической деятельности людей и в быту каждодневно создаются такие положения, когда возникает массовый спрос на обслуживание какого-либо специального вида, причем обслуживающая организация, располагая лишь ограниченным числом обслуживающих единиц, не всегда способна немедленно удовлетворять все поступающие заявки. Примеры такой ситуации хорошо известны каждому. Очереди у магазинных и билетных касс, в буфетах, парикмахерских и т. д.; невозможность получить билет на нужный поезд из-за его переполнения; задержка в посадке самолетов, вызываемая отсутствием свободных посадочных площадок; задержка в ремонте потерпевших аварию станков из-за нехватки ремонтных бригад — все эти и многие другие аналогичные, хорошо известные примеры, несмотря на существенные различия их реального содержания, с формальной стороны очень близки друг другу. Во всех подобных случаях перед теорией встает, в сущности, одна основная задача: установить с возможной точностью взаимную зависимость между числом обслуживающих единиц и качеством обслуживания. При этом качество обслуживания в различных случаях, естественно, измеряется различными показателями. Большей частью таким показателем служит либо процент заявок, получающих отказ (процент пассажиров, не получивших билетов на данный поезд), либо среднее время ожидания начала обслуживания (очереди различного рода). Разумеется, качество обслуживания во всех случаях тем выше, чем больше число обслуживающих единиц; однако столь же очевидно, что чрезмерный рост этого числа сопряжен с излишним расходом сил и материальных средств; практически поэтому вопрос обычно ставится так, что сначала устанавливается необходимый уровень качества обслуживания, а затем находится минимальное число обслуживающих единиц, при котором этот уровень может быть достигнут.

В задачах подобного рода почти всегда приходится учитывать влияние случайного элемента на течение изучаемого явления. Количество поступающих заявок не является, как правило, постоянным, а испытывает случайные колебания. Время обслуживания заявок в большинстве задач не является стандартным, а подвержено случайным колебаниям от одной заявки к другой. Все эти элементы случайности отнюдь не имеют характера небольших «возмущений», нарушающих собой плавный и закономерный ход явления; напротив, они составляют собой основную черту в картине изучаемых процессов. Естественно поэтому, что математическим инструментом теории массового обслуживания должны стать понятия и методы теории вероятностей — математической дисциплины, посвященной изучению закономерностей случая.

Цель настоящей книги — ознакомить читателя с основными идеями, методами и отдельными ходами мысли, господствующими в приложениях теории вероятностей к вопросам массового обслуживания. Необходимость создания монографии подобного рода уже давно ощущается как математиками, так и практическими работниками (в первую очередь связистами, в свое время вызвавшими к жизни изучаемую теорию и до сих пор остающимися ее главными потребителями). Эта потребность усиливается еще тем обстоятельством, что сколько-нибудь доступных изложений общей теории не имеется и за границей.

Само собой разумеется, что предлагаемая монография ни в какой мере не может претендовать на полноту сообщаемых сведений. Теория массового обслуживания в настоящее время разрослась очень широко, и для сколько-нибудь полного изложения хотя бы только важнейших из ее достижений понадобился бы толстый том. Однако такой цели я себе и не ставил. Моей задачей было — осветить на небольшом числе важнейших примеров общий характер, основной стиль использования вероятностных рассуждений в вопросах массового обслуживания. Исходя из этой «методологической» целевой установки, я и производил отбор материала для моей монографии. Особое внимание было при этом уделено трудам основоположника теории А. К. Эрланга, далеко не все исследования которого известны у нас в той степени, какой они заслуживают. В значительном количестве мною введены в книгу также идеи К. Пальма — крупнейшего современного продолжателя дела Эрланга. В известной мере я учел и пробивающееся все чаще в наши дни стремление искать более элементарные методы исследования в противовес почти безраздельно господствующим со времени классических трудов Эрланга аналитическим методам, связанным с составлением систем дифференциальных, разностно-дифференциальных и интегральных уравнений. В этом направлении я обращаю внимание читателя на § 2 и 25, а также на всю главу 11. Попытка создания общей элементарной теории предпринята недавно в интересной статье Лундквиста (Ericsson Technics, 1953).

Составление этой монографии было очень сильно затруднено тем обстоятельством, что вся основная литература принадлежит перу практических специалистов и в математическом отношении является неудовлетворительной. Чтобы придать всему изложению форму, сколько-нибудь приемлемую в математическом смысле, я не мог оставить почти ни одного рассуждения в его первоначальном виде; приходилось либо существенно дополнять приводимую автором аргументацию, либо выбрасывать ее и заменять другой. В равной степени и вводимые новые понятия во многих случаях пришлось определять по-иному, так как определения, даваемые авторами, оказывались недостаточно четкими.

Чтобы помочь читателю связывать общие понятия с конкретными представлениями, я на протяжении всей книги пользуюсь терминологией, заимствованной из телефонной практики, запросы которой до сих пор остаются важнейшим стимулом развития теории массового обслуживания. Так, я говорю о «вызовах» (вместо «требований» или «заявок»), о «потерях» (вместо «отказов»), о длительности «разговора» (вместо «обслуживания») и т. п. Однако во всех случаях все сказанное относится, разумеется, к любому виду массового обслуживания при соответствующем изменении терминологии.

Книга разделена на три части, из которых первая — самая большая — посвящена изучению «потока» поступающих вызовов и вовсе не содержит вопросов обслуживания (которым посвящены две остальные части). Такое отчетливое выделение в особую часть изучения входящего потока казалось мне целесообразным не только потому, что для правильного обслуживания надо прежде всего хорошо знать то, что обслуживаешь, хорошо знать все черты чередования поступающих заявок; другой, не менее важный повод к такому выделению состоял в том, что теория входящего потока вызовов есть не что иное, как общая теория случайного чередования однородных событий, которая находит себе широкий круг применений и в областях, ни с каким обслуживанием не связанных (например, в вопросах радиоактивного распада атомов).

Я стремился сделать книгу доступной всякому, кто владеет основными понятиями теории вероятностей и хотя бы кратким (втузовским) курсом математического анализа. Во многих случаях изложение могло бы быть сокращено, если бы я, вместо проведения полного доказательства того или иного утверждения, позволил себе сослаться на общие результаты соответствующей вероятностной теории (цепи Маркова, случайные процессы, эргодическая теория). Однако, желая сделать книгу более доступной возможно более широкому кругу читателей, я почти нигде не поддавался такого рода искушению.

В теории вероятностей принято называть законом распределения данной случайной величины ξ (неубывающую) функцию $F(x)$, выражающую собой вероятность неравенства $\xi < x$. В теории массового обслуживания часто оказывается более удобным называть законом распределения величины ξ (невозрастающую) функцию $\Phi(x)$, выражающую собой вероятность неравенства $\xi > x$. Мы будем, в зависимости от обстоятельств, пользоваться тем или другим пониманием закона распределения.

2 октября 1954 г. *А. Хинчин*

ЧАСТЬ I
ВХОДЯЩИЙ ПОТОК ВЫЗОВОВ

Глава 1
ТЕОРИЯ ПРОСТЕЙШЕГО ПОТОКА

Общая теория потоков однородных событий должна естественно начинаться с определения основных общих понятий, связанных с такими потоками. Однако мы отложим такой общий подход до главы 2, где он будет проведен в нужной широте. Мы предпочитаем сразу ввести читателя в круг конкретных исследований, связанных с потоками некоторого простейшего типа, чтобы тем самым с первых страниц дать ему наглядное представление об основных ходах мысли и математических орудиях теории массового обслуживания, о стиле этого учения как математической дисциплины. После того как эти конкретные представления будут в достаточной мере усвоены, изучение более абстрактной общей теории уже не должно будет показаться трудным.

Добавим к этому, что тот простейший тип потока, который мы будем изучать в настоящей главе, в течение долгого времени оставался почти единственным, употреблявшимся в приложениях; лишь в сравнительно недавнее время отчетливо выяснилась необходимость изучения потоков более общего типа; впрочем и в наши дни значительное большинство приложений теории массового обслуживания (в частности, и приложений к телефонному делу) исходит еще из предпосылки, что поступающий поток требований (вызовов) принадлежит простейшему типу. Приложения (особенно технические) теории простейшего потока за последние десятилетия настолько расширились, что в настоящее время даже элементарные курсы теории вероятностей, как правило, включают в свою программу специальные главы, посвященные этой теории.

§ 1. Определение и постановка задачи

Простейшим мы будем называть поток однородных событий, если он обладает следующими тремя свойствами.

1° Стационарность. Каковы бы ни были $t>0$ и целое $k\geqslant 0$, вероятность того, что за промежуток времени $(a, a+t)$ произойдет k событий, одна и та же для всех $a\geqslant 0$ (и, значит, зависит только от k и t); будем во всем дальнейшем обозначать эту вероятность через $v_k(t)$. На протяжении всей книги мы будем иметь дело только с такими потоками, в которых за конечный промежуток времени с вероятностью 1 происходит лишь конечное число событий. Мы будем, таким образом, всегда иметь $\sum\limits_{k=0}^{\infty} v_k(t)=1$ при любом t. Стационарность потока выражает собой неизменность его вероятностного режима во времени.

2° Отсутствие последействия. Вероятность $v_k(t)$ наступления k событий за промежуток времени $(a, a+t)$ не зависит от чередования событий до момента a; другими словами, условная вероятность наступления k событий за промежуток времени $(a, a+t)$, вычисленная при любом предположении о чередовании событий до момента a, равна безусловной вероятности $v_k(t)$ того же события. Отсутствие последействия выражает собой взаимную независимость протеканий потока в непересекающихся между собой промежутках времени.

3° Ординарность. Пусть для данного стационарного потока $\psi(t)$ означает вероятность того, что за (где угодно расположенный) промежуток времени длины t наступит по меньшей мере два события $\left[\text{очевидно, } \psi(t)=1-v_0(t)-\right.$ $\left.-v_1(t)=\sum\limits_{k=2}^{\infty} v_k(t)\right]$. Тогда мы имеем

$$\psi(t)=o(t) \quad (t\longrightarrow 0),$$

или, что то же,

$$\frac{\psi(t)}{t}\longrightarrow 0 \quad (t\longrightarrow 0).$$

Как мы увидим далее, ординарность потока выражает собой практическую невозможность совмещения двух или более событий в один и тот же момент времени.

Итак, ***простейшим потоком однородных событий мы называем всякий стационарный ординарный поток без последействия.***

Основная задача теории простейшего потока состоит в определении вида функций $v_k(t)$; иначе говоря, целью нашей будет отыскание закона распределения числа событий за промежуток времени длины t, рассматриваемого как случайная величина. При этом мы ради определенности реальной интерпретации изучаемого потока будем во всем дальнейшем предполагать, что речь идет о потоке вызовов, поступающих на некоторую телефонную установку, и в соответствии с этим называть наши однородные события «вызовами».

§ 2. Элементарное решение

Разобьем промежуток времени (0, 1) на произвольное число n равных промежутков длины $1/n$. Вероятность того, что в какой-либо из этих частей не произойдет ни одного вызова, равна $v_0(1/n)$; а так как наш поток — без последействия, то

$$v_0(1)=\left[v_0\left(\frac{1}{n}\right)\right]^n,$$

или, полагая $v_0(1)=\theta$,

$$v_0\left(\frac{1}{n}\right)=\theta^{\frac{1}{n}}.$$

Если мы имеем промежуток длины k/n ($k=1, 2, \ldots$), то он разбивается на k промежутков длины $\frac{1}{n}$, вследствие чего

$$v_0\left(\frac{k}{n}\right)=\left[v_0\left(\frac{1}{n}\right)\right]^k=\theta^{k/n}. \tag{2.1}$$

Пусть, наконец, t — любое положительное число и пусть натуральное число k определяется из неравенств

$$\frac{k-1}{n}<t\leqslant\frac{k}{n};$$

так как $v_0(t)$, очевидно, есть невозрастающая функция от t, то

$$v_0\left(\frac{k-1}{n}\right)\geqslant v_0(t)\geqslant v_0\left(\frac{k}{n}\right),$$

или, в силу (2.1),

$$\theta^{\frac{k-1}{n}} \geqslant v_0(t) \geqslant \theta^{\frac{k}{n}};$$

но при $n \to \infty$ мы имеем $\frac{k}{n} \to t$, вследствие чего крайние члены написанных неравенств стремятся к θ^t, и мы находим

$$v_0(t) = \theta^t$$

для любого $t > 0$. При этом постоянное число θ нами определено как $v_0(1)$, и следовательно, $0 \leqslant \theta \leqslant 1$. Однако случаи $\theta = 0$ и $\theta = 1$ не представляют интереса, и мы можем их не рассматривать. В самом деле, при $\theta = 1$ мы имеем $v_0(t) = 1$ при любом $t > 0$, что означает достоверное отсутствие вызовов в любом промежутке времени, т. е. отсутствие какого бы то ни было потока. Если же $\theta = 0$, то $v_0(t) = 0$ при любом $t > 0$; это означает, что вызовы с достоверностью будут получены в любом, сколь угодно малом промежутке времени; но тогда, сколь бы велико ни было k, число вызовов в любом промежутке с достоверностью будет больше, чем k; другими словами, число вызовов в любом промежутке бесконечно с вероятностью 1; но такие потоки мы в § 1 раз навсегда исключили из рассмотрения. Итак, мы можем считать, что $0 < \theta < 1$; поэтому можно положить $\theta = e^{-\lambda}$, где λ — постоянное положительное число, и писать

$$v_0(t) = e^{-\lambda t}. \tag{2.2}$$

Отметим, что при этом выводе мы нигде не пользовались ординарностью нашего потока, так что соотношение (2.2) имеет силу для любого стационарного потока без последействия; это замечание будет для нас важно в дальнейшем.

Теперь обращаемся к отысканию функций $v_k(t)$ при $k > 0$. Существует много различных способов решения этой задачи, и почти все они поучительны, так как заложенные в них методы позволяют решать и ряд более сложных задач. Мы начнем с наиболее элементарного способа. Будем считать чи ло t постоянным и разобьем промежуток $(0, t)$ на произвольное число $n > k$ равных частей (ячеек) длины $t/n = \delta$. Относительно расположения вызовов в этих ячейках возможны две гипотезы:

H_1 — ни в одной из n ячеек не будет более одного вызова;

H_2 — по крайней мере в одной из ячеек произойдет более одного вызова.

Мы, очевидно, имеем

$$v_k(t) = \mathsf{P}(H_1, k) + \mathsf{P}(H_2, k), \tag{2.3}$$

где $\mathsf{P}(H_i, k)$ $(i = 1, 2)$ означает вероятность двойного события: 1) реализуется гипотеза H_i и 2) в промежутке $(0, t)$ поступает k вызовов. Очевидно, что $\mathsf{P}\{H_1, k\}$ есть вероятность такого положения вещей, когда из наших n ячеек какие-либо k содержат по одному вызову, а остальные $n-k$ вообще вызовов не содержат, поэтому

$$\mathsf{P}(H_1, k) = \binom{n}{k} [v_1(\delta)]^k [v_0(\delta)]^{n-k}.$$

В силу формулы (2.2) и ординарности данного потока мы имеем при $n \to \infty$ $(\delta \to 0)$ и постоянном k

$$[v_0(\delta)]^{n-k} = e^{-\lambda\delta(n-k)} = e^{-\lambda t} e^{k\lambda\delta} = e^{-\lambda t}[1 + o(1)];$$

$$[v_1(\delta)]^k = [1 - e^{-\lambda\delta} - \psi(\delta)]^k = [1 - e^{-\lambda\delta} + o(\delta)]^k =$$
$$= (\lambda\delta)^k [1 + o(1)] = \frac{(\lambda t)^k}{n^k}[1 + o(1)],$$

и, следовательно,

$$\mathsf{P}(H_1, k) =$$
$$= e^{-\lambda t}\frac{(\lambda t)^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k}[1 + o(1)] \to e^{-\lambda t}\frac{(\lambda t)^k}{k!}$$

при $n \to \infty$.

С другой стороны, $\mathsf{P}(H_2, k)$, очевидно, не превосходит вероятности гипотезы H_2, т. е. того, что по меньшей мере одна из n ячеек содержит более одного вызова; так как для отдельной ячейки вероятность содержать более одного вызова есть $\psi(\delta)$, то поэтому

$$\mathsf{P}(H_2, k) \leqslant n\psi(\delta) = t\frac{\psi(\delta)}{\delta} \to 0 \quad (n \to \infty);$$

таким образом, правая часть равенства (2.3) при $n \to \infty$ имеет пределом

$$e^{-\lambda t}\frac{(\lambda t)^k}{k!};$$

а так как левая часть (2.3) от n не зависит, то

$$v_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \quad (k=0,\ 1,\ \ldots).$$

Таким образом, для простейшего потока число вызовов в промежутке длины t распределено по закону Пуассона с параметром λt.

§ 3. Метод дифференциальных уравнений

В специальной литературе поставленная нами задача решается обычно другим методом, менее элементарным, но зато легко распространяемым на более сложные задачи. Рассмотрим теперь этот метод. Пусть t и τ — любые положительные числа и пусть $k>0$. Пусть k_1 — число вызовов в промежутке $(0, t)$, а k_2 — в промежутке $(t, t+\tau)$. Для того чтобы в промежутке $(0, t+\tau)$ произошло k вызовов, необходимо и достаточно наступление одного из следующих двойных событий: $k_1=k$, $k_2=0$; $k_1=k-1$, $k_2=1$; $k_1=k-2$, $k_2=2$; ...; $k_1=0$, $k_2=k$. Но вероятности событий $k_1=l$ и $k_2=m$ соответственно равны $v_l(t)$ и $v_m(\tau)$; а так как эти события взаимно независимы (поток без последействия!), то

$$\begin{aligned} v_k(t+\tau) = v_k(t)\,v_0(\tau) + v_{k-1}(t)\,v_1(\tau) &+ \\ + v_{k-2}(t)\,v_2(\tau) + \cdots &+ v_0(t)\,v_k(\tau). \qquad (3.1) \end{aligned}$$

Но при $\tau \to 0$ мы имеем:

$$v_0(\tau) = e^{-\lambda\tau} = 1 - \lambda\tau + o(\tau);$$
$$v_1(\tau) = 1 - v_0(\tau) - \psi(\tau) = \lambda\tau + o(\tau);$$
$$\sum_{l=2}^{k} v_{k-l}(t)\,v_l(\tau) \leqslant \sum_{l=2}^{\infty} v_l(\tau) = \psi(\tau) = o(\tau).$$

и, следовательно, (3.1) дает

$$v_k(t+\tau) = v_k(t)(1-\lambda\tau) + v_{k-1}(t)\,\lambda\tau + o(\tau);$$
$$\frac{v_k(t+\tau) - v_k(t)}{\tau} = \lambda\left[v_{k-1}(t) - v_k(t)\right] + o(1).$$

Это показывает, что функция $v_k(t)$ дифференцируема при любом $t>0$ и что

$$v_k'(t) = \lambda\left[v_{k-1}(t) - v_k(t)\right] \quad (k=1,\ 2,\ \ldots); \qquad (3.2)$$

если положить для общности $v_{-1}(t)\equiv 0$, то уравнение (3.2), как мы непосредственно убеждаемся из (2.2), имеет место и при $k=0$.

Таким образом, для определения искомых функций $v_k(t)$ мы получили систему линейных дифференциальных уравнений (3.2). Эта система легко решается различными методами, из которых мы рассмотрим два наиболее поучительных для дальнейшего.

А. Метод замены искомых функций

Положим
$$v_k(t)=e^{-\lambda t}u_k(t) \qquad (k=0,\ 1,\ 2,\ \ldots),$$
отсюда
$$v_k'(t)=e^{-\lambda t}\left[u_k'(t)-\lambda u_k(t)\right] \qquad (k=0,\ 1,\ 2,\ \ldots);$$
вставляя же эти выражения $v_k(t)$ и $v_k'(t)$ в уравнения (3.2), легко находим
$$u_k'(t)=\lambda u_{k-1}(t) \qquad (k=0,\ 1,\ 2,\ \ldots),$$
где по определению $u_{-1}(t)\equiv 0$. Отсюда, интегрируя, находим
$$u_k(t)-u_k(0)=\lambda\int_0^t u_{k-1}(z)\,dz \qquad (k=0,\ 1,\ 2,\ \ldots).$$

Очевидно, мы имеем при любом $k\geqslant 0$
$$v_k(0)=u_k(0);$$
но по определению функций $v_k(t)$
$$v_0(0)=1, \quad v_k(0)=0 \quad (k>0);$$
поэтому и $u_0(0)=1$, $u_k(0)=0$ $(k=1,\ 2,\ \ldots)$, и мы находим при $k\geqslant 1$
$$u_k(t)=\lambda\int_0^t u_{k-1}(z)\,dz. \tag{3.3}$$

Замечая, что в силу (2.2) мы имеем по определению функций $u_k(t)$

мы по формуле (3.3) рекуррентно находим

$$u_1(t)=\lambda t;$$
$$u_2(t)=\frac{(\lambda t)^2}{2};$$
$$\ldots\ldots\ldots;$$
$$u_k(t)=\frac{(\lambda t)^k}{k!};$$
$$\ldots\ldots\ldots,$$

и, следовательно,

$$v_k(t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!},$$

т. е. получаем прежнее решение задачи.

В. Метод производящих функций

Положим

$$\sum_{k=0}^{\infty} v_k(t)\, x^k=\Phi(t,\ x);$$

ряд в левой части этого равенства во всяком случае абсолютно сходится при $|x|\leqslant 1$. Умножая на x^k все члены уравнения (3.2) и суммируя по k от 0 до ∞, мы легко находим

$$\frac{\partial\Phi}{\partial t}=\lambda\sum_{k=0}^{\infty} v_{k-1}(t)\, x^k-\lambda\Phi=\lambda(x-1)\Phi,$$

или

$$\frac{\partial \ln\Phi}{\partial t}=\lambda(x-1);$$

отсюда

$$\ln\Phi(t,\ x)-\ln\Phi(0,\ x)=\lambda(x-1)\,t; \tag{3.4}$$

но легко видеть, что при любом x

$$\Phi(0,\ x)=v_0(0)=1,$$

поэтому (3.4) дает

$$\Phi(t,\ x)=e^{\lambda(x-1)t}=e^{-\lambda t}e^{\lambda tx}=e^{-\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^k}{k!}x^k.$$

Сопоставляя это с определением функции $\Phi(t, x)$, мы непосредственно видим, что

$$v_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad (k = 0, 1, 2, \ldots),$$

т. е. снова приходим к прежнему решению задачи.

§ 4. Интенсивность простейшего потока

Полученные нами результаты показывают, что те три свойства (стационарность, отсутствие последействия, ординарность), которыми мы определили простейший поток, полностью характеризуют его структуру с точностью до значения параметра λ, которое может быть любым положительным числом. Два простейших потока могут отличаться друг от друга только значениями этого параметра.

Условимся обозначать в дальнейшем для любого стационарного потока через $w(t)$ вероятность того, что за промежуток времени t произойдет по меньшей мере один вызов. Очевидно, мы имеем

$$w(t) = 1 - v_0(t) = \sum_{k=1}^{\infty} v_k(t) = v_1(t) + \psi(t),$$

где $\psi(t) = \sum_{k=2}^{\infty} v_k(t)$ по-прежнему означает вероятность поступления по меньшей мере двух вызовов за промежуток времени длины t. Для простейшего потока с параметром λ $v_0(t) = e^{-\lambda t}$ и, значит, при $t \to 0$

$$w(t) = 1 - e^{-\lambda t} = \lambda t + o(t),$$

или, что то же,

$$\lim_{t \to 0} \frac{w(t)}{t} = \lambda. \tag{4.1}$$

Мы можем считать это соотношение *определением* параметра λ для данного потока. Мы узнаем в дальнейшем, что предел (4.1) существует у любого стационарного потока и определенный соотношением (4.1) параметр λ служит одной из важнейших характеристик этого потока.

Но вернемся к простейшему потоку и найдем теперь математическое ожидание числа вызовов, поступающих за промежуток времени длины t. Оно равно

$$\sum_{k=1}^{\infty} k v_k(t) = e^{-\lambda t} \sum_{k=1}^{\infty} \frac{(\lambda t)^k}{(k-1)!} = e^{-\lambda t} \lambda t \sum_{k=1}^{\infty} \frac{(\lambda t)^{k-1}}{(k-1)!} = \lambda t,$$

так как последняя сумма, очевидно, равна $e^{\lambda t}$. Мы могли бы предвидеть этот результат и заранее, так как известно, что математическое ожидание величины, распределенной по закону Пуассона, равно параметру этого закона, т. е. в данном случае равно λt.

Математическое ожидание числа вызовов в единицу времени называют **интенсивностью** данного потока; мы будем обозначать эту интенсивность через μ. Как мы только что установили, для простейшего потока $\mu = \lambda$. Однако для стационарных потоков более сложной структуры это равенство не только не очевидно, но и не всегда верно; в этом вопросе мы подробно разберемся в дальнейшем. Сейчас же убедимся только, что для любого стационарного потока $\mu \geqslant \lambda$. В самом деле, математическое ожидание числа вызовов за время t для данного потока равно

$$\mu t = \sum_{k=1}^{\infty} k v_k(t) \geqslant \sum_{k=1}^{\infty} v_k(t) = w(t),$$

$$\mu \geqslant \frac{w(t)}{t},$$

а так как левая часть этого неравенства от t не зависит, то в силу (4.1) $\mu \geqslant \lambda$; разумеется, самое существование предела (4.1) для любого стационарного потока еще должно быть доказано.

Итак, *интенсивность μ простейшего потока совпадает с его параметром λ; для произвольного же стационарного потока мы пока можем утверждать только, что $\mu \geqslant \lambda$.* При этом существование предела (4.1) для любого стационарного потока еще должно быть доказано*).

---

*) В сущности и параметр λ может быть реально интерпретирован как «интенсивность» данного потока, так как соотношение (4 1) показывает, что вероятность поступления вызовов в промежутке бесконечно малой длины t асимптотически пропорциональна t, и коэффициентом пропорциональности служит как раз параметр λ. Можно было бы говорить о «верхней интенсивности» μ и «нижней интенсивности» λ (так как всегда $\mu \geqslant \lambda$). В дальнейшем (§ 11) мы узнаем, что для ординарного потока всегда $\mu = \lambda$.

§ 5. Поток с переменным параметром

В этой книге мы будем изучать почти исключительно стационарные потоки вызовов. Однако для некоторых простейших задач решение в нестационарном случае настолько легко проводится и вместе с тем имеет столь ясное практическое значение, что было бы жаль оставить его совсем без рассмотрения. В частности, в настоящем параграфе мы подвергнем изучению потоки, не обладающие стационарностью, но являющиеся, подобно простейшему потоку, ординарными потоками без последействия. Мы сейчас более точно поясним смысл этих предпосылок.

Если поток не стационарен, то вероятность получить k вызовов за промежуток времени длины τ зависит не только от τ, но и от начального момента t этого промежутка: поэтому мы будем обозначать ее через $v_k(\tau, t)$. Таким образом, $v_k(\tau, t)$ есть вероятность того, что за промежуток времени $(t, t+\tau)$ произойдет k вызовов. По аналогии со стационарным случаем мы полагаем

$$1-v_0(\tau, t)=w(\tau, t); \quad 1-v_0(\tau, t)-v_1(\tau, t)=\psi(\tau, t).$$

Мы будем называть исследуемый поток ординарным, если при $\tau \to 0$ и любом постоянном $t \geqslant 0$ имеет место соотношение

$$\frac{\psi(\tau, t)}{\tau} \to 0.$$

Далее мы должны допустить, что *для любого* $t \geqslant 0$ *существует*

$$\lim_{\tau \to 0} \frac{w(\tau, t)}{\tau}=\lambda(t) \tag{5.1}$$

(мгновенное значение параметра).

Исходя из этих предпосылок, мы поставим себе задачей найти выражение функций $v_k(\tau, t)$. Рассмотрим, как и прежде, сначала случай $k=0$.

Так как мы имеем дело с потоком без последействия, то при $\Delta\tau > 0$

$$v_0(\tau+\Delta\tau, t)=v_0(\tau, t)\, v_0(\Delta\tau, t+\tau);$$

но по предположению при $\Delta\tau \to 0$ и постоянных t, τ

$$v_0(\Delta\tau, t+\tau)=1-w(\Delta\tau, t+\tau)=1-\lambda(t+\tau)\Delta\tau+o(\Delta\tau),$$

следовательно,

$$v_0(\tau+\Delta\tau, t)-v_0(\tau, t)=-v_0(\tau, t)\lambda(t+\tau)\Delta\tau+o(\Delta\tau);$$

это после почленного деления на $\Delta\tau$ в пределе приводит к соотношению

$$\frac{\partial v_0(\tau, t)}{\partial\tau}=-\lambda(t+\tau)v_0(\tau, t) \tag{5.2}$$

(причем существование производной, очевидно, попутно доказывается); отсюда же

$$\frac{\partial \ln v_0(\tau, t)}{\partial\tau}=-\lambda(t+\tau),$$

и, следовательно,

$$\ln v_0(\tau, t)-\ln v_0(0, t)=-\int_0^\tau \lambda(t+u)\,du;$$

а так как $\ln v_0(0, t)=0$, то

$$\ln v_0(\tau, t)=-\int_0^\tau \lambda(t+u)\,du=-\Lambda(\tau, t);$$

$$v_0(\tau, t)=e^{-\Lambda(\tau, t)}.$$

В стационарном случае мы имели показателем $-\lambda\tau$; в общем случае, как мы теперь видим, нужно заменить λ величиной

$$\frac{1}{\tau}\Lambda(\tau, t)=\frac{1}{\tau}\int_0^\tau \lambda(t+u)\,du,$$

которую естественно рассматривать как среднее значение «мгновенного параметра» $\lambda(t)$ в промежутке $(t, t+\tau)$.

Переходя теперь к случаю $k>0$, мы аналогично предыдущему [см. § 3] при $\Delta\tau\to 0$ и постоянных t, τ легко находим

$$v_k(\tau+\Delta\tau, t)=v_k(\tau, t)v_0(\Delta\tau, t+\tau)+\\ +v_{k-1}(\tau, t)v_1(\Delta\tau, t+\tau)+o(\Delta\tau),$$

где

$$v_0(\Delta\tau, t+\tau)=1-\lambda(t+\tau)\Delta\tau+o(\Delta\tau)$$

и

$$v_1(\Delta\tau, t+\tau)=1-v_0(\Delta\tau, t+\tau)-\psi(\Delta\tau, t+\tau)=\\ =\lambda(t+\tau)\Delta\tau+o(\Delta\tau),$$

так что

$$v_k(\tau+\Delta\tau, t)=v_k(\tau, t)[1-\lambda(t+\tau)\Delta\tau]+ \\ +v_{k-1}(\tau, t)\lambda(t+\tau)\Delta\tau+o(\Delta\tau),$$

откуда

$$\frac{v_k(\tau+\Delta\tau, t)-v_k(\tau, t)}{\Delta\tau}=\lambda(t+\tau)[v_{k-1}(\tau, t)-v_k(\tau, t)]+o(1),$$

и следовательно, в пределе

$$\frac{\partial v_k(\tau, t)}{\partial\tau}=\lambda(t+\tau)[v_{k-1}(\tau, t)-v_k(\tau, t)]. \quad (5.3)$$

Это соотношение, доказанное нами для любого $k>0$, остается, как показывает (5.2), верным и при $k=0$, если положить

$$v_{-1}(\tau, t)\equiv 0.$$

Мы найдем нужное нам решение системы (5.3), применяя метод производящих функций. Положим

$$F(t, \tau, x)=\sum_{k=0}^{\infty} v_k(\tau, t)x^k.$$

Умножая все члены уравнения (5.3) на x^k и суммируя по k от 0 до ∞, мы находим в точной аналогии с § 3

$$\frac{\partial F}{\partial\tau}=(x-1)\lambda(t+\tau)F,$$

или

$$\frac{\partial \ln F}{\partial\tau}=(x-1)\lambda(t+\tau),$$

откуда

$$\ln F(t, \tau, x)-\ln F(t, 0, x)=(x-1)\int_0^{\tau}\lambda(t+u)\,du= \\ =(x-1)\Lambda(\tau, t). \quad (5.4)$$

При любых x и t мы имеем

$$F(t, 0, x)=v_0(0, t)=1;$$

поэтому (5.4) дает

$$F(t, \tau, x)=e^{(x-1)\Lambda(\tau, t)}=e^{-\Lambda(\tau, t)}e^{x\Lambda(\tau, t)}= \\ =\sum_{k=0}^{\infty} e^{-\Lambda(\tau, t)}\frac{[\Lambda(\tau, t)]^k}{k!}x^k,$$

и сопоставление с определением функции $F(t, \tau, x)$ дает

$$v_k(\tau, t) = e^{-\Lambda(\tau, t)} \frac{[\Lambda(\tau, t)]^k}{k!} \quad (k = 0, 1, 2, \ldots). \tag{5.5}$$

Эти формулы полностью решают поставленную задачу. Мы видим, что и для потока с переменным параметром число вызовов в промежутке $(t, t+\tau)$ подчиняется закону Пуассона; однако параметр этого закона теперь зависит не только от длины τ данного промежутка, но и от его начального момента t. В случае стационарного потока мы имели закон Пуассона с параметром $\lambda\tau$; при переходе к нестационарному случаю мы должны, как видим, заменить постоянное число λ выражением

$$\frac{\Lambda(\tau, t)}{\tau} = \frac{1}{\tau} \int_0^\tau \lambda(t+u)\, du,$$

т. е. средним значением $\lambda(x)$ в промежутке $t \leqslant x \leqslant t+\tau$. Обратно, если функция $\lambda(x) = \lambda$ есть постоянная величина, то, очевидно, при любом t

$$\Lambda(\tau, t) = \int_0^\tau \lambda(t+u)\, du = \lambda\tau$$

и формулы (5.5) переходят в решения, полученные нами в § 3 для стационарного случая.

Заметим, наконец, что число вызовов в промежутке $(t, t+\tau)$, подчиняясь закону Пуассона (5.5), имеет своим математическим ожиданием параметр этого закона, т. е. величину

$$\Lambda(\tau, t) = \int_t^{t+\tau} \lambda(u)\, du;$$

поэтому величину $\frac{1}{\tau}\Lambda(\tau, t)$ можно понимать как *среднюю* интенсивность нашего потока в промежутке $(t, t+\tau)$; предел же этой величины при $\tau \to 0$ есть *мгновенная* интенсивность $\mu(t)$ данного потока в момент t; мы находим

$$\mu(t) = \lim_{\tau \to 0} \frac{1}{\tau} \Lambda(\tau, t) = \lim_{\tau \to 0} \frac{1}{\tau} \int_t^{t+\tau} \lambda(u)\, du = \lambda(t).$$

Таким образом, и в случае простейшего потока с переменным параметром мы имеем совпадение мгновенной интенсивности потока с мгновенным значением параметра.

Глава 2

ОБЩИЕ СВОЙСТВА СТАЦИОНАРНЫХ ПОТОКОВ *)

§ 6. Поток вызовов как случайный процесс

Как мы уже говорили в начале главы 1, потоки вызовов, с которыми мы встречаемся на практике, во многих случаях могут с достаточно хорошим приближением рассматриваться как простейшие; при изучении таких потоков поэтому обычно пользуются теми результатами, которые получены нами в главе 1. Однако в последние годы, когда усложняющаяся практика ставит перед наукой задачи все более сложные и требующие все более точного решения, встала настоятельная необходимость изучения потоков более общего типа. Непосредственных поводов для такого расширения изучаемой области имеется в основном два.

С одной стороны, статистика потоков вызовов даже в самых обычных условиях при возрастающей точности показывает, что выводы, основанные на предположении простейшего характера потока, недостаточно хорошо согласуются с опытными данными; можно без труда и теоретически предвидеть необходимость такого рода расхождений: нетрудно сообразить, например, что в действительности почти всегда следует ожидать в изучаемом потоке известного последействия и что не всегда этим последействием можно пренебрегать.

С другой стороны, имеются и такие случаи, когда изучаемый поток заведомо и принципиально отличается от простейшего; таковы все виды потоков переменной интенсивности (они не стационарны, см. § 5); таковы же и потоки, поступающие на вторую, третью, и т. д. линию «полнодоступного пучка», даже в том случае, когда на первую линию поступает простейший поток; все эти потоки обладают значитель-

*) Естественным продолжением результатов настоящей главы являются работы А. Я. Хинчина «Потоки случайных событий без последействия» и «О пуассоновских потоках случайных событий», помещенные в настоящей книге (стр. 170 и 190). — *Б. Г.*

ным последействием, возрастающим с номером линии, и учет этого последействия обязателен для теории (см. гл. 8). Поэтому современные исследования по теории массового обслуживания не могут ограничиться рассмотрением простейших потоков и вынуждены расширить в той или другой мере исходные предпосылки.

Переходя к исследованию потоков более общего типа, мы должны в целях строгости и недвусмысленной ясности изложения начать с точного определения основных понятий; мы не сделали этого в главе 1, так как рассматривали эту главу как вводную, имеющую целью на простейшем примере показать характерные для всей теории потоков образцы применяемых в ней математических методов.

Если мы обозначим через $x(t)$ число вызовов, поступающих за промежуток времени $(0, t)$, то для каждого фиксированного значения $t > 0$ $x(t)$ представляет собой случайную величину. При переменном t, $x(t)$ представляет собой однопараметрическое семейство случайных величин, которое называют *случайным процессом* или *случайной функцией*. Для функции $x(t)$ характерно то, что она: 1) может принимать только целые неотрицательные значения и 2) с возрастанием t никогда не убывает. График такой функции

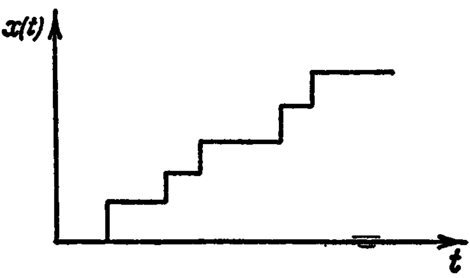

Рис. 1.

поэтому, независимо от случая, всегда имеет форму «лестницы», изображенную на рис. 1.

Для задания любого случайного процесса $x(t)$ как такового надо, чтобы для любой конечной группы положительных чисел $t_1, t_2, \ldots, t_n$ был задан n-мерный закон распределения вектора

$$x(t_1),\ x(t_2),\ \ldots,\ x(t_n).$$

Если процесс $x(t)$ представляет собой поток вызовов и, следовательно, $x(t)$ может принимать только целые неотрицательные значения, то для задания этого потока как случайного процесса надо задать для каждой группы положительных $t_1, t_2, \ldots, t_n$ и каждой группы целых неотрицательных чисел $k_1, k_2, \ldots, k_n$ вероятность системы равенств $x(t_1)=k_1$, $x(t_2)=k_2, \ldots, x(t_n)=k_n$; очевидно, эта вероятность может быть отличной от нуля только в том случае, если при $t_1<t_2<\ldots<t_n$ мы имеем и $k_1\leqslant k_2\leqslant\ldots\leqslant k_n$. В частности ($n=1$), для любого $t>0$ и любого целого неотрицательного k должна быть известна вероятность равенства $x(t)=k$, которую мы в главе 1 обозначали через $v_k(t)$ *). Таким образом, система функций $v_k(t)$ $(k=0, 1, 2, \ldots)$ входит как обязательный элемент в состав описания каждого потока вызовов. В общем случае, однако, задания этой системы функций для полной характеристики потока еще недостаточно.

Поток вызовов называется *стационарным*, если при $0<t_1<t_2<\ldots<t_n$ и при любом положительном a закон распределения вектора $x(t_i)$ $(1\leqslant i\leqslant n)$ совпадает с законом распределения вектора $x(a+t_i)-x(a)$ $(1\leqslant i\leqslant n)$; иначе говоря, закон распределения вектора $x(a+t_i)-x(a)$ $(1\leqslant i\leqslant n)$ зависит от чисел t_i, но не зависит от a. В частности ($n=1$), $v_k(t)$ для стационарного процесса означает вероятность поступления k вызовов в промежутке $(a, a+t)$, где $a\geqslant 0$ произвольно, т. е. в любом промежутке длины t **).

Данный поток вызовов называется *потоком без последействия*, если закон распределения вектора

---

*) Необходимо, впрочем, отметить, что в главе 1 $v_k(t)$ (в силу предположенной стационарности потока) означало вероятность поступления k вызовов в *любом* промежутке времени длины t, в то время как здесь речь идет лишь о промежутке $(0, t)$.

**) Для читателя, знакомого с теорией случайных процессов, заметим, что определенный таким образом стационарный поток вызовов не является, конечно, стационарным случайным процессом в общепринятом смысле этого термина. В теории случайных процессов наш стационарный поток принадлежит к числу «процессов со стационарными приращениями».

$x(a+t_i)-x(a)$ $(t_i>0,\ 1\leqslant i\leqslant n)$ при любом $a\geqslant 0$ не зависит от значений величины $x(t)$ при каких-либо значениях $t<a$. Это определение, очевидно, в точной форме выражает то требование, чтобы случайное течение потока вызовов после какого-либо момента времени a было независимым от его течения до момента a; но в этом и состоит отсутствие последействия в понимании теории вероятностей.

Легко видеть, что стационарный поток без последействия полностью характеризуется системой функций $v_k(t)$ $(k=0, 1, 2, \ldots)$, т. е. законом распределения числа вызовов, поступающих в течение (где угодно расположенного) промежутка времени длины t. В самом деле, так как система равенств

$$x(t_i)=k_i \quad (1\leqslant i\leqslant n,\ 0<t_1<t_2<\ldots<t_n),$$

очевидно, равносильна системе равенств

$$x(t_i)-x(t_{i-1})=k_i-k_{i-1} \quad (1\leqslant i\leqslant n),$$

где для общности положено $t_0=k_0=x(t_0)=0$, то (обозначая через $\mathrm{P}\{\ \}$ вероятность события, помещенного в фигурных скобках) мы будем иметь

$$\mathrm{P}\{x(t_i)=k_i,\ 1\leqslant i\leqslant n\}=\\ =\mathrm{P}\{x(t_i)-x(t_{i-1})=k_i-k_{i-1},\ 1\leqslant i\leqslant n\};$$

так как промежутки (t_0, t_1), (t_1, t_2), $\ldots$, (t_{n-1}, t_n) взаимно не перекрываются, то отсюда в силу отсутствия последействия

$$\mathrm{P}\{x(t_i)=k_i,\quad 1\leqslant i\leqslant n\}=\\ =\prod_{i=1}^{n}\mathrm{P}\{x(t_i)-x(t_{i-1})=k_i-k_{i-1}\},$$

а так как в силу стационарности потока

$$\mathrm{P}\{x(t_i)-x(t_{i-1})=k_i-k_{i-1}\}=\mathrm{P}\{x(t_i-t_{i-1})=k_i-\\ -k_{i-1}\}=v_{k_i-k_{i-1}}(t_i-t_{i-1}) \quad (1\leqslant i\leqslant n),$$

то *)

$$\mathrm{P}\{x(t_i)=k_i,\ 1\leqslant i\leqslant n\}=\prod_{i=1}^{n}v_{k_i-k_{i-1}}(t_i-t_{i-1}).$$

*) Здесь и в дальнейшем мы, разумеется, полагаем $v_r(t)=0$ для любого отрицательного индекса r.

Это показывает, что заданием системы функций $v_k(t)$ действительно однозначно определяются вероятности вида $\mathsf{P}\{x(t_i)=k_i,\quad 1\leqslant i\leqslant n\}$, т. е. полностью характеризуется данный поток как случайный процесс.

В частности, для простейшего потока с параметром λ мы имели (§ 2)

$$v_k(t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}\quad (k=0,\ 1,\ 2,\ \ldots);$$

поэтому для простейшего потока с параметром λ мы имеем при

$$0=t_0<t_1<\ldots<t_n,\ 0\leqslant k_0\leqslant k_1\leqslant\ldots\leqslant k_n$$

$$\mathsf{P}\{x(t_i)=k_i,\ 1\leqslant i\leqslant n\}=$$

$$=\prod_{i=1}^{n}e^{-\lambda(t_i-t_{i-1})}\frac{\lambda^{k_i-k_{i-1}}(t_i-t_{i-1})^{k_i-k_{i-1}}}{(k_i-k_{i-1})!}=$$

$$=e^{-\lambda t_n}\lambda^{k_n}\prod_{i=1}^{n}\frac{(t_i-t_{i-1})^{k_i-k_{i-1}}}{(k_i-k_{i-1})!}.$$

§ 7. Основное свойство стационарных потоков

Для любого стационарного потока условимся, как в главе 1, обозначать через $w(t)$ вероятность того, что в течение промежутка времени длины t произойдет по крайней мере один вызов, так что

$$w(t)=1-v_0(t)=\sum_{k=1}^{\infty}v_k(t).$$

В § 4 мы убедились, что для стационарного потока без последействия (в частности, для простейшего потока) отношение $w(t)/t$ при $t\longrightarrow 0$ стремится к определенному пределу λ, который мы называли параметром данного потока и который, как мы видели, имеет важнейшее значение для изучения основных свойств этого потока. Соотношение

$$\frac{w(t)}{t}\longrightarrow\lambda\quad (t\longrightarrow 0), \tag{7.1}$$

равносильное соотношению

$$w(t) = \lambda t + o(t) \quad (t \to 0), \tag{7.2}$$

часто выражают, говоря, что $w(t)$ при малых t «асимптотически пропорционально» t. В подавляющем большинстве изложений теории простейшего потока соотношение (7.2) [или (7.1)] прямо включается в определение простейшего потока *), что, как мы видели, является излишним, так как это соотношение выводится как простое следствие требований стационарности и отсутствия последействия.

Однако соотношение (7.1) на самом деле обладает еще значительно более широкой областью применимости; оно имеет место, как мы теперь убедимся, *для любого стационарного потока.* Параметром, как мы его определяли в главе 1, обладает, таким образом, каждый стационарный поток, независимо от наличия или отсутствия последействия. Это обстоятельство дает нам, как мы увидим, весьма удобный опорный пункт для изучения общих свойств стационарных потоков.

Доказательство существования предела (7.1) для любого стационарного потока опирается на следующую элементарную лемму теории пределов, которая, как мы увидим, пригодится нам и в дальнейшем.

Лемма. *Пусть функция $f(x)$ — неотрицательная и неубывающая в отрезке $0 < x \leqslant a$ и $f(x+y) \leqslant f(x) + f(y)$, если x, y и $x+y$ принадлежат отрезку* $(0, a)$. *Тогда отношение $f(x)/x$ при $x \to 0$ либо безгранично возрастает, либо стремится к некоторому пределу; этот предел равен нулю только в тривиальном случае* $f(a) = 0$.

Доказательство. Из неравенства $f(x+y) \leqslant f(x) + f(y)$ легко следует, что

$$f(x) \leqslant mf\left(\frac{x}{m}\right) \tag{7.3}$$

для $0 < x \leqslant a$ и любого натурального числа m; в частности, при $x = a$

$$\frac{f\left(\frac{a}{m}\right)}{\frac{a}{m}} \geqslant \frac{f(a)}{a} \quad (m = 1, 2, \ldots);$$

---

*) См. Erlang [7], Феллер [3], Фрай [4], Хинчин [5].

это показывает, что [за исключением тривиального случая $f(a)=0$]

$$\alpha=\overline{\lim_{x\to 0}}\frac{f(x)}{x}\geqslant\frac{f(a)}{a}>0,$$

причем не исключен случай $\alpha=+\infty$.

Допустим сначала, что $\alpha<+\infty$. Пусть число $c>0$ таково, что

$$\frac{f(c)}{c}>\alpha-\varepsilon, \tag{7.4}$$

где $\varepsilon>0$ — сколь угодно малое наперед заданное число, и пусть $0<x<c$. Определим натуральное число $m\geqslant 2$ из неравенств

$$\frac{c}{m}\leqslant x<\frac{c}{m-1};$$

тогда в силу (7.3) и предположенной монотонности $f(x)$

$$\frac{f(x)}{x}\geqslant\frac{f\left(\frac{c}{m}\right)}{\frac{c}{m-1}}=\frac{m-1}{m}\frac{mf\left(\frac{c}{m}\right)}{c}\geqslant\frac{m-1}{m}\frac{f(c)}{c}, \tag{7.5}$$

и, значит, в силу (7.4)

$$\frac{f(x)}{x}\geqslant\left(1-\frac{1}{m}\right)(\alpha-\varepsilon),$$

а так как ε произвольно мало и $m\to\infty$ при $x\to 0$, то

$$\lim_{x\to 0}\frac{f(x)}{x}=\alpha,$$

и лемма доказана. Рассуждение остается в принципе тем же при $\alpha=+\infty$. Мы берем произвольно большое $A>0$ и выбираем число c так, что $f(c)/c>A$; тогда из (7.5)

$$\frac{f(x)}{x}\geqslant\frac{m-1}{m}A,$$

и значит, $f(x)/x\to+\infty$ при $x\to 0$.

Теорема. *Для любого стационарного потока существует*

$$\lim_{t\to 0}\frac{w(t)}{t}=\lambda>0,$$

причем не исключен случай $\lambda=+\infty$.

Доказательство. Достаточно показать, что функция $w(t)$ в некотором отрезке $(0, a)$ удовлетворяет всем предпосылкам только что доказанной леммы. Очевидно, что $w(t)\geqslant 0$ и с возрастанием t не может убывать; очевидно также, что в силу стационарности потока $w(a)>0$, если мы отбросим тривиальный случай потока, в котором вызовы вообще невозможны. Наконец, если произошел по меньшей мере один вызов в промежутке $(0,\ t_1+t_2)$, то, очевидно, то же самое должно иметь место хотя бы для одного из двух промежутков $(0,\ t_1)$ и $(t_1,\ t_1+t_2)$, откуда

$$w(t_1+t_2)\leqslant w(t_1)+w(t_2)\quad (t_1>0,\ t_2>0,\ t_1+t_2<a);$$

таким образом, и эта последняя предпосылка оказывается выполненной. Применяя лемму, мы видим, что теорема доказана.

Если данный стационарный поток есть поток без последействия, то для него, как мы это показали в § 2, мы имеем

$$v_0(t)=e^{-\lambda t}\quad (\lambda>0 \text{ постоянная}),$$

за исключением случаев, когда в любом промежутке времени либо с достоверностью вовсе не поступает вызовов, либо с достоверностью поступает бесконечное множество вызовов; эти два случая, как не имеющие практического значения, мы условились выше исключить из рассмотрения. Отсюда для потока без последействия

$$w(t)=1-e^{-\lambda t},\quad \lim_{t\to 0}\frac{w(t)}{t}=\lambda;$$

предел, существование которого мы доказали в последней теореме, в случае потока без последействия всегда есть, таким образом, некоторое конечное положительное число. Но если допустить возможность последействия, то λ может обращаться в $+\infty$ и в случаях, не исключенных нами из рассмотрения.

Чтобы в этом убедиться, рассмотрим следующий пример стационарного потока. Вообразим себе простейший поток, параметр λ которого представляет собой случайную величину, распределенную по некоторому закону $F(x)$ $[F(+0)=0$, $F(+\infty)=1]$. В конце § 6 мы определили вероятность системы равенств $x(t_i)=k_i\ (1\leqslant i\leqslant n)$ для простейшего потока с данным значением параметра λ; будем теперь для краткости обозначать эту вероятность через $\mathsf{P}_\lambda(t_i,\ k_i)$. Если параметр λ есть случайная величина, распределенная по закону $F(x)$, то вероятность системы равенств $x(t_i)=$ $=k_i\ (1\leqslant i\leqslant n)$ будет (по формуле полной вероятности) равна

$$\mathsf{P}(t_i,\ k_i)=\int_0^\infty \mathsf{P}_\lambda(t_i,\ k_i)\,dF(\lambda). \tag{7.6}$$

Эта вероятность определена, таким образом, для любых n, t_i, k_i; а это, как мы знаем, означает задание определенного потока; этот поток будет, очевидно, стационарным; но, вообще говоря, он будет потоком с последействием. Если $w(t)$ и $v_k(t)$ имеют обычное значение для потока (7.6), а $w_\lambda(t)$ и $v_{k\lambda}(t)$ означают те же величины для простейшего потока с параметром λ, то мы, очевидно, имеем

$$w(t)=\int_0^\infty w_\lambda(t)\,dF(\lambda)$$

и

$$w_\lambda(t)=1-v_{0\lambda}(t)=1-e^{-\lambda t},$$

откуда

$$\frac{w(t)}{t}=\int_0^\infty \frac{1-e^{-\lambda t}}{t}\,dF(\lambda).$$

Если

$$\int_0^\infty \lambda\,dF(\lambda)<+\infty$$

[т. е. закон $F(x)$ имеет конечное математическое ожидание], то в силу неравенства $1-e^{-\lambda t}\leqslant\lambda t$ мы получаем

$$\frac{w(t)}{t}\leqslant\int_0^\infty \lambda\,dF(\lambda);$$

отношение $w(t)/t$ при $t \to 0$ ограничено, и рассматриваемый нами поток имеет конечный параметр. Но если интеграл

$$\int_0^\infty \lambda \, dF(\lambda) \tag{7.7}$$

расходится, то $w(t)/t \to +\infty$ при $t \to 0$. В самом деле, пусть $\varepsilon > 0$ произвольно мало и A столь велико, что

$$\int_0^A \lambda \, dF(\lambda) > \frac{1}{\varepsilon};$$

тогда в силу

$$\frac{w(t)}{t} \geqslant \int_0^A \frac{1 - e^{-\lambda t}}{t} \, dF(\lambda) \to \int_0^A \lambda \, dF(\lambda) \quad (t \to 0)$$

мы имеем

$$\lim_{t \to \infty} \frac{w(t)}{t} \geqslant \frac{1}{\varepsilon},$$

и значит, $w(t)/t \to +\infty$ при $t \to 0$. Таким образом, в случае расходимости интеграла (7.7) наш поток имеет бесконечное значение параметра. Вместе с тем этот поток с вероятностью 1 дает конечное число вызовов в любом конечном промежутке времени; мы имеем

$$v_k(t) = \int_0^\infty v_{k\lambda}(t) \, dF(\lambda),$$

или, так как $v_{k\lambda}(t) = e^{-\lambda t}(\lambda t)^k/k!$,

$$v_k(t) = \frac{t^k}{k!} \int_0^\infty \lambda^k e^{-\lambda t} \, dF(\lambda) > 0,$$

$$\sum_{k=0}^\infty v_k(t) = \int_0^\infty e^{-\lambda t} \sum_{k=0}^\infty \frac{(\lambda t)^k}{k!} \, dF(\lambda) = \int_0^\infty dF(\lambda) = 1.$$

§ 8. Общая форма стационарного потока без последействия

Как мы видели в главе 1, стационарный поток без последействия, если он сверх того обладает еще свойством ординарности, есть простейший поток, общая структура которого легко может быть установлена. Теперь мы поставим себе задачу найти общий вид стационарного потока без последействия, отбрасывая требование ординарности.

Как мы видели в § 6, стационарный поток без последействия однозначно определяется заданием функций $v_k(t)$ ($k=0, 1, 2, \ldots$) [причем всегда $v_0(t)=e^{-\lambda t}$]. Поэтому наша задача сводится к определению общего вида семейства функций $v_k(t)$ для стационарных потоков без последействия. С этой целью мы прежде всего установим, что для любого такого потока при любом $k>0$ отношение $v_k(t)/t$ при $t \to 0$ стремится к определенному пределу, который может быть либо нулем, либо положительным числом.

Обозначим через $\psi_k(t)$ ($k=0, 1, 2, \ldots$) вероятность того, что в промежутке длины t произойдет по меньшей мере k вызовов, так что для $k=0, 1, 2, \ldots$

$$\psi_k(t) = \sum_{i=k}^{\infty} v_i(t), \qquad v_k(t) = \psi_k(t) - \psi_{k+1}(t),$$

$$\psi_0(t) = 1, \qquad \psi_1(t) = w(t), \qquad \psi_2(t) = \psi(t).$$

В § 7 мы доказали, что для любого стационарного потока отношение $\psi_1(t)/t$ при $t \to 0$ стремится к определенному пределу, конечному или бесконечному. В случае потока без последействия этот результат является тривиальным, так как $\psi_1(t) = w(t) = 1 - e^{-\lambda t}$, откуда $\psi_1(t)/t \to \lambda$ при $t \to 0$. Но зато мы теперь убедимся, что в случае стационарного потока без последействия предел отношения $\psi_k(t)/t$ при $t \to 0$ существует для любого $k>0$. Так как $v_k(t) = \psi_k(t) - \psi_{k+1}(t)$, то отсюда будет следовать, что и предел отношения $v_k(t)/t$ при $t \to 0$ существует для любого $k>0$. В случае простейшего потока мы имеем, конечно, $v_1(t)/t \to \lambda$, $v_k(t)/t \to 0$ ($k>1$).

Подобно рассуждению § 7 мы начнем с доказательства одного элементарного вспомогательного предложения из теории пределов, представляющего собой некоторое усиление леммы § 7.

Лемма. *Пусть функция $f(x)$ — неотрицательная и неубывающая в отрезке $0 < x \leq a$, отношение $f(x)/x$ ограничено в этом отрезке и*

$$f(nx) \leq nf(x) + cn^2x^2, \tag{8.1}$$

где $c > 0$ — постоянная, n — любое натуральное число, и $0 < nx \leq a$; тогда отношение $f(x)/x$ при $x \to 0$ стремится к некоторому пределу $l \geq 0$.

Доказательство. Полагая в (8.1) $x = x_0/n$, где $0 < x_0 \leq a$, получаем

$$f(x_0) \leq nf\left(\frac{x_0}{n}\right) + cx_0^2,$$

откуда

$$\frac{f\left(\frac{x_0}{n}\right)}{\frac{x_0}{n}} \geq \frac{f(x_0)}{x_0} - cx_0. \tag{8.2}$$

Положим

$$\overline{\lim_{x \to 0}} \frac{f(x)}{x} = l;$$

мы можем принять $l > 0$, так как при $l = 0$ утверждение леммы тривиально.

Пусть $\varepsilon > 0$ задано произвольно. Выберем $x_0 < \frac{\varepsilon}{c}$ и так, чтобы

$$\frac{f(x_0)}{x_0} > l - \varepsilon.$$

Пусть $0 < x < x_0$ и натуральное число $n > 1$ определяется неравенствами

$$\frac{x_0}{n} \leq x < \frac{x_0}{n-1};$$

тогда в силу (8.2)

$$\frac{f(x)}{x} \geq \frac{f\left(\frac{x_0}{n}\right)}{\frac{x_0}{n-1}} = \frac{n-1}{n} \frac{f\left(\frac{x_0}{n}\right)}{\frac{x_0}{n}} \geq \frac{n-1}{n} \frac{f(x_0)}{x_0} - cx_0 \geq$$

$$\geq \left(1 - \frac{1}{n}\right)(l - \varepsilon) - \varepsilon > l - 3\varepsilon,$$

если x достаточно мало; так как, с другой стороны, $f(x)/x < l + \varepsilon$ для достаточно малых x, то

$$\frac{f(x)}{x} \to l \quad (x \to 0),$$

и лемма доказана.

Чтобы установить существование предела отношения $\psi_k(t)/t$ при $t \to 0$, нам надо только показать, что функция $\psi_k(t)$ при $k > 0$ удовлетворяет всем предпосылкам доказанной леммы. Неотрицательность и монотонность $\psi_k(t)$ в любом отрезке самоочевидны. Далее, из $\psi_k(t) \leqslant \psi_1(t) = w(t)$ и $w(t)/t \to \lambda$ $(t \to 0)$ вытекает ограниченность $\psi_k(t)/t$ в любом отрезке. Поэтому нам остается только убедиться, что $\psi_k(t)$ удовлетворяет соотношению (8.1).

Обозначим с этой целью через g_l верхнюю грань отношения $\psi_l(t)/t$ в области $0 < t < +\infty$ $(l \geqslant 1)$ и положим

$$A_k = \sum_{l=1}^{k-1} g_l g_{k-l}.$$

Мы утверждаем, что

$$\psi_k(nt) \leqslant n\psi_k(t) + A_k \frac{n(n-1)}{2} t^2, \tag{8.3}$$

откуда и будет следовать, что функция $\psi_k(t)$ удовлетворяет соотношению (8.1).

Неравенство (8.3) мы докажем с помощью индукции по n. При $n = 1$ оно тривиально. Допустим, что для некоторого n оно имеет место. Для того чтобы в отрезке $[0, (n+1)t]$ длины $(n+1)t = nt + t$ поступило не менее k вызовов [вероятность чего равна $\psi_k\{(n+1)t\}$], необходимо, чтобы при каком-либо l $(0 \leqslant l \leqslant k)$ имелось не менее l вызовов в отрезке $(0, t)$ и не менее $k - l$ вызовов в отрезке $[t, (n+1)t]$ (длины nt); поэтому

$$\psi_k[(n+1)t] \leqslant \sum_{l=0}^{k} \psi_l(t)\,\psi_{k-l}(nt) =$$

$$= \psi_0(t)\,\psi_k(nt) + \psi_k(t)\,\psi_0(nt) + \sum_{l=1}^{k-1} \psi_l(t)\,\psi_{k-l}(nt) \leqslant$$

$$\leqslant \psi_k(nt) + \psi_k(t) + nt^2 \sum_{l=1}^{k-1} g_l g_{k-l} =$$

$$= \psi_k(nt) + \psi_k(t) + A_k n t^2.$$

В силу (8.3) отсюда

$$\psi_k[(n+1)t] \leqslant (n+1)\psi_k(t) + A_k \frac{n(n+1)}{2} t^2,$$

а это и есть соотношение (8.3) с $n+1$ вместо n.

Таким образом, функция $\psi_k(t)$ при $k>0$ удовлетворяет всем предпосылкам доказанной леммы, и следовательно, при $t \longrightarrow 0$ отношение $\psi_k(t)/t$, а значит, и отношение $v_k(t)/t$ стремятся к определенному пределу. Так как $w(t)/t \longrightarrow \lambda > 0$ при $t \longrightarrow 0$, то и отношение $v_k(t)/w(t)$ при этом имеет определенный предел.

Положим

$$\lim_{t \to 0} [v_k(t)/w(t)] = p_k \qquad (k=1,\ 2,\ \ldots).$$

Отношение $v_k(t)/w(t)$ есть вероятность получить k вызовов в отрезке длины t, *если известно, что в этом отрезке вызовы существуют*. Предел этого отношения при $t \longrightarrow 0$, т. е. число p_k, можно поэтому рассматривать как вероятность получения k вызовов в определенный момент, если известно, что в этот момент вообще вызовы происходят (такое истолкование величины p_k возможно, но, конечно, не обязательно).

Перейдем теперь к определению общего вида функций $v_k(t)$ для стационарного потока без последействия. В § 3 мы установили для такого потока [см. (3.1)] общее соотношение

$$v_k(t+\tau) = \sum_{l=0}^{k} v_l(\tau)\, v_{k-l}(t) \quad (t>0,\ \tau>0,\ k=0,\ 1,\ 2,\ldots);$$

так как при $\tau \longrightarrow 0$

$$v_0(\tau) = e^{-\lambda\tau} = 1 - \lambda\tau + o(\tau),$$

то отсюда при $k>0$

$$v_k(t+\tau) = (1-\lambda\tau)\, v_k(t) + \sum_{l=1}^{k} v_l(\tau)\, v_{k-l}(t) + o(\tau),$$

и, следовательно,

$$\frac{v_k(t+\tau) - v_k(t)}{\tau} = -\lambda v_k(t) + \sum_{l=1}^{k} \frac{v_l(\tau)}{\tau} v_{k-l}(t) + o(1).$$

Но при $l>0$ и $\tau\to 0$ по доказанному выше

$$\frac{v_l(\tau)}{\tau}=\frac{v_l(\tau)}{w(\tau)}\frac{w(\tau)}{\tau}\to\lambda p_l;$$

поэтому предельный переход доказывает существование $v_k'(t)$ и дает

$$v_k'(t)=-\lambda v_k(t)+\lambda\sum_{l=1}^{k}p_l v_{k-l}(t)\qquad (k=1,\ 2,\ \ldots). \tag{8.4}$$

Добавляя сюда очевидное соотношение

$$v_0'(t)=-\lambda v_0(t), \tag{8.5}$$

мы получаем систему уравнений, позволяющих однозначно определить систему функций $v_k(t)$. В частности, к этому ведет путь замены неизвестных функций, которым мы пользовались в § 3. Полагая, как там,

$$v_k(t)=e^{-\lambda t}u_k(t)\qquad (k=0,\ 1,\ 2,\ \ldots),$$

мы легко приводим систему (8.4) к виду

$$u_k'(t)=\lambda\left[p_1u_{k-1}(t)+p_2u_{k-2}(t)+\ldots+p_ku_0(t)\right],$$

позволяющему рекуррентно определить все функции $u_k(t)$ [а значит, и $v_k(t)$]. Так, например, мы в силу $u_0(t)\equiv 1$ находим

$$u_1'(t)=\lambda p_1,$$

откуда

$$u_1(t)=\lambda p_1 t,\quad v_1(t)=e^{-\lambda t}\lambda p_1 t.$$

Мы не будем проводить здесь дальнейших выводов, так как значительно более простые и изящные результаты дает метод производящих функций, к применению которого мы теперь и переходим.

Положим

$$F(t,\ x)=\sum_{k=0}^{\infty}v_k(t)\,x^k;$$

отыскание системы функций $v_k(t)$ сводится к отысканию функции $F(t,\ x)$. Умножая соотношение (8.4) [для $k=0$ заменяемое соотношением (8.5)] на x^k и суммируя по k

от 0 до ∞, мы находим

$$\frac{\partial F}{\partial t} = -\lambda F + \lambda \sum_{k=1}^{\infty} x^k \sum_{l=1}^{k} p_l v_{k-l}(t) =$$

$$= -\lambda F + \lambda \sum_{l=1}^{\infty} p_l \sum_{q=0}^{\infty} v_q(t) x^{q+l} =$$

$$= -\lambda F + \lambda \sum_{l=1}^{\infty} p_l x^l \sum_{q=0}^{\infty} v_q(t) x^q,$$

или, полагая

$$\sum_{l=1}^{\infty} p_l x^l = \Phi(x),$$

получаем

$$\frac{\partial F}{\partial t} = \lambda [\Phi(x) - 1] F,$$

$$\frac{\partial \ln F}{\partial t} = \lambda [\Phi(x) - 1],$$

а так как при любом x

$$F(0, x) = v_0(0) = 1,$$

то интегрированием по t находим

$$F(t, x) = e^{\lambda [\Phi(x) - 1] t}, \tag{8.6}$$

и наша задача решена. Заметим, еще, что при любом t

$$F(t, 1) = \sum_{k=0}^{\infty} v_k(t) = 1,$$

вследствие чего (8.6) дает

$$\Phi(1) = \sum_{l=1}^{\infty} p_l = 1.$$

Таким образом, производящая функция $F(t, x)$ для любого стационарного потока без последействия имеет вид (8.6), где $\lambda > 0$ и $\Phi(x) = \sum_{l=1}^{\infty} p_l x^l$, $p_l \geqslant 0$, $\sum_{l=1}^{\infty} p_l = 1$.

Убедимся теперь, что и обратно, если числа λ и p_l ($l = 1, 2 \ldots$) подчиняются только что перечисленным требованиям, то существует стационарный поток без последействия, производящая функция которого дается формулой (8.6).

С этой целью мы допустим, что моменты времени, в которые происходят вызовы, образуют простейший поток с параметром λ, так что для вероятности $\overset{*}{v}_k(t)$ того, что за промежуток времени t произойдет k таких «вызывающих моментов», мы имеем обычное выражение

$$\overset{*}{v}_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Однако поток самих вызовов не будет, вообще говоря, простейшим, так как мы допустим, что в каждый вызывающий момент может с вероятностью, отличной от нуля, поступить и более одного вызова. Примем вероятность поступления в данный вызывающий момент ровно l вызовов равной p_l ($l = 1, 2, \ldots$), независимо от того, каков данный вызывающий момент и каково было течение этого потока до данного момента. Этим соглашением мы задаем некоторый определенный поток вызовов, который, очевидно, будет стационарным потоком без последействия. Покажем, что производящая функция $F(t, x)$ этого потока дается формулой (8.6).

Число вызовов, происходящих в любой вызывающий момент, мы определили как случайную величину, принимающую значение l с вероятностью p_l ($l = 1, 2, \ldots$); производящая функция такой величины есть

$$\sum_{l=1}^{\infty} p_l x^l = \Phi(x).$$

Возьмем теперь r любых различных между собой вызывающих моментов и обозначим через $\mathsf{P}_r(k)$ вероятность того, что в эти r моментов в совокупности произойдет k вызовов, так что суммарное число вызовов за r вызывающих моментов есть случайная величина с производящей функцией

$$\sum_{k=0}^{\infty} \mathsf{P}_r(k)\, x^k$$

[где, разумеется, $\mathsf{P}_r(k) = 0$ при $k < r$]. Но эта случайная величина есть сумма r взаимно независимых случайных величин, каждая из которых имеет производящую функцию $\Phi(x)$.

Так как при сложении взаимно независимых случайных величин их производящие функции перемножаются *), то поэтому

$$\sum_{k=0}^{\infty} \mathsf{P}_r(k)\, x^k = \{\Phi(x)\}^r.$$

А так как с другой стороны, очевидно, для рассматриваемого потока

$$v_k(t) = \sum_{r=0}^{\infty} v_r^*(t)\, \mathsf{P}_r(k),$$

то

$$F(t, x) = \sum_{k=0}^{\infty} v_k(t)\, x^k = \sum_{k=0}^{\infty} x^k \sum_{r=0}^{\infty} v_r^*(t)\, \mathsf{P}_r(k) =$$

$$= \sum_{r=0}^{\infty} v_r^*(t) \sum_{k=0}^{\infty} \mathsf{P}_r(k)\, x^k = \sum_{r=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^r}{r!} \{\Phi(x)\}^r =$$

$$= e^{-\lambda t} \sum_{r=0}^{\infty} \frac{\{\lambda t \Phi(x)\}^r}{r!} = e^{\lambda t\, [\Phi(x) - 1]},$$

что совпадает с формулой (8.6).

Мы можем формулировать результат нашего исследования так: совокупность всех стационарных потоков без последействия совпадает с совокупностью всех потоков, даваемых формулой (8.6), где $\lambda > 0$:

$$\Phi(x) = \sum_{l=1}^{\infty} p_l x^l, \quad p_l \geqslant 0 \quad (l = 1, 2, \ldots), \quad \sum_{l=1}^{\infty} p_l = 1.$$

С предметной точки зрения мы убедились, что для каждого стационарного потока без последействия поток вызывающих моментов является простейшим и для полного описания данного потока вызовов надо, кроме параметра λ этого простейшего потока, задать еще закон распределения $(p_1, p_2, \ldots, p_n, \ldots)$ числа вызовов, поступающих в любой

*) Это непосредственно вытекает из того, что для случайной величины ξ с законом распределения $\mathsf{P}\{\xi = n\} = q_n$ $(n = 0, 1, 2, \ldots)$ производящая функция $f(x) = \sum_{n=0}^{\infty} q_n x^n$ есть, очевидно, математическое ожидание величины x^{ξ}.

выбранный вызывающий момент. Очевидно, эти соображения делают совершенно прозрачной структуру самого общего стационарного потока без последействия.

Заметим еще, что в случае $p_1=1$, $p_k=0$ $(k>1)$ формула (8.6) дает нам производящую функцию простейшего потока с параметром λ:

$$F(t,\ x)=e^{\lambda t(x-1)}=\sum_{k=0}^{\infty}\left\{e^{-\lambda t}\frac{(\lambda t)^k}{k!}\right\}x^k.$$

Глава 3

ФУНКЦИИ ПАЛЬМА

§ 9. Определение и доказательство существования

После того как в конце предыдущей главы мы полностью выяснили строение стационарных потоков без последействия, мы должны теперь обратиться к исследованию потоков более общего типа. Для довольно широких классов таких потоков весьма удобным орудием исследования оказалась одна функция, введенная Пальмом [8] и примененная им с успехом к решению ряда задач. Пальм определяет эту функцию $\varphi_0(t)$ (для любого стационарного потока) как *условную вероятность отсутствия вызовов в промежутке* $(t_0,\ t_0+t)$, *если известно, что в момент* t_0 *произошел вызов*. Однако такое определение вряд ли можно считать достаточно удобным; то условие, при котором должна быть вычислена вероятность $\varphi_0(t)$, т. е. наличие вызова в некоторый момент t_0, само имеет во всех актуальных случаях вероятность 0, и это обстоятельство, как известно, не позволяет нам непосредственно определить функцию $\varphi_0(t)$ для заданного потока с помощью известных правил расчета условных вероятностей. Поэтому мы дадим этой функции другое, более сложное определение, которое позволит однозначно определять ее для любого стационарного потока. Вместе с тем мы определим не одну функцию, а целую последовательность $\varphi_k(t)$ $(k=0,\ 1,\ 2,\ \ldots)$ функций, которые будем называть функциями Пальма и которые в дальнейшем окажутся нам полезными при решении ряда важных задач.

Пусть мы имеем два последовательных промежутка времени, из которых первый имеет длину τ, а второй — t (в дальнейшем мы будем для краткости называть самые эти промежутки соответственно «промежутком τ» и «промежутком t»). Обозначим для данного стационарного потока через $H_k(\tau, t)$ ($k \geqslant 0$) вероятность следующего двойного события: 1) в промежутке τ произойдет по меньшей мере один вызов; 2) в промежутке t произойдет не более k вызовов. Эти два события, вообще говоря, будут взаимно зависимы; так как вероятность события 1) в наших старых обозначениях есть $w(\tau)$, то отношение

$$\frac{H_k(\tau, t)}{w(\tau)} \tag{9.1}$$

выражает собою условную вероятность события 2) при условии, что имело место событие 1), т. е. вероятность появления не более k вызовов в промежутке t при условии, что в промежутке τ появился по меньшей мере один вызов.

Если это отношение при $\tau \to 0$ (и при постоянном t) стремится к некоторому пределу, то этот предел естественно называть условной вероятностью появления не более чем k вызовов в промежутке t *при условии, что в начальный момент этого промежутка произошел вызов.*

Убедимся теперь, что предел отношения (9.1) при $\tau \to 0$ (и постоянном t) всегда существует, если только данный стационарный поток имеет конечный параметр λ. С этой целью рассмотрим сначала отношение $H_k(\tau, t)/t$. Чтобы доказать существование предела этого отношения при $\tau \to 0$, достаточно убедиться, что величина $H_k(\tau, t)$ как функция от τ удовлетворяет всем предпосылкам леммы § 7. Неотрицательность и монотонность этой функции самоочевидны. Пусть $\tau = \tau_1 + \tau_2$ и промежуток τ_1 предшествует промежутку τ_2. Тогда, если выполнено то двойное событие, вероятность которого мы обозначили $H_k(\tau, t)$, то, очевидно, выполняется по меньшей мере одно из следующих двух событий:

(A) В промежутке τ_2 имеется по меньшей мере один вызов, в промежутке t имеется не более k вызовов [вероятность события (A) равна $H_k(\tau_2, t)$].

(B) В промежутке τ_1 имеется по меньшей мере один вызов, в промежутке $\tau_2 + t$ имеется не более k вызовов [вероятность события (B) равна $H_k(\tau_1, \tau_2+t) \leqslant H_k(\tau_1, t)$ (так как при фиксированном τ, $H_k(\tau, t)$, очевидно, есть невозрастающая функция от t)]. Таким образом, мы находим

$$H_k(\tau_1+\tau_2, t) \leqslant H_k(\tau_1, t)+H_k(\tau_2, t),$$

т. е. функция $H_k(\tau, t)$ удовлетворяет (относительно τ) и последней предпосылке леммы § 7. Применяя эту лемму, мы находим, что отношение $H_k(\tau, t)/\tau$ при $\tau \to 0$ стремится к некоторому пределу или безгранично возрастает; однако последний случай исключается, так как, очевидно, всегда $H_k(\tau, t) \leqslant w(\tau)$, а отношение $w(\tau)/\tau$ по нашему предположению стремится к конечному пределу λ.

Наконец,

$$\frac{H_k(\tau, t)}{w(\tau)} = \frac{H_k(\tau, t)/\tau}{w(\tau)/\tau};$$

числитель и знаменатель этой дроби по доказанному стремятся при $\tau \to 0$ к определенным пределам; поэтому и

$$\lim_{\tau \to 0} \frac{H_k(\tau, t)}{w(\tau)} = \Phi_k(t) \tag{9.2}$$

существует; разумеется, этот предел является функцией от t.

Положим теперь

$$h_0(\tau, t) = H_0(\tau, t), \quad h_k(\tau, t) = H_k(\tau, t) - H_{k-1}(\tau, t) \quad (k > 0);$$

очевидно, $h_k(\tau, t)$ есть вероятность того, что 1) в промежутке τ имеется по меньшей мере один вызов и 2) в промежутке t имеется ровно k вызовов; отношение $h_k(\tau, t)/w(\tau)$ представляет собой условную вероятность иметь k вызовов в промежутке t при условии, что в промежутке τ имеется по меньшей мере один вызов. Из (9.2) следует

$$\lim_{\tau \to 0} \frac{h_0(\tau, t)}{w(\tau)} = \Phi_0(t), \quad \lim_{\tau \to 0} \frac{h_k(\tau, t)}{w(\tau)} = \Phi_k(t) - \Phi_{k-1}(t) \quad (k > 0).$$

Полагая

$$\varphi_0(t) = \Phi_0(t), \quad \varphi_k(t) = \Phi_k(t) - \Phi_{k-1}(t) \quad (k > 0).$$

мы имеем

$$\lim_{\tau \to 0} \frac{h_k(\tau, t)}{w(\tau)} = \varphi_k(t) \quad (k = 0, 1, 2, \ldots).$$

Функции $\varphi_k(t)$ мы и будем называть **функциями Пальма**. Функция $\varphi_k(t)$ может быть понимаема как *вероятность иметь k вызовов в промежутке длины t при условии, что в начальный момент этого промежутка произошел вызов*. Этим она отличается от функции $v_k(t)$, представляющей собою вероятность того же события при условии, что относительно начального момента ничего неизвестно. Проведенное нами рассуждение показывает, что вся совокупность функций Пальма однозначно определяется для любого стационарного потока с конечным параметром λ.

Заметим еще, что так как $H_k(\tau, t)$ относительно t есть функция невозрастающая, а $w(\tau)$ от t не зависит, то все функции $\Phi_k(t)$ [в частности, функция $\Phi_0(t) = \varphi_0(t)$] — невозрастающие в области $0 < t < +\infty$.

§ 10. Формулы Пальма

Функции Пальма связаны с основными функциями $v_k(t)$ данного стационарного потока простыми и важными формулами, которые нам предстоит теперь вывести.

Допустим, что данный стационарный поток — ординарный и имеет конечный параметр λ. Рассмотрим снова промежуток времени длины $\tau + t$, составленный из «промежутка τ» и непосредственно следующего за ним «промежутка t». Обозначим через n_1 и n_2 соответственно числа вызовов в промежутках τ и t (n_1, n_2 — случайные величины). Мы имеем, очевидно,

$$v_k(\tau + t) = \mathsf{P}\{n_1 + n_2 = k\} = \sum_{r=0}^{k} \mathsf{P}\{n_1 = r,\ n_2 = k - r\},$$

откуда в силу ординарности данного потока при $\tau \to 0$

$$v_k(\tau + t) = \mathsf{P}\{n_1 = 0,\ n_2 = k\} + \mathsf{P}\{n_1 = 1,\ n_2 = k - 1\} + o(\tau). \qquad (10.1)$$

Но

$$\mathsf{P}\{n_1 = 0,\ n_2 = k\} = \mathsf{P}\{n_2 = k\} - \mathsf{P}\{n_1 > 0,\ n_2 = k\} = v_k(t) - h_k(\tau, t), \qquad (10.2)$$

где мы пользуемся обозначениями § 9. С другой стороны,

$$\mathsf{P}\{n_1 = 1,\ n_2 = k - 1\} = \mathsf{P}\{n_1 > 0,\ n_2 = k - 1\} - \mathsf{P}\{n_1 > 1,\ n_2 = k - 1\} = h_{k-1}(\tau, t) + o(\tau). \qquad (10.3)$$

Вставляя (10.2) и (10.3) в (10.1), находим

$$v_k(t+\tau) = v_k(t) - h_k(\tau, t) + h_{k-1}(\tau, t) + o(\tau),$$

откуда

$$\frac{v_k(\tau+t) - v_k(t)}{\tau} = \frac{h_{k-1}(\tau, t)}{w(\tau)} \frac{w(\tau)}{\tau} - \frac{h_k(\tau, t)}{w(\tau)} \frac{w(\tau)}{\tau} + o(1).$$

В силу результатов § 9 отсюда вытекает дифференцируемость функции $v_k(t)$ и соотношение

$$v_k'(t) = \lambda[\varphi_{k-1}(t) - \varphi_k(t)] \quad (k > 0); \tag{10.4}$$

при $k=0$ это же рассуждение дает

$$v_0'(t) = -\lambda\varphi_0(t), \tag{10.5}$$

так что соотношение (10.4) имеет место и при $k=0$, если положить $\varphi_{-1}(t) \equiv 0$. Складывая соотношения (10.4) для $k = 0, 1, \ldots, m$ и обозначая через

$$V_m(t) = \sum_{k=0}^{m} v_k(t)$$

вероятность иметь в промежутке длины t *не более* m вызовов, мы находим

$$V_m'(t) = -\lambda\varphi_m(t) \quad (m = 0, 1, 2, \ldots). \tag{10.6}$$

Формулы (10.4) и (10.6) и были целью нашего вывода [у Пальма имеется лишь формула (10.5)]. Эти формулы иногда удобнее применять в интегральной форме. Интегрируя обе части (10.6) от 0 до t мы находим

$$V_m(+0) - V_m(t) = \lambda \int_0^t \varphi_m(u)\, du \quad (m = 0, 1, 2, \ldots),$$

но

$$V_m(+0) = 1;$$

в самом деле, мы имеем

$$V_m(+0) \geqslant V_0(+0) = v_0(+0) = 1 - w(+0);$$

а так как, при $t\to 0$, $w(t)/t\to\lambda$, то $w(+0)=0$. Итак, мы находим для любого $t>0$

$$1-V_m(t)=\lambda\int_0^t\varphi_m(u)\,du\quad(m=0,\,1,\,2,\,\ldots),\qquad(10.7)$$

откуда легко вытекает

$$\left.\begin{aligned}v_0(t)&=1-\lambda\int_0^t\varphi_0(u)\,du;\\ v_k(t)&=\lambda\int_0^t[\varphi_{k-1}(u)-\varphi_k(u)]\,du\quad(k=1,\,2,\ldots).\end{aligned}\right\}\qquad(10.8)$$

Эти формулы просто и непосредственно выражают функции $v_k(t)$ данного потока через функции Пальма $\varphi_k(t)$.

§ 11. Интенсивность стационарного потока. Теорема Королюка

В § 4 мы условились называть *интенсивностью* μ данного стационарного потока математическое ожидание числа вызовов в единицу времени; в силу аддитивности математических ожиданий мы имеем тогда, что математическое ожидание числа вызовов в промежутке длины t для стационарного потока пропорционально t, т. е.

$$\sum_{k=0}^{\infty}kv_k(t)=\mu t.$$

Там же мы убедились, что всегда $\mu\geqslant\lambda$, а для простейшего потока $\mu=\lambda$.

В работах прикладного характера совпадение параметров μ и λ обычно принимается как самоочевидный факт, не требующий даже оговорки, при исследовании стационарных потоков самого общего типа. Ввиду практического значения этого допущения представляется важным разобраться в его предпосылках и дать ему строгое обоснование там, где это возможно.

Остановимся сначала на случае стационарного потока без последействия. В § 8 мы видели, что для потоков этого рода производящая функция

$$F(t, x)=\sum_{k=0}^{\infty} v_k(t)\, x^k$$

имеет вид

$$e^{\lambda t[\Phi(x)-1]},$$

где $\lambda>0$, $\Phi(x)=\sum_{l=1}^{\infty} p_l x^l$, $p_l \geqslant 0$ $(l=1, 2, \ldots)$, $\sum_{l=1}^{\infty} p_l=1$.

Так как, очевидно,

$$\mu t=\sum_{k=0}^{\infty} k v_k(t)=\left\{\frac{\partial F(t, x)}{\partial x}\right\}_{x=1}$$

и так как

$$\frac{\partial F(t, x)}{\partial x}=F(t, x)\lambda t\,\Phi'(x), \quad F(t, 1)=\sum_{k=0}^{\infty} v_k(t)=1,$$

то

$$\mu t=\lambda t\,\Phi'(1)=\lambda t\,[p_1+2p_2+3p_3+\ldots],$$

откуда

$$\mu=\lambda[p_1+2p_2+3p_3+\ldots].$$

Так как

$$\sum_{l=1}^{\infty} p_l=1$$

и

$$p_l \geqslant 0 \quad (l \geqslant 1),$$

то мы непосредственно видим, что для равенства $\mu=\lambda$ необходимо и достаточно иметь $p_1=1$. Но при $p_1=1$ данный поток, как мы видели в конце § 8, является простейшим. Таким образом, *среди стационарных потоков без последействия только простейшие потоки удовлетворяют требованию* $\mu=\lambda$; *для всех остальных* $\mu>\lambda$.

Так как для стационарного потока без последействия ординарность есть необходимое и достаточное условие для того, чтобы этот поток был простейшим, то можно еще сказать, что *для стационарного потока без последействия, интенсивность которого конечна, необходимым и достаточным* *условием равенства* $\mu=\lambda$ *является ординарность этого потока* *).

Выведенные нами в § 10 формулы Пальма позволяют, как это показал В. С. Королюк, легко убедиться, что для любого стационарного потока ординарность влечет за собой равенство $\mu=\lambda$ (причем не исключается случай $\mu=\lambda=+\infty$).

В самом деле, мы имеем

$$\mu=\sum_{k=1}^{\infty} kv_k(1)=\sum_{k=1}^{\infty}[v_k(1)+v_{k+1}(1)+v_{k+2}(1)+\ldots]=$$
$$=\sum_{k=1}^{\infty}[1-V_{k-1}(1)]=\sum_{k=0}^{\infty}[1-V_k(1)],$$

откуда в силу формулы (10.7)

$$\mu=\lambda\sum_{k=0}^{\infty}\int_0^1 \varphi_k(u)\,du. \tag{11.1}$$

Но

$$\varphi_k(u)=\lim_{\tau\to 0}\frac{h_k(\tau,u)}{w(\tau)}, \tag{11.2}$$

а так как отношение $h_k(\tau, u)/w(\tau)$ есть условная вероятность иметь в промежутке u ровно k вызовов (при условии наличия вызовов в промежутке τ), то при любом $m>0$

$$\sum_{k=0}^{m}\frac{h_k(\tau,u)}{w(\tau)}\leqslant 1 \quad (0<u\leqslant 1),$$

а потому в силу (11.2)

$$\sum_{k=0}^{m}\varphi_k(u)\leqslant 1 \quad (0<u\leqslant 1);$$

---

*) Совсем простое доказательство этого предложения и пример, в силу которого для потоков с бесконечной интенсивностью из равенства $\mu=\lambda$ ординарность потока не вытекает, были даны Ф. Зитеком в работе «Заметка об одной теореме Королюка», Чехословац. матем. журнал, т. 7 (82), 1957, 318—319.— *Б. Г.*

следовательно, при любом $m>0$

$$\sum_{k=0}^{m}\int_{0}^{1}\varphi_k(u)\,du=\int_{0}^{1}\left\{\sum_{k=0}^{m}\varphi_k(u)\right\}du\leqslant 1;$$

а значит, и

$$\sum_{k=0}^{\infty}\int_{0}^{1}\varphi_k(u)\,du\leqslant 1,$$

и (11.1) дает $\mu\leqslant\lambda$; а так как еще в § 4 мы видели, что всегда $\mu\geqslant\lambda$, то $\mu=\lambda$, и наше утверждение доказано.

Глава 4

ПОТОКИ С ОГРАНИЧЕННЫМ ПОСЛЕДЕЙСТВИЕМ

§ 12. Другой способ описания потока

Способ задания потока вызовов, описанный нами в § 6, исходит из понимания потока как случайного процесса $x(t)$ и ничем не отличается от общего способа описания произвольного случайного процесса. Это автоматическое включение теории потоков в общую теорию случайных процессов, несомненно, имеет свои преимущества, так как позволяет применять к изучению потоков методы и результаты общей теории случайных процессов; однако, учитывая специфические свойства наших потоков как случайных процессов [и в первую очередь то, что величина $x(t)$ всегда монотонна и принимает лишь целые неотрицательные значения], мы можем надеяться найти для этих потоков хоть и менее общий, но зато более простой и удобный способ описания. Этим вопросом мы теперь и займемся.

Пусть снова начальный момент данного потока есть $t_0=0$; пусть $t_i\,(i=1,2,\ldots)$ есть момент i-го вызова, так что $t_{i-1}\leqslant t_i\,(i=1,2,\ldots)$; положим, наконец,

$$t_i-t_{i-1}=z_i\,(i=1,2,\ldots),$$

так что $z_1=t_1$, а z_i при $i>1$ означает величину промежутка времени между $(i-1)$-м и i-м вызовами. Очевидно, все t_i и z_i $(i=1,2,\ldots)$ представляют собой случайные величины, способные принимать лишь неотрицательные значения.

Условимся теперь считать поток заданным, если для любого $n>0$ задан n-мерный закон распределения вектора $(z_1, z_2, \ldots, z_n)$. Очевидно, этот способ описания потока более элементарен, чем выбранный нами в § 6, так как в основе его лежит представление о потоке не как о случайном процессе общего вида, а как о последовательности случайных величин. Убедимся теперь, что оба способа задания потока равносильны, т. е. что поток, заданный с помощью какого-либо одного из этих двух способов, будет тем самым однозначно описанным и в смысле другого способа.

Пусть, как в § 6, $x(t)$ означает число вызовов, предшествующих моменту t; очевидно, неравенства $t_k<u$ и $x(u)\geqslant k$ выражают собой одно и то же событие; то же самое имеет место и для неравенств $t_k\geqslant u$, $x(u)<k$, а значит, и для неравенств $u\leqslant t_k<v$, $x(u)<k\leqslant x(v)$. Отсюда далее следует, что система неравенств

$$u_k\leqslant t_k<v_k\,(1\leqslant k\leqslant n) \tag{12.1}$$

выражает то же событие, что и система неравенств

$$x(u_k)<k\leqslant x(v_k) \quad (1\leqslant k\leqslant n), \tag{12.2}$$

каковы бы ни были вещественные числа u_k, v_k $(1\leqslant k\leqslant n)$.

Если поток вызовов задан в смысле § 6, то при любом n и при любых u_k, v_k $(1\leqslant k\leqslant n)$ нам задан закон распределения $2n$-мерного вектора $[x(u_k),\ x(v_k);\ k=1, 2, \ldots, n]$ и, значит, однозначно определена вероятность системы (12.2), а следовательно, и равносильной ей системы (12.1). Но ввиду произвольности чисел u_k, v_k последнее означает, что однозначно задан закон распределения вектора $(t_1, t_2, \ldots, t_n)$, а значит, и вектора

$$(z_1, z_2, \ldots, z_n)=(t_1, t_2-t_1, \ldots, t_n-t_{n-1});$$

таким образом, данный поток однозначно определен и в новом смысле.

Обратно, если при любом n нам задан закон распределения вектора $(z_1, z_2, \ldots, z_n)$, то в силу

$$t_k=\sum_{i=1}^{k} z_i \quad (1\leqslant k\leqslant n)$$

тем самым однозначно определен и закон распределения вектора $(t_1, t_2, \ldots, t_n)$. Но при любых $u_1, u_2, \ldots, u_n$ и любых целых $k_1, k_2, \ldots, k_n$ система неравенств

$$t_{k_i} < u_i, \quad t_{k_{i+1}} \geqslant u_i \quad (1 \leqslant i \leqslant n) \tag{12.3}$$

равносильна системе равенств

$$x(u_i) = k_i \quad (1 \leqslant i \leqslant n); \tag{12.4}$$

так как вероятность системы (12.3) однозначно определена, то то же имеет место и для системы (12.4). А это означает, что при любом n и любых u_i $(1 \leqslant i \leqslant n)$ однозначно определен закон распределения вектора $[x(u_1), x(u_2), \ldots, x(u_n)]$, т. е. что наш процесс однозначно определен в смысле § 6.

Таким образом, указанный нами новый способ задания потока вызовов действительно равносилен принятому в § 6.

§ 13. Потоки с ограниченным последействием

Если данный поток — без последействия, то величины $z_1, z_2, \ldots, z_n$, очевидно, взаимно независимы. Однако обратное заключение, как мы узнаем в дальнейшем, было бы неверным. Взаимная независимость величин z_k в значительной степени ограничивает явление последействия, но не исключает его полностью. В части II мы узнаем, что как раз потоки с взаимно независимыми z_k, но с наличием последействия играют важнейшую роль в теории обслуживания вызовов полнодоступным пучком линий. Мы должны поэтому заняться теперь установлением некоторых основных свойств таких потоков.

Условимся (следуя Пальму) называть *потоком с ограниченным последействием* всякий поток, у которого $z_1, z_2, \ldots, z_n, \ldots$ есть последовательность взаимно независимых случайных величин*). Очевидно, для однозначного описания такого потока достаточно задать законы распределения всех величин z_k $(k = 1, 2, \ldots)$. Мы будем в дальнейшем обозначать эти законы через $F_k(x)$ $(k = 1, 2, \ldots)$.

Мы знаем, что если данный поток — стационарный и ординарный, то полное отсутствие последействия влечет за собой простейший характер потока, подробно изученный нами

*) В другом плане потоки с ограниченным последействием изучаются в теории восстановления, см. работу В. Смита «Теория восстановления и смежные с ней вопросы», сб. переводов «Математика», т. 5, вып. 3, 1961, 95—150.— *Б. Г.*

в главе 1. Поэтому стационарный и ординарный поток с ограниченным последействием мы можем рассматривать как некоторое обобщение простейшего потока*). Именно такого рода потоки представляют значительный интерес для теории обслуживания в случае систем с потерями (см. далее гл. 8). Стационарный ординарный поток с ограниченным последействием мы будем для краткости называть **потоком типа** P (или потоком Пальма).

В § 9 мы ввели для любого стационарного потока систему «функций Пальма» $\varphi_k(t)$ $(k=0, 1, 2, \ldots)$. Функция $\varphi_0(t)$, как мы теперь увидим, играет основную роль в теории потоков типа P. *Заданием этой функции поток типа P однозначно определяется.* В самом деле, так как в случае потока типа P величины $z_1, z_2, \ldots, z_n, \ldots$ между собою независимы, то для однозначного определения закона распределения каждого вектора $(z_1, z_2, \ldots, z_n)$ $(n=1, 2, \ldots)$, а значит, и для однозначного описания потока достаточно задать законы распределения $F_k(x)$ величин z_k $(k=1, 2, \ldots)$. Но эти законы однозначно определяются заданием функции Пальма $\varphi_0(t)$, как показывает следующее предложение.

Теорема. *Для потока типа* P

$$F_1(x)=\lambda\int_0^x \varphi_0(u)\,du, \quad F_k(x)=1-\varphi_0(x) \quad (k\geqslant 2).$$

Для лучшей обозримости мы разобьем доказательство на несколько этапов.

1. Так как $F_1(x)=\mathsf{P}\{t_1<x\}$ есть вероятность наличия вызовов в промежутке $(0, x)$ и, следовательно, в наших старых обозначениях равна

$$w(x)=1-v_0(x)=1-V_0(x),$$

то утверждаемое выражение для $F_1(x)$ непосредственно вытекает из формулы (10.7) при $m=0$; при этом параметр λ

---

*) Интересно заметить, что ограниченность последействия следует из отсутствия последействия лишь для ординарных потоков. Неординарный поток без последействия может не быть потоком с ограниченным последействием. Примером может служить поток, рассмотренный нами в конце § 8 при $p_1=p_2=\frac{1}{2}$. (Этим замечанием автор обязан П. И. Васильеву.)

данного потока определяется через функцию $\varphi_0(t)$ с помощью соотношения

$$\lambda\int_0^\infty \varphi_0(u)\,du = F_1(+\infty) = 1.$$

Нам остается, таким образом, рассмотреть случай $k>1$.

2. Положим, как в § 8,

$$\psi_k(t) = 1 - V_{k-1}(t) = v_k(t) + v_{k+1}(t) + \ldots$$

Тогда имеет место

Лемма. (См. примечание на стр. 68 — *Б. Г.*) *Для любого потока типа P и любого* $r>0$

$$\frac{\psi_{r+1}(u)}{\psi_r(u)} \to 0 \quad (u\to 0).$$

Доказательство. Обозначая через t_k момент k-го вызова и полагая, как прежде, $t_k - t_{k-1} = z_k$ $(k=1, 2, \ldots)$, мы очевидно имеем

$$\psi_{r+1}(u) = \mathsf{P}\{t_{r+1}<u\} \leqslant \mathsf{P}\{t_r<u,\ z_{r+1}<u\},$$

и следовательно, в силу независимости z_{r+1} от t_r

$$\psi_{r+1}(u) \leqslant \psi_r(u)\,F_{r+1}(u).$$

Для доказательства леммы достаточно поэтому убедиться что $F_{r+1}(u)\to 0$ при $u\to 0$.

Пусть $a>0$ столь велико, что $\psi_r(a)>0$. Пусть $x>0$ произвольно мало и n таково, что $(n-1)\,x<a\leqslant nx$. Условимся называть «ячейками» отрезки $[(k-1)\,x, kx]$ $(1\leqslant k\leqslant n)$. Если $t_r<nx$ и $z_{r+1}<x$, то моменты t_r и t_{r+1} лежат либо в одной ячейке, либо в двух соседних ячейках, так что по меньшей мере один из n отрезков

$$[(l-1)\,x, (l+1)\,x] \quad (l=1, 2, \ldots, n)$$

длины $2x$ содержит более одного вызова. Поэтому

$$\mathsf{P}\{t_r<nx,\ z_{r+1}<x\} = \psi_r(nx)\,F_{r+1}(x) \leqslant n\psi_2(2x) = \\ = 2nx\,\frac{\psi_2(2x)}{2x} < 2\,(a+x)\,\frac{\psi_2(2x)}{2x},$$

откуда при $x<a$

$$F_{r+1}(x) \leqslant \frac{2\,(a+x)}{\psi_r(nx)}\,\frac{\psi_2(2x)}{2x} \leqslant \frac{4a}{\psi_r(a)}\,\frac{\psi_2(2x)}{2x} \to 0$$

при $x \to 0$ в силу ординарности данного потока. Этим завершено доказательство нашей леммы.

3. Переходя теперь к доказательству теоремы, убедимся прежде всего, что $F_2(t) = 1 - \varphi_0(t)$. С этой целью рассмотрим введенную нами в главе 3 вероятность $h_0(\tau, t)$ того, что в некотором промежутке длины τ вызовы имеются, а в последующем за ним промежутке длины t вызовов нет. Если число вызовов в промежутке τ равно k, то из отсутствия вызовов в промежутке t следует $z_{k+1} > t$, вероятность чего есть $1 - F_{k+1}(t)$; поэтому

$$h_0(\tau, t) \leqslant \sum_{k=1}^{\infty} v_k(\tau)[1 - F_{k+1}(t)] \leqslant v_1(\tau)[1 - F_2(t)] + \psi_2(\tau).$$

С другой стороны, при том же условии (k вызовов в промежутке τ) из $z_{k+1} > t + \tau$ следует, что в промежутке t вызовов нет, поэтому

$$h_0(\tau, t) \geqslant \sum_{k=1}^{\infty} v_k(\tau)[1 - F_{k+1}(t+\tau)] \geqslant v_1(\tau)[1 - F_2(t+\tau)].$$

Таким образом,

$$v_1(\tau)[1 - F_2(t+\tau)] \leqslant h_0(\tau, t) \leqslant v_1(\tau)[1 - F_2(t)] + \psi_2(\tau).$$

Деля все части этих неравенств на $w(\tau)$ и замечая, что при $\tau \to 0$

$$\frac{v_1(\tau)}{w(\tau)} \to 1, \quad \frac{\psi_2(\tau)}{w(\tau)} \to 0, \quad \frac{h_0(\tau, t)}{w(\tau)} \to \varphi_0(t),$$

мы в пределе находим

$$1 - F_2(t+0) \leqslant \varphi_0(t) \leqslant 1 - F_2(t),$$

откуда $F_2(t) = 1 - \varphi_0(t)$ во всех точках непрерывности закона распределения $F_2(t)$.

4. Теперь мы с помощью индукции убедимся, что $F_{r+2}(t) = 1 - \varphi_0(t)$ для всех $r \geqslant 0$.

В силу стационарности данного потока закон распределения расстояния между двумя первыми вызовами, следующими за каким-либо моментом $a > 0$, совпадает с законом распределения $F_2(t)$ расстояния z_2 между первыми двумя вызовами, следующими за моментом 0. Но если в промежутке $(0, a)$ имеется r вызовов, то расстояние между первыми двумя вызовами, следующими за моментом a, есть z_{r+2}, и закон распределения его, равный $F_{r+2}(t)$, не зависит от предшествующего течения потока. Таким образом,

$$F_2(t) = \sum_{r=0}^{\infty} v_r(a) F_{r+2}(t). \tag{13.1}$$

Пусть теперь уже установлено, что

$$F_2(t) = F_3(t) = \cdots = F_{r+1}(t) = 1 - \varphi_0(t).$$

Тогда (13.1) дает

$$1 - \varphi_0(t) = [1 - \varphi_0(t)] \sum_{k=0}^{r-1} v_k(a) + v_r(a) F_{r+2}(t) + \\ + \sum_{k>r} v_k(a) F_{k+2}(t),$$

откуда в силу $v_r(a) = \psi_r(a) - \psi_{r+1}(a)$

$$[1 - \varphi_0(t)] \psi_r(a) - F_{r+2}(t) \psi_r(a) = \\ = -\psi_{r+1}(a) F_{r+2}(t) + \sum_{k>r} v_k(a) F_{k+2}(t).$$

Так как последняя сумма правой части не превосходит

$$\sum_{k>r} v_k(a) = \psi_{r+1}(a),$$

то мы получаем

$$\psi_r(a) \left| 1 - \varphi_0(t) - F_{r+2}(t) \right| \leqslant \psi_{r+1}(a);$$

$$\left| 1 - \varphi_0(t) - F_{r+2}(t) \right| \leqslant \frac{\psi_{r+1}(a)}{\psi_r(a)}.$$

Это неравенство имеет место при любых $t > 0$ и $a > 0$; но при $a \to 0$ правая часть по доказанной лемме стремится к нулю; а так как левая часть от a не зависит, то при любом $t > 0$

$$F_{r+2}(t) = 1 - \varphi_0(t),$$

что и требовалось доказать.

Глава 5

ПРЕДЕЛЬНАЯ ТЕОРЕМА

§ 14. Постановка задачи. Теорема Пальма

Как мы уже говорили в главе 1, значительное большинство исследований прикладного характера основывается на предположении, что первичный поток поступающих на данную установку вызовов является простейшим потоком. Однако давно уже известен целый ряд принципиальных соображений, заставляющих сомневаться в том, что предпосылки, которые представляют собой определение простейшего потока, с достаточной степенью точности выполняются в большинстве практически встречающихся случаев (в особенности это относится к требованию отсутствия последействия). Если поэтому наблюдения и опыт констатируют некоторое небольшое отклонение реально встречающихся потоков от простейших, то этому не следует удивляться; более того, удивление может вызвать тот факт, что отклонения такого рода в большинстве случаев бывают менее значительными, чем этого можно было бы ожидать из теоретических соображений. Таким образом, если обычно при сопоставлении выводов теории с опытными данными перед исследователем встает задача — объяснить причины отклонения реально протекающих явлений от теоретически предсказанного их течения, то в данном случае дело обстоит как раз наоборот: опытные данные согласуются с выводами построенной теории, как правило, лучше, чем этого можно было бы ожидать по принципиальным соображениям, и именно это «слишком хорошее» согласие требует объяснения.

Пальмом [8] сделана заслуживающая внимания попытка объяснения фактов этого рода, исходя из предположения, что данный поток представляет собой простую сумму (суперпозицию) большого числа взаимно независимых потоков малой интенсивности, причем каждый из слагаемых потоков является стационарным и ординарным, в отношении же последействия эти потоки могут вести себя произвольным образом. При этом оказывается, что в весьма широких предположениях суммарный поток по своему характеру должен быть близок к простейшему. Такая постановка задачи, по-видимому, во многих случаях близка к реальной ситуации. Так, есл к данной установке прикреплено большое число абонентов, то общий поток вызовов слагается из потоков (сравнительно весьма малой интенсивности), исходящих от отдельных абонентов, причем эти слагаемые потоки можно в первом приближении считать стационарными, ординарными и взаимно независимыми.

Мы приходим на этом пути к ряду своеобразных предельных теорем, которые способны в значительной степени объяснить исследуемое явление. Этим вопросом мы и займемся в настоящей главе.

Пусть исследуемый поток представляет собой суперпозицию n стационарных, ординарных и взаимно независимых потоков. Обозначим через λ_r интенсивность r-го потока, через $\varphi_r(t)$ его функцию Пальма [которую в гл. 3 мы обозначали через $\varphi_0(t)$] и через $v_{kr}(t)$ — вероятность поступления в промежутке $(0, t)$ k вызовов r-го потока. Те же величины для суммарного потока обозначим соответственно через Λ, $\Phi(t)$ и $V_k(t)$ (так что в частности $\Lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$). Мы будем исходить из следующих предпосылок:

1°. При $n \to \infty$, Λ остается постоянным, в то время как числа $\lambda_1, \lambda_2, \ldots, \lambda_n$ равномерно стремятся к нулю, так что для любого $\varepsilon > 0$ мы имеем $\lambda_r < \varepsilon$ $(r = 1, 2, \ldots, n)$, если n достаточно велико.

2°. При любом постоянном $t > 0$ и при $n \to \infty$ числа $\varphi_r(t)$ $(r = 1, 2, \ldots, n)$ равномерно стремятся к единице, так что для любого $\varepsilon > 0$ мы имеем $1 - \varphi_r(t) < \varepsilon$ $(1 \leqslant r \leqslant n)$, если n достаточно велико.

Предпосылка 2° требует некоторого пояснения. Ближайший анализ показывает, что одного равномерного уменьшения интенсивности слагаемых потоков, выражаемого предпосылкой 1°, еще недостаточно для того, чтобы суммарный поток приближался к простейшему; этому могут помешать скопления большого числа вызовов одного и того же потока на небольших участках — скопления, возможность которых создается тем, что последействие в каждом из слагаемых потоков мы не подвергали до сих пор никаким ограничениям. Предпосылка 2° имеет целью как раз уменьшить шансы такого рода скоплений. Она говорит, что при сколь угодно большом t вероятность не получить после некоторого вызова за время t ни одного нового вызова *того же потока* должна при $n\to\infty$ стремиться к единице равномерно по всем слагаемым потокам.

Прежде всего мы покажем, что при сделанных предпосылках вероятность $V_0(t)$ отсутствия вызовов суммарного потока в промежутке $(0, t)$ приближается при $n\to\infty$ к соответствующей вероятности для простейшего потока с параметром Λ *).

Теорема Пальма. *При постоянном* $t>0$ *и при* $n\to\infty$

$$V_0(t)\to e^{-\Lambda t}.$$

Доказательство. Формула (10.5) дает для r-го потока

$$1-v_{0r}(t)=\lambda_r\int_0^t \varphi_r(u)\,du=\lambda_r t-\lambda_r\int_0^t [1-\varphi_r(u)]\,du,$$

или

$$v_{0r}(t)=1-\lambda_r t+\lambda_r\int_0^t [1-\varphi_r(u)]\,du;$$

поэтому в силу 2° при достаточно большом n

$$v_{0r}(t)=1-\lambda_r t+\varepsilon\theta_r\lambda_r t,\quad |\theta_r|<1,\quad 1\leqslant r\leqslant n, \qquad (14.1)$$

---

*) В работе Г. А. Ососкова «Одна предельная теорема для потоков однородных событий» (журнал «Теория вероятностей и ее применения», т. 1, вып. 2, 1956, 274—282) рассуждениями, в значительной степени повторяющими рассуждения А. Я. Хинчина, доказана несколько более общая теорема:

Для того чтобы суммарный поток, о котором шла речь, при $n\to\infty$ сходился к простейшему потоку с параметром Λ, достаточно выполнения условия: при фиксированном $t>0$ и $n\to\infty$

$$\frac{1}{\Lambda}\sum_{k=1}^{n}\lambda_k\varphi_k(t)\to 1. \qquad (1)$$

Если же параметры слагаемых потоков бесконечно малы, т. е. при любом $\varepsilon>0$ для достаточно больших n

$$\max_{1\leqslant k\leqslant n}\lambda_k\leqslant\varepsilon,$$

то условие (1) является и необходимым.

Несколько иная формулировка теоремы Г. А. Ососкова и иное доказательство даны Б. И. Григелионисом.— *Б. Г.*

отсюда легко находим

$$|\ln v_{0r}(t)+\lambda_r t|<c(t)\,\varepsilon\lambda_r \quad (1\leqslant r\leqslant n),$$

где $c(t)>0$ зависит только от t. Так как в силу взаимной независимости потоков

$$V_0(t)=\prod_{r=1}^{n} v_{0r}(t),$$

то отсюда

$$|\ln V_0(t)+\Lambda t|=\left|\sum_{r=1}^{n}[\ln v_{0r}(t)+\lambda_r t]\right|\leqslant$$
$$\leqslant\sum_{r=1}^{n}|\ln v_{0r}(t)+\lambda_r t|<c(t)\,\varepsilon\Lambda.$$

Так как $\varepsilon>0$ сколь угодно мало при достаточно большом n, то при $n\to\infty$

$$\ln V_0(t)\to-\Lambda t,\quad V_0(t)\to e^{-\Lambda t},$$

что и требовалось доказать.

Без достаточных оснований Пальм полагает, что доказанная теорема уже влечет за собой приближенно простейший характер суммарного потока *). Разумеется, этой теоремы еще далеко не достаточно, и мы должны теперь перейти к дальнейшему исследованию вопроса.

§ 15. Предельное поведение функций $V_k(t)$

Мы должны теперь в первую очередь убедиться, что и при любом $k>0$ функция $V_k(t)$ нашего суммарного потока при $n\to\infty$ стремится к соответствующей функции простейшего потока с параметром Λ, т. е. к $e^{-\Lambda t}(\Lambda t)^k/k!$ С этой целью нам понадобится следующая общая

Лемма. *Пусть мы имеем стационарный и ординарный поток с интенсивностью λ и функцией Пальма $\varphi(t)$ и пусть $\psi(t)$, как прежде, означает вероятность поступления не менее двух вызовов за время t. Тогда при любом*

*) Мы должны при этом отметить, что доказательство Пальма, с одной стороны, налагает на функции $\varphi_r(t)$ излишние требования, а с другой — содержит пробелы.

$t > 0$

$$\psi(t) \leqslant \lambda t\,[1 - \varphi(t)].$$

Доказательство. Разобьем отрезок $(0,\ t)$ на m равных между собой частей (ячеек)

$$\Delta_k = \left(\frac{k-1}{m}t,\quad \frac{k}{m}t\right) \quad (1 \leqslant k \leqslant m).$$

Поступление в отрезке $(0,\ t)$ по меньшей мере двух вызовов, очевидно, влечет за собой наступление по меньшей мере одного из следующих двух событий:

(A) Существует по меньшей мере одна ячейка Δ_k, содержащая не менее двух вызовов.

(B) Существует такая ячейка Δ_k $(k < m)$, что как в Δ_k, так и в отрезке $\left(\frac{k}{m}t,\ t\right)$ содержатся вызовы.

Поэтому мы имеем

$$\psi(t) \leqslant \mathrm{P}(A) + \mathrm{P}(B). \tag{15.1}$$

Прежде всего мы имеем в силу ординарности данного потока

$$\mathrm{P}(A) \leqslant m\psi\left(\frac{t}{m}\right) = t\,\frac{\psi\left(\frac{t}{m}\right)}{\frac{t}{m}} \longrightarrow 0 \quad (m \longrightarrow \infty). \tag{15.2}$$

Далее в § 9 мы обозначали через $h_0(\tau,\ t)$ вероятность того, что в промежутке длины τ имеются вызовы, а в следующем за ним промежутке длины t вызовов нет. $w(\tau) - h_0(\tau, t)$ есть поэтому вероятность того, что вызовы имеются как в τ, так и в t, следовательно,

$$\mathrm{P}(B) \leqslant \sum_{k=1}^{m-1}\left[w\left(\frac{t}{m}\right) - h_0\left(\frac{t}{m},\ \frac{m-k}{m}t\right)\right] \leqslant$$
$$\leqslant m\left[w\left(\frac{t}{m}\right) - h_0\left(\frac{t}{m},\ t\right)\right] =$$
$$= mw\left(\frac{t}{m}\right)\left[1 - \frac{h_0\left(\frac{t}{m};\ t\right)}{w\left(\frac{t}{m}\right)}\right]. \tag{15.3}$$

Так как при $m \to \infty$

$$mw\left(\frac{t}{m}\right) = t\,\frac{w\left(\frac{t}{m}\right)}{\frac{t}{m}} \to \lambda t,$$

то правая часть последних неравенств при $m \to \infty$ стремится к $\lambda t[1-\varphi(t)]$. А так как $\psi(t)$ от m не зависит, то из (15.1), (15.2) и (15.3) вытекает в пределе при $m \to \infty$

$$\psi(t) \leqslant \lambda t[1-\varphi(t)],$$

что и требовалось доказать.

Введем теперь следующие обозначения для событий:

A_k — в промежутке $(0, t)$ поступает k вызовов суммарного потока;

H_1 — ни один из слагаемых потоков не дает в $(0, t)$ более одного вызова;

H_2 — по меньшей мере один из слагаемых потоков дает в $(0, t)$ более одного вызова.

Нашей целью является исследование асимптотического поведения величины $V_k(t) = \mathrm{P}(A)$. Но

$$\mathrm{P}(A_k) = \mathrm{P}(H_1A_k) + \mathrm{P}(H_2A_k),$$

и, обозначая функцию $\psi(t)$ для r-го слагаемого потока через $\psi_r(t)$, в силу доказанной леммы

$$\mathrm{P}(H_2A_k) \leqslant \mathrm{P}(H_2) \leqslant \sum_{r=1}^{n} \psi_r(t) \leqslant \sum_{r=1}^{n} \lambda_r t[1-\varphi_r(t)];$$

а так как в силу предпосылки 2° мы имеем при достаточно большом n

$$1-\varphi_r(t) < \varepsilon \quad (r = 1, 2, \ldots, n),$$

то при достаточно большом n

$$\mathrm{P}(H_2A_k) \leqslant \varepsilon t\Lambda,$$

и, следовательно, при $n \to \infty$

$$\mathrm{P}(H_2A_k) \rightrightarrows 0, \quad V_k(t) = \mathrm{P}(H_1A_k) + o(1). \tag{15.4}$$

Но событие H_1A_k состоит, очевидно, в том, что из n слагаемых потоков какие-то k дают в промежутке $(0, t)$ по одному вызову, тогда как остальные $n-k$ в этом промежутке вызовов не дают. Поэтому, если $C(r_1, r_2, \ldots, r_k)$ означает произвольное сочетание из k различных между собой чисел ряда $1, 2, \ldots, n$, то

$$\mathsf{P}(H_1A_k)=\sum_C \frac{v_{1r_1}(t)\,v_{1r_2}(t)\,\ldots\,v_{1r_k}(t)}{v_{0r_1}(t)\,v_{0r_2}(t)\,\ldots\,v_{0r_k}(t)}\prod_{l=1}^{n} v_{0l}(t)=$$

$$=V_0(t)\sum_C \prod_{p=1}^{k} \frac{v_{1r_p}(t)}{v_{0r_p}(t)}, \tag{15.5}$$

где суммирование производится по всем сочетаниям описанного типа.

Теперь мы можем приступить к доказательству нашего основного утверждения.

Т е о р е м а. *При* $n \to \infty$

$$V_k(t) \to e^{-\Lambda t}\frac{(\Lambda t)^k}{k!} \quad (k=0,\ 1,\ 2,\ \ldots).$$

Д о к а з а т е л ь с т в о. Из (14.1) следует, что при достаточно большом n

$$|w_r(t)-\lambda_r t| = |1-v_{0r}(t)-\lambda_r t| < \varepsilon\lambda_r t \quad (1 \leqslant r \leqslant n);$$

а так как $v_{1r}(t)=w_r(t)-\psi_r(t)$ и в силу доказанной леммы, при достаточно большом n, $\psi_r(t)<\varepsilon\lambda_r t$, то мы можем при постоянном t писать

$$\left.\begin{aligned} v_{1r_p}(t) &= \lambda_{r_p} t + q_1\varepsilon\lambda_{r_p} t, \\ v_{0r_p}(t) &= 1-\lambda_{r_p} t + q_2\varepsilon\lambda_{r_p} t = 1+q_3\varepsilon \end{aligned}\right\} \quad (p=1, 2, \ldots, k),$$

где q_1, q_2, q_3 (как и q_4, q_5, ... в дальнейшем) ограничены при $n \to \infty$. Отсюда

$$\frac{v_{1r_p}(t)}{v_{0r_p}(t)} = \frac{\lambda_{r_p} t\,(1+q_1\varepsilon)}{1+q_3\varepsilon} = \lambda_{r_p} t\,(1+q_4\varepsilon) \quad (p=1,\ 2,\ \ldots\ k),$$

и, следовательно,

$$\prod_{p=1}^{k} \frac{v_{1r_p}(t)}{v_{0r_p}(t)} = \lambda_{r_1}\lambda_{r_2}\ldots\lambda_{r_k} t^k (1+q_5\varepsilon).$$

В силу (15.5) и теоремы Пальма (§ 14) поэтому

$$\mathsf{P}(H_1A_k) = e^{-\Lambda t}\, t^k (1+q_\bullet\varepsilon) \sum_C \lambda_{r_1}\lambda_{r_2}\ldots\lambda_{r_k}.$$

Так как ε произвольно мало при достаточно большом n, то для доказательства теоремы нам в силу (15.4) достаточно убедиться, что при $n\to\infty$ *)

$$S_k = \sum_{C_k} \lambda_{r_1}\lambda_{r_2}\ldots\lambda_{r_k} \to \frac{\Lambda^k}{k!}. \tag{15.6}$$

Это мы теперь и сделаем. При $k=1$ соотношение (15.6) тривиально. Пусть поэтому для некоторого $k>1$ при $n\to\infty$

$$S_{k-1} = \sum_{C_{k-1}} \lambda_{r_1}\lambda_{r_2}\ldots\lambda_{r_{k-1}} = \frac{\Lambda^{k-1}}{(k-1)!} + o(1). \tag{15.7}$$

Умножим каждый член суммы S_{k-1} на сумму всех λ_i, не входящих в него, т. е. на величину

$$\Lambda - \lambda_{r_1} - \lambda_{r_2} - \cdots - \lambda_{r_{k-1}}.$$

Тогда после раскрытия всех скобок мы получим сумму произведений вида

$$\lambda_{r_1}\lambda_{r_2}\ldots\lambda_{r_k},$$

где индексы $r_1, r_2, \ldots, r_k$ попарно различны между собой.

Каждое такое произведение есть один из членов суммы S_k; обратно, любой член суммы S_k будет, очевидно, получен при этой операции и притом в точности k раз [член $\lambda_{r_1}\lambda_{r_2}\ldots\lambda_{r_{k-1}}\lambda_{r_k}$ получается как $(\lambda_{r_1}\lambda_{r_2}\ldots\lambda_{r_{k-1}})\lambda_{r_k}$, как $(\lambda_{r_1}\lambda_{r_2}\ldots\lambda_{r_{k-2}}\lambda_{r_k})\lambda_{r_{k-1}}$ и т. д., наконец, как $(\lambda_{r_2},\ldots,\lambda_{r_k})\lambda_{r_1}$]. Так как, очевидно, при достаточно большом n и при любой комбинации C_{k-1}

$$\Lambda-(k-1)\varepsilon \leqslant \Lambda - \lambda_{r_1} - \lambda_{r_2} - \cdots - \lambda_{r_{k-1}} \leqslant \Lambda,$$

то из нашего подсчета следует

$$(\Lambda - k\varepsilon)S_{k-1} \leqslant kS_k \leqslant \Lambda S_{k-1},$$

*) Мы здесь для большей отчетливости обозначаем различные сочетания по k из чисел 1, 2, ..., n через C_k (вместо прежнего обозначения C).

и следовательно, из (15.7) при $n \to \infty$

$$(\Lambda - k\varepsilon)\left[\frac{\Lambda^{k-1}}{(k-1)!} + o(1)\right] \leqslant kS_k \leqslant \Lambda\left[\frac{\Lambda^{k-1}}{(k-1)!} + o(1)\right];$$

а так как ε при достаточно большом n как угодно мало, то $kS_k \to \Lambda^k/(k-1)!$ при $n \to \infty$, т. е.

$$\lim_{n \to \infty} S_k = \frac{\Lambda^k}{k!},$$

и наша теорема доказана индукцией.

§ 16. Предельная теорема

Только что доказанная теорема устанавливает, что в рассматриваемых нами условиях функции $V_k(t)$ для суммарного потока стремятся при $n \to \infty$ к соответствующим функциям простейшего потока с параметром Λ. Это, однако, еще не означает, что наш суммарный поток сам приближается к этому простейшему потоку. Дело в том, что, как мы видели в § 6, совокупность функций $V_k(t)$ $(k=0, 1, 2, \ldots)$ однозначно определяет собой данный поток лишь при условии, что это есть поток без последействия; мы же пока не рассматривали вопроса о последействии в нашем суммарном потоке. Поэтому вопрос о приближении этого суммарного потока к простейшему потоку с параметром Λ требует дальнейшего исследования.

Как показывает заключительная формула § 6, для простейшего потока с параметром Λ определяющей является формула

$$\mathsf{P}\{x(t_i) = k_i,\ 1 \leqslant i \leqslant m\} = e^{-\Lambda t_m}\Lambda^{k_m}\prod_{i=1}^{m}\frac{(t_i - t_{i-1})^{k_i - k_{i-1}}}{(k_i - k_{i-1})!}, \tag{16.1}$$

где $t_0 = k_0 = 0$, m — любое натуральное число, $0 < t_1 < t_2 < \ldots < t_m$, $0 \leqslant k_1 \leqslant k_2 \leqslant \ldots \leqslant k_m$ и все k_i — неотрицательные целые числа. Мы можем поэтому считать наш суммарный поток стремящимся к простейшему потоку с параметром Λ, если для этого суммарного потока вероятность, стоящая в левой части равенства (16.1), при любых m, t_i, k_i $(1 \leqslant i \leqslant m)$

и при безгранично возрастающем n имеет своим пределом правую часть этого равенства. Это мы теперь и установим.

Введем сначала более удобные для данной цели обозначения. Положим $t_i - t_{i-1} = u_i$, $k_i - k_{i-1} = l_i$, $t_m = \sum_{i=1}^{m} u_i = u$, $k_m = \sum_{i=1}^{m} l_i = k$ и обозначим через $n(u_i)$ число вызовов, поступающих в промежутке $u_i = (t_{i-1}, t_i)$. Тогда, очевидно, равенство (16.1) равносильно равенству

$$\mathsf{P}\{n(u_i) = l_i,\ 1 \leqslant i \leqslant m\} = e^{-\Lambda u} \Lambda^k \prod_{i=1}^{m} \frac{u_i^{l_i}}{l_i!}. \tag{16.2}$$

В § 15 мы рассматривали событие H_1A_k, состоящее в том, что за некоторый промежуток времени $(0, u)$ происходит k вызовов и что все эти вызовы принадлежат различным слагающим потокам. Пусть отрезок $(0, u)$ разбит на m частей $u_1, u_2, \ldots, u_m$, длины которых мы будем обозначать теми же буквами, и пусть событие H_1A_k совершилось. В силу стационарности и взаимной независимости слагающих потоков для любого из поступивших в промежутке $(0, u)$ k вызовов вероятность попасть в отрезок $u_i (1 \leqslant i \leqslant m)$ тогда равна u_i/u, каково бы ни было положение этого отрезка и каковы бы ни были моменты остальных поступивших вызовов. Обозначим через B событие

$$n(u_i) = l_i \quad (1 \leqslant i \leqslant m);$$

тогда из только что сказанного следует, что

$$\mathsf{P}_{H_1A_k}(B) = \frac{k!}{l_1!\,l_2!\ldots l_m!} \left(\frac{u_1}{u}\right)^{l_1} \left(\frac{u_2}{u}\right)^{l_2} \cdots \left(\frac{u_m}{u}\right)^{l_m}.$$

Но в § 15 мы доказали, что при $n \to \infty$

$$V_k(u) = \mathsf{P}(A_k) \to e^{-\Lambda u} \frac{(\Lambda u)^k}{k!},$$

$$\mathsf{P}(H_1A_k) \to e^{-\Lambda u} \frac{(\Lambda u)^k}{k!}, \quad \mathsf{P}(H_2A_k) \to 0.$$

Поэтому, учитывая, что в силу $\sum_{i=1}^{m} l_i = k$ событие A_k есть следствие события B, мы находим при $n \to \infty$

$$
\begin{aligned}
\mathrm{P}(B) = \quad & \mathrm{P}(A_k B) = \mathrm{P}(H_1 A_k B) + \mathrm{P}(H_2 A_k B) = \\
&= \mathrm{P}(H_1 A_k)\, \mathrm{P}_{H_1 A_k}(B) + \mathrm{P}(H_2 A_k)\, \mathrm{P}_{H_2 A_k}(B) \to \\
&\to e^{-\Lambda u} \frac{(\Lambda u)^k}{k!} \cdot \frac{k!}{l_1! l_2! \ldots l_m!} \left(\frac{u_1}{u}\right)^{l_1} \left(\frac{u_2}{u}\right)^{l_2} \ldots \left(\frac{u_m}{u}\right)^{l_m} = \\
&= e^{-\Lambda u} \Lambda^k \prod_{i=1}^{m} \frac{u_i^{l_i}}{l_i!}.
\end{aligned}
$$

Так как правая часть этого соотношения совпадает с правой частью равенства (16.2), то наша предельная теорема доказана. Мы можем, таким образом, утверждать, что при рассматриваемых нами условиях суммарный поток действительно стремится к простейшему потоку с параметром Λ.

П р и м е ч а н и е к л е м м е н а с т р. 55. Формулировка леммы неявно предполагает, что при любом r для каждого $u > 0$ функция $\psi_r(u)$ положительна. В действительности же существуют потоки, у которых $\psi_r(u) = 0$ для всех $u \leqslant a_r$. Это обстоятельство было отмечено в заметке Ю. Лукашевич «Замечание об одной теореме Хинчина из теории случайных потоков», Colloquium mathematicum 7:2, 285—287, 1960. Для последующего изложения лемма Хинчина должна быть заменена на следующую: *Для любого потока P и любого* $r = 1, 2, \ldots$

$$
\lim_{u \to a_r + 0} \frac{\psi_{r+1}(u)}{\psi_r(u)} = 0,
$$

где $a_r = \sup\, [u,\ \psi_r(u) = 0]$.

Доказательство этой измененной леммы является буквальным повторением рассуждений А. Я. Хинчина.— *Б. Г.*

ЧАСТЬ II

СИСТЕМЫ С ПОТЕРЯМИ

§ 17. Вводные замечания

После того как в первой части книги подробно изучены свойства потока поступающих вызовов, мы в последующих разделах должны рассмотреть основные вопросы, связанные с обслуживанием этого потока. Каждая станция (пункт, в который поступают вызовы) снабжена несколькими приборами, предназначенными для обслуживания вызовов; эти приборы могут быть весьма различного типа соответственно возлагаемым на них функциям; в частности, таким «прибором» может быть и человек (телефонистка, продавец в магазине, врач в амбулатории и т. п.); мы условимся ради единства терминологии во всех случаях называть обслуживающие приборы *линиями*. Процесс обслуживания протекает так, что всякий поступающий вызов *занимает* на некоторое время одну из имеющихся в момент его поступления свободных (незанятых) линий; пока линия занята каким-либо вызовом, она недоступна для вновь поступающих вызовов; период занятия какой-либо линии одним вызовом мы будем (опять-таки только ради единства терминологии) называть *разговором*.

Для задач обслуживания основное значение имеет различение двух типов устройства станций. Если в момент поступления какого-либо вызова имеются свободные линии, то вызов при любом устройстве станции занимает какую-либо одну из них и приступает к разговору. Различия возникают лишь в том случае, когда поступающий вызов застает все линии занятыми. При одном устройстве (*системы с потерями*) такой вызов просто получает отказ (или, как говорят, «теряется»), и все дальнейшее течение процесса обслуживания идет так, как если бы этот вызов вообще не поступал; при другом устройстве (системы с ожиданием) вызов, заставший все линии занятыми, сохраняется как претендент на будущий разговор и в дальнейшем занимает одну из освободившихся линий. Эти два типа устройств отличаются друг от друга не только способом решения основных задач, но и самой их постановкой: дело в том, что уже показатели качества обслуживания в этих двух случаях совершенно различны. Для систем с потерями основным показателем является, очевидно, вероятность отказа (потери) — понятие, лишенное смысла для систем с ожиданием. Напротив, для систем с ожиданием центральной задачей служит изучение времени ожидания как случайной величины; эта задача, очевидно, не имеет никакого смысла для систем с потерями.

Ввиду столь существенных различий мы, естественно, должны рассматривать эти два типа устройств обособленно друг от друга. Системам с потерями мы посвятим вторую, а системам с ожиданием — третью часть книги. Здесь же заметим еще только, что мыслимы и системы промежуточного типа; так, например, можно заставлять неудачные вызовы ожидать, покуда число ожидающих не превзойдет определенной границы, после чего вновь поступающие вызовы уже получают отказ. Теория таких имеющих несомненное практическое значение «смешанных» систем в настоящее время еще почти не разработана.

Кроме указанного основного различия, устройства станций отличаются друг от друга еще многими другими чертами, вследствие чего число различных между собой употребительных устройств становится очень большим. Само собой понятно, что в нашей краткой монографии мы имеем возможность остановиться лишь на небольшом числе типов таких устройств и что поэтому при решении каждой задачи мы вынуждены исходить из некоторых определенных предпосылок, которые в действительности отнюдь не являются единственно возможными. Мы считаем полезным здесь же рассмотреть еще некоторые важные черты, которыми могут отличаться друг от друга различные системы обслуживания вызовов.

1° Будем предполагать, что все имеющиеся линии в равной степени доступны всем поступающим вызовам (такую систему называют «полнодоступным пучком линий»). В противоположность этому в действительности нередко приходится встречать случаи, когда определенные категории вызовов «прикреплены» к определенным линиям и не могут занимать других линий.

2° Мы почти всегда будем предполагать поток поступающих вызовов простейшим. В части I мы сделали достаточно указаний относительно того, в какой мере это допущение соответствует действительности.

3° Назовем полнодоступный пучок упорядоченным, если его линии перенумерованы так, что поступающий вызов всегда занимает линию с наименьшим номером из числа тех, которые свободны в момент его поступления (т. е. первую, если она свободна; вторую, если первая занята, а вторая свободна, и т. д.). В неупорядоченном пучке линии занимаются в случайном порядке. На практике встречаются устройства обоих типов. Следует отметить, что для ряда основных задач вопрос об упорядоченности или неупорядоченности пучка не играет никакой роли; такова, например, вся проблема Эрланга, которой будут посвящены главы 6 и 7. Однако есть и такие задачи, которые, напротив, получают смысл только для упорядоченного пучка, такова, например, задача Пальма, которая будет рассмотрена в главе 8.

4° Для систем с ожиданием имеет во многих задачах существенное значение вопрос о порядке обслуживания ожидающих вызовов. Это обслуживание может проводиться либо в порядке очереди, либо в случайном порядке, причем в ряде задач подсчеты приводят в этих двух случаях к различным результатам.

5° Наконец, весьма важное значение для большинства задач теории обслуживания имеет вопрос о законе распределения длительности занятий (разговоров). В значительном большинстве исследований этот закон предполагается *показательным* (т. е. вероятность того, что длительность разговора будет больше t, принимается равной $e^{-\beta t}$, где $\beta > 0$ — постоянная). Этот выбор обусловлен главным образом тем, что он значительно облегчает необходимые расчеты. Можно без преувеличения сказать, что заметное большинство задач теории обслуживания решается сравнительно просто при показательном распределении длительности разговоров и, напротив, приводит к неодолимым трудностям при почти всяком ином предположении о форме этого закона распределения.

Такое исключительное положение показательного распределения длительности разговоров обусловлено главным образом одним его важным свойством, которым мы не раз будем пользоваться в дальнейшем. Пусть $f_a(t)$ есть вероятность того, что разговор, длящийся уже a секунд, продлится еще не менее t секунд, так что при показательном распределении $f_0(t)=e^{-\beta t}$. Так как, очевидно, всегда

$$f_0(a+t)=f_0(a)f_a(t),$$

то при показательном распределении

$$e^{-\beta(a+t)}=e^{-\beta a}\cdot f_a(t),$$

откуда

$$f_a(t)=e^{-\beta t}.$$

Это означает, что *при показательном распределении длительности разговоров закон распределения оставшейся части разговора не зависит от его «возраста», т. е. от того, сколько времени он уже длится*. Именно это свойство показательного распределения в большинстве случаев и упрощает производимые расчеты. Вместе с тем оно же заставляет думать, что в практических ситуациях гипотеза показательного распределения длительности разговоров вряд ли может рассчитывать на точное осуществление и в лучшем случае способна служить лишь более или менее хорошим приближением к действительности.

Глава 6

ЗАДАЧА ЭРЛАНГА ДЛЯ КОНЕЧНОГО ПУЧКА

§ 18. Постановка задачи

В этой главе мы будем иметь дело с полнодоступным пучком (упорядоченным или нет — безразлично) из n линий, на который поступает простейший поток вызовов с параметром λ; мы допустим, что длительность разговоров подчиняется показательному закону распределения $1-e^{-x}$. Так как в случае общего показательного закона $1-e^{-\beta x}$ средняя длительность разговора равна $1/\beta$, то выбор $\beta=1$ означает просто, что мы принимаем эту среднюю длительность разговора за единицу времени, что, конечно, ни в какой мере не ограничивает общности исследования.

Если известно, что в некоторый момент 0 было занято ровно k линий данного пучка ($0 \leqslant k \leqslant n$), то число $N(t)$ занятых линий в какой-либо последующий момент t есть случайная величина, значение которой определяется рядом случайных факторов: моментами окончания тех k разговоров, которые ведутся в момент 0, моментами поступления новых вызовов между 0 и t и длинами тех разговоров, которые ведутся этими вызовами. Число $N(t)$ представляет собой, таким образом, однопараметрическое семейство случайных величин, или, как говорят, *случайный процесс*. Этот процесс обладает, при сделанных нами предпосылках, одним важным свойством, позволяющим применить к его изучению хорошо разработанные методы.

Пусть $N(t_0)=i$, т. е. в момент t_0 занято i линий. Тогда последующее течение процесса в вероятностном смысле независимо от всего, что происходило до момента t_0. В самом деле, это дальнейшее течение, как мы уже отметили, однозначно определяется следующими тремя факторами:

1. Моментами окончания тех i разговоров, которые ведутся в момент t_0.
2. Моментами появления новых вызовов после t_0.
3. Длительностями разговоров для вызовов, упомянутых в 2.

Но легко видеть, что ни один из этих трех случайных факторов не зависит от того, что происходило до момента t_0. Для фактора 1 это вытекает из принятого нами показательного распределения длительности разговоров, при котором, как мы видели в § 17, длительность остающейся части разговора не зависит от его возраста. Для фактора 2 это следует из того, что поступающий поток вызовов — простейший и, следовательно, не обладает последействием. Наконец, для фактора 3 это очевидно само собой. Таким образом, действительно все три перечисленных фактора не зависят от «прошлого» нашей системы, т. е. от течения процесса до момента t_0, а следовательно, не зависит от прошлого и течение процесса после момента t_0, ибо оно однозначно определяется указанными тремя факторами.

Таким образом, случайный процесс $N(t)$ обладает следующим свойством: если известно $N(t_0)$, то течение процесса после момента t_0 в вероятностном смысле независимо от его течения до момента t_0 (коротко: если известно настоящее, то будущее не зависит от прошедшего). Случайные процессы, обладающие этим свойством, называют **процессами Маркова**.

Если в некоторый момент t занято i линий пучка [т. е. $N(t)=i$], то мы будем говорить, что в этот момент система находится в «состоянии i» ($i=0, 1, \ldots, n$); всего, таким образом, для системы возможно $n+1$ различных состояний. Обозначим через $\mathsf{P}_{ik}(t)$ ($t>0$, $0\leqslant i\leqslant n$, $0\leqslant k\leqslant n$) условную вероятность того, что система, находившаяся в некоторый момент в состоянии i, по истечении t единиц времени перейдет в состояние k [вероятность $N(a+t)=k$ при условии $N(a)=i$]. Эти «переходные» вероятности играют основную роль в исследовании процессов Маркова. Очевидно, всегда

$$\mathsf{P}_{ik}(t)\geqslant 0, \quad \sum_{k=0}^{n}\mathsf{P}_{ik}(t)=1 \quad (0\leqslant i\leqslant n,\ 0\leqslant k\leqslant n).$$

Если $t_1>0$, $t_2>0$, $0\leqslant i\leqslant n$, $0\leqslant k\leqslant n$, то имеет место соотношение

$$\mathsf{P}_{ik}(t_1+t_2)=\sum_{r=0}^{n}\mathsf{P}_{ir}(t_1)\,\mathsf{P}_{rk}(t_2). \tag{18.1}$$

В самом деле, для того чтобы за время t_1+t_2 перейти из состояния i в состояние k, система должна сначала перейти за время t_1 из состояния i в некоторое состояние r, а потом за время t_2 перейти из состояния r в состояние k, так что соотношение (18. 1) есть результат простого применения формулы полной вероятности. Очень важно отметить, что эта формула имеет место только для процессов Маркова; в самом деле, только для процессов Маркова переходную вероятность $\mathsf{P}_{rk}(t_2)$ можно считать независимой от i; если бы наш процесс не был процессом Маркова, то на месте $\mathsf{P}_{rk}(t_2)$ должна была бы стоять вероятность перехода за время t_2 из состояния r в состояние k *при дополнительном условии, что система предварительно за время t_1 перешла в состояние r из состояния i.* Для процессов же Маркова вероятность $\mathsf{P}_{rk}(t_2)$ не зависит от того, что происходило до этого перехода, благодаря чему и имеет место формула (18.1).

Формула (18.1), иногда называемая уравнением Чэпмана — Колмогорова, лежит в основании всех исследований о процессах Маркова; в нашем изложении она также будет играть значительную роль.

Если в начальный момент 0 система находится в данном состоянии i, то вероятность застать ее в момент $t>0$ в состоянии k равна $\mathsf{P}_{ik}(t)$. Мы можем, однако, сделать относительно начального момента допущение более общего характера: в момент 0 мы можем считать известным не состояние системы, а лишь **начальные вероятности** $\mathsf{P}_i(0)$ $(0 \leqslant i \leqslant n)$ различных состояний; этот общий случай, конечно, сводится к упомянутому нами частному случаю, когда из чисел $\mathsf{P}_i(0)$ какое-нибудь одно равно единице (а остальные равны нулю). Вероятность $\mathsf{P}_k(t)$ застать систему в момент t в состоянии k по формуле полной вероятности равна

$$\mathsf{P}_k(t) = \sum_{i=0}^{n} \mathsf{P}_i(0)\, \mathsf{P}_{ik}(t); \tag{18.2}$$

эта вероятность зависит как от t, так и от начальных данных $\mathsf{P}_i(0)$ $(0 \leqslant i \leqslant n)$.

Если в уравнении (18.1) помножить обе части на $\mathsf{P}_i(0)$ и просуммировать по i от 0 до n, то в силу (18.2) мы получим

$$\mathsf{P}_k(t_1+t_2) = \sum_{r=0}^{n} \mathsf{P}_r(t_1)\, \mathsf{P}_{rk}(t_2) \quad (0 \leqslant k \leqslant n). \tag{18.3}$$

Задача Эрланга, которой мы посвятим настоящую главу состоит в отыскании вероятности застать систему в том или ином данном состоянии. В свете изложенного нами выше такая постановка вопроса требует пояснений; случайный процесс $N(t)$ нестационарен, вероятности

$$\mathsf{P}\{N(t)=k\} = \mathsf{P}_k(t) \quad (0 \leqslant k \leqslant n)$$

меняются с течением времени и, кроме того, зависят еще от начальных данных, т. е. от чисел $\mathsf{P}_i(0)$ $(0 \leqslant i \leqslant n)$; представляется поэтому, что искомые в задаче Эрланга вероятности $\mathsf{P}_k(t)$ могут быть определены лишь при данных t и $\mathsf{P}_i(0)$ $(0 \leqslant i \leqslant n)$. В приложениях, однако, обычно считают возможным говорить о вероятности p_k застать систему в состоянии k, независимо от выбранного момента времени и от начальных данных. Чтобы теоретически оправдать такую практику, можно попытаться установить, что процесс $N(t)$ при $t \to \infty$ безгранично приближается к некоторому стационарному процессу, не зависящему от начальных данных; говоря более конкретно, надо установить, что вероятности $\mathsf{P}_k(t)$ при $t \to \infty$ стремятся к некоторым постоянным числам $p_k (0 \leqslant k \leqslant n)$, не зависящим от начальных данных. Эти числа p_k мы тогда, естественно, и принимаем за искомые в задаче Эрланга вероятности застать систему в том или другом определенном состоянии, ибо число p_k, с одной стороны, не зависит от начальных данных задачи, а с другой — становится сколь угодно близким к реальной вероятности $\mathsf{P}_k(t)$, если процесс продолжается достаточно долгое время.

Итак, нашей задачей является показать, что *при* $t \to \infty$ *функции* $\mathsf{P}_k(t)$ *стремятся к числам* $p_k (0 \leqslant k \leqslant n)$, не зависящим от начальных данных. Разумеется, при этом числа p_k должны быть нами найдены. Для практики особо важное значение имеет число p_n — вероятность застать все линии занятыми. Это есть вероятность потери (отказа), являющаяся для систем с потерями важнейшим показателем качества обслуживания.

Мы прежде всего редуцируем поставленную нами задачу к другой, более удобной для применения уравнения Чэпмана — Колмогорова, с помощью следующего вспомогательного предложения.

Лемма. *Для того чтобы вероятности* $\mathsf{P}_k(t)$ *при* $t \to \infty$ *стремились к не зависящим от начальных данных числам* p_k $(0 \leqslant k \leqslant n)$, *необходимо и достаточно, чтобы к тем же пределам стремились соответственно переходные вероятности* $\mathsf{P}_{ik}(t) (0 \leqslant k \leqslant n)$ *при любом значении* i.

Доказательство. Оба утверждения леммы почти очевидны в силу (18.2).

1. Пусть $\mathsf{P}_k(t) \to p_k$ $(t \to \infty,\ 0 \leqslant k \leqslant n)$, где p_k не зависит от начальных данных; выбирая тогда $\mathsf{P}_i(0) = 1$, $\mathsf{P}_l(0) = 0$ $(l \neq i)$, мы в силу (18.2) имеем $\mathsf{P}_k(t) = \mathsf{P}_{ik}(t)$, и, следовательно, $\mathsf{P}_{ik}(t) \to p_k$ $(t \to \infty,\ 0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$.

2. Пусть, обратно, $\mathsf{P}_{ik}(t) \to p_k$ $(t \to \infty,\ 0 \leqslant i \leqslant n,$ $0 \leqslant k \leqslant n)$; тогда в силу (18.2) при любом выборе вероят-

ностей $\mathsf{P}_i(0)$ мы имеем

$$\mathsf{P}_k(t) \longrightarrow \sum_{i=0}^{n} \mathsf{P}_i(0) p_k = p_k \quad (t \longrightarrow \infty,\ 0 \leqslant k \leqslant n),$$

так как

$$\sum_{i=0}^{n} \mathsf{P}_i(0) = 1.$$

В силу этой леммы наша ближайшая задача сводится к доказательству того, что при $t \longrightarrow \infty$ переходные вероятности $\mathsf{P}_{ik}(t)$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$ стремятся к пределам p_k, не зависящим от i.

§ 19. Теорема Маркова

Задачу, упомянутую в конце предыдущего параграфа, можно было бы пытаться решить, найдя выражения переходных вероятностей $\mathsf{P}_{ik}(t)$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$ для специального интересующего нас процесса $N(t)$ и стремясь затем анализом этих выражений установить существование нужных нам пределов и одновременно найти эти пределы. Однако мы предпочтем другой путь, при котором трудная задача отыскания функций $\mathsf{P}_{ik}(t)$ может быть обойдена. В настоящем параграфе мы, не пытаясь найти переходные вероятности $\mathsf{P}_{ik}(t)$ и их пределы, установим только самый факт существования этих пределов; такой путь становится возможным потому, что эта теорема существования представляет собой свойство очень широкого класса процессов Маркова, отнюдь не являясь характеристикой нашего специального процесса $N(t)$. После того как это будет сделано, мы в следующем параграфе, опираясь на доказанное уже существование пределов, сможем найти эти пределы для специально интересующего нас процесса, снова минуя явные выражения функций $\mathsf{P}_{ik}(t)$.

Условимся называть процесс Маркова, характеризуемый переходными вероятностями $\mathsf{P}_{ik}(t)$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$, т р а н з и т и в н ы м, если существует такое $t > 0$, что $\mathsf{P}_{ik}(t) > 0$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$. Таким образом, для транзитивного процесса существует такой промежуток времени, в течение которого возможен переход системы из любого состояния в любое другое. Непосредственно ясно, что интересующий нас процесс $N(t)$ транзитивен, причем в качестве t может быть выбрано любое положительное число.

Теорема Маркова*). *Для любого транзитивного процесса Маркова* $\mathsf{P}_{ik}(t)$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$ *предел*

$$\lim_{t\to\infty} \mathsf{P}_{ik}(t) = p_k \qquad (0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$$

существует и не зависит от i.

Доказательство. Во всем дальнейшем k будет означать произвольное закрепленное число ряда $0, 1, \ldots, n$. Положим

$$\operatorname*{Max}_{0 \leqslant i \leqslant n} \mathsf{P}_{ik}(t) = M_k(t), \quad \operatorname*{Min}_{0 \leqslant i \leqslant n} \mathsf{P}_{ik}(t) = m_k(t).$$

В силу (18.1) для любого $i\,(0 \leqslant i \leqslant n)$ и любых $t > 0,\ \tau > 0$

$$\mathsf{P}_{ik}(t+\tau) = \sum_{r=0}^{n} \mathsf{P}_{ir}(\tau)\,\mathsf{P}_{rk}(t) \leqslant M_k(t) \sum_{r=0}^{n} \mathsf{P}_{ir}(\tau) = M_k(t),$$

а следовательно,

$$M_k(t+\tau) \leqslant M_k(t),$$

т. е. $M_k(t)$ есть невозрастающая функция от t. Подобным же образом легко убедиться, что $m_k(t)$ есть неубывающая функция от t. Отсюда следует, что $M_k(t)$ и $m_k(t)$ при $t \to \infty$ стремятся к определенным пределам. Теорема, очевидно, будет доказана, если мы покажем, что эти пределы совпадают между собой, а для этого необходимо и достаточно иметь

$$\Delta_k(t) = M_k(t) - m_k(t) \to 0 \qquad (t \to \infty).$$

В дальнейшем все суммы по всем индексам будут распространяться на значения $0, 1, \ldots, n$ этих индексов, вследствие чего мы можем не указывать области суммирования. Пусть $\mathsf{P}_{ir}(t_0) > 0$ $(0 \leqslant i \leqslant n,\ 0 \leqslant k \leqslant n)$ (такое t_0 найдется в силу транзитивности процесса). Положим,

$$d_{il}^{(r)} = \mathsf{P}_{ir}(t_0) - \mathsf{P}_{lr}(t_0) \qquad (0 \leqslant i, l, r \leqslant n),$$

---

*) У Маркова эта теорема доказана для «цепей», т. е. процессов с дискретным временем; однако доказательство переносится без всяких изменений на интересующий нас случай непрерывного времени.

и будем в дальнейшем обозначать через $\sum'$ (соответственно $\sum''$) суммы, распространенные лишь на область положительных (соответственно неположительных) $d_{il}^{(r)}$. Тогда в силу

$$\sum_r \mathsf{P}_{ir}(t_0) = \sum_r \mathsf{P}_{lr}(t_0) = 1 \quad (0 \leqslant i,\ l \leqslant n)$$

мы имеем

$$0 = \sum_r d_{il}^{(r)} = \sum_r{}' |d_{il}^{(r)}| - \sum_r{}'' |d_{il}^{(r)}| \quad (0 \leqslant i,\ l \leqslant n),$$

или

$$\sum_r{}' |d_{il}^{(r)}| = \sum_r{}'' |d_{il}^{(r)}| = h_{il} \quad (0 \leqslant i,\ l \leqslant n);$$

при этом в силу $\mathsf{P}_{lr}(t_0) > 0$ $(0 \leqslant l,\ r \leqslant n)$

$$h_{il} = \sum_r{}' d_{il}^{(r)} = \sum_r{}' [P_{ir}(t_0) - \mathsf{P}_{lr}(t_0)] < \sum_r{}' \mathsf{P}_{ir}(t_0) \leqslant \sum_r \mathsf{P}_{ir}(t_0) = 1.$$

Это неравенство имеет место для любых i и l, вследствие чего и

$$h = \max_{0 \leqslant i,\ l \leqslant n} h_{il} < 1.$$

Пусть теперь q — любое натуральное число. Тогда при $0 \leqslant i \leqslant n$ $0 \leqslant l \leqslant n$ в силу (18.1)

$$\begin{aligned}
\mathsf{P}_{ik}(qt_0 + t_0) - \mathsf{P}_{lk}(qt_0 + t_0) &= \\
&= \sum_r \mathsf{P}_{ir}(t_0)\mathsf{P}_{rk}(qt_0) - \sum_r \mathsf{P}_{lr}(t_0)\mathsf{P}_{rk}(qt_0) = \\
&= \sum_r [\mathsf{P}_{ir}(t_0) - \mathsf{P}_{lr}(t_0)]\mathsf{P}_{rk}(qt_0) = \\
&= \sum_r d_{il}^{(r)}\mathsf{P}_{rk}(qt_0) = \sum_r{}' d_{il}^{(r)}\mathsf{P}_{rk}(qt_0) - \sum_r{}'' |d_{il}^{(r)}|\mathsf{P}_{rk}(qt_0) \leqslant \\
&\leqslant M_k(qt_0)h_{il} - m_k(qt_0)h_{il} = h_{il}\Delta_k(qt_0) \leqslant h\Delta_k(qt_0).
\end{aligned}$$

Так как это неравенство имеет место для любых i, l, то можно принять

$$\mathsf{P}_{ik}(qt_0 + t_0) = M_k(qt_0 + t_0), \qquad \mathsf{P}_{lk}(qt_0 + t_0) = m_k(qt_0 + t_0);$$

тогда мы получаем

$$\Delta_k(qt_0 + t_0) \leqslant h\Delta_k(qt_0);$$

рекуррентное же применение этого неравенства дает

$$\Delta_k(qt_0) \leqslant h^{q-1}\Delta_k(t_0) \leqslant h^{q-1} \to 0 \quad (q \to \infty).$$

В силу моннотонности функции $\Delta_k(t)$ отсюда очевидно следует, что

$$\Delta_k(t) \longrightarrow 0 \quad (t \longrightarrow \infty).$$

Этим теорема Маркова доказана.

§ 20. Уравнения и формулы Эрланга

Теперь мы переходим к ставшему классическим методу Эрланга определения величин p_k, существование которых нами только что доказано. В отличие от предыдущего параграфа, мы будем при этом иметь в виду исключительно наш конкретный процесс $N(t)$.

Во всем дальнейшем нам придется иметь дело с промежутком времени бесконечно малой длины τ. Условимся для краткости обозначать через $o(\tau)$ всякую бесконечно малую порядка выше τ и соединять знаком $\approx$ всякие две величины, разность которых есть величина вида $o(\tau)$.

Согласно принятым нами в § 18 предпосылкам вероятность $w(\tau)$ поступления по меньшей мере одного вызова за промежуток времени τ есть величина $\approx \lambda\tau$, а вероятность $\psi(\tau)$ поступления более одного вызова — величина вида $o(\tau)$. С другой стороны, если какая-либо линия в данный момент занята, то вероятность оставаться занятой еще в течение τ секунд (или более) для нее равна $e^{-\tau}$: если занято k линий, то вероятность того, что все они останутся занятыми в течение промежутка времени τ, равна поэтому $e^{-k\tau}$; вероятность же того, что в течение промежутка времени τ по меньшей мере одна из этих линий освободится, равна

$$1 - e^{-k\tau} \approx k\tau.$$

Поступление вызовов и освобождение линий представляют собой элементарные события, в моменты которых скачкообразно меняется величина $N(t)$. Из того, что мы до сих пор установили для вероятностей таких элементарных событий, с очевидностью следует, что вероятность наступления в промежутке длины τ по меньшей мере одного элементарного события (того или другого типа) при $\tau \longrightarrow 0$ асимптотически пропорциональна τ; вероятность же наступления в промежутке длины двух или более элементарных событий (все равно, каких типов) есть величина вида $o(\tau)$ (или, что то же, ≈ 0).

Эти замечания позволяют легко найти асимптотические выражения переходных вероятностей $P_{ik}(\tau)$ при $\tau \to 0$. Прежде всего, если $|i-k|>1$, то переход из состояния i в состояние k требует, очевидно, наступления по меньшей мере двух элементарных событий; поэтому в силу вышесказанного при $\tau \to 0$

$$P_{ik}(\tau) \approx 0 \quad (|i-k|>1). \tag{20.1}$$

Далее, для перехода из состояния $k<n$ в состояние $k+1$ требуется либо наступление одного вызова, либо наступление более чем одного элементарного события; поэтому в силу вышесказанного при $\tau \to 0$

$$P_{k,\, k+1}(\tau) \approx \lambda\tau \quad (0 \leqslant k < n).$$

Чтобы система перешла из состояния $k>0$ в состояние $k-1$, требуется либо освобождение одной из линий, либо наступление более чем одного элементарного события; так как вероятность освобождения одной из k занятых линий за время τ при $\tau \to 0$, как мы видели выше, $\approx k\tau$, то мы находим

$$P_{k,\, k-1}(\tau) \approx k\tau \quad (0 < k \leqslant n).$$

Наконец, в силу (20.1) мы имеем при $\tau \to 0$

$$P_{kk}(\tau) \approx 1 - P_{k,\, k+1}(\tau) - P_{k,\, k-1}(\tau) \quad (0 \leqslant k \leqslant n),$$

где при $k=n$ второй, а при $k=0$ третий член правой части надо заменить нулем; это дает

$$\begin{aligned} &P_{00}(\tau) \approx 1 - \lambda\tau; \\ &P_{kk}(\tau) \approx 1 - \lambda\tau - k\tau \quad (1 \leqslant k \leqslant n-1); \\ &P_{nn}(\tau) \approx 1 - n\tau. \end{aligned}$$

Таким образом, для всех вероятностей $P_{ik}(\tau)$ нами установлены очень простые асимптотические выражения с точностью до величины вида $o(\tau)$.

Теперь мы обратимся к уравнению (18.3), в силу которого при любом постоянном $t \geqslant 0$

$$P_k(t+\tau) = \sum_r P_r(t)\, P_{rk}(\tau).$$

Применяя к вероятностям $P_{rk}(\tau)$ в правой части этого равенства найденные нами асимптотические оценки, мы находим

$$\begin{aligned}
P_0(t+\tau) &= P_0(t)\,P_{00}(\tau)+P_1(t)\,P_{10}(\tau)+o(\tau)=\\
&=(1-\lambda\tau)\,P_0(t)+\tau P_1(t)+o(\tau);\\
P_k(t+\tau) &= P_{k-1}(t)\,P_{k-1,\,k}(\tau)+\\
&\quad +P_k(t)\,P_{kk}(\tau)+P_{k+1}(t)\,P_{k+1,\,k}(\tau)+o(\tau)=\\
&=\lambda\tau\,P_{k-1}(t)+(1-\lambda\tau-k\tau)\,P_k(t)+(k+1)\tau\,P_{k+1}(t)+o(\tau)\\
&\qquad (0<k<n);\\
P_n(t+\tau) &= P_{n-1}(t)\,P_{n-1,\,n}(\tau)+P_n(t)\,P_{nn}(\tau)+o(\tau)=\\
&=\lambda\tau P_{n-1}(t)+(1-n\tau)P_n(t)+o(\tau).
\end{aligned}$$

Отсюда

$$\begin{aligned}
\frac{P_0(t+\tau)-P_0(t)}{\tau} &= -\lambda\,P_0(t)+P_1(t)+o(1);\\
\frac{P_k(t+\tau)-P_k(t)}{\tau} &= \lambda P_{k-1}(t)-(\lambda+k)P_k(t)+\\
&\quad +(k+1)P_{k+1}(t)+o(1)\quad (0<k<n);\\
\frac{P_n(t+\tau)-P_n(t)}{\tau} &= \lambda P_{n-1}(t)-nP_n(t)+o(1).
\end{aligned}$$

Если мы заставим теперь τ стремиться к нулю (сохраняя t постоянным), то убеждаемся в существовании производных всех функций $P_k(t)$ ($k=0,\ 1,\ \ldots,\ n$) и находим в пределе

$$\left.\begin{aligned}
&P_0'(t)=-\lambda P_0(t)+P_1(t);\\
&P_k'(t)=\lambda P_{k-1}(t)-(\lambda+k)\,P_k(t)+(k+1)P_{k+1}(t)\ (0<k<n);\\
&P_n'(t)=\lambda P_{n-1}(t)-nP_n(t).
\end{aligned}\right\}(\mathcal{E})$$

Система $n+1$ уравнений ($\mathcal{E}$) с $n+1$ неизвестными функциями $P_k(t)$ ($k=0,\ 1,\ \ldots,\ n$) называется **системой Эрланга**. Так как все уравнения этой системы однородны, то искомые функции содержат произвольный постоянный множитель, который может быть определен из очевидного «нормировочного» условия

$$\sum_{k=0}^{n} P_k(t)=1.$$

Как мы уже говорили, нам нет надобности искать решения системы дифференциальных уравнений ($\mathcal{E}$). В § 19 мы доказали, что для любого $k\,(0 \leqslant k \leqslant n)$ существует предел

$$\lim_{t \to \infty} \mathrm{P}_k(t) = p_k.$$

Отсюда следует, что правые части всех уравнений ($\mathscr{E}$) при $t \longrightarrow \infty$ имеют пределы. Переходя к левым частям, мы видим, что все производные $\mathrm{P}'_k(t)$ при $t \longrightarrow \infty$ стремятся к пределам; но такой предел может быть только нулем, так как, если бы какое-нибудь $\mathrm{P}'_k(t)$ стремилось к числу, отличному от нуля, соответствующее $|\mathrm{P}_k(t)|$ при $t \longrightarrow \infty$ возрастало бы безгранично, что (независимо от реального смысла величин $\mathrm{P}_k(t)$ как вероятностей) невозможно уже в силу теоремы Маркова. Таким образом, мы приходим к выводу, что

$$\mathrm{P}'_k(t) \longrightarrow 0 \quad (t \longrightarrow \infty) \quad (0 \leqslant k \leqslant n),$$

вследствие чего система ($\mathscr{E}$) в пределе при $t \longrightarrow \infty$ дает

$$\left.\begin{aligned} -\lambda p_0 + p_1 &= 0; \\ \lambda p_{k-1} - (\lambda + k) p_k + (k+1) p_{k+1} &= 0 \ (0 < k < n); \\ \lambda p_{n-1} - n p_n &= 0. \end{aligned}\right\} \qquad (20.2)$$

Эта простая система линейных уравнений вместе с нормировочным условием $\sum\limits_{k=0}^{n} p_k = 1$ может служить для однозначного определения искомых чисел p_k.

Если положить

$$\lambda p_{k-1} - k p_k = z_k \quad (1 \leqslant k \leqslant n),$$

система (20.2) может быть записана в виде

$$z_1 = 0, \quad z_k - z_{k+1} = 0 \quad (0 < k < n), \quad z_n = 0,$$

откуда $z_k = 0$ $(1 \leqslant k \leqslant n)$; это же дает

$$p_k = \frac{\lambda}{k} p_{k-1} \quad (1 \leqslant k \leqslant n),$$

и, следовательно,

$$p_k = \frac{\lambda^k}{k!} p_0 \quad (1 \leqslant k \leqslant n).$$

Применяя нормировочное условие, находим

$$p_0=\frac{1}{\sum\limits_{l=0}^{n}\frac{\lambda^l}{l!}},$$

и, следовательно,

$$p_k=\frac{\frac{\lambda^k}{k!}}{\sum\limits_{l=0}^{n}\frac{\lambda^l}{l!}}\quad (k=0,\ 1,\ \ldots,\ n). \tag{20.3}$$

Формулы (20.3), называемые обычно **формулами Эрланга**, полностью решают поставленную нами задачу. В частности, вероятность «потери» дается формулой

$$p_n=\frac{\frac{\lambda^n}{n!}}{\sum\limits_{l=0}^{n}\frac{\lambda^l}{l!}}.$$

Полезно отметить, насколько решающую роль во всем проведенном исследовании играла предпосылка о показательном распределении длительности разговоров; только при этом допущении процесс $N(t)$ становится процессом Маркова; при отказе от этого допущения все развитые нами в § 19 и 20 методы становятся принципиально неприменимыми. В специальной литературе имеется целый ряд попыток доказать, что формулы Эрланга остаются в силе и при любом другом распределении длительности разговоров. Однако, насколько мы видим, эти попытки не привели до сих пор к сколько-нибудь законченным результатам*).

*) Вскоре после опубликования этой работы сам А. Я. Хинчин (см. «О формулах Эрланга в теории массового обслуживания», стр. 199) и Б. А. Севастьянов («Эргодическая теорема для марковских процессов и ее приложение к телефонным задачам с отказами», Теор. вероят., т. 2, вып. 1, 1957, стр. 106—116) доказали это утверждение.— *Б. Г.*

§ 21. Эргодическая теорема

Любая вероятность в любой теории получает реальный смысл лишь в том случае, если известна реальная совокупность объектов, в которой эта вероятность интерпретирует собой долю того или другого признака. Каков же реальный смысл тех вероятностей p_k, которыми мы занимались в последних параграфах? Что означает, в частности, «вероятность потери» p_n?

В подавляющем большинстве специальных исследований реальная интерпретация этих вероятностей носит один и тот же вполне определенный характер: их истолковывают как *средние относительные времена пребывания* системы в соответствующих состояниях. Это означает следующее. Обозначим через $x_k(t)$ величину, равную 1, если система в момент t находится в состоянии k, и равную 0 в противном случае. Тогда интеграл

$$\int_0^T x_k(t)\,dt$$

представляет собой суммарную длину тех промежутков времени (между 0 и T), в течение которых система находится в состоянии k, а отношение

$$\frac{1}{T}\int_0^T x_k(t)\,dt$$

дает нам среднее относительное время пребывания системы в состоянии k (за промежуток времени $(0, T)$). Под вероятностью p_k застать систему в состоянии k понимают тогда предел

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T x_k(t)\,dt. \tag{21.1}$$

Представляется, однако, очевидным, что определенные нами в предшествующих параграфах величины p_k непосредственно не допускают подобного истолкования. Это видно уже из того, что величина $x_k(t)$ представляет собой с принятой нами точки зрения случайную функцию (случайный процесс), а следовательно, предел (21.1) (если он существует) — случайную величину, которая поэтому не может отождествляться с вероятностью p_k, по сущности своей, не зависящей от случая. С другой стороны, вероятности p_k нами определены как пределы при $t \to \infty$ вероятностей $\mathsf{P}_k(t)$; если все вероятности понимать как средние времена пребывания системы в том или ином состоянии, то величины $\mathsf{P}_k(t)$ не допускают, как легко видеть, никакой разумной интерпретации.

Таким образом, избранное нами определение вероятностей p_k, а также и развитый нами метод их вычисления непосредственно не дают никаких оснований для отождествления их с пределами вида (21.1) вопреки установившейся во всей прикладной литературе практике. Если это отождествление невозможно, то все же с практической точки зрения представляется весьма желательным найти соображения, позволяющие с известным основанием считать интегралы

$$X_k(t) = \frac{1}{T} \int_0^T x_k(t)\, dt$$

при больших значениях T хотя бы приближенно совпадающими с определенными нами вероятностями p_k; если бы это удалось, то такое сближение в значительной степени оправдало бы общепринятую в специальной литературе практику понимания вероятностей p_k как средних относительных времен пребывания — практику, очень удобную в прикладных задачах.

Так как $X_k(t)$ есть с нашей точки зрения случайная величина, то близость ее к не зависящей от случая величине p_k при самых благоприятных обстоятельствах может утверждаться лишь с некоторой (достаточно большой) вероятностью. В настоящем параграфе мы докажем предложение, идущее в указанном направлении так далеко, как только можно было бы надеяться; именно, имеет место

Теорема. *Как бы мало ни было* $\varepsilon > 0$,

$$\lim_{T \to \infty} \mathsf{P}\{|X_k(t) - p_k| > \varepsilon\} = 0.$$

Предварительное замечание. Предложения подобного рода, в которых вероятность некоторого состояния системы, первоначально определенная как доля в большой совокупности систем одинакового строения, сближается затем со средним временем пребывания в данном состоянии для какой-либо одной системы за большой промежуток времени, в теоретической физике называются обычно эргодическими теоремами. Доказываемое нами предложение представляет собой весьма типичный пример эргодической теоремы.

Доказательство. Во всем дальнейшем начальные данные (вероятности $\mathsf{P}_k(0)$, $0 \leqslant k \leqslant n$) предполагаются произвольными, но твердо установленными. Математическое ожидание случайной величины ξ мы будем обозначать через $\mathsf{M}\xi$ или $\mathsf{M}\{\xi\}$. Так как величина $x_k(t)$ может принимать значения 1 и 0 с соответственными вероятностями $\mathsf{P}_k(t)$ и $1-\mathsf{P}_k(t)$, то

$$\mathsf{M}x_k(t)=\mathsf{P}_k(t). \tag{21.2}$$

Так как $\mathsf{P}_k(t) \to p_k$ при $t \to \infty$, то

$$\int_0^T [\mathsf{P}_k(t)-p_k]\,dt = o(T) \quad (T \to \infty).$$

Поэтому при достаточно большом T

$$\mathsf{P}\{|X_k(T)-p_k|>\varepsilon\} = P\left\{\left|\int_0^T [x_k(t)-p_k]\,dt\right|>\varepsilon T\right\} < \\ < \mathsf{P}\left\{\left|\int_0^T [x_k(t)-\mathsf{P}_k(t)]\,dt\right|>\frac{\varepsilon}{2}T\right\}.$$

Отсюда, применяя неравенство Чебышева, находим в силу (21.2)

$$\mathsf{P}\{|X_k(T)-p_k|>\varepsilon\} < \frac{4}{\varepsilon^2T^2}\mathsf{M}\left\{\left[\int_0^T [x_k(t)-\mathsf{P}_k(t)]\,dt\right]^2\right\} = \\ = \frac{4}{\varepsilon^2T^2}\mathsf{M}\left\{\int_0^T\int_0^T [x_k(u)-\mathsf{P}_k(u)][x_k(v)-\mathsf{P}_k(v)]\,du\,dv\right\} = \\ = \frac{4}{\varepsilon^2T^2}\int_0^T\int_0^T [\mathsf{M}\{x_k(u)\,x_k(v)\}-\mathsf{P}_k(u)\,\mathsf{P}_k(v)]\,du\,dv = \\ = \frac{8}{\varepsilon^2T^2}\iint\limits_{0\leqslant u\leqslant v\leqslant T} [\mathsf{M}\{x_k(u)\,x_k(v)\}-\mathsf{P}_k(u)\,\mathsf{P}_k(v)]\,du\,dv.$$

Так как величина $x_k(u)\,x_k(v)$ может принимать лишь значения 1 и 0, то $\mathsf{M}\{x_k(u)\,x_k(v)\}=\mathsf{P}\{x_k(u)\,x_k(v)=1\}$ есть вероятность застать систему в состоянии k как в момент u, так и в момент v; поэтому при $v>u$

$$\mathsf{M}\{x_k(u)\,x_k(v)\}=\mathsf{P}_k(u)\,\mathsf{P}_{kk}(v-u),$$

и мы находим

$$\mathsf{P}\{|X_k(T)-p_k|>\varepsilon\}<\frac{8}{\varepsilon^2T^2}\iint\limits_{0\leqslant u\leqslant v\leqslant T}\mathsf{P}_k(u)\,[\mathsf{P}_{kk}(v-u)-$$

$$-\mathsf{P}_k(v)]\,du\,dv=$$

$$=\frac{8}{\varepsilon^2T^2}\int\limits_0^T\mathsf{P}_k(u)\,du\int\limits_u^T[\mathsf{P}_{kk}(v-u)-\mathsf{P}_k(v)]\,dv=$$

$$=\frac{8}{\varepsilon^2T^2}\int\limits_0^T\mathsf{P}_k(u)\,du\int\limits_0^{T-u}[\mathsf{P}_{kk}(z)-\mathsf{P}_k(z+u)]\,dz.$$

Пусть теперь $\delta>0$ произвольно мало и A настолько велико, что при $z>A$

$$|\mathsf{P}_{ik}(z)-p_k|<\frac{\delta}{2}\qquad(0\leqslant i\leqslant n).$$

Тогда при $z>A$, $u>0$

$$|\mathsf{P}_{kk}(z)-p_k|<\frac{\delta}{2},$$

$$|\mathsf{P}_k(z+u)-p_k|=|\sum_i\mathsf{P}_i(0)\,\mathsf{P}_{ik}(z+u)-p_k|=$$

$$=|\sum_i\mathsf{P}_k(0)\,[\mathsf{P}_{ik}(z+u)-p_k]|\leqslant\frac{\delta}{2}\sum_i\mathsf{P}_i(0)=\frac{\delta}{2},$$

и следовательно,

$$|\mathsf{P}_{kk}(z)-\mathsf{P}_k(z+u)|<\delta.$$

Поэтому при $T>A$

$$\mathsf{P}\{|X_k(T)-p_k|>\varepsilon\}\leqslant$$

$$\leqslant\frac{8}{\varepsilon^2T^2}\int\limits_0^T du\int\limits_0^T|\mathsf{P}_{kk}(z)-\mathsf{P}_k(z+u)|\,dz\leqslant$$

$$\leqslant\frac{8}{\varepsilon^2T^2}\int\limits_0^T du\left\{\int\limits_0^A dz+\int\limits_A^T\delta\,dz\right\}\leqslant\frac{8}{\varepsilon^2T^2}\{AT+\delta T^2\}=\frac{8A}{\varepsilon^2T}+\frac{8\delta}{\varepsilon^2}.$$

Беря сначала δ достаточно малым, выбирая затем A описанным выше образом и беря, наконец, T достаточно большим, мы видим, что правая часть последнего неравенства сколь угодно мала при достаточно большом T. Этим наша теорема доказана.

Глава 7

ЗАДАЧА ЭРЛАНГА ДЛЯ БЕСКОНЕЧНОГО ПУЧКА

§ 22. Уравнение для производящей функции

Если число линий в пучке бесконечно, то рассматриваемую установку нельзя уже причислять к «системам с потерями», так как потери становятся невозможными. Расчет вероятностей различных состояний сохраняет, однако, практическое значение и в этом случае, так как на практике встречаются такие положения, когда потери недопустимы и число линий должно быть достаточно большим для того, чтобы вероятность потери оказалась пренебрегаемо малой; в таких случаях вероятности различных состояний дают возможность оценить степень использования системы, что в свою очередь имеет значение для расчета быстроты износа и других экономических показателей.

Если в формулах Эрланга (20.3) перейти к пределу при $n \to \infty$, то мы получаем

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!} \quad (k = 0,\ 1,\ 2,\ \ldots);$$

можно поэтому предвидеть, что это пуассоновское распределение и даст нам вероятности различных состояний в случае бесконечного пучка. Однако метод, которым мы пришли к формулам (20.3), в случае бесконечного пучка оказывается неприменимым, так как теорема Маркова, на которой он основан, существенным образом предполагает число состояний конечным. Мы покажем, что случай бесконечного пучка может быть очень просто изучен методом производящих функций.

Те рассуждения, которые привели нас в § 20 к системе ($\mathcal{E}$) уравнений Эрланга, сохраняются почти полностью и в случае бесконечного пучка. Различие состоит, очевидно, лишь в том, что группа уравнений системы ($\mathcal{E}$), построенная нами в § 20 для $0<k<n$, теперь имеет место для любого $k>0$, последнее же из уравнений системы ($\mathscr{E}$) отпадает совсем. Таким образом, мы получаем для вероятностей различных состояний в случае бесконечного пучка систему уравнений

$$\left.\begin{aligned} &\mathrm{P}_0'(t)=-\lambda\mathrm{P}_0(t)+\mathrm{P}_1(t);\\ &\mathrm{P}_k'(t)=\lambda\mathrm{P}_{k-1}(t)-(\lambda+k)\,\mathrm{P}_k(t)+(k+1)\,\mathrm{P}_{k+1}(t)\\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (k>0). \end{aligned}\right\}\qquad (\mathscr{E}^*)$$

Эту систему можно считать более простой, чем система ($\mathscr{E}$), так как здесь для всех $k>0$ мы имеем уравнения одинакового типа; именно это обстоятельство и позволяет применить к решению системы ($\mathscr{E}^*$) метод производящих функций (к системе ($\mathscr{E}$) он непосредственно применен быть не может).

Положим

$$\Phi(t,\,x)=\sum_{k=0}^{\infty}\mathrm{P}_k(t)\,x^k$$

[ряд, очевидно, сходится при $|x|\leqslant 1$ и любом t]; положим еще для $k>0$

$$\lambda\mathrm{P}_{k-1}(t)-k\mathrm{P}_k(t)=Q_k(t);$$

тогда система уравнений ($\mathscr{E}^*$) получает более краткий вид:

$$\left.\begin{aligned} &\mathrm{P}_0'(t)=-Q_1(t);\\ &\mathrm{P}_k'(t)=Q_k(t)-Q_{k+1}(t)\quad (k>0). \end{aligned}\right\}$$

Отсюда

$$\begin{aligned} \frac{\partial\Phi}{\partial t}=\sum_{k=0}^{\infty}\mathrm{P}_k'(t)\,x^k=-Q_1(t)+\sum_{k=1}^{\infty}\left[Q_k(t)-Q_{k+1}(t)\right]x^k=\\ =(x-1)\sum_{k=1}^{\infty}Q_k(t)\,x^{k-1}=\\ =(x-1)\left\{\lambda\sum_{k=1}^{\infty}\mathrm{P}_{k-1}(t)\,x^{k-1}-\sum_{k=1}^{\infty}k\mathrm{P}_k(t)\,x^{k-1}\right\}=\\ =(x-1)\left\{\lambda\Phi-\frac{\partial\Phi}{\partial x}\right\}, \end{aligned}$$

или

$$\frac{\partial\Phi}{\partial t}+(x-1)\frac{\partial\Phi}{\partial x}-\lambda(x-1)\,\Phi=0. \qquad (22.1)$$

Это простое уравнение с частными производными первого порядка и послужит нам для определения функции $\Phi(t, x)$. Прежде всего мы его еще несколько упростим преобразованием неизвестной функции. Положим

$$e^{\lambda(x-1)(1-e^{-t})} = G(t, x)$$

и

$$\Phi(t, x) = G(t, x) F(t, x),$$

где $F(t, x)$ — новая неизвестная функция. Мы будем иметь

$$\frac{\partial \Phi}{\partial t} = G\frac{\partial F}{\partial t} + FG\lambda(x-1)e^{-t};$$
$$\frac{\partial \Phi}{\partial x} = G\frac{\partial F}{\partial x} + FG\lambda(1-e^{-t}),$$

вследствие чего

$$\frac{\partial \Phi}{\partial t} + (x-1)\frac{\partial \Phi}{\partial x} - \lambda(x-1)\Phi = G\left\{\frac{\partial F}{\partial t} + \lambda(x-1)e^{-t}F + \right.$$
$$\left. + (x-1)\frac{\partial F}{\partial x} + \lambda(x-1)(1-e^{-t})F - \lambda(x-1)F\right\} =$$
$$= G\left\{\frac{\partial F}{\partial t} + (x-1)\frac{\partial F}{\partial x}\right\},$$

и уравнение (22.1) равносильно уравнению

$$\frac{\partial F}{\partial t} + (x-1)\frac{\partial F}{\partial x} = 0. \tag{22.2}$$

Положим теперь

$$(x-1)e^{-t} = L(t, x)$$

и составим функциональный определитель

$$\frac{D(F, L)}{D(t, x)} = e^{-t}\left\{\frac{\partial F}{\partial t} + (x-1)\frac{\partial F}{\partial x}\right\};$$

уравнение (22.2) равносильно уравнению

$$\frac{D(F, L)}{D(t, x)} = 0,$$

и значит, общим решением его служит соотношение

$$F = R(L) = R[(x-1)e^{-t}],$$

где R — произвольная дифференцируемая функция своего аргумента. Для искомой функции Φ мы отсюда находим выражение

$$\Phi(t, x) = e^{\lambda(x-1)(1-e^{-t})} R[(x-1)e^{-t}]. \qquad (22.3)$$

Это — общее решение уравнения (22.1). Для решения нашей задачи мы должны с помощью начальных данных определить вид функции R.

§ 23. Решение задачи

Пусть в начальный момент $t=0$ мы имеем

$$\mathrm{P}_k(0) = a_k \quad (k=0,\ 1,\ \ldots);$$

в частности, если известно, что при $t=0$ система находится в состоянии i, то $a_i = 1$, $a_k = 0$ $(k \neq i)$; во всех случаях $\sum\limits_{k=0}^{\infty} a_k = 1$. Полагая тогда $t=0$ в выражении (22.3) функции $\Phi(t, x)$, мы находим

$$\Phi(0, x) = \sum_{k=0}^{\infty} \mathrm{P}_k(0)\, x^k = \sum_{k=0}^{\infty} a_k x^k = R(x-1),$$

откуда

$$R(z) = \sum_{k=0}^{\infty} a_k (z+1)^k.$$

В частности,

$$R[(x-1)e^{-t}] = \sum_{k=0}^{\infty} a_k [1 + (x-1)e^{-t}]^k, \qquad (23.1)$$

и для производящей функции мы получаем уже окончательное выражение

$$\Phi(t, x) = e^{\lambda(x-1)(1-e^{-t})} \sum_{k=0}^{\infty} a_k [1 + (x-1)e^{-t}]^k\,^{*)}.$$

Наша задача состоит в том, чтобы, пользуясь этим выражением, найти пределы вероятностей $\mathrm{P}_k(t)$ при $t \to \infty$.

Как показывает соотношение (23.1), величина $R[(x-1)e^{-t}]$ может быть представлена в виде (сходящегося при $t \geqslant 0$,

---

*) Этот ряд заведомо сходится при $t \geqslant 0$, $|x| \leqslant 1$. В самом деле, легко убедиться, что при этих условиях и $|1 + (x-1)e^{-t}| \leqslant 1$.

$|x| \leqslant 1$) ряда по степеням x:

$$R[(x-1)e^{-t}] = \sum_{k=0}^{\infty} b_k(t) x^k.$$

Убедимся теперь, что при $t \to \infty$ мы имеем

$$b_0(t) \to 1, \quad b_k(t) \to 0 \quad (k > 0).$$

В самом деле, в силу (23.1)

$$\sum_{k=0}^{\infty} b_k(t) x^k = R[(x-1)e^{-t}] = \sum_{k=0}^{\infty} a_k (xe^{-t} + 1 - e^{-t})^k; \qquad (23.2)$$

так как $a_k \geqslant 0$, $e^{-t} > 0$ и $1 - e^{-t} \geqslant 0$, то сравнение левой части с правой прежде всего показывает, что все $b_k(t) \geqslant 0$. Далее, полагая в этом равенстве $x = 0$, мы находим

$$b_0(t) = \sum_{k=0}^{\infty} a_k (1 - e^{-t})^k,$$

откуда при $t \to \infty$

$$b_0(t) \to \sum_{k=0}^{\infty} a_k = 1;$$

наконец, полагая в (23.2) $x = 1$, мы находим при любом $t \geqslant 0$

$$\sum_{k=0}^{\infty} b_k(t) = \sum_{k=0}^{\infty} a_k = 1,$$

откуда в силу предыдущего $b_k(t) \to 0$ $(k > 0, t \to \infty)$, и наше утверждение доказано.

В силу (22.3) мы теперь имеем

$$\Phi(t, x) = \sum_{n=0}^{\infty} P_n(t) x^n = R[(x-1)e^{-t}] e^{\lambda(x-1)(1-e^{-t})} =$$

$$= \sum_{k=0}^{\infty} b_k(t) x^k e^{-\lambda(1-e^{-t})} \sum_{r=0}^{\infty} \frac{\lambda^r (1-e^{-t})^r}{r!} x^r,$$

откуда

$$P_n(t) = e^{-\lambda(1-e^{-t})} \sum_{k=0}^{n} b_k(t) \frac{\lambda^{n-k}(1-e^{-t})^{n-k}}{(n-k)!} \quad (n \geqslant 0).$$

Если теперь при фиксированном n заставить t безгранично возрастать, то, как мы видели выше, $b_0(t) \longrightarrow 1$, $b_k(t) \longrightarrow 0$ $(k > 0)$, а потому

$$\mathsf{P}_n(t) \longrightarrow e^{-\lambda} \frac{\lambda^n}{n!} \quad (t \longrightarrow \infty, \; n = 0, 1, 2, \ldots),$$

что полностью решает поставленную задачу. Мы приходим при этом к тем самым формулам, которые мы в начале § 22 в порядке гипотезы получили предельным переходом из формул Эрланга для конечного пучка.

§ 24. Поток с переменным параметром

Методы, изложенные нами в §§ 22 и 23, позволяют легко решить задачу Эрланга для бесконечного пучка и в том случае, когда параметр поступающего потока вызовов меняется с течением времени. Мы уже имели дело с этим случаем в § 5. Будем в дальнейшем и здесь обозначать через $\lambda(t)$ «мгновенное значение» параметра в момент t, определяемое формулой (5.1).

Так как уравнения системы $(\mathscr{E}^*)$ § 22 имеют чисто локальную природу (относятся к некоторому определенному моменту времени t), то проведенный нами вывод этих уравнений остается в полной силе и в случае переменного параметра, только на место постоянного λ становится «мгновенное значение» $\lambda(t)$ этого параметра, вообще говоря, различное в различные моменты времени t. Таким образом, в качестве исходной системы уравнений для определения функций $\mathsf{P}_k(t)$ мы имеем

$$\left.\begin{aligned} &\mathsf{P}_0'(t) = -\lambda(t)\,\mathsf{P}_0(t) + \mathsf{P}_1(t);\\ &\mathsf{P}_k'(t) = \lambda(t)\,\mathsf{P}_{k-1}(t) - [\lambda(t) + k]\,\mathsf{P}_k(t) +\\ &\qquad\qquad + (k+1)\,\mathsf{P}_{k+1}(t) \quad (k > 0). \end{aligned}\right\} \tag{24.1}$$

Полагая

$$\Phi(t, x) = \sum_{k=0}^{\infty} \mathsf{P}_k(t)\, x^k,$$

мы из системы (24.1) в точности как в § 22 находим для производящей функции $\Phi(t, x)$ уравнение с частными производными

$$\frac{\partial \Phi}{\partial t} + (x-1)\frac{\partial \Phi}{\partial x} - \lambda(t)(x-1)\,\Phi = 0, \tag{24.2}$$

отличающееся от уравнения (22.1) только тем, что на месте постоянного параметра λ стоит теперь функция $\lambda(t)$. Это отличие, несущественное для вывода уравнения (24.2), в значительной мере влияет на его решение, заставляя нас обратиться к другой замене искомой функции.

Положим во всем дальнейшем

$$e^{-t}\int_0^t e^u\lambda(u)\,du = \Lambda(t);$$

$$e^{(x-1)\Lambda(t)} = G(t, x);$$

$$\Phi(t, x) = G(t, x)\,F(t, x),$$

где $F(t, x)$ — новая неизвестная функция. Производя эту замену в (24.2), мы легко находим для функции $F(t, x)$ уравнение

$$\frac{\partial F}{\partial t} + (x-1)\frac{\partial F}{\partial x} = 0,$$

в точности совпадающее с уравнением (22.2). В § 22 мы видели, что общим решением этого уравнения служит

$$F(t, x) = R\left[(x-1)\,e^{-t}\right],$$

где R — произвольная дифференцируемая функция своего аргумента. Отсюда

$$\Phi(t, x) = e^{(x-1)\Lambda(t)}R\left[(x-1)\,e^{-t}\right]. \tag{24.3}$$

Теперь мы должны определить вид функции R с помощью начальных данных $P_k(0) = a_k$ $(k = 0, 1, 2, \ldots)$. Мы находим, как в § 23,

$$R(z) = \sum_{k=0}^{\infty} a_k (z+1)^k,$$

откуда, в частности,

$$R\left[(x-1)\,e^{-t}\right] = \sum_{k=0}^{\infty} a_k\left[1 + (x-1)\,e^{-t}\right]^k, \tag{24.4}$$

где ряд заведомо сходится при $t \geqslant 0$, $|x| \leqslant 1$. Отсюда с помощью (24.3) мы получаем окончательное выражение производящей функции

$$\Phi(t, x) = e^{(x-1)\Lambda(t)}\sum_{k=0}^{\infty} a_k\left[1 + (x-1)\,e^{-t}\right]^k. \tag{24.5}$$

В § 23 мы полагали

$$R\left[(x-1)e^{-t}\right]=\sum_{k=0}^{\infty} b_k(t)\,x^k, \tag{24.6}$$

и убедились, что при $t\to\infty$

$$b_0(t)\to 1,\quad b_k(t)\to 0,\quad (k>0).$$

Эти выводы сохраняют силу и в нашем новом случае, так как новая функция $R(z)$ ничем не отличается от прежней.

Мы имеем в силу (24.5), (24.4) и (24.6)

$$\Phi(t,x)=\sum_{n=0}^{\infty}\mathrm{P}_n(t)x^n=e^{(x-1)\Lambda(t)}\sum_{k=0}^{\infty}b_k(t)\,x^k=$$

$$=e^{-\Lambda(t)}\sum_{r=0}^{\infty}\frac{[\Lambda(t)]^r}{r!}x^r\sum_{k=0}^{\infty}b_k(t)\,x^k=$$

$$=e^{-\Lambda(t)}\sum_{n=0}^{\infty}\left\{\sum_{r=0}^{n}\frac{[\Lambda(t)]^r}{r!}b_{n-r}(t)\right\}x^n,$$

и следовательно,

$$\mathrm{P}_n(t)=e^{-\Lambda(t)}\sum_{r=0}^{n}\frac{[\Lambda(t)]^r}{r!}b_{n-r}(t)\qquad(n=0,\,1,\,2,\,\ldots).$$

Допустим теперь, что при $t\to\infty$ параметр $\lambda(t)$ остается ограниченным;

$$\lambda(t)<\Lambda\qquad(t\geqslant 0).$$

Тогда и

$$\Lambda(t)=e^{-t}\int_0^t e^{u}\lambda(u)\,du<\Lambda\qquad(t\geqslant 0);$$

поэтому для любого постоянного n при $t\to\infty$ в сумме

$$\sum_{r=0}^{n}\frac{[\Lambda(t)]^r}{r!}b_{n-r}(t)$$

все члены, кроме последнего ($r=n$), стремятся к нулю, в то время как последний бесконечно мало отличается от $[\Lambda(t)]^n/n!$. Поэтому при $t\to\infty$ и постоянном n

$$\mathrm{P}_n(t)-e^{-\Lambda(t)}\frac{[\Lambda(t)]^n}{n!}\to 0,$$

т. е. закон распределения $P_n(t)$ безгранично приближается к закону Пуассона с (переменным) параметром

$$\Lambda(t) = e^{-t}\int_0^t e^u \lambda(u)\, du.$$

Этим поставленная задача решена.

§ 25. Бесконечный пучок при произвольном законе распределения длин разговоров

В §§ 22—24 мы убедились, что для задачи Эрланга бесконечный пучок в известном смысле представляет собой более простой объект исследования, чем конечный (в частности, в § 24 мы до конца рассмотрели случай входящего потока с переменным параметром — задача, которая, насколько нам известно, еще не решена для конечного пучка). В нашем изложении сравнительная простота трактовки случая бесконечного пучка все время связывалась с возможностью применения метода производящих функций; при этом мы ради соблюдения самой тесной аналогии с теорией конечного пучка исходили всегда из системы дифференциальных уравнений Эрланга. На самом деле, задача Эрланга для бесконечного пучка представляет собой чрезвычайно простую проблему, которая может быть решена и вполне элементарными средствами; при этом оказывается, что показательный закон распределения длин разговоров, служивший важной предпосылкой в методе уравнений Эрланга, при элементарной трактовке задачи без существенных усложнений может быть заменен любым другим законом. Ради этого важного обобщения мы и остановимся в настоящем параграфе на элементарном выводе формул Эрланга для бесконечного пучка.

Мы сохраним все предпосылки § 22 с той разницей, что вероятность $F(x)$ для наудачу выбранного разговора иметь длину $>x$ мы будем предполагать произвольной невозрастающей функцией, подчиненной только требованиям

$$F(0) = 1, \quad F(+\infty) = 0, \quad -\int_0^\infty x\, dF(x) = \int_0^\infty F(x)\, dx = 1,$$

из которых последнее выражает собой соглашение принимать за единицу времени среднюю длительность разговора. Будем, как прежде, обозначать через $\mathsf{P}_k(t)$ вероятность того, что в момент $t>0$ ведется k разговоров (или, что то же, занято k линий). Наша задача — показать, что

$$\lim_{t\to\infty} \mathsf{P}_k(t)=e^{-\lambda}\frac{\lambda^k}{k!} \quad (k=0, 1, 2, \ldots).$$

Пусть $X=(x_1, x_2, \ldots, x_r)$ — произвольный r-мерный вектор, принадлежащий области $D_r(0<x_1<x_2<\ldots<x_r<t)$ r-мерного пространства. Условимся называть **гипотезой** (r, X) событие, состоящее в том, что в промежутке $(0, t)$ происходит r вызовов и моменты $t_i(0<t_1<t_2<\ldots<t_r<t)$ этих вызовов удовлетворяют неравенствам

$$x_i<t_i<x_i+dx_i \qquad (1\leqslant i\leqslant r).$$

Так как поток вызовов — простейший с параметром λ, то вероятность гипотезы (r, X) с точностью до бесконечно малых высших порядков выражается при малых dx_i формулой

$$\mathsf{P}(r, X)= \\ =\lambda e^{-\lambda x_1}dx_1\lambda e^{-\lambda(x_2-x_1)}dx_2\ldots\lambda e^{-\lambda(x_r-x_{r-1})}dx_r e^{-\lambda(t-x_r)},$$

где последний множитель есть вероятность того, что между моментами x_r и t вызовов не поступает. Отсюда

$$\mathsf{P}(r, X)=e^{-\lambda t}\lambda^r dx_1\,dx_2\ldots dx_r \tag{25.1}$$

(и, в частности, не зависит от вектора X). Пусть $\mathsf{P}_{k,(r,X)}(t)$ — условная вероятность застать k разговоров в момент t, если имеет место гипотеза (r, X). Очевидно, $\mathsf{P}_{k,(r,X)}(t)$ может быть отличной от нуля лишь при $r\geqslant k$. Так как для разговора, начавшегося в момент $x\,(0<x<t)$, вероятность не закончиться к моменту t равна $F(t-x)$, то

$$\mathsf{P}_{k,(r,X)}(t)=\sum_C\prod_{i=1}^{k}F(t-x_{s_i})\prod_{\substack{s=1\\ s\neq s_i}}^{r}[1-F(t-x_s)], \tag{25.2}$$

где C — любое сочетание $(s_1, s_2, \ldots, s_k)$ по k чисел ряда $(1, 2, \ldots, r)$ и где суммирование производится по всем таким сочетаниям.

По формуле полной вероятности мы имеем

$$\mathsf{P}_k(t)=\sum_{r,X}\mathsf{P}_r(r,X)\,\mathsf{P}_{k,(r,X)}(t),$$

где по r мы должны суммировать от k до ∞, а по X интегрировать по области D_r. Поэтому мы находим в силу (25.1) и (25.2)

$$\mathsf{P}_k(t)=e^{-\lambda t}\sum_{r=k}^{\infty}\lambda^r\int\limits_{D_r}\sum_C\prod_{i=1}^{k}F(t-x_{s_i})\prod_{\substack{s=1\\ s\neq s_i}}^{r}[1-F(t-x_s)]\,dx_1\ldots dx_r.$$

Подынтегральная функция здесь, очевидно, симметрична относительно переменных $x_1, x_2, \ldots, x_r$. Интеграл поэтому не изменится, если в определении области D_r мы заменим порядок $x_1<x_2<\ldots<x_r$ переменных интеграции любым другим порядком; а так как всех таких порядков возможно $r!$ и так как соединение всех получаемых таким образом областей D_r дает r-мерный куб $K_r[0<x_i<t,\ i=1, 2, \ldots, r]$, то мы можем интегрировать по всему этому кубу с последующим делением на $r!$. Таким образом, мы получаем

$$\mathsf{P}_k(t)=e^{-\lambda t}\sum_{r=k}^{\infty}\frac{\lambda^r}{r!}\sum_C\int\limits_{K_r}\prod_{i=1}^{k}F(t-x_{s_i})\prod_{\substack{s=1\\ s\neq s_i}}^{r}[1-F(t-x_s)]\,dx_1\ldots dx_r.$$

Здесь интеграл по кубу K_r распадается на произведение r простых интегралов и равен

$$\left\{\int_0^t F(t-u)\,du\right\}^k\left\{\int_0^t[1-F(t-u)]\,du\right\}^{r-k}=\left\{\int_0^t F(z)\,dz\right\}^k\left\{\int_0^t[1-F(z)]\,dz\right\}^{r-k}.$$

Полагая

$$\int_t^{\infty}F(z)\,dz=\varepsilon(t),$$

мы имеем

$$\int_0^t F(z)\,dz = 1 - \varepsilon(t), \quad \int_0^t [1 - F(z)]\,dz = t - 1 + \varepsilon(t),$$

и следовательно $\left(\text{так как число сочетаний } C \text{ равно } \binom{r}{k}\right)$,

$$\begin{aligned} \mathrm{P}_k(t) &= e^{-\lambda t} \sum_{r=k}^{\infty} \frac{\lambda^r}{r!} \binom{r}{k} [1 - \varepsilon(t)]^k [t - 1 + \varepsilon(t)]^{r-k} = \\ &= e^{-\lambda t} \frac{\lambda^k}{k!} [1 - \varepsilon(t)]^k \sum_{r=k}^{\infty} \frac{\lambda^{r-k}}{(r-k)!} [t - 1 + \varepsilon(t)]^{r-k} = \\ &= e^{-\lambda t} \frac{\lambda^k}{k!} [1 - \varepsilon(t)]^k e^{\lambda[t - 1 + \varepsilon(t)]} = \\ &= e^{-\lambda} \frac{\lambda^k}{k!} [1 - \varepsilon(t)]^k e^{\lambda \varepsilon(t)}. \end{aligned}$$

Так как $\varepsilon(t) \to 0$ при $t \to \infty$, то отсюда

$$\lim_{t\to\infty} \mathrm{P}_k(t) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad (k = 0, 1, 2, \ldots),$$

что и требовалось доказать*).

Глава 8

ЗАДАЧА ПАЛЬМА

§ 26. Постановка задачи

Для решения задачи Эрланга в гл. 6 и 7 нам было безразлично, рассматривается ли данный пучок линий как упорядоченный или нет. В настоящей главе, напротив, мы будем иметь дело с задачей, которая имеет смысл лишь для упорядоченного пучка. Мы будем, таким образом, все время исходить из предположения, что линии данного пучка перенумерованы и что каждый поступающий вызов занимает линию с наименьшим номером из числа тех, которые свободны в момент его поступления. Очевидно, что при таком порядке

---

*) В этом доказательстве мы неявно предполагали, что поток вызовов начинается с момента 0, т. е. что $\mathrm{P}_0(0) = 1$, $\mathrm{P}_k(0) = 0$ $(k > 0)$. Доказательство лишь немного усложнилось бы при произвольных начальных данных.

средняя загруженность будет различной для различных линий: наиболее загруженной окажется первая линия, за ней вторая, и т. д.

С другой стороны, задачи, рассматриваемые в настоящей главе, таковы, что уже по смыслу их постановки безразлично, имеем ли мы дело с конечным или бесконечным пучком; таким образом, это различение, столь важное, как мы видели, для задачи Эрланга, для целей настоящей главы совершенно несущественно.

Во всем остальном мы сохраняем предпосылки предшествующих глав. Входящий поток вызовов мы предполагаем простейшим с параметром λ. Длительности разговоров предполагаются не зависящими ни друг от друга, ни от каких-либо данных о поступлении вызовов и распределенными по показательному закону со средним значением 1.

Условимся обозначать через L_r линию с номером r. Для всех рассуждений настоящей главы имеет основное значение тот простой и самоочевидный факт, что совокупность линий $L_1, L_2, \ldots, L_r$ (при любом r, не превосходящем общего числа линий пучка) мы можем рассматривать как самостоятельный полнодоступный пучок. Каждый вызов, «потерянный» на этом пучке (т. е. заставший первые r линий занятыми), поступает на линию L_{r+1} (если, конечно, таковая имеется), и обратно — для того, чтобы вызов поступил на L_{r+1}, необходимо, чтобы он был потерян на пучке $(L_1, L_2, \ldots, L_r)$. Вероятность потери на этом пучке есть доля времени, в течение которой все линии $L_1, L_2, \ldots, L_r$ заняты; она, очевидно, совершенно не зависит от того, существуют ли еще линии с более высокими номерами и сколько таких линий; она может быть вычислена по формуле Эрланга для вероятности потери на пучке из r линий и равна

$$\frac{\frac{\lambda^r}{r!}}{\sum_{i=0}^{r} \frac{\lambda^i}{i!}}.$$

В частности, при $r=1$ мы получаем для вероятности потери вызова на линии L_1 выражение

$$\frac{\lambda}{1+\lambda}, \qquad (26.1)$$

а при $r=2$ — для вероятности потери на пучке (L_1, L_2) выражение

$$\frac{\frac{\lambda^2}{2}}{1+\lambda+\frac{\lambda^2}{2}}=\frac{\lambda^2}{\lambda^2+2\lambda+2}. \tag{26.2}$$

Но потерю вызова на пучке (L_1, L_2) мы можем рассматривать как двойное событие: 1) потеря на L_1 и 2) потеря на L_2. Вероятность первого из этих событий равна $\lambda/(1+\lambda)$. Чтобы найти условную вероятность второго события при условии, что первое состоялось, заметим, что это есть вероятность потери на L_2 для вызова, потерянного на L_1 (или, что то же, поступившего на L_2). Если обозначить через λ' интенсивность потока вызовов, поступающих на L_2, то искомая условная вероятность второго события будет поэтому по формуле (26.1) равна $\lambda'/(1+\lambda')$ (так как для этого потока L_2 служит первой линией). Но λ' есть число вызовов, теряющихся на L_1 в единицу времени; так как поступает на L_1 в единицу времени λ вызовов, а доля потерь составляет $\lambda/(1+\lambda)$, то

$$\lambda'=\lambda\frac{\lambda}{1+\lambda}=\frac{\lambda^2}{1+\lambda},$$

и искомая условная вероятность второго события равна

$$\frac{\lambda'}{1+\lambda'}=\frac{\lambda^2}{1+\lambda+\lambda^2}.$$

Для вероятности потери на пучке (L_1, L_2) мы таким образом получаем выражение

$$\frac{\lambda}{1+\lambda}\,\frac{\lambda^2}{1+\lambda+\lambda^2}=\frac{\lambda^3}{(1+\lambda)(1+\lambda+\lambda^2)}, \tag{26.3}$$

отличное от непосредственно даваемого формулами Эрланга выражения (26.2). Следовательно, формула (26.3) для вероятности потери на пучке (L_1, L_2) является ошибочной *). Анализируя цепь рассуждений, приведших нас к этой формуле, мы легко находим источник ошибки. Обозначив через λ' интенсивность потока вызовов, поступающих на L_2, мы

*) Указанием на этот поучительный «парадокс» я обязан В. К. Лезерсону.

приняли, следуя формуле Эрланга (26.1), вероятность потери на L_2 для поступающего на эту линию вызова равной $\lambda'/(1+\lambda')$; тем самым мы неявно допустили, что поток вызовов, поступающих на L_2, является *простейшим*, так как формулы Эрланга установлены нами лишь в этом предположении. То, что мы пришли к неверному результату, доказывает, что это предположение было ошибочным. Мы можем, таким образом, считать установленным, что *поток вызовов, поступающих на L_2 (или, что то же, теряемых на L_1), не есть простейший поток*. Тем более, конечно, у нас нет оснований ожидать, чтобы простейшими оказались потоки вызовов, поступающих на L_3, L_4, ... Этот отрицательного характера вывод поучителен тем, что ясно показывает, насколько важно уже для решения самых элементарных задач не ограничиваться изучением одних только простейших поступающих потоков.

Детальное, до конца идущее исследование природы потока вызовов, поступающих на любую линию L_r данного упорядоченного полнодоступного пучка, представляет собой теоретически интересную и практически важную задачу, решению которой и будет посвящена настоящая глава. Все основные результаты в этом направлении были получены Пальмом [8] в 1943 г.

§ 27. Элементарные расчеты

Как мы уже указали в предыдущем параграфе, для вызова, поступающего на данный пучок (или, что то же, на линию L_1), вероятность оказаться потерянным на пучке (L_1, L_2, ..., L_r) равна по формуле Эрланга

$$E_r = \frac{\frac{\lambda^r}{r!}}{\sum_{k=0}^{r} \frac{\lambda^k}{k!}} \quad (r=1,\ 2,\ \ldots).$$

Очевидно, что вызов теряется на линии L_r тогда и только тогда, когда он теряется на пучке (L_1, L_2, ..., L_r). Поэтому можно также сказать, что число E_r выражает собой вероятность потери на линии L_r; однако при этом необходимо отчетливо иметь в виду, что речь идет о вероятности потери на L_r *для вызова, поступающего на* L_1; вероятность же Π_r потери на L_r *для вызова, поступающего на* L_r, имеет другую величину, которую мы теперь должны постараться найти. Мы можем при этом допустить, что $r>1$, так как, очевидно, $\Pi_1=E_1$.

Вероятность E_r того, что вызов, поступивший на L_1, будет потерян на L_r, очевидно, может быть представлена в виде произведения двух множителей: вероятности E_{r-1} того, что он будет потерян на L_{r-1}, и условной вероятности его потери на L_r, если известно, что он потерян на L_{r-1} (или, что то же, поступил на L_r). Но эта условная вероятность и есть Π_r; поэтому

$$E_r=E_{r-1}\Pi_r,$$

или

$$\Pi_r=\frac{E_r}{E_{r-1}} \qquad (r>1).$$

Несмотря на свою кажущуюся простоту, эта формула для расчетов неудобна тем, что содержит одновременно E_r и E_{r-1}. Поэтому удобнее заменить ее формулой

$$\Pi_r=\frac{\lambda}{r+\lambda E_{r-1}} \qquad (r>1), \tag{27.1}$$

которую мы сейчас докажем. Положим для краткости

$$\sum_{k=0}^{q}\frac{\lambda^k}{k!}=S_q \qquad (q\geqslant 0),$$

так что

$$E_r=\frac{1}{S_r}\frac{\lambda^r}{r!} \qquad (r\geqslant 1);$$

следовательно, при $r>1$

$$\frac{1}{\Pi_r}=\frac{E_{r-1}}{E_r}=\frac{r}{\lambda}\frac{S_r}{S_{r-1}}=\frac{r}{\lambda}\left[1+\frac{\lambda^r/r!}{S_{r-1}}\right]=\frac{1}{\lambda}(r+\lambda E_{r-1}),$$

а это и есть (27.1). При этом необходимо иметь в виду, что λ в формуле (27.1) означает интенсивность первичного пот ка вызовов, поступающих на L_1.

Представляет существенный интерес сравнить между собой вероятности потери на различных линиях при одинаковой интенсивности поступающих на них потоков. Произведенные подсчеты во всех случаях показывают, что эта вероятность возрастает с номером линии. В цитированной нами работе Пальм утверждает, что это непосредственно вытекает из формулы (27.1). Мы не видим, однако, как можно было бы это показать. Более того, нам вообще неизвестно, верно ли это утверждение в его общей формулировке*). Нам удалось доказать в этом направлении только следующее значительно более скромное предложение.

Теорема. *Вероятность потери на L_r при $r>1$ всегда больше, чем вероятность потери на L_1, если поступающие на эти линии потоки имеют одинаковую интенсивность.*

Доказательство. Если интенсивность поступающего на L_1 потока равна λ, то вероятность поступить на L_r для вызова, поступившего на L_1, равна E_{r-1} (так как это есть вероятность потерпеть потерю на линиях $L_1, L_2, \ldots, L_{r-1}$). Поэтому среди λ вызовов, поступающих в среднем на L_1 в единицу времени, на L_r будет в среднем поступать λE_{r-1} вызовов, т. е. поток вызовов, поступающих на L_r, будет иметь интенсивность λE_{r-1}. При этом, как мы знаем, вероятность потери на L_r для вызовов этого потока равна

$$\Pi_r=\frac{\lambda}{r+\lambda E_{r-1}} \qquad (r>1).$$

Если бы поток той же интенсивности λE_{r-1} падал на L_1, то вероятность потери на L_1 для вызовов этого потока согласно формуле (26.1) была бы

$$\frac{\lambda E_{r-1}}{1+\lambda E_{r-1}} \qquad (r>1).$$

Для доказательства нашей теоремы достаточно поэтому убедиться, что при $r>1$ и $\lambda>0$ мы всегда имеем

$$\frac{\lambda E_{r-1}}{1+\lambda E_{r-1}}<\frac{\lambda}{r+\lambda E_{r-1}}, \quad \text{или} \quad r+\lambda E_{r-1}<\lambda+\frac{1}{E_{r-1}},$$

или, наконец,

$$\lambda E_{r-1}+r-\lambda<\frac{1}{E_{r-1}} \qquad (r>1). \tag{27.2}$$

*) Этот факт доказан в статье Т. А. Азларова «Обобщение одной теоремы А. Я. Хинчина», Труды Ташкент. гос. университета им. В. И. Ленина, вып. 189, 113—118, 1961.— *Б. Г.*

Пусть сначала $r=2$, $E_{r-1}=E_1=\frac{\lambda}{1+\lambda}$. Тогда (27.2) получает вид

$$\frac{\lambda^2}{1+\lambda}+2-\lambda<\frac{1+\lambda}{\lambda},$$

что равносильно

$$\frac{\lambda^2}{1+\lambda}+1-\lambda<\frac{1}{\lambda},$$

или

$$\frac{1}{1+\lambda}<\frac{1}{\lambda},$$

и, следовательно, выполняется при любом $\lambda>0$.

Пусть теперь неравенство (27.2) справедливо при каком-нибудь $r>1$ (и любом $\lambda>0$); покажем, что в таком случае оно останется верным и при замене r на $r+1$, т. е. что при любом $\lambda>0$ будет иметь место неравенство

$$\lambda E_r+r+1-\lambda<\frac{1}{E_r}.$$

В силу (27.1) [помня, что $\Pi_r=E_r/E_{r-1}$] мы имеем

$$E_r=\frac{\lambda E_{r-1}}{r+\lambda E_{r-1}}.$$

Отсюда легко находим

$$\lambda E_r+r+1-\lambda-\frac{1}{E_r}=r\left\{\frac{\lambda E_{r-1}+r-\lambda}{\lambda E_{r-1}+r}-\frac{1}{\lambda E_{r-1}}\right\}=$$
$$=\frac{rH_r(\lambda)}{\lambda E_{r-1}(\lambda E_{r-1}+r)},$$

где

$$H_r(\lambda)=\lambda E_{r-1}(\lambda E_{r-1}+r-\lambda)-(\lambda E_{r-1}+r)=$$
$$=\lambda E_{r-1}\left(\lambda E_{r-1}+r-\lambda-\frac{1}{E_{r-1}}\right)-(\lambda E_{r-1}+r-\lambda)=$$
$$=(\lambda E_{r-1}-1)(\lambda E_{r-1}+r-\lambda)-\lambda.$$

Наша теорема будет доказана, если мы убедимся, что $H_r(\lambda)<0$ при любом $\lambda>0$. Положим для краткости

$$\lambda E_{r-1}-1=A,\quad \lambda E_{r-1}+r-\lambda=B,$$

так что

$$H_r(\lambda)=AB-\lambda.$$

Мы можем допустить, что $AB > 0$, так как в противном случае тривиальным образом $H_r(\lambda) < 0$. Если $A > 0$ и $B > 0$, то, так как в силу (27.2) $B < 1/E_{r-1}$, мы имеем

$$AB < \frac{\lambda E_{r-1} - 1}{E_{r-1}} < \lambda;$$

если же $A < 0$, $B < 0$, то, очевидно,

$$|A| < 1, \quad |B| = -B = \lambda - r - \lambda E_{r-1} < \lambda, \quad AB = |A|\,|B| < \lambda.$$

Таким образом, $H_r(\lambda) < 0$ во всех случаях, и наша теорема доказана.

§ 28. Основная теорема Пальма

Мы убедились в § 26, что если на линию L_1 поступает простейший поток вызовов, то поток, поступающий на линию L_r $(r > 1)$, уже не будет простейшим. Важнейшая основная теорема теории Пальма состоит в том, что при этом *на любую линию L_r поступает поток типа* Р, *т. е. стационарный, ординарный и с ограниченным последействием.* От простейшего такой поток отличается, следовательно, только тем, что требование отсутствия последействия заменяется более общим требованием ограниченности последействия.

Для доказательства этой теоремы нам не понадобится никаких расчетов; достаточно лишь более внимательно вглядеться в картину происходящего. Так как для L_1 утверждение теоремы тривиально, то надо только показать, что, если оно верно для L_r, оно остается верным и для L_{r+1} $(r = 1, 2, \ldots)$; а так как вызовы, поступающие на L_{r+1}, совпадают с вызовами, теряющимися на L_r, то мы должны доказать следующее: *если на линию L_r поступает поток вызовов типа* Р, *то потерянные на L_r вызовы также образуют поток типа* Р. При такой постановке задачи, очевидно, самое существование линии L_{r+1} является несущественным.

Обозначим для краткости через A поток вызовов, поступающих на L_r, и через B — поток вызовов, теряющихся на L_r. Течение потока B после произвольно выбранного момента t_0 будет однозначно определено, если станет известно, сколько времени будет еще длиться занимающий линию L_r в момент t_0 разговор, а также каковы моменты поступающих после t_0 вызовов потока A и какова длительность начинаемых этими вызовами разговоров. Но все эти факторы в свою очередь однозначно определяются течением потока A, который есть поток типа Р и, следовательно, стационарен. Поэтому все перечисленные факторы не зависят от выбранного момента t_0, а вместе с ними не зависит от t_0 и дальнейшее течение потока B; другими словами, поток B также стационарен.

Ординарность потока B с самоочевидностью вытекает из того, что он составляет собой часть (ординарного) потока A.

Покажем, наконец, что поток B — с ограниченным последействием. Пусть $t_0=\tau_0=0$; обозначим через t_i $(i=1,2,\ldots)$ моменты следующих за t_0 вызовов потока A и через $\tau_1, \tau_2, \ldots$ — моменты следующих за t_0 вызовов потока B и положим

$$\tau_k-\tau_{k-1}=\zeta_k \quad (k=1,2,\ldots).$$

Наша задача — показать, что закон распределения величины ζ_{k+1} не зависит от значений величин $\zeta_1, \zeta_2, \ldots, \zeta_k$, или, что то же, от значений величин $\tau_1, \tau_2, \ldots, \tau_k$. Но величина ζ_{k+1} будет однозначно определена, если станут известны:

1) расстояния от τ_k до поступающих на L_r после момента τ_k дальнейших вызовов;

2) остаточная длительность разговора, занимающего линию L_r в момент τ_k;

3) длительности разговоров, занимающих линию L_r после момента τ_k.

Все эти три фактора независимы от значений, принимаемых величинами $\tau_1, \tau_2, \ldots, \tau_k$. Для первого это следует из того, что поток A — с ограниченным последействием; для второго — из показательного закона распределения длин разговоров, а для третьего это самоочевидно. Но так как этими тремя факторами значение случайной величины ζ_{k+1} определяется однозначно, то и эта случайная величина не зависит от величин $\tau_1, \tau_2, \ldots, \tau_k$, или, что то же, от величин $\zeta_1, \zeta_2, \ldots$ $\ldots, \zeta_k$. Но это и означает, что поток B — с ограниченным последействием. Таким образом, основная теорема Пальма доказана.

Отсюда, очевидно, следует, что для упорядоченного полнодоступного пучка с простейшим входящим потоком и показательным распределением длин разговоров поток вызовов, поступающих на любую линию L_r этого пучка, представляет собой поток типа Р. Но такой поток (см. § 13) однозначно определяется заданием функции Пальма $\varphi(t)$. Условимся в дальнейшем обозначать через $\varphi_r(t)$ $(r=0, 1, 2, \ldots)$ функцию Пальма для потока вызовов, поступающих на линию L_{r+1} (так что, в частности, $\varphi_0(t)=e^{-\lambda t}$). Мы видим, что наша задача сводится к определению функции $\varphi_r(t)$ для любого $r>0$; этому и будет посвящено все дальнейшее.

§ 29. Вывод основной системы уравнений

Пусть в момент t_0 теряется вызов на линии L_r (или, что то же, поступает вызов на линию L_{r+1}). Тогда $\varphi_r(t)$ есть вероятность того, что в промежутке (t_0, t_0+t) ни один вызов не будет потерян на L_r (не поступит на L_{r+1}). Но это событие может произойти двумя способами:

(A) В промежутке (t_0, t_0+t) на L_r не поступит ни одного вызова.

(B) В промежутке (t_0, t_0+t) на L_r будут поступать вызовы, но ни один из них не будет потерян.

Мы имеем поэтому

$$\varphi_r(t)=\mathsf{P}(A)+\mathsf{P}(B);$$

при этом, в случае $r>0$, $\mathsf{P}(A)=\varphi_{r-1}(t)$, так как, с одной стороны, вызов, потерянный в момент t_0 на L_r, был потерян и на L_{r-1}, а с другой — поток вызовов, поступающих на L_r, совпадает с потоком вызовов, теряемых на L_{r-1}.

Переходим к определению $\mathsf{P}(B)$. Пусть первый вызов, поступающий на L_r после момента t_0, происходит в промежутке (t_0+x, t_0+x+dx); вероятность этого события равна $\varphi_{r-1}(x)-\varphi_{r-1}(x+dx)=-d\varphi_{r-1}(x)$. Для того чтобы этот вызов не был потерян на L_r, очевидно, необходимо и достаточно, чтобы линия L_r до момента t_0+x освободилась от того разговора, которым она была занята в момент t_0; вероятность этого события равна $1-e^{-x}$. Таким образом, вероятность того, что первый после момента t_0 вызов поступит на L_r в промежутке (t_0+x, t_0+x+dx) и что этот вызов не будет потерян, равна

$$-(1-e^{-x})\,d\varphi_{r-1}(x).$$

Мы утверждаем теперь, что если наступило все описанное и если $x<t$, то вероятность того, что в остающемся промежутке (t_0+x, t_0+t) не будет потеряно на L_r ни

одного вызова, равна $\varphi_r(t-x)$. Это вытекало бы непосредственно из определения функции $\varphi_r(t)$, если бы вызов, поступивший на L_r в промежутке $(t_0+x,\ t_0+x+dx)$, был потерян на этой линии; но на самом деле этот вызов по нашему допущению не теряется, так что наше утверждение требует обоснования. Будет ли вызов, поступивший на L_r в момент t_0+x [точнее: в промежутке $(t_0+x,\ t_0+x+dx)$], потерян или нет, во всяком случае, раз этот вызов произошел, линия L_r с момента t_0+x будет занята. Когда она освободится — это в силу показательного закона распределения длин разговоров совершенно не зависит от того, была ли она занята поступившим в момент t_0+x вызовом или была занята ранее (и поступивший в момент t_0+x вызов был потерян). С другой стороны, моменты дальнейших (после t_0+x) поступающих на L_r вызовов зависят от того факта, что такой вызов поступил в момент t_0+x, но совершенно не зависят от судьбы этого вызова (от того, был ли он потерян или нет); не зависят, конечно, от этой судьбы и длины тех разговоров, которые начинаются этими последующими вызовами. Таким образом, ни один из факторов, определяющих собой наличие или отсутствие потерь на L_r в промежутке $(t_0+x,\ t_0+t)$, не зависит от того, какая судьба постигла вызов, поступивший в момент t_0+x. И хотя этот вызов по нашему предположению не был потерян, вероятность того, что в промежутке $(t_0+x,\ t_0+t)$ потерь на линии L_r не будет, такова же, как если бы он был потерян, т. е. равна $\varphi_r(t-x)$. Сопоставляя это с тем, что было установлено ранее, мы приходим к следующему выводу: при $x<t$ вероятность того, что первый после момента t_0 вызов поступит на L_r в промежутке $(t_0+x,\ t_0+x+dx)$ и что между t_0 и t_0+t ни один вызов не будет потерян на L_r, равна

$$-(1-e^{-x})\,d\varphi_{r-1}(x)\,\varphi_r(t-x).$$

Но чтобы получить вероятность события B, мы, очевидно, должны просуммировать все такие вероятности по x от 0 до t. Это дает

$$\mathsf{P}(B)=-\int_0^t(1-e^{-x})\,\varphi_r(t-x)\,d\varphi_{r-1}(x),$$

и следовательно,

$$\varphi_r(t)=\varphi_{r-1}(t)-\int_0^t (1-e^{-x})\,\varphi_r(t-x)\,d\varphi_{r-1}(x)$$

$$(r \geqslant 1). \qquad (29.1)$$

Это и есть исходная система уравнений теории Пальма. Из (29.1) непосредственно видно, что при любом $t>0$

$$\varphi_r(t) \geqslant \varphi_{r-1}(t)$$

— неравенство, которое является очевидным и само по себе.

§ 30. Преобразование Лапласа

С целью решения основной системы уравнений (29.1) мы теперь заменим искомые функции $\varphi_r(t)$ их преобразованиями Лапласа. Положим

$$\psi_r(t)=\int_0^\infty e^{-tx}\varphi_r(x)\,dx \quad (r \geqslant 0).$$

Напишем уравнение (29.1) в виде

$$\varphi_r(x)=\varphi_{r-1}(x)-\int_0^x (1-e^{-z})\,\varphi_r(x-z)\,d\varphi_{r-1}(z),$$

умножим обе части на e^{-tx} и проинтегрируем по x от 0 до ∞. Это дает

$$\psi_r(t)=\psi_{r-1}(t)-\int_0^\infty e^{-tx}\,dx\int_0^x (1-e^{-z})\,\varphi_r(x-z)\,d\varphi_{r-1}(z)=$$

$$=\psi_{r-1}(t)-\int_0^\infty (1-e^{-z})\,e^{-zt}\,d\varphi_{r-1}(z)\int_z^\infty e^{-t(x-z)}\,\varphi_r(x-z)\,dx=$$

$$=\psi_{r-1}(t)-\int_0^\infty (1-e^{-z})\,e^{-zt}\,d\varphi_{r-1}(z)\int_0^\infty e^{-ty}\,\varphi_r(y)\,dy=$$

$$=\psi_{r-1}(t)-\psi_r(t)\int_0^\infty (1-e^{-z})\,e^{-zt}\,d\varphi_{r-1}(z). \qquad (30.1)$$

Интеграция по частям легко дает

$$\int_0^\infty e^{-zt}\,d\varphi_{r-1}(z) = -1 + t\psi_{r-1}(t),$$

и, следовательно,

$$\int_0^\infty e^{-z(t+1)}\,d\varphi_{r-1}(z) = -1 + (t+1)\,\psi_{r-1}(t+1);$$

поэтому из (30.1) мы получаем

$$\psi_r(t) = \psi_{r-1}(t) - \psi_r(t)\left[t\psi_{r-1}(t) - (t+1)\,\psi_{r-1}(t+1)\right],$$

откуда

$$\psi_r(t) = \frac{\psi_{r-1}(t)}{1 + t\psi_{r-1}(t) - (t+1)\,\psi_{r-1}(t+1)}. \tag{30.2}$$

Таким образом, для определения функций $\psi_r(t)$ мы получаем простую рекуррентную формулу. Так как $\varphi_0(x) = e^{-\lambda x}$, то

$$\psi_0(t) = \int_0^\infty e^{-(\lambda+t)x}\,dx = \frac{1}{t+\lambda},$$

и соотношение (30.2) позволяет последовательно определить все функции $\psi_r(t)$. В частности, мы непосредственно видим, что все эти функции рациональны. Однако мы не можем удовлетвориться этим, так как для обратного перехода к функциям $\varphi_r(x)$ нам важно более детально знать свойства рациональных дробей $\psi_r(t)$. В частности, для этого обратного перехода существенное значение имеет разложение функций $\psi_r(t)$ на простые дроби и, следовательно, природа и расположение корней их знаменателей. Этими вопросами мы и должны будем теперь заняться.

Заметим еще, что простым преобразованием искомых функций

$$t\psi_r(t) = \chi_r(t) \quad (r = 0, 1, 2, \ldots)$$

мы можем привести рекуррентные формулы (30.2) к более краткому виду:

$$\chi_r(t) = \frac{\chi_{r-1}(t)}{1 + \chi_{r-1}(t) - \chi_{r-1}(t+1)};$$

впрочем, в дальнейшем мы этим замечанием пользоваться не будем.

§ 31. Определение функций $\psi_r(t)$

Обозначим через $B_r(t)$ $(r = -1, 0, 1, 2, \ldots)$ многочлен степени $r+1$:

$$B_r(t) = \lambda^{r+1} + \sum_{l=0}^{r} \binom{r+1}{l} t(t+1)\ldots(t+r-l)\lambda^l,$$

где $\sum\limits_{l=0}^{-1} = 0$, и следовательно, $B_{-1}(t) = \lambda^0 = 1$.

Лемма. *Многочлены $B_r(t)$ связаны рекуррентной формулой*

$$B_r(t) = tB_{r-1}(t+1) + \lambda B_{r-1}(t) \quad (r = 0, 1, \ldots) \tag{31.1}$$

Доказательство. Мы имеем

$$\lambda B_{r-1}(t) = \lambda^{r+1} + \sum_{l=0}^{r-1} \binom{r}{l} t(t+1)\ \ldots\ (t+r-1-l)\lambda^{l+1};$$

$$tB_{r-1}(t+1) = t\lambda^r + \sum_{l=0}^{r-1} \binom{r}{l} t(t+1)\ldots(t+r-l)\lambda^l =$$

$$= \sum_{l=0}^{r} \binom{r}{l} t(t+1)\ldots(t+r-l)\lambda^l.$$

Так как $\binom{r}{0} = \binom{r+1}{0}$ и, при $l > 0$, $\binom{r}{l-1} + \binom{r}{l} = \binom{r+1}{l}$, то отсюда

$$\lambda B_{r-1}(t) + tB_{r-1}(t+1) =$$

$$= \lambda^{r+1} + \sum_{l=1}^{r} \binom{r+1}{l} t(t+1)\ldots(t+r-l)\lambda^l +$$

$$+ \binom{r}{0} t(t+1)\ldots(t+r) =$$

$$= \lambda^{r+1} + \sum_{l=0}^{r} \binom{r+1}{l} t(t+1)\ldots(t+r-l)\lambda^l = B_r(t),$$

что и доказывает лемму.

Теперь мы можем найти явное выражение для рациональных функций $\psi_r(t)$.

Теорема.

$$\psi_r(t) = \frac{B_{r-1}(t+1)}{B_r(t)} \quad (t = 0, 1, 2, \ldots). \tag{31.2}$$

Доказательство. Так как $B_{-1}(t) = 1$ и, как легко видеть, $B_0(t) = t + \lambda$, то при $r = 0$ доказываемое соотношение (31.2) имеет вид:

$$\psi_0(t) = \frac{1}{t+\lambda},$$

и было нами установлено уже в § 30. Допустим поэтому, что $r > 0$ и соотношение (31.2) уже установлено для $\psi_{r-1}(t)$. Убедимся, что оно остается справедливым и для $\psi_r(t)$; этим, очевидно, теорема будет доказана.

В силу принятого допущения мы имеем

$$1 + t\psi_{r-1}(t) - (t+1)\psi_{r-1}(t+1) = \\ = 1 + \frac{tB_{r-2}(t+1)}{B_{r-1}(t)} - \frac{(t+1)B_{r-2}(t+2)}{B_{r-1}(t+1)} = \\ = \frac{K_r(t)}{B_{r-1}(t)\,B_{r-1}(t+1)},$$

где

$$K_r(t) = B_{r-1}(t+1)\,B_{r-1}(t) + tB_{r-2}(t+1)\,B_{r-1}(t+1) - \\ - (t+1)\,B_{r-2}(t+2)\,B_{r-1}(t). \tag{31.3}$$

Поэтому в силу (30.2)

$$\psi_r(t) = \frac{B_{r-2}(t+1)}{B_{r-1}(t)} \frac{B_{r-1}(t)\,B_{r-1}(t+1)}{K_r(t)} = \\ = \frac{B_{r-2}(t+1)\,B_{r-1}(t+1)}{K_r(t)}.$$

Подлежащее доказательству соотношение (31.2) поэтому равносильно соотношению

$$K_r(t) = B_r(t)\,B_{r-2}(t+1),$$

которое в свою очередь в силу (31.1) равносильно соотношению

$$K_r(t) = tB_{r-1}(t+1)\,B_{r-2}(t+1) + \lambda B_{r-1}(t)\,B_{r-2}(t+1);$$

подставляя вместо $K_r(t)$ его выражение (31.3), мы находим, что доказать надо следующее равенство:

$$B_{r-1}(t+1)B_{r-1}(t)-(t+1)B_{r-2}(t+2)B_{r-1}(t)=\\=\lambda B_{r-1}(t)B_{r-2}(t+1),$$

или

$$B_{r-1}(t+1)=(t+1)B_{r-2}(t+2)+\lambda B_{r-2}(t+1).$$

Но это соотношение мы непосредственно получаем, заменяя в рекуррентной формуле (31.1) r на $r-1$ и t на $t+1$. Таким образом, наша теорема доказана.

Мы видим, что каждая функция $\psi_r(t)$ представляет собой правильную рациональную дробь, числитель которой есть многочлен степени r, а знаменатель — степени $r+1$.

§ 32. Разложение функций $\psi_r(t)$ на простые дроби

Чтобы определить функции $\varphi_r(x)$, лапласовскими преобразованиями которых служат функции $\psi_r(t)$, мы должны теперь посмотреть, как рациональные функции $\psi_r(t)$ разлагаются на простые дроби, и с этой целью исследовать корни многочленов $B_r(t)$, служащих знаменателями этих рациональных функций.

Многочлен $B_r(t)$ степени $r+1$ имеет, как непосредственно видно из его определения, положительные коэффициенты, причем коэффициент при t^{r+1} равен 1. Отсюда уже следует, что все вещественные корни этого многочлена отрицательны и что он может быть представлен в виде:

$$B_r(t)=(t+a_{r0})(t+a_{r1})\ldots(t+a_{rr}),$$

где числа $a_{ri}\,(0\leqslant i\leqslant r)$ положительны или мнимы. Мы утверждаем, что *числа a_{ri} все положительны и что, если они расположены в порядке возрастания, то*

$$a_{ri}>a_{r,\,i-1}+1\quad(1\leqslant i\leqslant r),$$

т. е. расстояние между двумя соседними корнями превосходит единицу.

Докажем это утверждение посредством индукции. Мы уже видели, что $B_0(t)=t+\lambda$, так что для $r=0$ наше утверждение верно. Допустим, что оно верно для $B_r(t)\,(r\geqslant 0)$, и покажем, что в таком случае оно справедливо и для

$B_{r+1}(t)$. Для этого мы воспользуемся рекуррентной формулой (31.1), в силу которой

$$B_{r+1}(t) = tB_r(t+1) + \lambda B_r(t) =$$
$$= t(t+1+a_{r0})\ldots(t+1+a_{rr}) + \lambda(t+a_{r0})\ldots(t+a_{rr});$$

при этом в силу нашего предположения все $a_{rk} > 0$ и $a_{rk} > a_{r,k-1} + 1$ $(1 \leqslant k \leqslant r)$. Эта формула (как, впрочем, и непосредственное определение) показывает, что $B_{r+1}(0) > 0$. С другой стороны, она дает

$$B_{r+1}(-a_{r0}) = -a_{r0}(a_{r1}+1-a_{r0}) \ \ldots \ (a_{rr}+1-a_{r0}) < 0.$$

Это показывает, что $B_{r+1}(t)$ имеет корень между 0 и $-a_{r0}$. Пусть теперь k — одно из чисел ряда $0, 1, \ldots, r-1$. Тогда

$$B_{r+1}(-a_{rk}-1) = \lambda(a_{r0}-a_{rk}-1) \ \ldots \ (a_{r,k-1}-a_{rk}-1) \times$$
$$\times (a_{rk}-a_{rk}-1)(a_{r,k+1}-a_{rk}-1)\ldots(a_{rr}-a_{rk}-1)$$

имеет, как легко подсчитать, $k+1$ отрицательных множителей и, следовательно, знак $(-1)^{k+1}$; напротив

$$B_{r+1}(-a_{r,k+1}) = -a_{r,k+1}(-a_{r,k+1}+1+a_{r0})\ldots$$
$$\ldots(-a_{r,k+1}+1+a_{rk})(-a_{r,k+1}+1+a_{r,k+1})\ldots$$
$$\ldots(-a_{r,k+1}+1+a_{rr})$$

имеет $k+2$ отрицательных множителей и, следовательно, знак $(-1)^k$. Таким образом, многочлен $B_{r+1}(t)$ для любого $k\,(0 \leqslant k \leqslant r-1)$ имеет корень в промежутке между $-a_{rk}-1$ и $-a_{r,k+1}$. Наконец, так как старший член многочлена $B_{r+1}(t)$ есть t^{r+2}, то при отрицательных t, достаточно больших по абсолютной величине, $B_{r+1}(t)$ имеет знак $(-1)^{r+2}$; в то же время в выражении

$$B_{r+1}(-a_{rr}-1) = \lambda(-a_{rr}-1+a_{r0}) \ \ldots \ (-a_{rr}-1+a_{rr})$$

все скобки отрицательны, так что оно имеет знак $(-1)^{r+1}$. Это показывает, что $B_{r+1}(t)$ имеет корень между $-a_{rr}-1$ и $-\infty$.

Сопоставляя все полученное, мы видим, что многочлен $B_{r+1}(t)$ имеет по меньшей мере по одному корню в каждом из $r+2$ промежутков:

$$(0, -a_{r0}), \ (-a_{rk}-1, \ -a_{r,k+1}) \ (0 \leqslant k \leqslant r-1),$$
$$(-a_{rr}-1, \ -\infty),$$

которые попарно не имеют общих точек. Так как $B_{r+1}(t)$ есть многочлен степени $r+2$, то этим его корни исчерпаны; а так как рассмотренные нами $r+2$ промежутков таковы, что расстояние между двумя соседними из них равно единице, то расстояние между двумя соседними корнями многочлена $B_{r+1}(t)$ всегда превосходит единицу. Таким образом, наше утверждение о расположении корней многочленов $B_r(t)$ полностью доказано.

Из этого следует, что разложение функции $\psi_r(t)$ на простые дроби имеет вид:

$$\psi_r(t) = \sum_{k=0}^{r} \frac{C_{rk}}{t+a_{rk}},$$

где числители C_{rk} легко могут быть выражены через числа a_{rk} хорошо известными методами.

§ 33. Заключение

Так как функция $\frac{C_{rk}}{t+a_{rk}}$ служит, как легко непосредственно убедиться, преобразованием Лапласа функции $C_{rk}e^{-a_{rk}x}$, то функция $\psi(t)$, вид которой мы только что нашли, есть преобразование Лапласа функции

$$\sum_{k=0}^{r} C_{rk}e^{-a_{rk}x}. \tag{33.1}$$

Но мы определили $\psi_r(t)$ как преобразование Лапласа искомой функции $\varphi_r(x)$. Можем ли мы отсюда заключить, что $\varphi_r(x)$ совпадает с функцией (33.1)?

Теория преобразований Лапласа (в детали которой мы здесь не можем входить) показывает, что среди функций, обладающих данным преобразованием Лапласа, может быть только одна ограниченная и неотрицательная при $0 \leqslant x < +\infty$; а так как функция $\varphi_r(x)$, очевидно, обладает обоими этими свойствами, то для ее совпадения с функцией (33.1) достаточно убедиться в положительности этой последней; для этого же в свою очередь, очевидно, достаточно показать, что $C_{rk} \geqslant 0$ $(0 \leqslant k \leqslant r)$. В своем цитированном нами исследовании Пальм дает явное выражение чисел C_{rk} через числа a_{rk} и путем анализа этих выражений действительно доказывает положительность всех коэффициентов C_{rk}. Таким образом, мы имеем для всех $r \geqslant 0$

$$\varphi_r(x) = \sum_{k=0}^{r} C_{rk} e^{-a_{rk}x} \quad (0 \leqslant x < +\infty),$$

и задача Пальма может считаться полностью решенной. Мы видим, что для любого r функция Пальма $\varphi_r(x)$, однозначно определяющая собой поток вызовов, теряемых на L_r (т. е. поступающих на L_{r+1}), представляет собой линейную комбинацию $r+1$ показательных функций. В частности, Пальм приводит следующее явное выражение для $\varphi_1(x)$:

$$\varphi_1(x) = \left(\frac{1}{2} + \frac{1}{4\sqrt{\lambda + \frac{1}{4}}}\right) e^{-\left(\lambda + \frac{1}{2} - \sqrt{\lambda + \frac{1}{4}}\right)x} + \\ + \left(\frac{1}{2} - \frac{1}{4\sqrt{\lambda + \frac{1}{4}}}\right) e^{-\left(\lambda + \frac{1}{2} + \sqrt{\lambda + \frac{1}{4}}\right)x}.$$

ЧАСТЬ III

СИСТЕМЫ С ОЖИДАНИЕМ

Глава 9

СЛУЧАЙ ПОКАЗАТЕЛЬНОГО РАСПРЕДЕЛЕНИЯ ДЛИН РАЗГОВОРОВ

§ 34. Вероятности состояний

Мы обращаемся теперь к изучению полнодоступного пучка с ожиданием. Поступившему вызову приходится ждать разговора тогда и только тогда, когда он застает все n линий пучка занятыми. Поэтому здесь наша прежняя «вероятность потери» (т. е. вероятность застать все линии занятыми) может быть названа «вероятностью ожидания». Эта величина понятным образом играет известную роль в оценке качества работы пучка. Однако для систем с ожиданием эта роль сравнительно невелика, так как, если даже значительному большинству вызовов приходится ожидать, обслуживание должно быть признано вполне удовлетворительным во всех тех случаях, когда промежутки ожидания оказываются в своем большинстве очень малыми. Решающую роль здесь играет не частота ожиданий («потерь»), а природа времени ожидания γ как случайной величины; частота же ожиданий дает нам только один штрих этой картины — вероятность неравенства $\gamma > 0$. Понятно поэтому, что конечной целью в исследовании систем с ожиданием всегда служит отыскание *закона распределения времени ожидания* γ.

Для всех задач, которые мы будем рассматривать, безразлично, является ли данный пучок линий упорядоченным или нет. Входящий поток вызовов мы будем всегда предполагать простейшим с параметром λ. Вызовы обслуживаются в порядке их поступления. Длины разговоров всегда будут мыслиться независимыми как друг от друга, так и от течения потока вызовов. Что касается закона распределения этих длин, то именно он составляет собой основной момент различия в задачах теории систем с ожиданием. Обычно бывает так, что при различных распределениях длин разговоров к исследованию времени ожидания приходится подходить различными методами. В настоящей главе мы рассмотрим самый простой случай, когда длины l разговоров подчиняются показательному закону

$$\mathsf{P}\{l>t\}=e^{-\beta t}, \quad t>0, \quad \beta>0 \text{ — постоянная.}$$

Для этого случая полное решение задачи было дано еще Эрлангом [7].

Условимся обозначать через $\mathsf{P}_k(t)$ вероятность того, что в момент t система находится в «состоянии k», т. е. имеется всего k «наличных» (говорящих или ожидающих) вызовов. При $k\leqslant n$ занято k линий, ожидающих нет; при $k>n$ все n линий заняты и имеется $k-n$ ожидающих.

Если $k<n$, то мы находимся в тех же условиях, что и в случае систем с потерями, так как при $k<n$ нет ни потерь, ни ожиданий. Все соображения, приведенные нами в § 20, поэтому остаются в силе и, как там, приводят нас к системе уравнений

$$\begin{aligned} &\mathsf{P}_0'(t)=-\lambda\mathsf{P}_0(t)+\beta\mathsf{P}_1(t);\\ &\mathsf{P}_k'(t)=\lambda\mathsf{P}_{k-1}(t)-(\lambda+k\beta)\mathsf{P}_k(t)+(k+1)\beta\mathsf{P}_{k+1}(t)\\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad (0<k<n) \end{aligned}$$

[ср. ($\mathscr{E}$) § 20, где мы полагали $\beta=1$]. Но последнее уравнение системы ($\mathscr{E}$) должно быть теперь заменено другим, так как теперь возможен переход в состояние n из состояния $n+1$, которое не имело смысла в случае системы с потерями.

Проведем общее рассмотрение для любого $k\geqslant n$. Будем, как прежде, обозначать через $\mathsf{P}_{rs}(\tau)$ вероятность перехода системы из состояния r в состояние s за время τ и понимать знак $\approx$ как равенство с точностью до бесконечно малых вида $o(\tau)$ при $\tau\to 0$. Тогда мы легко находим, аналогично § 20, при $k\geqslant n$ и $\tau\to 0$

$$\mathsf{P}_{k-1,k}(\tau)\approx\lambda\tau, \quad \mathsf{P}_{k+1,k}(\tau)\approx n\beta\tau;$$
$$\mathsf{P}_{kk}(\tau)\approx 1-\lambda\tau-n\beta\tau, \quad \mathsf{P}_{ik}(\tau)\approx 0 \quad (|i-k|>1)$$

[отличие от случая $k<n$ состоит в том, что теперь мы имеем $\mathrm{P}_{k+1,k}(\tau)\approx n\beta\tau$, а не $\approx(k+1)\beta\tau$, как прежде, так как при $k\geqslant n$ в состоянии $k+1$ мы имеем n, а не $k+1$ занятых линий]. Эти оценки с помощью рассуждений, в точности аналогичных проведенным нами в § 20, приводят к соотношению

$$\mathrm{P}_k(t+\tau)\approx\mathrm{P}_{k-1}(t)\lambda\tau+\mathrm{P}_k(t)(1-\lambda\tau-n\beta\tau)+\mathrm{P}_{k+1}(t)n\beta\tau,$$

откуда мы, снова в тесной аналогии с рассуждениями § 20, получаем с помощью предельного перехода

$$\mathrm{P}_k'(t)=\lambda\mathrm{P}_{k-1}(t)-(\lambda+n\beta)\mathrm{P}_k(t)+n\beta\mathrm{P}_{k+1}(t)\qquad(k\geqslant n).$$

Таким образом, в целом система ($\mathscr{E}$) § 20 теперь заменяется (бесконечной) системой уравнений

$$\left.\begin{aligned}
&\mathrm{P}_0'(t)=-\lambda\mathrm{P}_0(t)+\beta\mathrm{P}_1(t);\\
&\mathrm{P}_k'(t)=\lambda\mathrm{P}_{k-1}(t)-(\lambda+k\beta)\mathrm{P}_k(t)+(k+1)\beta\mathrm{P}_{k+1}(t)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad(0<k<n);\\
&\mathrm{P}_k'(t)=\lambda\mathrm{P}_{k-1}(t)-(\lambda+n\beta)\mathrm{P}_k(t)+n\beta\mathrm{P}_{k+1}(t)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad(k\geqslant n).
\end{aligned}\right\}\quad(34.1)$$

Как и в случае систем с потерями, мы принимаем в качестве вероятностей состояний пределы, к которым стремятся вероятности $\mathrm{P}_k(t)$ при $t\to\infty$. В случае систем с потерями мы доказали во всей полноте существование этих пределов [теорема Маркова, § 19]. В случае систем с ожиданием [т. е. в случае уравнений (34.1)] такое доказательство также может быть проведено; однако здесь оно несравненно более сложно и требует существенно новых идей, так как метод Маркова тесно связан с предположением конечного числа возможных состояний системы. Мы не можем поместить этого доказательства здесь и вынуждены ограничиться ссылкой на его возможность *). Необходимо еще отметить, что в нашем новом случае мы, кроме существования пределов величин $\mathrm{P}_k(t)$ при $t\to\infty$, должны еще доказать возможность предельного

*) Мы не знаем в литературе ни одного изложения этого доказательства, хотя сама задача рассматривалась приводимым здесь методом много раз (Эрланг [7], Фрай [4], Колмогоров [2], Феллер [3]). Эрланг вводит возможность предельного перехода как особый постулат; все остальные упомянутые авторы ограничиваются кратким указанием на возможность доказательства.

перехода во всей системе (34.1) — вопрос, который в § 20 у нас не возникал, так как там мы имели дело с конечной системой уравнений.

Итак, мы допускаем, что при $t \to \infty$ существуют пределы

$$\lim_{t \to \infty} \mathsf{P}_k(t) = p_k \quad (k = 0, 1, \ldots)$$

и что соответствующий предельный переход возможен одновременно во всех уравнениях системы (34.1). Так как при этом левые части всех этих уравнений в пределе обращаются в нуль (это доказывается в точности так же, как и в § 20), то мы приходим к системе линейных уравнений

$$\left.\begin{aligned} -\lambda p_0 + \beta p_1 &= 0; \\ \lambda p_{k-1} - (\lambda + k\beta) p_k + (k+1)\beta p_{k+1} &= 0 \quad (0 < k < n); \\ \lambda p_{k-1} - (\lambda + n\beta) p_k + n\beta p_{k+1} &= 0 \quad (k \geqslant n), \end{aligned}\right\} \tag{34.2}$$

из которой, в соединении с нормирующим условием

$$\sum_{k=0}^{\infty} p_k = 1$$

мы и должны определить числа p_k.

Полагая

$$\lambda p_{k-1} - k\beta p_k = z_k \qquad (1 \leqslant k \leqslant n),$$

мы легко находим из системы (34.2)

$$z_1 = 0, \; z_k - z_{k+1} = 0 \quad (1 \leqslant k \leqslant n),$$

откуда

$$z_k = 0 \quad (0 \leqslant k \leqslant n),$$

или

$$p_k = \frac{\lambda}{k\beta} p_{k-1} \quad (0 < k \leqslant n).$$

Это дает, если для краткости положить еще $\lambda / \beta = y$,

$$p_k = \frac{y^k}{k!} p_0 \quad (0 \leqslant k \leqslant n);$$

в частности, при $k = n$ мы находим

$$p_n = \frac{y^n}{n!} p_0. \tag{34.3}$$

Чтобы найти значения p_k при $k>n$, мы обращаемся к последней группе уравнений (34.2). Перепишем их в виде:

$$n\beta(p_{k+1}-p_k)=\lambda(p_k-p_{k-1}) \quad (k\geqslant n),$$

и просуммируем по k от n до $n+r$:

$$n\beta(p_{n+r+1}-p_n)=\lambda(p_{n+r}-p_{n-1}),$$

откуда

$$n\beta p_{n+r+1}+z_n=\lambda p_{n+r},$$

или, так как $z_n=0$, $\lambda/\beta=y$,

$$p_{n+r+1}=\frac{y}{n}p_{n+r} \quad (r\geqslant 0),$$

и следовательно, в силу (34.3)

$$p_{n+r+1}=\left(\frac{y}{n}\right)^{r+1}\frac{y^n}{n!}p_0 \quad (r\geqslant 0).$$

Таким образом, для любого $k\geqslant n$ мы имеем

$$p_k=\left(\frac{y}{n}\right)^{k-n}p_n=\frac{y^k}{n!\,n^{k-n}}p_0. \tag{34.4}$$

Соединяя этот результат с полученным прежде для $k\leqslant n$, мы находим

$$\left.\begin{aligned} p_k&=\frac{y^k}{k!}p_0 \qquad (0\leqslant k\leqslant n);\\ p_k&=\frac{y^k}{n!\,n^{k-n}}p_0 \qquad (k\geqslant n). \end{aligned}\right\} \tag{34.5}$$

Нам остается найти p_0. Нормирующее условие

$$\sum_{k=0}^{\infty}p_k=1$$

дает

$$\frac{1}{p_0}=\sum_{k=0}^{n-1}\frac{y^k}{k!}+\frac{n^n}{n!}\sum_{k=n}^{\infty}\left(\frac{y}{n}\right)^k=s_{n-1}(y)+\frac{y^n}{(n-1)!\,(n-y)},$$

где положено $s_m(y)=\sum_{k=0}^{m}(y^k/k!)$. Вспомним для сравнения, что в случае системы с потерями мы имели (§ 20)

$$\frac{1}{p_0}=s_n(y)=s_{n-1}(y)+\frac{y^n}{n!}.$$

Отметим еще, что вероятность найти все линии занятыми («вероятность ожидания») равна в силу (34.4)

$$\Pi=\sum_{k=n}^{\infty}p_n=\frac{n^n p_0}{n!}\sum_{k=n}^{\infty}\left(\frac{y}{n}\right)^k=\frac{y^n}{n!}\frac{p_0}{1-\frac{y}{n}}. \qquad (34.6)$$

§ 35. Закон распределения времени ожидания

Теперь мы уже легко можем найти вероятность $\mathrm{P}\{\gamma>t\}$ того, что для поступившего в произвольно выбранный момент вызова время ожидания будет больше чем t. Обозначим через $\mathrm{P}_k\{\gamma>t\}$ условную вероятность того же неравенства в предположении, что произведенный вызов застал систему в состоянии k. По формуле полной вероятности мы имеем

$$\mathrm{P}\{\gamma>t\}=\sum_{k=0}^{\infty}p_k\mathrm{P}_k\{\gamma>t\},$$

или, так как, очевидно, $\mathrm{P}_k\{\gamma>t\}=0$ при $k<n$ и $t\geqslant 0$,

$$\mathrm{P}\{\gamma>t\}=\sum_{k=n}^{\infty}p_k\mathrm{P}_k\{\gamma>t\}. \qquad (35.1)$$

Величины p_k нам известны; остается определить величины $\mathrm{P}_k\{\gamma>t\}$ при всех $k\geqslant n$.

Положим $k-n=\nu$ $(\nu=0,\ 1,\ 2,\ldots)$. Наша задача — найти вероятность неравенства $\gamma>t$ при условии, что в момент вызова все линии были заняты и сверх того имелось ν ожидающих. Очевидно, что при этом наш вызов получает разговор после $(\nu+1)$-го освобождения линии. Искомая вероятность есть поэтому вероятность того, что за время t после появления нашего вызова произойдет не более чем ν освобождений линии. Пусть $q_r(t)$ $(0\leqslant r\leqslant\nu)$ есть вероятность того, что за это время произойдет ровно r освобождений;

тогда в силу $k-n=\nu$

$$\mathsf{P}_k\{\gamma>t\}=\sum_{r=0}^{k-n}q_r(t)\qquad(k\geqslant n).$$

Но поток освобождений за время ожидания нашего вызова представляет собой в силу показательного закона распределения длин разговоров простейший поток с параметром $n\beta$, так как вероятность того, что не произойдет ни одного освобождения за время t после такого момента, когда все линии заняты, равна $(e^{-\beta t})^n=e^{-\beta nt}$. Величина $q_r(t)$ есть вероятность того, что за время t наступит r событий этого потока; в силу формул главы 1 поэтому

$$q_r(t)=e^{-n\beta t}\frac{(n\beta t)^r}{r!}\qquad(0\leqslant r\leqslant\nu),$$

и мы находим

$$\mathsf{P}_k\{\gamma>t\}=\sum_{r=0}^{k-n}e^{-n\beta t}\frac{(n\beta t)^r}{r!}\qquad(k\geqslant n).$$

Возвращаясь к формуле (35.1) и используя соотношение (34.4), мы находим

$$\begin{aligned}\mathsf{P}\{\gamma>t\}&=\sum_{k=n}^{\infty}p_k\sum_{r=0}^{k-n}e^{-n\beta t}\frac{(n\beta t)^r}{r!}=\\&=e^{-n\beta t}\sum_{k=n}^{\infty}\left(\frac{y}{n}\right)^{k-n}p_n\sum_{r=0}^{k-n}\frac{(n\beta t)^r}{r!}=\\&=p_ne^{-n\beta t}\sum_{r=0}^{\infty}\frac{(n\beta t)^r}{r!}\sum_{k=n+r}^{\infty}\left(\frac{y}{n}\right)^{k-n}=\\&=p_ne^{-n\beta t}\sum_{r=0}^{\infty}\frac{(n\beta ty)^r}{n^r\ !}\sum_{k=n+r}^{\infty}\left(\frac{y}{n}\right)^{k-n-r}=\\&=\frac{p_ne^{-n\beta t}}{1-\frac{y}{n}}\sum_{r=0}^{\infty}\frac{(\lambda t)^r}{r!}=\frac{p_n}{1-\frac{y}{n}}e^{-(n\beta-\lambda)t},\end{aligned}$$

или, так как в силу (34.3) и (34.6) $p_n=\frac{y^n}{n!}p_0=\Pi\left(1-\frac{y}{n}\right)$

$$\mathsf{P}\{\gamma>t\}=\Pi e^{-(n\beta-\lambda)t}\qquad(t\geqslant 0).$$

Этим наша задача решена. Мы видим, что в принятых нами условиях время ожидания подчиняется показательному закону распределения с параметром $n\beta - \lambda$. Вместе с тем мы получаем

$$\mathsf{P}\{\gamma > 0\} = \Pi,$$

как оно и должно быть: через Π мы как раз обозначили в конце § 34 вероятность застать все линии занятыми («вероятность ожидания»).

Глава 10

ОДНОЛИНЕЙНЫЕ СИСТЕМЫ В СЛУЧАЕ СТАНДАРТНОЙ ДЛИНЫ РАЗГОВОРА

§ 36. Разностно-дифференциальное уравнение задачи

За пределами показательного распределения длин разговоров исследование систем с ожиданием сопряжено с большими трудностями. Простые и законченные результаты здесь удается получить лишь в некоторых частных предположениях. Особенно важным в практическом отношении является случай систем с одной линией (короче: однолинейных систем), для которых задача исследования времени ожидания может быть продвинута весьма далеко при статистических предпосылках широкой общности. Этим случаем мы и будем теперь заниматься до конца книги.

В настоящей главе мы изложим созданную Эрлангом интересную как оригинальностью метода, так и законченностью результатов теорию однолинейных систем в предположении, что все разговоры имеют в точности одну и ту же длину τ. Во всех других отношениях мы сохраняем статистические предпосылки предшествующей главы.

Рассмотрим какие-либо два последовательных вызова. Пусть γ_0 — время ожидания первого вызова, а γ — второго. Обозначим через z расстояние между этими двумя вызовами. Очевидно, что если $z \geqslant \gamma_0 + \tau$, то при появлении второго вызова (единственная) линия свободна и $\gamma = 0$; если же $z < \gamma_0 + \tau$, то при появлении второго вызова линия занята и ему приходится ждать в течение промежутка времени $\gamma = \gamma_0 + \tau - z$. Таким образом, мы находим, что при данном

$z=u$

$$\gamma=\begin{cases} 0 & (\gamma_0+\tau-u\leqslant 0); \\ \gamma_0+\tau-u & (\gamma_0+\tau-u\geqslant 0), \end{cases}$$

или, что то же,

$$\gamma=\max(0,\ \gamma_0+\tau-u).$$

Пусть теперь t — любое положительное число. Убедимся, что при данном $z=u$ неравенства $\gamma<t$ и $\gamma_0<t+u-\tau$ равносильны между собой. В самом деле, если $\gamma<t$, то либо $\gamma=0$, и тогда $\gamma_0\leqslant u-\tau<u+t-\tau$, либо $\gamma=\gamma_0+\tau-u$, и тогда $\gamma_0+\tau-u<t$, и значит, $\gamma_0<t+u-\tau$; обратно, если $\gamma_0<t+u-\tau$, то либо $\gamma_0<u-\tau$, $\gamma_0+\tau-u<0$, $\gamma=0<t$, либо $\gamma_0>u-\tau$, $\gamma=\gamma_0+\tau-u<t$. Таким образом, если P_u означает условную вероятность, вычисленную в предположении, что $z=u$, то мы имеем при любом $u>0$

$$\mathsf{P}_u\{\gamma<t\}=\mathsf{P}_u\{\gamma_0<t+u-\tau\}\quad(t>0);$$

но γ_0 (время ожидания первого вызова), очевидно, не зависит (как случайная величина) от того, когда последует второй вызов, т. е. какое значение получит величина z; поэтому условная вероятность $\mathsf{P}_u\{\gamma_0<t+u-\tau\}$ неравенства $\gamma_0<t+u-\tau$ равна безусловной вероятности $\mathsf{P}\{\gamma_0<t+u-\tau\}$ того же неравенства, и мы получаем

$$\mathsf{P}_u\{\gamma<t\}=\mathsf{P}\{\gamma_0<t+u-\tau\}\quad(t>0).\tag{36.1}$$

Обозначим через $f(t)$ закон распределения величины γ, полагая

$$f(t)=\mathsf{P}\{\gamma<t\}.$$

Так как поток вызовов мы предполагаем простейшим с параметром λ, то вероятность неравенств $u<z<u+du$ (с точностью до бесконечно малых высших порядков) равна $\lambda e^{-\lambda u}\,du$, и мы по формуле полной вероятности находим

$$f(t)=\int_0^\infty \lambda e^{-\lambda u}\,\mathsf{P}_u\{\gamma<t\}\,du,$$

или в силу (36.1) при $t\geqslant 0$

$$f(t)=\int_0^\infty \lambda e^{-\lambda u}\,\mathsf{P}\{\gamma_0<t+u-\tau\}\,du.$$

Но закон распределения величины γ_0 дается той же функцией $f(t)$, что и для величины γ; поэтому мы находим

$$f(t)=\int_0^\infty \lambda e^{-\lambda u} f(t+u-\tau)\,du \quad (t\geqslant 0).$$

Это уравнение и будет служить нам основой для определения искомой функции $f(t)$. Прежде всего, предполагая эту функцию дифференцируемой, мы находим

$$f'(t)=\int_0^\infty \lambda e^{-\lambda u} f'(t+u-\tau)\,du \quad (t\geqslant 0),$$

и интеграция по частям дает

$$f'(t)=\lambda e^{-\lambda u} f(t+u-\tau)\Big|_0^\infty+\lambda^2\int_0^\infty e^{-\lambda u} f(t+u-\tau)\,du=$$
$$=-\lambda f(t-\tau)+\lambda f(t),$$

или

$$f'(t)=\lambda\,[f(t)-f(t-\tau)] \quad (t\geqslant 0).$$

Мы получаем, таким образом, для определения функции $f(t)$ разностно-дифференциальное уравнение простого вида. Это уравнение может быть еще упрощено преобразованием искомой функции

$$f(t)=e^{\lambda t}g(t),$$

что, как легко видеть, дает для новой неизвестной функции уравнение

$$g'(t)=-\lambda e^{-\lambda\tau}g(t-\tau) \quad (t>0). \tag{36.2}$$

§ 37. Закон распределения времени ожидания

Рассмотрим сначала отрезок времени $0<t\leqslant\tau$; в силу $t>0$ при этом имеют место все соотношения, выведенные в § 36. Но при $t\leqslant\tau$ мы имеем $f(t-\tau)=\mathsf{P}\{\gamma<t-\tau\}=0$, и следовательно, соотношение (36.2) дает

$$g'(t)=0,\ g(t)=c=\text{const} \quad (0<t\leqslant\tau),$$

откуда

$$f(t)=ce^{\lambda t} \quad (0<t\leqslant\tau).$$

Чтобы определить постоянную c, заставим в последнем равенстве t стремиться к нулю; в пределе мы получим $c=f(+0)$, но $f(-0)=0$, а потому $c=f(+0)-f(-0)=$ $=\mathsf{P}\{\gamma=0\}$; это есть вероятность того, что поступившему в произвольно выбранный момент вызову не придется ожидать (линия окажется свободной); поэтому $c=1-\alpha$, где α означает вероятность в произвольно выбранный момент застать линию занятой. Иначе говоря, α есть математическое ожидание суммарной длительности всех разговоров, ведущихся в течение единицы времени. Но математическое ожидание числа разговоров в единицу времени равно λ, а длительность каждого разговора равна τ; следовательно,

$$\alpha=\lambda\tau, \quad c=1-\alpha=1-\lambda\tau.$$

Таким образом, окончательно

$$f(t)=(1-\alpha)\,e^{\lambda t} \qquad (\alpha=\lambda\tau,\ 0<t\leqslant\tau),$$

и закон распределения времени ожидания найден нами для отрезка $0<t\leqslant\tau$.

Мы теперь докажем, что *для любого неотрицательного целого числа n в отрезке $n\tau<t\leqslant(n+1)\tau$ функция $g(t)$ определяется формулой*

$$g(t)=(1-\alpha)\sum_{k=0}^{n} e^{-\lambda k\tau}\frac{(k\tau-t)^k}{k!}\lambda^k. \tag{37.1}$$

Так как функция $f(t)$ просто выражается через функцию $g(t)$, то этим мы получим явное выражение для закона распределения времени ожидания.

При $n=0$ формула (37.1) дает

$$g(t)=1-\alpha \quad (0<t\leqslant\tau),$$

что нами доказано выше. Поэтому мы можем доказать формулу (37.1) с помощью индукции. Допустим, что она верна для какого-либо числа $n\geqslant 0$, и покажем, что в таком случае она останется верной и для числа $n+1$.

Итак, пусть соотношение (37.1) верно при $n\tau<t\leqslant$ $\leqslant(n+1)\tau$; если $(n+1)\tau<t\leqslant(n+2)\tau$, то в силу (36.2)

мы будем иметь

$$g'(t) = -\lambda e^{-\lambda\tau} g(t-\tau) =$$

$$= -\lambda e^{-\lambda\tau}(1-\alpha) \sum_{k=0}^{n} e^{-\lambda k\tau} \frac{(k\tau - t + \tau)^k}{k!} \lambda^k;$$

интегрируя это равенство по t от $(n+1)\tau$ до числа, которое мы снова обозначим через t, найдем

$$g(t) - g[(n+1)\tau] =$$

$$= -\lambda e^{-\lambda\tau}(1-\alpha) \sum_{k=0}^{n} \frac{e^{-\lambda k\tau}\lambda^k}{k!} \int_{(n+1)\tau}^{t} [(k+1)\tau - u]^k \, du.$$

Здесь

$$\int_{(n+1)\tau}^{t} [(k+1)\tau - u]^k \, du =$$

$$= -\frac{1}{k+1} \{[(k+1)\tau - t]^{k+1} - [(k-n)\tau]^{k+1}\},$$

а потому

$$g(t) - g[(n+1)\tau] =$$

$$= \lambda e^{-\lambda\tau}(1-\alpha) \sum_{k=0}^{n} \frac{e^{-\lambda k\tau}\lambda^k}{(k+1)!} \{[(k+1)\tau - t]^{k+1} -$$

$$- [(k+1)\tau - (n+1)\tau]^{k+1}\} =$$

$$= (1-\alpha) \sum_{k=0}^{n} \frac{e^{-\lambda(k+1)\tau}\lambda^{k+1}}{(k+1)!} \{[(k+1)\tau - t]^{k+1} -$$

$$- [(k+1)\tau - (n+1)\tau]^{k+1}\} =$$

$$= (1-\alpha) \sum_{r=1}^{n+1} \frac{e^{-\lambda r\tau}\lambda^r}{r!} \{[r\tau - t]^r - [r\tau - (n+1)\tau]^r\} =$$

$$= (1-\alpha) \sum_{r=0}^{n+1} \frac{\lambda^r e^{-\lambda r\tau}}{r!} (r\tau - t)^r -$$

$$- (1-\alpha) \sum_{r=0}^{n+1} \frac{\lambda^r e^{-\lambda\tau r}}{r!} [r\tau - (n+1)\tau]^r, \qquad (37.2)$$

где суммирование по r можно вести от нуля, так как члены с $r=0$ в обоих суммах, очевидно, взаимно уничтожаются.

Формула (37.1) по нашему допущению верна при $n\tau < t \leqslant (n+1)\tau$. Полагая в ней $t=(n+1)\tau$, мы находим

$$g[(n+1)\tau]=(1-\alpha)\sum_{r=0}^{n+1}\frac{\lambda^r e^{-\lambda r\tau}}{r!}[r\tau-(n+1)\tau]^r, \quad (37.3)$$

где суммирование можно вести до $n+1$ потому, что член суммы с $r=n+1$, очевидно, равен нулю. Наконец, складывая между собой равенства (37.2) и (37.3), мы находим

$$g(t)=(1-\alpha)\sum_{r=0}^{n+1}\frac{\lambda^r e^{-\lambda r\tau}}{r!}(r\tau-t)^r$$

при любом t в промежутке $(n+1)\tau < t \leqslant (n+2)\tau$, что и требовалось доказать.

Таким образом, мы получаем для любого $t>0$

$$\mathsf{P}\{\gamma < t\}=f(t)=e^{\lambda t}(1-\lambda\tau)\sum_{k=0}^{n}e^{-\lambda k\tau}\frac{(k\tau-t)^k}{k!}\lambda^k,$$

где целое неотрицательное число n определяется неравенствами

$$n\tau < t \leqslant (n+1)\tau.$$

Глава 11

ОБЩАЯ ТЕОРИЯ ОДНОЛИНЕЙНЫХ СИСТЕМ

§ 38. Постановка задачи и обозначения

В двух предшествующих главах мы, следуя классическим работам Эрланга, нашли закон распределения времени ожидания в двух наиболее простых предположениях относительно закона распределения длин разговоров: в случае показательного распределения (см. гл. 9) и (для однолинейных систем) в случае стандартной длительности разговора (см. гл. 10). Однако можно считать, что на практике мы встречаемся с тем или другим из этих двух простейших распределений лишь в очень редких случаях. В большинстве же приложений мы можем в лучшем случае рассчитывать лишь на некоторое приближение реального распределения длин разговоров к той или другой из наших двух предпосылок; и даже для такого расчета в очень многих случаях мы не имеем в сущности никаких оснований. Поэтому представляется весьма желательным построить метод, позволяющий определить закон распределения времени ожидания (или хотя бы важнейшие его статистические характеристики) при возможно широких предпосылках относительно распределения длин разговоров.

В своей общей постановке эта задача приводит к расчетам, трудно обозримым по своей сложности. Поэтому во всем дальнейшем мы сосредоточим наше внимание на практически весьма важном случае системы с одной линией*). Зато в отношении закона распределения длин разговоров мы ограничимся естественным требованием существования конечного математического ожидания, оставляя этот закон во всем остальном совершенно произвольным. Мы увидим, что при этих предпосылках задача отыскания закона распределения времени ожидания принципиально решается до конца сравнительно несложными приемами.

В течение всей настоящей главы мы будем иметь дело с пучком из одной линии, на которую поступает простейший поток вызовов с параметром λ. Вызовы, заставшие линию занятой, ожидают ее освобождения и занимают ее в порядке их поступления. Длины разговоров не зависят ни друг от друга, ни от числа ожидающих. Вероятность того, что длина произвольно выбранного разговора окажется больше чем t, мы будем обозначать через $F(t)$, так что средняя длитель-

---

*) В последние годы появились многочисленные работы, посвященные однолинейным системам, в которых, с одной стороны, были получены интересные новые общие результаты, а с другой, намечены другие подходы к решению возникающих вопросов. Мы здесь ограничимся указанием лишь на небольшое число статей:

D. J. Kendall. Some problems in the theory of queues. Journ. Royal Stat. Soc., Ser. B, 1951, 151—173.

D. V. Lindley. The theory of queues with a single server, Proc. Camr. Phil. Soc., v. 48, 1952, 277—289.

L. Takács. Investigation of waiting time problems by reduction to Markov processes, Acta Math. Acad. Sci. Hungar, v. 6, 1955, 101—129. — *Б. Г.*

ность разговора будет

$$-\int_0^\infty t\,dF(t)=\int_0^\infty F(t)\,dt=s.$$

Кроме того, мы вводим следующие обозначения, которыми будем пользоваться на протяжении всей главы:

α — вероятность того, что в произвольно выбранный момент времени линия окажется занятой; иначе говоря математическое ожидание суммарного времени занятости линии за 1 час (часом мы будем условно называть принятую единицу времени);

π_k $(k=0, 1, 2, \ldots)$ — вероятность того, что в начале произвольно выбранного разговора мы будем иметь k ожидающих;

$v_k(t)=e^{-\lambda t}(\lambda t)^k/k!$ — вероятность того, что в течение промежутка времени длины t на линию поступит k вызовов;

γ (случайная величина) — время ожидания для вызова, поступившего в произвольно выбранный момент времени.

Дальнейшие обозначения будут объяснены по мере их введения.

§ 39. Вспомогательные предложения

Лемма 1. *Пусть $0<a<b$. Вероятность застать линию в произвольно выбранный момент времени занятой разговором, длина которого заключена между a и b, равна*

$$-\lambda\int_a^b u\,dF(u).$$

Доказательство. Обозначим через $\mathsf{P}_T(a, b)$ вероятность застать линию занятой разговором длины, заключенной между a и b *), в момент времени, произвольно выбранный в промежутке $(0, T)$. Лемма 1 утверждает тогда, что

$$\lim_{T\to\infty}\mathsf{P}_T(a, b)=-\lambda\int_a^b u\,dF(u).$$

Обозначим далее через $L_T^*(a, b)$ суммарную длительность всех разговоров и частей разговоров длины (a, b), веду-

*) В дальнейшем мы такой разговор для краткости будем называть «разговором длины (a, b)».

щихся в промежутке $(0, T)$. $L_T^*(a, b)$ есть случайная величина; если она принимает какое-либо значение l, то соответствующая условная вероятность застать разговор длины (a, b), очевидно, равна l/T. Поэтому в силу формулы полной вероятности

$$\mathsf{P}_T(a, b) = \sum_l \mathsf{P}\{L_T^*(a, b) = l\} \frac{l}{T} = \frac{1}{T} \mathsf{M} L_T^*(a, b),$$

где M — символ математического ожидания.

Для доказательства леммы 1 достаточно поэтому установить, что при $T \to \infty$

$$\lim \frac{1}{T} \mathsf{M} L_T^*(a, b) = -\lambda \int_a^b u\, dF(u). \tag{39.1}$$

Пусть $L_T(a, b)$ означает суммарную длительность разговоров длины (a, b), *начинающихся* в отрезке $(0, T)$. Так как, очевидно,

$$|L_T(a, b) - L_T^*(a, b)| < b,$$

а следовательно, и

$$|\mathsf{M} L_T(a, b) - \mathsf{M} L_T^*(a, b)| < b,$$

то (39.1) равносильно соотношению

$$\frac{1}{T} \mathsf{M} L_T(a, b) \to -\lambda \int_a^b u\, dF(u) \quad (T \to \infty);$$

а так как $\mathsf{M} L_T(a, b)$, очевидно, пропорционально T, то последнее соотношение означает просто, что при любом $T > 0$

$$\frac{1}{T} \mathsf{M} L_T(a, b) = -\lambda \int_a^b u\, dF(u).$$

В частности, это соотношение будет установлено (и, значит, лемма 1 доказана), если мы убедимся, что

$$\lim_{T \to \infty} \frac{1}{T} \mathsf{M} L_T(a, b) = -\lambda \int_a^b u\, dF(u). \tag{39.2}$$

Пусть u — любое положительное число и $T<u$. В промежутке $(0, T)$ может тогда начаться не более одного разговора длины $(u, u+du)$, так что случайная величина $L_T(u, u+du)$ может, кроме значения 0, принять только значение вида $u+\theta\,du$, где $0<\theta<1$; такое значение она примет, очевидно, если (1°) в отрезке $(0, T)$ начнется по меньшей мере один разговор и (2°) последний из начавшихся в $(0, T)$ разговоров будет длины $(u, u+du)$. Вероятность (2°) есть $F(u)-F(u+du)$, а вероятность (1°) при $T\to 0$ имеет вид $\lambda T+o(T)$, где λ — среднее число начал разговоров в единицу времени, совпадающее, очевидно, с параметром входящего потока вызовов. Таким образом, случайная величина $L_T(u, u+du)$ в рассматриваемом нами случае либо равна нулю, либо принимает значение вида $u+\theta\,du$, причем последнее — с вероятностью $[\lambda T+o(T)][F(u)-F(u+du)]$. Таким образом,

$$\mathsf{M}L_T(u, u+du)=(u+\theta\,du)[\lambda T+o(T)][F(u)-F(u+du)],$$

и следовательно, суммируя по u от a до b,

$$\mathsf{M}L_T(a, b)=[-\lambda T+o(T)]\int_a^b u\,dF(u).$$

Отсюда следует (39.2), а значит, и лемма 1.

Лемма 1, таким образом, доказана. При $a=0$, $b=+\infty$ рассматриваемая в ней вероятность, очевидно, есть вероятность α того, что в произвольно выбранный момент времени линия окажется занятой. Мы приходим, таким образом, к важной формуле

$$\alpha=-\lambda\int_0^\infty t\,dF(t)=\lambda s$$

(которую, впрочем, можно было бы усмотреть и непосредственным рассуждением).

Условимся называть *разговором типа* k $(k=0, 1, 2, \ldots)$ разговор, в начале которого имеется k ожидающих, так что введенную нами в § 38 величину π_k можно определить как долю разговоров типа k среди всех ведущихся разговоров (вероятность того, что наудачу выбранный разговор окажется разговором типа k).

Лемма 2. *Если в некоторый момент линия оказалась занятой, то условная вероятность того, что она занята разговором типа k, равна π_k.*

Доказательство. По определению величины π_k среди λ разговоров, ведущихся в среднем в единицу времени, мы будем иметь в среднем $\lambda\pi_k$ разговоров типа k. Так как длительность разговора, как случайная величина, не зависит от числа ожидающих в его начале (т. е. от его типа), то средняя длительность разговора типа k равна s, а следовательно, суммарная длительность разговоров типа k, ведущихся в единицу времени, в среднем равна $\lambda\pi_k s = \alpha\pi_k$; но условная вероятность, о которой идет речь в лемме 2, есть отношение этой средней суммарной длительности ко времени занятости линии, т. е. к α; таким образом, эта условная вероятность равна π_k, и лемма 2 доказана.

Выберем теперь произвольный разговор длины $t>0$ и найдем вероятность $u_k(z)$ $(0<z\leqslant t)$ того, что через промежуток времени z после начала этого разговора число ожидающих будет равно k $(k=0, 1, 2, \ldots)$. Так как длина и тип разговора взаимно независимы, то вероятность того, что в начале выбранного разговора будет r ожидающих, равна π_r; если же это случится, то для того, чтобы по истечении времени z число ожидающих достигло k, необходимо и достаточно, чтобы было $r\leqslant k$ и чтобы за этот промежуток времени z поступило $k-r$ новых вызовов, вероятность чего есть

$$v_{k-r}(z) = e^{-\lambda z}\frac{(\lambda z)^{k-r}}{(k-r)!}.$$

Мы приходим таким образом к следующему предложению.

Лемма 3. *Вероятность $u_k(z)$ того, что через промежуток времени $z\leqslant t$ после начала произвольно выбранного разговора длины t мы будем иметь k ожидающих, равна*

$$u_k(z) = \sum_{r=0}^{k}\pi_r v_{k-r}(z) \quad (k=0, 1, 2, \ldots).$$

Условимся теперь обозначать через τ промежуток времени между произвольно выбранным моментом и окончанием того разговора, который мы в этот момент застали (остаточная длительность случайно застигнутого разговора). Имеет место

Лемма 4. *Пусть $k \geqslant 0$, $x > 0$ и $dx > 0$ мало. Вероятность застать в произвольно выбранный момент k ожидающих и разговор с $x < \tau < x + dx$ равна*

$$-\lambda\, dx \int_x^{\infty} u_k(t-x)\, dF(t) + o(dx).$$

Доказательство. Трудность обоснования леммы 4 заключается в том, что закон распределения случайной величины τ зависит от того, сколько ожидающих мы застаем в выбранный нами момент времени. Если мы обозначим через

a — событие, состоящее в том, что в выбранный момент имеется k ожидающих;

b — событие, состоящее в том, что в выбранный момент ведется разговор с $x < \tau < x + dx$,

то эти события взаимно зависимы; наша задача — определить вероятность $\mathsf{P}(ab)$ их совместного наступления. С этой целью рассмотрим сначала условную вероятность $\mathsf{P}_t(ab)$ той же пары событий при условии, что застигнутый разговор имеет длину t. Мы можем записать в легко понятных обозначениях

$$\mathsf{P}_t(ab) = \mathsf{P}_t(b)\, \mathsf{P}_{tb}(a). \tag{39.3}$$

Так как выбранный нами момент, если он застал разговор длины t, с одинаковой вероятностью мог попасть в любой момент этого разговора, то

$$\mathsf{P}_t(b) = \begin{Bmatrix} dx/t & (t > x); \\ 0 & (t \leqslant x). \end{Bmatrix} \tag{39.4}$$

С другой стороны событие b равносильно событию

$$t - x - dx < t - \tau < t - x,$$

где $t - \tau$ есть возраст застигнутого разговора. Таким образом, $\mathsf{P}_{tb}(a)$ есть условная вероятность застать в произвольно выбранный момент k ожидающих, если известно, что в этот момент мы застали разговор длины t и возраста, заключенного между $t - x - dx$ и $t - x$. Согласно определению функции $u_k(x)$ (см. лемму 3) эта вероятность при $dx \to 0$ асимптотически равна $u_k(t-x)$ *).

Таким образом, в силу (39.3) и (39.4) мы имеем

$$P_t(ab) = \begin{cases} u_k(t-x)\,dx/t + o(dx) & (t > x); \\ 0 & (t \leqslant x). \end{cases}$$

Такова условная вероятность совместного наступления событий a и b *при условии, что застигнут разговор длины* t. Нам остается теперь освободиться от этого условия. Согласно лемме 1 закон распределения длины застигнутого разговора есть

$$G(t) = -\lambda \int_t^\infty u\,dF(u),$$

и мы имеем по формуле полной вероятности

$$\begin{aligned} P(a, b) &= -\int_0^\infty P_t(a, b)\,dG(t) = \\ &= -\int_x^\infty u_k(t-x)\frac{dx}{t}\,dG(t) + o(dx) = \\ &= -\lambda\int_x^\infty u_k(t-x)\frac{dx}{t}\,t\,dF(t) + o(dx) = \\ &= -\lambda\,dx\int_x^\infty u_k(t-x)\,dF(t) + o(dx), \end{aligned}$$

что и требовалось доказать.

Лемма 5.

$$-\int_0^\infty u_{k+1}(x)\,dF(x) = \pi_k \quad (k > 0);$$

$$-\int_0^\infty u_1(x)\,dF(x) = \pi_0\left\{1 + \int_0^\infty e^{-\lambda x}\,dF(x)\right\}.$$

---

*) Непрерывность функции $u_k(z)$ вытекает из леммы 3.

Доказательство. По смыслу функции $u_{k+1}(x)$ [см. формулировку леммы 3] интеграл

$$-\int_0^\infty u_{k+1}(x)\,dF(x)$$

означает вероятность того, что в конце произвольно выбранного разговора мы будем иметь $k+1$ ожидающих. В случае $k>0$ это равносильно тому, что следующий разговор начнется при k ожидающих, вероятность чего равна π_k. Этим доказано первое из двух утверждаемых равенств.

В случае $k=0$ вероятность того, что выбранный разговор окончится при одном ожидающем, будет *меньше* чем π_0;-дело в том, что разговор без ожидающих может начаться не только при окончании предшествующего разговора с одним ожидающим, но и другим способом — появлением вызова в момент, когда линия свободна. Поэтому

$$-\int_0^\infty u_1(x)\,dF(x)=\pi_0-\varrho,$$

где $\varrho>0$. Для определения ϱ мы сложим между собой все установленные нами соотношения, замечая при этом, что

$$\sum_{k=0}^\infty \pi_k=1 \quad \text{и} \quad \sum_{k=0}^\infty u_k(x)=1$$

при любом x; поэтому мы получаем

$$-\int_0^\infty [1-u_0(x)]\,dF(x)=1-\varrho;$$

а так как в силу леммы 3 $u_0(x)=\pi_0 v_0(x)=\pi_0 e^{-\lambda x}$, то отсюда

$$\varrho=-\pi_0\int_0^\infty e^{-\lambda x}\,dF(x), \tag{39.5}$$

и лемма 5 доказана полностью.

Лемма 6. *При* $|a|\leqslant 1$, $z>0$

$$\sum_{k=0}^\infty a^k u_k(z)=e^{\lambda z(a-1)}\sum_{k=0}^\infty \pi_k a^k.$$

Доказательство. В силу леммы 3

$$\sum_{k=0}^{\infty} a^k u_k(z) = \sum_{k=0}^{\infty} a^k \sum_{r=0}^{k} \pi_r e^{-\lambda z} \frac{(\lambda z)^{k-r}}{(k-r)!} =$$

$$= e^{-\lambda z} \sum_{r=0}^{\infty} \pi_r \sum_{k=r}^{\infty} a^k \frac{(\lambda z)^{k-r}}{(k-r)!} = e^{-\lambda z} \sum_{r=0}^{\infty} \pi_r a^r \sum_{k=r}^{\infty} \frac{(a\lambda z)^{k-r}}{(k-r)!} =$$

$$= e^{\lambda z(a-1)} \sum_{k=0}^{\infty} \pi_k a^k,$$

что и требовалось доказать.

Обозначим через $\psi(\xi)$ характеристическую функцию длительности разговора, т. е. положим

$$\psi(\xi) = -\int_0^{\infty} e^{i\xi t}\, dF(t);$$

положим, далее, для любого вещественного ξ

$$\chi(\xi) = \sum_{k=0}^{\infty} \pi_k [\psi(\xi)]^k;$$

наконец, пусть

$$\eta = \eta(\xi) = \frac{\lambda}{i}[\psi(\xi) - 1].$$

Лемма 7.

$$\chi(\xi) = (1-\alpha)\frac{1-\psi(\xi)}{\psi(\eta) - \psi(\xi)}.$$

Доказательство. В силу леммы 5 мы имеем [определяя ϱ формулой (39.5)]

$$\chi(\xi) = \sum_{k=0}^{\infty} [\psi(\xi)]^k \pi_k = -\sum_{k=0}^{\infty} [\psi(\xi)]^k \int_0^{\infty} u_{k+1}(x)\, dF(x) + \varrho =$$

$$= \varrho - \frac{1}{\psi(\xi)} \int_0^{\infty} \left\{ \sum_{k=1}^{\infty} [\psi(\xi)]^k u_k(x) \right\} dF(x),$$

откуда в силу леммы 6

$$\chi(\xi)=\varrho-\frac{1}{\psi(\xi)}\int_0^\infty\left\{e^{\lambda x[\psi(\xi)-1]}\sum_{k=0}^\infty \pi_k[\psi(\xi)]^k-u_0(x)\right\}dF(x)=$$

$$=\varrho-\frac{1}{\psi(\xi)}\int_0^\infty\{e^{\lambda x[\psi(\xi)-1]}\chi(\xi)-\pi_0 e^{-\lambda x}\}\,dF(x)=$$

$$=\varrho-\frac{\chi(\xi)}{\psi(\xi)}\int_0^\infty e^{\lambda x[\psi(\xi)-1]}\,dF(x)+\frac{1}{\psi(\xi)}\int_0^\infty \pi_0 e^{-\lambda x}\,dF(x)=$$

$$=\varrho-\frac{\chi(\xi)}{\psi(\xi)}\int_0^\infty e^{i\eta x}\,dF(x)-\frac{\varrho}{\psi(\xi)}=\varrho\frac{\psi(\xi)-1}{\psi(\xi)}+\frac{\chi(\xi)\psi(\eta)}{\psi(\xi)};$$

$$\chi(\xi)\frac{\psi(\xi)-\psi(\eta)}{\psi(\xi)}=\varrho\,\frac{\psi(\xi)-1}{\psi(\xi)};$$

$$\chi(\xi)=\varrho\,\frac{1-\psi(\xi)}{\psi(\eta)-\psi(\xi)}.$$

Для доказательства леммы 7 остается показать, что $\varrho=1-\alpha$. Но ϱ не зависит от ξ, а так как $\chi(0)=1$, то из последней формулы

$$\varrho=\lim_{\xi\to 0}\frac{\psi(\eta)-\psi(\xi)}{1-\psi(\xi)},$$

или, по правилу Лопиталя,

$$\varrho=\lim_{\xi\to 0}\frac{\psi'(\eta)\dfrac{d\eta}{d\xi}-\psi'(\xi)}{-\psi'(\xi)}=1-\lim_{\xi\to 0}\frac{\psi'(\eta)}{\psi'(\xi)}\frac{d\eta}{d\xi};$$

но $\psi'(0)=is$, $\dfrac{d\eta}{d\xi}=\dfrac{\lambda}{i}\psi'(\xi)\longrightarrow\lambda s=\alpha$ при $\xi\longrightarrow 0$, и мы действительно находим $\varrho=1-\alpha$.

§ 40. Характеристическая функция времени ожидания

Целью нашего исследования является отыскание закона распределения случайной величины γ — времени ожидания для вызова, производимого в произвольно выбранный момент времени. Так как всякий закон распределения однозначно определяется соответствующей характеристической функцией, а отыскание важнейших свободных характеристик случайной величины по ее характеристической функции обычно бывает проще, чем по ее закону распределения, то мы естественно будем считать нашу цель достигнутой, если нам удастся найти характеристическую функцию $\varphi(\xi)$ величины γ, т. е. математическое ожидание величины $e^{i\gamma\xi}$ как функцию вещественного параметра ξ. К этому и будут направлены наши усилия.

Пусть в дальнейшем M есть символ математического ожидания, так что $\varphi(\xi) = \mathsf{M}e^{i\gamma\xi}$. Мы будем различать следующие возможности. Во-первых, в произвольно выбранный момент мы можем застать линию свободной; вероятность этого равна $1-\alpha$, и в этом случае наверняка $\gamma = 0$, $e^{i\gamma\xi} = 1$. Во-вторых, мы можем застать линию занятой при k ожидающих ($k = 0, 1, 2, \ldots$); вероятность этого мы обозначим через $\mathsf{P}(k)$, а математические ожидания, вычисленные при этом условии, будем обозначать через M_k. Очевидно, мы имеем

$$\varphi(\xi) = \mathsf{M}e^{i\gamma\xi} = 1 - \alpha + \sum_{k=0}^{\infty} \mathsf{P}(k)\,\mathsf{M}_k e^{i\gamma\xi}.$$

Если наш вызов застал линию занятой при k ожидающих, то время ожидания γ для него слагается из двух частей: 1) остаточная длительность τ того разговора, который ведется в момент появления нашего вызова, и 2) суммарная длительность T разговоров тех k ожидающих, которых наш вызов застал при своем появлении. Мы имеем поэтому

$$\gamma = \tau + T, \quad \mathsf{M}_k e^{i\gamma\xi} = \mathsf{M}_k (e^{i\tau\xi} e^{iT\xi}).$$

Но очевидно, что *при данном* k величины τ и T взаимно независимы. Поэтому мы имеем

$$\varphi(\xi) = 1 - \alpha + \sum_{k=0}^{\infty} \mathsf{P}(k)\,\mathsf{M}_k e^{i\tau\xi}\,\mathsf{M}_k e^{iT\xi}.$$

Величина $\mathsf{M}_k e^{iT\xi}$ вычисляется непосредственно. В самом деле, величина T есть сумма k взаимно независимых случайных величин, распределенных по закону $F(x)$ с характеристической функцией $\psi(\xi)$. Поэтому

$$\mathsf{M}_k e^{iT\xi} = [\psi(\xi)]^k \quad (k = 0, 1, 2, \ldots)$$

и мы находим

$$\varphi(\xi) = 1 - \alpha + \sum_{k=0}^{\infty} [\psi(\xi)]^k \mathsf{P}(k)\, \mathsf{M}_k e^{i\tau\xi}.$$

Обозначим через $Q_k(x)$ закон распределения величины τ (вероятность неравенства $\tau > x$) при k ожидающих. Тогда

$$\mathsf{M}_k e^{i\tau\xi} = -\int_0^{\infty} e^{i\xi x}\, dQ_k(x);$$

$$\mathsf{P}(k)\, \mathsf{M}_k e^{i\tau\xi} = -\int_0^{\infty} e^{i\xi x}\, \mathsf{P}(k)\, dQ_k(x).$$

Но величина

$$\mathsf{P}(k)\,[Q_k(x) - Q_k(x+dx)]$$

есть, очевидно, вероятность того, что наш вызов застанет линию занятой при k ожидающих и что при этом будет $x < \tau < x + dx$. Эта вероятность в силу леммы 4 с точностью до малых высших порядков при $dx \to 0$ равна

$$-\lambda\, dx \int_x^{\infty} u_k(t-x)\, dF(t).$$

Поэтому мы получаем

$$\mathsf{P}(k)\, \mathsf{M}_k e^{i\xi\tau} = -\lambda \int_0^{\infty} e^{i\xi x}\, dx \int_x^{\infty} u_k(t-x)\, dF(t),$$

и следовательно,

$$\varphi(\xi) = 1 - \alpha - \lambda \int_0^{\infty} e^{i\xi x}\, dx \int_x^{\infty} \left\{ \sum_{k=0}^{\infty} [\psi(\xi)]^k u_k(t-x) \right\} dF(t),$$

откуда в силу леммы 6

$$\varphi(\xi) = 1 - \alpha - \lambda \int_0^{\infty} e^{i\xi x}\, dx \int_x^{\infty} e^{\lambda(t-x)[\psi(\xi)-1]} \sum_{k=0}^{\infty} \pi_k [\psi(\xi)]^k\, dF(t) =$$

$$= 1 - \alpha - \lambda\chi(\xi) \int_0^{\infty} e^{i\xi x}\, dx \int_x^{\infty} e^{\lambda(t-x)[\psi(\xi)-1]}\, dF(t),$$

где положено, как в § 39, $\sum_{k=0}^{\infty} \pi_k [\psi(\varepsilon)]^k = \chi(\xi)$. Если мы еще в согласии с § 39 положим

$$\frac{\lambda}{i}[\psi(\xi)-1]=\eta=\eta(\xi),$$

то получим

$$\varphi(\xi)=1-\alpha-\lambda\chi(\xi)\int_0^\infty e^{i(\xi-\eta)x}\,dx\int_x^\infty e^{i\eta t}\,dF(t)=$$

$$=1-\alpha-\lambda\chi(\xi)\int_0^\infty e^{i\eta t}\,dF(t)\int_0^t e^{i(\xi-\eta)x}\,dx=$$

$$=1-\alpha-\lambda\chi(\xi)\int_0^\infty e^{i\eta t}\,dF(t)\,\frac{e^{i(\xi-\eta)t}-1}{i(\xi-\eta)}=$$

$$=1-\alpha-\frac{\lambda\chi(\xi)}{i(\xi-\eta)}\int_0^\infty [e^{i\xi t}-e^{i\eta t}]\,dF(t)=$$

$$=1-\alpha+\frac{\lambda}{i}\chi(\xi)\frac{\psi(\xi)-\psi(\eta)}{\xi-\eta}.$$

Наконец, выражая $\chi(\xi)$ согласно лемме 7, получаем

$$\varphi(\xi)=1-\alpha+(1-\alpha)\frac{\lambda}{i}\frac{\psi(\xi)-1}{\xi-\eta}=$$

$$=(1-\alpha)\left\{1+\frac{\lambda}{i}\frac{\psi(\xi)-1}{\xi-\frac{\lambda}{i}[\psi(\xi)-1]}\right\}=$$

$$=(1-\alpha)\frac{\frac{i\xi}{\lambda}}{\frac{\xi}{\lambda}-[\psi(\xi)-1]}=\frac{1-\alpha}{1-\alpha\frac{\psi(\xi)-1}{is\xi}}.$$

Полученная таким образом формула

$$\varphi(\xi)=\frac{1-\alpha}{1-\alpha\frac{\psi(\xi)-1}{is\xi}} \tag{40.1}$$

может считаться полным решением поставленной задачи, так как характеристическая функция $\varphi(\xi)$ искомого закона распределения времени ожидания нами выражена через данные постоянные s и $\alpha = \lambda s$ и через характеристическую функцию $\psi(\xi)$ данного закона распределения $F(t)$ длин разговоров.

В качестве первой иллюстрации формулы (40.1) рассмотрим, какое выражение с ее помощью получает среднее значение $\bar{\gamma}$ времени ожидания. Мы имеем в виду $\varphi(0) = 1$

$$i\bar{\gamma} = \varphi'(0) = \frac{\varphi'(0)}{\varphi(0)};$$

а так как в силу (40.1)

$$\varphi(\xi) = \frac{(1-\alpha)\, is\xi}{is\xi - \alpha[\psi(\xi) - 1]},$$

то

$$\frac{\varphi'(\xi)}{\varphi(\xi)} = \frac{d\ln\varphi(\xi)}{d\xi} = \frac{1}{\xi} - \frac{is - \alpha\psi'(\xi)}{is\,\xi - \alpha[\psi(\xi) - 1]} =$$

$$= \frac{\alpha\{\xi\psi'(\xi) - [\psi(\xi) - 1]\}}{\xi\{is\,\xi - \alpha[\psi(\xi) - 1]\}}.$$

Предел этого выражения при $\xi \to 0$ легко находится по правилу Лопиталя и равен

$$\frac{\alpha\psi''(0)}{2(1-\alpha)\, is}.$$

Величина

$$\psi''(0) = \int_0^\infty t^2\, dF(t) = -s_2$$

есть взятое с обратным знаком математическое ожидание квадрата длины разговора. Мы находим таким образом

$$i\bar{\gamma} = -\frac{\alpha}{2i(1-\alpha)}\frac{s_2}{s},$$

или

$$\bar{\gamma} = \frac{\alpha}{2(1-\alpha)}\frac{s_2}{s}.$$

Эта простая формула, между прочим, показывает, что при данной загрузке линии α и при данной средней длине разговора s время ожидания будет в среднем тем меньше, чем меньше дисперсия длин разговоров, т. е. чем более стандартизованы эти длины.

В качестве второй иллюстрации рассмотрим простейший случай показательного распределения

$$F(t)=e^{-\beta t}$$

длин разговоров. Мы легко находим в этом случае

$$s=\frac{1}{\beta}, \quad \psi(\xi)=\frac{\beta}{\beta-i\xi}=\frac{1}{1-i\xi s},$$

откуда

$$\frac{\psi(\xi)-1}{is\xi}=\frac{1}{1-i\xi s}=\psi(\xi),$$

и, значит, в силу (40.1)

$$\varphi(\xi)=(1-\alpha)\frac{1-i\xi s}{1-\alpha-i\xi s}=1-\alpha+\alpha\frac{1-\alpha}{1-\alpha-i\xi s}. \quad (40.2)$$

Но

$$\frac{1-\alpha}{1-\alpha-i\xi s}=\frac{\frac{1-\alpha}{s}}{\frac{1-\alpha}{s}-i\xi}$$

есть характеристическая функция показательного закона распределения с параметром $(1-\alpha)/s$. Поэтому (40.2) показывает, что закон распределения времени ожидания есть

$$\mathsf{P}\{\gamma\geqslant t\}=(1-\alpha)E(t)+\alpha e^{-\frac{1-\alpha}{s}t} \quad (t\geqslant 0),$$

где

$$E(t)=\begin{cases} 1 \ (t\leqslant 0); \\ 0 \ (t>0). \end{cases}$$

Это согласуется с формулой

$$\mathsf{P}\{\gamma>t\}=\Pi e^{-(n\beta-\lambda)t},$$

полученной нами в конце главы 10 для случая n линий. В самом деле, при $n=1$ мы имеем $\Pi=\alpha=\lambda s$, $n\beta-\lambda=$ $=\frac{1}{s}-\lambda=\frac{1-\alpha}{s}$. Член же $(1-\alpha)E(t)$ при $t>0$ равен нулю, а при $t=0$ обращается в $1-\alpha=\mathsf{P}\{\gamma=0\}$.

ЛИТЕРАТУРНЫЕ УКАЗАНИЯ

(числа в прямых скобках относятся к списку литературы в конце настоящих указаний на стр. 148).

§ 1. Эти элементарные сведения помещаются почти во всех современных курсах теории вероятностей; см., в частности, Гнеденко [1], Феллер [3], Фрай [4], Хинчин [5], Erlang [7].

§ 2. По-видимому, публикуется впервые.

§ 3, 4. См. § 1.

§ 5. См. Фрай [4].

§ 6. В этой форме публикуется впервые.

§ 7. Публикуется впервые.

§ 8. См. Redheffer [9], где в другой форме решается та же задача.

§ 9, 10. Относительно функции $\varphi_0(t)$ см. Palm [8]. Относительно функций $\varphi_k(t)$ $(k > 0)$ публикуется впервые.

§ 11. Публикуется впервые.

§ 12. Публикуется впервые.

§ 13. Понятие потока с ограниченным последействием принадлежит Пальму [8]. В остальном содержание параграфа публикуется впервые.

§ 14. См. Palm [8].

§ 15, 16. Публикуется впервые.

§ 17, 18, 19, 20. См. Феллер [3], Фрай [4], Erlang [7].

§ 21. Эта эргодическая теорема является частным случаем известных общих теорем эргодической теории. В такой редакции, по-видимому, публикуется впервые.

§ 22, 23. Erlang [7].

§ 24. Публикуется впервые. Частные случаи см. Palm [8].

§ 25. Публикуется впервые.

§ 26. См. Palm [8].

§ 27. Публикуется впервые.

§ 28, 29, 30, 31, 32, 33. См. Palm [8].
§ 34, 35. См Колмогоров [2], Феллер [3], Фрай [4], Erlang [7].
§ 36, 37 См. Erlang [7].
§ 38, 39, 40. Вся глава II представляет собой существенную переработку статьи автора [6].

ЛИТЕРАТУРА

1. Гнеденко Б. В., Курс теории вероятностей, изд. 2-е, Гостехиздат, 1954.
2. Колмогоров А. Н., Sur le problème d'attente, Мат. сборн., 1931, 38, № 1—2, 101—106.
3. Феллер В., Введение в теорию вероятностей и ее приложения, Изд. иностр. лит., 1952.
4. Фрай Т., Теория вероятностей для инженеров, Гостехиздат, 1934.
5. Хинчин А. Я., Асимптотические законы теории вероятностей, ОНТИ, 1936
6. Хинчин А Я., Математическая теория стационарной очереди, Мат. сборн 1932, 39, № 4, 73—81.
7. Erlang A. K. The Life and Works of The Copenhagen Telephone Company, 1948.
8. Palm C., Intensitätsschwankungen im Fernsprechverkehr, Ericsson Technics, 1943, 44, 1—189
9. Redheffer R. M., A note on the Poisson law, Math. Magazine, 1953, 26, № 4, 185—188.

МАТЕМАТИЧЕСКАЯ ТЕОРИЯ СТАЦИОНАРНОЙ ОЧЕРЕДИ*)

Предлагаемое исследование было предпринято с целью ответить на некоторые вопросы, возникшие в связи со строительством и эксплуатацией телефонных станций; однако, результаты его имеют совершенно общий характер и могут найти применение к случаю любой очереди в периоде ее стационарности, т. е. при условии, что очередь не имеет тенденции к нарастанию или убыванию; следует отметить, что применяемые здесь методы, безусловно, могут служить орудием исследования и нестационарных очередей; однако, эта (впрочем, практически менее важная) часть работы в настоящее время еще не выполнена.

В частности, я полагаю, что результаты настоящего исследования могут служить основанием для расчетов основных показателей времени простоя станков и других установок в некоторых отраслях промышленности — всякий раз, когда большое число установок поручается наблюдению одного рабочего. В этом случае очередь образуется станками, ожидающими ремонтной операции. Однако в случае, когда число станков, поручаемых одному рабочему, не велико (лежит в пределах одного десятка), формулы настоящей статьи становятся неточными, и следует избегать пользоваться ими. Этот случай, более сложный, требует специального исследования, которое я надеюсь изложить в другой работе.

Опытная проверка результатов настоящего исследования ведется Московской городской телефонной станцией. Результаты ее, а также практические правила расчета времени

*) Матем. сб., т. 39 (1932), № 4, стр. 73—84.

ожидания, вытекающие из полученных формул, будут опубликованы в соответствующей научно-технической печати.

Некоторые из результатов настоящей работы содержатся в качестве частных случаев в работах Pollaczeck'a 1930 г.*) Однако я счел желательным все же их опубликовать, так как чрезвычайно громоздкий метод (позволяющий, правда, в то же время решать более общие задачи) делает чтение его работ крайне затруднительным, в особенности для инженера. Некоторые частные случаи формулы для среднего времени ожидания (§ 4) были найдены Thornton Fry'ем**), которому, по-видимому, принадлежит и идея применяемого мною рекуррентного метода, хотя Fry применяет его к другим объектам и в других предположениях.

В дальнейшем речь идет о вызовах, разговорах, абонентах и т. п. Эта терминология, взятая из телефонной практики, выбрана лишь для определенности и ни в какой мере не должна ограничивать собою области приложимости полученных результатов.

§ 1. Определения, обозначения и постановка задачи

Телефонистка, обслуживающая данную (многочисленную) группу абонентов, получает, в среднем, определенное число вызовов (требований) в час. Мы допускаем, что вероятность поступления данного числа k вызовов в данный промежуток времени t зависит только от длины этого промежутка, но не зависит ни от момента его начала, ни от того, сколько абонентов находилось в этот момент в состоянии ожидания разговора. Это условие и является предпосылкой стационарности очереди. Если вызов застает телефонистку занятой, абонент должен выждать, пока она переговорит со всеми абонентами, ранее него занявшими очередь. Длительность разговора между абонентом и телефонисткой предполагается случайной величиною; в дальнейшем через $f(t)\,dt$ обозначается вероятность того, что эта длительность заключена между t и $t+dt$ (таким образом, $f(t)\,dt$ есть относительное число разговоров, имеющих данную длительность). Средняя

*) Math. Zeitschr., т. 32, 1930, № 64, стр. 729.

**) См. его книгу «Теория вероятностей для инженеров», ГТТИ, 1934.

продолжительность разговора равна

$$\mu_1 = \int_0^\infty t f(t)\, dt.$$

Если в единицу времени (за которую в телефонной практике обычно принимается час) происходит n вызовов, то $n\mu_1 = \alpha$ есть среднее время занятости (нагрузка) телефонистки. Во всем дальнейшем предполагается $\alpha < 1$, так как в противном случае очередь, как легко видеть, безгранично растет и не может быть стационарной (такого рода установка была бы в качестве постоянной установки практически бессмысленной).

Таким образом данными нашей задачи служат: с одной стороны, функция $f(t)$ (закон распределения длительности разговоров), с другой,— число α (нагрузка телефонистки) или n (число вызовов в час). Ищется, в конечном счете, закон распределения времени ожидания, т. е. вероятность того, что абонент, производящий вызов в произвольно выбранный момент времени, должен будет прождать более чем t часов, прежде чем вступит в разговор с телефонисткой; наиболее важными показателями являются, однако, *среднее значение* этого времени ожидания и его *среднее квадратическое отклонение*.

В дальнейшем я обозначаю через π_k $(k = 1, 2, \ldots)$ вероятность того, что при начале наудачу выбранного разговора*) число ожидающих очереди абонентов упадет с k до $k-1$; иначе говоря, π_k есть относительное число разговоров, начинающихся в указанных условиях. Через π_0 я подобным же образом буду обозначать относительное число разговоров, начинающихся без предварительного ожидания.

Очевидно, мы можем также сказать, что π_k $(k = 0, 1, 2, \ldots)$ есть относительное число таких *окончаний* разговоров, в момент которых число занятых абонентов падает с $k+1$ до k; в случае $k > 0$ это непосредственно ясно, так как

---

*) Идея рассмотрения моментов начала разговоров оказалась весьма плодотворной. О возникшей в связи с этим теории вложенных цепей Маркова см. работу Д. Кендалла «Стохастические процессы, встречающиеся в теории очередей, и их анализ методом вложенных цепей Маркова», сб. переводов «Математика», т. 3, вып. 6, 1959, стр. 97—111.— *Б. Г.*

момент окончания одного разговора есть, вместе с тем, момент начала следующего разговора; в случае же $k=0$ это следует из того, что за каждым окончанием в отсутствии ожидающих следует в точности одно начало без ожидания.

Но непосредственно очевидно, что число моментов, когда очередь переходит от $k+1$ участников к k, равно числу моментов обратного перехода от k к $k+1$; так как момент первого рода наступает только при окончании разговора, а момент второго рода — только при появлении нового вызова, то π_k $(k=0, 1, 2, \ldots)$ есть относительное число вызовов, в момент которых число занятых абонентов повышается с k до $k+1$.

Эти числа π_k мы будем называть *начальными индексами* нашей задачи; их роль будет чисто вспомогательной; однако, мы должны начать с их определения.

§ 2. Рекуррентное определение начальных индексов

В тех условиях, которые нами приняты относитель о поступления вызовов, вероятность того, что в течение некоторого данного промежутка времени t произойдет k вызовов, выражается, как известно, формулой Пуассона*)

$$\nu_k(t) = e^{-\frac{\alpha}{\mu_1}t} \frac{\left(\frac{\alpha}{\mu_1}t\right)^k}{k!}.$$

Найдем прежде всего вероятность η_k того, что в течение данного, наудачу выбранного разговора произойдет k вызовов (иначе говоря, η_k есть относительное число тех разговоров, в течение которых происходит k вызовов); вероятность того, что выбранный разговор будет иметь длительность, заключенную между t и $t+dt$, равна по предположению $f(t)\,dt$; следовательно, по формуле полной вероятности:

$$\eta_k = \int_0^\infty \nu_k(t) f(t)\,dt.$$

---

*) Это есть общая теорема теории вероятностей, находящая себе применение не только в эксплуатации установок общего пользования, но, например, также в теории радиоактивного излучения. Доказательство можно найти, например, в книге Фрая «Теория вероятностей для инженеров».

Мы обозначим через π_0 вероятность того, что наудачу выбранный разговор начнется без предварительного ожидания; для того чтобы это случилось, необходимо и достаточно, чтобы предыдущий разговор начался в момент, когда никто не ждет (вероятность чего есть $\pi_0+\pi_1$), и чтобы, сверх того, в течение этого разговора не последовало ни одного вызова (вероятность чего есть η_0). Отсюда

$$\pi_0=(\pi_1+\pi_0)\,\eta_0. \tag{1}$$

В случае $k>0$ π_k есть вероятность того, что наудачу выбранный разговор начнется в момент, когда число ожидающих падает с k до $k-1$; для того чтобы это случилось, необходимо и достаточно осуществление какой-нибудь одной из следующего ряда предпосылок:

1. Предыдущий разговор начался в момент, когда ожидающих не было (вероятность $\pi_0+\pi_1$), и в течение этого разговора произошло k вызовов (вероятность η_k).

2. Предыдущий разговор начался в момент, когда число ожидающих уменьшилось с двух до одного (вероятность π_2), и в течение этого разговора произошло $k-1$ вызовов (вероятность η_{k-1}).

. . . .

$k+1$. Предыдущий разговор начался в момент, когда число ожидающих упало с $k+1$ до k (вероятность π_{k+1}), и в течение этого разговора не последовало ни одного вызова (вероятность η_0).

По формуле полной вероятности мы, таким образом, получаем:

$$\pi_k=\pi_{k+1}\eta_0+\pi_k\eta_1+\pi_{k-1}\eta_2+\ \dots\ +\pi_2\eta_{k-1}+(\pi_1+\pi_0)\,\eta_k \quad (k>0). \tag{2}$$

Формулы (1) и (2) позволяют последовательно выразить все π_k через π_0, а так как $\sum\limits_{k=0}^{\infty}\pi_k=1$, то и π_0 является определенным. Явные выражения этих начальных индексов через данные элементы нам в дальнейшем не понадобятся, но полученные рекуррентные формулы призваны играть основную роль.

§ 3. Основной закон стационарной очереди

Пусть $P(k)$ означает вероятность того, что в произвольно выбранный момент времени в точности k абонентов окажутся занятыми (т. е. что в случае $k>0$ один говорит и $k-1$ ожидают). Практически эта вероятность измеряется относительной длительностью того времени, в течение которого занято k абонентов.

Основным законом стационарной очереди назовем соотношение

$$P(k)=\pi_k \quad (k=0,\ 1,\ 2,\ \ldots)$$

справедливость которого и непосредственно довольно ясна; в самом деле, в § 2 мы видели, что π_k есть относительное число вызовов, встречающих при своем возникновении k занятых абонентов; очевидно, что это число при наших предпосылках совпадает с относительной длительностью $P(k)$ того времени, в течение которого выполняется это условие.

Однако этот непосредственный вывод основного закона все же не является вполне убедительным; поэтому мы дадим формальное доказательство этого закона, тем более, что нужный для этого математический аппарат нам все равно понадобится в дальнейшем.

Итак, мы постараемся выразить величину $P(k)$ через начальные индексы π_i. С этой целью заметим следующее: вероятность того, что в произвольно выбранный момент времени телефонистка занята, равна α; если это случилось, то вероятность того, что ведущийся разговор имеет длительность, заключенную между t и $t+dt$, равна относительной общей продолжительности таких разговоров, т. е.

$$\frac{t\,f(t)\,dt}{\mu_1};$$

если, наконец, разговор действительно имеет эту длительность, то вероятность $P_t(x<u<x+dx)$ того, что уже протекшая до нашего вызова часть u этого разговора заключена между x и $x+dx$, составляет $\frac{dx}{t}$ (если $x<t$) и 0

(если $x > t$). Отсюда:

$$P(x < u < x + dx) = \alpha \int_x^\infty \frac{t f(t)\, dt}{\mu_1} \frac{dx}{t} =$$

$$= \frac{\alpha\, dx}{\mu_1} \int_x^\infty f(t)\, dt = \frac{\alpha}{\mu_1} [1 - F(x)]\, dx.$$

Здесь левая часть есть вероятность того, что в произвольно выбранный момент времени мы застанем разговор, уже протекшая часть которого u заключена между x и $x + dx$, а в правой части через $F(x)$ обозначен интегральный закон распределения длительности разговоров. Вероятность же того, что в течение этой, уже протекшей, части разговора состоялось j вызовов, равна, очевидно

$$\omega_j = \int_0^\infty P(x < u < x + dx)\, v_i(x) = \frac{\alpha}{\mu_1} \int_0^\infty v_j(x) [1 - F(x)]\, dx.$$

Но, с другой стороны, очевидно, что

$$P(k) = \pi_k \omega_0 + \pi_{k-1} \omega_1 + \ldots + (\pi_1 + \pi_0) \omega_{k-1},$$

откуда

$$P(k) = \frac{\alpha}{\mu_1} \int_0^\infty [1 - F(x)]\, u_{k-1}(x)\, dx, \tag{3}$$

где положено:

$$u_{k-1}(x) = \pi_k v_0(x) + \pi_{k-1} v_1(x) + \ldots + (\pi_1 + \pi_0) v_{k-1}(x).$$

Функции $u_k(x)$, как легко непосредственно проверить, удовлетворяют разностно-дифференциальному уравнению

$$u_k'(x) = \frac{\alpha}{\mu_1} [u_{k-1}(x) - u_k(x)];$$

поэтому, написав соотношение (2) в виде:

$$\pi_k = \int_0^\infty u_k(x) f(x)\, dx,$$

и интегрируя по частям, мы находим:

$$\pi_k = u_k(0) + \int_0^\infty u'_k(x)[1-F(x)]\,dx =$$

$$= \pi_{k+1} + \frac{\alpha}{\mu_1}\int_0^\infty [u_{k+1}(x) - u_k(x)][1-F(x)]\,dx =$$

$$= \pi_{k+1} + P(k) - P(k+1),$$

откуда

$$\pi_k - P(k) = \pi_{k+1} - P(k+1) = \text{const}.$$

Но π_k и $P(k)$ — члены сходящихся рядов, в силу чего эта постоянная должна обращаться в нуль, и таким образом, основной закон доказан.

В силу этого закона рекуррентные формулы (1) и (2) имеют место для величин $P(k)$ и позволяют в случае надобности их последовательно определять; однако, получающееся отсюда общее выражение для $P(k)$ неудобно для расчетов и для теоретического употребления, вследствие чего мы предпочитаем в дальнейшем пользоваться соотношениями (1) и (2) в их первоначальном виде:

$$P(k) = \int_0^\infty u_k(x) f(x)\,dx. \tag{4}$$

§ 4. Определение среднего времени ожидания

Время ожидания γ (случайная величина), т. е. время, протекающее от момента вызова до начала разговора, слагается из двух составляющих: 1) время τ от момента вызова до окончания того разговора, который ведется в момент вызова, и 2) суммарная длительность T разговоров всех тех абонентов, которые в момент произведенного вызова находились в состоянии ожидания (разумеется, $T=0$, если вызов произошел в момент, когда нет ожидающих, и $T=\tau=\gamma=0$, если вызов застал телефонистку свободной).

Таким образом:

$$\gamma = \tau + T,$$

где τ и T — также случайные величины. Для метода последующих расчетов важно иметь в виду следующее замечание.

Очень легко найти закон распределения для величины τ; зная величины $P(k)$, найденные нами в § 3, столь же нетрудно найти закон распределения величины T; однако, это не дало бы нам возможности написать закон распределения для величины γ, потому что величины τ и T *взаимно зависимы*, и в этом — главная трудность проблемы. В самом деле, если, например, τ мало, то это есть при прочих равных условиях указание на то, что разговор длится уже относительно долго; это же, в свою очередь, указывает на то, что, вероятно, скопилось много ожидающих, так что T получает шансы стать большим. Каким образом может быть преодолено это затруднение, мы увидим в дальнейшем; здесь же, где речь идет только о математическом ожидании величины γ, это обстоятельство не играет, конечно, никакой роли, так как (обозначая математическое ожидание случайной величины тою же буквой, но с чертой сверху):

$$\bar{\gamma}=\bar{\tau}+\bar{T}.$$

Чтобы найти $\bar{\tau}$, заметим, что вероятность застать разговор длительности, заключенной между t и $t+dt$, равна, как мы уже видели,

$$\frac{\alpha}{\mu_1} tf(t)\,dt;$$

при этом же условии математическое ожидание величины τ, очевидно, равно $\frac{1}{2}t$; отсюда

$$\bar{\tau}=\frac{\alpha}{2\mu_1}\int_0^\infty t^2 f(t)\,dt=\frac{\alpha\mu_2}{2\mu_1},$$

где μ_2 означает математическое ожидание квадрата длительности разговора.

Что касается величины T, то при условии, что в момент вызова занято $j \geqslant 1$ абонентов, ее математическое ожидание, очевидно, становится равным $(j-1)\mu_1$; отсюда:

$$\bar{T}=\mu_1\sum_{j=1}^{\infty}(j-1)P(j);$$

а так как $\sum\limits_{j=1}^{\infty} jP(j)$ есть математическое ожидание числа занятых абонентов в произвольно выбранный момент времени, очевидно, совпадающее с суммарным временем занятости всех абонентов в течение часа, то мы имеем:

$$\sum_{j=1}^{\infty} jP(j) = n(\mu_1 + \bar{\gamma}) = \alpha + \frac{\alpha}{\mu_1}\bar{\gamma};$$

с другой стороны, очевидно

$$\sum_{j=1}^{\infty} P(j) = 1 - P(0) = \alpha,$$

так что

$$\bar{T} = \mu_1\left(\alpha + \frac{\alpha}{\mu_1}\bar{\gamma} - \alpha\right) = \alpha\bar{\gamma};$$

таким образом

$$\bar{\gamma} = \frac{\alpha}{2}\frac{\mu_2}{\mu_1} + \alpha\bar{\gamma},$$

откуда

$$\bar{\gamma} = \frac{\alpha}{2(1-\alpha)}\frac{\mu_2}{\mu_1}. \tag{5}$$

Эта простая формула полностью решает вопрос о среднем времени ожидания *); мы видим, что $\bar{\gamma}$ зависит не только от средней длительности разговора, но и от дисперсии. Практически это лишний раз подчеркивает важность стандартизации разговора между абонентом и телефонисткой.

В качестве числового примера рассмотрим случай нулевой дисперсии ($\mu_2 = \mu_1^2$), в налаженной телефонной практике близкий к действительности; положим $\alpha = 0{,}6$, $\mu_1 = 8$ сек.; это дает $\bar{\gamma} = 6$ сек.; если при тех же α и μ_1 предположить распределение длительности разговоров по показательному закону (ввиду его простоты часто, хотя и без достаточных оснований принимаемому за базу теоретических изысканий), то $\mu_2 = 2\mu_1^2$, и среднее время ожидания увеличивается вдвое: $\bar{\gamma} = 12$ сек.

---

*) Мы считаем нужным указать, что имеющая некоторое распространение среди практиков формула д-ра Маттиаса неверна, и вывод ее основан на ошибочных предпосылках.

§ 5. Формулы для основных вероятностей

Мы видели, что среднее время ожидания $\bar{\gamma}$ находится вполне элементарно; однако, уже математическое ожидание величины γ^2 этим методом не удалось бы установить ввиду отмеченной зависимости между величинами τ и T. Здесь необходим более глубокий анализ.

Прежде всего, поставим следующую задачу: если известно, что в некоторый момент времени происходит разговор, длительность которого равна t, то какова вероятность того, что в этот момент занято j абонентов, и что в то же время уже протекшая часть разговора u заключена между x и $x+dx$? Обозначая эту вероятность через $P_t(j,\ x<u<x+dx)$, мы, очевидно, будем иметь в легко понятных обозначениях:

$$P_t(j,\ x<u<x+dx)=P_t(j)\,P_{t,\,j}(x<u<x+dx)=$$
$$=P_t(x<u<x+dx)=P_{t,\ x<u<x+dx}(j).$$

Правая часть этих равенств легко может быть вычислена; в самом деле, во-первых (ср. § 3):

$$P_t(x<u<x+dx)=\begin{cases}\dfrac{dx}{t}, & \text{если } x<t,\\ 0, & \text{если } x>t;\end{cases}$$

во-вторых, последний множитель совпадает с $P_{x<u<x+dx}(j)$ (так как при фиксированной протекшей части разговора его общая длительность, очевидно, не может влиять на степень скученности в данный момент), и следовательно, равен:

$$\pi_j v_0(x)+\pi_{j-1}v_1(x)+\ldots+(\pi_1+\pi_0)\,v_{j-1}(x)=u_{j-1}(x).$$

Таким образом:

$$P_t(j)\,P_{t,\,j}(x<u<x+dx)=u_{j-1}(x)\frac{dx}{t}$$

в случае $x<t$ и 0 в случае $x>t$.

Но при данном t неравенства $x<u<x+dx$ эквивалентны неравенствам $t-x-dx<\tau<t-x$, а потому:

$$P_t(j)\,P_{t,\,j}(x<\tau<x+dx)=u_{j-1}(t-x)\frac{dx}{t}$$

при $x<t$ и 0 при $x>t$. Чтобы получить отсюда величину $P(j)\,P_j(x<\tau<x+dx)$, т. е. вероятность того, что

в произвольно выбранный момент времени мы найдем j занятых абонентов и что застигнутый нами при этом разговор продлится еще в течение времени, заключенного между x и $x+dx$, надо, очевидно, помножить найденную вероятность на $\frac{\alpha}{\mu_1} tf(t)\,dt$ и проинтегрировать по t от x до ∞ (ибо при $t<x$ подынтегральная функция равна нулю). Таким образом:

$$P(j)\,P_j(x<\tau<x+dx)=\frac{\alpha}{\mu_1}\,dx\int_x^\infty u_{j-1}(t-x)\,f(t)\,dt, \tag{6}$$

где $j=1,\ 2,\ 3,\ \ldots$

§ 6. Характеристическая функция времени ожидания

Как известно, для изучения всех вопросов, связанных с законом распределения случайной величины γ, в частности, для вычисления ее моментов любого порядка, удобнее всего найти ее характеристическую функцию, т. е. математическое ожидание выражения $e^{i\gamma\xi}$, рассматриваемое как функция параметра ξ. Так мы и поступим. Пользуясь для обозначения математического ожидания символом E, мы будем иметь:

$$\varphi(\xi)=E(e^{i\gamma\xi})=E(e^{i\tau\xi}\,e^{iT\xi}).$$

Однако математическое ожидание произведения в данном случае не равно произведению математических ожиданий, ввиду отмеченной зависимости величин τ и T. Поэтому для отыскания характеристической функции величины γ мы должны прибегнуть к более сложному приему.

Если мы условимся обозначать через E_j математические ожидания, вычисленные в предположении, что вызов застал j занятых абонентов, то в силу известных правил теории вероятностей:

$$\varphi(\xi)=E(e^{i\gamma\xi})=\sum_{j=0}^{\infty}P(j)\,E_j(e^{i\gamma\xi})=\sum_{j=0}^{\infty}P(j)\,E_j(e^{i\tau\xi}\,e^{iT\xi}).$$

Но при данном значении j величины τ и T взаимно независимы, так как, очевидно, длительность разговоров $j-1$ ожидающих абонентов не зависит от того, сколько еще времени будет длиться застигнутый разговор (величина T

только потому зависела от τ, что j зависело от τ). Поэтому:

$$\varphi(\xi)=\sum_{j=0}^{\infty} P(j)\, E_j(e^{i\tau\xi})\, E_j(e^{iT\xi})=$$
$$=P(0)+\sum_{j=1}^{\infty} P(j)\, E_j(e^{i\tau\xi})\, E_j(e^{iT\xi}).$$

После достигнутого таким образом разделения величин τ и T дальнейший метод вычисления уже ясен. Прежде всего, легко вычисляется $E_j(e^{iT\xi})$. В самом деле, если мы обозначим через $\psi(\xi)$ характеристическую функцию длительности разговора, т. е. положим

$$\psi(\xi)=\int_0^{\infty} e^{iT\xi} f(t)\, dt,$$

и заметим, что T есть суммарная длительность $j-1$ взаимно независимых разговоров, то мы без всяких вычислений можем написать:

$$E_j(e^{iT\xi})=[\psi(\xi)]^{j-1} \quad (j=1,\ 2,\ \ldots).$$

Далее, в силу формулы (6):

$$P(j)\, E_j(e^{i\tau\xi})=\int_0^{\infty} P(j)\, e^{ix\xi} P_j(x<\tau<x+dx)=$$
$$=\frac{\alpha}{\mu_1}\int_0^{\infty} e^{ix\xi}\, dx \int_x^{\infty} u_{j-1}(t-x) f(t)\, dt$$
$$(j=1,\ 2,\ \ldots),$$

откуда

$$\varphi(\xi)=P(0)+\frac{\alpha}{\mu_1}\int_0^{\infty} e^{ix\xi}\, dx \int_x^{\infty} f(t) \left\{\sum_{j=0}^{\infty} [\psi(\xi)]^j u_j(t-x)\right\} dt. \quad (7)$$

Пользуясь выражением функции $u_j(x)$, очень легко показать, что

$$\sum_{j=0}^{\infty} a^j u_j(z)=e^{\frac{z}{\mu_1}(a-1)z}\left\{P(0)+\sum_{k=1}^{\infty} a^{k-1} P(k)\right\}, \quad (8)$$

откуда в силу формулы (7)

$$\varphi(\xi) = P(0) + \frac{\alpha}{\mu_1}\chi(\xi)\int_x^\infty e^{ix\xi}\,dx\int_x^\infty e^{\frac{\alpha}{\mu_1}[\psi(\xi)-1](t-x)} f(t)\,dt,$$

где положено:

$$\chi(\xi) = P(0) + \sum_{k=1}^{\infty}[\psi(\xi)]^{k-1}P(k).$$

Полагая

$$\frac{\alpha}{i\mu_1}[\psi(\xi)-1] = \eta.$$

находим:

$$\varphi(\xi) = P(0) + \frac{\alpha}{\mu_1}\chi(\xi)\int_0^\infty e^{ix(\xi-\eta)}\,dx\int_x^\infty e^{it\eta} f(t)\,dt =$$

$$= P(0) + \frac{\alpha}{\mu_1}\chi(\xi)\int_0^\infty e^{it\eta} f(t)\,dt\int_0^t e^{ix(\xi-\eta)}\,dx =$$

$$= P(0) + \frac{\alpha}{\mu_1}\frac{\chi(\xi)}{(\eta-\xi)i}\int_0^\infty e^{it\eta} f(t)\left[1 - e^{it(\xi-\eta)}\right]dt =$$

$$= P(0) + \frac{\alpha}{i\mu_1}\chi(\xi)\frac{\psi(\eta)-\psi(\xi)}{\eta-\xi}$$

Но в силу формул (4) и (8):

$$\chi(\xi) = P(0) + \frac{1}{\psi(\xi)}\int_0^\infty f(x)\left\{\sum_{k=0}^{\infty}[\psi(\xi)]^k u_k(x) - P(0)\right\}dx =$$

$$= P(0)\,\frac{\psi(\xi)-1}{\psi(\xi)} + \frac{\chi(\xi)}{\psi(\xi)}\int_0^\infty e^{\frac{\alpha}{\mu_1}[\psi(\xi)-1]x} f(x)\,dx =$$

$$= P(0)\,\frac{\psi(\xi)-1}{\psi(\xi)} + \chi(\xi)\,\frac{\psi(\eta)}{\psi(\xi)},$$

откуда

$$\chi(\xi) = P(0)\,\frac{\psi(\xi)-1}{\psi(\xi)-\psi(\eta)},$$

и следовательно:

$$\varphi(\xi) = P(0)\left\{1 - \frac{\alpha}{i\mu_1}\frac{\psi(\xi)-1}{\eta-\xi}\right\} =$$
$$= (1-\alpha)\left\{1 - \frac{\alpha[\psi(\xi)-1]}{\alpha[\psi(\xi)-1-i\mu_1\xi]}\right\},$$

или

$$\varphi(\xi) = \frac{1-\alpha}{1-\alpha\left\{\dfrac{\psi(\xi)-1}{i\mu_1\xi}\right\}}.$$

Эта замечательная по своей простоте формула позволяет с большою легкостью находить выражения моментов величины γ через моменты длительности разговора; для этого надо только вычислять последовательные производные $\varphi(\xi)$ и, деля их на соответствующие степени i, полагать в них $\xi = 0$. Так, $\frac{1}{i}\varphi'(0) = \bar{\gamma}$ дает нам прежнюю формулу (5). Далее,

$$-\varphi''(0) = E(\gamma^2) = 2\bar{\gamma}^2 + \frac{\alpha\mu_3}{3(1-\alpha)\mu_1},$$

где

$$\mu_3 = \int_0^\infty t^3 f(t)\,dt.$$

Отсюда дисперсия:

$$\sigma^2(\gamma) = E(\gamma^2) - \bar{\gamma}^2 = \bar{\gamma}^2 + \frac{\alpha\mu_3}{3(1-\alpha)\mu_1}. \tag{9}$$

Эта формула показывает, что среднее квадратическое отклонение величины γ всегда превышает ее среднее значение; этим хорошо объясняется то явление, что даже при небольшом среднем времени ожидания мы часто имеем дело со сравнительно большими задержками. Так, например, при условии стандартной длительности разговора 8 сек. и при $\alpha = 0{,}6$ мы имели $\bar{\gamma} = 6$ сек. При тех же условиях формула (9) дает $\sigma(\gamma) > 8$ сек., так что при среднем времени ожидания 6 сек. не редкостью будет задержка в 15 сек. и более, что является уже ощутительным.

О СРЕДНЕМ ВРЕМЕНИ ПРОСТОЯ СТАНКОВ*)

1. Решаемая в настоящей статье общая задача имеет значение для всех тех отраслей промышленности, где наблюдению одного рабочего одновременно поручается несколько станков (или других установок) одинакового типа, причем функция рабочего состоит в выполнении операций, надобность в которых для каждого отдельного станка возникает в случайном порядке (ликвидация неполадок, мелкий ремонт и т. п.).

Момент, в который станок начинает требовать той или иной операции, мы будем называть моментом его остановки. Время, требующееся для совершения той или другой операции, может быть различным для различных операций и допускать какие угодно вариации для одной и той же операции.

В дальнейшем мы принимаем следующие обозначения:

λ — число станков, поручаемых одному рабочему;

$\varkappa\, dt$ — вероятность того, что станок, работающий в момент t, остановится ранее момента $t+dt$ (dt предполагается весьма малым);

$f(t)\, dt$ — относительное число операций, для которых потребное время заключено между t и $t+dt$.

Эти три элемента являются д а н н ы м и. И щ е т с я среднее время ожидания γ, т. е. среднее значение промежутка времени, протекающего от момента остановки до начала соответствующей операции; при этом предполагается, что рабочий обслуживает станки в порядке очереди, т. е. в порядке их остановок.

В качестве вспомогательных величин нам понадобятся следующие:

---

*) Матем. сб., т. 40 (1933), № 2, стр. 119—123.

μ — средняя длительность одной операции, n — число остановок одного станка в единицу времени;

$\varphi(t)$ — вероятность того, что станок, работающий в момент t_0, не остановится до момента t_0+t;

π_k — относительное число остановок, происходящих в такие моменты, когда число уже стоящих станков равно k.

Все эти величины могут быть определены из данных нашей задачи.

Следует заметить, что формулами, которые даются в настоящей статье, целесообразно пользоваться лишь для небольших значений λ (в пределах 1—2 десятков). При большем числе станков можно с успехом применять, в порядке приближения, значительно более простые методы расчета, которые нами установлены в другой статье*).

2. Единицу времени, которая может быть выбрана произвольно, целесообразно принять весьма большою; для краткости условимся называть ее часом.

Для средней длительности операции мы, очевидно, имеем формулу:

$$\mu = \int_0^\infty t f(t)\,dt.$$

Определим, далее, функцию $\varphi(t)$; для этого заметим, что, очевидно,

$$\varphi(t+dt) = \varphi(t)(1-\varkappa\,dt),$$

откуда

$$\varphi'(t) = -\varkappa\varphi(t),$$

или

$$\varphi(t) = Ce^{-\varkappa t},$$

причем очевидное условие $\varphi(0)=1$ дает $C=1$, так что

$$\varphi(t) = e^{-\varkappa t}.$$

Вероятность же того, что станок, работающий в начале некоторого промежутка длины t, остановится в течение этого промежутка, равна

$$1-\varphi(t) = 1-e^{-\varkappa t},$$

*) «Математическая теория стационарной очереди», Матем. сб., т. 39 (1932), № 4, стр. 73.

и далее, вероятность того, что для станка, работающего в момент t_0, первая остановка последует в течение малого промежутка времени $(t_0+t,\ t_0+t+dt)$, равна

$$\varphi(t)\varkappa\,dt=\varkappa e^{-\varkappa t}dt;$$

отсюда мы можем определить среднее значение рабочего периода станка, т. е. времени, протекающего от момента пуска (конца операции) до ближайшей остановки:

$$\int_0^\infty t\varkappa e^{-\varkappa t}dt=\frac{1}{\varkappa}.$$

Заметим, что практически величину $\varkappa$ лучше всего находить именно из этой формулы, так как средний рабочий период станка легко поддается непосредственному измерению.

3. Теперь мы можем перейти к установлению формулы, определяющей среднее время ожидания γ. С этой целью заметим, что рабочий час для каждого станка слагается из чередующихся периодов трех различных категорий: рабочий период $\left(\text{средней длины } \frac{1}{\varkappa}\right)$, период ожидания (средней длины γ) и период операции (средней длины μ). Так как мы обозначили через n число остановок станка в течение часа, то

$$n\left(\frac{1}{\varkappa}+\gamma+\mu\right)=1. \tag{1}$$

В этом уравнении, кроме γ, неизвестно еще и n. Поэтому мы должны искать второе соотношение, связывающее те же неизвестные. С этой целью заметим, что для рабочего его рабочий час слагается из чередующихся периодов двух категорий — периодов занятости и периодов свободы. Сумму длин всех периодов занятости легко найти. Число станков равно λ, каждый станок требует в час n операций, средняя длительность операции равна μ. Отсюда время загрузки рабочего равно $n\lambda\mu$. Сложнее найти время, в течение которого рабочий свободен. Заметим прежде всего, что число моментов освобождения рабочего, очевидно, равно числу моментов, когда он от свободного состояния переходит к работе (так как моменты этих двух категорий строго чередуются). Но второе число есть, вместе с тем, число остановок, происходящих в периоды, когда все остальные станки работают (и потому рабочий свободен). Согласно нашим обозначениям это число есть $\lambda n\pi_0$; таково, следовательно, число периодов свободы для рабочего. Что касается средней длительности такого периода, то она может быть найдена из следующих соображений. Если в данный момент все станки работают, то вероятность того, что и по истечении времени t все они еще будут работать, равна $e^{-\lambda\varkappa t}$; вероятность же того, что какой-нибудь из них остановится в промежуток времени $(t, t+dt)$, равна (пренебрегая малыми высших порядков) $\lambda\varkappa\, dt$. Следовательно, вероятность того, что длительность данного периода свободы для рабочего будет заключена между t и $t+dt$, равна $e^{-\lambda\varkappa t}\lambda\varkappa\, dt$, а значит, средняя длительность периода свободы рабочего есть

$$\int_0^\infty te^{-\lambda\varkappa t}\,\lambda\varkappa dt=\frac{1}{\lambda\varkappa}.$$

Отсюда сумма длин всех периодов свободы равна

$$\lambda n\pi_0\cdot\frac{1}{\lambda\varkappa}=n\frac{\pi_0}{\varkappa},$$

и мы получаем соотношение

$$n\left(\lambda\mu+\frac{\pi_0}{\varkappa}\right)=1. \qquad (2)$$

Исключая n из уравнений (1) и (2), мы находим:

$$\gamma=(\lambda-1)\,\mu-\frac{1-\pi_0}{\varkappa}.$$

Этот результат полностью решал бы поставленную задачу, если бы π_0 было нам известно. Определение величины π_0, которому будет посвящено все дальнейшее, и составляет в сущности, математическое содержание предлагаемого решения, так как все предшествующее было элементарным расчетом.

4. Число π_k, представляющее собою, по определению, относительное число остановок, в момент которых число стоящих станков увеличивается с k до $k+1$, очевидно, есть, вместе с тем, относительное число таких операций, в конце которых число стоящих станков уменьшается

с $k+1$ до k. Обозначим через η_{kj} вероятность того, что в течение операции, в начале которой число стоящих станков равно k, остановится еще j станков. Для того чтобы некоторая операция началась без предварительного ожидания (вероятность π_0), необходимо и достаточно, чтобы: 1) предшествующая операция началась в отсутствии ожидающих (вероятность $\pi_1+\pi_0$) и 2) в течение этой операции ни один станок не остановился (вероятность $\eta_{1,0}$). Отсюда:

$$\pi_0=(\pi_0+\pi_1)\,\eta_{1,0}. \tag{3}$$

Подобным же образом, если $0<k<\lambda-1$, то для того, чтобы какая-либо операция началась в момент, когда число стоящих станков уменьшается с $k+1$ до k *) (вероятность π_k), необходимо и достаточно, чтобы: либо предшествующая операция началась в момент, когда число стоящих станков уменьшается с $k+2$ до $k+1$ (вероятность π_{k+1}), и в течение этой операции не произошло ни одной остановки (вероятность $\eta_{k+1,0}$); либо предшествующая операция началась в момент, когда число стоящих станков уменьшается с $k+1$ до k (вероятность π_k), и в течение этой операции остановился один станок (вероятность $\eta_{k,1}$), и т. д.; либо, наконец, чтобы предшествующая операция началась в отсутствии ожидающих (вероятность $\pi_1+\pi_0$) и в течение этой операции произошло k остановок (вероятность $\eta_{1,k}$). Таким образом мы находим:

$$\pi_k=\pi_{k+1}\eta_{k+1,0}+\pi_k\eta_{k,1}+\ldots+(\pi_1+\pi_0)\,\eta_{1,k} \qquad (0<k<\lambda-1). \tag{4}$$

Наконец, в случае $k=\lambda-1$ мы аналогичным рассуждением находим:

$$\pi_{\lambda-1}=\pi_{\lambda-1}\eta_{\lambda-1,1}+\ldots+(\pi_1+\pi_0)\,\eta_{1,\lambda-1}. \tag{5}$$

Если предполагать все величины $\eta_{k,j}$ известными, то, очевидно, из уравнений (3) и (4) можно последовательно выразить все π_k линейно через π_0; уравнение (5) представляет собою следствие соотношений (3) и (4) и новых результатов

*) Заметим, что при $k>0$ момент начала данной операции есть, вместе с тем, момент окончания предшествующей операции.

не дает, но очевидное соотношение

$$\sum_{k=0}^{\lambda-1} \pi_k = 1$$

позволяет определить и π_0.

Таким образом, задача сводится к определению коэффициентов $\eta_{k,j}$, что может быть сделано без затруднений. В самом деле, если данная операция длится t часов и если в начале этой операции стоят k станков, то вероятность того, что в течение этой операции из $\lambda - k$ работающих станков остановятся ровно j, очевидно, есть

$$C_{\lambda-k}^{j}(1-e^{-xt})^{j}(e^{-xt})^{\lambda-k-j}.$$

А так как вероятность того, что данная операция будет иметь длительность, заключенную между t и $t+dt$, есть $f(t)\,dt$, то

$$\eta_{k,j} = C_{\lambda-k}^{j} \int_{0}^{\infty} (1-e^{-xt})^{j}(e^{-xt})^{\lambda-k-j} f(t)\,dt,$$

что и решает задачу.

5. Зная среднее время простоя, мы без труда можем решать все вопросы, связанные с наиболее выгодным распределением станков между рабочими. Для практических расчетов, основанных на изложенной теории, необходимо, разумеется, создать специальный аппарат в виде номограммы или таблицы, что, вероятно, не представит затруднений. Теоретически же задача приведенными нами соображениями решается до конца.

ПОТОКИ СЛУЧАЙНЫХ СОБЫТИЙ БЕЗ ПОСЛЕДЕЙСТВИЯ*)

§ 1. Постановка задачи

Поток однородных случайных событий представляет собой случайный процесс $x(t)$ $(t \geqslant 0)$, в котором функция $x(t)$ (число событий, наступающих в промежутке времени**) $(0, t)$) — неубывающая и способна принимать лишь целые неотрицательные значения. Полное стохастическое описание такого потока мы получаем, задавая для любого натурального числа n и для любой группы из n моментов времени $0 < t_1 < t_2 < \ldots < t_n$ n — мерный закон распределения вектора $[x(t_1), x(t_2), \ldots, x(t_n)]$. Вместо этого можно, очевидно, задать и закон распределения вектора $[x(t_1), x(t_2) - x(t_1), \ldots, x(t_n) - x(t_{n-1})]$.

Поток событий называется *стационарным,* если закон распределения вектора $[x(a+t_i) - x(a), 1 \leqslant i \leqslant n]$ зависит только от чисел $t_1, t_2, \ldots, t_n$, но не зависит от a (и, следовательно, при любом $a \geqslant 0$ совпадает с законом распределения вышеприведенного вектора $x(t_i)$, $1 \leqslant i \leqslant n$). Поток называется *потоком без последействия,* если для любой системы (u_i, v_i) $(1 \leqslant i \leqslant n)$ попарно не пересекающихся промежутков времени разности $x(v_i) - x(u_i)$ $(1 \leqslant i \leqslant n)$ представляют собой взаимно независимые случайные величины. Если $v_k(\alpha, \beta)$ $(0 \leqslant \alpha < \beta$, $k \geqslant 0$ целое) означает вероятность равенства $x(\beta) - x(\alpha) = k$ (т. е. вероятность того, что в течение промежутка (α, β) наступит k событий данного потока), то поток без последействия, очевидно, полностью описывается заданием системы функций $v_k(\alpha, \beta)$ $(0 \leqslant \alpha < \beta,$

---

*) Теория вероятностей и ее применения, т. 1, вып. 1, 1956, стр. 3—17.

**) Промежутком времени (a, b) мы во всем дальнейшем называем совокупность всех t, для которых $a \leqslant t < b$.

$k \geqslant 0$), так как тем самым задается распределение разности $x(\beta) - x(\alpha)$ для любого промежутка (α, β). Если данный поток к тому же стационарный, то величина $v_k(\alpha, \alpha + t)$ зависит только от k и t и может быть обозначена через $v_k(t)$ (вероятность наступления k событий в произвольном промежутке длины t).

Если к требованиям стационарности и отсутствия последействия добавить еще условие

$$\lim_{t \to 0} \frac{1 - v_0(t) - v_1(t)}{t} = 0, \qquad (1)$$

то, как известно ([1], глава 1),

$$v_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad (k = 0, 1, \ldots);$$

распределение величины $x(t)$ при сделанных трех допущениях определяется, таким образом, однозначно с точностью до произвольного положительного числа λ (это число называют *интенсивностью* данного потока).

Недавно Редхеффером [2] и автором ([1], § 8) найдена общая форма стационарных потоков без последействия без дополнительного требования (1). С другой стороны, уже давно известна форма потока без последействия, удовлетворяющего требованию (1), но, вообще говоря, нестационарного *).

За последнее время, с одной стороны, обнаружилось, что для приложений имеет значение исследование потоков без последействия весьма общего вида. С другой стороны, удалось найти методы, сделавшие возможным исследования такого рода для некоторых специальных задач (см., например, рассмотрение известной задачи Эрланга у Пальма [3] или у автора [1], § 24; см. также работу Такача [4] о дробовом эффекте). Поэтому естественно встает вопрос об общей форме потоков без последействия. Настоящая работа имеет целью дать ответ на этот вопрос; полученное решение задачи можно признать удовлетворительным, так как все элементы произвола искомой структуры выявлены в нем с исчерпывающей ясностью.

---

*) Правда, при этом обычно предполагается еще существование некоторой «мгновенной интенсивности» $\lambda(t)$ потока любого момента времени t.

Пусть $\Lambda(t)$ означает (конечное или бесконечное) математическое ожидание числа событий данного потока, наступающих в промежутке $(0, t)$. Очевидно, $\Lambda(t)$ есть неотрицательная и неубывающая функция от t, которую мы будем называть *ведущей функцией* данного потока. Будем называть этот поток *финитным*, если $\Lambda(t) < +\infty$ $(t \geqslant 0)$. Во всем дальнейшем мы будем рассматривать только финитные потоки. Случай $\Lambda(t) = +\infty$ представляет собой задачу совсем особого рода, вряд ли способную отвечать запросам большинства приложений.

Если $\Lambda(t)$ — любая неотрицательная, конечная, неубывающая функция от t, то формулы

$$v_k(\alpha, \beta) = \exp\{-[\Lambda(\beta) - \Lambda(\alpha)]\} \frac{[\Lambda(\beta) - \Lambda(\alpha)]^k}{k!} \qquad (2)$$

$$(0 \leqslant \alpha < \beta, \quad k \geqslant 0)$$

всегда определяют некоторый поток без последействия, ведущая функция которого совпадает с $\Lambda(t)$. Именно этот поток обычно и рассматривают, когда исходят от ведущей функции $\Lambda(t)$. Однако такое ограничение, вообще говоря, представляется необоснованным. Как мы увидим, потоки типа (2) образуют лишь частный случай финитных потоков без последействия, так что, ограничиваясь потоками типа (2), мы подменили бы общее исследование проблемы рассмотрением частных случаев. Правда, потоки типа (2) образуют собою, как мы увидим, в известном смысле фундамент всей совокупности финитных потоков без последействия; но это как раз и надо показать с полной ясностью для того, чтобы специальное исследование потоков типа (2) получило хотя бы временное обоснование.

§ 2. Леммы о финитных потоках

Для решения поставленной задачи нам понадобится ряд элементарных и по большей части легко доказуемых вспомогательных предложений. Многие из этих лемм не связаны с отсутствием последействия и имеют силу для всех финитных потоков; такими леммами мы и займемся в настоящем параграфе; во всех последующих предложениях этого параграфа данный поток предполагается финитным, а во всем остальном — произвольным.

Лемма 1. *При* $0 \leqslant \alpha < \beta$ *всегда* $\sum\limits_{k=0}^{\infty} v_k(\alpha, \beta) = 1.$

Доказательство. Очевидно, разность $1 - \sum\limits_{i=0}^{k} v_i(\alpha, \beta)$ есть вероятность того, что в промежутке (α, β) наступит более k событий данного потока. Поэтому на основании неравенства Чебышева

$$1 - \sum_{i=0}^{k} v_i(\alpha, \beta) < \frac{\Lambda(\beta) - \Lambda(\alpha)}{k},$$

и, следовательно,

$$1 - \sum_{i=0}^{k} v_i(\alpha, \beta) \to 0 \quad (k \to \infty),$$

ч. т. д.

Лемма 2. *При* $k \geqslant 0$ *и* $t > 0$ *существует предел*

$$\lim_{h \to +0} v_k(t-h, t+h) = p_k(t). \tag{3}$$

Доказательство. Положим для $k \geqslant 0$, $0 \leqslant \alpha < \beta$

$$\sum_{r=k}^{\infty} v_r(\alpha, \beta) = \psi_k(\alpha, \beta).$$

Так как $\psi_k(\alpha, \beta)$ есть вероятность того, что в промежутке (α, β) наступит по меньшей мере k событий, то величина $\psi(t-h, t+h)$ при положительном и убывающем h не может возрастать. Поэтому всегда существует

$$\lim_{h \to +0} \psi_k(t-h, t+h).$$

В силу соотношения $v_k(\alpha, \beta) = \psi_k(\alpha, \beta) - \psi_{k+1}(\alpha, \beta)$ существует и предел (3), ч. т. д.

Условимся называть точку (момент времени) t *регулярной* или *сингулярной* точкой данного финитного потока, смотря по тому, будет ли $p_0(t) = 1$ или $p_0(t) < 1$.

Лемма 3. *Все точки непрерывности функции* $\Lambda(t)$ — *регулярные, а все ее точки разрыва — сингулярные точки данного потока.*

Доказательство. Так как

$$1-v_0(t-h,\ t+h)=\sum_{k=1}^{\infty} v_k(t-h,\ t+h)\leqslant$$

$$\leqslant\sum_{k=1}^{\infty} kv_k(t-h,\ t+h)=\Lambda(t+h)-\Lambda(t-h),$$

то в случае $\Lambda(t+0)=\Lambda(t-0)$ мы имеем

$$1-p_0(t)=\lim_{h\to+0}[1-v_0(t-h,\ t+h)]=0,$$

что доказывает первое утверждение леммы 3.

Далее, при $t>0$, $0<h<t$ и любом целом $m>0$

$$\Lambda(t+h)-\Lambda(t-h)=\sum_{k=1}^{\infty} kv_k(t-h,\ t+h)=$$

$$=\sum_{k=1}^{\infty}\psi_k(t-h,\ t+h)=\sum_{k=1}^{m}\psi_k(t-h,\ t+h)+$$

$$+\sum_{k>m}\psi_k(t-h,\ t+h).$$

Пусть $\varepsilon>0$ задано произвольно. Закрепим сначала некоторое h_0, $0<h_0<t$, и выберем число m столь большим, чтобы было

$$R_m(h_0)=\sum_{k>m}\psi_k(t-h_0,\ t+h_0)<\varepsilon;$$

тогда и подавно при $0<h\leqslant h_0$

$$R_m(h)<\varepsilon,$$

и, следовательно,

$$\Lambda(t+h)-\Lambda(t-h)<\sum_{k=1}^{m}\psi_k(t-h,\ t+h)+\varepsilon$$

$$(0<h\leqslant h_0). \qquad (4)$$

Но здесь

$$\sum_{k=1}^{m}\psi_k(t-h,\ t+h)\leqslant m\psi_1(t-h,\ t+h)\leqslant$$

$$\leqslant m[1-v_0(t-h,\ t+h)].$$

Если $p_0(t)=1$ (т. е. t — регулярная точка потока), то

$$1-v_0(t-h,\ t+h)\to 0\ \ (h\to+0);$$

поэтому при достаточно малом $h>0$

$$\sum_{k=1}^{m} \psi_k(t-h,\ t+h)<\varepsilon,$$

и неравенство (4) дает при достаточно малом $h>0$

$$\Lambda(t+h)-\Lambda(t-h)<2\varepsilon;$$

следовательно, $\Lambda(t+0)=\Lambda(t-0)$, и второе утверждение леммы 3 также доказано.

Лемма 4. *Финитный поток может иметь не более счетного множества сингулярных точек.*

Доказательство непосредственно вытекает из леммы 3.

Лемма 5. *Пусть L — финитный поток, и L' — поток тех событий потока L, моменты наступления которых отличны от некоторого фиксированного момента t_0. Тогда t_0 есть регулярная точка потока L'.*

Доказательство. Пусть $0<h_0<t_0$. Обозначим через w_n вероятность того, что в области

$$\frac{h_0}{2^n}<|t-t_0|\leqslant\frac{h_0}{2^{n-1}} \qquad (\Delta_n)$$

наступит по меньшей мере одно событие потока L'. Если промежуток $(t_0-h_0,\ t_0+h_0)$ содержит события потока L', то то же самое должно иметь место и для по меньшей мере одной из областей (Δ_n) $(n=1,\ 2,\ \ldots)$. Поэтому, если обозначить вообще через $w(\alpha,\ \beta)$ вероятность наличия событий потока L' в промежутке $(\alpha,\ \beta)$, мы будем иметь

$$w(t_0-h_0,\ t_0+h_0)\leqslant\sum_{n=1}^{\infty} w_n.$$

Аналогичным образом находим для любого натурального числа k

$$w\left(t_0-\frac{h_0}{2^k},\ t_0+\frac{h_0}{2^k}\right)\leqslant\sum_{n=k+1}^{\infty} w_n. \qquad (5)$$

Пусть теперь v_{nk} означает вероятность того, что в область (Δ_n) попадет k событий потока L'; тогда

$$w_n=\sum_{k=1}^{\infty} v_{nk}\leqslant\sum_{k=1}^{\infty} kv_{nk}=\mu_n, \qquad (6)$$

где μ_n означает математическое ожидание числа событий потока L' в области (Δ_n). Так как поток L (а следовательно, и поток L') — финитный, то ряд $\sum_{n=1}^{\infty}\mu_n$ сходится; в силу (6) сходится поэтому и ряд $\sum_{n=1}^{\infty}w_n$. Но тогда (5) показывает, что $w(t_0-h,\ t_0+h)\to 0$ $(h\to+0)$, и, следовательно, t_0 есть регулярная точка потока L', ч. т. д.

Условимся во всем дальнейшем называть финитный поток *регулярным*, если он не имеет ни одной сингулярной точки, так что любое $t\,(0\leqslant t<+\infty)$ является для него регулярной точкой. Тогда из леммы 3 непосредственно следует

Лемма 6. *Необходимым и достаточным условием регулярности потока служит непрерывность его ведущей функции.*

Наконец, нам понадобится еще следующая

Лемма 7. *Если* $0<\alpha<\beta$ *и данный поток — регулярный, то равномерно в отрезке* $\alpha\leqslant t\leqslant\beta$

$$v_0(t-h,\ t+h)\to 1 \quad (h\to+0).$$

Доказательство. В силу леммы 1

$$1-v_0(t-h,\ t+h)=\sum_{k=1}^{\infty}v_k(t-h,\ t+h)\leqslant$$

$$\leqslant\sum_{k=1}^{\infty}kv_k(t-h,\ t+h)=\Lambda(t+h)-\Lambda(t-h),$$

и утверждение леммы 7 вытекает из равномерной непрерывности функции $\Lambda(t)$ в любом конечном отрезке.

§ 3. Леммы о регулярных потоках без последействия

Лемма 8. *Для любого регулярного потока без последействия и любых* $\alpha,\ \beta\ (0\leqslant\alpha<\beta)$

$$v_0(\alpha,\ \beta)>0.$$

Доказательство. Допустим, что $v_0(\alpha,\ \beta)=0$, и положим $\gamma=\frac{1}{2}(\alpha+\beta)$. Так как данный поток — без последействия, то $v_0(\alpha,\ \beta)=v_0(\alpha,\ \gamma)\,v_0(\gamma,\ \beta)$ и, следовательно, либо $v_0(\alpha,\ \gamma)=0$, либо $v_0(\gamma,\ \beta)=0$. Безграничное продолжение этого процесса известным образом приводит нас к точке ξ, обладающей следующим свойством: при любом $h>0$ промежуток $(\xi-h, \xi+h)$ содержит другой промежуток (a, b), для которого $v_0(a, b)=0$; но тогда и подавно при любом $h>0$ мы имеем $v_0(\xi-h, \xi+h)=0$, откуда $p_0(\xi)=0$, что стоит в противоречии с регулярностью данного потока. Лемма 8 доказана.

Условимся во всем дальнейшем называть функцию

$$\Phi(x, \alpha, \beta)=\sum_{k=0}^{\infty} v_k(\alpha, \beta)\, x^k$$

(где ряд в правой части при любых α, β сходится по меньшей мере для $|x|\leqslant 1$) *производящей функцией* данного потока. Если этот поток — регулярный и без последействия, то по лемме 8 $v_0(\alpha, \beta)>0$, и, следовательно, разложение

$$\ln\Phi(x, \alpha, \beta)=\sum_{k=0}^{\infty} \omega_k(\alpha, \beta)\, x^k$$

функции $\ln\Phi(x, \alpha, \beta)$ по степеням x сходится в некоторой окрестности точки $x=0$. По известным свойствам производящих функций мы имеем для потоков без последействия при $\alpha<\gamma<\beta$

$$\Phi(x, \alpha, \beta)=\Phi(x, \alpha, \gamma)\,\Phi(x, \gamma, \beta),$$

и, следовательно,

$$\ln\Phi(x, \alpha, \beta)=\ln\Phi(x, \alpha, \gamma)+\ln\Phi(x, \gamma, \beta),$$

что в свою очередь имеет следствием

$$\omega_k(\alpha, \beta)=\omega_k(\alpha, \gamma)+\omega_k(\gamma, \beta) \quad (k\geqslant 0).$$

Поэтому, если положить $\omega_k(0, t)=\chi_k(t)$ $(t\geqslant 0, \ k\geqslant 0)$, то $\omega_k(\alpha, \beta)=\chi_k(\beta)-\chi_k(\alpha)$, и мы получаем в некоторой окрестности точки $x=0$ разложение

$$\ln\Phi(x, \alpha, \beta)=\sum_{k=0}^{\infty}[\chi_k(\beta)-\chi_k(\alpha)]\, x^k. \qquad (7)$$

Для установления общего вида регулярного потока без последействия имеют существенное значение свойства функций $\chi_k(t)$; изучению этих свойств и будет посвящен конец настоящего параграфа.

Лемма 9. *При* $0 \leqslant \alpha < \beta$, $n \geqslant 1$ *и достаточно малом* $x > 0$

$$\frac{\partial^n \ln \Phi}{\partial x^n} = \frac{\Phi^{(n)}}{\Phi} + \sum a_{i_1 \ldots i_n} \frac{\Phi^{(i_1)} \ldots \Phi^{(i_n)}}{\Phi^n}; \tag{8}$$

здесь положено $\Phi = \Phi(x, \alpha, \beta)$, $\Phi^{(i)} = \frac{\partial^i \Phi(x, \alpha, \beta)}{\partial x^i}$; *коэффициенты* $a_{i_1 \ldots i_n}$ *зависят только от* n *и индексов* $i_1, \ldots, i_n$ *(и не зависят от* α, β *и* x*); суммирование распространяется на все системы индексов* $i_1, \ldots, i_n$, *подчиненные условиям*

$$0 \leqslant i_k < n \quad (1 \leqslant k \leqslant n), \quad \sum_{k=1}^{n} i_k = n.$$

Соотношение (8) представляет собою элементарную формулу дифференциального исчисления, имеющую место для любой n-кратно дифференцируемой функции $\Phi(x)$ $[\Phi(0) = 0]$. Доказательство автоматически проводится индукцией по n [при $n = 1$ формула (8) тривиальна] и может быть предоставлено читателю.

Лемма 10. *Пусть* $0 \leqslant \alpha < \beta$, $n \geqslant 1$ *и данный поток — регулярный и без последействия. Пусть* $\alpha = \alpha_0 < \alpha_1 < \ldots < \alpha_s = \beta$, $d(s) = \max\limits_{1 \leqslant r \leqslant s} (\alpha_r - \alpha_{r-1})$. *Тогда в формуле* (7)

$$\chi_n(\beta) - \chi_n(\alpha) = \lim_{d(s) \to 0} \sum_{r=1}^{s} \frac{v_n(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}. \tag{9}$$

Предварительное замечание. В силу леммы 8 в правой части (9) все знаменатели отличны от нуля.

Доказательство. В ходе последующего рассуждения мы под $\Phi^{(i)}$ $(i = 0, 1, \ldots)$ всегда будем понимать значение производной $\frac{\partial^i \Phi(x, \alpha, \beta)}{\partial x^i}$ при $x = 0$.

Так как в силу (7)

$$\chi_n(\beta) - \chi_n(\alpha) = \frac{1}{n!} \left\{ \frac{\partial^n \ln \Phi(x, \alpha, \beta)}{\partial x^n} \right\}_{x=0} \quad (n \geqslant 0),$$

то соотношение (8) при $n \geqslant 1$ дает

$$\chi_n(\beta) - \chi_n(\alpha) = \frac{1}{n!} \left\{ \frac{\Phi^{(n)}}{\Phi} + \sum a_{i_1 \ldots i_n} \frac{\Phi^{(i_1)} \ldots \Phi^{(i_n)}}{\Phi^n} \right\}.$$

Так как, очевидно, $\Phi^{(i)} = i! v_i(\alpha, \beta)$ $(i \geqslant 0)$, то отсюда

$$\chi_n(\beta) - \chi_n(\alpha) = \frac{v_n(\alpha, \beta)}{v_0(\alpha, \beta)} + \\ + \sum b_{i_1 \ldots i_n} \frac{v_{i_1}(\alpha, \beta) \ldots v_{i_n}(\alpha, \beta)}{[v_0(\alpha, \beta)]^n}, \qquad (10)$$

где область суммирования определяется соотношениями

$$0 \leqslant i_k < n \quad (1 \leqslant k \leqslant n), \qquad \sum_{k=1}^{n} i_k = n,$$

а коэффициенты $b_{i_1 \ldots i_n}$, подобно прежним $a_{i_1 \ldots i_n}$, не зависят от α и β. Если мы разобьем промежуток (α, β) на s промежутков (α_{i-1}, α_i) $(1 \leqslant i \leqslant s)$, как это описано в формулировке леммы 10, то соотношение (10) имеет место для всех промежутков (α_{i-1}, α_i), и притом с одними и теми же коэффициентами $b_{i_1 \ldots i_n}$. Написав все эти соотношения и складывая их почленно, мы получаем

$$\chi_n(\beta) - \chi_n(\alpha) = \sum_{r=1}^{s} \frac{v_n(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} + \\ + \sum b_{i_1 \ldots i_n} \sum_{r=1}^{s} \frac{v_{i_1}(\alpha_{r-1}, \alpha_r) \ldots v_{i_n}(\alpha_{r-1}, \alpha_r)}{[v_0(\alpha_{r-1}, \alpha_r)]^n}. \qquad (11)$$

Чтобы доказать лемму 10, нам надо только установить, что при $d(s) = \max\limits_{1 \leqslant r \leqslant s} (\alpha_r - \alpha_{r-1}) \to 0$ второе слагаемое правой части формулы (11) становится бесконечно малым. Это же в свою очередь будет доказано, если мы убедимся, что для любой фиксированной, принадлежащей области суммирования системы индексов $i_1, i_2, \ldots, i_n$ величина

$$S = \sum_{r=1}^{s} \frac{v_{i_1}(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} \frac{v_{i_2}(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)} \cdots \frac{v_{i_n}(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}$$

стремится к нулю при $d(s) \to 0$.

С этой целью заметим прежде всего, что при $n = 1$ область суммирования $0 \leqslant i_k < n$ $(1 \leqslant k \leqslant n)$, $\sum\limits_{k=1}^{n} i_k = n$ становится пустой (второе слагаемое правой части (11). Поэтому мы допустим, что $n>1$ и что для всех чисел $<n$ лемма 10 уже доказана).

Обратимся теперь к исследованию величины S. В силу леммы 7 при любом $\varepsilon>0$ и достаточно малом $d(s)$

$$1-v_0(\alpha_{r-1}, \alpha_r)<\varepsilon \quad (1\leqslant r\leqslant s),$$

и, следовательно,

$$\frac{1-v_0(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}<\frac{\varepsilon}{1-\varepsilon} \quad (1\leqslant r\leqslant s).$$

Но при $i_k>0$ мы имеем

$$v_{i_k}(\alpha_{r-1}, \alpha_r)\leqslant 1-v_0(\alpha_{r-1}, \alpha_r),$$

и, следовательно,

$$\frac{v_{i_k}(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}<\frac{\varepsilon}{1-\varepsilon} \quad (1\leqslant r\leqslant s),$$

в то время как при $i_k=0$ это отношение равно 1. Но в силу условий $i_k<n\,(1\leqslant k\leqslant n)$ и $\sum\limits_{k=1}^{n} i_k=n$ среди индексов $i_k\,(1\leqslant k\leqslant n)$ найдется по крайней мере два положительных. Пусть, например, такими будут индексы i и l. Тогда приведенные выше оценки дают

$$S\leqslant\sum_{r=1}^{s}\frac{v_i(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}\,\frac{v_l(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}<\frac{\varepsilon}{1-\varepsilon}\sum_{r=1}^{s}\frac{v_l(\alpha_{r-1}, \alpha_r)}{v_0(\alpha_{r-1}, \alpha_r)}.$$

Так как $l<n$, то мы можем принять, что лемма 10 уже доказана с l вместо n; поэтому последняя сумма при $d(s)\to 0$ имеет пределом $\chi_l(\beta)-\chi_l(\alpha)$. А так как ε при достаточно малом $d(s)$ может быть выбрано сколь угодно малым, то $S\to 0$ при $d(s)\to 0$, чем лемма 10 и доказана.

Лемма 11. *При $n\geqslant 1$, $\chi_n(t)$ есть неубывающая функция от t; ряд*

$$\sum_{n=0}^{\infty} n\,[\chi_n(\beta)-\chi_n(\alpha)]$$

сходится при любых α, β; *для всех* $t \geqslant 0$

$$\sum_{n=0}^{\infty} \chi_n(t) = 0.$$

Доказательство. Первое утверждение непосредственно вытекает из леммы 10, в силу которой $\chi_n(\beta) - \chi_n(\alpha) \geqslant 0$ $(\beta \geqslant \alpha)$. Далее, из леммы 10 следует при любом $m \geqslant 1$

$$\sum_{l=1}^{m} l\,[\chi_l(\beta) - \chi_l(\alpha)] = \lim_{d(s)\to 0} \sum_{r=1}^{s} \frac{1}{v_0(\alpha_{r-1}, \alpha_r)} \sum_{l=1}^{m} l v_l(\alpha_{r-1}, \alpha_r) \leqslant$$

$$\leqslant \lim_{d(s)\to 0} \sup \sum_{r=1}^{s} \frac{\Lambda(\alpha_r) - \Lambda(\alpha_{r-1})}{v_0(\alpha_{r-1}, \alpha_r)} = \Lambda(\beta) - \Lambda(\alpha)$$

[последнее равенство вытекает из того, что в силу леммы 7 $v_0(\alpha_{r-1}, \alpha_r)$ при $d(s) \to 0$ стремится к 1 равномерно относительно r $(1 \leqslant r \leqslant s)$]. Так как правая часть этого равенства не зависит от m, то доказано и второе утверждение леммы. Наконец, из этого второго утверждения следует, что формула (7) имеет место при $|x| \leqslant 1$. В частности, при $x = 1$ это дает

$$0 = \ln \Phi(1, \alpha, \beta) = \sum_{k=0}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)],$$

чем доказано и третье утверждение леммы 11.

Лемма 12. *Каждая функция* $\chi_k(t)$ $(k = 0, 1, \ldots)$ *всюду непрерывна.*

Доказательство. Допустим, что для некоторого $k \geqslant 1$ и некоторого $t \geqslant 0$

$$\chi_k(t+0) - \chi_k(t-0) > 0.$$

Так как все функции $\chi_k(t)$ $(k \geqslant 1)$ — неубывающие, то тогда

$$\chi_0(t+0) - \chi_0(t-0) = -\sum_{k=1}^{\infty} [\chi_k(t+0) - \chi_k(t-0)] < 0.$$

Но

$$\ln \Phi(0, t-h, t+h) = \chi_0(t+h) - \chi_0(t-h).$$

Поэтому при достаточно малом $h>0$

$$\ln \Phi(0,\ t-h,\ t+h)<-\alpha,$$

или

$$\Phi(0,\ t-h,\ t+h)=v_0(t-h,\ t+h)<e^{-\alpha},$$

где α — положительная постоянная. А отсюда следует, что

$$\lim_{h\to+0} v_0(t-h,\ t+h)\leqslant e^{-\alpha}<1,$$

что противоречит допущению о регулярности данного потока. Этим лемма 12 доказана.

§ 4. Общая форма регулярных потоков без последействия

Леммы § 3 в своей совокупности показывают, что производящая функция $\Phi(x,\ \alpha,\ \beta)$ любого регулярного потока без последействия имеет вид

$$\Phi(x,\ \alpha,\ \beta)=\exp\left\{\sum_{k=0}^{\infty}[\chi_k(\beta)-\chi_k(\alpha)]\,x^k\right\}, \tag{12}$$

где функции $\chi_k(t)$ $(\chi_k(0)=0)$ обладают следующими свойствами:

1° Все функции $\chi_k(t)$ всюду непрерывны.

2° При любом k функция $\chi_k(t)$ монотонна (неубывающая при $k>0$ и невозрастающая при $k=0$).

3° Ряд $\sum\limits_{k=0}^{\infty} k\chi_k(t)$ сходится при любом $t\geqslant 0$.

4° Сумма $\sum\limits_{k=0}^{\infty} \chi_k(t)=0$ при любом $t\geqslant 0$.

Теперь мы докажем обратное предложение: *если* $\Phi(x,\ \alpha,\ \beta)$ *есть функция вида* (12), *где функции* $\chi_k(t)$ *удовлетворяют требованиям* 1°—4°, *то существует регулярный поток без последействия, производящая функция которого совпадает с функцией* $\Phi(x,\ \alpha,\ \beta)$.

Так как в силу свойства 2° разность $\chi_k(\beta)-\chi_k(\alpha)$ при $k>0$ и $\alpha<\beta$ неотрицательна, то, как известно, существует поток без последействия L_k, для которого

$$v_m^{(k)}(\alpha,\ \beta)=\begin{cases}\exp\{-[\chi_k(\beta)-\chi_k(\alpha)]\}\dfrac{[\chi_k(\beta)-\chi_k(\alpha)]^r}{r!} & (m=kr)\\ 0\ \ (m\not\equiv 0 \bmod k)\end{cases}$$

[$v_m^{(k)}(\alpha, \beta)$ означает вероятность того, что в промежутке (α, β) наступит m событий потока L_k]; требуемое для этого условие, чтобы при $\alpha < \gamma < \beta$ всего было

$$v_m^{(k)}(\alpha, \beta) = \sum_{i=0}^{m} v_i^{(k)}(\alpha, \gamma)\, v_{m-i}^{(k)}(\gamma, \beta),$$

легко проверяется непосредственным подсчетом. Производящая функция потока L_k равна

$$\Phi_k(x, \alpha, \beta) = \sum_{r=0}^{\infty} v_{rk}^{(k)}(\alpha, \beta)\, x^{rk} =$$

$$= \exp\{-[\chi_k(\beta) - \chi_k(\alpha)]\} \sum_{r=0}^{\infty} \frac{[\chi_k(\beta) - \chi_k(\alpha)]^r}{r!} x^{rk} =$$

$$= \exp\{[\chi_k(\beta) - \chi_k(\alpha)](x^k - 1)\}.$$

Рассмотрим теперь суперпозицию L определенных нами потоков $L_1, L_2, \ldots, L_n, \ldots$, которые будем предполагать взаимно независимыми. Легко убедиться, что производящая функция потока L (который, очевидно, есть поток без последействия) совпадает с

$$F(x, \alpha, \beta) = \prod_{k=1}^{\infty} \Phi_k(x, \alpha, \beta).$$

В самом деле, если положить

$$F(x, \alpha, \beta) = \sum_{r=0}^{\infty} v_r(\alpha, \beta)\, x^r$$

(этим соотношением определяется число $v_r(\alpha, \beta)$), то, очевидно, для любого $r \geqslant 0$

$$v_r(\alpha, \beta) = \sum v_{r_1}^{(1)}(\alpha, \beta)\, v_{r_2}^{(2)}(\alpha, \beta) \ldots v_{r_k}^{(k)}(\alpha, \beta) \ldots, \tag{13}$$

где суммирование распространяется на область

$$r_k \geqslant 0 \quad (k \geqslant 1), \quad \sum_{k=1}^{\infty} r_k = r.$$

Но в силу взаимной независимости потоков L_k сумма в правой части (13), очевидно, равна вероятности того, что в промежутке (α, β) наступит r событий суммарного потока L. В силу (13), число $v_r(\alpha, \beta)$ совпадает с этой вероятностью, а это и значит, что $F(x, \alpha, \beta)$ есть производящая функция потока L.

Но в силу свойства 4° мы имеем

$$F(x, \alpha, \beta) = \prod_{k=1}^{\infty} \Phi_k(x, \alpha, \beta) =$$
$$= \exp\left\{\sum_{k=1}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)](x^k - 1)\right\} =$$
$$= \exp\left\{\sum_{k=1}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)]x^k - \sum_{k=1}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)]\right\} =$$
$$= \exp\left\{\sum_{k=0}^{\infty} [\chi_k(\beta) - \chi_k(\alpha)]x^k\right\} = \Phi(x, \alpha, \beta).$$

Этим показано, что данная функция $\Phi(x, \alpha, \beta)$ действительно есть производящая функция некоторого потока без последействия L. Нам остается только убедиться, что построенный нами поток L — регулярный.

В силу свойства 3° для любых $\varepsilon > 0$ и $t > 0$ найдется такое $N > 0$, что

$$\sum_{k>N} k\chi_k(t) < \varepsilon. \tag{14}$$

Так как поток L_k $(k = 1, 2, \ldots)$ имеет непрерывную (в силу свойства 1°) ведущую функцию $k\chi_k(t)$, то он регулярен в силу леммы 6. Следовательно, для любой фиксированной точки γ $(0 < \gamma < t)$ при достаточно малом $h_0 > 0$ и при $0 < h < h_0$

$$\sum_{k=1}^{N} [1 - v_0^{(k)}(\gamma - h, \gamma + h)] < \varepsilon. \tag{15}$$

Но при $0 \leqslant \alpha < \beta$

$$1 - v_0^{(k)}(\alpha, \beta) = \sum_{p=1}^{\infty} v_p^{(k)}(\alpha, \beta) \leqslant \sum_{p=1}^{\infty} p v_p^{(k)}(\alpha, \beta) =$$
$$= k[\chi_k(\beta) - \chi_k(\alpha)].$$

[так как $k\chi_k(t)$ есть ведущая функция потока L_k]. Из (15) и (14) следует поэтому при $0<h<h_0$, $0<\gamma-h<\gamma+h<t$

$$\sum_{k=1}^{\infty}[1-v_0^{(k)}(\gamma-h,\ \gamma+h)]\leqslant\sum_{k=1}^{N}[1-v_0^{(k)}(\gamma-h,\gamma+h)]+$$
$$+\sum_{k>N}[1-v_0^{(k)}(0,\ t)]<\varepsilon+\sum_{k>N}k\chi_k(t)<2\varepsilon.$$

Поэтому для суммарного потока L мы получаем при $0<h<h_0$

$$1-v_0(\gamma-h,\ \gamma+h)\leqslant\sum_{k=1}^{\infty}[1-v_0^{(k)}(\gamma-h,\ \gamma+h)]<2\varepsilon$$

[$1-v_0(\gamma-h,\ \gamma+h)$ есть вероятность того, что в промежутке $(\gamma-h,\ \gamma+h)$ наступит по меньшей мере одно событие потока L]. Но это означает, что γ — регулярная точка потока L; а так как выбор этой точки произволен, то поток L — регулярный, ч. т. д.

Мы можем, таким образом, сформулировать результаты §§ 3 и 4 в виде следующего предложения.

Теорема I. *Каждый регулярный поток без последействия имеет производящую функцию* $\Phi(x,\ \alpha,\ \beta)$ *вида* (12), *где функции* $\chi_k(t)$ *удовлетворяют требованиям* 1°—4°; *обратно, каждая такая функция* $\Phi(x,\ \alpha,\ \beta)$ *есть производящая функция некоторого регулярного потока без последействия.*

Посмотрим теперь еще, какой вид принимает общая форма (12) для некоторых известных классов потоков без последействия.

1. Пусть данный поток при любом $\alpha\geqslant 0$ удовлетворяет требованию

$$1-v_0(\alpha,\ \beta)-v_1(\alpha,\ \beta)=\sum_{k=2}^{\infty}v_k(\alpha,\ \beta)=o(\beta-\alpha)$$
$$(\beta\to\alpha+0)$$

ср. [1], § 1), и пусть при любом $\alpha\geqslant 0$ существует предел

$$\lim_{\beta\to\alpha+0}\frac{1-v_0(\alpha,\ \beta)}{\beta-\alpha}=\lambda(\alpha).$$

Тогда формула (9) леммы 10 легко дает

$$\chi_1(\beta)-\chi_1(\alpha)=\int_\alpha^\beta \lambda(u)\,du, \quad \chi_k(\beta)-\chi_k(\alpha)=0 \quad (k>1),$$

откуда в силу свойства 4°

$$\chi_0(\beta)-\chi_0(\alpha)=-\int_\alpha^\beta \lambda(u)\,du,$$

и формула (12) получает вид

$$\Phi(x,\ \alpha,\ \beta)=\exp\left\{(x-1)\int_\alpha^\beta \lambda(u)\,du\right\}.$$

Это — давно известный результат (см. [1], стр. 17 *), или [6], стр. 183).

2. Пусть данный поток — стационарный; положим

$$\lim_{t\to+0}\frac{1-v_0(t)}{t}=\lambda, \qquad \lim_{t\to+0}\frac{v_k(t)}{t}=\lambda p_k \quad (k>0);$$

тогда формула (9) леммы 10 дает при любом $t\geqslant 0$

$$\chi_k(t)=\lambda t p_k \quad (k\geqslant 1), \quad \chi_0(t)=-\lambda t,$$

и формула (12) получает вид

$$\Phi(x,\ 0,\ t)=\exp\left\{\lambda t\left(\sum_{k=1}^{\infty} p_k x^k-1\right)\right\}.$$

Это полностью совпадает с формулой (8.6) ([1], стр. 31) и является известным результатом Редхеффера [2] и автора.

§ 5. Общая форма финитных потоков без последействия

До сих пор мы рассматривали только регулярные потоки, т. е. такие, ведущая функция $\Lambda(t)$ которых всюду непрерывна. Теперь мы обратимся к рассмотрению любого финитного потока без последействия, так что функция $\Lambda(t)$ всюду

*) Пользуюсь случаем указать на грубую опечатку: под знаком последней суммы стр. 17 [1] пропущен множитель x^k.

конечна, но может быть как угодно разрывной. Как мы знаем из § 2, точки разрыва функции $\Lambda(t)$ — это сингулярные точки данного потока.

Условимся теперь называть поток S *сингулярным*, если он обладает следующими тремя свойствами:

1) События потока S могут наступать лишь в некоторые заранее определенные моменты времени; эти моменты, которые мы будем называть *ступенями* потока S, образуют не более чем счетное множество.

2) Числа событий, падающих на различные ступени, представляют собою взаимно независимые случайные величины.

3) Пусть $q_k^{(i)}$ ($k \geqslant 0$, $i \geqslant 1$) означает вероятность того, что на ступень t_i придется k событий; тогда при любом $t > 0$

$$\sum_{0 \leqslant t_i \leqslant t} \sum_{k=0}^{\infty} k q_k^{(i)} < +\infty.$$

Свойство 2) выражает собой отсутствие последействия, а свойство 3) — финитность потока S. Очевидно, сингулярный поток однозначно определяется заданием чисел t_i и $q_k^{(i)}$ ($i \geqslant 1$, $k \geqslant 0$). Структура такого потока представляется поэтому совершенно прозрачной: выбор чисел t_i и $q_k^{(i)}$ подчинен только требованиям $t_i \geqslant 0$, $q_k^{(i)} \geqslant 0$ и $\sum\limits_{k=0}^{\infty} q_k^{(i)} = 1$ ($i = 1, 2, \ldots$), в остальном же остается произвольным.

Пусть теперь нам дан произвольный финитный поток K без последействия. Обозначим через $t_1, t_2, \ldots$ сингулярные точки этого потока, занумерованные в произвольном порядке. Мы будем рассматривать поток K как суперпозицию двух потоков R и S. Будем считать некоторое событие потока K принадлежащим потоку S, если оно наступает в один из моментов $t_1, t_2, \ldots$, в противном случае мы будем считать его принадлежащим потоку R. Можно убедиться (чего мы здесь делать не будем), что оба потока S и R — без последействия, и что они взаимно независимы (ср. работу Марчевского [5]).

Докажем в первую очередь регулярность потока R. Пусть K_i ($i \geqslant 1$) означает совокупность событий потока K, моменты наступления которых отличны от t_i. В силу леммы 5 точка t_i является регулярной для потока K_i, а значит и подавно — для потока $R \subset K_i$. Но точка t, отличная от всех t_i, по самому определению чисел t_i регулярна для потока K, а значит и для $R \subset K$. Таким образом, поток R не имеет сингулярных точек и, следовательно, регулярен.

Что касается потока S, то сингулярность его очевидна из самого определения; ступенями его служат числа t_i $(i \geqslant 1)$. Мы покажем еще, что его характеристические вероятности $q_k^{(i)}$ соответственно совпадают с определенными в § 2 числами

$$p_k(t_i) = \lim_{h \to +0} v_k(t_i - h,\ t_i + h) \quad (k \geqslant 0,\ i \geqslant 1).$$

Пусть поток K_i $(i \geqslant 1)$ определяется как выше, и пусть $v_k^{(i)}(\alpha, \beta)$ означает вероятность того, что в промежутке (α, β) наступит k событий потока K_i. Тогда для потока K мы, очевидно, имеем при $i \geqslant 1$, $k \geqslant 0$

$$v_k(t_i - h,\ t_i + h) = \sum_{l=0}^{k} q_l^{(i)} v_{k-l}^{(i)}(t_i - h,\ t_i + h). \tag{16}$$

Так как t_i (в силу леммы 5) есть регулярная точка потока K_i, то при $h \to +0$

$$v_{k-l}^{(i)}(t_i - h, t_i + h) \to \begin{cases} 1 & (l = k), \\ 0 & (l < k). \end{cases}$$

Поэтому соотношение (16) при $h \to +0$ дает в пределе

$$p_k(t_i) = \lim_{h \to +0} v_k(t_i - h,\ t_i + h) = q_k^{(i)} \quad (i \geqslant 1,\ k \geqslant 0),$$

ч. т. д.

Резюмируя результаты настоящего параграфа, мы приходим к следующему предложению.

Теорема II. *Всякий финитный поток без последействия может рассматриваться как суперпозиция двух взаимно независимых потоков того же типа, из которых один — регулярный, а другой — сингулярный; ступенями t_i сингулярной компоненты служат при этом сингулярные точки данного потока, а характеристическими вероятностями $q_k^{(i)}$ — соответствующие числа $p_k(t_i)$ $(i \geqslant 1, k \geqslant 0)$ данного потока.*

Непосредственно очевидно, что и обратно — любая суперпозиция описанного типа представляет собою некоторый финитный поток без последействия.

Теоремы I и II вместе взятые дают легко обозримое описание совокупности всех финитных потоков случайных событий без последействия, что и было целью настоящего исследования.

ЛИТЕРАТУРА

[1] А. Я. Хинчин, Математические методы теории массового обслуживания, Труды Матем. инст. им. Стеклова, 49, (1955), 1—122.
[2] R. M. Redheffer, Math. Magazine, 26, (1953), 185—188.
[3] C. Palm, Intensitätsschwankungen im Fernsprechverkehr, Ericsson Technics, 44 (1943), 1—189.
[4] L. Takacs, Acta math. hung., 5, (1954), 203—236.
[5] D. Marzewski, Studia math. (Warszawa), 13, No 1, (1953), 130—136.
[6] Фрай Торнтон, Теория вероятностей для инженеров М. —Л., ГТТИ, 1934.

О ПУАССОНОВСКИХ ПОТОКАХ СЛУЧАЙНЫХ СОБЫТИЙ*)

§ 1. Постановка задачи

В классических приложениях теории потоков случайных событий (например, в теории распада атомов и теории массового обслуживания) рассматривается, как известно, в большинстве случаев лишь простейший вид таких потоков, когда вероятность наступления k событий в произвольно взятом промежутке времени длины t дается формулой

$$v_k(t) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \qquad (1)$$

где $\lambda > 0$ — постоянная; такой поток предполагается обычно потоком *без последействия*; это означает, что для любой конечной подгруппы попарно не пересекающихся промежутков времени числа наступающих в этих промежутках событий представляют собой взаимно независимые случайные величины.

Причиной столь широкой распространенности потоков специальной формы (1) служит, как известно, то, что эта форма для некоторого весьма обширного класса потоков фактически является единственно возможной. Именно, если к требованиям стационарности и отсутствия последействия добавить еще условие

$$\psi_2(t) = o(t) \qquad (t \to 0), \qquad (2)$$

где $\psi_2(t)$ означает вероятность наступления по меньшей мере двух событий в промежутке времени длины t, то рассматриваемый поток обязательно будет иметь форму (1)

---

*) Теория вероятностей и ее применения, т. 1, вып. 3, 1956, стр. 320—327.

(см., например, [1], гл. 1). Обратно, всякий поток без последействия, имеющий форму (1), очевидно стационарен и, как легко убедиться, удовлетворяет условию (2). Таким образом, три перечисленных свойства — отсутствие последействия, стационарность и условие (2) — полностью характеризуют собою класс «пуассоновских» потоков (1).

Однако в приложениях все чаще появляются задачи, требующие исследования *нестационарных* потоков событий без последействия. За исходный пункт такого исследования обычно бывает целесообразно принять «ведущую функцию» $\Lambda(t)$ данного потока — математическое ожидание числа событий, наступающих в полуоткрытом промежутке $[0,\ t)$. В стационарном случае $\Lambda(t)=\lambda t$, где $\lambda>0$ — постоянная; в общем случае $\Lambda(t)$ — любая неубывающая функция от t $[\Lambda(t)=0$ при $t\leqslant 0]$. Обычно при этом предполагают, что поток имеет «пуассоновскую» форму

$$v_k(\alpha,\ \beta)=e^{-[\Lambda(\beta)-\Lambda(\alpha)]}\frac{[\Lambda(\beta)-\Lambda(\alpha)]^k}{k!} \tag{3}$$

$[v_k(\alpha,\ \beta)$ означает вероятность того, что в промежутке $(\alpha,\ \beta)$ наступит k событий данного потока; задание всех функций $v_k(\alpha,\ \beta)$ $(k\geqslant 0,\ 0\leqslant\alpha<\beta)$, очевидно, дает полное описание данного потока без последействия]. Однако форма (3) при заданной ведущей функции $\Lambda(t)$ отнюдь не является единственно возможной: ведь даже в стационарном случае, как было упомянуто выше, форма (3) [в этом случае совпадающая с (1)] имеет место лишь при выполнении дополнительного требования (2). С другой стороны, в моей недавней работе [2] непосредственно показано, что общая форма потоков без последействия с данной ведущей функцией $\Lambda(t)$, даже в случае непрерывной $\Lambda(t)$ выходит далеко за пределы формулы (3).

Таким образом, естественно возникает вопрос о том, какими свойствами должен обладать поток без последействия для того, чтобы иметь простейшую (почти всегда принимаемую в приложениях) «пуассоновскую» форму (3); дело идет при этом об отыскании возможно простого (необходимого и достаточного) дополнительного требования, подобно условию (2) в стационарном случае. Решение этой задачи означало бы распространение на нестационарный случай той фундаментальной теоремы, которую мы приводили в самом начале для случая стационарного. От данного потока при этом, кроме отсутствия последействия, мы будем требовать только существование ведущей функции $\Lambda(t)$ [т. е. $\Lambda(t) < +\infty$ при $t < +\infty$].

Как покажет исследование, наиболее интересным при этом будет случай, когда функция $\Lambda(t)$ непрерывна; точки же разрыва этой функции, если они имеются, осложняют решение задачи лишь в весьма незначительной степени.

§ 2. Регулярный случай

Условимся ради краткости называть поток без последействия регулярным, если его ведущая функция $\Lambda(t)$ всюду непрерывна. В этом параграфе мы будем рассматривать только регулярные потоки. В частности, регулярным будет любой стационарный поток без последействия [так как для такого потока $\Lambda(t) = \lambda t$, $0 < \lambda < +\infty$].

Пусть для любого стационарного потока и любого $k \geqslant 0$, $\psi_k(t)$ означает вероятность того, что в промежутке длины t наступит по меньшей мере k событий данного потока. Очевидно, мы всегда имеем

$$\psi_k(t) = \sum_{l=k}^{\infty} v_l(t) \quad (k \geqslant 0).$$

Убедимся прежде всего, что для стационарного потока условие (2) равносильно соотношению

$$\frac{\psi_2(t)}{\psi_1(t)} \to 0 \quad (t \to 0). \tag{4}$$

С этой целью заметим, что, как известно ([1], стр. 24), для любого стационарного потока при $t \to +0$

$$\frac{\psi_1(t)}{t} \to \alpha \quad (0 < \alpha \leqslant +\infty);$$

но в нашем случае

$$\psi_1(t) = \sum_{k=1}^{\infty} v_k(t) \leqslant \sum_{k=1}^{\infty} k v_k(t) = \Lambda(t) = \lambda t$$

и, следовательно, $\alpha \leqslant \lambda < +\infty$; таким образом, при $t \to +0$ отношение $\psi_1(t)/t$ в нашем случае стремится к некоторому конечному положительному пределу, и соотношение

$$\frac{\psi_2(t)}{t}=\frac{\psi_2(t)}{\psi_1(t)}\frac{\psi_1(t)}{t}$$

показывает, что требования (2) и (4) действительно равносильны.

Условимся теперь называть любой (вообще говоря, нестационарный) поток ординарным, если при любых $t>0$ и $\varepsilon>0$ найдется такое $\delta>0$, что для $0\leqslant\alpha<\beta<\alpha+\delta<t$ имеет место неравенство

$$\psi_2(\alpha,\ \beta)\leqslant\varepsilon\psi_1(\alpha,\ \beta). \tag{5}$$

При этом $\psi_k(\alpha,\ \beta)$ означает вероятность того, что в промежутке $(\alpha,\ \beta)$ наступит по меньшей мере k событий данного потока. Вместо этого несколько сложного определения ординарности мы могли бы и просто сказать, что

$$\frac{\psi_2(\alpha,\ \beta)}{\psi_1(\alpha,\ \beta)}\longrightarrow 0\quad(\beta-\alpha\longrightarrow 0)$$

равномерно в области $0\leqslant\alpha<\beta\leqslant t$; если бы мы не были вынуждены учитывать возможность равенства $\psi_1(\alpha,\beta)=0$.

Для стационарного потока ординарность, очевидно, равносильна требованию (4) (а значит, как мы видели, и требованию (2)). Приведенная нами в § 1 хорошо известная теорема может поэтому быть сформулирована так, что для стационарного потока без последействия ординарность служит необходимой и достаточной предпосылкой формы (1). Теперь мы покажем, что эта теорема распространяется на все регулярные потоки: *необходимой и достаточной предпосылкой пуассоновской формы* (3) *регулярного потока является его* [понимаемая в смысле условия (5)] *ординарность.*

Доказательство. Пусть данный поток — регулярный и ординарный. Его производящая функция

$$\Phi(x,\ \alpha,\ \beta)=\sum_{k=0}^{\infty}v_k(\alpha,\ \beta)\,x^k,$$

как я показал в моей работе [2], может быть представлена в виде

$$\Phi(x,\ \alpha,\ \beta)=\exp\left\{\sum_{k=0}^{\infty}[\chi_k(\beta)-\chi_k(\alpha)]\,x^k\right\},$$

где функции $\chi_k(t)$ ($k=0, 1, \ldots$) всюду непрерывны и при $k>0$ — не убывающие; при этом

$$\chi_k(0)=0\ (k\geqslant 0),\ \sum_{k=0}^{\infty}\chi_k(t)=0,\ \sum_{k=1}^{\infty}k\chi_k(t)<+\infty\ (t\geqslant 0).$$

Там же мною показано, что при $k>0$

$$\chi_k(\beta)-\chi_k(\alpha)=\lim_{d\to 0}\sum_{r=1}^{s}\frac{v_k(\alpha_{r-1},\ \alpha_r)}{v_0(\alpha_{r-1},\ \alpha_r)},\tag{6}$$

где $\alpha=\alpha_0<\alpha_1<\ldots<\alpha_s=\beta$ — любое разбиение промежутка (α, β), и $d=\max\limits_{1\leqslant r\leqslant s}(\alpha_r-\alpha_{r-1})$.

В нашем случае, в силу предположенной ординарности данного потока, для любых $t>0$ и $\varepsilon>0$ найдется такое $\delta>0$, что при $0\leqslant t_1<t_2<t_1+\delta<t$ всегда

$$\psi_2(t_1,\ t_2)\leqslant\varepsilon\psi_1(t_1,\ t_2).$$

Отсюда

$$(1-\varepsilon)\psi_2(t_1,\ t_2)\leqslant\varepsilon[\psi_1(t_1,\ t_2)-\psi_2(t_1,\ t_2)]=\varepsilon v_1(t_1,\ t_2)$$

и, следовательно, при $\varepsilon<{}^1/_2$

$$\psi_2(t_1,\ t_2)\leqslant 2\varepsilon v_1(t_1,\ t_2).$$

Поэтому, если δ достаточно мало, мы имеем при $d<\delta$

$$\psi_2(\alpha_{r-1},\ \alpha_r)\leqslant 2\varepsilon v_1(\alpha_{r-1},\ \alpha_r)\quad(1\leqslant r\leqslant s),$$

и при $k>1$ из (6) следует

$$\chi_k(\beta)-\chi_k(\alpha)\leqslant\limsup_{d\to 0}\sum_{r=1}^{s}\frac{\psi_k(\alpha_{r-1},\ \alpha_r)}{v_0(\alpha_{r-1},\ \alpha_r)}\leqslant$$

$$\leqslant\limsup_{d\to 0}\sum_{r=1}^{s}\frac{\psi_2(\alpha_{r-1},\ \alpha_r)}{v_0(\alpha_{r-1},\ \alpha_r)}\leqslant 2\varepsilon\limsup_{d\to 0}\sum_{r=1}^{s}\frac{v_1(\alpha_{r-1},\ \alpha_r)}{v_0(\alpha_{r-1},\ \alpha_r)}=$$
$$=2\varepsilon[\chi_1(\beta)-\chi_1(\alpha)].$$

А так как ε может быть выбрано сколь угодно малым, то $\chi_k(\beta)-\chi_k(\alpha)=0$ ($k>1$), и мы находим

$$\Phi(x,\ \alpha,\ \beta)=\exp\{[\chi_0(\beta)-\chi_0(\alpha)]+x[\chi_1(\beta)-\chi_1(\alpha)]\},$$

или, полагая $\chi_1(t) = -\chi_0(t) = \Lambda(t)$,

$$\Phi(x, \alpha, \beta) = \exp\{[\Lambda(\beta) - \Lambda(\alpha)](x-1)\},$$

что равносильно (3).

Пусть мы теперь имеем любой регулярный поток вида (3). Полагая $\Lambda(\beta) - \Lambda(\alpha) = h$, мы находим

$$\psi_1(\alpha, \beta) = 1 - e^{-h}, \quad \psi_2(\alpha, \beta) = 1 - e^{-h} - he^{-h}; \tag{7}$$

так как функция $\Lambda(t)$ равномерно непрерывна в любом конечном интервале, то величина h при $\beta - \alpha \to 0$ бесконечно мала равномерно относительно α и β. Неравенство (5) тривиально при $h = 0$; если же $h > 0$, то в силу (7) $\psi_1(\alpha, \beta) = h + o(h)$, $\psi_2(\alpha, \beta) = o(h)$, откуда снова вытекает (5), если $\beta - \alpha$ (а следовательно, и h) достаточно мало. Таким образом, данный поток — ординарный, и наше утверждение доказано полностью.

§ 3. Общий случай

Переходим теперь к случаю, когда $\Lambda(t)$ — любая неубывающая функция $[\Lambda(t) = 0$ при $t < 0]$. В моей уже цитированной работе [2] я называл данный поток *сингулярным*, если образующие его события могут наступать лишь в заранее определенные моменты времени t_i $(i = 1, 2, \ldots)$ (в конечном или счетном числе), а числа событий, наступающих в различные моменты t_i, представляют собою взаимно независимые случайные величины. В частности, если вероятность наступления k сотытий в момент t_i имеет вид

$$e^{-\alpha_i} \frac{\alpha_i^k}{k!} \quad (\alpha_i > 0, \ i = 1, 2, \ldots), \tag{8}$$

т. е. если каждая из только что упомянутых случайных величин распределена по некоторому закону Пуассона, то мы будем называть данный сингулярный поток *пуассоновским*. Для полного описания такого потока достаточно, таким образом, задать все его «ступени» t_i и соответствующие им значения α_i пуассоновского параметра.

Если имеется только одна ступень t_i и число наступающих в момент t_i событий подчиняется пуассоновскому закону (8), то производящая функция числа событий, наступающих в промежутке (α, β), равна

$$\Gamma_i(x, \alpha, \beta) = \begin{cases} \exp\{\alpha_i(x-1)\} & (\alpha \leqslant t_i < \beta), \\ 1 & \text{в противном случае} \end{cases}$$

Но в общем случае пуассоновский сингулярный поток представляет собою, очевидно, суперпозицию конечного или счетного числа элементарных потоков только что описанного типа, независимых между собою. Таким образом, для производящей функции общего пуассоновского сингулярного потока мы находим выражение

$$\Gamma(x, \alpha, \beta) = \exp\left\{(x-1)\sum_{\alpha \leqslant t_i < \beta} \alpha_i\right\}.$$

При этом необходимо допустить, что ряд

$$\sum_{0 \leqslant t_i < t} \alpha_i = C(t)$$

сходится при любом $t \geqslant 0$; мы имеем

$$\Gamma(x, \alpha, \beta) = \exp\{[C(\beta) - C(\alpha)](x-1)\}.$$

Составим теперь суперпозицию данного пуассоновского сингулярного потока с регулярным ординарным потоком (имеющим, согласно § 2, форму (3)), ведущую функцию которого мы обозначим через $\Lambda(t)$, а производящую функцию через

$$\Phi(x, \alpha, \beta) = \exp\{[\Lambda(\beta) - \Lambda(\alpha)](x-1)\}.$$

Эта суперпозиция будет, очевидно, иметь тот же тип производящей функции, что и ее компоненты, с $C(t) + \Lambda(t)$ в качестве ведущей функции; образуемый ею поток имеет, следовательно, форму (3). Мы видим, таким образом, что *суперпозиция двух взаимно независимых потоков, один из которых — регулярный ординарный, а другой — пуассоновский сингулярный с рядом значений параметра, сходящимся в любом конечном интервале, всегда является потоком вида* (3).

Рассмотрим теперь произвольный поток без последействия вида (3), где $\Lambda(t)$ означает некоторую неубывающую функцию $[\Lambda(t) = 0$ при $t < 0]$. Пусть $t_1, t_2, \ldots$ занумерованные в любом порядке точки разрыва функции $\Lambda(t)$. Положим

$$\Lambda(t_i + 0) - \Lambda(t_i - 0) = \alpha_i \; (i = 1, 2, \ldots), \quad \sum_{0 \leqslant t_i < t} \alpha_i = C(t),$$

так что $\Lambda(t)-C(t)=\Lambda^*(t)$ непрерывная неубывающая функция. Положим далее

$$\exp\{[C(\beta)-C(\alpha)](x-1)\}=\Gamma(x, \alpha, \beta),$$
$$\exp\{[\Lambda^*(\beta)-\Lambda^*(\alpha)](x-1)\}=\Phi^*(x, \alpha, \beta),$$
$$\exp\{[\Lambda(\beta)-\Lambda(\alpha)](x-1)\}=\Phi(x, \alpha, \beta),$$

откуда

$$\Phi(x, \alpha, \beta)=\Gamma(x, \alpha, \beta)\,\Phi^*(x, \alpha, \beta).$$

Оба множителя правой части этого равенства представляют собою производящие функции потоков типа (3), ведущие функции которых соответственно равны $C(t)$ и $\Lambda^*(t)$. Так как функция $\Lambda^*(t)$ непрерывна, то второй из этих потоков — регулярный; а так как это поток типа (3), то в силу результатов § 2 он должен быть ординарным.

Рассмотрим теперь первый поток с производящей функцией $\Gamma(x, \alpha, \beta)$. Так как

$$C(\beta)-C(\alpha)=\sum_{\alpha\leqslant t_i<\beta}\alpha_i,$$

то

$$\Gamma(x, \alpha, \beta)=\prod_{\alpha\leqslant t_i<\beta}\exp\{\alpha_i(x-1)\};$$

отдельный множитель произведения; стоящего в правой части этого равенства, имеет вид

$$\exp\{\alpha_i(x-1)\}=e^{-\alpha_i}\sum_{k=0}^{\infty}\frac{\alpha_i^k}{k!}x^k,$$

и, следовательно, представляет собою производящую функцию распределения Пуассона с параметром α_i; следовательно, $\Gamma(x, \alpha, \beta)$ есть производящая функция пуассоновского сингулярного потока со ступенями t_i и параметрами $\alpha_i=\Lambda(t_i+0)-\Lambda(t_i-0)$, так что ряд

$$\sum_{0\leqslant t_i<t}\alpha_i\leqslant\Lambda(t)$$

сходится при любом $t\geqslant 0$.

Мы приходим, таким образом, к заключению, что *любой поток типа* (3) *может быть представлен в виде суперпозиции двух взаимно независимых потоков, из которых один регулярный и ординарный, а другой — пуассоновский сингулярный со сходящимися в любом конечном интервале рядом параметров.*

Сопоставляя это с результатом, полученным выше, мы приходим к следующему предложению, которое и решает поставленную в настоящей статье задачу.

Теорема. *Для того чтобы поток без последействия имел форму* (3), *необходимо и достаточно, чтобы он представлял собою суперпозицию двух взаимно независимых потоков, из которых один регулярный и ординарный, а другой — пуассоновский сингулярный, с рядом параметров, сходящимся в каждом конечном интервале.*

Разложение, о котором здесь идет речь, непосредственно определяется ведущей функцией $\Lambda(t)$ данного потока: ступенями сингулярной компоненты служат точки разрыва функции $\Lambda(t)$, в качестве параметра α_i для ступени t_i фигурирует величина $\Lambda(t_i+0)-\Lambda(t_i-0)$; наконец, ведущими функциями компонент служат $C(t)=\sum\limits_{t_i<t}\alpha_i$ (для сингулярной компоненты) и $\Lambda^*(t)=\Lambda(t)-C(t)$ (для регулярной компоненты).

ЛИТЕРАТУРА

[1] А. Я. Хинчин, Математические методы теории массового обслуживания. Труды Матем. ин-та. АН СССР, 49 (1955), стр. 1—122.

[2] А. Я. Хинчин, Теория вероятностей и ее применения, 1 вып. 1 (1956), стр. 3—17.

О ФОРМУЛАХ ЭРЛАНГА В ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ *)

1. Постановка задачи. Все изложенное в настоящей статье относится в равной мере к любым установкам массового использования, и лишь ради краткости мы будем пользоваться терминологией, принятой в приложениях теории вероятностей к телефону.

В теории массового обслуживания принято называть законом распределения $f(x)$ случайной величины ξ вероятность неравенства $\xi > x$; в этом смысле законы распределения и будут пониматься во всем дальнейшем.

Пусть мы имеем дело с телефонной установкой, имеющей n обслуживающих устройств, которые мы для краткости будем называть *линиями*. На эти линии поступают требования (*вызовы*). Если в момент такого вызова имеется свободная (незанятая) линия, то вызов *занимает* ее (одну из них, если свободных линий несколько). Период занятия линии одним вызовом называется *разговором*; по окончании

---

*) Теория вероятностей и ее применения, т. 7, вып. 3, 1962, стр. 330—335.

Рукопись настоящей статьи А. Я. Хинчина обнаружена мной в его бумагах 12.6.60 г. Рукопись была полностью подготовлена к печати: перепечатана на машинке, формулы вставлены рукой автора. Бумага от времени пожелтела. Судя по тому, что известная монография Хинчина «Математические методы теории массового обслуживания» была закончена летом 1954 г. и напечатана в первой половине 1955 г., а начиная с мая 1956 г. А. Я. Хинчин уже тяжело заболел, эта рукопись относится к периоду август 1954 г.— апрель 1956 г.

Основной результат настоящей работы был получен Б. А. Севастьяновым в работе «Эргодическая теорема для марковских процессов и ее приложение к телефонным системам с отказами» (Теория вероятностей и ее применения, т. 2, вып. 1, 1957, стр. 106—116).

Однако метод А. Я. Хинчина может оказаться полезным при дальнейших исследованиях. — *Б. Г.*

разговора линия освобождается и может быть занята новыми вызовами. Если в момент поступления вызова все n линий заняты, то вызов получает отказ (потерю) и все дальнейшее происходит так, как если бы этого вызова не было. Если в данный момент занято k из общего числа n линий ($k=0, 1, \ldots, n$), то мы для краткости будем говорить, что система находится в состоянии k.

Одним из важнейших показателей качества обслуживания для данной установки служат вероятности различных ее состояний. Под вероятностью состояния k при этом всегда понимается доля времени, в течение которого система находится в этом состоянии. При этом имеется в виду, что промежуток времени T, в течение которого ведется наблюдение, очень велик. В частности, вероятность состояния n, т. е. доля времени, в течение которого поступающие вызовы получают отказы, называется вероятностью потери (или опасным временем). Очевидно, что вероятности состояний зависят как от природы поступающего потока вызовов, так и от закона распределения длительности разговоров. Поступающий поток вызовов обычно предполагается простейшим; это значит, что для любого момента времени a вероятность отсутствия вызовов в промежутке $(a, a+t)$ равна $e^{-\lambda t}$ (где λ — постоянное положительное число) и не зависит от всего предшествующего течения потока. Что касается длительности разговоров, то, прежде всего, предполагается, что длины различных разговоров не зависят ни друг от друга, ни от того, как протекает поток вызовов.

Пусть $F(t)$ — закон распределения длительности разговоров, т. е. доля тех разговоров, длительность которых превосходит t. Эрланг выводил свои известные формулы вероятностей состояний в предположении $F(t)=e^{-\beta t}$ ($\beta>0$ — постоянная); это допущение, как известно, вообще значительно облегчает исследование вопросов теории массового обслуживания. Однако ввиду важности задачи позднее был сделан ряд попыток показать, что формулы, найденные Эрлангом, сохраняют силу при любом законе $F(t)$ (иногда этот закон подчиняется некоторым требованиям общего характера, например условию непрерывности) *). Во всех этих попытках

*) См. [1], [2.].

используется сложный аналитический аппарат и, насколько мы можем судить, так и не приводит к окончательному решению задачи. Лишь в 1953 г. появилась работа Лундквиста [3], знаменующая собой некоторый сдвиг в этом направлении; с помощью нового, простого и элементарного метода Лундквисту удалось показать, что формулы Эрланга сохраняют силу и в случае, когда все разговоры имеют одну и ту же длину τ, т. е. в случае

$$F(t)=\begin{cases} 1 & (t<\tau), \\ 0 & (t\geqslant\tau). \end{cases}$$

Настоящая работа ставит себе целью показать, что некоторое усовершенствование метода Лундквиста позволяет установить формулы Эрланга и для любого закона распределения $F(t)$. При этом все рассуждение не только не усложняется, но, напротив, становится более кратким и обозримым.

2. Обозначения. Значение символов λ и $F(t)$ определено в n°1; обозначим через s среднюю длительность разговора, так что

$$s=-\int_0^\infty t\,dF(t)=\int_0^\infty F(t)\,dt.$$

Отрезок времени, в течение которого система находится в состоянии k, мы будем обозначать через Δ_k. Число всех отрезков Δ_k за время T обозначается через N_k. Очевидно, что каждый отрезок Δ_k начинается и кончается либо вызовом, либо освобождением линии. Обозначим через $A_k(B_k)$ долю отрезков Δ_k, начинающихся вызовом (освобождением); через $a_k(b_k)$ — долю отрезка Δ_k, кончающихся вызовом (освобождением), и, наконец, через Aa_k — долю отрезков Δ_k, начинающихся и кончающихся вызовом.

Ab_k, Ba_k, Bb_k определяются по аналогии очевидным образом.

Условимся говорить, что данный отрезок Δ_k *принадлежит типу* A, если он начинается с вызова. Тогда непосредственно ясно, что означает принадлежность отрезка Δ_k типам B, a, b, Aa, Ab, Ba, Bb.

Далее, обозначим через $M_k(MA_k$, $MB_k)$ среднюю длительность отрезка Δ_k (отрезка Δ_k типа A, отрезка Δ_k типа B), так что

$$M_k=A_kMA_k+B_kMB_k. \tag{1}$$

Наконец, пусть $[k]$ $(k=0,\ 1,\ \ldots,\ n)$ означает вероятность состояния k, т. е. отношение суммарной длительности всех отрезков Δ_k за (большой) промежуток времени T к длине T этого промежутка.

3. Элементарная статистика отрезков Δ_k. Из определений n°2 с непосредственной очевидностью вытекают соотношения $(0 \leqslant k \leqslant n)$

$$A_k + B_k = a_k + b_k = 1 \tag{2}$$

$$\left.\begin{aligned} Aa_k + Ab_k = A_k, \quad Ba_k + Bb_k = B_k, \\ Aa_k + Ba_k = a_k, \quad Ab_k + Bb_k = b_k. \end{aligned}\right\} \tag{3}$$

Далее, непосредственно очевидно, что между двумя последовательными отрезками Δ_k типа Aa должен обязательно встретиться отрезок Δ_k типа Bb, и обратно; из этого следует важное соотношение

$$Aa_k = Bb_k; \tag{4}$$

отсюда и из формул (3) следует

$$a_k = Aa_k + Ba_k = Bb_k + Ba_k = B_k \tag{5}$$

и аналогично

$$b_k = A_k.$$

4. Исходная рекуррентная формула. За каждым отрезком $\Delta_{k-1} (1 \leqslant k \leqslant n)$ типа a следует отрезок Δ_k типа A; обратно, каждому отрезку Δ_k типа A предшествует некоторый отрезок Δ_{k-1} типа a. Отсюда

$$N_{k-1} a_{k-1} = N_k A_k,$$

или в силу (5)

$$N_{k-1} B_{k-1} = N_k A_k. \tag{6}$$

Так как суммарная длительность всех отрезков Δ_k равна $T[k]$, то

$$M_k = \frac{T[k]}{N_k}, \quad N_k = \frac{T[k]}{M_k} \quad (k = 0,\ 1,\ \ldots,\ n).$$

Поэтому из (6) следует

$$\frac{B_{k-1} T[k-1]}{M_{k-1}} = \frac{A_k T[k]}{M_k},$$

или

$$\frac{A_k}{M_k}[k] = \frac{B_{k-1}}{M_{k-1}}[k-1] \qquad (k = 1, 2, \ldots, n). \tag{7}$$

Эта элементарная рекуррентная формула и служит отправным пунктом при выводе формул Эрланга. Для ее применения необходимо найти величины A_k, B_k и M_k. К их постепенному отысканию мы теперь и переходим.

5. Закон распределения оставшейся части разговора. Пусть в некоторый произвольно выбранный момент времени мы застали ведущийся разговор (наряду, быть может, с другими разговорами). Обозначим через $\varphi(t)$ вероятность того, что по истечении *t сек* этот разговор еще не будет закончен. Чтобы подойти к определению функций $\varphi(t)$, разобьем отрезок времени $(0, T)$ на очень малые части одинаковой длины τ (ячейки). Отрезок какого-либо разговора, приходящийся на одну такую ячейку длины τ, условимся называть *элементом*. Если под нормальным элементом понимать такой, у которого содержащий его разговор по истечении *t сек* еще не будет закончен, то $\varphi(t)$ можно статистически определить как отношение числа нормальных элементов к числу всех элементов (точнее, как предел этого отношения при $\tau \to 0$). Число всех элементов равно суммарной длительности всех ведущихся в отрезке $(0, T)$ разговоров, разделенной на τ, т. е. равно

$$\frac{Ns}{\tau}, \tag{8}$$

где N — число всех разговоров за время T. Подсчитаем теперь, сколько среди этих элементов будет нормальных.

Число разговоров в отрезке $(0, T)$, длина которых заключена между u и $u + du$, будет

$$N[F(u) - F(u + du)] = -NdF(u).$$

При $u < t$ разговор такой длины, очевидно, нормальных элементов содержать не будет. Если же $u > t$, то нормальными элементами будут те, которые отстоят менее чем на $u - t$ от начала разговора (более чем на t от его конца). Число нормальных элементов для одного разговора длины $u > t$ поэтому равно $(u - t)/\tau$, а число таких элементов для всех разговоров длины u равно

$$-\frac{N(u-t)\,dF(u)}{\tau} \quad (u>t).$$

Суммируя по u это выражение, мы находим, что число всех нормальных элементов равно

$$-\frac{N}{\tau}\int_t^{\infty}(u-t)\,dF(u); \tag{9}$$

для функции $\varphi(t)$, являющейся отношением величин (9) и (8), мы отсюда находим выражение

$$\varphi(t)=-\frac{1}{s}\int_t^{\infty}(u-t)\,dF(u).$$

Интеграция по частям дает

$$\varphi(t)=\frac{1}{s}\int_t^{\infty}F(u)\,du. \tag{10}$$

6. Законы распределения длин отрезков Δ_k различных типов. Рассмотрим какой-либо отрезок Δ_k типа A и допустим, что $0<k<n$ (отрезков Δ_0 типа A не существует, а случай $k=n$ требует особого рассмотрения, которое будет проведено в дальнейшем). Для того чтобы этот отрезок имел длину $>t$, необходимо и достаточно выполнение следующих трех условий:

1. В течение t *сек* после начала Δ_k не поступает ни одного вызова; вероятность этого условия равна $e^{-\lambda t}$.

2. Разговор того вызова, которым начинается Δ_k, имеет длительность $>t$; вероятность этого условия равна $F(t)$.

3. Каждый из тех $k-1$ разговоров, которые уже велись в начальный момент Δ_k, продолжается еще $>t$ *сек*. Вероятность этого условия ввиду взаимной независимости длительностей различных разговоров равна $[\varphi(t)]^{k-1}$.

Так как эти три условия взаимно независимы, то для отрезка Δ_k типа A вероятность длины $>t$ (закон распределения) равна

$$\psi_A(t)=e^{-\lambda t}F(t)[\varphi(t)]^{k-1} \qquad (0<k<n).$$

Для отрезка Δ_k типа B дело обстоит несколько иначе, так как здесь для всех k ведущихся разговоров время их начала неизвестно; поэтому рассуждение, аналогичное только что нами проведенному, для длин отрезков Δ_k типа B дает, как легко видеть, закон распределения

$$\psi_B(t) = e^{-\lambda t}[\varphi(t)]^k \quad (0 \leqslant k < n)$$

(эта формула пригодна и при $k=0$).

С другой стороны, нас будет интересовать вероятность того, что отрезок Δ_k того или другого типа закончится поступлением нового вызова в возрасте между t и $t+dt$. Чтобы это произошло, необходимо выполнение следующих двух условий: 1) ни один из k разговоров, ведущихся в данном Δ_k, не должен закончиться за время t после начала Δ_k; вероятность этого условия равна, как мы уже видели, $F(t)[\varphi(t)]^{k-1}$ для Δ_k типа A $(0<k<n)$ и $[\varphi(t)]^k$ для Δ_k типа B $(0 \leqslant k < n)$; 2) первый после начала данного Δ_k вызов должен поступить через промежуток времени, заключенный между t и $t+dt$; вероятность этого условия равна $\lambda e^{-\lambda t}\,dt$. Таким образом, вероятность закончиться вызовом в возрасте между t и $t+dt$ равна

$$\lambda e^{-\lambda t}F(t)[\varphi(t)]^{k-1}dt = \lambda\psi_A(t)\,dt \quad (0<k<n) \text{ (для } \Delta_k \text{ типа } A),$$
$$\lambda e^{-\lambda t}[\varphi(t)]^k dt = \lambda\psi_B(t)\,dt \quad (0 \leqslant k<n) \quad \text{(для } \Delta_k \text{ типа } B).$$

Вероятность же закончиться вызовом независимо от возраста (т. е. принадлежать типу a) для отрезка Δ_k того или другого типа A или B получится интегрированием того или другого из только что написанных выражений по t от 0 до ∞. Но эта вероятность есть для типа A условная вероятность принадлежать типу Aa, а для типа B — условная вероятность принадлежать типу Ba. Эти же две условные вероятности в принятых нами обозначениях выражаются соответственно в виде Aa_k/A_k и Ba_k/B_k. Таким образом,

$$\left.\begin{aligned} \frac{Aa_k}{A_k} &= \lambda\int_0^\infty \psi_A(t)\,dt \qquad (0<k<n), \\ \frac{Ba_k}{B_k} &= \lambda\int_0^\infty \psi_B(t)\,dt \qquad (0 \leqslant k<n). \end{aligned}\right\} \tag{11}$$

7. Определение чисел A_k, B_k и M_k. Мы можем теперь перейти к определению A_k, B_k и M_k, входящих в рекуррентное соотношение (7). При этом мы по-прежнему предполагаем $k < n$.

Так как $\psi_A(t)$ есть закон распределения длин отрезков Δ_k типа A, то среднее значение MA_k этих длин равно

$$MA_k = -\int_0^\infty t\, d\psi_A(t) = \int_0^\infty \psi_A(t)\, dt \qquad (0 < k < n),$$

и аналогично

$$MB_k = \int_0^\infty \psi_B(t)\, dt \qquad (0 \leqslant k < n);$$

поэтому соотношения (11) дают

$$\frac{Aa_k}{A_k} = \lambda MA_k \quad (0 < k < n), \quad \frac{Ba_k}{B_k} = \lambda MB_k \quad (0 \leqslant k < n). \tag{12}$$

Из этих же формул в силу (4) получаем

$$\lambda MB_k = \frac{Ba_k}{B_k} = 1 - \frac{Bb_k}{B_k} = 1 - \frac{Aa_k}{B_k} = 1 - \lambda \frac{A_k}{B_k} MA_k$$
$$(0 \leqslant k < n). \tag{13}$$

С другой стороны, интеграция по частям и формула (10) дают

$$\lambda MB_k = \lambda \int_0^\infty \psi_B(t)\, dt = \int_0^\infty \lambda e^{-\lambda t} [\varphi(t)]^k dt =$$
$$= 1 + k \int_0^\infty e^{-\lambda t} [\varphi(t)]^{k-1} d\varphi(t) =$$
$$= 1 - \frac{k}{s} \int_0^\infty e^{-\lambda t} [\varphi(t)]^{k-1} F(t)\, dt =$$
$$= 1 - \frac{k}{s} \int_0^\infty \psi_A(t)\, dt = 1 - \frac{k}{s} MA_k \quad (0 \leqslant k < n). \tag{14}$$

Приравнивая друг другу правые части равенств (13) и (14), мы находим

$$\frac{A_k}{B_k} = \frac{k}{s\lambda} \qquad (0 \leqslant k < n);$$

это же в силу (2) дает

$$A_k = \frac{k}{k+s\lambda}, \quad B_k = \frac{s\lambda}{k+s\lambda} \quad (0 \leqslant k < n).$$

Наконец, в силу (12) и (5) формула (1) дает при $0 \leqslant k < n$

$$M_k = A_k M A_k + B_k M B_k = \frac{1}{\lambda}[A a_k + B a_k] = \frac{a_k}{\lambda} = \frac{B_k}{\lambda} = \frac{s}{k+s\lambda};$$

мы нашли, таким образом, числа A_k, B_k и M_k для всех $k < n$.

Нам остается рассмотреть случай $k = n$. Легко видеть, что в этом случае

$$A_n = 1, \quad B_n = 0, \quad \psi_A(t) = F(t)[\varphi(t)]^{n-1},$$

вследствие чего в силу (10)

$$M_n = MA_n = \int_0^\infty \psi_A(t)\,dt = \int_0^\infty [\varphi(t)]^{n-1} F(t)\,dt =$$

$$= -s\int_0^\infty [\varphi(t)]^{n-1}\,d\varphi(t) = \frac{s}{n}.$$

8. Формулы Эрланга. Подставляя в рекуррентное соотношение (7) вместо A_k, B_{k-1}, M_k и M_{k-1} найденные нами значения этих величин, мы легко получаем (случай $k = n$ не составляет исключения)

$$[k] = \frac{\lambda s}{k}[k-1] \quad (1 \leqslant k \leqslant n),$$

откуда

$$[k] = \frac{(\lambda s)^k}{k!}[0] \quad (0 \leqslant k \leqslant n).$$

Присоединяя же сюда нормирующее соотношение

$$\sum_{k=0}^{n}[k] = 1,$$

находим

$$[0] = \left\{\sum_{k=0}^{n}\frac{(\lambda s)^k}{k!}\right\}^{-1},$$

что и дает известные формулы Эрланга.

Впрочем, ничто в предыдущем не мешает нам считать число линий n бесконечно большим. В этом случае мы получаем также хорошо известные формулы Пуассона

$$[k] = e^{-\lambda s} \frac{(\lambda s)^k}{k!} \quad (k = 0, 1, 2, \ldots).$$

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] A. E. Vaulot, Rev. gén. de l'électricité, 1927, 22, № 26, стр. 1164—1171.
[2] C. Palm, Ericsson Technics, 1938, 6, № 4, стр. 39—58.
[3] K. Lundkvist, Ericsson Technics, 1953, 9, № 2, стр. 111—140.

ТЕОРИЯ СПАРЕННЫХ АППАРАТОВ*)

§ 1. Постановка задачи и предварительные замечания

В настоящей работе исследуется вопрос о числе потерь и времени ожидания в случае двух абонентов, пользующихся одним и тем же проводом. Для простоты предполагаем, что все разговоры имеют одинаковую длительность, которую мы будем обозначать через T (единица времени произвольна, мы будем условно называть ее **часом**). Это допущение может быть, конечно, оправдано лишь в самом грубом приближении, так как, по-видимому, дисперсия длительности разговоров может оказывать довольно существенное влияние на интересующие нас величины; однако развиваемый нами метод позволяет принципиально решить все поставленные вопросы и в случае любого закона распределения длительности разговоров; различие заключается лишь в том, что в случае переменной длительности расчеты становятся значительно сложнее; только по этой причине мы принимаем в настоящем исследовании (основная цель которого — выработка метода) длительность всех разговоров одинаковой. Далее, пусть каждый из двух абонентов в течение часа производит n_1 вызовов и сам получает n_2 вызовов (мы допускаем, таким образом, что эти числа для обоих абонентов одинаковы; если бы они были различны, это вызвало бы в дальнейших расчетах лишь самые незначительные усложнения). Величинами n_1, n_2 и T исчерпываются данные задачи: все остальные числа, с которыми мы будем встречаться, должны быть через них выражены.

---

*) Рукопись настоящей статьи А. Я. Хинчина была обнаружена мною в бумагах автора. Она написана в 1934 г., но не была опубликована. В настоящее время эта тематика приобретает актуальный интерес.— *Б. Г.*

Мы допускаем, что в случае, когда вызов поступает в момент занятости провода, вызывающий (абонент или посторонний) ожидает его освобождения, после чего приступает к разговору; мы допускаем далее, что в случае скопления нескольких вызовов они обслуживаются проводом в порядке очереди; это последнее допущение, по-видимому, не оказывает никакого влияния на число потерь и среднее время ожидания и потому может быть сделано без всяких оговорок.

Если в момент t один из абонентов свободен, то вероятность того, что он произведет вызов до наступления момента $t+dt$, с точностью до малых высших порядков равна $\varkappa dt$, где $\varkappa$ — постоянная, которая в дальнейшем должна быть определена из данных нашей задачи. Отсюда, как известно, следует, что если абонент в данный момент свободен, то вероятность того, что он не произведет вызова в течение конечного промежутка времени t, равна $e^{-\varkappa t}$.

Что касается вызовов, возникающих извне, то мы, как обычно, предполагаем их режим совершенно независимым от имеющегося в данный момент состояния скученности. Как известно, это имеет своим следствием так называемый закон Пуассона: вероятность появления j вызовов в течение промежутка времени t выражается формулой

$$e^{-\nu t}\frac{(\nu t)^j}{j!}.$$

Здесь ν означает (среднее) число внешних вызовов в единицу времени и, следовательно, $\nu=2n_2$.

Нам важно иметь удобное обозначение для состояния скученности в данный момент; условимся обозначать через a ожидающего абонента и через e — ожидающий внешний вызов; запись $eaee$ будет означать, что в данный момент имеется четверо ожидающих, причем первым в очереди стоит внешний вызов, вторым — абонент, а затем — еще два внешних вызова.

Нам надо также иметь обозначение для вероятности того, что в течение данного разговора, в начале которого скученность известна, произойдут те или другие определенные вызовы; так как на число этих вызовов может оказать влияние только наличие или отсутствие в начале разговора ожидающего абонента и так как более одного ожидающего абонента, очевидно, быть не может, то все разговоры мы должны разделить на два класса: 1) те, в начале которых нет ожидающего абонента, и 2) те, в начале которых имеется один ожидающий абонент; эти два класса мы будем различать соответственно верхними индексами 0 и 1; мы введем следующие обозначения:

v_0^0 — вероятность того, что в течение разговора первого класса не поступит ни одного вызова;

v_a^0 — вероятность того, что в течение разговора первого класса поступит один вызов абонента (и ни одного внешнего вызова);

v_e^0 — вероятность того, что в течение разговора первого класса поступит один вызов внешний;

v_{ae}^0 — вероятность того, что в течение разговора первого класса поступит сначала вызов абонента, а затем — внешний;

v_{ea}^0 — вероятность того, что в течение разговоров первого класса поступят те же вызовы в обратном порядке, и т. д.;

v_0^1 — вероятность того, что в течение разговора второго класса не поступит ни одного вызова;

v_e^1 — вероятность того, что в течение разговора второго класса поступит один внешний вызов;

v_{ee}^1 — вероятность того, что в течение разговора второго класса поступят два внешних вызова, и т. д.

(Очевидно, что в течение разговора второго класса абонент вызывать не может). Все эти величины легко выражаются через данные нашей задачи и через величину $\varkappa$ (которую, как уже замечено, мы впоследствии также выразим через эти данные). В самом деле, полагая для краткости

$$e^{-\alpha T} = \alpha,$$

$$e^{-\nu T} \frac{(\nu T)^j}{j!} = \tau_j,$$

мы, очевидно, будем иметь:

$$v_0^0 = \alpha\tau_0, \qquad v_e^1 = \tau_1;$$

$$v_a^0 = (1-\alpha)\tau_0, \qquad v_{ee}^1 = \tau_2,$$

$$v_e^0 = \alpha\tau_1, \qquad v_{eee}^1 = \tau_3 \text{ и т. д.}$$

$$v_0^1 = \tau_0,$$

Несколько сложнее выражаются такие вероятности, как v^0_{ae} и v^0_{ea}; однако и этот расчет не вызывает существенных затруднений. Вероятность того, что в течение данного разговора произойдет один вызов, и притом как раз в элементарный промежуток времени между t и $t+dt$ (считая от начала разговора), равна, очевидно, $\tau_1 \frac{dt}{T}$; вероятность того, что абонент, свободный в начале разговора, не произведет вызова до момента t, но произведет его в промежутке между t и T, равна

$$e^{-\varkappa t}\left[1-e^{-\varkappa(T-t)}\right],$$

так что

$$v^0_{ea}=\frac{\tau_1}{T}\int\limits_0^T e^{-\varkappa t}\left[1-e^{-\varkappa(T-t)}\right]dt=$$

$$=\frac{\tau_1(1-\alpha)}{\varkappa T}-\alpha\tau_1;$$

аналогично легко находим:

$$v^0_{ae}=\tau_1-\frac{\tau_1}{\varkappa T}(1-\alpha).$$

Подобным образом могут быть вычислены и более сложные вероятности того же типа.

Наконец, мы переходим к обозначению величин, являющихся основным предметом нашего исследования. Обозначим через π_0 относительное число разговоров, начинающихся без ожидания, или — что то же — относительное число вызовов, застающих провод свободным. Относительное число потерь (точнее — задержанных вызовов) равно тогда $1-\pi_0$; таким образом, определение величины π_0, естественно, является основной задачей нашего исследования.

Обозначим далее через π_a относительное число таких разговоров, перед началом которых ожидающим является один абонент; через π_e — относительное число таких разговоров, перед началом которых ожидающим является один внешний вызов; в обоих случаях непосредственно после начала данного разговора ожидающих не имеется. Вообще, через $\pi_{aee}\ldots$ мы будем обозначать относительное число таких разговоров, непосредственно перед началом которых скученность характеризуется схемой *aee*...; в момент начала такого разговора первый охлаждающий выбывает из очереди, и следовательно, непосредственно после начала разговора скученность определяется схемой *ee*...

§ 2. Основная система уравнений

Как уже сказано, нашей основной целью является определение величины π_0. Однако, как и во всех подобного рода задачах, эту величину удается определить лишь из системы линейных уравнений, в которую наряду с π_0 в качестве неизвестных входят все величины: $\pi_a, \pi_e, \pi_{ae}, \pi_{ea}, \pi_{aee}, \ldots$, т. е. из системы бесконечного числа линейных уравнений с бесконечным числом неизвестных. К составлению этой системы мы теперь и переходим.

Для того чтобы некоторый разговор начался без предварительного ожидания (вероятность π_0), необходимо и достаточно, чтобы 1) в начале предыдущего разговора не было ни одного ожидающего (вероятность $\pi_0+\pi_a+\pi_e$) и 2) в течение этого предыдущего разговора не поступило ни одного вызова (вероятность v_0^0); поэтому, полагая для краткости

$$\pi_0+\pi_a+\pi_e=x,$$

мы будем иметь:

$$\pi_0=v_0^0x;$$

это — первое уравнение нашей системы.

Далее, для того чтобы некоторый разговор начался в момент, когда скученность характеризуется схемой a (вероятность π_a), необходимо и достаточно выполнение одной из следующих двух предпосылок — либо I: 1) предшествующий разговор вначале протекает без ожидающих (вероятность x) и 2) в течение этого разговора происходит один вызов от абонента (вероятность v_a^0), либо II: 1) предшествующий разговор вначале протекает при одном ожидающем абоненте (вероятность π_{ea}) и 2) в течение этого разговора никаких вызовов не происходит; отсюда

$$\pi_a=v_a^0x+v_0^1\pi_{ea}$$

Это — второе уравнение нашей системы.

Совершенно аналогичными рассуждениями мы получаем и дальнейшие уравнения нашей системы; вот несколько из них:

$$\pi_e = v_e^0 x + v_0^0 \pi_{ee} + v_0^0 \pi_{ae},$$
$$\pi_{ea} = v_{ea}^0 x + v_a^0 \pi_{ee} + v_a^0 \pi_{ae} + v_0^0 \pi_{eea},$$
$$\pi_{ae} = v_{ae}^0 x + v_e^0 \pi_{ea} + v_0^1 \pi_{eae},$$
$$\pi_{ee} = v_{ee}^0 x + v_e^0 \pi_{ae} + v_e^0 \pi_{ee} + v_0^0 (o_{aee} + \pi_{eee})$$

и т. д.

Прежде всего заметим, что искомая величина π_0 во все уравнения, кроме первого, входит только в комбинации $\pi_0 + \pi_a + \pi_e = x$; поэтому целесообразно ввести x в качестве новой неизвестной вместо π_0; при этом мы сначала выбросим первое уравнение нашей системы, поставив своей основной целью определение x; когда это будет сделано, мы вернемся к этому первому уравнению, чтобы определить из него π_0.

Далее, заметим, что наша система, будучи однородной, позволит нам в лучшем случае определить взаимные отношения чисел π; чтобы найти сами эти числа, мы должны, естественно, воспользоваться нормирующим соотношением

$$\pi_0 + \pi_a + \pi_e + \pi_{ae} + \pi_{ee} + \ldots = 1.$$

Таким образом, наша система получает следующий вид:

$$x + \pi_{ae} + \pi_{ea} + \pi_{ee} + \ldots = 1,$$

$$\left.\begin{aligned} \pi_a &= v_a^0 x + v_0^1 \pi_{ea}, \\ \pi_e &= v_e^0 x + v_0^0 \pi_{ee} + v_0^0 \pi_{ae}, \\ \pi_{ea} &= v_{ea}^0 x + v_a^0 \pi_{ee} + v_a^0 \pi_{ae} + v_0^0 \pi_{eea}, \\ \pi_{ae} &= v_{ae}^0 x + v_e^0 \pi_{ea} + v_0^1 \pi_{eae}, \\ \pi_{ee} &= v_{ee}^0 x + v_e^0 \pi_{ae} + v_e^0 \pi_{ee} + v_0^0 (\pi_{aee} + \pi_{eee}). \end{aligned}\right\} \quad (1)$$

Эта система, в отличие от других, встречаемых в более простых задачах телефонии, не может быть решена методом последовательного (рекуррентного) определения неизвестных;

точное ее решение представляет поэтому значительные трудности. В дальнейшем мы увидим, что для практических целей достаточно дать приближенное решение, получаемое весьма просто.

К системе (1) надо присоединить еще одно добавочное уравнение

$$\pi_0 = v_0^0 x, \tag{2}$$

определяющее величину π_0.

§ 3. Приближенное решение основной системы

Все величины, входящие в нашу основную систему (1) (как данные, так и искомые), можно рассматривать как функции от T; на практике T обычно представляет собой небольшую дробь (примерно $^1/_{36}$). При $T \to 0$ величины v_0^0 и v_0^1 стремятся к единице, а величины v с буквенными нижними индексами — к нулю; при этом v_a и v_e являются (независимо от верхних индексов) малыми первого порядка, v_{ae}, v_{ea}, v_{ee} — второго и т. д. Аналогично этому $\pi_0 \to 1$, а все π с буквенными индексами бесконечно малы, причем порядок малости совпадает с числом нижних индексов.

Это обстоятельство, естественно, наводит нас на мысль искать приближенное решение системы (1), разлагая все данные и искомые величины по степеням T и ограничиваясь в этом разложении членами того или иного порядка. В дальнейшем мы проведем все вычисления, пренебрегая всеми, содержащими третью и более высокие степени T. Первым следствием этого соглашения является то, что система уравнений становится конечной, причем и в оставшихся уравнениях ряд членов исчезает. Первое, четвертое, пятое и шестое уравнения обращаются в систему:

$$x + \pi_{ae} + \pi_{ea} + \pi_{ee} = 1,$$
$$\pi_{ea} = v_{ea}^0 x,$$
$$\pi_{ae} = v_{ae}^0 x,$$
$$\pi_{ee} = v_{ee}^0 x,$$

которая может быть решена без затруднений; мы получаем:

$$x = \frac{1}{1 + v_{ae}^0 + v_{ea}^0 + v_{ee}^0} = \frac{1}{1 + \tau_1(1-\alpha) + \alpha\tau_2}$$

и, следовательно, с точностью до малых второго порядка

$$x = 1 - \tau_1(1-\alpha) - \tau_2\alpha;$$

уравнение (2) дает нам поэтому

$$\pi_0 = \alpha\tau_0[1 - \tau_1(1-\alpha) - \tau_2\alpha].$$

Если теперь принять во внимание, что

$$\alpha = e^{-\varkappa T} = 1 - \varkappa T + \frac{\varkappa^2 T^2}{2} + o(T^2),$$
$$\tau_0 = e^{-\nu T} = 1 - \nu T + \frac{\nu^2 T^2}{2} + o(T^2),$$
$$\tau_1 = \nu T e^{-\nu T} = \nu T - \nu^2 T^2 + o(T^2),$$
$$\tau_2 = \frac{\nu^2 T^2}{2} e^{-\nu T} = \frac{\nu^2 T^2}{2} + o(T^2),$$

то мы после элементарных вычислений получаем с точностью до малых второго порядка

$$\pi_0 = 1 - (\nu + \varkappa)\, T + \frac{1}{2}\varkappa^2 T^2. \tag{3}$$

Однако эта формула еще не решает нашей задачи, как и вообще, с помощью одних только уравнений (1) и (2) ее невозможно решить. Дело в том, что если величина ν нам прямо дана ($\nu = 2n_2$), то величина $\varkappa$, входящая в формулу (3), нам неизвестна; эта величина не принадлежит к числу априорных данных, а напротив, сама является функцией создающегося режима. Поэтому мы обращаемся к выводу второго соотношения, связывающего между собой величины π_0 и $\varkappa$.

Если в некоторый момент t провод свободен, то вероятность появления вызова со стороны одного (определенного) абонента до момента $t + dt$ есть $\varkappa dt$; вероятность же появления за этот промежуток времени внешнего вызова есть νdT; таким образом, вероятность того, что к моменту $t + dt$ провод окажется занятым (при условии, что в момент t он был свободен), составляет $(2\varkappa + \nu)\, dt$. С помощью обычного рассуждения мы отсюда находим: если в момент 0 провод был свободен, то вероятность того, что ближайшее занятие его произойдет в промежутке времени между t и $t+dt$, равна

$$e^{-(2\varkappa+\nu)t}(2\varkappa+\nu)\,dt,$$

а это позволяет найти среднюю величину периода свободы:

$$\beta=\int_0^\infty te^{-(2\varkappa+\nu)t}(2\varkappa+\nu)\,dt=\frac{1}{2\varkappa+\nu};$$

но общее число периодов свободы, очевидно, равно числу разговоров, начинающихся без ожидания, т. е. $2(n_1+n_2)\pi_0$. Поэтому суммарная длительность периодов свободы, если положить $n_1+n_2=n$, составит

$$\frac{2n\pi_0}{2\varkappa+\nu},$$

а так как суммарная длительность всех разговоров есть $2nT$, то мы, очевидно, имеем:

$$\frac{2n\pi_0}{2\varkappa+\nu}+2nT=1.$$

Это и дает нам искомое второе соотношение между π_0 и $\varkappa$; отсюда

$$\pi_0=(2\varkappa+\nu)\left(\frac{1}{2n}-T\right);$$

сопоставляя это с формулой (3), мы находим:

$$(2\varkappa+\nu)\left(\frac{1}{2n}-T\right)=1-(\nu+\varkappa)T+\frac{1}{2}\varkappa^2T^2,$$

а подстановка этого выражения в формулу (3) приводит нас к окончательному результату:

$$\pi_0=1-(n_1+2n_2)T-(n_1+2n_2)\frac{n_1}{2}T^2. \tag{4}$$

Числовой пример: $T=0{,}03$; $n_1=2$; $n_2=2{,}5$;

$$1-\pi_0=0{,}2163=21{,}63\%.$$

§ 4. Вычисление среднего времени ожидания

Найти среднее время ожидания в данной схеме гораздо труднее, чем в обычных схемах телефонии. Мы ограничимся расчетом приближенной формулы, аналогичной той, которую мы нашли для числа потерь, и для практических целей вполне достаточной.

Обозначим через $P(k)$ относительное время, в течение которого имеется k ожидающих (абонентов или внешних вызовов — безразлично). Найдем сначала приближенное выражение для $P(0)$ — времени, протекающего в отсутствие ожидающих. Очевидно, что $P(0)$ слагается из следующих компонент:

1) время, когда провод свободен $\sim 1-2nT$;

2) время, когда провод занят, но ожидающих нет; это время мы должны вычислить.

Заметим, что относительное число разговоров, в начале которых не имеется ожидающих, равно

$$x=\pi_0+\pi_a+\pi_e.$$

Если в момент 0 начался разговор этого типа, то вероятность того, что в течение малого промежутка dt последует вызов, равна $(\varkappa+\nu)\,dt$, а потому, как известно, вероятность того, что до момента t вызова не последует, равна $e^{-(\varkappa+\nu)t}$ [1]; вероятность того, что ближайший вызов последует в промежутке $(t,\ t+dt)$, равна

$$e^{-(\varkappa+\nu)t}(\varkappa+\nu)\,dt.$$

Поэтому математическое ожидание той части разговора, которая протекает в отсутствие ожидающих, равно

$$\int_0^T e^{-(\varkappa+\nu)t}(\varkappa+\nu)\,t\,dt+Te^{-(\varkappa+\nu)T},$$

что по вычислении дает

$$T-\frac{\nu+\varkappa}{2}T^2+o(T^2).$$

---

[1]) В эти формулы входят $\varkappa$, а не $2\varkappa$ (как прежде) потому, что один из абонентов занят разговором, и вызывать может только другой.

Таким образом, суммарная длительность тех частей разговоров, в течение которых не имеется ожидающих, равна с точностью до малых второго порядка

$$2nx\left\{T-\frac{\nu+\varkappa}{2}T^2\right\}=2nTx\left\{1-\frac{\nu+\varkappa}{2}T\right\};$$

принимая же во внимание сказанное выше, мы находим:

$$P(0)=1-2nT+2nTx\left\{1-\frac{\nu+\varkappa}{2}T\right\}. \tag{5}$$

Теперь мы должны были бы перейти к вычислению величины $P(1)$; это можно произвести теми же методами, хотя и несравненно сложнее. Однако нетрудно сообразить и без всяких вычислений, как должно выразиться $P(1)$ с точностью, какую мы приняли. В самом деле, обозначая через $P(>1)$ время $\sum\limits_{k=2}^{\infty} P(k)$, в течение которого число ожидающих превосходит единицу, мы непосредственно видим, что все это время принадлежит таким разговорам, при окончании которых (т. е. перед началом следующего разговора) число ожидающих не менее чем 2. Относительное число таких разговоров равно $1-x$ и есть, как мы знаем, малая второго порядка, а так как суммарная длительность таких разговоров получается умножением их числа на T, то она есть $n(1-x)T$; следовательно, она (а тем более и не превышающая ее величина $P(>1)$) есть малая третьего порядка. Но если так, то с точностью до малых второго порядка мы должны иметь:

$$P(0)+P(1)=1,$$

или в силу формулы (5)

$$P(1)=2n(1-x)T+n(\nu+\varkappa)T^2x,$$

или проще

$$P(1)+n(\nu+\varkappa)T^2x,$$

так как $2n(1-x)T$ есть малая третьего порядка. Но, с другой стороны, очевидно, что, обозначая через γ среднее время ожидания, мы должны иметь:

$$2n\gamma=P(1),$$

откуда

$$\gamma = \frac{\nu + \varkappa}{2} x T^2;$$

в пределах требуемой точности мы вправе положить $x = 1$ и $\varkappa = n_1$, что дает

$$\gamma = \left(\frac{n_1}{2} + n_2\right) T^2.$$

В случае числового примера § 3 мы находим:

$$\gamma = 0{,}00315 \text{ часа} = 11{,}34 \text{ секунды}.$$

О НЕКОТОРЫХ ПОСТАНОВКАХ ЗАДАЧ И РЕЗУЛЬТАТАХ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

Б. В. Гнеденко

Со времени выхода в свет первого издания монографии А. Я. Хинчина по теории массового обслуживания интерес к этому направлению исследований неизмеримо возрос как в Советском Союзе, так и за его пределами. За этот срок появились многочисленные оригинильные исследования, посвященные решению как чисто прикладных, так и теоретических вопросов. Методы исследования стали более разнообразны и гибки. Одновременно круг применений теории стал значительно более широким. Если в двадцатые годы отдельные вопросы теории массового обслуживания возникали лишь в практике организации телефонных сетей, то теперь к вопросам телефонного дела добавились разнообразнейшие и интересные задачи ядерной физики, автоматизации производства, эксплуатации аэропортов, морских причалов, автомобильных хозяйств, железнодорожных станций, больниц, ремонтных пунктов, предприятий массового обслуживания населения (магазинов, пунктов проката, билетных касс и пр.). В связи с таким увеличением прикладной роли теории массового обслуживания за последние несколько лет появились хорошие обзоры, приспособленные преимущественно к нуждам практиков. В первую очередь при этом, пожалуй, следует назвать обзорную статью Т. Саати [1], позднее перепечатанную в несколько измененном виде в его книге [2], посвященной изложению содержания новой научной дисциплины, известной под названием исследования операций. Более математический характер носят статья Л. Такача [1] и обзор, составивший главу 9 недавно появившейся книги Баруча Райда [1].

Из довольно многочисленных теперь монографий по теории массового обслуживания хотелось бы упомянуть книги Авондо

Бодино и Бромбилла [1], Кокса и В. Смита [1], Л. Такача [2], Р. Сиски [1], Т. Саати [3], Дж. Риордэна [1]. Каждая из названных книг представляет известный интерес, однако в первую очередь хотелось бы выделить среди них монографии Р. Сиски и Т. Саати. Первая из них представляет собой подробную сводку результатов по теории массового обслуживания, связанных с задачами телефонного дела. Вторая книга ставит перед собой более широкие цели — изложение основных результатов, полученных относительно важнейших в теоретическом и прикладном отношениях задач теории массового обслуживания.

В существующей монографической литературе по теории массового обслуживания книга А. Я. Хинчина занимает значительное место, поскольку в ней впервые было систематически изучено строение входящего потока требований, а также распределение времени ожидания для системы с очередью при обслуживании ее одним прибором. Заметим, что именно в этой монографии А. Я. Хинчин изложил свои достаточные условия близости суммарного потока, слагаемые которого независимы и равномерно малы, к простейшему потоку. По сути дела, как это выяснил ученик А. Я. Хинчина Г. А. Ососков, условия Хинчина являются и необходимыми.

Ниже мы изложим некоторые вопросы, решению которых последнее время уделялось значительное внимание.

С точки зрения математика теория массового обслуживания может быть включена в общую теорию случайных процессов и полей. При этом интересующие ее процессы будут весьма частного вида. Однако при таком подходе многие практически важные задачи могут приобретать весьма причудливый характер. В результате целесообразно выделить новый раздел теории случайных процессов, разрешив в нем собственную терминологию, способствующую естественной формулировке специфических проблем. Получается ситуация, близко напоминающая собой то, что произошло с самой теорией вероятностей: формально ее можно превратить в раздел теории функций действительного переменного, однако при этом будет потеряна специфика теории вероятностей, столь ценная как для самой математики, так и для многообразных применений. В теории массового обслуживания теория случайных процессов становится орудием исследования, не подминая при этом под себя специфику возникающих вопросов.

Первичным понятием теории массового обслуживания, изучению которого большое внимание уделил А. Я. Хинчин, является понятие потока требований, поступающих на систему обслуживания. Кратко мы будем называть его входящим потоком. В том виде, в каком понятие потока до сих пор встречалось в литературе, оно укладывается в общее понятие случайного процесса. Это обстоятельство четко отмечено в § 6 книги Хинчина, где сказано, что поток требований является не чем иным, как дискретным случайным процессом, принимающим только целочисленные неотрицательные значения. Нужно, однако, отметить, что такой подход становится теперь уже недостаточным, так как в ряде важных задач необходимо изучать поток требований не только во времени, но и в пространстве. Отсюда возникает мысль о следующем определении.

Пусть задано множество элементарных событий X, а также σ-алгебра его измеримых подмножеств $\mathfrak{A}_X$. Случайным потоком $\eta(A)$ с фазовым пространством $(X, \mathfrak{A}_X)$ называется система случайных величин $\eta(A)$, определенных на элементах $A \subset \mathfrak{A}_X$ и обладающих следующими двумя свойствами:

1) $\eta(A)$ — абсолютно аддитивная функция множества A;

2) $\eta(A)$ принимает лишь неотрицательные целочисленные значения.

Для примера, если нас интересует изучение потока землетрясений, то указания только момента t наступления землетрясения недостаточно, необходимо дополнительно указать и координаты его эпицентра φ, θ, h — широту, долготу и глубину. Таким образом, в качестве множества X следует принять множество точек (t, φ, θ, h), где $0 \leqslant \varphi \leqslant 2\pi$, $0 \leqslant \theta \leqslant \pi$, $0 < h < 6000$ *км*, $0 < t < \infty$.

Важным частным примером случайного потока является пуассоновский поток. В соответствии с предложенным общим подходом, пуассоновским потокам следует дать такое определение. Пусть $\eta(A)$ — абсолютно аддитивная, неотрицательная функция множества A, определенная на $A \in \mathfrak{A}_X$. Для любой системы $A_1, A_2, \ldots, A_n$ попарно непересекающихся множеств величины $\eta(A_1), \eta(A_2), \ldots, \eta(A_n)$ взаимно независимы и для каждого допустимого множества имеют место равенства

$$P\{\eta(A) = k\} = \frac{[\Lambda(A)]^k}{k!} e^{-\Lambda(A)}.$$

Не изменяя известных рассуждений монографии Хинчина, легко доказать, что поток $\eta(A)$ тогда и только тогда определяется равенствами

$$P\{\eta(A)=k\}=\frac{(\lambda|A|)^k}{k!}e^{-\lambda|A|}.$$

(простейший поток), где $\lambda>0$ — постоянная, $|A|$ — мера множества A, когда выполнены условия стационарности, ординарности и отсутствия последействия.

Более того, многие теоремы, доказанные Хинчиным относительно строения потоков и помещенные в настоящей книге, легко переносятся и на приведенное здесь определение потока.

Предельная теорема, вскрывшая одну из существенных причин, в силу которой простейший поток может служить в широких границах хорошим приближением к истинному потоку требований, была предметом исследований ряда ученых. Прежде всего, как я упомянул в примечании к основному тексту (см. стр. 60), Г. А. Ососков обнаружил, что условие Хинчина является не только достаточным, но по сути дела и необходимым.

Позднее И. Б. Погожев заинтересовался вопросом быстроты сближения с пуассоновским распределением в зависимости от числа слагаемых потоков. Для случая одинаково распределенных независимых стационарных потоков он нашел первый член асимптотического разложения по степеням n^{-1} (работа не опубликована, докладывалась на семинаре по теории массового обслуживания, МГУ, 1961 г.). Более детальное исследование этого вопроса в тех же условиях под влиянием доклада И. Б. Погожева было проведено П. Франкеном. Б. И. Григелионис применил к доказательству предельной теоремы Хинчина-Ососкова методы классической теории суммирования. Случай суммирования нестационарных потоков изучен Б. И. Григелионисом. Им найдены необходимые и достаточные условия сближения суммарного потока с нестационарным пуассоновским; эти условия особенно просты в том случае, когда суммируются потоки типа процессов восстановления. Им же рассмотрен во всей полноте вопрос асимптотического разложения.

В системах с многоступенчатым обслуживанием, когда требование, после того как оно обслужено прибором первой группы, поступает на прибор второй группы и т. д., интерес представляет не только поток входящих требований, но и те новые потоки, которые образуются требованиями, уже обслуженными приборами первой группы (или же получившими на них отказ), а также второй и последующими группами приборов. С такой ситуацией приходится иметь дело во многих реальных проблемах, в частности при расчете автоматических линий. К сожалению, это направление исследований находится совсем в зачаточном состоянии. Кроме результата Пальма, сформулированного и доказанного в монографии А. Я. Хинчина «Математические методы теории массового обслуживания» (стр. 107, § 28), можно отметить несколько статей Н. В. Яровицкого [1], [2]. Н. В. Яровицкий подверг систематическому изучению поток выходящих после обслуживания требований и ввел в рассмотрение односвязно зависимые потоки. Мы направляем читателя за определением потоков этого типа и их свойствами к указанным работам автора.

Во многих случаях первоначальный поток, проходя через ряд последовательных установок, постепенно теряет часть своих требований. Так обстоит дело на автоматической линии, когда после каждой операции происходит отбраковка обрабатываемых изделий. Подобная же картина наблюдается при исправлении опечаток несколькими корректорами. Поток опечаток при этом редеет. Перед исследователями неоднократно возникал вопрос относительно свойств таких редеющих потоков. Одно из предположений состояло в том, что при весьма общих условиях, касающихся процесса разряжения, поток требований должен подходить все ближе и ближе к семейству пуассоновских. Мы сформулируем сейчас теорему, полученную в этом направлении А. Реньи [1].

Пусть поток A подвергнут следующей операции: каждое требование потока остается с вероятностью p и выбрасывается из потока с вероятностью $q=1-p$; одновременно изменяется масштаб, причем за единицу масштаба времени принимается величина $1/q$. Обозначим эту операцию через T_p.

Если к потоку A с ограниченным последействием и конечной интенсивностью λ применить последовательность операций разряжения T_{p_1}, T_{p_2}, ..., T_{p_n} и при $n\to\infty$

$$p_1 p_2 \dots p_n \to 0,$$

то поток $T_{p_1}T_{p_2}\ldots T_{p_n}$ стремится к простейшему потоку интенсивности λ.

Обобщение этого результата было дано Ю. К. Беляевым.

Им же [1] был указан один тип потоков, представляющий интерес для теории массового обслуживания.

В последнее время значительное число работ было посвящено вопросам обслуживания неординарного потока, поскольку в реальных задачах приходится считаться с тем, что требования могут появляться не поодиночке, а группами. Изучением неординарных потоков без последействия систематически занимается П. И. Васильев.

Системы с потерями. Классическая проблема обслуживания с потерями за последние годы получила значительное развитие и в некоторых аспектах может считаться уже исчерпывающе разрешенной. Так, в примечаниях к тексту второй части монографии Хинчина было отмечено (стр. 84), что формулы Эрланга сохраняют свой вид в случае, когда обслуживается стационарный пуассоновский поток n приборами одинаковой производительности, а функция распределения длительности обслуживания произвольна с конечным математическим ожиданием. Метод доказательства, примененный Б. А. Севастьяновым [1], оказался особенно удачным. Выяснилось, как мы отчасти увидим позднее, что его рассуждения применимы ко многим задачам более сложной природы.

Другое направление обобщений формул Эрланга было развито Пальмом [1] и Такачем [2], [3]. Они рассмотрели случай, когда на обслуживающую систему поступает произвольный поток типа Пальма, а длительность обслуживания каждого требования (на каждом приборе) имеет распределение $H(x)=1-e^{-\mu x}$.

В работе [3] Л. Такач обобщил эту постановку задачи на следующую схему: требование остается в системе, если имеется хотя бы один свободный обслуживающий прибор или если число ожидающих требований не превосходит заданного числа m; в противном случае требование покидает систему обслуживания.

А. Шахбазов рассмотрел случай обслуживания пуассоновского потока требований (с потерями) приборами разной производительности.

В последних работах В. Бенеша [1] имеются попытки изучить вероятность потери при обслуживании произвольного нестационарного потока.

Еще в 1918 г., вскоре после опубликования работы К. Эрланга, Энгсет обобщил результат Эрланга на случай обслуживания конечного множества потребителей, каждый из которых может предъявить требование в случайный момент времени. К этой задаче возвращались неоднократно, в частности при изучении простоев механизмов, обслуживаемых одним рабочим или бригадой. Мы укажем на большое исследование Ж. Коэна [1]. В теоретическом и прикладном отношениях исключительно интересно при этом учесть то обстоятельство, что требование, получившее отказ, вновь и вновь предъявляет свои претензии к системе обслуживания.

Полезно указать еще на два направления обобщений постановок вопросов об обслуживании требований с потерями. Первое из них состоит в том, что на систему приборов поступают несколько потоков, причем потоки с большими номерами обладают преимуществами по сравнению с предшествующими. Это преимущество состоит в том, что требование с большим номером, поступившее в систему и заставшее все приборы занятыми, вытесняет требование с минимальным номером (это последнее может теряться или становиться в очередь) и может быть потеряно лишь в том случае, когда все приборы заняты обслуживанием требований либо того же номера, либо более высоких номеров.

В частности, именно с такой постановкой задачи приходится иметь дело, когда сами обслуживающие приборы могут выходить из рабочего состояния. Тогда следует учитывать сразу два потока: поток требований и поток поломок, причем поток поломок обладает преимуществом. Недавно в работе Т. П. Марьяновича [1] методом Севастьянова была изучена следующая задача: простейший поток интенсивности λ поступает на систему, состоящую из n одинаковых приборов. Каждый прибор независимо от других может выйти из рабочего состояния в любой момент времени в период, когда он занят обслуживанием. Период непрерывной работы прибора — случайная величина с функцией распределения $K(x)$. Время восстановления прибора —

случайная величина с функцией распределения $G(x)$. Длительность обслуживания требования — случайная величина с функцией распределения $H(x)$. Предполагается, что $H(x)$ имеет конечный первый момент, а также хотя бы одно из распределений $K(x)$ или $G(x)$ имеет конечный первый момент. В этих предположениях доказано существование пределов $p_{ij}=\lim\limits_{t\to\infty} P_{ij}(t)$, где $P_{ij}(t)$ — вероятность того, что в момент t обслуживанием заняты i приборов, а j приборов находятся в ремонте ($i, j=0, 1, 2, \ldots, n$; $i+j\leqslant n$). Требование теряется, если оно поступило в момент, когда все приборы заняты (или находятся в ремонте), а также тогда, когда обслуживающий прибор вышел из рабочего состояния. В последнем случае требование теряется даже тогда, когда в системе имеются свободные приборы.

Только что описанная задача была обобщена Т. П. Марьяновичем на тот случай, когда в системе имеются, помимо приборов работающих, также приборы, находящиеся в резерве разного порядка [2]. Понятно, что эта задача имеет самое непосредственное отношение к вопросам теории надежности.

Второе направление обобщений касается выбора оптимальных режимов обслуживания.

Имеется n приборов разной производительности: среднее время обслуживания требования для них соответственно равно $a_1, a_2, \ldots, a_n$. Приборы расположены в определенном порядке, и требования потока поступают на свободный прибор с наименьшим номером в только что определенном порядке. Требование, заставшее все приборы в работе, теряется. Спрашивается, как надо перенумеровать приборы, чтобы вероятность потери была минимальной? Ответ таков: приборы следует нумеровать так, чтобы с ростом номера уменьшалась их производительность.

Следующая задача в том же круге идей была рассмотрена Г. П. Климовым. Имеется группа однотипных приборов и заданного объема накопитель, в котором требования могут ожидать освобождения приборов. Требования, попавшие в систему, когда накопитель полон, теряются. На систему обслуживания поступает пуассоновский поток с параметром $\lambda(t)$. Спрашивается, как выбрать $\lambda(t)$, чтобы итог работы системы за время $(0, T)$ был максимален. Итог оценивается функционалом $I(\lambda(t), \mu)$. Длительность обслуживания случайна и имеет показательное распределение: $1-e^{-\mu t}$.

Таким итогом работы могут быть общая производительность системы, количество потерь и пр.

Оказывается, что экстремум достигается на классе ступенчатых функций $\lambda(t)$. В частности, для получения максимальной производительности оборудования следует выбрать сначала $\lambda(t)$ возможно большим, а затем сделать минимально допустимым. Этим достигается быстрое заполнение накопителя. Наличие требований в накопителе обеспечивает бесперебойную работу механизмов системы.

Системы с ожиданием. Пожалуй, максимальное число работ, относящихся к теории массового обслуживания, посвящено изучению образования очередей в системах с ожиданием. Распределение длительности ожидания, условия образования увеличивающейся очереди, распределение длительности непрерывной занятости прибора — вот примерный перечень вопросов, которым было уделено много внимания. Особенно интересовал исследователей случай одного обслуживающего прибора. Мы знаем, что для пуассоновского потока и произвольного распределения длительности обслуживания общие результаты были получены А. Я. Хинчиным. Изящные результаты были получены здесь также Линдли, Л. Такачем, А. Шахбазовым [1] и др. Линдли получил интегральное уравнение для распределения длительности ожидания начала обслуживания в случае одного прибора для произвольного пальмовского потока и произвольного распределения длительности обслуживания. Это интегральное уравнение дало ему возможность заключить, когда очередь будет расти неограничено и когда распределение величины очереди стремится к стационарному состоянию.

Обслуживание одним прибором неординарного потока было предметом исследований Гэйвера [1], Миллера [1] и Шахбазова [1]. А. Шахбазов ввел усложнение, представляющее интерес для теории надежности, а именно: он предположил дополнительно, что прибор, начиная обслуживание требования без предварительного ожидания его освобождения, требует времени на «разогрев».

Киффер и Вольфовиц [1], [2] исследовали обслуживание произвольного пальмовского потока при условии произвольного распределения длительности обслуживания n одинаковыми приборами. Ими составлены интегральные уравнения задачи, получено условие существования стационарного решения.

Известно, что если интенсивность поступления требований в систему равна средней производительности системы, то длина очереди со временем возрастает неограничено (за исключением тривиального случая регулярного поступления требований, равного длительности обслуживания). Возникает вопрос изучения асимптотического поведения различных характеристик обслуживания. Этот вопрос интересовал многих исследователей. Я приведу здесь результаты И. Н. Коваленко.

Пусть λ — интенсивность входящего потока, τ — средняя длительность обслуживания одного требования, число обслуживающих приборов равно единице и $\varrho = \lambda\tau = 1 - \varepsilon$, где $\varepsilon > 0$. Если $\varepsilon \to 0$, то средняя длительность времени ожидания начала обслуживания будет расти, имея порядок $1/\varepsilon$. Функция распределения длительности ожидания γ подчинена такому предельному распределению

$$P\{\varepsilon\gamma > x\} \to e^{-\frac{2\tau x}{\sigma^2+\tau^2}} \qquad (\varepsilon \to 0),$$

где σ^2 — дисперсия длительности обслуживания.

Точно так же, если дисперсия времени обслуживания конечна, то для распределения числа ν требований в системе имеет место предельная теорема: при $\varepsilon \to 0$

$$P\{\varepsilon\nu > x\} \to e^{-\frac{2\tau^2 x}{\sigma^2+\tau^2}}.$$

Большое внимание уделялось последние годы изучению обслуживания с очередью при наличии преимущественных требований. Система обслуживания может быть описана таким образом: имеется k классов требований, каждый из которых образует простейший поток; потоки между собой независимы. Внутри каждого класса обслуживание производится в порядке строгой очередности. Различаются две схемы: обслуживание без прерывания (английский термин — head of the line) и обслуживание с прерыванием (preemtive priority). Предположим, что в некоторый момент времени происходит обслуживание требований i-го класса и поступает требование класса j $(j > i)$; для первой схемы вновь поступившее требование ожидает окончания обслуживания низшего класса, но зато немедленно занимает первую освободившуюся линию, если только перед ним не было требований того же или более высокого класса. Вторая схема в свою очередь подразделяется на две категории: в первой из них учитывается время, уже потраченное на его обслуживание (preemptive resume), во второй (preemptive repeat) это время теряется и обслуживание начинается сначала. Отметим здесь лишь работу Миллера [2], в которой указаны предшествующие исследования, а также для случая $k=2$ даны такие характеристики: производящая функция стационарных вероятностей числа преимущественных и непреимущественных требований, находящихся в пункте обслуживания, преобразования Лапласа для функций распределения времени обслуживания и периода занятости, а также для числа требований, обслуженных за один период занятости.

Как частный случай указанной схемы, можно рассмотреть такую задачу, имеющую очевидный практический интерес: изучить систему обслуживания с очередью при дополнительном предложении, что обслуживающие приборы могут выходить из рабочего состояния и их восстановление требует, вообще говоря, случайного времени. Здесь мы вновь имеем два потока: поток требований и поток поломок, причем поток поломок пользуется преимуществом. Иногда полезно учитывать также необходимость профилактических мероприятий.

В ряде телефонных вопросов приходится встречаться с такими постановками задач, когда один поток обслуживается по схеме ожидания, а другой по схеме потерь.

Для ряда практически важных задач интересно рассмотрение задачи обслуживания с очередью при дополнительном вмешательстве регулирующей силы. Если иметь в виду интересы работников морских портов, то задача может быть поставлена так: имеется группа близких портов, в которых имеются соответственно n_1, n_2, ..., n_k причалов. В том случае, когда какой-либо порт перегружен и имеется опасность возникновения очереди, диспетчер отдает распоряжение о подходе очередного судна в тот порт, в котором имеются свободные причалы или же причалы должны освободиться в ближайшее время. Само собой разумеется, что в экономическом отношении эта задача представляет значительный интерес. Кстати следует сказать, что представляет интерес изучение различных характеристик экономической эффективности использования портов. Так, некоторые экономисты ставят себе в особую заслугу то, что удается добиться рекордного времени использования причалов. При этом никогда не отмечается, как долго ждут погрузки или разгрузки суда, как велики их непроизводительные стоянки в ожидании освобождения причала.

Обслуживание с ограничениями. Как ни важны в прикладном отношении и как ни интересны в теоретическом отношении задачи, возникающие при изучении обслуживания с очередью и с потерями, в настоящее время, по-видимому, большее значение приобретает объединяющая обе эти задачи схема.

Мы называем эту схему обслуживанием с ограничениями. Она включает в себя большое число разнообразных и практически интересных вопросов.

Все возникающие здесь задачи так или иначе связаны с тем, что требование, попавшее в систему обслуживания и заставшее все приборы занятыми, не обязательно теряется, а оставшись в очереди, не обязательно доводит ожидание до начала обслуживания.

Как правило, требование ограничено в своих возможностях ожидания и даже пребывания в системе (ожидание плюс обслуживание). Так, для примера, человек, пришедший на междугороднюю телефонную станцию, готов уделить некоторое время на ожидание начала разговора, однако он не может ожидать сколь угодно долго, так как либо занят другим делом, либо же слишком длительное ожидание приведет к тому, что надобность в разговоре отпадет. Точно так же самолет, вставший в очередь на посадку, ограничен в своих возможностях ожидания многими обстоятельствами: количеством горючего в баках, состоянием экипажа и пр.

Последние годы многие авторы систематически подвергали изучению такую схему, которая учитывает возможность очереди ограниченного размера (всякое требование, заставшее в системе m ожидающих, теряется). Далее, рассматривался случай, когда требование, заставшее все приборы занятыми, остается в системе лишь с вероятностью, зависящей от длины очереди. Исходя из разных прикладных задач, почти одновременно ряд авторов пришли к постановке близких задач теории обслуживания в таких условиях:

1) требование, поступившее в систему обслуживания, ожидает начала обслуживания в течение времени τ (вообще говоря, случайное);

2) требование остается в системе обслуживания не более чем время τ (даже в том случае, если его начали обслуживать, но длительность обслуживания превышает τ).

Понятно, что обе эти постановки задач включают в себя как крайние случаи схемы обслуживания с потерями и с очередью.

Случай $n=1$ был детально исследован И. Н. Коваленко [1]. В этой работе он рассмотрел несколько более широкую схему, чем мы только что наметили. В ней учтено то обстоятельство, что длительность пребывания зависит от того, сколько времени требование ожидало начала обслуживания.

Теория обслуживания с ограничениями находится еще в начальной стадии, и несомненно, ближайшее будущее принесет многочисленные полезные результаты, которые не только дадут возможность решить те или иные частные прикладные задачи, но и создать общую теорию, способную охватить многочисленные постановки вопросов единым подходом, единой точкой зрения.

Обратим внимание на то, что подавляющее число задач теории массового обслуживания приводит к необходимости рассмотрения процессов, отличных от марковских. Нередко удается тем или иным способом (например, путем введения дополнительной переменной) превратить эти процессы в марковские и тем самым получить возможность использовать хорошо разработанный аналитический аппарат теории марковских процессов. На этом пути получено большое число ценных результатов.

Отметим, что в этой небольшой обзорной статье не было возможности даже только упомянуть все наиболее интересные работы. Довольно полная библиография (до 1957 г.) составлена Дойгом [1]. Она может служить хорошим дополнением к библиографии, имеющейся в уже упомянутых книгах Сиски и Саати.

ЛИТЕРАТУРА

A. T. Bharucha—Reid [1], Elements of Markov processes and the app.ications, McGraw Hill, 1960.

Беляев Ю. К. [1], Линейчатые марковские процессы и их приложение к задачам теории надежности.

V. E. Benes [1], General Stochastic Processes in Traffic Systems with One Server, Bell System. Techn. J., 1960, 127—159.

G. A. Bodino, F. Brombilla [1], Teoria delle code, Milano — Vares, 1959, 1—219.

D. P. Gaver [1]. Imbedded Markov chain analysis of a Waiting line Processes in continuous time, Ann. Math. Statistics, v. 30, 1959, 698—720.

A. Doig [1], A bib'iography on the theory of queues, Biometrica 44, № 3—4, 1957. 490—514.

Kiefer and Wolfowitz [1], On the theory of queues with many servers, Trans. Amer. Math. Soc. 78, 1955, 1—18.

[2] On the characteristics of the general queuing process with applications to random walk, Ann. Math. Statistics 27, № 1, 1956 г. 147—161.

Коваленко И. Н. [1], Некоторые задачи массового обслуживания с ограничением, Теория вероят. и ее применения, т. 6, № 1, 1961, 222—228.

Cox and Walter L. Smith [1], Queues, Methuen & Co, London, 1961.

Cohen J. W. [1], The generalized Engset Formulae, Philips Telecommunication Rev., 18, № 4, 1957, 158—170.

Марьянович Т. П. [1], Обобщение формул Эрланга на случай, когда приборы могут выходить из строя и восстанавливаться, Укр. матем. ж., т. 12, № 3, 1960, 279—286.

[2]. Надійність системи зі змішаним резервом, Доповіді УРСР, № 8, 1961, 994—997.

Rupert G. Miller, [1] A contribution to the theory of Bulk Queues, J. Roy. Statist. Soc. B, v. 21, № 2, 1959, 320—337.

[2] Priority Queues, Ann. Math. Statistics, 31, № 1, 1960, 86—106.

Palm C., [1], Intensitätsshwankungen im Fernsprechverkehr, Ericsson Techn., 44, 1943, 1—189.

Rény A., [1], A Poisson—folyamat egy jellemzése, Тр. Матем. ин-та Венгрии, v. 1, fasc. 4. 1956, 519—527.

J. Riordan, [1], Stochastic Service Systems, Wiley and Sons, 1962, 1—139.

Saaty T. L., [1] Résumé of useful formulas in Queuing Theory, Operat. Res., v. 5, 1957, 162—187.

[2] Mathematical methods of operations research, McGraw-Hill, 1959, 331—374.

[3] Elements of Queuing theory with applications, McGraw-Hill, 1961, 1—423.

R. Syski, [1], Introduction to congestion theory in telephone systems, Oliver and Boyd, Ediburgh and London, 1960, 1—742.

Севастьянов Б. А. [1], Эргодическая теорема для марковских процессов и ее приложение к телефонным системам с отказами; Теория вероят. и ее применен., т. 2, вып. 1, 1957, 106—116.
Takács L., [1] Некоторые вероятностные задачи в телефонии, Математика (сб. переводов), 4:6, 1960, 93—144.
[2] Introduction to the theory of queues, Oxford University Press, 1962, 1—268.
[3] On the generalization of Erlang's formula, Acta Math. Acad. scient. Hung., 7, 1956, 419—433.
[4] On a combined waiting time and loss problem concerning telephone traffic, Ann. Univ. Sc., t. 1, Budapest, 1958, 73—82.
Шахбазов А., [1] Обслуживание неординарного потока; Докл. АН СССР, т. 145, № 2.
Яровицкий Н. В., [1] Протвихідний потік однолінійної системи обслуговвання з втратами, Доповіді АН УРСР, № 10, 1961, 1251—1254.
[2] О некоторых свойствах односвязно-зависимых потоков, Укр. матем. ж., т. 14, № 2, 1962, 170—179.

Александр Яковлевич Хинчин

Работы по математической теории массового обслуживания

М., Физматгиз, 1963 г., 236 стр.

Редактор *Морозова И. Е.*
Техн редактор *Михлин Э. И.*
Корректор *Володяева И. А.*

Сдано в набор 30/VI 1962 г. Подписано к печати 16/I 1963 г. Бумага $84 \times 108^{1}/_{32}$. Физ. печ. л. 7,375. Условн. печ. л. 12,92. Уч.-изд. л. 11,34. Тираж 15 000 экз. Т-10925. Цена книги 72 коп. Заказ № 3121.

Государственное издательство физико-математической литературы, Москва, В-71, Ленинский проспект, 15.

Первая Образцовая типография имени А. А. Жданова Московского городского совнархоза. Москва, Ж-54, Валовая, 28.

Отпечатано с матриц в Гос. тип. «Вайздас», Вильнюс, Страздялио 1. Заказ № 238.

следовательно, их сумму можно переписать в виде $\sum_{j,k}(x_j+y_k)\,p(x_j,y_k)$, что по определению является математическим ожиданием $\mathbf{X}+\mathbf{Y}$. Тем самым доказательство закончено. ▶

Ясно, что аналогичная общая теорема для произведений неверна. Например, $\mathbf{E}(\mathbf{X}^2)$, вообще говоря, отличается от $(\mathbf{E}(\mathbf{X}))^2$. Так, если $\mathbf{X}$ — число очков, выпавшее при бросании правильной игральной кости, то

$$\mathbf{E}(\mathbf{X})=7/2, \quad \text{но } \mathbf{E}(\mathbf{X}^2)=(1+4+9+16+25+36)/6=91/6.$$

Однако для независимых случайных величин справедливо простое правило умножения.

Теорема 3. *Если $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины с конечными математическими ожиданиями, то их произведение является случайной величиной с конечным математическим ожиданием и*

$$\mathbf{E}(\mathbf{X}\mathbf{Y})=\mathbf{E}(\mathbf{X})\,\mathbf{E}(\mathbf{Y}). \tag{2.6}$$

Доказательство. Для нахождения $\mathbf{E}(\mathbf{X}\mathbf{Y})$ нужно умножить каждое возможное значение x_jy_k на соответствующую вероятность. Следовательно,

$$\mathbf{E}(\mathbf{X}\mathbf{Y})=\sum_{j,k}x_jy_kf(x_j)\,g(y_k)=\left\{\sum_j x_jf(x_j)\right\}\left\{\sum_k y_kg(y_k)\right\}, \tag{2.7}$$

причем перестановка членов возможна в силу абсолютной сходимости рядов. ▶

По индукции это правило умножения доказывается для произвольного числа взаимно независимых случайных величин.

Для математического ожидания условного распределения вероятностей удобно иметь свое обозначение. Если $\mathbf{X}$ и $\mathbf{Y}$ — две случайные величины с совместным распределением (1.3), то *условное математическое ожидание* $\mathbf{E}(\mathbf{Y}|\mathbf{X})$ *величины* $\mathbf{Y}$ *при заданном* $\mathbf{X}$ *есть функция, принимающая в точке* x_j *значение*

$$\sum_k y_k\mathbf{P}\{\mathbf{Y}=y_k\,|\,\mathbf{X}=x_j\}=\left(\sum_k y_kp(x_j,\,y_k)\right)\Big/f(x_j). \tag{2.8}$$

Это определение имеет смысл только в том случае, когда ряд в правой части сходится абсолютно и $f(x_j)>0$ при всех j.

Условное математическое ожидание $\mathbf{E}(\mathbf{Y}|\mathbf{X})$ является новой случайной величиной. Для вычисления ее математического ожидания нужно умножить (2.8) на $f(x_j)$ и просуммировать по x_j. В результате получим

$$\mathbf{E}(\mathbf{E}(\mathbf{Y}\,|\,\mathbf{X}))=\mathbf{E}(\mathbf{Y}). \tag{2.9}$$

§ 3. ПРИМЕРЫ И ПРИЛОЖЕНИЯ

а) *Биномиальное распределение.* Пусть $\mathbf{S}_n$ — число успехов в n испытаниях Бернулли с вероятностью успеха p. Мы знаем, что $\mathbf{S}_n$ имеет биномиальное распределение $\{b(k;\ n,\ p)\}$; следовательно, $\mathbf{E}(\mathbf{S}_n) = \sum kb(k;\ n,\ p) = np \sum b(k-1;\ n-1,\ p)$. Последняя сумма состоит из всех членов биномиального распределения для $n-1$ испытаний и поэтому равна 1. Итак, *математическое ожидание биномиального распределения* равно

$$\mathbf{E}(\mathbf{S}_n) = np. \tag{3.1}$$

Этот результат можно получить без вычислений с помощью метода, который часто оказывается полезным. Пусть $\mathbf{X}_k$ — число успехов в k-м испытании. Эта случайная величина принимает только значения 0 и 1 с соответствующими вероятностями q и p. Следовательно,

$$\mathbf{E}(\mathbf{X}_k) = 0 \cdot q + 1 \cdot p = p,$$

а так как

$$\mathbf{S}_n = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n, \tag{3.2}$$

то мы получаем (3.1) непосредственно из (2.4).

б) *Распределение Пуассона.* Если $\mathbf{X}$ имеет распределение Пуассона $p(k;\ \lambda) = e^{-\lambda}\lambda^k/k!$, где $k = 0,\ 1,\ \ldots$, то

$$\mathbf{E}(\mathbf{X}) = \sum kp(k;\ \lambda) = \lambda \sum p(k-1;\ \lambda).$$

Последний ряд содержит все члены распределения, и поэтому его сумма равна единице. Следовательно, λ *есть математическое ожидание распределения Пуассона* $\{e^{-\lambda}\lambda^k/k!\}$.

в) *Отрицательное биномиальное распределение.* Пусть $\mathbf{X}$ — случайная величина, имеющая *геометрическое распределение* $\mathbf{P}\{\mathbf{X} = k\} = q^k p$, где $k = 0,\ 1,\ 2,\ \ldots$. Тогда $\mathbf{E}(\mathbf{X}) = qp(1 + 2q + 3q^2 + \ldots)$. Справа в скобках стоит производная геометрического ряда, поэтому $\mathbf{E}(\mathbf{X}) = qp(1-q)^{-2} = q/p$. В гл. VI, 8 мы видели, что $\mathbf{X}$ можно трактовать как число неудач до первого успеха в последовательности испытаний Бернулли. В более общем случае рассматривается пространство элементарных событий, соответствующее испытаниям Бернулли, которые продолжаются до n-го успеха. Для $r < n$ положим $\mathbf{X}_1 = \mathbf{X}$ и обозначим через $\mathbf{X}_r$ число неудач между $(r-1)$-м и r-м успехами. Тогда каждая величина $\mathbf{X}_\nu$ имеет геометрическое распределение $\{q^k p\}$ и $\mathbf{E}(\mathbf{X}_\nu) = q/p$. Сумма

$$\mathbf{Y}_r = \mathbf{X}_1 + \ldots + \mathbf{X}_r$$

дает число неудач до r-го успеха. Иначе говоря, $\mathbf{Y}_r$ есть случайная величина с отрицательным биномиальным распределением, определяемым одной из двух эквивалентных формул (8.1) или (8.2) гл. VI. Отсюда следует, что *математическое ожидание отрицательного* *биномиального распределения равно* rq/p. Это можно проверить непосредственными вычислениями. Из формулы (8.2) гл. VI ясно, что

$$kf(k;\ r,\ p)=rp^{-1}\,qf(k-1;\ r+1,\ p),$$

и сумма членов распределения $\{f(k-1;\ r+1,\ p)\}$ дает единицу. Эти непосредственные вычисления хороши тем, что они применимы и в случае *нецелых* значений r. С другой стороны, первоначальные рассуждения позволяют получить тот же результат, не зная точного вида распределения $\mathbf{X}_1+\ldots+\mathbf{X}_r$.

г) *Время ожидания при выборе.* Из генеральной совокупности, содержащей N различных элементов, производится выбор с возвращением. Вследствие возможных повторений случайная выборка объема r в общем случае будет состоять из менее чем r различных элементов. При увеличении объема выборки новые элементы будут появляться в ней все реже и реже. Нас интересует, каков должен быть объем выборки $\mathbf{S}_r$, чтобы в ней содержалось r различных элементов. (Как частный случай рассмотрим генеральную совокупность из $N=365$ возможных дней рождения; здесь $\mathbf{S}_r$ представляет собой число людей, попавших в выборку до момента, когда в выборке окажется r различных дней рождения. Аналогичная трактовка возможна и при случайном размещении шаров по ящикам. Рассматриваемая задача представляет особый интерес для тех, кто собирает купоны и другие предметы, так как новое приобретение можно сравнить со случайным выбором [1].)

Для упрощения изложения назовем очередное извлечение элемента успешным, если выбранный элемент появляется в выборке впервые. Тогда $\mathbf{S}_r$ представляет собой число извлечений до r-го успеха включительно. Положим $\mathbf{X}_k=\mathbf{S}_{k+1}-\mathbf{S}_k$. Тогда $\mathbf{X}_k-1$ есть число неудачных извлечений между k-м и $(k+1)$-м успехами. Во время этих извлечений в генеральной совокупности было $N-k$ элементов, которые еще не появлялись в выборке, и поэтому $\mathbf{X}_k-1$ равно числу неудач до первого успеха в испытаниях Бернулли с $p=(N-k)/N$. Следовательно, согласно примеру в), имеем $\mathbf{E}(\mathbf{X}_k)=1+q/p=N/(N-k)$. Наконец, поскольку $\mathbf{S}_r=1+\mathbf{X}_1+\ldots+\mathbf{X}_r$, получаем

$$\mathbf{E}(\mathbf{S}_r)=N\ \{1/N+1/(N-1)+\ldots+1/(N-r+1)\}. \tag{3.3}$$

В частности, $\mathbf{E}(\mathbf{S}_N)$ есть ожидаемое число извлечений, необходимых для того, чтобы в выборку вошли все элементы генеральной совокупности. При $N=10$ имеем $\mathbf{E}(\mathbf{S}_5)\approx 6{,}5$ и $\mathbf{E}(\mathbf{S}_{10})\approx 29{,}3$. Это означает, что для выборки первой половины совокупности из 10 элементов в

---

[1]) Polya G., Eine Warscheinlichkeitsaufgabe zur Kundenwerbung, Zeitschrift für Angewandte Mathematik und Mechanik, 10 (1930), 96—97. Пойа другими методами разбирает несколько более общую задачу. Существует обширная литература, посвященная различным вариантам задачи о собирании купонов. (См. задачи 24, 25, а также задачи 12—14 в гл. XI, 7 и задачу 12 в гл. II, 11.)

среднем достаточно семи извлечений, а на вторую половину потребуется еще 23 извлечения.

Чтобы получить приближенную формулу для (3.3), будем трактовать $(N-k)^{-1}$ как площадь прямоугольника, основанием которого является интервал единичной длины с центром в точке $N-k$, а высотой — значение функции x^{-1} в этой точке. Заменяя площадь этого прямоугольника на площадь области, лежащей ниже графика функции x^{-1}, получаем приближение

$$\mathbf{E}(\mathbf{S}_r) \approx N \int_{N-r+1/2}^{N+1/2} x^{-1}\,dx = N \log\left[(N+1/2)/(N-r+1/2)\right]. \quad (3.4)$$

В качестве примера использования этого приближения возьмем произвольное $\alpha<1$ и рассмотрим *ожидаемое число извлечений, необходимых для получения в выборке доли α всей генеральной совокупности*. Оно равно $\mathbf{E}(\mathbf{S}_r)$, где r — наименьшее целое $\geqslant \alpha N$. При $N\to\infty$ погрешность, допускаемая при использовании (3.4), стремится к нулю и искомое математическое ожидание в пределе равно $N\log(1-\alpha)^{-1}$. Заметим, что все эти результаты получены без использования самого распределения вероятностей величины $\mathbf{S}_r$. [Его легко найти по вероятностям (2.3) гл. IV, полученным в задаче о размещении.]

д) *Задача об оценке*. Некоторая урна содержит шары с номерами от 1 до N. Пусть $\mathbf{X}$ — *наибольший номер, полученный при n извлечениях*, если производится случайный выбор с возвращением. Событие $\mathbf{X}\leqslant k$ означает, что каждый из n вынутых номеров не превышает k, и, следовательно, $\mathbf{P}\{\mathbf{X}\leqslant k\}=(k/N)^n$. Таким образом, распределение вероятностей случайной величины $\mathbf{X}$ определяется формулой

$$\begin{aligned} p_k &= \mathbf{P}\{\mathbf{X}=k\} = \mathbf{P}\{\mathbf{X}\leqslant k\} - \mathbf{P}\{\mathbf{X}\leqslant k-1\} = \\ &= \{k^n-(k-1)^n\}\,N^{-n}. \end{aligned} \quad (3.5)$$

Отсюда следует, что

$$\begin{aligned} \mathbf{E}(\mathbf{X}) &= \sum_{k=1}^{N} k p_k = N^{-n}\sum_{k=1}^{N}\{k^{n+1}-(k-1)^{n+1}-(k-1)^n\} = \\ &= N^{-n}\left\{N^{n+1}-\sum_{k=1}^{N}(k-1)^n\right\}. \end{aligned} \quad (3.6)$$

При больших N последняя сумма приближенно равна площади области, лежащей ниже кривой $y=x^n$ и ограниченной прямыми $x=0$ и $x=N$, т. е. $N^{n+1}/(n+1)$. Отсюда следует, что при больших N

$$\mathbf{E}(\mathbf{X}) \approx [n/(n+1)]\,N. \quad (3.7)$$

Если в городе имеется $N=1000$ машин и производится случайная выборка объема $n=10$, то математическое ожидание максимального наблюденного регистрационного номера приближенно равно 910. В прикладной статистике наблюденный максимум выборки используется для оценки неизвестного истинного значения N. Этот метод применялся во время последней войны для оценки объема производства противника (см. задачи 8 и 9).

е) *Использование в статистическом критерии.* Этот пример [1]) иллюстрирует практическое применение понятия математического ожидания, позволяющее избежать сложных вычислений, связанных с нахождением вероятностных распределений.

Споры гриба *Sordaria* образуются цепочками по восемь элементов. Каждая цепочка может разделиться на несколько частей, и в конечном счете споры отбрасываются группами, содержащими от 1 до 8 спор. Есть основания предполагать, что разрывы семи соединений между спорами в цепочке стохастически независимы и вероятность разрыва одинакова для всех соединений и равна p. При этих предположениях теоретически возможно найти совместное распределение групп из одной споры, из двух и т. д., но потребуются утомительные вычисления. С другой стороны, для эмпирической проверки гипотез достаточно знать ожидаемые числа групп из одной споры, из двух и т. д., а их легко найти. Например, споры, находящиеся на концах цепочки, с вероятностью p становятся одиночными, в то время как для всех других спор эта вероятность равна p^2. Следовательно, по правилу сложения математическое ожидание числа одиночных спор, образованных из одной цепочки, определяется формулой $\varepsilon_1=2p+6p^2$. Аналогичные рассуждения показывают, что математическое ожидание числа групп из двух спор равно $\varepsilon_2=2qp+5qp^2$, где $q=1-p$. Подобным образом получаем $\varepsilon_3=2q^2p+4q^2p^2, \ldots, \varepsilon_8=q^7$. Математическое ожидание числа всех возможных групп равно $\varepsilon_1+\ldots+\varepsilon_8=1+7p$. (Это очевидно без вычислений, так как математическое ожидание числа разрывов соединений равно $7p$ и каждый разрыв увеличивает количество групп на 1.)

При реальных наблюдениях в полевых условиях было найдено 7251 спор, образовавшихся, очевидно, из $N=907$ цепочек (5 спор не было обнаружено). Если применима наша вероятностная модель, то должно иметь место равенство $(1+7p)\ N\approx 7251$, или $p=0,168$. (Этот довод опирается на интуитивное представление о математическом ожидании и обосновывается законом больших чисел.) Наблюденное число f_k групп должно быть близко к математическому ожиданию $N\varepsilon_k$. Как показывает табл. 3, различия невелики и нет оснований отвергать нашу модель. ►

[1]) Этот пример взят из речи Д. Р. Кокса при его вступлении в должность в Бербек-Колледже (Лондон) в 1961 г. Кокс ссылается на работу Ingold C. T., Hadland S. A., New Phytologist, 58 (1959), 46—57.

Таблица 3

Наблюденные числа f_k и математические ожидания Ne_k чисел групп по k спор в примере «е»

k	f_k	Ne_k	k	f_k	Ne_k
1	490	458,3	5	200	170,6
2	343	360,8	6	134	131,7
3	265	281,8	7	72	101,1
4	199	219,7	8	272	250,3

§ 4. ДИСПЕРСИЯ

Пусть $\mathbf{X}$ — случайная величина с распределением $\{f(x_j)\}$, и пусть $r \geqslant 0$ — некоторое целое число. *Если математическое ожидание случайной величины* $\mathbf{X}^r$, *т. е.*

$$\mathbf{E}(\mathbf{X}^r) = \sum x_j^r f(x_j) \tag{4.1}$$

существует, то оно называется моментом порядка r случайной величины $\mathbf{X}$. Если ряд (4.1) не является абсолютно сходящимся, то говорят, что момент порядка r не существует. Поскольку $|\mathbf{X}|^{r-1} \leqslant |\mathbf{X}|^r + 1$, *из существования момента порядка r вытекает существование момента порядка $r-1$ и, следовательно, всех моментов меньшего порядка.*

Моменты играют важную роль в общей теории, но в этом томе будет использоваться только момент второго порядка. Если он существует, то существует и математическое ожидание

$$\mu = \mathbf{E}(\mathbf{X}). \tag{4.2}$$

Тогда естественно заменить случайную величину $\mathbf{X}$ ее *отклонением от математического ожидания* $\mathbf{X}-\mu$. Поскольку $(x-\mu)^2 \leqslant 2(x^2+\mu^2)$, момент второго порядка величины $\mathbf{X}-\mu$ существует, если существует $\mathbf{E}(\mathbf{X}^2)$, и определяется по формуле

$$\mathbf{E}((\mathbf{X}-\mu)^2) = \sum_j (x_j^2 - 2\mu x_j + \mu^2) f(x_j). \tag{4.3}$$

Разбивая правую часть на три отдельные суммы, получаем эквивалентную запись в виде $\mathbf{E}(\mathbf{X}^2) - 2\mu\mathbf{E}(\mathbf{X}) + \mu^2 = \mathbf{E}(\mathbf{X}^2) - \mu^2$.

Определение. *Пусть* $\mathbf{X}$ — *случайная величина, имеющая момент второго порядка* $\mathbf{E}(\mathbf{X}^2)$, *и пусть* $\mu = \mathbf{E}(\mathbf{X})$ — *ее математическое ожидание. Определим величину, называемую дисперсией случайной величины* $\mathbf{X}$, *формулой*

$$\operatorname{Var}(\mathbf{X}) = \mathbf{E}((\mathbf{X}-\mu)^2) = \mathbf{E}(\mathbf{X}^2) - \mu^2. \tag{4.4}$$

Положительное значение корня квадратного из этого числа (или нуль) называется стандартным отклонением случайной величины **X**.

Для простоты будем часто говорить о дисперсии распределения, не указывая соответствующей случайной величины.

Примеры. а) Если **X** принимает значение $\pm c$, каждое с вероятностью 1/2, то $\operatorname{Var}(\mathbf{X}) = c^2$.

б) Если **X** — число очков, выпавших при бросании симметричной игральной кости, то $\operatorname{Var}(\mathbf{X}) = (1/6)(1^2+2^2+\ldots+6^2) - (7/2)^2 = 35/12$.

в) Для *распределения Пуассона* $p(k;\lambda)$ среднее равно λ (см. пример 3, б)), и поэтому для дисперсии получаем

$$\sum k^2 p(k;\lambda) - \lambda^2 = \lambda \sum kp(k-1;\lambda) - \lambda^2 = \\ = \lambda \sum (k-1)\,p(k-1;\lambda) + \lambda \sum p(k-1;\lambda) - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

В этом случае математическое ожидание и дисперсия совпадают.

г) Для *биномиального распределения* (см. пример 3, а)) аналогичные вычисления показывают, что дисперсия равна

$$\sum k^2 b(k;n,p) - (np)^2 = np \sum kb(k-1;n-1,p) - (np)^2 = \\ = np\{(n-1)p+1\} - (np)^2 = npq. \qquad \blacktriangleright$$

Полезность понятия дисперсии выявится постепенно, в частности, в связи с предельными теоремами в гл. X. Здесь мы отметим, что дисперсия является, хотя и довольно грубой, *мерой разброса* значений случайной величины. Действительно, если дисперсия $\operatorname{Var}(\mathbf{X}) = \sum (x_j-\mu)^2 f(x_j)$ мала, то каждый член этой суммы тоже мал. Следовательно, значение x_j, при котором $|x_j-\mu|$ велико, должно иметь малую вероятность. Иначе говоря, при малой дисперсии большие отклонения **X** от среднего μ маловероятны. Обратно, большая дисперсия указывает на то, что не все значения случайной величины **X** лежат вблизи от математического ожидания.

Возможно, некоторым читателям поможет следующая механическая интерпретация. Предположим, что механическая система состоит из единичной массы, распределенной вдоль оси x так, что в точке x_j сконцентрирована масса $f(x_j)$. Тогда среднее μ есть абсцисса *центра тяжести*, а дисперсия — *момент инерции*. Ясно, что различные распределения массы могут иметь одинаковые абсциссу центра тяжести и момент инерции, но хорошо известно, что этими двумя величинами могут быть описаны некоторые важные механические свойства такой системы.

Если **X** представляет собой некоторую величину, которую можно измерить, типа длины или температуры, то ее численные значения зависят от начала отсчета и единицы измерения. Изменение последних означает переход от **X** к новой величине $a\mathbf{X}+b$, где a и b — постоянные. Ясно, что $\operatorname{Var}(\mathbf{X}+b) = \operatorname{Var}(\mathbf{X})$, и поэтому

$$\operatorname{Var}(a\mathbf{X}+b) = a^2 \operatorname{Var}(\mathbf{X}). \tag{4.5}$$

Выбор начала отсчета и единицы измерения является в значительной степени произвольным, и часто за начало отсчета удобнее всего брать среднее, а за единицу измерения — стандартное отклонение. Так мы поступили в гл. VII, 3, когда ввели нормированное число успехов $\mathbf{S}_n^* = (\mathbf{S}_n - np)/\sqrt{npq}$. В общем случае, если $\mathbf{X}$ имеет среднее μ и дисперсию σ^2, то $\mathbf{X}-\mu$ имеет нулевое среднее и дисперсию σ^2, и поэтому у *случайной величины*

$$\mathbf{X}^* = (\mathbf{X}-\mu)/\sigma \quad (\sigma > 0) \tag{4.6}$$

среднее равно 0 *и дисперсия* 1. *Она называется нормированной случайной величиной, соответствующей* $\mathbf{X}$. На языке физики переход от $\mathbf{X}$ к $\mathbf{X}^*$ трактовался бы как введение безразмерной величины.

§ 5. КОВАРИАЦИЯ; ДИСПЕРСИЯ СУММЫ

Пусть $\mathbf{X}$ и $\mathbf{Y}$ — две случайные величины, определенные на одном пространстве элементарных событий. Тогда $\mathbf{X}+\mathbf{Y}$ и $\mathbf{XY}$ тоже случайные величины, и их распределения могут быть найдены путем простой группировки членов совместного распределения $\mathbf{X}$ и $\mathbf{Y}$. Наша цель состоит теперь в вычислении $\operatorname{Var}(\mathbf{X}+\mathbf{Y})$. Для этого введем понятие ковариации, которое будет подробнее проанализировано в § 8. Если $\{p(x_j, y_k)\}$ — совместное распределение $\mathbf{X}$ и $\mathbf{Y}$, то математическое ожидание $\mathbf{XY}$ определяется формулой

$$\mathrm{E}(\mathbf{XY}) = \sum x_j y_k p(x_j, y_k) \tag{5.1}$$

при условии, конечно, что ряд сходится абсолютно. Далее, $|x_j y_k| \leqslant (x_j^2 + y_k^2)/2$, и, следовательно, $\mathrm{E}(\mathbf{XY})$ обязательно существует, если существуют $\mathrm{E}(\mathbf{X}^2)$ и $\mathrm{E}(\mathbf{Y}^2)$. В этом случае существуют также математические ожидания

$$\mu_x = \mathrm{E}(\mathbf{X}), \quad \mu_y = \mathrm{E}(\mathbf{Y}) \tag{5.2}$$

и величины $\mathbf{X}-\mu_x$ и $\mathbf{Y}-\mu_y$ имеют нулевые средние. Для их произведения по теореме сложения из § 2 имеем

$$\begin{aligned}\mathrm{E}((\mathbf{X}-\mu_x)(\mathbf{Y}-\mu_y)) &= \mathrm{E}(\mathbf{XY}) - \mu_x \mathrm{E}(\mathbf{Y}) - \mu_y \mathrm{E}(\mathbf{X}) + \mu_x\mu_y = \\ &= \mathrm{E}(\mathbf{XY}) - \mu_x\mu_y.\end{aligned} \tag{5.3}$$

Определение. *Ковариация случайных величин* $\mathbf{X}$ *и* $\mathbf{Y}$ *определяется формулой*

$$\operatorname{Cov}(\mathbf{X}, \mathbf{Y}) = \mathrm{E}((\mathbf{X}-\mu_x)(\mathbf{Y}-\mu_y)) = \mathrm{E}(\mathbf{XY}) - \mu_x\mu_y. \tag{5.4}$$

Это определение имеет смысл, если $\mathbf{X}$ *и* $\mathbf{Y}$ *имеют конечные дисперсии.*

Из § 2 мы знаем, что для независимых случайных величин $\mathrm{E}(\mathbf{XY}) = \mathrm{E}(\mathbf{X})\,\mathrm{E}(\mathbf{Y})$. Следовательно, из (5.4) вытекает следующая теорема.

Теорема 1. *Если* $\mathbf{X}$ *и* $\mathbf{Y}$ *независимы, то* $\operatorname{Cov}(\mathbf{X}, \mathbf{Y}) = 0$.

Заметим, что *обратное утверждение неверно*. Табл. 1 дает пример двух зависимых случайных величин, ковариация которых тем не менее равна нулю. Мы вернемся к этому вопросу в § 8. Следующая теорема важна, и правило сложения (5.6) для независимых случайных величин постоянно применяется.

Теорема 2. *Если* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — *случайные величины с конечными дисперсиями* $\sigma_1^2, \ldots, \sigma_n^2$ *и* $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$, *то*

$$\operatorname{Var}(\mathbf{S}_n) = \sum_{k=1}^{n} \sigma_k^2 + 2\sum_{j,k} \operatorname{Cov}(\mathbf{X}_j, \mathbf{X}_k), \tag{5.5}$$

причем последняя сумма состоит из $\binom{n}{2}$ *пар* $(\mathbf{X}_j, \mathbf{X}_k)$ *с* $j < k$.

В частности, если $\mathbf{X}_j$ *взаимно независимы, то*

$$\operatorname{Var}(\mathbf{S}_n) = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2. \tag{5.6}$$

Доказательство. Положим $\mu_k = \mathbf{E}(\mathbf{X}_k)$ и $m_n = \mu_1 + \ldots + \mu_n = \mathbf{E}(\mathbf{S}_n)$. Тогда $\mathbf{S}_n - m_n = \sum (\mathbf{X}_k - \mu_k)$ и

$$(\mathbf{S}_n - m_n)^2 = \sum (\mathbf{X}_k - \mu_k)^2 + 2\sum (\mathbf{X}_j - \mu_j)(\mathbf{X}_k - \mu_k). \tag{5.7}$$

Вычисляя математическое ожидание от обеих частей (5.7), получаем (5.5). ▶

Примеры. а) *Биномиальное распределение* $\{b(k;\, n, p)\}$. В примере 3, а) случайные величины $\mathbf{X}_k$ взаимно независимы. Имеем

$$\mathbf{E}(\mathbf{X}_k^2) = 0^2 \cdot q + 1^2 \cdot p = p$$

и $\mathbf{E}(\mathbf{X}_k) = p$. Следовательно, $\sigma_k^2 = p - p^2 = pq$, и из (5.6) видно, что *дисперсия биномиального распределения* равна npq. Этот же результат был получен в примере 4, г) непосредственным вычислением.

б) *Испытания Бернулли с переменными вероятностями успеха.* Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — взаимно независимые случайные величины, такие, что $\mathbf{X}_k$ принимает значения 1 и 0 с вероятностями соответственно p_k и $q_k = 1 - p_k$. Тогда $\mathbf{E}(\mathbf{X}_k) = p_k$ и $\operatorname{Var}(\mathbf{X}_k) = p_k - p_k^2 = p_k q_k$; снова полагая

$$\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n,$$

из (5.6) имеем

$$\operatorname{Var}(\mathbf{S}_n) = \sum_{k=1}^{n} p_k q_k. \tag{5.8}$$

Как в примере 1, д), величину $\mathbf{S}_n$ можно интерпретировать как общее число успехов в n независимых испытаниях, каждое из которых заканчивается либо успехом, либо неудачей. Тогда $p = (p_1 + \ldots + p_n)/n$ есть средняя вероятность успеха, и вполне естественно сравнить описанную ситуацию с испытаниями Бернулли с постоянной вероятностью успеха p. Такое сравнение приводит к поразительному результату. Равенство (5.8) можно переписать в виде

$$\mathrm{Var}(\mathbf{S}_n) = np - \sum p_k^2.$$

Далее, легко показать (элементарными вычислениями или по индукции), что из всех комбинаций $\{p_k\}$, таких, что $\sum p_k = np$, сумма $\sum p_k^2$ имеет наименьшее значение, когда все p_k равны. Отсюда следует, что если средняя вероятность успеха p остается постоянной, то дисперсия $\mathrm{Var}(\mathbf{S}_n)$ *достигает максимума при* $p_1 = \ldots = p_n = p$. Итак, получен удивительный результат: *изменение* p_k, *или их неодинаковость, уменьшает величину случайных флуктуаций*, которая характеризуется дисперсией[1]). Например, число пожаров в городе за год можно рассматривать как случайную величину; при фиксированном среднем значении дисперсия максимальна, если вероятности пожара для всех строений одинаковы. Если для n машин задана средняя производительность p, *то производительность всего комплекса будет наименее равномерной тогда, когда все машины одинаковы*. (Приложение к современному образованию очевидно, но оно не обнадеживает.)

в) *Совпадение карт.* Колода из n занумерованных карт раскладывается в случайном порядке так, что все $n!$ возможных расположений равновероятны. Число совпадений (число карт, попавших на место, соответствующее их номеру) представляет собой случайную величину $\mathbf{S}_n$ со значениями $0, 1, \ldots, n$. Ее распределение вероятностей найдено в гл. IV, 4. Среднее и дисперсию можно вычислить, исходя из этого распределения, однако следующий способ проще и поучительней.

Определим случайную величину $\mathbf{X}_k$, принимающую значения 1 или 0 в зависимости от того, попала ли карта с номером k на k-е место или нет. Тогда $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. Далее, каждая карта с вероятностью $1/n$ оказывается на k-м месте. Следовательно, $\mathbf{P}\{\mathbf{X}_k = 1\} = 1/n$ и $\mathbf{P}\{\mathbf{X}_k = 0\} = (n-1)/n$. Поэтому $\mathbf{E}(\mathbf{X}_k) = 1/n$, и отсюда следует, что $\mathbf{E}(\mathbf{S}_n) = 1$: в среднем на колоду приходится одно совпадение.

Чтобы найти $\mathrm{Var}(\mathbf{S}_n)$, вычислим сначала дисперсию σ_k^2 случайной величины $\mathbf{X}_k$:

$$\sigma_k^2 = 1/n - (1/n)^2 = (n-1)/n^2. \tag{5.9}$$

Затем найдем $\mathbf{E}(\mathbf{X}_j \mathbf{X}_k)$. Произведение $\mathbf{X}_j \mathbf{X}_k$ равно 0 или 1, причем последнее имеет место в том случае, когда обе карты с номерами j и k оказываются на своих местах, а вероятность этого события равна

[1]) Более сильные результаты в том же направлении см. в работе Hoeffding W., On the distribution of the number of successes in independent trials, Ann. Math. Statist., 27 (1956), 713—721. О приближении распределением Пуассона см. пример гл. XI, 6, б).

$1/n(n-1)$. Следовательно,

$$\mathbf{E}(\mathbf{X}_j\mathbf{X}_k) = 1/[n(n-1)],$$
$$\operatorname{Cov}(\mathbf{X}_j, \mathbf{X}_k) = 1/[n(n-1)] - 1/n^2 = 1/[n^2(n-1)]. \tag{5.10}$$

Итак, наконец,

$$\operatorname{Var}(\mathbf{S}_n) = n\frac{n-1}{n^2} + 2\binom{n}{2}\frac{1}{n^2(n-1)} = 1. \tag{5.11}$$

Мы видим, что и среднее, и дисперсия числа совпадений равны единице. Этот результат можно использовать в задаче об *угадывании карт*, которая обсуждалась в гл. IV, 4. Там рассматривались три метода угадывания, первый из которых соответствовал совпадению карт. Второй можно описать как последовательность n испытаний Бернулли с вероятностью $p=1/n$; в этом случае математическое ожидание числа верных догадок равно $np=1$ и дисперсия равна $npq=(n-1)/n$. Математические ожидания в обоих случаях одинаковы, но бо́льшая дисперсия при первом методе указывает на бо́льшие случайные флуктуации около среднего и тем самым обещает несколько более острую игру. (При колодах более сложного состава различие между двумя дисперсиями станет несколько больше, хотя никогда не будет действительно значительным.) При последнем способе угадывания испытуемый все время называет одну и ту же карту; число верных догадок, очевидно, равно единице, а случайных колебаний вообще нет (нулевая дисперсия). Мы видим, что способ угадывания не может повлиять на математическое ожидание числа верных догадок, но от него зависит величина случайных флуктуаций.

г) *Выбор без возвращения.* Предположим, что генеральная совокупность состоит из b черных и g зеленых элементов и производится случайная выборка объема r (без возвращения). Число $\mathbf{S}_k$ черных элементов в выборке есть случайная величина с *гипергеометрическим распределением* (см. гл. II, 6), поэтому математическое ожидание и дисперсия могут быть найдены прямым вычислением. Однако предпочтительнее другой метод. Введем случайную величину $\mathbf{X}_k$, принимающую значения 1 или 0 в зависимости от того, является ли k-й элемент в выборке черным или нет ($k \leqslant r$). Из соображений симметрии вероятность того, что $\mathbf{X}_k=1$, равна $b/(b+g)$, и поэтому

$$\mathbf{E}(\mathbf{X}_k) = b/(b+g), \quad \operatorname{Var}(\mathbf{X}_k) = bg/(b+g)^2. \tag{5.12}$$

Далее, при $j \neq k$ произведение $\mathbf{X}_j\mathbf{X}_k=1$, если j-й и k-й элементы выборки черные, в противном случае $\mathbf{X}_j\mathbf{X}_k=0$. Вероятность того, что $\mathbf{X}_j\mathbf{X}_k=1$, равна $b(b-1)/[(b+g)(b+g-1)]$, и поэтому

$$\mathbf{E}(\mathbf{X}_j\mathbf{X}_k) = b(b-1)/[(b+g)(b+g-1)],$$
$$\operatorname{Cov}(\mathbf{X}_j\mathbf{X}_k) = -bg/[(b+g)^2(b+g-1)]. \tag{5.13}$$

Таким образом,

$$\mathbf{E}(\mathbf{S}_r)=rb/(b+g), \quad \operatorname{Var}(\mathbf{S}_r)=\frac{rbg}{(b+g)^2}\left\{1-\frac{r-1}{b+g-1}\right\}. \tag{5.14}$$

При выборке с возвращением мы имели бы такое же среднее и немного бо́льшую дисперсию, а именно $rbg/(b+g)^2$.

§ 6. НЕРАВЕНСТВО ЧЕБЫШЕВА [1]

Мы видим, что малость дисперсии указывает на малую вероятность больших отклонений от математического ожидания. Это утверждение уточняется неравенством Чебышева, которое является чрезвычайно полезным. Неравенство Чебышева предполагает существование момента второго порядка.

Теорема. *Для произвольного* $t>0$

$$\mathbf{P}\{|\mathbf{X}|\geqslant t\}\leqslant t^{-2}\mathbf{E}(\mathbf{X}^2). \tag{6.1}$$

В частности, если $\mathbf{E}(\mathbf{X})=\mu$, *то*

$$\mathbf{P}\{|\mathbf{X}-\mu|\geqslant t\}\leqslant t^{-2}\operatorname{Var}(\mathbf{X}). \tag{6.2}$$

Доказательство. Если первое неравенство применить к случайной величине $\mathbf{X}-\mu$, то получаем второе неравенство. Используя обозначения § 4, имеем

$$\mathbf{P}\{|\mathbf{X}|>t\}=\sum_{|x_j|\geqslant t} f(x_j)\leqslant t^{-2}\sum_{|x_j|\geqslant t} x_j^2 f(x_j), \tag{6.3}$$

суммирование производится по всем x_j, абсолютная величина которых не меньше t. Последняя сумма не превышает $\mathbf{E}(\mathbf{X}^2)$, и поэтому (6.1) справедливо.

Неравенство Чебышева нужно рассматривать скорее как теоретический инструмент, чем как практический метод оценивания. Важность этого неравенства вытекает из его универсальности, но от утверждений большой общности нельзя ожидать точных результатов в отдельных случаях.

Примеры. а) Если $\mathbf{X}$ — число очков, выпавших при бросании правильной игральной кости, то (см. пример 4, б)) $\mu=7/2$, $\sigma^2=35/12$. Максимум отклонения $\mathbf{X}$ от μ равен $2{,}5\approx 3\sigma/2$. Вероятность бо́льших отклонений равна нулю, в то время как неравенство Чебышева утверждает лишь, что эта вероятность меньше чем 0,47.

б) Для биномиального распределения $\{b(k;\ n,\ p)\}$ имеем (см. пример 5, а)) $\mu=np$, $\sigma^2=npq$. При больших n мы знаем, что

$$\mathbf{P}\{|\mathbf{S}_n-np|>x\sqrt{npq}\}\approx 1-\mathfrak{N}(x)+\mathfrak{N}(-x). \tag{6.4}$$

Неравенство Чебышева утверждает лишь, что левая часть меньше чем x^{-2}; это, очевидно, намного более слабая оценка, чем (6.4).

[1] П. Л. Чебышев (1821—1894).

§ 7*). НЕРАВЕНСТВО КОЛМОГОРОВА

В качестве примера более точной оценки докажем следующее утверждение.

Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — *взаимно независимые случайные величины с математическими ожиданиями* $\mu_k = \mathbf{E}(\mathbf{X}_k)$ *и дисперсиями* σ_k^2. *Положим*

$$\mathbf{S}_k = \mathbf{X}_1 + \ldots + \mathbf{X}_k, \tag{7.1}$$

$$m_k = \mathbf{E}(\mathbf{S}_k) = \mu_1 + \ldots + \mu_k,$$
$$s_k^2 = \operatorname{Var}(\mathbf{S}_k) = \sigma_1^2 + \ldots + \sigma_k^2. \tag{7.2}$$

Для каждого $t > 0$ *вероятность того, что одновременно выполняются* n *неравенств*

$$|\mathbf{S}_k - m_k| < ts_n, \quad k = 1, 2, \ldots, n, \tag{7.3}$$

не меньше, чем $1 - t^{-2}$.

При $n = 1$ эта теорема сводится к неравенству Чебышева. При $n > 1$ неравенство Чебышева дает ту же самую оценку для вероятности единственного соотношения $|\mathbf{S}_n - m_n| < ts_n$, так что неравенство Колмогорова значительно сильнее.

Доказательство. Оценим вероятность x того, что хотя бы одно из неравенств (7.3) не выполняется. Теорема утверждает, что $x \leqslant t^{-2}$.

Определим n случайных величин $\mathbf{Y}_\nu$ следующим образом: $\mathbf{Y}_\nu = 1$, если

$$|\mathbf{S}_\nu - m_\nu| \geqslant ts_n, \tag{7.4}$$

но

$$|\mathbf{S}_k - m_k| < ts_n \quad \text{для } k = 1, 2, \ldots, \nu - 1; \tag{7.5}$$

$\mathbf{Y}_\nu = 0$ во всех остальных случаях. Словесно это выражается так: $\mathbf{Y}_\nu$ равно 1 для всех тех элементарных событий, для которых первым из невыполняющихся неравенств (7.3) является ν-е. Тогда для каждого элементарного события не более чем одна из величин $\mathbf{Y}_k$ равна 1 и сумма $\mathbf{Y}_1 + \mathbf{Y}_2 + \ldots + \mathbf{Y}_n$ может принимать только значения 0 и 1. Эта сумма равна 1 тогда и только тогда, когда хотя бы одно из неравенств (7.3) не выполняется, и поэтому

$$x = \mathbf{P}\{\mathbf{Y}_1 + \ldots + \mathbf{Y}_n = 1\}. \tag{7.6}$$

Поскольку $\mathbf{Y}_1 + \ldots + \mathbf{Y}_n$ равно 0 или 1, имеем $\sum \mathbf{Y}_k \leqslant 1$. Умножая обе части последнего неравенства на $(\mathbf{S}_n - m_n)^2$ и вычисляя математическое ожидание, получаем

$$\sum_{k=1}^{n} \mathbf{E}(\mathbf{Y}_k (\mathbf{S}_n - m_n)^2) \leqslant s_n^2. \tag{7.7}$$

*) В этом параграфе рассматривается специальный вопрос, и при первом чтении его можно пропустить.

Для оценки слагаемых в левой части положим

$$\mathbf{U}_k = (\mathbf{S}_n - m_n) - (\mathbf{S}_k - m_k) = \sum_{\nu=k+1}^{n} (\mathbf{X}_\nu - \mu_\nu); \tag{7.8}$$

тогда

$$\mathbf{E}(\mathbf{Y}_k (\mathbf{S}_n - m_n)^2) = \mathbf{E}(\mathbf{Y}_k (\mathbf{S}_k - m_k)^2) + 2\mathbf{E}(\mathbf{Y}_k \mathbf{U}_k (\mathbf{S}_k - m_k)) + \mathbf{E}(\mathbf{Y}_k \mathbf{U}_k^2). \tag{7.9}$$

Далее, $\mathbf{U}_k$ зависит только от $\mathbf{X}_{k+1}, \ldots, \mathbf{X}_n$, в то время как $\mathbf{Y}_k$ и $\mathbf{S}_k$ зависят только от $\mathbf{X}_1, \ldots, \mathbf{X}_k$. Следовательно, $\mathbf{U}_k$ не зависит от $\mathbf{Y}_k (\mathbf{S}_k - m_k)$, и поэтому $\mathbf{E}(\mathbf{Y}_k \mathbf{U}_k (\mathbf{S}_k - m_k)) = \mathbf{E}(\mathbf{Y}_k (\mathbf{S}_k - m_k)) \times \mathbf{E}(\mathbf{U}_k) = 0$, так как $\mathbf{E}(\mathbf{U}_k) = 0$. Таким образом, из (7.9) находим

$$\mathbf{E}(\mathbf{Y}_k (\mathbf{S}_n - m_n)^2) \geqslant \mathbf{E}(\mathbf{Y}_k (\mathbf{S}_k - m_k)^2). \tag{7.10}$$

Но $\mathbf{Y}_k \neq 0$ только тогда, когда $|\mathbf{S}_k - m_k| \geqslant t s_n$, так что $\mathbf{Y}_k (\mathbf{S}_k - m_k)^2 \geqslant t^2 s_n^2 \mathbf{Y}_k$. Следовательно, объединяя (7.7) и (7.10), получаем

$$s_n^2 \geqslant t^2 s_n^2 \mathbf{E}(\mathbf{Y}_1 + \ldots + \mathbf{Y}_n). \tag{7.11}$$

Поскольку $\mathbf{Y}_1 + \ldots + \mathbf{Y}_n$ равняется либо 0, либо 1, математическое ожидание правой части равно вероятности x, определяемой формулой (7.6). Итак, $xt^2 \leqslant 1$, что и утверждалось. ►

§ 8*). КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Пусть $\mathbf{X}$ и $\mathbf{Y}$ — две произвольные случайные величины со средними μ_x и μ_y и положительными дисперсиями σ_x^2 и σ_y^2. Введем соответствующие нормированные случайные величины $\mathbf{X}^*$ и $\mathbf{Y}^*$, определяемые формулой (4.6). Их ковариация называется *коэффициентом корреляции случайных величин* $\mathbf{X}, \mathbf{Y}$ *и обозначается* $\rho(\mathbf{X}, \mathbf{Y})$. Итак, используя (5.4), имеем

$$\rho(\mathbf{X}, \mathbf{Y}) = \mathrm{Cov}(\mathbf{X}^*, \mathbf{Y}^*) = \mathrm{Cov}(\mathbf{X}, \mathbf{Y}) / (\sigma_x \sigma_y). \tag{8.1}$$

◄ Ясно, что коэффициент корреляции не зависит от выбора начала отсчета и единицы измерения, т. е. для произвольных постоянных a_1, a_2, b_1, b_2, где $a_1 > 0$, $a_2 > 0$, имеем $\rho(a_1\mathbf{X} + b_1, a_2\mathbf{Y} + b_2) = \rho(\mathbf{X}, \mathbf{Y})$.

Использование коэффициента корреляции равносильно особой форме записи ковариации [1]). К сожалению, термин «корреляция» создает неправильное представление о свойствах этого коэффициента. Из § 5 известно, что $\rho(\mathbf{X}, \mathbf{Y}) = 0$, если $\mathbf{X}$ и $\mathbf{Y}$ независимы. Важно понять, что обратное утверждение неверно. На самом деле *коэффициент корреляции* $\rho(\mathbf{X}, \mathbf{Y})$ *может быть равен нулю тогда, когда* $\mathbf{Y}$ *является функцией от* $\mathbf{X}$.

*) В этом параграфе рассматривается специальный вопрос, и при первом чтении его можно пропустить.

[1]) Физик определил бы коэффициент корреляции как «безразмерную ковариацию».

Примеры. а) Пусть X принимает значения ± 1, ± 2 с вероятностью 1/4 каждое. Предположим, что $\mathbf{Y}=\mathbf{X}^2$. Совместное распределение определяется соотношениями $p(-1,1)=p(1,1)=p(2,4)=p(-2,4)=1/4$. Из соображений симметрии $\rho(\mathbf{X}, \mathbf{Y})=0$, хотя между $\mathbf{Y}$ и $\mathbf{X}$ существует функциональная зависимость.

б) Пусть $\mathbf{U}$ и $\mathbf{V}$ имеют одинаковые распределения и $\mathbf{X}=\mathbf{U}+\mathbf{V}$, $\mathbf{Y}=\mathbf{U}-\mathbf{V}$. Тогда $\mathbf{E}(\mathbf{XY})=\mathbf{E}(\mathbf{U}^2)-\mathbf{E}(\mathbf{V}^2)=0$ и $\mathbf{E}(\mathbf{Y})=0$. Следовательно, $\mathrm{Cov}(\mathbf{X}, \mathbf{Y})=0$, и поэтому также $\rho(\mathbf{X}, \mathbf{Y})=0$. Например, $\mathbf{X}$ и $\mathbf{Y}$ могут быть соответственно суммой и разностью очков, выпавших на двух костях. Тогда величины $\mathbf{X}$ и $\mathbf{Y}$ либо обе четны, либо обе нечетны и, следовательно, зависимы. ►

Отсюда вытекает, что коэффициент корреляции никоим образом не является исчерпывающей мерой зависимости между $\mathbf{X}$ и $\mathbf{Y}$. Однако $\rho(\mathbf{X}, \mathbf{Y})$ связано с *линейной* зависимостью между $\mathbf{X}$ и $\mathbf{Y}$.

Теорема. *Всегда* $|\rho(\mathbf{X}, \mathbf{Y})|\leqslant 1$, *причем* $\rho(\mathbf{X}, \mathbf{Y})=\pm 1$ *только тогда, когда существуют постоянные* a *и* b, *такие, что справедливо равенство* $\mathbf{Y}=a\mathbf{X}+b$, *за исключением, быть может, тех значений* $\mathbf{X}$, *которые принимаются с нулевой вероятностью.*

Доказательство. Пусть $\mathbf{X}^*$ и $\mathbf{Y}^*$ — нормированные случайные величины. Тогда

$$\mathrm{Var}(\mathbf{X}^*\pm\mathbf{Y}^*)=\mathrm{Var}(\mathbf{X}^*)\pm 2\,\mathrm{Cov}(\mathbf{X}^*, \mathbf{Y}^*)+\mathrm{Var}(\mathbf{Y}^*)= =2(1\pm\rho(\mathbf{X}, \mathbf{Y})). \qquad (8.2)$$

Левая часть не может быть отрицательной, следовательно, $|\rho(\mathbf{X}, \mathbf{Y})|\leqslant 1$. Для $\rho(\mathbf{X}, \mathbf{Y})=1$ необходимо, чтобы $\mathrm{Var}(\mathbf{X}^*-\mathbf{Y}^*)=0$, а это значит, что случайная величина $\mathbf{X}^*-\mathbf{Y}^*$ принимает с вероятностью единица одно значение. В этом случае $\mathbf{X}^*-\mathbf{Y}^*=\mathrm{const}$, и поэтому $\mathbf{Y}=a\mathbf{X}+\mathrm{const}$, где $a=\sigma_y/\sigma_x$. Аналогичное доказательство применимо и к случаю $\rho(\mathbf{X}, \mathbf{Y})=-1$. ►

§ 9. ЗАДАЧИ

1. Семь шаров случайно распределяются по семи ящикам. Пусть $\mathbf{X}_i$ — число ящиков, содержащих ровно i шаров. Используя вероятности из таблицы в § 5 гл. II, выписать совместное распределение $\mathbf{X}_2$ и $\mathbf{X}_3$.

2. Бросаются две правильные кости. Пусть $\mathbf{X}$ — число очков на первой кости и $\mathbf{Y}$ — бо́льшее из двух выпавших чисел. а) Выписать совместное распределение $\mathbf{X}$ и $\mathbf{Y}$. б) Найти математические ожидания, дисперсии и ковариацию.

3. Пусть $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ — соответственно число выпадений герба, число серий гербов и длина максимальной из этих серий при пяти бросаниях монеты. Составить таблицу из 32 элементарных событий с соответствующими значениями $\mathbf{X}$, $\mathbf{Y}$ и $\mathbf{Z}$.

Простым подсчетом найти совместное распределение пар $(\mathbf{X}, \mathbf{Y})$, $(\mathbf{X}, \mathbf{Z})$, $(\mathbf{Y}, \mathbf{Z})$ и распределения величин $\mathbf{X}+\mathbf{Y}$ и $\mathbf{XY}$. Вычислить математические ожидания, дисперсии и ковариации этих случайных величин.

4. Пусть $\mathbf{X}$, $\mathbf{Y}$ и $\mathbf{Z}$ — независимые случайные величины, имеющие одно и то же геометрическое распределение $\{q^kp\}$. Найти а) $\mathbf{P}\{\mathbf{X}=\mathbf{Y}\}$; б) $\mathbf{P}\{\mathbf{X}\geqslant 2\mathbf{Y}\}$; в) $\mathbf{P}\{\mathbf{X}+\mathbf{Y}\leqslant\mathbf{Z}\}$.

5. *Продолжение.* Пусть $\mathbf{U}$ наименьшая из величин $\mathbf{X}$ и $\mathbf{Y}$. Положим $\mathbf{V}=\mathbf{X}-\mathbf{Y}$. Показать, что $\mathbf{U}$ и $\mathbf{V}$ независимы [1]).

6. Пусть $\mathbf{X}_1$ и $\mathbf{X}_2$ — независимые случайные величины, имеющие распределения Пуассона $\{p(k;\ \lambda_1)\}$ и $\{p(k;\ \lambda_2)\}$.

а) Доказать, что $\mathbf{X}_1+\mathbf{X}_2$ имеет распределение Пуассона $\{p(k;\ \lambda_1+\lambda_2)\}$.

б) Доказать, что *условное распределение* $\mathbf{X}_1$ *при заданной сумме* $\mathbf{X}_1+\mathbf{X}_2$ *есть биномиальное распределение*, а именно что

$$\mathbf{P}\{\mathbf{X}_1=k \mid \mathbf{X}_1+\mathbf{X}_2=n\}=b(k;\ n,\ \lambda_1/(\lambda_1+\lambda_2)). \tag{9.1}$$

7. Пусть $\mathbf{X}_1$ и $\mathbf{X}_2$ независимы и имеют одно и то же геометрическое распределение $\{q^k p\}$ (как в задаче 4). Не вычисляя, показать, что *условное распределение* $\mathbf{X}_1$ *при заданной сумме* $\mathbf{X}_1+\mathbf{X}_2$ *равномерно*, т. е. что

$$\mathbf{P}\{\mathbf{X}_1=k \mid \mathbf{X}_1+\mathbf{X}_2=n\}=1/(n+1),\quad k=0,\ \ldots,\ n. \tag{9.2}$$

8. Пусть $\mathbf{X}_1,\ \mathbf{X}_2,\ \ldots,\ \mathbf{X}_r$ — взаимно независимые случайные величины, каждая из которых имеет *равномерное распределение* $\mathbf{P}\{\mathbf{X}_j=k\}=1/N$ для $k=1,\ 2,\ \ldots,\ N$. Пусть $\mathbf{U}_n$ — наименьшая из величин $\mathbf{X}_1,\ \ldots,\ \mathbf{X}_n$, а $\mathbf{V}_n$ — наибольшая из них. Найти распределения $\mathbf{U}_n$ и $\mathbf{V}_n$. Как связана эта задача с задачей об оценке (пример 3, д))?

9. *Дополнение к задаче об оценке* (пример 3, д)). а) Найти совместное распределение максимального и минимального из наблюденных значений. Особо разобрать случай $n=2$. (*Указание.* Сначала вычислить $\mathbf{P}\{\mathbf{X}\leqslant r,\ \mathbf{Y}\geqslant s\}$.)

б) Найти условную вероятность того, что первые два наблюдения дали j и k, если известно, что $\mathbf{X}=r$.

в) Найти $\mathbf{E}(\mathbf{X}^2)$ и затем асимптотическое выражение для $\mathrm{Var}(\mathbf{X})$ при $N\longrightarrow\infty$ (при фиксированном n).

10. *Имитация симметричной монеты.* Для данной несимметричной монеты, такой, что вероятность выпадения герба равна α, имитируем правильную монету следующим образом. Бросаем несимметричную монету дважды. Считаем комбинацию $ГР$ за успех, а $РГ$ — за неудачу; если ни одно из этих событий не произошло, то повторяем бросания до осуществления любого из них. а) Показать, что эта модель приводит к испытаниям Бернулли с $p=1/2$. б) Найти распределение и математическое ожидание числа бросаний до появления одного из событий $ГР$ или $РГ$.

11. *Задача Банаха о спичечных коробках* (пример гл. VI, 8, а)). Показать, что математическое ожидание распределения $\{u_r\}$ определяется по формуле $\mu=(2N+1)u_0-1$. Используя формулу Стирлинга, доказать, что оно приближенно равно $2\sqrt{N/\pi}-1$. (При $N=50$ среднее примерно равно 7,04.)

Указание. Начать с соотношения

$$(N-r)\,u_r=(1/2)\,(2N+1)\,u_{r+1}-(1/2)\,(r+1)\,u_{r+1}.$$

Использовать тот факт [2]), что $\sum u_r=1$.

12. *Выборочный контроль.* Предположим, что изделия, каждое из которых с вероятностью p стандартно, подвергаются контролю с вероятностью p' каждое. Имеем четыре класса изделий, а именно «стандартные и проверенные», «стандартные, но не проверенные» и т. д. с соответствующими вероятностями $pp',\ pq',\ p'q,\ qq'$, где $q=1-p,\ q'=1-p'$. Итак, мы имеем дело со сложными испытаниями Бернулли [см. пример гл. VI, 9, в)]. Пусть $\mathbf{N}$ — число изделий, прошедших стол контроля (как проверенных, так и непроверенных), прежде чем обнаружено первое бракованное изделие, и пусть $\mathbf{K}$ — число

---

[1]) Геометрическое распределение является единственным определенным на целых числах распределением вероятностей, для которого это утверждение верно. См. Ferguson T. S., A characterization of the geometric distribution, Amer. Math. Monthly, 72 (1965), 256—260.

[2]) Этот факт аналитически не очевиден; его можно доказать индукцией по N.

необнаруженных бракованных изделий среди них. Найти совместное распределение $\mathbf{N}$ и $\mathbf{K}$ и их маргинальные распределения.

13. *Продолжение.* Найти $\mathbf{E}(\mathbf{K}/(\mathbf{N}+1))$ и $\mathrm{Cov}(\mathbf{K}, \mathbf{N})$. [При промышленном производстве обнаруженное бракованное изделие заменяется стандартным, так что $\mathbf{K}/(\mathbf{N}+1)$ представляет собой долю брака в партии и является мерой ее качества. Заметьте, что $\mathbf{E}(\mathbf{K}/(\mathbf{N}+1))$ не равно $\mathbf{E}(\mathbf{K})/\mathbf{E}(\mathbf{N}+1)$.]

14. Пусть $\mathbf{X}$ — длина серии (успехов или неудач), начавшейся в первом испытании, в последовательности испытаний Бернулли. а) Найти распределение $\mathbf{X}$, $\mathbf{E}(\mathbf{X})$, $\mathrm{Var}(\mathbf{X})$. б) Пусть $\mathbf{Y}$ — длина *второй* серии. Найти распределение $\mathbf{Y}$, $\mathbf{E}(\mathbf{Y})$, $\mathrm{Var}(\mathbf{Y})$ и совместное распределение $\mathbf{X}$, $\mathbf{Y}$.

15. Пусть $\mathbf{X}$ и $\mathbf{Y}$ имеют одно и то же отрицательное биномиальное распределение. Найти условную вероятность $\mathbf{P}\{\mathbf{X}=j \mid \mathbf{X}+\mathbf{Y}=k\}$ и показать, что равенство (12.16) гл. II теперь очевидно без всяких вычислений[1]).

16. Доказать, что если две случайные величины $\mathbf{X}$ и $\mathbf{Y}$ принимают только два значения каждая и $\mathrm{Cov}(\mathbf{X}, \mathbf{Y})=0$, то $\mathbf{X}$ и $\mathbf{Y}$ независимы.

---

17. *Дни рождения.* Для группы из n человек найти математическое ожидание числа дней в году, на каждый из которых приходятся дни рождения ровно k человек. (Предполагается, что в году 365 дней и все размещения дней рождения равновероятны.)

18. *Продолжение.* Найти математическое ожидание числа дней в году, являющихся днем рождения по меньшей мере двух человек. Как велико должно быть n, чтобы это среднее было больше 1?

19. Человек, имеющий n ключей, хочет отпереть свою дверь, испытывая ключи независимо один от другого и в случайном порядке. Найти математическое ожидание и дисперсию числа испытаний, а) если неподошедшие ключи не исключаются из дальнейших испытаний, б) если они исключаются. (Предполагается, что к двери подходит только один ключ. Точное распределение приведено в гл. II, 7, но для решения этой задачи оно не требуется.)

20. Пусть $(\mathbf{X}, \mathbf{Y})$ — случайные величины, совместное распределение которых является триномиальным, т. е. определяется формулой (1.8). Найти $\mathbf{E}(\mathbf{X})$, $\mathrm{Var}(\mathbf{X})$ и $\mathrm{Cov}(\mathbf{X}, \mathbf{Y})$ а) непосредственным вычислением, б) представляя $\mathbf{X}$ и $\mathbf{Y}$ как суммы n случайных величин и используя методы § 5.

21. Найти ковариацию числа выпадений единицы и шестерки при n бросаниях кости.

22. В задаче об отлове животных (задача 24 гл. VI, 10) доказать, что математическое ожидание числа животных, пойманных при ν-м отлове, равно $nqp^{\nu-1}$.

23. Пусть $\mathbf{X}$ имеет *геометрическое* распределение $\mathbf{P}\{\mathbf{X}=k\}=q^k p$, где $k=0, 1, \ldots$. Показать, что $\mathrm{Var}(\mathbf{X})=qp^{-2}$. Вывести отсюда, что дисперсия *отрицательного* биномиального распределения $\{f(k; r, p)\}$ равна rqp^{-2} при условии, что r — положительное целое число. Непосредственным вычислением доказать, что это утверждение остается справедливым при всех $r>0$.

24. В *задаче о времени ожидания* (пример 3, г)) доказать, что

$$\mathrm{Var}(\mathbf{S}_r)=N\left\{\frac{1}{(N-1)^2}+\frac{2}{(N-2)^2}+\ldots+\frac{r-1}{(N-r+1)^2}\right\}.$$

Проверить, что $N^{-2}\mathbf{E}(\mathbf{S}_N)\sim\sum k^{-2}$. (Сумма этого ряда равна $\pi^2/6$.) *Указание.* Использовать выражение для дисперсии геометрического распределения, найденное в предыдущей задаче.

---

[1]) Этот вывод допускает обобщения на более чем две случайные величины. Он принадлежит Wisniewski T. K. M., Amer. Statistician, 20 (1966), 25.

25. *Продолжение.* Пусть $\mathbf{Y}_r$ — число извлечений, необходимых для получения заранее намеченных r элементов (а не любых r различных элементов, как в § 3). Найти $\mathrm{E}(\mathbf{Y}_r)$ и $\mathrm{Var}(\mathbf{Y}_r)$. (*Замечание.* Точное распределение $\mathbf{Y}_r$ найдено в задаче 12 гл. II, 11, но для решения данной задачи оно не требуется.)

26. *Задача об анализе крови* [1]). Большому числу N людей нужно сделать анализ крови. Это можно организовать двумя способами. (i) Кровь каждого человека исследуется отдельно. В этом случае потребуется N анализов. (ii). Кровь k человек смешивается, и анализируется смесь. Если результат анализа *отрицателен*, то одного этого анализа достаточно для k человек. Если же он *положителен*, то кровь каждого из k человек нужно исследовать отдельно, и для k человек всего потребуется $k+1$ анализов.

Предположим, что вероятность p положительного результата анализа одна и та же для всех людей и что результаты анализов для различных людей стохастически независимы.

а) Найти вероятность того, что анализ смешанной крови k человек даст положительный результат.

б) Чему равно математическое ожидание числа $\mathbf{X}$ анализов при втором способе исследования?

в) Найти уравнение для такого значения k, при котором минимально математическое ожидание числа анализов при втором способе исследования. (Не пытайтесь решить это уравнение.)

г) Показать, что это значение k близко к $1/\sqrt{p}$ и поэтому минимум математического ожидания числа анализов примерно равен $2N\sqrt{p}$. (Это замечание принадлежит М. С. Раффу.)

27. *Структура выборки.* Генеральная совокупность состоит из r классов, количества элементов в которых относятся как $p_1:p_2:\ldots:p_r$. Производится случайная выборка объема n с возвращением. Найти математическое ожидание числа классов, *не* представленных в этой выборке.

28. Пусть $\mathbf{X}$ — число серий альф при случайном размещении r_1 альф и r_2 бет. Распределение $\mathbf{X}$ дается в задаче 23 гл. II, 11. Найти $\mathrm{E}(\mathbf{X})$ и $\mathrm{Var}(\mathbf{X})$.

29. Пусть в *урновой схеме Пойа* (пример гл. V, 2, в)) $\mathbf{X}_n$ равно единице или нулю в зависимости от того, какой шар извлечен при n-м испытании — черный или красный. Доказать, что $\rho(\mathbf{X}_n, \mathbf{X}_m) = c/(b+r+c)$ при $n \neq m$.

30. *Продолжение.* Пусть $\mathbf{S}_n$ — общее число черных шаров, извлеченных при первых n испытаниях (т. е. $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$). Найти $\mathrm{E}(\mathbf{S}_n)$ и $\mathrm{Var}(\mathbf{S}_n)$. Проверить результат с помощью реккурентной формулы из задачи 22 гл. V, 8. *Указание.* Использовать результаты задач 19 и 20 гл. V, 8.

31. *Выбор по группам.* Город разбит на n кварталов, в n_j из которых насчитывается по x_j жителей ($n_1+n_2+\ldots=n$). Пусть $m=\sum n_j x_j/n$ — среднее число жителей в квартале. Положим $a^2=n^{-1}\sum n_j x_j^2 - m^2$. Выбором без возвращения случайно отбираются r кварталов, и в каждом из них подсчитывается число жителей. Пусть эти числа равны $\mathbf{X}_1, \ldots, \mathbf{X}_r$ соответственно. Показать, что

$$\mathrm{E}(\mathbf{X}_1+\ldots+\mathbf{X}_r)=mr, \qquad \mathrm{Var}(\mathbf{X}_1+\ldots+\mathbf{X}_r)=\frac{a^2r\,(n-r)}{n-1}.$$

(При выборе с возвращением дисперсия была бы больше, а именно a^2r.)

[1]) Эта задача основана на методике, разработанной Р. Дорфманом во время второй мировой войны. При использовании своей методики в армейской практике Дорфман добился экономии на 80%. Когда эта задача появилась в первом издании, она привлекла широкое внимание и привела к различным обобщениям, а также к новым приложениям в биологии и промышленности. Основное улучшение заключается во введении более чем двух этапов. См., например, Sobel M., Groll P. A., Group testing to eliminate efficiently all defectives in a binomial sample, The Bell System Journal, 38 (1959), 1179—1252; Watson G. S., A study of the group screening method, Technometrics, 3 (1961), 371—388; Finucan H. M., The blood-testing problem, Applied Statistics, 13 (1964), 43—50.

32. *Длина случайной цепи* [1]). Цепь на плоскости x, y состоит из n звеньев единичной длины. Угол между двумя последовательными звеньями равен $\pm\alpha$, где α — положительная постоянная. Каждая из возможностей имеет вероятность 1/2, и величины последовательных углов взаимно независимы. Расстояние $\mathbf{L}_n$ от начала до конца цепи есть случайная величина, и мы хотим доказать, что

$$\mathbf{E}(\mathbf{L}_n^2) = n\frac{1+\cos\alpha}{1-\cos\alpha} - 2\cos\alpha\frac{1-\cos^n\alpha}{(1-\cos\alpha)^2}. \tag{9.3}$$

Не ограничивая общности, можно предполагать, что первое звено расположено в положительном направлении оси x. Угол между k-м звеном и положительным направлением оси x есть случайная величина $\mathbf{S}_{k-1}$, где $\mathbf{S}_0 = 0$, $\mathbf{S}_k = \mathbf{S}_{k-1} + \mathbf{X}_k\alpha$, а $\mathbf{X}_k$ — взаимно независимые случайные величины, принимающие значения ± 1 с вероятностями 1/2. Проекциями k-го звена на оси x и y будут соответственно $\cos\mathbf{S}_{k-1}$ и $\sin\mathbf{S}_{k-1}$. Следовательно, для $n \geqslant 1$

$$\mathbf{L}_n^2 = \left(\sum_{k=0}^{n-1}\cos\mathbf{S}_k\right)^2 + \left(\sum_{k=0}^{n-1}\sin\mathbf{S}_k\right)^2. \tag{9.4}$$

Доказать последовательно по индукции, что при $m < n$

$$\mathbf{E}(\cos\mathbf{S}_n) = \cos^n\alpha, \quad \mathbf{E}(\sin\mathbf{S}_n) = 0, \tag{9.5}$$

$$\mathbf{E}((\cos\mathbf{S}_m)\cdot(\cos\mathbf{S}_n)) = \cos^{n-m}\alpha\cdot\mathbf{E}(\cos^2\mathbf{S}_m), \tag{9.6}$$

$$\mathbf{E}((\sin\mathbf{S}_m)\cdot(\sin\mathbf{S}_n)) = \cos^{n-m}\alpha\cdot\mathbf{E}(\sin^2\mathbf{S}_m), \tag{9.7}$$

$$\mathbf{E}(\mathbf{L}_n^2) - \mathbf{E}(\mathbf{L}_{n-1}^2) = 1 + 2\cos\alpha\cdot(1-\cos^{n-1}\alpha)/(1-\cos\alpha) \tag{9.8}$$

(где $\mathbf{L}_0 = 0$) и отсюда получить наконец (9.3).

---

33. Испытания Бернулли проводятся до тех пор, пока число успехов не достигнет r, где r — фиксированное целое число. Пусть $\mathbf{X}$ — число испытаний, которые для этого потребовались. Найти [2]) $\mathbf{E}(r/\mathbf{X})$. (Формула для этого математического ожидания содержит бесконечный ряд, для которого может быть получено конечное выражение.)

34. При случайном размещении r шаров по n ящикам вероятность того, что m ящиков окажутся пустыми, удовлетворяет реккурентной формуле (11.8) гл. II. Пусть m_r — математическое ожидание числа пустых ящиков. *Используя рекуррентную формулу*, доказать, что

$$m_{r+1} = (1-n^{-1})\,m_r$$

и вывести отсюда, что $m_r = n(1-1/n)^r$.

35. Пусть $\mathbf{S}_n$ — число успехов в n испытаниях Бернулли. Доказать, что

$$\mathbf{E}(|\mathbf{S}_n - np|) = 2\nu q b(\nu;\ n,\ p),$$

где ν — целое число, такое, что $np < \nu \leqslant np+1$.

---

[1]) Это двухмерный аналог химической задачи о длине *длинной полимерной молекулы*. Задача иллюстрирует применение к случайным величинам, которые не представляются в виде суммы простых случайных величин.

[2]) Этот пример иллюстрирует влияние *остановки в произвольный момент времени*. Если число испытаний n фиксировано, то отношение числа успехов N к числу испытаний n есть случайная величина со средним p. Часто ошибочно предполагают, что то же самое верно и в нашем примере, где число успехов r фиксировано, а число испытаний зависит от случая. Если $p = 1/2$ и $r = 2$, то $\mathbf{E}(2/\mathbf{X}) = 0{,}614$ вместо 0,5; при $r = 3$ находим $\mathbf{E}(3/\mathbf{X}) = 0{,}579$.

Указание. Величина в левой части равна $2\sum_{k=0}^{\nu-1}(np-k)\binom{n}{k}p^kq^{n-k}$.

Другая возможность: использовать формулу (10.7) гл. VI.

36. Пусть $\{\mathbf{X}_k\}$ — последовательность взаимно независимых одинаково распределенных случайных величин. Предположим, что $\mathbf{X}_k$ принимает только положительные значения и что существуют математические ожидания $\mathbf{E}(\mathbf{X}_k)=a$ и $\mathbf{E}(\mathbf{X}_k^{-1})=b$. Положив $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$, доказать, что $\mathbf{E}(\mathbf{S}_n^{-1})$ конечно и $\mathbf{E}(\mathbf{X}_k\mathbf{S}_n^{-1})=n^{-1}$ при $k=1, 2, \ldots, n$.

37. *Продолжение* [1]. Доказать, что

$$\mathbf{E}(\mathbf{S}_m/\mathbf{S}_n)=m/n, \qquad \text{если } m\leqslant n;$$

$$\mathbf{E}(\mathbf{S}_m/\mathbf{S}_n)=1+(m-n)a\mathbf{E}(\mathbf{S}_n^{-1}), \qquad \text{если } m\geqslant n.$$

38. Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — взаимно независимые одинаково распределенные случайные величины со средним m и дисперсией σ^2. Положив $\overline{\mathbf{X}}=(\mathbf{X}_1+\ldots+\mathbf{X}_n)/n$, доказать, что [2]

$$[1/(n-1)]\,\mathbf{E}\left(\sum_{k=1}^{n}(\mathbf{X}_k-\overline{\mathbf{X}})^2\right)=\sigma^2.$$

39. Пусть $\mathbf{X}_1, \ldots, \mathbf{X}_n$ — взаимно независимые случайные величины, $\mathbf{U}$ — функция от $\mathbf{X}_1, \ldots, \mathbf{X}_k$ и $\mathbf{V}$ — функция от $\mathbf{X}_{k+1}, \ldots, \mathbf{X}_n$. Доказать, что $\mathbf{U}$ и $\mathbf{V}$ — взаимно независимые случайные величины.

---

40. *Обобщенное неравенство Чебышева.* Пусть $\varphi(x)>0$ монотонно возрастает при $x>0$; предположим, что существует $\mathbf{E}(\varphi(|\mathbf{X}|))=M$. Доказать, что

$$\mathbf{P}\{|\mathbf{X}|\geqslant t\}\leqslant M/\varphi(t).$$

41. *Неравенство Шварца.* Для любых двух случайных величин $\mathbf{X}$ и $\mathbf{Y}$ с конечными дисперсиями справедливо неравенство $\mathbf{E}^2(\mathbf{X}\mathbf{Y})\leqslant\mathbf{E}(\mathbf{X}^2)\mathbf{E}(\mathbf{Y}^2)$. Доказать это, используя тот факт, что квадратичная форма $\mathbf{E}((t\mathbf{X}+\mathbf{Y})^2)$ неотрицательна.

---

[1]) Чжун Кайлай заметил, что результаты этой задачи можно получить из результата задачи 36.

[2]) Иначе это можно выразить так: $\sum(\mathbf{X}_k-\overline{\mathbf{X}})^2/(n-1)$ является *несмещенной оценкой для* σ^2.

ЗАКОНЫ БОЛЬШИХ ЧИСЕЛ

§ 1. ОДИНАКОВО РАСПРЕДЕЛЕННЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Предельные теоремы для испытаний Бернулли, установленные в гл. VII и VIII, являются частными случаями общих предельных теорем, которые в данном томе не рассматриваются. Однако для того, чтобы установить новую точку зрения на математическое ожидание случайной величины, мы рассмотрим здесь хотя бы некоторые формулировки закона больших чисел.

Связь между испытаниями Бернулли и теорией случайных величин станет яснее, если мы рассмотрим зависимость числа успехов $\mathbf{S}_n$ от числа испытаний n. При каждом испытании $\mathbf{S}_n$ возрастает на 1 или 0, и мы можем записать это в виде

$$\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n, \tag{1.1}$$

где случайная величина $\mathbf{X}_k$ принимает значение 1 в случае успеха k-го испытания и значение 0 в случае неудачи. Следовательно, $\mathbf{S}_n$ есть сумма n взаимно независимых случайных величин, каждая из которых принимает значение 1 или 0 с вероятностями p и q соответственно. Отсюда остается только один шаг до того, чтобы рассмотреть суммы вида (1.1), где $\mathbf{X}_k$ — взаимно независимые случайные величины с произвольными распределениями. Согласно (слабому) закону больших чисел из гл. VI, 4, для больших n весьма правдоподобно, что среднее число успехов $\mathbf{S}_n/n$ близко к p. Это утверждение является частным случаем следующей теоремы.

Закон больших чисел. *Пусть* $\{\mathbf{X}_k\}$ *— последовательность взаимно независимых одинаково распределенных случайных величин. Если математическое ожидание* $\mu = \mathbf{E}(\mathbf{X}_k)$ *существует, то для любого* $\varepsilon > 0$ *при* $n \to \infty$

$$\mathbf{P}\{|(\mathbf{X}_1 + \ldots + \mathbf{X}_n)/n - \mu| > \varepsilon\} \to 0. \tag{1.2}$$

Иначе говоря, вероятность того, что среднее $\mathbf{S}_n/n$ будет отличаться от математического ожидания меньше чем на произвольно заданное ε, стремится к 1.

В таком общем виде теорема впервые была доказана А. Я. Хинчиным[1]). Предыдущие доказательства проводились при излишнем

[1]) Khintchin A. Comptes rendus de l'Académie des Sciences, Paris, 189 (1929), 477—479. Между прочим, читатель должен иметь в виду предостережение, сделанное в связи с законом больших чисел для испытаний Бернулли в конце § 4 гл. VI.

ограничении, состоявшем в требовании конечности дисперсии [1]) $\mathrm{Var}(\mathbf{X}_k)$. Однако для этого случая существует гораздо более точный результат, обобщающий предельную теорему Муавра — Лапласа для испытаний Бернулли, а именно следующая теорема.

Центральная предельная теорема. *Пусть* $\{\mathbf{X}_k\}$ *— последовательность взаимно независимых одинаково распределенных случайных величин. Предположим, что математическое ожидание* $\mu = \mathbf{E}(\mathbf{X}_k)$ *и дисперсия* $\sigma^2 = \mathrm{Var}(\mathbf{X}_k)$ *существуют, и положим* $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$. *Тогда для любого фиксированного* β

$$\mathbf{P}\{(\mathbf{S}_n - n\mu)/(\sigma\sqrt{n}) < \beta\} \to \mathfrak{N}(\beta). \tag{1.3}$$

Здесь $\mathfrak{N}(x)$ — нормальное распределение, определенное в гл. VII, 1. Эта теорема принадлежит Линдебергу [2]); А. М. Ляпунов и другие авторы доказывали ее раньше при более ограничительных условиях. Следует отметить, что эта теорема является только частным случаем гораздо более общей теоремы, формулировка и доказательство которой приводятся в томе 2. Здесь мы отметим, что утверждение (1.3) сильнее, чем (1.2), так как (1.3) дает оценку для вероятности того, что разность $|n^{-1}\mathbf{S}_n - \mu|$ больше чем $\sigma/\sqrt{n}$. С другой стороны, закон больших чисел (1.2) справедлив даже в том случае, когда случайные величины $\mathbf{X}_k$ не имеют конечной дисперсии, и, следовательно, он применим к более общему случаю, чем центральная предельная теорема. По этой причине мы дадим независимое доказательство закона больших чисел, но сначала проиллюстрируем эти две теоремы примерами.

Примеры. а) Пусть в последовательности независимых бросаний симметричной кости $\mathbf{X}_k$ — число очков, выпавших при k-м бросании. Тогда

$$\mathbf{E}(\mathbf{X}_k) = (1+2+3+4+5+6)/6 = 3{,}5$$

и $\mathrm{Var}(\mathbf{X}_k) = (1^2+2^2+3^2+4^2+5^2+6^2)/6 - (3{,}5)^2 = 35/12$. Закон больших чисел утверждает: правдоподобно, что для больших n среднее число очков $\mathbf{S}_n/n$ будет близко к 3,5. Согласно центральной предельной теореме,

$$\mathbf{P}\{|\mathbf{S}_n - 3{,}5n| < \alpha\sqrt{35n/12}\} \approx \mathfrak{N}(\alpha) - \mathfrak{N}(-\alpha). \tag{1.4}$$

В случае $n = 1000$ и $\alpha = 1$ получаем $\mathbf{P}\{3450 < \mathbf{S}_n < 3550\} \approx 0{,}68$. Для $\alpha = 0{,}6744\ldots$ правая часть в (1.4) равна 1/2, поэтому для $\mathbf{S}_n$ шансы находиться внутри интервала 3500 ± 36 или вне его примерно одинаковы.

---

[1]) А. А. Марков показал, что достаточно существования $\mathbf{E}(|\mathbf{X}_k|^{1+a})$ для некоторого $a > 0$.

[2]) Lindeberg J. W., Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, Mathematische Zeitschrift, 15 (1922), 211—225.

б) *Выборка*. Предположим, что в генеральной совокупности, состоящей из N семей, N_k семей имеют ровно по k детей ($k=0, 1, \ldots$; $\sum N_k = N$). Если семья выбирается случайным образом, то число детей в ней является случайной величиной, принимающей значение ν с вероятностью $p_\nu = N_\nu/N$. При выборе с возвращением выборку объема n можно рассматривать как n независимых случайных величин или «наблюдений» $\mathbf{X}_1, \ldots, \mathbf{X}_n$, имеющих одинаковое распределение; $\mathbf{S}_n/n$ является *средним значением выборки*. Согласно закону больших чисел, среднее значение достаточно большой случайной выборки будет, вероятно, близким к $\mu = \sum \nu p_\nu = \sum \nu N_\nu/N$, т. е. к среднему значению генеральной совокупности. Центральная предельная теорема позволяет оценить вероятную величину расхождения между этими величинами и определить размер выборки, необходимый для надежных оценок. На практике μ и σ^2 неизвестны, однако во многих случаях нетрудно получить предварительную оценку для σ^2, и всегда можно заключить σ^2 в надежные границы. Если необходимо, чтобы с вероятностью не меньшей 0,99 среднее значение выборки $\mathbf{S}_n/n$ отличалось от неизвестного среднего значения генеральной совокупности μ меньше, чем на 1/10, то объем выборки должен быть взят таким, чтобы

$$\mathbf{P}\{|(\mathbf{S}_n - n\mu)/n| < 1/10\} \geqslant 0{,}99. \tag{1.5}$$

Корень уравнения $\mathfrak{N}(x) - \mathfrak{N}(-x) = 0{,}99$ равен $x = 2{,}57\ldots$ и, следовательно, число n должно удовлетворять неравенству $\sqrt{n}/(10\sigma) \geqslant 2{,}57$, или $n \geqslant 660\sigma^2$. Осторожная предварительная оценка для σ^2 дает возможность найти требуемый объем выборки. Аналогичные ситуации встречаются довольно часто. Так, когда экспериментатор определяет среднее из n измерений, он также полагается на закон больших чисел и использует среднее значение выборки в качестве оценки для неизвестного математического ожидания. Надежность оценки может быть выражена только через σ^2, и обычно мы вынуждены использовать довольно грубые оценки для σ^2.

в) *Распределение Пуассона*. В гл. VII, 5 мы показали, что при больших λ распределение Пуассона $\{p(k; \lambda)\}$ можно аппроксимировать нормальным распределением. На самом деле этот факт является непосредственным следствием центральной предельной теоремы. Предположим, что случайные величины $\mathbf{X}_k$ имеют распределение Пуассона $\{p(k; \gamma)\}$. Тогда $\mathbf{S}_n$ имеет распределение Пуассона $\{p(k; n\gamma)\}$ с математическим ожиданием и дисперсией, равными $n\gamma$. Написав λ вместо $n\gamma$, мы приходим к выводу, что при $n \to \infty$

$$\sum_{k < \lambda + \beta\sqrt{\lambda}} e^{-\lambda}\lambda^k/k! \to \mathfrak{N}(\beta), \tag{1.6}$$

где суммирование производится по всем k, меньшим чем $\lambda + \beta\sqrt{\lambda}$. Теперь очевидно, что утверждение (1.6) справедливо и в том случае, когда $\lambda \to \infty$ произвольным образом. Эта теорема используется в теории суммирования расходящихся рядов и представляет общий интерес. Оценки разности выражений, стоящих справа и слева в равенстве (1.6), можно получить из общей теории. ▶

Замечание о случайных величинах, не имеющих математического ожидания

Если математическое ожидание μ не существует, то и закон больших чисел, и центральная предельная теорема становятся бессмысленными, но их можно заменить более общими теоремами, дающими аналогичную информацию. В современной теории случайные величины, не имеющие математических ожиданий, играют важную роль, и многие времена ожидания и возвращения в физике оказываются величинами такого типа. Это справедливо даже для простой игры с бросанием монеты.

Предположим, что n монет подбрасываются поодиночке. Обозначим через $\mathbf{X}_k$ время ожидания до первого момента, когда у k-й монеты число выпавших решеток будет равно числу выпавших гербов. Случайные величины $\mathbf{X}_k$ взаимно независимы и одинаково распределены; каждая случайная величина $\mathbf{X}_k$ принимает только положительные четные значения и $\mathbf{P}\{\mathbf{X}_k=2r\}=f_{2r}$, причем распределение вероятностей $\{f_{2r}\}$ определяется формулой (3.7) гл. III. Сумма $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ имеет такое же распределение, как и время ожидания до n-го момента, когда число выпавших гербов равно числу выпавших решеток. Такое же распределение имеет момент n-го возвращения в начало в симметричном случайном блуждании. Как показано в теореме 4 гл. III, 7, распределение $\mathbf{S}_n$ имеет вид

$$\mathbf{P}\{\mathbf{S}_n < n^2x\} \to 2\left[1-\mathfrak{N}\left(1/\sqrt{x}\right)\right]. \tag{1.7}$$

Итак, мы получили предельную теорему того же типа, что и центральная предельная теорема, с той заметной разницей, что теперь *предельное распределение имеет случайная величина* $\mathbf{S}_n/n^2$, *а не* $\mathbf{S}_n/n$, *как ранее.* С точки зрения физики случайные величины $\mathbf{X}_k$ можно интерпретировать как независимые измерения некоторого параметра, и наша предельная теорема утверждает, что *среднее*

$$(\mathbf{X}_1+\ldots+\mathbf{X}_n)/n$$

возрастает по вероятности линейно вместе с n. Этот парадоксальный результат нельзя отбросить как патологический случай, поскольку оказывается, что $\mathbf{X}_k$ — типичные времена ожидания, получающиеся при рассмотрении многих физических и экономических процессов. Предельная теорема (1.7) является также типичным представителем многих современных предельных теорем для случайных величин, не имеющих математического ожидания [1]).

[1]) Аналоги закона больших чисел для случайных величин, не имеющих математического ожидания, приводятся в § 4 и в задаче 13. Неожиданные следствия утверждения (1.7) подробно обсуждались в гл. III.

§ 2*). ДОКАЗАТЕЛЬСТВО ЗАКОНА БОЛЬШИХ ЧИСЕЛ

Без ограничения общности можно предположить, что $\mu=\mathbf{E}(\mathbf{X}_k)=0$. В противном случае вместо $\mathbf{X}_k$ следует рассматривать $\mathbf{X}_k-\mu$, что привело бы просто к замене обозначений. В частном случае, когда дисперсия $\sigma^2=\operatorname{Var}(\mathbf{X}_k)$ существует, закон больших чисел является тривиальным следствием неравенства Чебышева (6.1), гл. IX, согласно которому

$$\mathbf{P}\{|\mathbf{S}_n|>t\}\leqslant n\sigma^2/t^2. \tag{2.1}$$

При $t=\varepsilon n$ правая часть стремится к 0 при $n\to\infty$, следовательно, утверждение (1.2) справедливо.

Более сложным является случай, когда дисперсия не существует. При этом предположении доказательство основывается на универсальном *методе усечения*, который обычно используется при выводе различных предельных теорем. Пусть δ — фиксированное положительное число, которое мы определим позже. Для каждого n определим следующие n пар случайных величин:

$$\begin{aligned} &\mathbf{U}_k=\mathbf{X}_k, \quad \mathbf{V}_k=0, \qquad \text{если } |\mathbf{X}_k|\leqslant\delta n,\\ &\mathbf{U}_k=0, \quad \mathbf{V}_k=\mathbf{X}_k, \qquad \text{если } |\mathbf{X}_k|>\delta n. \end{aligned} \tag{2.2}$$

Здесь $k=1, \ldots, n$ и следует помнить о зависимости случайных величин $\mathbf{U}_k$ и $\mathbf{V}_k$ от n. Тогда

$$\mathbf{X}_k=\mathbf{U}_k+\mathbf{V}_k. \tag{2.3}$$

Для доказательства закона больших чисел достаточно показать, что для заданного $\varepsilon>0$ постоянная δ может быть выбрана таким образом, что при $n\to\infty$

$$\mathbf{P}\{|\mathbf{U}_1+\ldots+\mathbf{U}_n|>(1/2)\,\varepsilon n\}\to 0 \tag{2.4}$$

и

$$\mathbf{P}\{|\mathbf{V}_1+\ldots+\mathbf{V}_n|>(1/2)\,\varepsilon n\}\to 0. \tag{2.5}$$

Обозначим возможные значения случайных величин $\mathbf{X}_j$ через $x_1, x_2, \ldots$, а вероятности, с которыми $\mathbf{X}_j$ принимают эти значения, через $f(x_j)$. Пусть $a=\mathbf{E}(|\mathbf{X}_j|)$, т. е.

$$a=\sum_j |x_j|\,f(x_j). \tag{2.6}$$

Случайная величина $\mathbf{U}_1$ ограничена величиной δn, и, следовательно,

$$\mathbf{E}(\mathbf{U}_1^2)<a\delta n. \tag{2.7}$$

Так как случайные величины $\mathbf{U}_1, \ldots, \mathbf{U}_n$ взаимно независимы и имеют одинаковые распределения, то

$$\operatorname{Var}(\mathbf{U}_1+\ldots+\mathbf{U}_n)=n\operatorname{Var}(\mathbf{U}_1)\leqslant n\mathbf{E}(\mathbf{U}_1^2)\leqslant a\delta n^2. \tag{2.8}$$

*) Этот параграф может быть опущен при первом чтении.

С другой стороны, из определения случайных величин $\mathbf{U}_k$ следует, что при $n \to \infty$

$$\mathbf{E}(\mathbf{U}_1) \to \mathbf{E}(\mathbf{X}_1)=0. \tag{2.9}$$

Отсюда получаем, что при достаточно большом n имеет место неравенство

$$\mathbf{E}((\mathbf{U}_1+\ldots+\mathbf{U}_n)^2)\leqslant 2a\delta n^2. \tag{2.10}$$

Теперь соотношение (2.4) сразу следует из неравенства Чебышева (6.1) гл. IX, согласно которому

$$\mathbf{P}\{|\mathbf{U}_1+\ldots+\mathbf{U}_n|>(1/2)\varepsilon n\}\leqslant 8a\delta/\varepsilon^2. \tag{2.11}$$

Выбирая δ достаточно малым, мы можем сделать правую часть сколь угодно малой, и поэтому соотношение (2.4) справедливо.

Что касается соотношения (2.5), то заметим, что, согласно основному неравенству (7.6) гл. I,

$$\mathbf{P}\{\mathbf{V}_1+\ldots+\mathbf{V}_n\neq 0\}\leqslant n\mathbf{P}\{\mathbf{V}_1\neq 0\}. \tag{2.12}$$

Для произвольного $\delta>0$ имеем

$$\mathbf{P}\{\mathbf{V}_1\neq 0\}=\mathbf{P}\{|\mathbf{X}_1|>\delta n\} = \sum_{|x_j|>\delta n} f(x_j)\leqslant$$
$$\leqslant[1/(\delta n)]\sum_{|x_j|>\delta n}|x_j|f(x_j). \tag{2.13}$$

Последняя сумма стремится к 0 при $n\to\infty$. Следовательно, выражение, стоящее в левой части неравенства (2.12), также стремится к 0. Этим утверждением, более сильным, чем (2.5), и завершается доказательство теоремы. ►

§ 3. ТЕОРИЯ «БЕЗОБИДНЫХ» ИГР

При дальнейшем анализе смысла закона больших чисел мы будем пользоваться традиционной терминологией игроков, хотя наши рассуждения в равной степени допускают и менее легкомысленные приложения, а два наших основных предположения более реальны в статистике и физике, чем в игорном доме. Во-первых, предположим, что игрок обладает *неограниченным капиталом*, поэтому никакой проигрыш не приведет к окончанию игры. (Отказ от этого предположения приводит к задаче о *разорении* игрока, которая всегда интригует изучающих теорию вероятностей. Эта задача играет важную роль в последовательном анализе Вальда и в теории стохастических процессов, и мы вернемся к ней в гл. XIV.) Во-вторых, предположим, что игрок *не имеет права прервать игру в произвольный момент*; *число испытаний n должно быть фиксировано заранее* и не зависеть от хода игры. (В действительности игрок, обладающий неограниченным капиталом, может дождаться серии удач и в подходящий момент прекратить игру. Такого игрока интересует не вероятное состояние в заданный момент, а только максимальные флуктуации, которые можно считать правдоподобными при большом числе партий и которые описываются законом повторного логарифма, а не законом больших чисел (см. гл. VIII, 5).)

В дальнейшем случайная величина $\mathbf{X}_k$ будет обозначать выигрыш (положительный или отрицательный) игрока в k-й партии. Тогда сумма $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ является общим выигрышем в n независимых партиях. Если за право играть игрок уплачивает (не обязательно положительный) взнос μ' перед каждой игрой, то общий уплаченный взнос в n партиях и *общий чистый выигрыш* соответственно равны $n\mu'$ и $\mathbf{S}_n-n\mu'$. Когда существует математическое ожидание $\mu=\mathbf{E}(\mathbf{X}_k)$, применим закон больших чисел, и поэтому, грубо говоря, при больших n весьма правдоподобно, что разность $\mathbf{S}_n-n\mu$ окажется малой по сравнению с n. Поэтому, если μ' меньше μ, то при больших n игрок, вероятно, будет иметь выигрыш порядка $n(\mu-\mu')$. По тем же соображениям взнос $\mu'>\mu$ практически наверняка приводит к убытку. Короче говоря, случай $\mu'<\mu$ *благоприятен* для игрока, тогда как случай $\mu'>\mu$ *неблагоприятен*.

Заметим, что мы еще ничего не говорили о случае $\mu'=\mu$. В этом случае *единственно* возможный вывод состоит в том, что при достаточно большом n общий выигрыш или проигрыш $\mathbf{S}_n-n\mu$ будет с очень большой вероятностью малым по сравнению с n. Но неизвестно, что вероятнее для $\mathbf{S}_n-n\mu$: оказаться положительным или отрицательным, т. е. неизвестно, будет ли игра выгодной или разорительной. Это не было учтено в классической теории, которая называла взнос $\mu'=\mu$ «безобидной» ценой и игру с таким взносом *«безобидной»*. Столь многообещающее и привлекательное название приводило к многочисленным заблуждениям. Следует понимать, что *«безобидная» игра может быть явно разорительной для игрока.* ►

В игровых и других простых ситуациях, в которых случайная величина $\mathbf{X}_k$ имеет конечную дисперсию, можно оправдать понятие «безобидная игра», но, когда дисперсия бесконечна, употребление этого термина совершенно не оправдано. Нет оснований считать, что общий чистый выигрыш $\mathbf{S}_n-n\mu'$ будет колебаться около нуля. Действительно, *можно привести примеры «безобидных» игр* [1], *в которых вероятность того, что игрок потерпит чистый убыток, стремится к единице.* Закон больших чисел утверждает, что, вероятно, этот убыток будет величиной меньшего порядка, чем n. Однако, кроме этого, ничего более утверждать нельзя. Если a_n — произвольная последовательность, такая, что $a_n/n\to 0$, то можно придумать «безобидную» игру, в которой вероятность того, что в n партиях общий чистый убыток превысит a_n, стремится к единице. В задаче 15 приводится пример игры, в которой игрок практически

[1]) Feller W., Note on the law of large numbers and «fair» games, Ann. Math. Statist., 16 (1945), 301—304.

может быть уверен, что его убыток превысит $n/\log n$. Эта игра является «безобидной», и взнос за участие в каждой партии равен единице. Трудно себе представить, что игрок будет считать игру «безобидной», если он практически уверен в том, что будет нести постоянно увеличивающийся убыток.

Было бы ошибкой считать эти явления неестественными или не имеющими практического значения. Пренебрежение случайными величинами, не имеющими математического ожидания, нанесло большой ущерб приложениям, так как эти случайные величины играют существенную роль даже в простейших случайных процессах. Например, случайное блуждание (или игра с бросанием монеты), обсуждавшееся в гл. III, служит прототипом многих стохастических процессов в физике и экономике. Как было показано в гл. III, время ожидания и время первого возвращения в этом случайном блуждании не имеют математических ожиданий и поэтому подвержены случайным флуктуациям, так что возникает парадокс, не согласующийся с нашей интуицией. Несовершенная интуиция, так же как и многие современные приложения теории вероятностей, находятся под сильным влиянием традиционного недопонимания смысла закона больших чисел и распространенного представления о так называемом законе о среднем. Это наследство классической теории, в которой математический анализ неизбежно переплетался с эмпирическими и метафизическими соображениями и в которой с различными предельными теоремами связывалось нечто таинственное [1]).

Вернемся к «нормальной» ситуации, при которой существует не только математическое ожидание $\mathbf{E}(\mathbf{X}_k)$, но и дисперсия $\operatorname{Var}(\mathbf{X}_k)$. В этом случае закон больших чисел дополняется центральной предельной теоремой, из которой следует, что при «безобидной» игре весьма правдоподобно, что чистый выигрыш $\mathbf{S}_n-n\mu$ в результате продолжительной игры будет иметь величину порядка $\sqrt{n}$ и при больших n он с равными шансами будет положительным или отрицательным. Таким образом, если центральная предельная теорема применима, то термин «безобидная игра» оказывается оправданным, но даже в этом случае мы имеем дело с предельной теоремой, что подчеркивается словами «в результате продолжительной игры».

Чтобы проиллюстрировать это, рассмотрим игровой аппарат, опустив в который доллар игрок может с вероятностью 10^{-6} выиграть 10^6-1 долларов либо потерять опущенный доллар. Здесь мы имеем испытания Бернулли, и игра «безобидна». Сыграв миллион раз, игрок уплатит за это миллион долларов. Он может выиграть 0, 1, 2,... раз. Согласно приближению Пуассона для би-

[1]) Изучающие современную теорию вероятностей, возможно, будут удивлены, узнав, что еще в 1934 г. ведущие специалисты могли сомневаться в возможности сформулировать основные предельные теоремы теории вероятностей чисто аналитически.

номиального распределения, вероятность выиграть ровно k раз с точностью до нескольких десятичных знаков равна $e^{-1}/k!$. Таким образом, с вероятностью 0,368... игрок потеряет миллион и с той же вероятностью он окупит свои расходы; с вероятностью 0,184... он может выиграть один миллион и т. д. Здесь 10^6 испытаний эквивалентны одному-единственному испытанию при игре с выигрышем, имеющим распределение Пуассона. Такая игра может быть осуществлена сравнением двух больших колод карт, как это описано в гл. IV, 4.

На практике бессмысленно ожидать выполнимость закона больших чисел после трех или четырех сравнений. По той же причине не имеет смысла применять закон больших чисел в примере с игровым аппаратом, пока не произведено много миллионов испытаний. К этой же схеме относится страхование от пожара, автомобильных катастроф и т. д.; риску подвергается огромная сумма, но соответствующая вероятность очень мала. Кроме того, страхование происходит только один раз в год, так что число испытаний n никогда не становится большим. Для застрахованного игра будет «небезобидной», но тем не менее экономически выгодной; закон больших чисел здесь ни при чем. Что касается страховой компании, то она имеет дело с большим числом игр; но так как дисперсия велика, то возникают случайные флуктуации. Страховые премии должны быть такими, чтобы предотвратить огромный убыток в отдельные годы, и поэтому компания имеет дело скорее с задачей о разорении, чем с законом больших чисел.

§ 4*). ПЕТЕРБУРГСКАЯ ИГРА

В классической теории понятие математического ожидания не было четко отделено от определения вероятности и в обращении с ним не было достаточной математической строгости. Поэтому изучение случайных величин, не имеющих математических ожиданий, сталкивалось с непреодолимыми трудностями, и даже сравнительно недавние дискуссии кажутся странными тому, кто изучает современную теорию вероятностей. Важность случайных величин, не имеющих математических ожиданий, была подчеркнута в предыдущем параграфе, и вполне естественно привести здесь пример аналога закона больших чисел для этих величин. С этой целью мы рассмотрим известный парадокс петербургской игры [1]).

В петербургской игре каждое испытание состоит в бросании правильной монеты до тех пор, пока не выпадет герб; если это случится при r-м бросании, то игрок получает 2^r долларов. Иначе говоря, мы имеем дело с независимыми случайными величинами, ко-

*) Этот параграф может быть опущен при первом чтении.

[1]) Этот парадокс был рассмотрен Даниилом Бернулли (1700—1782). Отметим, что испытания Бернулли были названы в честь Якова Бернулли.

торые принимают значения $2^1, 2^2, 2^3, \ldots$ с вероятностями $2^{-1}, 2^{-2}, 2^{-3}, \ldots$ соответственно. Математическое ожидание формально определяется суммой $\sum x_r f(x_r)$, в которой $x_r = 2^r$ и $f(x_r) = 2^{-r}$, так что каждое слагаемое равно единице. Таким образом, в этой игре не существует конечного математического ожидания и закон больших чисел неприменим. Ясно, что игра станет менее благоприятной для игрока, если изменить правила и установить, что игрок не получает ничего, если желаемый результат не будет достигнут после N бросаний (т. е. решетка выпадет N раз подряд). В этой измененной игре выигрыш имеет конечное математическое ожидание, равное N, и закон больших чисел применим. Из этого следует, что исходная игра будет «благоприятной» для игрока, даже если он платит взнос, равный N, за право играть в одной партии. Это справедливо для каждого N; однако чем больше N, тем больше должно быть число испытаний, чтобы был вероятен положительный выигрыш, так что бессмысленно говорить о «благоприятной» игре. Классическая теория утверждала, что $\mu' = \infty$ является «безобидным» взносом, но современный студент с трудом поймет туманные рассуждения, приводившие к этому «парадоксу».

Оказывается вполне возможным определить взнос за право участия в петербургской игре таким образом, чтобы она имела все свойства «безобидной» игры в классическом смысле, за исключением того, что этот взнос будет зависеть от номера испытания, вместо того чтобы оставаться постоянным. Переменный взнос неудобен в игорном доме, однако петербургская игра и без того неосуществима вследствие ограниченности имеющихся денежных средств. В случае конечного математического ожидания $\mu = \mathbf{E}(\mathbf{X}_k) > 0$ игра называется «безобидной», если при больших n отношение общего выигрыша $\mathbf{S}_n$ к общему уплаченному взносу e_n будет, вероятно, близким к единице (т. е. если разность $\mathbf{S}_n - e_n$ будет, вероятно, иметь порядок, меньший чем e_n). Если математическое ожидание $\mathbf{E}(\mathbf{X}_k)$ не существует, то нельзя сохранять взнос за право участия в игре постоянным и надо определить e_n иначе. Мы будем говорить, что *игра с общим взносом e_n является безобидной в классическом смысле, если для каждого* $\varepsilon > 0$

$$\mathbf{P}\{|\mathbf{S}_n/e_n - 1| > \varepsilon\} \to 0. \tag{4.1}$$

Это полный аналог закона больших чисел, где $e_n = n\mu'$. Закон больших чисел понимается физиком в том смысле, что среднее из n независимых измерений оказывается близким к μ. В данном примере среднее из n измерений оказывается близким к e_n/n. В тех случаях, когда предельная теорема (4.1) применима, она имеет такое же теоретическое и практическое значение, что и закон больших чисел.

Покажем теперь [1]), что *петербургская игра становится «без-*

[1]) Это частный случай закона больших чисел, из которого легко вывести необходимые и достаточные условия выполнения (4.1); см. Feller W., Acta Scientiarum Litterarum Univ. Szeged, 8 (1937), 191—201.

обидной» в классическом смысле, если положить $e_n = n \operatorname{Log} n$, где логарифм берется по основанию 2, т. е. $2^{\operatorname{Log} n} = n$.

Доказательство. Используем метод усечения из § 2, однако на этот раз определим случайные величины $\mathbf{U}_k$ и $\mathbf{V}_k$ $(k = 1, 2, \ldots, n)$ следующим образом:

$$\begin{aligned} &\mathbf{U}_k = \mathbf{X}_k, \quad \mathbf{V}_k = 0, \quad \text{если } \mathbf{X}_k \leqslant n \operatorname{Log} n, \\ &\mathbf{U}_k = 0, \quad \mathbf{V}_k = \mathbf{X}_k, \quad \text{если } \mathbf{X}_k > n \operatorname{Log} n. \end{aligned} \tag{4.2}$$

Тогда

$$\mathbf{P}\{|e_n^{-1}\mathbf{S}_n - 1| > \varepsilon\} \leqslant \mathbf{P}\{|\mathbf{U}_1 + \ldots + \mathbf{U}_n - e_n| > \varepsilon e_n\} + \mathbf{P}\{\mathbf{V}_1 + \ldots + \mathbf{V}_n \neq 0\}, \tag{4.3}$$

так как событие, стоящее в левой части неравенства, не может произойти до тех пор, пока не произойдет хотя бы одно из событий, стоящих в правой части. Итак,

$$\mathbf{P}\{\mathbf{V}_1 + \ldots + \mathbf{V}_n \neq 0\} \leqslant n\mathbf{P}\{\mathbf{X}_1 > n \operatorname{Log} n\} \leqslant 2/\operatorname{Log} n \to 0. \tag{4.4}$$

Поэтому, чтобы проверить (4.1), достаточно доказать, что

$$\mathbf{P}\{|\mathbf{U}_1 + \ldots + \mathbf{U}_n - n \operatorname{Log} n| > \varepsilon n \operatorname{Log} n\} \to 0. \tag{4.5}$$

Положим $\mu_n = \mathbf{E}(\mathbf{U}_k)$ и $\sigma_n^2 = \operatorname{Var}(\mathbf{U}_k)$. Эти величины зависят от n, но одинаковы для $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_n$. Если r — наибольшее из целых чисел, удовлетворяющих неравенству $2^r \leqslant n \operatorname{Log} n$, то $\mu_n = r$, и, следовательно, для достаточно больших n

$$\operatorname{Log} n < \mu_n \leqslant \operatorname{Log} n + \operatorname{Log} \operatorname{Log} n. \tag{4.6}$$

Аналогично

$$\sigma_n^2 < \mathbf{E}(\mathbf{U}_k^2) = 2 + 2^2 + \ldots + 2^r < 2^{r+1} \leqslant 2n \operatorname{Log} n. \tag{4.7}$$

Так как сумма $\mathbf{U}_1 + \ldots + \mathbf{U}_n$ имеет математическое ожидание $n\mu_n$ и дисперсию $n\sigma_n^2$, из неравенства Чебышева получаем

$$\mathbf{P}\{|\mathbf{U}_1 + \ldots + \mathbf{U}_n - n\mu_n| > \varepsilon n\mu_n\} \leqslant \frac{n\sigma_n^2}{\varepsilon^2 n^2 \mu_n^2} < \frac{2}{\varepsilon^2 \operatorname{Log} n} \to 0. \tag{4.8}$$

Теперь, согласно (4.6), имеем $\mu_n \sim \operatorname{Log} n$, и, следовательно, (4.8) эквивалентно (4.5). ►

§ 5. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ С РАЗЛИЧНЫМИ РАСПРЕДЕЛЕНИЯМИ

До сих пор мы рассматривали только случайные величины $\mathbf{X}_k$, имеющие одинаковые распределения. Эта ситуация соответствует повторению одной и той же игры; однако еще интереснее выяснить, что случится, если вид игры будет изменяться на каждом шаге. Для этого не обязательно представлять себе игорный дом; статистик, применяющий статистические критерии, занят вполне достойным видом «игры», а в его случае распределение случайных величин меняется от одного шага к другому.

Для определенности представим себе, что задана бесконечная последовательность распределений вероятностей, так что при каждом n мы имеем n взаимно независимых случайных величин $\mathbf{X}_1, \ldots, \mathbf{X}_n$ с заданными распределениями. Предположим, что математические ожидания и дисперсии существуют, и положим

$$\mu_k = \mathbf{E}(\mathbf{X}_k), \quad \sigma_k^2 = \operatorname{Var}(\mathbf{X}_k). \tag{5.1}$$

Сумма $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ имеет математическое ожидание m_n и дисперсию s_n^2, определяемые формулами

$$m_n = \mu_1 + \ldots + \mu_n, \quad s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2 \tag{5.2}$$

(см. формулы (2.4) и (5.6) гл. IX). В частном случае одинаковых распределений мы имели $m_n = n\mu$, $s_n^2 = n\sigma^2$.

Говорят, что для последовательности $\{\mathbf{X}_k\}$ *выполняется (слабый) закон больших чисел, если при каждом* $\varepsilon > 0$

$$\mathbf{P}\{|\mathbf{S}_n - m_n|/n > \varepsilon\} \to 0. \tag{5.3}$$

О последовательности $\{\mathbf{X}_k\}$ *говорят, что она удовлетворяет центральной предельной теореме, если при любых фиксированных* $\alpha < \beta$

$$\mathbf{P}\{\alpha < (\mathbf{S}_n - m_n)/s_n < \beta\} \to \mathfrak{N}(\beta) - \mathfrak{N}(\alpha). \tag{5.4}$$

Одна из характерных особенностей теории вероятностей заключается в том, что и закон больших чисел, и центральная предельная теорема выполняются для поразительно широкого класса последовательностей $\{\mathbf{X}_k\}$. В частности, *закон больших чисел справедлив каждый раз, когда случайные величины* $\mathbf{X}_k$ *равномерно ограничены*, т. е. существует такая константа A, что $|\mathbf{X}_k| < A$ при всех k. *Более общее условие, достаточное для того, чтобы выполнялся закон больших чисел, состоит в том, что*

$$s_n/n \to 0. \tag{5.5}$$

Это утверждение является прямым следствием неравенства Чебышева, и доказательство, приведенное в самом начале § 2, здесь полностью применимо. Заметим, однако, что условие (5.5) не является необходимым (см. задачу 14).

Известны различные достаточные условия для центральной предельной теоремы, но все они следуют из *теоремы Линдеберга* [1], *согласно которой центральная предельная теорема выполняется, если при каждом* $\varepsilon > 0$ *усеченные случайные величины* $\mathbf{U}_k$, *определенные формулами*

$$\begin{aligned} \mathbf{U}_k &= \mathbf{X}_k - \mu_k, \quad \text{если } |\mathbf{X}_k - \mu_k| \leq \varepsilon s_n, \\ \mathbf{U}_k &= 0, \quad \text{если } |\mathbf{X}_k - \mu_k| > \varepsilon s_n, \end{aligned} \tag{5.6}$$

[1]) См. примечание 2 на стр. 258.

удовлетворяют условиям $s_n \to \infty$ *и*

$$(1/s_n^2) \sum_{k=1}^{n} \mathbf{E}(\mathbf{U}_k^2) \to 1. \qquad (5.7)$$

Если случайные величины $\mathbf{X}_k$ равномерно ограничены, т. е. если $|\mathbf{X}_k| < A$, то случайные величины $\mathbf{U}_k = \mathbf{X}_k - \mu_k$ для всех n настолько больших, что $s_n > 2A\varepsilon^{-1}$. Тогда левая часть (5.7) равна 1. Поэтому из теоремы Линдеберга следует, что *любая равномерно ограниченная последовательность* $\{\mathbf{X}_k\}$ *взаимно независимых случайных величин удовлетворяет центральной предельной теореме* в предположении, конечно, что $s_n \to \infty$. Позднее было показано [1]), что условие Линдеберга также является и необходимым для справедливости (5.4). Доказательство этого утверждения будет дано в томе 2; там же приводятся оценки разности между величинами, стоящими в правой и левой частях соотношения (5.4).

Мы показали, что в случае, когда случайные величины $\mathbf{X}_k$ имеют одинаковые распределения, центральная предельная теорема сильнее, чем закон больших чисел. Однако в общем случае это не так, и мы увидим, что центральная предельная теорема применима и к последовательностям, не удовлетворяющим закону больших чисел.

Примеры. а) Пусть $\lambda > 0$ фиксировано, и пусть $\mathbf{X}_k = \pm k^\lambda$, причем каждое значение принимается с вероятностью 1/2 (например, ставка при k-м бросании монеты равна $\pm k^\lambda$). Здесь $\mu_k = 0$, $\sigma_k^2 = k^{2\lambda}$ и

$$s_n^2 = 1^{2\lambda} + 2^{2\lambda} + 3^{2\lambda} + \ldots + n^{2\lambda} \sim n^{2\lambda+1}/(2\lambda+1). \qquad (5.8)$$

Условие (5.5) справедливо при $\lambda < 1/2$. Следовательно, закон больших чисел выполняется при $\lambda < 1/2$; покажем, что при $\lambda \geqslant 1/2$ закон больших чисел не выполняется.

При $k = 1, 2, \ldots, n$ имеем $|\mathbf{X}_k| = k^\lambda \leqslant n^\lambda$, поэтому при $n > (2\lambda+1)\varepsilon^{-2}$ усеченные переменные $\mathbf{U}_k$ совпадают с $\mathbf{X}_k$. Таким образом, условие Линдеберга выполняется и

$$\mathbf{P}\{\alpha < \sqrt{(2\lambda+1)/n^{2\lambda+1}}\, \mathbf{S}_n < \beta\} \to \mathfrak{N}(\beta) - \mathfrak{N}(\alpha). \qquad (5.9)$$

Следовательно, вероятнее всего, что сумма $\mathbf{S}_n$ будет иметь порядок роста $n^{\lambda+1/2}$, так что при $\lambda \geqslant 1/2$ закон больших чисел не выполняется. В этом примере *центральная предельная теорема*

[1]) Feller W., Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung, Mathematische Zeitschrift, 40 (1935), 521—559. Получены также обобщения центральной предельной теоремы, применимые к случайным величинам, не имеющим математических ожиданий. Отметим, что здесь мы рассматриваем только независимые случайные величины; для зависимых случайных величин условие Линдеберга не является ни необходимым, ни достаточным.

применима для всех $\lambda > 0$, *а закон больших чисел — только при* $\lambda < 1/2$.

б) Рассмотрим две независимые последовательности по 1000 бросаний монеты (или извлечение по 1000 монет из двух урн) и исследуем *разность* $\mathbf{D}$ чисел выпадения герба. Пусть бросания двух последовательностей занумерованы числами от 1 до 1000 и от 1001 до 2000 соответственно, а 2000 случайных величин определены следующим образом: если при k-м бросании выпала решетка, то $\mathbf{X}_k = 0$, если же выпал герб, то $\mathbf{X}_k = 1$ для $k \leqslant 1000$ и $\mathbf{X}_k = -1$ для $k > 1000$. Тогда $\mathbf{D} = \mathbf{X}_1 + \ldots + \mathbf{X}_{2000}$. Случайные величины $\mathbf{X}_k$ имеют среднее значение $\mu_k = \pm 1/2$ и дисперсию $\sigma_k^2 = 1/4$, поэтому $\mathbf{E}(\mathbf{D}) = 0$ и $\mathrm{Var}(\mathbf{D}) = 500$. Следовательно, вероятность того, что разность $\mathbf{D}$ будет иметь значение в пределах $\pm\sqrt{500}\alpha$, приблизительно равна $\mathfrak{N}(\alpha) - \mathfrak{N}(-\alpha)$, и случайная величина $\mathbf{D}$ сравнима с отклонением $\mathbf{S}_{2000} - 1000$ числа выпадений герба в 2000 бросаний от среднего числа выпадений 1000.

в) Приложения к *теории наследственности* иллюстрируют большое разнообразие выводов, основанных на центральной предельной теореме. В гл. V, 5 изучались свойства организма, которые по существу зависят только от одной пары генов (аллелей). Мы предполагали, что другие свойства (такие, например, как рост), зависят от совместного действия многих пар генов. Для простоты изложения предположим, что для каждой отдельной пары генов существуют три генотипа AA, Aa или aa. Пусть x_1, x_2 и x_3 — соответствующие эффекты этих генотипов. Генотип индивидуума является случайным событием, и вклад отдельной пары генов в конечный результат, рост, есть случайная величина $\mathbf{X}$, принимающая три значения: x_1, x_2, x_3 с некоторыми вероятностями. Рост зависит от совместных вкладов многих таких случайных величин $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, и так как каждый из этих вкладов мал, то в первом приближении можно считать, что рост равен *сумме* $\mathbf{X}_1 + \ldots + \mathbf{X}_n$. Конечно, не *все* $\mathbf{X}_k$ взаимно независимы. Но центральная предельная теорема справедлива также и для широкого класса зависимых случайных величин, и, кроме того, большинство случайных величин $\mathbf{X}_k$, вероятнее всего, можно считать независимыми. Эти рассуждения можно уточнить; они приведены здесь только для того, чтобы показать, как центральная предельная теорема объясняет тот факт, что эмпирические распределения многих биометрических показателей, таких, как рост, близки к нормальному распределению. Эта теория позволяет также предсказывать свойства, которые передаются по наследству, например оценивать средний рост детей, зная рост их родителей. Такие биометрические исследования были начаты Ф. Гальтоном и К. Пирсоном [1]).

[1]) Френсис Гальтон (1822—1911); Карл Пирсон (1857—1936).

§ 6*). ПРИЛОЖЕНИЯ К КОМБИНАТОРНОМУ АНАЛИЗУ

Приведем два примера приложений центральной предельной теоремы к задачам, которые не связаны непосредственно с теорией вероятностей. Оба примера относятся к $n!$ перестановкам элементов $a_1, \ldots, a_n$, причем каждой перестановке приписывается вероятность $1/n!$.

а) **Инверсии.** Говорят, что элемент a_k образует в данной перестановке r инверсий, если он стоит впереди r элементов с меньшими номерами (т. е. элементов, которые предшествуют a_k в натуральном порядке). Например, в перестановке $(a_3a_6a_1a_5a_2a_4)$ элементы a_1 и a_2 не образуют ни одной инверсии, элемент a_3 образует две, a_4 — ни одной, a_5 — две и a_6 — четыре инверсии. В перестановке $(a_6a_5a_4a_3a_2a_1)$ элемент a_k образует $k-1$ инверсий, а общее число инверсий равно 15. Число инверсий $\mathbf{X}_k$, образованных элементом a_k, является случайной величиной, а сумма $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ равна общему числу инверсий. Здесь случайные величины $\mathbf{X}_k$ принимают значения $0, 1, \ldots, k-1$, каждое с вероятностью $1/k$; поэтому

$$\begin{gathered}\mu_k = (k-1)/2,\\ \sigma_k^2 = [1 + 2^2 + \ldots + (k-1)^2]/k - [(k-1)/2]^2 = (k^2-1)/12.\end{gathered} \tag{6.1}$$

Число инверсий, образованных элементом a_k, не зависит от относительного расположения элементов $a_1, a_2, \ldots, a_{k-1}$, и, следовательно, случайные величины $\mathbf{X}_k$ взаимно независимы. Из (6.1) получаем

$$m_n = [1 + 2 + \ldots + (n-1)]/2 = n(n-1)/4 \sim n^2/4 \tag{6.2}$$

и

$$s_n^2 = (1/12)\sum_{k=1}^{n}(k^2-1) = (2n^3 + 3n^2 - 5n)/72 \sim n^3/36. \tag{6.3}$$

При больших n имеем $\varepsilon s_n > n \geqslant \mathbf{U}_k$, и поэтому случайные величины $\mathbf{U}_k$, входящие в условие Линдеберга, совпадают с $\mathbf{X}_k$. Следовательно, применима центральная предельная теорема, и мы приходим к выводу, что *число* $\mathbf{N}_n$ *перестановок, для которых число инверсий лежит внутри интервала* $n^2/4 \pm (\alpha/6)\sqrt{n^3}$, *асимптотически равно* $n!\{\mathfrak{N}(\alpha) - \mathfrak{N}(-\alpha)\}$. В частности, примерно для половины всех перестановок число инверсий лежит внутри интервала $n^2/4 \pm 0{,}11\sqrt{n^3}$.

б) **Циклы.** Каждую перестановку можно разбить на циклы, т. е. группы элементов, переставляемых между собой. Так, в перестановке $(a_3a_6a_1a_5a_2a_4)$ элементы a_1 и a_3 меняются местами и оставшиеся четыре элемента переставляются между собой; эта перестановка содержит два цикла. Если элемент стоит на своем естественном месте, то он образует цикл, так что тождественная перестановка $(a_1,$

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

$a_2, \ldots, a_n$) содержит столько же циклов, сколько и элементов. С другой стороны, циклические перестановки $(a_2, a_3, \ldots, a_n, a_1)$, $(a_3, a_4, \ldots, a_n, a_1, a_2)$ и т. д. содержат по одному циклу каждая. При изучении циклов удобно описывать перестановки, указывая стрелками, на какие места переходят элементы. Например, запись $1 \to 3 \to 4 \to 1$ означает, что элемент a_1 переходит на третье место, элемент a_3 — на четвертое место, a_4 — на первое, и третий шаг, таким образом, завершает цикл. Чтобы продолжить это описание, надо начинать со следующего в естественном порядке элемента, т. е. с a_2. В новых обозначениях перестановку $(a_4, a_8, a_1, a_3, a_2, a_5, a_7, a_6)$ можно записать в виде $1 \to 3 \to 4 \to 1$; $2 \to 5 \to 6 \to 8 \to 2$; $7 \to 7$. Иначе говоря, мы строим перестановку $(a_1, \ldots, a_n)$ при помощи n последовательных выборов. Сначала мы находим место i, занятое a_1, затем — место, занятое a_i, и т. д. На первом, втором, ..., n-м шаге имеется $n, n-1, \ldots, 1$ возможностей, и только одна из них завершает цикл.

Пусть случайная величина $\mathbf{X}_k$ равна 1, если при таком построении цикл завершается на k-м шаге, и равна 0 в противном случае. (В предыдущем примере $\mathbf{X}_3=\mathbf{X}_7=\mathbf{X}_8=1$ и $\mathbf{X}_1=\mathbf{X}_2=\mathbf{X}_4=\mathbf{X}_5=\mathbf{X}_6=0$.) Очевидно, что $\mathbf{X}_1=1$ тогда и только тогда, когда элемент a_1 стоит на первом месте. Из нашего построения следует, что $\mathbf{P}\{\mathbf{X}_k=1\}=1/(n-k+1)$, $\mathbf{P}\{\mathbf{X}_k=0\}=(n-k)/(n-k+1)$ и случайные величины $\mathbf{X}_k$ взаимно независимы [1]). Математические ожидания и дисперсии величин $\mathbf{X}_k$ соответственно равны

$$\mu_k = 1/(n-k+1), \quad \sigma_k^2 = (n-k)/(n-k+1)^2, \tag{6.4}$$

откуда следует, что

$$m_n = 1 + 1/2 + 1/3 + \ldots + 1/n \sim \log n \tag{6.5}$$

и

$$s_n^2 = \sum_{k=1}^{n} (n-k)/(n-k+1)^2 \sim \log n. \tag{6.6}$$

Сумма $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ равна общему числу циклов. *Среднее число циклов равно m_n; число перестановок, для которых число циклов заключено между $\log n + \alpha\sqrt{\log n}$ и $\log n + \beta\sqrt{\log n}$, приблизительно равно $n!\{\mathfrak{N}(\beta) - \mathfrak{N}(\alpha)\}$.* Более точные оценки даются усовершенствованными вариантами центральной предельной теоремы [2]).

[1]) Формально распределение случайной величины $\mathbf{X}_k$ зависит не только от k, но также и от n. Однако достаточно переупорядочить $\mathbf{X}_k$ в обратном порядке, чтобы получить случайные величины с распределением, зависящим только от индекса. (См. также пример гл. XI, 2, д).)

[2]) Много различных асимптотических оценок, относящихся к комбинаторному анализу, получено в работе Гончаров В. Л., Из области комбинаторного анализа, Изв. АН СССР, серия матем., 8 (1944), 3—48. Использованный здесь метод проще, но имеет меньшую область применимости; см. Feller W., The fundamental limit theorems in probability, Bull. Amer. Math. Soc., 51 (1945), 800—832.

§ 7*). УСИЛЕННЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ

Из (слабого) закона больших чисел (5.3) следует, что для каждого достаточно большого n весьма правдоподобно, что отклонение $|S_n - m_n|$ окажется малым по сравнению с n. В связи с испытаниями Бернулли уже было указано (гл. VIII), что из этого не следует, что $|S_n - m_n|/n$ остается малым для всех больших n; может случиться, что закон больших чисел применим, но $|S_n - m_n|/n$ продолжает колебаться между конечными или бесконечными границами. Закон больших чисел позволяет утверждать только то, что большие значения $|S_n - m_n|/n$ появляются очень редко.

Будем говорить, что последовательность $\mathbf{X}_k$ *удовлетворяет усиленному закону больших чисел, если каждой паре* $\varepsilon>0$, $\delta>0$ *соответствует* N, *для которого с вероятностью* $1-\delta$ *или большей для любого* $r>0$ *выполняются все* $r+1$ *неравенств*

$$|\mathbf{S}_n - m_n|/n < \varepsilon, \quad n = N,\ N+1,\ \ldots,\ N+r. \tag{7.1}$$

Грубо говоря, (7.1) означает, что с вероятностью, близкой к единице, $|\mathbf{S}_n - m_n|/n$ *остается* малым[1]) при всех $n > N$.

Критерий Колмогорова. *Сходимость ряда*

$$\sum \sigma_k^2/k^2 \tag{7.2}$$

является достаточным условием для того, чтобы усиленный закон больших чисел был применим к последовательности взаимно независимых случайных величин $\mathbf{X}_k$ *с дисперсией* σ_k^2.

Доказательство. Пусть событие A_ν состоит в том, что неравенство (7.1) *не* выполняется хотя бы при одном n из интервала $2^{\nu-1} < n \leqslant 2^\nu$. Очевидно достаточно доказать, что при достаточно большом ν и при всех r

$$\mathbf{P}\{A_\nu\} + \mathbf{P}\{A_{\nu+1}\} + \ldots + \mathbf{P}\{A_{\nu+r}\} < \delta,$$

т. е. доказать сходимость ряда $\sum \mathbf{P}\{A_\nu\}$. Но если событие A_ν произошло, то это означает, что при некотором n из интервала $2^{\nu-1} < n \leqslant 2^\nu$ выполняется неравенство

$$|\mathbf{S}_n - m_n| \geqslant \varepsilon \cdot 2^{\nu-1}, \tag{7.3}$$

и по неравенству Колмогорова (гл. IX, 7) имеем

$$\mathbf{P}\{A_\nu\} \leqslant 4\varepsilon^{-2} \cdot s_{2^\nu}^2 \cdot 2^{-2\nu}. \tag{7.4}$$

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

[1]) В общей теории вводится пространство элементарных событий, соответствующее бесконечной последовательности $\{\mathbf{X}_k\}$. Усиленный закон больших чисел утверждает тогда, что $|\mathbf{S}_n - m_n|/n$ стремится к нулю с вероятностью единица. В терминологии теории функций действительного переменного усиленный закон больших чисел означает сходимость почти всюду, а слабый закон больших чисел эквивалентен сходимости по мере.

Поэтому

$$\sum_{\nu=1}^{\infty} \mathbf{P}\{A_\nu\} \leqslant 4\varepsilon^{-2} \sum_{\nu=1}^{\infty} 2^{-2\nu} \sum_{k=1}^{2^\nu} \sigma_k^2 = 4\varepsilon^{-2} \sum_{k=1}^{\infty} \sigma_k^2 \sum_{2^\nu \geqslant k} 2^{-2\nu} \leqslant$$
$$\leqslant 8\varepsilon^{-2} \sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2}, \tag{7.5}$$

чем и завершается доказательство. ▶

Как пример типичного приложения критерия Колмогорова докажем следующую теорему.

Теорема. *Если взаимно независимые случайные величины* $\mathbf{X}_k$ *имеют одинаковые распределения вероятностей* $\{f(x_j)\}$ *с конечным математическим ожиданием* $\mu = \mathbf{E}(\mathbf{X}_k)$, *то последовательность* $\{\mathbf{X}_k\}$ *удовлетворяет усиленному закону больших чисел.*

Данная теорема, конечно, сильнее, чем слабый закон больших чисел из § 1. Эти две теоремы приводятся отдельно, так как их доказательства интересны с методологической точки зрения. Относительно обратных утверждений см. задачи 17 и 18.

Доказательство. Снова используем метод усечения. Определим две новые последовательности случайных величин, полагая

$$\begin{array}{lll} \mathbf{U}_k = \mathbf{X}_k, & \mathbf{V}_k = 0, & \text{если } |\mathbf{X}_k| < k, \\ \mathbf{U}_k = 0, & \mathbf{V}_k = \mathbf{X}_k, & \text{если } |\mathbf{X}_k| \geqslant k. \end{array} \tag{7.6}$$

Случайные величины $\mathbf{U}_k$ взаимно независимы, и мы покажем, что они удовлетворяют критерию Колмогорова. Если $\sigma_k^2 = \operatorname{Var}(\mathbf{U}_k)$, то

$$\sigma_k^2 \leqslant \mathbf{E}(\mathbf{U}_k^2) = \sum_{|x_j| < k} x_j^2 f(x_j). \tag{7.7}$$

Для краткости положим

$$a_\nu = \sum_{\nu-1 \leqslant |x_j| < \nu} |x_j| f(x_j). \tag{7.8}$$

Так как математические ожидания $\mathbf{E}(\mathbf{X}_k)$ существуют, ряд $\sum a_\nu$ сходится. Кроме того, из (7.7) следует, что

$$\sigma_k^2 \leqslant a_1 + 2a_2 + 3a_3 + \ldots + ka_k \tag{7.9}$$

и

$$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} \leqslant \sum_{k=1}^{\infty} \frac{1}{k^2} \sum_{\nu=1}^{k} \nu a_\nu = \sum_{\nu=1}^{\infty} \nu a_\nu \sum_{k=\nu}^{\infty} \frac{1}{k^2} < 2 \sum_{\nu=1}^{\infty} a_\nu < \infty. \tag{7.10}$$

Таким образом, для последовательности $\{\mathbf{U}_k\}$ ряд (7.2) сходится. Поскольку

$$\mathbf{E}(\mathbf{U}_k) = \mu_k = \sum_{|x_j| < k} x_j f(x_j), \tag{7.11}$$

то $\mu_k \to \mu$ и, следовательно, $(\mu_1 + \ldots + \mu_n)/n \to \mu$. Поэтому, применяя к последовательности $\{\mathbf{U}_k\}$ усиленный закон больших чисел, убеждаемся, что неравенство

$$\left| n^{-1} \sum_{k=1}^{n} \mathbf{U}_k - \mu \right| < \varepsilon \tag{7.12}$$

выполняется с вероятностью, не меньшей чем $1-\delta$ при *всех* $n > N$, если N выбрано достаточно большим. Остается доказать, что аналогичное утверждение будет справедливо, если $\mathbf{U}_k$ заменить на $\mathbf{X}_k$. Для этого, очевидно, достаточно показать, что N можно выбрать столь большим, что событие $\mathbf{U}_k = \mathbf{X}_k$ для всех $k > N$ имеет вероятность сколь угодно близкую к единице. Это в свою очередь сводится к утверждению, состоящему в том, что с вероятностью единица только конечное число случайных величин $\mathbf{V}_k$ отлично от нуля. Согласно первой лемме Бореля — Кантелли из гл. VIII, 3, для этого нужно доказать сходимость ряда $\sum \mathbf{P}\{\mathbf{V}_k \neq 0\}$. Легко видеть, что

$$\mathbf{P}\{\mathbf{V}_n \neq 0\} = \sum_{|x_j| \geq n} f(x_j) \leqslant \frac{a_{n+1}}{n} + \frac{a_{n+2}}{n+1} + \frac{a_{n+3}}{n+2} + \ldots, \tag{7.13}$$

и, следовательно,

$$\sum \mathbf{P}\{\mathbf{V}_n \neq 0\} \leqslant \sum_{n=1}^{\infty} \sum_{\nu=n}^{\infty} \frac{a_{\nu+1}}{\nu} = \sum_{\nu=1}^{\infty} \frac{a_{\nu+1}}{\nu} \sum_{n=1}^{\nu} 1 = \sum_{\nu} a_{\nu+1} < \infty, \tag{7.14}$$

что и утверждалось.

§ 8. ЗАДАЧИ

1. Доказать, что закон больших чисел применим в примере 5, а) также и при $\lambda \leqslant 0$. Центральная предельная теорема выполняется, если $\lambda \geqslant -1/2$.

2. Установить, будут ли выполнены закон больших чисел и центральная предельная теорема для последовательности взаимно независимых случайных величин $\mathbf{X}_k$ с распределениями, заданными следующим образом ($k \geqslant 1$):
   а) $\mathbf{P}\{\mathbf{X}_k = \pm 2^k\} = 1/2$;
   б) $\mathbf{P}\{\mathbf{X}_k = \pm 2^k\} = 2^{-(2k+1)}$, $\mathbf{P}\{\mathbf{X}_k = 0\} = 1 - 2^{-2k}$;
   в) $\mathbf{P}\{\mathbf{X}_k = \pm k\} = 1/(2\sqrt{k})$, $\mathbf{P}\{\mathbf{X}_k = 0\} = 1 - 1/\sqrt{k}$.

3. *Условие Ляпунова* (1901). Показать, что выполняется условие Линдеберга, если при некотором $\delta > 0$

$$(1/s_n^{2+\delta}) \sum_{k=1}^{n} \mathbf{E}(|\mathbf{X}_k|^{2+\delta}) \to 0.$$

4. Пусть $\mathbf{X}_k$ — взаимно независимые случайные величины, причем $\mathbf{X}_k$ принимает $2k+1$ значений $0, \pm L_k, \pm 2L_k, \ldots, \pm kL_k$ с вероятностью $1/(2k+1)$ каждое. Найти условия для констант L_k, обеспечивающие выполнение закона больших чисел и/или центральной предельной теоремы для последовательности $\{\mathbf{X}_k\}$.

5. Решить ту же задачу, если $\mathbf{X}_k$ принимает значения a_k, $-a_k$ и 0 с вероятностями p_k, p_k, $1-2p_k$ соответственно.

Замечание. В следующих семи задачах рассматривается закон больших чисел для зависимых случайных величин.

6. В задаче 13 гл. V, 8, положим $X_k=1$, если при k-м бросании выпала красная грань, и $X_k=0$—в противном случае. Показать, что закон больших чисел неприменим.

7. Пусть $\{X_k\}$ взаимно независимы и имеют одинаковые распределения с математическим ожиданием μ и конечной дисперсией. Доказать, что если $S_n=X_1+\ldots+X_n$, то закон больших чисел не выполняется для последовательности $\{S_n\}$, но выполняется для последовательности a_nS_n, если $na_n\longrightarrow 0$. *Указание.* Вычислить $\operatorname{Var}(S_1, \ldots, S_n)/n$.

8. Пусть $\{X_k\}$—последовательность случайных величин, такая, что X_k может зависеть только от X_{k-1} и X_{k+1}, но не зависит от всех других X_j. Показать, что закон больших чисел выполняется, если X_k имеют конечные дисперсии.

9. Доказать, что если совместное распределение величин $(X_1, \ldots, X_n)$ определено при каждом n, причем дисперсии ограничены, а ковариации отрицательны, то применим закон больших чисел.

10. *Продолжение.* Заменим условие $\operatorname{Cov}(X_j, X_k)\leqslant 0$ предположением, что $\operatorname{Cov}(X_j, X_k)\longrightarrow 0$ равномерно при $|j-k|\longrightarrow\infty$. Доказать, что выполняется закон больших чисел.

11. Доказать, что если $|S_n|<cn$, а $\operatorname{Var}(S_n)>\alpha n^2$, то закон больших чисел не применим к $\{X_k\}$.

12. В урновой схеме Пойа (пример гл. V, 2, в)) положим, что X_k равно 1 или 0 в зависимости от того, был ли k-й извлеченный шар черным или красным. Тогда S_n—число черных шаров при n извлечениях. Доказать, что закон больших чисел к $\{X_k\}$ не применим. *Указание.* Использовать результат предыдущей задачи и задачи 30 гл. IX, 9.

---

13. Взаимно независимые случайные величины X_k принимают значения $r=2, 3, 4, \ldots$ с вероятностями $p_r=c/(r^2\log r)$, где постоянная c выбрана так, что $\sum p_r=1$. Показать, что выполняется обобщенный закон больших чисел (4.1), если положить $e_n=c\cdot n\log\log n$.

14. Пусть $\{X_n\}$—последовательность взаимно независимых случайных величин, таких, что $X_n=\pm 1$ с вероятностями $(1-2^{-n})/2$ и $X_n=\pm 2^n$ с вероятностями 2^{-n-1}. Доказать, что к $\{X_k\}$ применимы как слабый закон больших чисел, так и усиленный закон больших чисел. (*Замечание.* Это означает, что условие (5.5) не является необходимым.)

15. *Пример разорительной «безобидной» игры.* Пусть возможные значения выигрыша при каждом испытании будут 0, 2, 2^2, 2^3, ...; вероятность того, что выигрыш равен 2^k, равна

$$p_k=1/[2^k k(k+1)], \tag{8.1}$$

а вероятность нулевого выигрыша равна $p_0=1-(p_1+p_2+\ldots)$. Тогда математическое ожидание величины выигрыша равно

$$\mu=\sum 2^k p_k=(1-1/2)+(1/2-1/3)+(1/3-1/4)+\ldots=1. \tag{8.2}$$

Предположим, что игрок при каждом испытании уплачивает за право участия в игре доллар, так что после n испытаний его чистый выигрыш или проигрыш равен S_n-n. Показать, что при каждом $\varepsilon>0$ *вероятность того, что в n испытаниях игрок проиграет более чем $(1-\varepsilon)\,n/\operatorname{Log} n$ долларов, стремится к единице,* причем Log означает логарифм по основанию 2. Иначе говоря, надо доказать, что

$$P\{S_n-n<-(1-\varepsilon)\,n/\operatorname{Log} n\}\longrightarrow 1. \tag{8.3}$$

Указание. Использовать метод усечения из § 4, но в (4.2) заменить границу $n\operatorname{Log} n$ границей $n/\operatorname{Log} n$. Показать, что вероятность того, что $U_k=X_k$

при всех $k \leqslant n$, стремится к единице, и доказать, что

$$\mathbf{P}\{|\mathbf{U}_1 + \ldots + \mathbf{U}_n - n\mathbf{E}(\mathbf{U}_1)| < \varepsilon n/\operatorname{Log} n\} \longrightarrow 1; \tag{8.4}$$

$$1 - 1/\operatorname{Log} n \geqslant \mathbf{E}(\mathbf{U}_1) \geqslant 1 - (1+\varepsilon)/\operatorname{Log} n. \tag{8.5}$$

Подробности см. в статье, указанной в примечании к § 3.

16. Пусть $\{\mathbf{X}_n\}$ — последовательность взаимно независимых одинаково распределенных случайных величин. Предположим, что $\mathbf{X}_n$ не имеет конечного математического ожидания и что A — некоторая положительная постоянная. Доказать, что с вероятностью единица осуществится бесконечно много событий $|\mathbf{X}_n| > An$.

17. *Обращение усиленного закона больших чисел.* Доказать, что в предположениях задачи 16 с вероятностью единица $|\mathbf{S}_n| > An$ для бесконечно многих n.

18. *Обращение критерия Колмогорова.* Доказать, что если ряд $\sum \sigma_k^2/k^2$ расходится, то существует последовательность $\{\mathbf{X}_k\}$ взаимно независимых случайных величин с $\operatorname{Var}\{\mathbf{X}_k\} = \sigma_k^2$, к которой неприменим закон больших чисел. *Указание.* Сначала показать, что сходимость ряда $\sum \mathbf{P}\{|\mathbf{X}_n| > \varepsilon n\}$ является необходимым условием для выполнения усиленного закона больших чисел.

ЦЕЛОЧИСЛЕННЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ. ПРОИЗВОДЯЩИЕ ФУНКЦИИ

§ 1. ОБЩИЕ ПОЛОЖЕНИЯ

Среди дискретных случайных величин особенно важны величины, принимающие только целые значения $k=0, 1, 2, \ldots$. Наиболее общим методом изучения таких случайных величин является метод производящих функций. Как мы убедимся в дальнейшем, этот метод — частный случай метода характеристических функций, который служит основой для решения многих задач теории вероятностей. С более общей точки зрения метод производящих функций относится к области операционных методов, широко применяемых в теории дифференциальных и интегральных уравнений. В теории вероятностей производящие функции применялись со времен Муавра и Лапласа, но в полной мере возможности этого метода использовались редко.

Определение. *Пусть $a_0, a_1, a_2, \ldots$— последовательность действительных чисел. Если ряд*

$$A(s)=a_0+a_1s+a_2s^2+\ldots \tag{1.1}$$

сходится в некотором интервале $-s_0<s<s_0$, то функция $A(s)$ называется производящей функцией последовательности $\{a_j\}$.

Переменная s сама по себе ничего не обозначает. Если последовательность $\{a_j\}$ ограничена, то сравнение с геометрической прогрессией показывает, что ряд (1.1) сходится хотя бы при $|s|<1$.

Примеры. Если $a_j=1$ для всех j, то $A(s)=1/(1-s)$. Производящей функцией последовательности $(0, 0, 1, 1, 1, \ldots)$ будет функция $s^2/(1-s)$. Последовательность $a_j=1/j!$ имеет производящую функцию e^s. При фиксированном n последовательность $a_j=\binom{n}{j}$ имеет производящую функцию $(1+s)^n$. Если $\mathbf{X}$ — число очков, выпавших при бросании правильной кости, то распределение вероятностей случайной величины $\mathbf{X}$ имеет производящую функцию $(s+s^2+s^3+s^4+s^5+s^6)/6$. ►

Пусть случайная величина $\mathbf{X}$ принимает значения $0, 1, 2, \ldots$. Удобно ввести специальные обозначения как для распределения величины $\mathbf{X}$, так и для «хвостов» распределения. Мы будем писать

$$\mathbf{P}\{\mathbf{X}=j\}=p_j,\quad \mathbf{P}\{\mathbf{X}>j\}=q_j; \tag{1.2}$$

тогда

$$q_k=p_{k+1}+p_{k+2}+\ldots,\quad k\geqslant 0. \tag{1.3}$$

Производящими функциями последовательностей $\{p_j\}$ и $\{q_k\}$ будут функции

$$P(s) = p_0 + p_1 s + p_2 s^2 + p_3 s^3 + \ldots, \tag{1.4}$$

$$Q(s) = q_0 + q_1 s + q_2 s^2 + q_3 s^3 + \ldots . \tag{1.5}$$

Так как $P(1) = 1$, *то ряд* (1.4) *сходится абсолютно хотя бы при* $-1 \leqslant s \leqslant 1$. Коэффициенты ряда (1.5) меньше единицы, поэтому *он сходится хотя бы при* $-1 < s < 1$.

Теорема 1. *При* $-1 < s < 1$ *имеем*

$$Q(s) = (1 - P(s))/(1 - s). \tag{1.6}$$

Доказательство. Коэффициент при s^n в разложении $(1-s)\cdot Q(s)$ равен $q_n - q_{n-1} = -p_n$, если $n \geqslant 1$, и равен $q_0 = p_1 + p_2 + \ldots = 1 - p_0$, если $n = 0$. Поэтому $(1-s)\cdot Q(s) = 1 - P(s)$, что и утверждалось. ▶

Рассмотрим теперь производную

$$P'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1}. \tag{1.7}$$

Здесь ряд сходится хотя бы при $-1 < s < 1$. При $s = 1$ правая часть формально сводится к $\sum k p_k = \mathbf{E}(\mathbf{X})$. Если математическое ожидание существует, то производная $P'(s)$ будет непрерывной в замкнутом интервале $-1 \leqslant s \leqslant 1$. Если же ряд $\sum k p_k$ расходится, то $P'(s) \to \infty$ при $s \to 1$. В этом случае мы будем говорить, что $\mathbf{X}$ *имеет бесконечное математическое ожидание*, и писать $P'(1) = \mathbf{E}(\mathbf{X}) = \infty$. (Если все рассматриваемые величины положительны, то употребление символа ∞ не приводит к противоречиям). Применяя теорему о среднем к числителю правой части соотношения (1.6), получаем, что $Q(s) = P'(\sigma)$, где число σ заключено между s и 1. Так как обе функции монотонно возрастают, из этого следует, что $P'(s)$ и $Q(s)$ имеют одинаковые конечные или бесконечные пределы, которые мы будем обозначать через $P'(1)$ или $Q(1)$. Итак, доказана следующая теорема.

Теорема 2. *Математическое ожидание* $\mathbf{E}(\mathbf{X})$ *удовлетворяет соотношениям*

$$\mathbf{E}(\mathbf{X}) = \sum_{j=1}^{\infty} j p_j = \sum_{k=0}^{\infty} q_k, \tag{1.8}$$

или в терминах производящих функций

$$\mathbf{E}(\mathbf{X}) = P'(1) = Q(1). \tag{1.9}$$

Дифференцируя формулу (1.7) и используя соотношение $P'(s) = Q(s) - (1-s)Q'(s)$, тем же способом находим

$$\mathbf{E}(\mathbf{X}(\mathbf{X}-1)) = \sum k(k-1)p_k = P''(1) = 2Q'(1). \tag{1.10}$$

Для того чтобы получить дисперсию $\mathbf{X}$, к этому выражению надо прибавить $\mathrm{E}(\mathbf{X})-\mathrm{E}^2(\mathbf{X})$, что приводит нас к следующей теореме.

Теорема 3. *Справедливо равенство*

$$\mathrm{Var}(\mathbf{X}) = P''(1)+P'(1)-P'^2(1) = \\ = 2Q'(1)+Q(1)-Q^2(1). \tag{1.11}$$

В случае бесконечной дисперсии $P''(s)\to\infty$ при $s\to 1$.

Формулы (1.9) и (1.11) часто дают самый простой способ вычисления $\mathrm{E}(\mathbf{X})$ и $\mathrm{Var}(\mathbf{X})$.

§ 2. СВЕРТКИ

Если случайная величина $\mathbf{X}$ принимает только неотрицательные целые значения, то $s^{\mathbf{X}}$ является новой случайной величиной и производящую функцию распределения величины $\mathbf{X}$ можно компактно записать в виде $\mathrm{E}(s^{\mathbf{X}})$. Если случайные величины $\mathbf{X}$ и $\mathbf{Y}$ независимы, то независимыми будут и случайные величины $s^{\mathbf{X}}$ и $s^{\mathbf{Y}}$; поэтому

$$\mathrm{E}(s^{\mathbf{X}+\mathbf{Y}}) = \mathrm{E}(s^{\mathbf{X}})\,\mathrm{E}(s^{\mathbf{Y}}).$$

Мы приведем другое доказательство этого важного результата, так как оно позволит нам получить полезное обобщение.

Пусть $\mathbf{X}$ и $\mathbf{Y}$ — неотрицательные независимые целочисленные случайные величины с распределениями вероятностей $\mathbf{P}\{\mathbf{X}=j\}=a_j$ и $\mathbf{P}\{\mathbf{Y}=j\}=b_j$. Событие $(\mathbf{X}=j,\ \mathbf{Y}=k)$ имеет вероятность a_jb_k. Сумма $\mathbf{S}=\mathbf{X}+\mathbf{Y}$ есть новая случайная величина, и событие $\mathbf{S}=r$ есть объединение несовместных событий

$$(\mathbf{X}=0,\ \mathbf{Y}=r),\ (\mathbf{X}=1,\ \mathbf{Y}=r-1),\ \ldots,\ (\mathbf{X}=r,\ \mathbf{Y}=0).$$

Поэтому распределение $c_r=\mathbf{P}\{\mathbf{S}=r\}$ задается формулой

$$c_r = a_0b_r + a_1b_{r-1} + a_2b_{r-2} + \ldots + a_{r-1}b_1 + a_rb_0. \tag{2.1}$$

Операция (2.1), сопоставляющая двум последовательностям $\{a_k\}$ и $\{b_k\}$ новую последовательность $\{c_k\}$, встречается так часто, что удобно ввести для нее специальное название и обозначение.

Определение. *Пусть $\{a_k\}$ и $\{b_k\}$ — две произвольные числовые последовательности (не обязательно распределения вероятностей). Новая последовательность $\{c_r\}$, определяемая формулой (2.1), называется сверткой* [1]) *последовательностей $\{a_k\}$ и $\{b_k\}$ и обозначается*

$$\{c_k\}=\{a_k\}*\{b_k\}. \tag{2.2}$$

Пример. а) Если $a_k=b_k=1$ для всех $k\geqslant 0$, то $c_k=k+1$. Если $a_k=k$, $b_k=1$, то $c_k=1+2+\ldots+k=k(k+1)/2$. Наконец, если $a_0=a_1=1/2$, $a_k=0$ для $k\geqslant 2$, то $c_k=(b_k+b_{k-1})/2$ и т. д. ►

[1]) Некоторые авторы предпочитают немецкий термин Faltung; французский термин — composition. [В оригинале «convolution» — основной термин, принятый в англоязычной литературе.— *Перев.*]

Пусть последовательности $\{a_k\}$ и $\{b_k\}$ имеют производящие функции $A(s)=\sum a_k s^k$ и $B(s)=\sum b_k s^k$ соответственно. Произведение $A(s)B(s)$ можно получить почленным перемножением степенных рядов для $A(s)$ и $B(s)$. Собирая члены с одинаковыми степенями s, убеждаемся, что коэффициент c_r при s^r в разложении $A(s)B(s)$ задается формулой (2.1). Таким образом, имеет место следующая теорема.

Теорема. *Если $\{a_k\}$ и $\{b_k\}$ — последовательности с производящими функциями $A(s)$ и $B(s)$, а $\{c_k\}$ — их свертка, то производящая функция $C(s)=\sum c_k s^k$ равна произведению*

$$C(s)=A(s)B(s). \tag{2.3}$$

Если $\mathbf{X}$ *и* $\mathbf{Y}$ *— неотрицательные целочисленные независимые случайные величины с производящими функциями $A(s)$ и $B(s)$, то их сумма* $\mathbf{X}+\mathbf{Y}$ *имеет производящую функцию $A(s)B(s)$.*

Пусть теперь $\{a_k\}$, $\{b_k\}$, $\{c_k\}$, $\{d_k\}$, ... — какие-либо последовательности. Мы можем образовать свертку $\{a_k\}*\{b_k\}$, затем свертку этой новой последовательности и последовательности $\{c_k\}$ и т. д. Производящая функция $\{a_k\}*\{b_k\}*\{c_k\}*\{d_k\}$ равна $A(s)B(s)\times C(s)D(s)$, и из этого следует, что порядок, в котором образуются свертки, безразличен. Например, $\{a_k\}*\{b_k\}*\{c_k\}=\{c_k\}*\{b_k\}*\{a_k\}$ и т. д. *Таким образом, свертка является ассоциативной и коммутативной операцией* (точно так же, как сложение случайных величин).

При изучении сумм независимых случайных величин $\mathbf{X}_n$ особый интерес представляет тот частный случай, когда эти величины имеют одинаковые распределения. *Если $\{a_j\}$ — общее распределение вероятностей величин $\mathbf{X}_n$, то распределение суммы $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ будем обозначать $\{a_j\}^{n*}$.* Таким образом,

$$\{a_j\}^{2*}=\{a_j\}*\{a_j\}, \quad \{a_j\}^{3*}=\{a_j\}^{2*}*\{a_j\}, \ \ldots \tag{2.4}$$

и вообще

$$\{a_j\}^{n*}=\{a_j\}^{(n-1)*}*\{a_j\}. \tag{2.5}$$

Иначе говоря, *$\{a_j\}^{n*}$ есть последовательность чисел с производящей функцией $A^n(s)$.* В частности, $\{a_j\}^{1*}$ совпадает с $\{a_j\}$, а *$\{a_j\}^{0*}$ определяется как последовательность с производящей функцией $A^0(s)=1$, т. е. как последовательность* $(1, 0, 0, 0, \ldots)$.

Примеры. б) *Биномиальное распределение.* Производящая функция биномиального распределения $b(k;\, n,\, p)=\binom{n}{k}p^k q^{n-k}$ равна

$$\sum_{k=0}^{n}\binom{n}{k}(ps)^k q^{n-k}=(q+ps)^n. \tag{2.6}$$

Тот факт, что эта производящая функция есть n-я степень бинома $q+ps$, означает, что $\{b(k;\, n,\, p)\}$ есть распределение суммы $\mathbf{S}_n=$

$= \mathbf{X}_1 + \ldots + \mathbf{X}_n$ n независимых случайных величин с одинаковыми производящими функциями $q+ps$; каждая величина принимает значение нуль с вероятностью q и значение единица с вероятностью p. Таким образом,

$$\{b(k;\ n,\ p)\} = \{b(k;\ 1,\ p\}^{n*}. \tag{2.7}$$

Представление $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ уже несколько раз использовалось нами (см. примеры гл. IX, 3, а) и 5, а)). Предыдущее рассуждение можно обратить и использовать для нового вывода биномиального распределения. Из мультипликативного свойства $(q+ps)^m (q+ps)^n = (q+ps)^{m+n}$ следует равенство

$$\{b(k;\ m,\ p)\} * \{b(k;\ n,\ p)\} = \{b(k;\ m+n,\ p)\}, \tag{2.8}$$

которое эквивалентно формуле (10.4) гл. VI. Дифференцируя производящую функцию $(q+ps)^n$, можно совсем просто получить доказательство того, что $\mathbf{E}(\mathbf{S}_n) = np$ и $\operatorname{Var}(\mathbf{S}_n) = npq$.

в) *Распределение Пуассона.* Производящая функция распределения $p(k;\ \lambda) = e^{-\lambda}\lambda^k/k!$ равна

$$\sum_{k=0}^{\infty} e^{-\lambda} (\lambda s)^k/k! = e^{-\lambda+\lambda s}. \tag{2.9}$$

Отсюда следует равенство

$$\{p(k;\ \lambda)\} * \{p(k;\ \mu)\} = \{p(k;\ \lambda+\mu)\}, \tag{2.10}$$

которое эквивалентно формуле (10.5) гл. VI. Дифференцированием убеждаемся, что и среднее значение, и дисперсия распределения Пуассона равны λ (см. пример гл. IX, 4, в)).

г) *Геометрическое и отрицательное биномиальное распределения.* Пусть случайная величина $\mathbf{X}$ имеет геометрическое распределение

$$\mathbf{P}\{\mathbf{X} = k\} = q^k p, \quad k = 0,\ 1,\ 2,\ \ldots, \tag{2.11}$$

где p и q — положительные постоянные, причем $p+q=1$. Соответствующая производящая функция равна

$$p \sum_{k=0}^{\infty} (qs)^k = p/(1-qs). \tag{2.12}$$

Пользуясь результатами § 1, легко убедиться в том, что $\mathbf{E}(\mathbf{X}) = q/p$ и $\operatorname{Var}(\mathbf{X}) = q/p^2$ в соответствии с результатами, полученными в примере гл. IX, 3, в).

Вероятность того, что в последовательности испытаний Бернулли *первый успех* имеет место после k неудач (т. е. при $(k+1)$-м испытании), равна $q^k p$, так что $\mathbf{X}$ можно интерпретировать как *время ожидания первого успеха.* Строго говоря, такая интерпретация относится к бесконечному пространству элементарных событий, и преимущество формального определения (2.11) и терминологии случайных величин состоит в том, что мы можем не беспокоиться о структуре исходного пространства элементарных событий. То же самое верно и для *времени ожидания r-го успеха*. Если $\mathbf{X}_k$ — число неудач между $(k-1)$-м и k-м успехами, то $\mathbf{S}_r = \mathbf{X}_1 + \ldots + \mathbf{X}_r$ — полное число неудач, предшествующих r-му успеху (а $\mathbf{S}_r + r$ — число испытаний до r-го успеха включительно). Поскольку мы имеем дело с испытаниями Бернулли, то величины $\mathbf{X}_k$ взаимно независимы и имеют одинаковое распределение (2.11), так что можно принять это свойство за *определение* величин $\mathbf{X}_k$. Тогда $\mathbf{S}_r$ имеет производящую функцию

$$[p/(1-qs)]^r, \tag{2.13}$$

и из формулы разложения бинома (8.7) гл. II сразу получаем, что коэффициент при s^k равен

$$f(k;\ r,\ p) = \binom{-r}{k} p^r (-q)^k, \quad k = 0,\ 1,\ 2,\ \ldots. \tag{2.14}$$

Отсюда следует, что $\mathbf{P}\{\mathbf{S}_r = k\} = f(k;\ r,\ p)$ в соответствии с формулой для числа неудач, предшествующих r-му успеху, которая была выведена в гл. VI, 8. Этот результат можно сформулировать иначе, сказав, что *распределение* $\{f(k;\ r,\ p)\}$ *есть r-кратная свертка геометрического распределения*, т. е.

$$\{f(k;\ r,\ p)\} = \{q^k p\}^{r*}. \tag{2.15}$$

До сих пор мы считали r целым, но, как это было доказано в гл. VI, 8, *отрицательное биномиальное распределение* $\{f(k;\ r,\ p)\}$ сохраняет смысл при любом неотрицательном r, не обязательно целом. Производящая функция по-прежнему задается формулой (2.13), и мы видим, что при $r>0$ *математическое ожидание и дисперсия отрицательного биномиального распределения равны rq/p и rq/p^2 и что*

$$\{f(k;\ r_1,\ p)\} * \{f(k;\ r_2,\ p)\} = \{f(k;\ r_1 + r_2,\ p)\}. \tag{2.16}$$

д) *Циклы*. В примере гл. X, 6, б) мы рассматривали число циклов $\mathbf{S}_n$ в случайной перестановке n элементов. Было показано, что эту случайную величину можно представить в виде суммы $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ n независимых величин, таких, что каждая случайная величина $\mathbf{X}_k$ может принимать два значения, 1 и 0, с вероятностями $(n-k+1)^{-1}$ и $(n-k)(n-k+1)^{-1}$ соответственно. Из этого сразу следует, что производящая функция случайной величины $\mathbf{S}_n$ равна произведению

$$\frac{n-1+s}{n} \cdot \frac{n-2+s}{n-1} \cdots \frac{1+s}{2} \cdot \frac{s}{1} = (-1)^n \binom{-s}{n}. \tag{2.17}$$

Коэффициенты этого многочлена определяют распределение вероятностей случайной величины $\mathbf{S}_n$, однако для их явного представления нужно знать числа Стирлинга. Здесь перед нами пример весьма обычной ситуации, когда производящая функция оказывается проще самого распределения вероятностей. Поэтому весьма благоприятно то обстоятельство, что сама производящая функция может дать много полезной информации. ►

§ 3. ВОЗВРАЩЕНИЕ В НАЧАЛО И ВРЕМЕНА ОЖИДАНИЙ В ИСПЫТАНИЯХ БЕРНУЛЛИ

В этом параграфе обсуждается несколько важных задач, представляющих методологический интерес, решение которых иллюстрирует силу и гибкость метода производящих функций. Результаты играют важную роль в теории случайных блужданий и могут рассматриваться как прототип аналогичных результатов в теории диффузий. Они будут получены и другими методами в гл. XIV (см., в частности, § 4 и 9). Для частного случая $p=1/2$ эти результаты были получены комбинаторными методами в гл. III. Сравнение этих методов проясняет суть дела [1].

Рассмотрим схему испытаний Бернулли с вероятностью успеха p. Положим $\mathbf{X}_k=+1$, если k-е испытание привело к успеху, и $\mathbf{X}_k=-1$ в противном случае. Иначе говоря, объектом нашего изучения является последовательность взаимно независимых случайных величин, принимающих значения $+1$ и -1 с соответствующими вероятностями p и q. Это описание является самым простым и наиболее естественным, но приводит к несчетному пространству элементарных событий, поскольку предполагает бесконечные испытания. Фактически мы будем вычислять только некоторые вероятности событий, которые определяются конечным числом испытаний, и поэтому никаких принципиальных трудностей не возникает. Можно говорить о фиксированном числе испытаний N и полагать $N\to\infty$, но это было бы излишней педантичностью, вредной для вероятностной интуиции.

Как обычно, положим

$$\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n, \quad \mathbf{S}_0=0. \tag{3.1}$$

В переводе на образный «игровой» язык это означает, что Петр и Павел играют по единичной ставке, а $\mathbf{S}_n$ — чистый выигрыш Петра после n партий. В терминологии теории случайных блужданий $\mathbf{S}_n$ обозначает положение «частицы», которая передвигается на единицу вправо или влево через одинаковые интервалы времени. Случайное блуждание будет несимметричным, если $p\neq 1/2$.

а. Время ожидания для игры. Событие

$$\mathbf{S}_1\leqslant 0, \ \ldots, \ \mathbf{S}_{n-1}\leqslant 0, \quad \mathbf{S}_n=1 \tag{3.2}$$

в игровых терминах означает, что чистый выигрыш Петра впервые оказался положительным в n-й партии. В терминологии теории случайных блужданий *первое попадание в* $+1$ произошло при n-м испытании; более естественное описание использует терминологию теории физической диффузии и называет (3.2) *первым прохождением* через единицу [2]). Мы ищем вероятность φ_n этого события. Точнее,

---

[1]) Из этого объяснения должно быть ясно, что настоящий параграф приводится в целях иллюстрации, хотя представляет и самостоятельный интерес; в дальнейшем он не используется.

[2]) Или первым достижением единицы; см. примечание 1 на с. 108.—*Прим. перев.*

мы ищем производящую функцию

$$\Phi(s) = \sum_{n=0}^{\infty} \varphi_n s^n, \tag{3.3}$$

где для удобства принято $\varphi_0 = 0$[1]. По определению $\varphi_1 = p$. Если (3.2) выполняется для некоторого $n > 1$, то $S_1 = -1$ и найдется *наименьший* индекс $\nu < n$, такой, что $S_\nu = 0$. Исходы первых n партий в игровой терминологии описываются следующим образом. (1) В первой партии Петр проигрывает условную единицу. (2) Петру потребуется сыграть еще $\nu - 1$ партий, чтобы восстановить исходную ситуацию. (3) Петру потребуется сыграть еще ровно $n - \nu$ партий, чтобы добиться положительного чистого выигрыша. Эти три события зависят от непересекающейся группы партий и поэтому являются взаимно независимыми. Из определения ясно, что события (2) и (3) имеют вероятности $\varphi_{\nu-1}$ и $\varphi_{n-\nu}$ соответственно и что вероятность одновременного осуществления всех трех событий равна произведению $q\varphi_{\nu-1}\varphi_{n-\nu}$. Итак, событие (3.2) осуществляется тогда и только тогда, когда происходят события (1)—(3) при некотором $\nu < n$. Суммируя по всем возможным ν, получаем

$$\varphi_n = q(\varphi_1\varphi_{n-2} + \varphi_2\varphi_{n-3} + \ldots + \varphi_{n-2}\varphi_1). \tag{3.4}$$

Следует помнить, что это соотношение справедливо только при $n > 1$ и что $\varphi_1 = p$, а $\varphi_0 = 0$. Умножая (3.4) на s^n и суммируя по $n = 2, 3, \ldots$, мы получаем в левой части $\Phi(s) - ps$. Величина, стоящая в скобках, представляет собой $(n-1)$-й член свертки $\{\varphi_n\} * \{\varphi_n\}$ и, следовательно, по теореме из § 2 правая часть равна $qs \cdot \Phi^2(s)$. Таким образом, мы видим, что *производящая функция Φ удовлетворяет квадратному уравнению*

$$\Phi(s) - ps = qs\Phi^2(s). \tag{3.5}$$

Один из двух корней этого уравнения неограничен в окрестности $s = 0$, и единственное ограниченное решение имеет вид

$$\Phi(s) = \left(1 - \sqrt{1 - 4pqs^2}\right)/(2qs), \tag{3.6}$$

где $\sqrt{}$ означает положительный корень. Формула разложения бинома (8.7) гл. II позволяет записать коэффициенты в виде

$$\varphi_{2k-1} = \frac{(-1)^{k-1}}{2q}\binom{1/2}{k}(4pq)^k, \quad \varphi_{2k} = 0. \tag{3.7}$$

Таким образом, мы получили явные выражения для вероятностей φ_k, но они имеют лишь второстепенное значение; более поучи-

[1] Как будет показано в дальнейшем, производящую функцию Φ можно получить непосредственно из простых вероятностных рассуждений. Следующий ниже менее изящный вывод приводится только потому, что дает хороший пример того, как нужно обращаться с уравнениями в свертках, которые встречаются в различных контекстах вне теории вероятностей (см., например, задачу 6).

тельно извлечь нужную информацию непосредственно из производящей функции.

Заметим сначала, что сумма $\sum \varphi_n$ равна

$$\Phi(1) = (1 - |p-q|)/(2q), \tag{3.8}$$

и поэтому

$$\sum \varphi_n = \begin{cases} p/q, & \text{если } p < q, \\ 1, & \text{если } p \geqslant q. \end{cases} \tag{3.9}$$

Иначе говоря, *если $p<q$, то вероятность того, что сумма $\mathbf{S}_n$ будет всегда отрицательной, равна $(q-p)/q$*. Если $p \geqslant q$, то эта вероятность равна нулю, поэтому с вероятностью единица сумма $\mathbf{S}_n$ рано или поздно станет положительной. Сколько времени пройдет до этого момента? Простые вычисления показывают, что $\Phi'(1) = (p-q)^{-1}$, если $p>q$, и $\Phi'(1) = \infty$, если $p=q=1/2$. Можно сделать вывод, что *при $p=1/2$ число партий, предшествующих моменту, когда сумма $\mathbf{S}_n$ впервые будет положительной, имеет бесконечное математическое ожидание.*

Имеет смысл прокомментировать на игровом языке этот заслуживающий внимания результат. Из него следует, что в игре с бросанием симметричной монеты Петр теоретически уверен, что рано или поздно добьется положительного чистого выигрыша, но математическое ожидание числа испытаний, необходимых для достижения этой цели, равно бесконечности. Следовательно, игрок с ограниченным капиталом никогда не может быть уверен в том, что он когда-либо добьется чистого положительного выигрыша. Мы вернемся к этому вопросу в гл. XIV в связи с задачей о разорении игрока.

В вероятностных терминах вывод квадратного уравнения (3.5) для Φ можно описать компактнее. Обозначим через $\mathbf{N}$ первый индекс, при котором $\mathbf{S_N} > 0$. Тогда $\mathbf{N}$ будет случайной величиной в несколько обобщенном смысле, а именно эта случайная величина не определена для события $S_n \leqslant 0$ при *всех n*. (В терминологии гл. XIII нам следовало бы назвать $\mathbf{N}$ *дефектной величиной*.) Производящую функцию Φ можно теперь записать в виде $\Phi(s) = \mathbf{E}(s^{\mathbf{N}})$. Если $\mathbf{X}_1 = -1$, то $\mathbf{N} = 1 + \mathbf{N}_1 + \mathbf{N}_2$, где $\mathbf{N}_1$ — число испытаний, необходимых для увеличения частных сумм $\mathbf{S}_k$ от -1 до 0, а $\mathbf{N}_2$ — число последующих испытаний, необходимых для увеличения $\mathbf{S}_k$ от 0 до 1. Эти величины независимы и имеют то же распределение, что и $\mathbf{N}$. Поэтому для условных математических ожиданий случайной величины $s^{\mathbf{N}}$ получаем

$$\mathbf{E}(s^{\mathbf{N}} \mid \mathbf{X}_1 = -1) = \mathbf{E}(s^{1+\mathbf{N}_1+\mathbf{N}_2} \mid \mathbf{X}_1 = -1) = s\Phi^2(s),$$
$$\mathbf{E}(s^{\mathbf{N}} \mid \mathbf{X}_1 = 1) = s.$$

Но

$$\mathbf{E}(s^{\mathbf{N}}) = p\mathbf{E}(s^{\mathbf{N}} \mid \mathbf{X}_1 = 1) + q\mathbf{E}(s^{\mathbf{N}} \mid \mathbf{X}_1 = -1), \tag{3.10}$$

что при $\Phi(s) = \mathbf{E}(s^{\mathbf{N}})$ совпадает с квадратным уравнением (3.5)

6. Возвращение в начало. Равенство сумм числа успехов и неудач имеет место при k-м испытании, если $\mathbf{S}_k=0$. Заимствуя термин из теории диффузии, мы назовем это событие возвращением в состояние равновесия (т. е. в начало). Число испытаний будет случайным, и вероятность возвращения при $2n$-м испытании равна

$$u_{2n}=\binom{2n}{n}p^nq^n=(-1)^n\binom{-1/2}{n}(4pq)^n. \tag{3.11}$$

Из формулы разложения бинома (8.7) гл. II получаем производящую функцию

$$U(s)=\sum_{n=0}^{\infty}u_{2n}s^{2n}=1/\sqrt{1-4pqs^2}. \tag{3.12}$$

Отметим, что $\{u_n\}$ *не* является распределением вероятностей, так как возвращение в начало может происходить неоднократно.

в. Первое возвращение в начало происходит при $2n$-м испытании, если $S_{2n}=0$, но $S_k\neq 0$ при $k=1, \ldots, 2n-1$. Обозначим вероятность этого события через f_{2n}. (Очевидно, $f_{2n-1}=0$.) Рассмотрим отдельно два подсобытия с $X_1=1$ и $X_1=-1$ и обозначим их вероятности через f_{2n}^+ и f_{2n}^- соответственно. Из сказанного в п. «а» ясно, что $f_{2n}^-=q\varphi_{2n-1}$, так как первые $2n-2$ частных сумм $X_2+X_3+\ldots+X_k$ не превышают нуля, но следующая сумма положительна. Поэтому, используя (3.6), получаем

$$F^-(s)=\sum_{n=1}^{\infty}f_{2n}^-s^{2n}=qs\Phi(s)=(1-\sqrt{1-4pqs^2})/2. \tag{3.13}$$

В силу симметричности производящую функцию последовательности $\{f_n^+\}$ можно получить взаимной перестановкой p и q. Из этого следует, что $F^+=F^-$, и поэтому окончательно[1])

$$F(s)=\sum_{n=1}^{\infty}f_ns^n=1-\sqrt{1-4pqs^2}. \tag{3.14}$$

Отсюда можно получить интересные выводы, не используя явного вида коэффициентов f_n. Очевидно, что $F(1)$ означает вероятность того, что рано или поздно происходит возвращение в начало. Итак, $F(1)=1-|p-q|$, и поэтому $|p-q|$ *есть вероятность того, что возвращение в начало не происходит никогда*, т. е. $S_k\neq 0$ для всех $k>0$. Только в симметричном случае $p=1/2$ несомненно происходит возвращение в начало. В этом случае последовательность $\{f_n\}$ представляет распределение вероятностей для времени ожидания первого возвращения. Это время имеет *бесконечное математическое ожидание*.

В симметричном случае $p=1/2$ имеем

$$U(s)=(1-F(s))/(1-s^2). \tag{3.15}$$

Так как U и F являются степенными рядами по s^2, то это соотношение отличается от (1.6) только обозначениями, и по теореме

[1]) Другой вывод можно найти в примере гл. XIII, 4, б).

1 § 1

$$u_{2n} = f_{2n+2} + f_{2n+4} + \dots, \tag{3.16}$$

т. е. *при $p=1/2$ вероятность того, что $S_{2n}=0$, совпадает с вероятностью того, что $2n$ сумм $S_1, \dots, S_{2n}$ отличается от нуля.* Этот результат был получен различными методами в гл. III, 3 и играл основную роль в анализе парадоксальной природы флуктуаций при игре с бросанием монеты.

г. Первые достижения и последующие возвращения. Мы говорим, что *первое достижение точки $r>0$ происходит при n-м испытании, если $S_n = r$, но $S_k < r$ для всех $k<n$.* Вероятность этого события обозначим через $\varphi_n^{(r)}$. Испытания, следующие за первым достижением точки $\nu>0$, с вероятностной точки зрения являются копией всей последовательности, и поэтому число испытаний, следующих за первым достижением точки ν до первого достижения точки $\nu+1$ включительно, имеет такое же распределение $\{\varphi_n\}$, как число испытаний до первого достижения точки 1. Если $p<q$, то сумма φ_n не равна единице, но все же имеет смысл говорить о том, что время ожидания первого достижения является случайной величиной с (возможно дефектным) распределением $\{\varphi_n\}$. Времена ожиданий между последовательными первыми достижениями взаимно независимы, и поэтому полное время ожидания для первого достижения точки r является суммой r независимых случайных величин с общим распределением $\{\varphi_n\}$. *Следовательно, производящая функция последовательности вероятностей первого достижения $\varphi_n^{(r)}$ равна r-й степени* Φ. (Начинающим следует проверить это утверждение непосредственно, получив уравнение свертки для $\varphi_n^{(2)}$, аналогичное (2.4), а затем воспользовавшись индукцией.)

Аналогичные рассуждения справедливы для вероятности $f_n^{(r)}$ того, что r-е *возвращение в начало происходит при n-м испытании. Производящая функция последовательности $\{f_n^{(r)}\}$ равна r-й степени F.* Сравнивая (3.6) и (3.14), сразу же замечаем, что

$$f_n^{(r)} = (2q)^r \varphi_{n-r}^{(r)}. \tag{3.17}$$

В частном случае $p=q=1/2$ этот результат содержится в теореме 4 гл. III,7.

При помощи производящих функций легко получить различные приближения и предельные теоремы, но при этом используются преобразования Лапласа, которые будут рассматриваться только в гл. XIII тома 2. Систематического способа получения явных выражений для $f_n^{(r)}$ из производящей функции F^r не существует, но правильное предположение легко проверить, исходя из вида производящей функции. Исходя из теоремы 4 гл. III, 7, можно предположить, что

$$f_{n2}^{(r)} = \frac{r}{2n-r} \binom{2n-r}{n} 2^r (pq)^n. \tag{3.18}$$

Для проверки этого предположения достаточно заметить, что из тождества

$$F^r(s) = 2F^{r-1}(s) - 4pqs^2F^{r-2}(s)$$

следует рекуррентное соотношение

$$f_{2n}^{(r)} = 2f_{2n}^{(r-1)} - 4pqf_{2n-2}^{(r-2)}, \tag{3.19}$$

которому также удовлетворяет правая часть (3.18). Таким образом, по индукции убеждаемся в справедливости (3.18). По поводу эквивалентных выражений, имеющих, однако, другой вид, см. задачу 13 гл. XIV, 9.

§ 4. РАЗЛОЖЕНИЕ НА ПРОСТЫЕ ДРОБИ

Если известна производящая функция $P(s) = \sum p_k s^k$, то коэффициенты p_k можно найти дифференцированием по очевидной формуле $p_k = P^{(k)}(0)/k!$. На практике не всегда возможно получить явные выражения для p_k; кроме того, эти выражения порой так сложны, что разумное приближение оказывается предпочтительнее. Наиболее распространенный метод получения таких приближений основан на разложении на простые дроби. Из теории функции комплексного переменного известно, что такие разложения возможны для широкого класса функций, но мы ограничимся рассмотрением простого случая *рациональных функций*.

Предположим, что производящая функция имеет вид

$$P(s) = U(s)/V(s), \tag{4.1}$$

где U и V — многочлены, не имеющие общих корней. Допустим для простоты, что степень U меньше степени V, причем степень V равна m. Кроме того, предположим, что уравнение $V(s)=0$ имеет m различных (действительных или мнимых) корней $s_1, \ldots, s_m$. Тогда

$$V(s) = (s-s_1)(s-s_2)\ldots(s-s_m), \tag{4.2}$$

и из алгебры известно, что $P(s)$ можно разложить на *простые дроби*

$$P(s) = \rho_1/(s_1-s) + \rho_2/(s_2-s) + \ldots + \rho_m/(s_m-s), \tag{4.3}$$

где $\rho_1, \rho_2, \ldots, \rho_m$ — постоянные. Чтобы найти ρ_1, умножим (4.3) на s_1-s; при $s \to s_1$ произведение $(s_1-s)P(s)$ стремится к ρ_1. С другой стороны, из (4.1) и (4.2) получаем

$$(s_1-s)P(s) = -U(s)/[(s-s_2)(s-s_3)\ldots(s-s_m)]. \tag{4.4}$$

При $s \to s_1$ числитель в (4.4) стремится к $-U(s_1)$, а знаменатель — к произведению $(s_1-s_2)(s_1-s_3)\ldots(s_1-s_m)$, равному $V'(s_1)$. Таким образом, $\rho_1 = -U(s_1)/V'(s_1)$. Это рассуждение применимо ко всем корням, поэтому при $k \leqslant m$

$$\rho_k = -U(s_k)/V'(s_k). \tag{4.5}$$

Для заданного ρ_k мы можем легко получить точное выражение для коэффициента при s^n в $P(s)$. Напишем

$$\frac{1}{s_k - s} = \frac{1}{s_k} \cdot \frac{1}{1 - s/s_k}. \tag{4.6}$$

При $|s|<|s_k|$ последнюю дробь можно разложить в геометрический ряд

$$\frac{1}{1-s/s_k}=1+\frac{s}{s_k}+\left(\frac{s}{s_k}\right)^2+\left(\frac{s}{s_k}\right)^3+\ldots. \tag{4.7}$$

Подставляя эти выражения в (4.3), получим *коэффициент* p_n *при* s^n в виде

$$p_n=\rho_1/s_1^{n+1}+\rho_2/s_2^{n+1}+\ldots+\rho_m/s_m^{n+1}. \tag{4.8}$$

Таким образом, для получения p_n нужно сначала найти корни знаменателя $s_1, s_2, \ldots, s_m$, а затем по (4.5) определить коэффициенты $\rho_1, \ldots, \rho_m$.

Формула (4.8) дает *точное* выражение для вероятности p_n. Вычислить все m корней обычно бывает затруднительно, и поэтому формула (4.8) представляет преимущественно теоретический интерес. К счастью, даже один член в (4.8) почти всегда обеспечивает удовлетворительное приближение. Действительно, пусть s_1 — *наименьший* по абсолютной величине корень. Тогда первый знаменатель в (4.8) является наименьшим. Ясно, что при возрастании n относительная доля остальных членов убывает и преобладает первое слагаемое. Иначе говоря, *если* s_1 — *наименьший по абсолютной величине корень уравнения* $V(s)=0$, *то при* $n\to\infty$

$$p_n\sim\rho_1/s_1^{n+1} \tag{4.9}$$

(знак $\sim$ означает, что отношение правой и левой частей стремится к единице). Обычно эта формула дает хорошее приближение даже при сравнительно малых значениях n. Главное преимущество формулы (4.9) состоит в том, что в ней требуется вычислить только один корень алгебраического уравнения.

Легко освободиться от тех ограничений, при которых мы вывели асимптотическую формулу (4.9). Во-первых степень числителя в (4.1) может превосходить степень m знаменателя. Пусть $U(s)$ имеет степень $m+r$ ($r\geqslant 0$); с помощью деления $P(s)$ приводится к сумме многочлена степени r и дроби $U_1(s)/V(s)$, в которой степень многочлена $U_1(s)$ меньше m. Этот многочлен влияет только на первые $r+1$ членов распределения $\{p_n\}$, а $U_1\{s\}/V(s)$ можно, как и выше, разложить на простые дроби. Итак, формула (4.9) остается верной. Во-вторых, ограничение, что $V(s)$ имеет только простые корни, также не является необходимым. Из алгебры известно, что любая рациональная функция допускает разложение на простые дроби. Если s_k — двойной корень $V(s)$, то разложение на простые дроби (4.3) будет содержать дополнительный член вида $a/(s-s_k)^2$, и это внесет член вида $a(n+1)s_k^{-(n+2)}$ в точное выражение (4.8) для p_n. Но если s_1 — простой корень, то это не повлияет на асимптотическое разложение (4.9). Для удобства дальнейших ссылок сформулируем этот результат в виде теоремы.

Теорема. *Если наименьший по абсолютной величине корень s_1 знаменателя рациональной функции $P(s)$ является простым корнем, то коэффициент p_n при s^n асимптотически равен $\rho_1 s_1^{-(n+1)}$ где ρ_1 определяется по формуле* (4.5).

Аналогичное асимптотическое представление существует также и в случае, когда s_1 является кратным корнем (см. задачу 25).

Примеры[1]. а) Пусть a_n — вероятность того, что *число успехов в n испытаниях Бернулли четно*. Это событие может произойти в двух случаях: если за неудачей в первом испытании следует четное число успехов или же за успехом в первом испытании следует нечетное число успехов в остальных испытаниях. Поэтому для $n \geqslant 1$

$$a_n = qa_{n-1} + p(1 - a_{n-1}), \quad a_0 = 1. \tag{4.10}$$

Умножая на s^n и суммируя полученные соотношения по $n = 1, 2, \ldots$, получаем следующее уравнение для производящей функции:

$$A(s) - 1 = qsA(s) + ps(1-s)^{-1} - psA(s),$$

или

$$2A(s) = [1-s]^{-1} + [1-(q-p)s]^{-1}.$$

Разлагая правую часть последнего соотношения в геометрическую прогрессию, мы, наконец, получаем a_n в явном виде

$$2a_n = 1 + (q-p)^n, \tag{4.11}$$

который с любой точки зрения предпочтительней очевидного ответа

$$a_n = b(0;\ n,\ p) + b(2;\ n,\ p) + \ldots .$$

б) Пусть q_n — вероятность того, что при n бросаниях правильной монеты не появится ни одна серия из трех последовательных выпадений герба. (Заметим, что $\{q_n\}$ не является распределением вероятностей; если p_n — вероятность того, что первая серия из трех последовательных гербов окончится на n-м испытании, то $\{p_n\}$ есть распределение вероятностей, а q_n представляет его «хвосты»: $q_n = p_{n+1} + p_{n+2} + \ldots$.)

Легко показать, что q_n удовлетворяет рекуррентной формуле

$$q_n = (1/2)\, q_{n-1} + (1/4)\, q_{n-2} + (1/8)\, q_{n-3}, \quad n \geqslant 3. \tag{4.12}$$

Действительно, событие, состоящее в том, что при n испытаниях ни разу не появится последовательность *ГГГ*, может произойти только в том случае, если испытания начинаются с *Р*, *ГР*, или *ГГР*. Вероятности того, что последующие испытания не приведут

---

[1]) Хорошую иллюстрацию применения простых дробей для численных приближений дает теория серий успехов (гл. XIII, 7). Явные выражения для вероятности разорения в гл. XIV, 5 и вероятностей перехода в гл. XVI, 1 также получены методом разложения на простые дроби.

10*

к серии $ГГГ$, равны соответственно q_{n-1}, q_{n-2} и q_{n-3}, и поэтому правая часть (4.12) содержит вероятности трех взаимно исключающих друг друга возможностей, при которых может произойти событие «ни одной серии $ГГГ$».

Очевидно, $q_0=q_1=q_2=1$, и, следовательно, остальные q_n можно последовательно вычислять по формуле (4.12). Чтобы получить производящую функцию $Q(s)=\sum q_n s^n$, умножим обе части равенства (4.12) на s^n и просуммируем по $n\geqslant 3$. Получим

$$Q(s)-1-s-s^2=(1/2)\,s\,\{Q(s)-1-s\}+(1/4)\,s^2\{Q(s)-1\}+ \\ +(1/8)\,s^3 Q(s),$$

или

$$Q(s)=\frac{2s^2+4s+8}{8-4s-2s^2-s^3}. \tag{4.13}$$

Знаменатель имеет действительный корень $s_1=1{,}0873778\ldots$, и два комплексных корня. При $|s|<s_1$ имеем $|4s+2s^2+s^3|<4s_1+2s_1^2+s_1^3=8$, и это же неравенство выполняется при $|s|=s_1$, за исключением случая $s=s_1$. Следовательно, другие два корня превосходят s_1 по абсолютной величине. Таким образом, из (4.9) получим

$$q_n\sim 1{,}236840/(1{,}0873778)^{n+1}, \tag{4.14}$$

где числитель равен $(2s_1^2+4s_1+8)/(4+4s_1+3s_1^2)$. Эта формула дает удивительно хорошее приближение даже при небольших значениях n. Вместо точного значения $q_3=0{,}875$ она дает $0{,}8847$, а вместо $q_4=0{,}8125$ дает $0{,}81360$. Относительная ошибка монотонно убывает, и для $q_{12}=0{,}41626\ldots$ точное и приближенное значения совпадают до пятого знака. ▶

§ 5. ДВОЙНЫЕ ПРОИЗВОДЯЩИЕ ФУНКЦИИ

Для пары целочисленных случайных величин $\mathbf{X}$, $\mathbf{Y}$ с совместным распределением вида

$$\mathbf{P}\{\mathbf{X}=j,\ \mathbf{Y}=k\}=p_{jk},\quad j,\ k=0,\ 1,\ \ldots \tag{5.1}$$

мы определим производящую функцию от двух переменных формулой

$$P(s_1,\ s_2)=\sum_{j,\,k} p_{jk} s_1^j s_2^k. \tag{5.2}$$

Такую производящую функцию для краткости будем называть двойной.

Методы доказательств, которыми мы пользовались в первых двух параграфах, можно применить здесь без существенных изменений, поэтому достаточно отметить три свойства, вытекающие из (5.2).

а) *Производящие функции маргинальных распределений* $\mathbf{P}\{\mathbf{X}=j\}$ *и* $\mathbf{P}\{\mathbf{Y}=k\}$ *равны* $A(s)=p(s, 1)$ *и* $B(s)=P(1, s)$.

б) *Производящая функция суммы* $\mathbf{X}+\mathbf{Y}$ *равна* $P(s, s)$.

в) *Случайные величины* $\mathbf{X}$ *и* $\mathbf{Y}$ *независимы тогда и только тогда, когда* $P(s_1, s_2)=A(s_1)\,B(s_2)$ *для всех* s_1, s_2.

Примеры. а) *Двумерное распределение Пуассона.* Очевидно, что функция

$$P(s_1, s_2)=e^{-a_1-a_2-b+a_1s_1+a_2s_2+bs_1s_2}, \quad a_i>0,\ b>0 \tag{5.3}$$

разлагается в степенной ряд с положительными коэффициентами, дающими в сумме единицу. Поэтому $P(s_1, s_2)$ представляет собой производящую функцию распределения некоторой пары случайных величин, каждая из которых имеет распределение Пуассона со средним значением, равным соответственно a_1+b и a_2+b. Но сумма $\mathbf{X}+\mathbf{Y}$ имеет производящую функцию $e^{-a_1-a_2-b+(a_1+a_2)s+bs^2}$ и поэтому является случайной величиной, имеющей распределение, отличающееся от распределения Пуассона. (Это так называемое обобщенное распределение Пуассона; см. гл. XII, 2.)

в) *Полиномиальное распределение.* Рассмотрим последовательность n независимых испытаний, каждое из которых приводит к одному из исходов E_0, E_1 или E_2, вероятности которых соответственно равны p_0, p_1 и p_2. Если $\mathbf{X}_i$ — случайная величина, равная числу испытаний, в которых произошло событие E_i, то $(\mathbf{X}_1, \mathbf{X}_2)$ имеет триномиальное распределение с производящей функцией $(p_0+p_1s_1+p_2s_2)^n$. ▶

§ 6*). ТЕОРЕМА НЕПРЕРЫВНОСТИ

В гл. VI было показано, что распределение Пуассона $\{e^{-\lambda}\lambda^k/k!\}$ есть предельная форма биномиального распределения для случая, когда вероятность успеха p зависит от n, причем $np\to\lambda$ при $n\to\infty$. В этом случае

$$b(k;\ n,\ p)\to e^{-\lambda}\lambda^k/k!.$$

Производящая функция распределения $\{b(k;\ n,\ p)\}$ равна $(q+ps)^n=\{1-\lambda(1-s)/n\}^n$. Логарифмируя, убеждаемся, что эта производящая функция стремится к функции $e^{-\lambda(1-s)}$, являющейся производящей функцией распределения Пуассона. Покажем, что это свойство производящей функции сохраняется и в общем случае: последовательность распределений вероятностей сходится к предельному распределению тогда и только тогда, когда соответствующие производящие функции сходятся. К сожалению, эта теорема имеет ограниченное применение, так как наиболее интересными предельными

*) Теорема непрерывности будет использоваться только при обсуждении общего вида безгранично делимых распределений в гл. XII, 2 и получении общего числа потомков в ветвящихся процессах в гл. XII, 5.

формами дискретных распределений являются непрерывные распределения (например, нормальное распределение как предельная форма биномиального распределения).

Теорема непрерывности. *Предположим, что при любом фиксированном n последовательность $a_{0,n}, a_{1,n}, a_{2,n}, \ldots$ является распределением вероятностей, т. е.*

$$a_{k,n} \geqslant 0, \quad \sum_{k=0}^{\infty} a_{k,n} = 1. \tag{6.1}$$

Для того чтобы при любом фиксированном $k \geqslant 0$ существовал предел

$$a_k = \lim_{n \to \infty} a_{k,n}, \tag{6.2}$$

необходимо и достаточно, чтобы при любом s из открытого интервала $0 < s < 1$ существовал предел

$$A(s) = \lim_{n \to \infty} \sum_{k=0}^{\infty} a_{k,n} s^k. \tag{6.3}$$

В этом случае автоматически получается

$$A(s) = \sum_{k=0}^{\infty} a_k s^k. \tag{6.4}$$

Очевидно, что $a_k \geqslant 0$ и $\sum a_k \leqslant 1$. Отметим тем не менее, что сумма может быть строго меньше 1. Например, если $a_{k,n} = f_{k+n}$, то $a_k = 0$ для всех k.

Доказательство[1]). Пусть $A_n(s)$ — ряд, стоящий в правой части (6.3).

(i) Предположим, что (6.2) выполнено, и определим $A(s)$ посредством (6.4). Так как $|a_{k,n} - a_k| \leqslant 1$, то для $0 < s < 1$

$$|A_n(s) - A(s)| \leqslant \sum_{k=0}^{r} |a_{k,n} - a_k| + s^r/(1-s). \tag{6.5}$$

Если выбрать r настолько большим, что $s^r < \varepsilon(1-s)$, то правая часть будет меньше 2ε для всех достаточно больших n. Таким образом, левую часть можно сделать сколь угодно малой, и, следовательно, (6.3) справедливо.

(ii) Предположим, что выполнено (6.3). Очевидно что $A(s)$ является монотонной функцией s, и поэтому $A(0)$ определяется

---

[1]) Эта теорема представляет собой частный случай теоремы непрерывности для преобразований Лапласа — Стилтьеса и доказывается так же, как и эта более общая теорема непрерывности. В литературе теорема непрерывности для производящих функций часто формулируется и доказывается при излишних ограничениях.

как предел $A(s)$ при $s \to 0$. Теперь

$$a_{0,n} \leqslant A_n(s) \leqslant a_{0,n} + s/(1-s). \qquad (6.6)$$

Из этого следует, что при $n \to \infty$ все предельные значения $a_{0,n}$ лежат между $A(0)$ и $A(s) - s/(1-s)$. Устремляя s к 0, видим, что $a_{0,n} \to A(0)$, и поэтому (6.2) выполняется при $k=0$.

Это рассуждение можно применить ко всем k. Действительно, для $0<s<1$

$$(A_n(s) - a_{0,n})/s \to (A(s) - A(0))/s. \qquad (6.7)$$

В левой части мы имеем степенной ряд с неотрицательными коэффициентами, и (6.7) во всех отношениях аналогично (6.3). При помощи рассуждений, подобных проведенным выше, убеждаемся в том, что производная $A'(0)$ существует и что $a_{1,n} \to A'(0)$. По индукции получаем, что (6.2) справедливо для всех k. ►

Примеры. а) *Отрицательное биномиальное распределение.* В примере 2, г), мы видели, что производящая функция распределения $\{f(k;\ r,\ p)\}$ имеет вид $p^r(1-qs)^{-r}$. Пусть теперь λ фиксировано, и пусть $p \to 1$, $q \to 0$ и $r \to \infty$, причем $q \sim \lambda/r$. Тогда

$$[p/(1-qs)]^r = [(1-\lambda/r)/(1-\lambda s/r)]^r. \qquad (6.8)$$

Логарифмируя, убеждаемся, что правая часть этого равенства стремится к функции $e^{-\lambda+\lambda s}$, являющейся производящей функцией распределения Пуассона $\{e^{-\lambda}\lambda^k/k!\}$. Поэтому, *если $r \to \infty$ и $rq \to \lambda$, то*

$$f(k;\ r,\ p) \to e^{-\lambda}\lambda^k/k!. \qquad (6.9)$$

б) *Испытания Бернулли с переменными вероятностями* [1]). Рассмотрим n независимых испытаний, таких, что k-е испытание оканчивается успехом с вероятностью p_k и неудачей с вероятностью $q_k = 1 - p_k$. Число успехов S_n можно записать в виде суммы $\mathbf{S}_n = \mathbf{X}_1 + \ldots + \mathbf{X}_n$ n взаимно независимых случайных величин $\mathbf{X}_k$, имеющих распределение

$$\mathbf{P}\{\mathbf{X}_k = 0\} = q_k, \quad \mathbf{P}\{\mathbf{X}_k = 1\} = p_k.$$

Производящая функция случайной величины $\mathbf{X}_k$ равна $q_k + p_k s$, и, следовательно, производящая функция величины $\mathbf{S}_n$ равна

$$P(s) = (q_1 + p_1 s)(q_2 + p_2 s)\ldots(q_n + p_n s). \qquad (6.10)$$

В качестве приложения этой схемы допустим, что в каждом доме некоторого города в течение дня может возникнуть пожар с небольшой вероятностью p_k. Сумма $p_1 + \ldots + p_n$, где n — число домов в городе, будет ожидаемым числом пожаров. В гл. VI мы видели, что если все p_k равны между собой и случаи возникновения пожара в различных домах стохастически независимы, то число

[1]) См. также примеры гл. IX, 1, д) и гл. IX, 5, б).

пожаров является случайной величиной, имеющей распределение, близкое к распределению Пуассона. Сейчас мы покажем, что это заключение остается в силе также и при более реальном предположении, что вероятности p_k не равны между собой. Этот результат еще больше убеждает нас в том, что распределение Пуассона дает адекватное описание явлений, представляющих собой сочетание нескольких маловероятных событий («успехов»). Примерами служат несчастные случаи и телефонные вызовы.

Воспользуемся уже знакомой нам схемой, в которой число величин n возрастает, причем вероятности p_k меняются вместе с n таким образом, что наибольшая из вероятностей p_k стремится к нулю, а сумма $p_1+p_2+\ldots+p_n=\lambda$ остается постоянной. Тогда, согласно (6.10),

$$\log P(s)=\sum_{k=1}^{n}\log\{1-p_k(1-s)\}. \tag{6.11}$$

Так как $p_k\to 0$, можно воспользоваться тем, что $\log(1-x)=-x-\theta x$, где $\theta\to 0$ при $x\to 0$. Отсюда следует, что

$$\log P(s)=-(1-s)\left\{\sum_{k=1}^{n}(p_k+\theta_k p_k)\right\}\to-\lambda(1-s), \tag{6.12}$$

так что $P(s)$ стремится к производящей функции распределения Пуассона. Следовательно, $\mathbf{S}_n$ *имеет в пределе распределение Пуассона*. Мы заключаем, что при большом n и не очень большом $\lambda=p_1+\ldots+p_n$ распределение величины $\mathbf{S}_n$ может быть приближенно представлено распределением Пуассона. ▶

§ 7. ЗАДАЧИ

1. Пусть $\mathbf{X}$ — случайная величина с производящей функцией $P(s)$. Найти производящие функции случайных величин $\mathbf{X}+1$ и $2\mathbf{X}$.

2. Найти производящие функции следующих последовательностей: а) $\mathbf{P}\{\mathbf{X}\leqslant n\}$; б) $\mathbf{P}\{\mathbf{X}<n\}$; в) $\mathbf{P}\{\mathbf{X}\geqslant n\}$; г) $\mathbf{P}\{\mathbf{X}>n+1\}$; д) $\mathbf{P}\{\mathbf{X}=2n\}$.

3. Рассмотрим последовательности испытаний Бернулли. Обозначим через u_n вероятность того, что комбинация *УН* впервые появится при $(n-1)$-м и n-м испытаниях. Найти производящую функцию, среднее значение и дисперсию.

4. Выяснить, какие из формул гл. II, 12 включают свертки и где можно было использовать производящие функции.

5. Пусть a_n — число способов, которыми можно получить сумму в n очков при произвольном числе бросаний правильной кости. Показать, что производящая функция последовательности $\{a_n\}$ равна $\{1-s-s^2-s^3-s^4-s^5-s^6\}^{-1}-1$.

6. Пусть a_n — число способов, которыми $n-2$ (непересекающихся) диагоналей разбивают выпуклый $(n+1)$-угольник $P_0P_1\ldots P_n$ на треугольники[1]. Положим $a_1=1$. Показать, что при $n\geqslant 2$

$$a_n=a_1a_{n-1}+a_2a_{n-2}+\ldots+a_{n-1}a_1.$$

Найти производящую функцию и явные выражения для a_n.

[1] Эта задача обсуждается в книге Polya G. Mathematics and plausible reasoning, Princeton University Press, 1954, p. 102. [Имеется перевод: Пойа Д., Математика и правдоподобные рассуждения.— М.: ИЛ, 1957; М.: Наука, 1975, с. 123.]

Указание. Предположить, что одна из диагоналей проходит через P_0, и принять за k наименьший индекс, такой, что P_0P_k является диагональю.

Замечание. Задачи 7—11 относятся к § 3. Производящие функции Φ, U и F соответствуют первым достижениям точки 1, возвращениям в начало и первым возвращениям; см. формулы (3.6), (3.12) и (3.14). Вычисления здесь не требуются.

7. а) Вероятность того, что возвращение в начало происходит не позже момента n-го испытания, равна $(1-s)^{-1}F(s)$.

б) Следствие. Производящая функция последовательности вероятностей того, что $\mathbf{S}_j \neq 0$, $j=1, \ldots, n$, равна $\sqrt{(1+s)/(1-s)} = (1+s)\,U(s)$.

в) Показать, что это эквивалентно утверждению, приведенному после формулы (3.16).

8. Производящая функция последовательности вероятностей того, что после n-го испытания (исключая само n-е испытание) не происходит возвращения в начало, равна $(1-s)^{-1}\,U(s)\,|\,p-q\,|$.

9. а) Производящая функция последовательности вероятностей $\mathbf{P}\{\mathbf{S}_n=r\}$ (при фиксированном $r>0$) равна $\Phi^r(s)\,U(s)$.

б) При $p=1/2$ это выражение является производящей функцией последовательности вероятностей того, что $\mathbf{S}_k=r$ в точности для одного индекса $k \leqslant n$.

10. а) Найти производящую функцию последовательности вероятностей того, что событие $\mathbf{S}_n=r$ произойдет ровно k раз ($r>0$ и $k>0$ фиксированы).

б) Решить предыдущую задачу, заменив «ровно» на «самое большее».

11. а) Найти производящую функцию последовательности вероятностей того, что первое возвращение в начало, следующее за первым достижением точки $r>0$, происходит при испытании с номером r.

б) Решить предыдущую задачу, опустив слово «первое».

12. Найти производящую функцию величины $\mathbf{S}_r$ (r фиксировано) в *задаче о времени ожидания* (пример гл. IX, 3, г)). Проверить формулу (3.3) гл. IX для среднего значения и вычислить дисперсию.

13. *Продолжение.* Другой метод решения той же задачи. Пусть $p_n(r) = \mathbf{P}\{\mathbf{S}_r=n\}$. Доказать рекуррентную формулу

$$p_{n+1}(r) = [(r-1)/N]\,p_n(r) + [(N-r+1)/N]\,p_n(r-1). \tag{7.1}$$

Непосредственно из (7.1) вывести производящую функцию.

14. Решить две предыдущие задачи для r заранее выбранных элементов (вместо r произвольных).

15[1]. Назовем *циклом* последовательность испытаний Бернулли до первой неудачи включительно. Найти производящую функцию и распределение вероятностей общего числа $\mathbf{S}_r$ успехов в r циклах.

16. *Продолжение.* а) Пусть $\mathbf{R}$ — число последовательных циклов до ν-го успеха (т. е. ν-й успех принадлежит $\mathbf{R}$-му циклу). Найти $\mathbf{E}(\mathbf{R})$ и $\operatorname{Var}(\mathbf{R})$. Доказать, что

$$\mathbf{P}\{\mathbf{R}=r\} = p^\nu q^{r-1}\binom{r+\nu-2}{\nu-1}.$$

---

[1]) Задачи 15 и 16 имеют самое непосредственное отношение к игре в биллиард. Вероятность попадания p служит мерой искусства игрока. Игрок продолжает игру до первого промаха. Поэтому число забитых им шаров есть длина «цикла попаданий». Игра заканчивается, когда один из игроков забьет N шаров. Таким образом, в задаче 15 ищется распределение вероятностей для числа циклов, в течение которых игрок забивает k шаров, в задаче 16 — средняя продолжительность игры и вероятность ничьей. Дальнейшие подробности см. в статье Bottema O., Van Veen S. C., Kansberekningen bij het biljartspel, Nieuw Archief voor Wiskunde (на датском языке), 22 (1943), 16—33, 123—158.

б) Рассмотрим *две* последовательности испытаний Бернулли с вероятностями p_1, q_1 и p_2, q_2 соответственно. Найти вероятность совпадения номеров циклов, при которых произойдет N-й успех в каждой последовательности.

---

17. Пусть случайная величина $\mathbf{X}$ принимает значения $0, 1, \ldots, r-1$ каждое с вероятностью $1/r$. Показать, что если r является составным числом, например $r=ab$, то $\mathbf{X}$ можно представить в виде суммы независимых целочисленных случайных величин.

18. Пусть $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ — сумма взаимно независимых случайных величин, каждая из которых принимает значения $1, 2, \ldots, a$ с вероятностями $1/a$. Показать, что производящая функция величины $\mathbf{S}_n$ равна

$$P(s)=\{s(1-s^a)/[a(1-s)]\}^n,$$

откуда для $j \geqslant n$

$$\mathbf{P}\{\mathbf{S}_n=j\}=a^{-n}\sum_{\nu=0}^{\infty}(-1)^{\nu+j-n-a\nu}\binom{n}{\nu}\binom{-n}{j-n-a\nu}=$$

$$=a^{-n}\sum_{\nu=0}^{\infty}(-1)^{\nu}\binom{n}{\nu}\binom{j-a\nu-1}{n-1}.$$

(Лишь конечное число членов этой суммы отлично от нуля.)

Замечание. При $a=6$ мы получим вероятность выпадения $j+n$ очков при бросании n костей. Решение этой задачи восходит к Муавру.

19. *Продолжение.* Вероятность $\mathbf{P}\{\mathbf{S}_n \leqslant j\}$ имеет производящую функцию $P(s)/(1-s)$, и поэтому

$$\mathbf{P}\{\mathbf{S}_n \leqslant j\}=(1/a^n)\sum_{\nu}(-1)^{\nu}\binom{n}{\nu}\binom{j-a\nu}{-n}.$$

20. *Продолжение: предельная форма.* Если $a \to \infty$ и $j \to \infty$, причем $j/a \to x$, то

$$\mathbf{P}\{\mathbf{S}_n \leqslant j\} \to (1/n!)\sum_{\nu}(-1)^{\nu}\binom{n}{\nu}(x-\nu)^n,$$

где сумма берется по всем ν из интервала $0 \leqslant \nu < x$.

Замечание. Этот результат принадлежит Лагранжу. В теории геометрических вероятностей правая часть представляет функцию распределения суммы n независимых случайных величин с «равномерным» распределением на интервале $(0, 1)$.

21. Пусть u_n — вероятность того, что число успехов в n испытаниях Бернулли делится на 3. Найти рекуррентное соотношение для u_n, а из него — производящую функцию.

22. *Продолжение: другой метод.* Пусть v_n и w_n — вероятности того, что $\mathbf{S}_n$ имеет соответственно вид $3v+1$ и $3v+2$ (так что $u_n+v_n+w_n=1$). Найти три одновременно выполняющиеся рекуррентные соотношения, а из них — три уравнения для производящих функций.

23. Пусть $\mathbf{X}$ и $\mathbf{Y}$ — независимые случайные величины с производящими функциями $U(s)$ и $V(s)$. Показать, что $\mathbf{P}\{\mathbf{X}-\mathbf{Y}=j\}$ равно коэффициенту при s^j в разложении функции $U(s)V(1/s)$, где $j=0, \pm 1, \pm 2, \ldots$.

24. *Производящие функции моментов.* Пусть $\mathbf{X}$ — случайная величина с производящей функцией $P(s)$, и пусть ряд $\sum p_n s^n$ сходится при некотором $s_0 > 1$. Тогда все моменты $m_r=\mathbf{E}(\mathbf{X}^r)$ существуют и производящая функция $F(s)$ последовательности $m_r/r!$ сходится по меньшей мере при $|s| < \log s_0$. Кроме

того,

$$F(s)=\sum_{r=0}^{\infty}(m_r/r!)\,s^r=P(e^s).$$

Замечание. $F(s)$ обычно называют *производящей функцией моментов*, хотя на самом деле она производит последовательность $m_r/r!$.

25. Предположим, что $A(s)=\sum a_n s^n$ — рациональная функция $U(s)/V(s)$ и что s_1 — наименьший по абсолютной величине корень многочлена $V(s)$. Показать, что если s_1 имеет кратность r, то

$$a_n \sim \frac{\rho_1}{s_1^{n+r}}\binom{n+r-1}{r-1},$$

где $\rho_1=(-1)^r r!\,U(s_1)/V^{(r)}(s_1)$.

26. *Двойное отрицательное биномиальное распределение.* Показать, что при положительных значениях параметров выражение $p_0^a\{1-p_1s_1-p_2s_2\}^{-a}$ представляет собой производящую функцию распределения пары случайных величин $(\mathbf{X}, \mathbf{Y})$, таких, что маргинальные распределения $\mathbf{X}$, $\mathbf{Y}$ и $\mathbf{X}+\mathbf{Y}$ являются отрицательными биномиальными распределениями [1].

[1] Распределение этого типа использовали Дж. Бейтс и Дж. Нейман при исследованиях частоты катастроф; см. University of California Publications in Statistics, 1 (1952).

СЛОЖНЫЕ РАСПРЕДЕЛЕНИЯ[1]. ВЕТВЯЩИЕСЯ ПРОЦЕССЫ

Значительная часть теории вероятностей связана с суммами независимых случайных величин, и порою число слагаемых в таких суммах само является случайной величиной. Мы ограничимся здесь рассмотрением ситуации, когда слагаемые являются целочисленными случайными величинами, чтобы показать, как используются производящие функции, и подготовиться к изучению (в томе 2) безгранично делимых распределений и процессов с независимыми приращениями.

В качестве примера особенно привлекательных применений мы опишем некоторые результаты изящной теории ветвящихся процессов.

§ 1. СУММЫ СЛУЧАЙНОГО ЧИСЛА ВЕЛИЧИН

Пусть $\{\mathbf{X}_k\}$ — последовательность взаимно независимых случайных величин с одним и тем же распределением $\mathbf{P}\{\mathbf{X}_k=j\}=f_j$ и производящей функцией $f(s)=\sum f_i s^i$. Мы будем рассматривать суммы

$$\mathbf{S}_\mathbf{N}=\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_\mathbf{N},$$

в которых число $\mathbf{N}$ слагаемых является случайной величиной, не зависящей от $\mathbf{X}_j$. Пусть $\mathbf{N}$ имеет распределение $\mathbf{P}\{\mathbf{N}=n\}=g_n$ и производящую функцию $g(s)=\sum g_n s^n$. Распределение $\{h_j\}$ суммы $\mathbf{S}_\mathbf{N}$ можно найти по основной формуле для условных вероятностей

$$h_j=\mathbf{P}\{\mathbf{S}_\mathbf{N}=j\}=\sum_{n=0}^{\infty}\mathbf{P}\{\mathbf{N}=n\}\,\mathbf{P}\{\mathbf{X}_1+\ldots+\mathbf{X}_n=j\}. \qquad (1.1)$$

Если число возможных значений $\mathbf{N}$ конечно, то случайная величина $\mathbf{S}_\mathbf{N}$ определена на пространстве элементарных событий, соответствующем конечному числу $\mathbf{X}_k$. В противном случае для вероятностного определения случайной величины $\mathbf{S}_\mathbf{N}$ как суммы необходимо

---

*) Материал этой главы не будет использоваться в дальнейшем.

[1]) В оригинале compound distributions. Здесь имеются в виду распределения, производящие функции которых являются сложными функциями (compound functions) вида $h(s)=g(f(s))$, где g и f — производящие функции (иначе говоря, h представляет собой суперпозицию g и f). — *Прим. перев.*

ввести пространство элементарных событий, соответствующее бесконечной последовательности $\{\mathbf{X}_k\}$. Однако нас будет интересовать только функция распределения $\mathbf{S_N}$, и мы рассматриваем распределение (1.1) как определение случайной величины $\mathbf{S_N}$ на пространстве элементарных событий с точками 0, 1, 2, ...

Для фиксированного n распределение $\mathbf{X}_1 + \ldots + \mathbf{X}_n$ является n-кратной сверткой $\{f_i\}$ с самим собой, поэтому (1.1) можно записать в компактном виде

$$\{h_j\} = \sum_{n=0}^{\infty} g_n \{f_j\}^{n*}. \tag{1.2}$$

Эту формулу можно упростить, используя производящие функции. Производящей функцией $\{f_j\}^{n*}$ является $f^n(s)$, и из (1.2) следует, что производящая функция суммы $\mathbf{S_N}$ вычисляется по формуле

$$h(s) = \sum_{j=0}^{\infty} h_j s^j = \sum_{n=0}^{\infty} g_n f^n(s). \tag{1.3}$$

Правая часть представляет собой ряд Тейлора для $g(s)$, в котором s заменено на $f(s)$; значит, она равна $g(f(s))$. Тем самым доказано следующее утверждение.

Теорема. *Производящей функцией суммы* $\mathbf{S_N} = \mathbf{X}_1 + \ldots + \mathbf{X_N}$ *является сложная функция* $g(f(s))$.

Доказательство можно повторить в терминах условных математических ожиданий. По определению

$$\mathrm{E}(s^{\mathbf{S_N}} \mid \mathbf{N} = n) = f^n(s), \tag{1.4}$$

и, чтобы получить $h(s) = \mathrm{E}(s^{\mathbf{S_N}})$, нам нужно умножить (1.4) на $\mathbf{P}\{\mathbf{N} = n\}$ и просуммировать по n (см. формулу (2.9) гл. IX).

Представляют интерес два частных случая.

а) Если $\mathbf{X}_i$—бернуллиевские случайные величины с $\mathbf{P}\{\mathbf{X}_i = 1\} = p$ и $\mathbf{P}\{\mathbf{X}_i = 0\} = q$, то $f(s) = q + ps$, и поэтому $h(s) = g(q + ps)$.

б) Если $\mathbf{N}$ имеет распределение Пуассона со средним t, то $h(s) = e^{-t + tf(s)}$. Распределение с такой производящей функцией будем называть *обобщенным распределением Пуассона*[1]). В частности, если $\mathbf{X}_i$—бернуллиевские величины и $\mathbf{N}$ имеет распределение Пуассона, то $h(s) = e^{-tp + tps}$; иначе говоря, *сумма* $\mathbf{S_N}$ *имеет распределение Пуассона со средним* tp.

Примеры. а) В примере гл. VI, 7, в) мы отмечали, что облучение рентгеновскими лучами вызывает перестройку хромосом в клетках; при фиксированных дозе и времени облучения число $\mathbf{N}$ изменений имеет распределение Пуассона. Каждое изменение с

---

[1]) Мы используем принятый в нашей литературе термин; в оригинале compound Poisson distribution.— *Прим. перев.*

вероятностью q залечивается, а с вероятностью $p=1-q$ приводит к гибели клетки. Здесь $\mathbf{S}_{\mathbf{N}}$ — число *наблюдаемых* изменений [1]) — имеет распределение Пуассона со средним tp.

б) Пусть в экспериментах с отловом животных [2]) g_n означает вероятность того, что численность животных данного вида равна n. Если для каждого животного вероятность быть пойманным равна p, то (в предположении стохастической независимости) число попавших в выборку представителей данного вида есть случайная величина $\mathbf{S}_{\mathbf{N}}$ с производящей функцией $g(q+ps)$. Эту схему можно изменять многими способами. Например, пусть g_n — вероятность того, что насекомое откладывает n яиц, а p — вероятность появления из яйца личинки. Тогда $\mathbf{S}_{\mathbf{N}}$ есть число появившихся личинок. Или пусть g_n — вероятность наличия в семье n детей, а отношение чисел мальчиков и девочек в популяции равно $p:q$. Тогда $\mathbf{S}_{\mathbf{N}}$ — число мальчиков в семье.

в) Каждое растение дает много семян, но каждое семя имеет малую вероятность прорастания, и поэтому разумно предполагать, что число потомков отдельного растения имеет распределение Пуассона. Если $\{g_n\}$ — распределение числа растений, то $g(e^{-\lambda+\lambda s})$ — производящая функция числа проросших семян.

г) *Суммарное время обслуживания.* Рассмотрим телефонную линию, прилавок или любое другое обслуживающее устройство, для которого длительности обслуживания последовательно появляющихся клиентов могут рассматриваться как независимые одинаково распределенные случайные величины $\mathbf{X}_1, \mathbf{X}_2, \ldots$. Число клиентов (или заявок), поступающих за день, само является случайной величиной $\mathbf{N}$, и суммарное время их обслуживания равно поэтому случайной сумме $\mathbf{X}_1+\ldots+\mathbf{X}_{\mathbf{N}}$.

§ 2. ОБОБЩЕННОЕ РАСПРЕДЕЛЕНИЕ ПУАССОНА

Среди случайных сумм $\mathbf{S}_{\mathbf{N}}=\mathbf{X}_1+\ldots+\mathbf{X}_{\mathbf{N}}$ наиболее важны для приложений те, в которых $\mathbf{N}$ имеет распределение Пуассона. По причинам, которые станут ясны позднее, мы обозначим математическое ожидание $\mathbf{N}$ через λt. Если $\mathbf{X}_j$ имеют одно и то же распределение $\{f_i\}$, то $\mathbf{S}_{\mathbf{N}}$ имеет *обобщенное распределение Пуассона*

$$\{h_i\}_t=e^{-\lambda t}\sum_{n=0}^{\infty}[(\lambda t)^n/n!]\,\{f_i\}^{n*} \tag{2.1}$$

с производящей функцией

$$h_t(s)=e^{-\lambda t+\lambda t f(s)}. \tag{2.2}$$

---

[1]) См. Catcheside D. G., Genetic effects of radiations, Advances in Genetics, edited by M. Demerec, v. 2, Acad. Press, New York, 1948, 271—358, особенно с. 339.

[2]) Kendall D. G., On some modes of population growth leading to R. A. Fisher's logarithmic series distribution, Biometrika, 35 (1948), 6—15.

Примеры. а) *Суммарный ущерб.* Предположим, что число ударов молнии в любом интервале времени длительности t является пуассоновской случайной величиной с математическим ожиданием λt. Если $\{f_n\}$ — распределение вероятностей ущерба, причиненного одним ударом молнии, то (при условии стохастической независимости) суммарный ущерб за время t имеет обобщенное распределение Пуассона (2.1).

б) *Ливни в космических лучах.* Принято считать, что число $\mathbf{N}$ ливней в космических лучах в интервале времени длительности t имеет распределение Пуассона с математическим ожиданием λt. Для любого счетчика число зарегистрированных частиц одного ливня является случайной величиной с распределением $\{f_i\}$. Общее число частиц, зарегистрированных за время t, снова есть случайная сумма $\mathbf{S}_\mathbf{N}$ с обобщенным распределением Пуассона (2.1).

в) *В экологии* предполагается, что число выводков животных на данном участке имеет распределение Пуассона с математическим ожиданием, пропорциональным площади t участка. Пусть $\{f_k\}$ — распределение числа животных в выводке; предположим, что размеры выводков независимы. При этих условиях число животных на данном участке подчиняется обобщенному распределению Пуассона (2.1). Эта модель широко используется на практике. ▶

Заметим, что все три примера тесно связаны с явлением, обсуждавшимся в гл. VI, 6 в связи с распределением Пуассона. В первых двух примерах каждому интервалу времени соответствует случайная величина $\mathbf{S}_\mathbf{N}$. (Это верно и для примера в), если мы согласимся рассматривать площадь как своеобразное время.) В такой модели предполагается, что если интервал разбит на два непересекающихся интервала, то их вклады стохастически независимы и в сумме дают $\mathbf{S}_\mathbf{N}$. Применительно к производящей функции (2.2) это означает, что

$$h_{t+\tau}(s) = h_t(s)\, h_\tau(s). \tag{2.3}$$

Производящая функция (2.2) любого обобщенного распределения Пуассона удовлетворяет (2.3). Покажем теперь, что верно и обратное: семейство производящих функций h_t, удовлетворяющее условию (2.3), обязательно имеет вид (2.2). (Следует иметь в виду, что это утверждение верно только для целочисленных случайных величин. Понятие обобщенного распределения Пуассона остается осмысленным даже тогда, когда $\mathbf{X}_j$ имеют произвольное распределение, а аналог соотношения (2.3) играет важную роль в общей теории случайных процессов с независимыми приращениями. Однако распределения таких процессов могут и не быть обобщенными распределениями Пуассона.)

Следующие определение и теорема в действительности относятся к распределениям вероятностей на множестве целых чисел

0, 1, ..., однако для простоты они формулируются в терминах соответствующих производящих функций.

Определение. *Вероятностная производящая функция h называется безгранично делимой, если для любого натурального числа n корень n-й степени $\sqrt[n]{h}$ тоже является вероятностной производящей функцией.*

Следующая теорема показывает, что это утверждение остается справедливым даже когда $n>0$ не является целым числом. Если семейство вероятностных производящих функций удовлетворяет (2.3), то $\sqrt[n]{h_t}=h_{t/n}$, и поэтому h_t безгранично делима. Обратной этому утверждению является следующая теорема.

Теорема [1]. *Все безгранично делимые производящие функции имеют вид* (2.2), *причем* $\{f_i\}$ — *распределение вероятностей на* 0, 1, ...

Доказательство. Положим $h(s)=\sum h_k s^k$ и допустим, что $\sqrt[n]{h}$ является вероятностной производящей функцией при каждом $n\geqslant 1$. Тогда $h_0>0$, поскольку в противном случае свободный член в степенном ряду для $\sqrt[n]{h}$ был бы равен нулю, а это, в свою очередь, означало бы, что $h_0=h_1=\ldots=h_{n-1}=0$. Следовательно, $\sqrt[n]{h(s)}\to 1$ для любого $0\leqslant s\leqslant 1$, и поэтому

$$\log\sqrt[n]{h(s)/h_0}=\log\left[1+\left(\sqrt[n]{h(s)/h_0}-1\right)\right]\sim\sqrt[n]{h(s)/h_0}-1, \quad (2.4)$$

где знак $\sim$ означает, что отношение правой части к левой стремится к 1. Из (2.4) и его частного случая при $s=1$ мы получаем (учитывая, что $h(1)=1$):

$$\frac{\log h(s)-\log h_0}{-\log h_0}=\frac{\log\sqrt[n]{h(s)/h_0}}{\log\sqrt[n]{1/h_0}}\sim\frac{\sqrt[n]{h(s)}-\sqrt[n]{h_0}}{1-\sqrt[n]{h_0}}. \quad (2.5)$$

Правая часть представляет собой степенной ряд с положительными коэффициентами, и, полагая $s=1$, легко проверить, что сумма этих коэффициентов равна 1. Таким образом, при любом n правая часть — вероятностная производящая функция, и поэтому левая часть является пределом последовательности вероятностных производящих функций. Согласно теореме непрерывности гл. XI,6, это означает, что левая часть — производящая функция неотрицательной последовательности $\{f_j\}$. Полагая $s=1$, убеждаемся в том, что $\sum f_j=1$. Значит, h имеет вид (2.2) с $\lambda t=-\log h_0$. ▶

Переформулировкой этой теоремы является следующий критерий.

---

[1]) Это простой частный случай важной теоремы П. Леви.

Критерий. *Функция h является безгранично делимой вероятностной производящей функцией тогда и только тогда, когда $h(1)=1$ и*

$$\log[h(s)/h(0)] = \sum_{k=1}^{\infty} a_k s^k, \quad \text{где } a_k \geqslant 0, \quad \sum a_k = \lambda < \infty. \quad (2.6)$$

Действительно, если выполнено (2.6), то достаточно положить $f_k = a_k/\lambda$, чтобы привести h к канонической форме (2.2) (с $t=1$), а она, в свою очередь, является производящей функцией обобщенного распределения Пуассона, определенного в (2.1).

Примеры. г) Сравнивая (2.2) с теоремой предыдущего параграфа, мы видим, что если распределение $\mathbf{N}$ безгранично делимо, то таким же является и распределение случайной суммы $\mathbf{S_N}$.

д) *Отрицательное биномиальное* распределение с производящей функцией

$$h_t(s) = [p/(1-qs)]^t, \quad p+q=1, \quad (2.7)$$

обладает свойством (2.3) и поэтому безгранично делимо. Переходя к логарифмам, сразу убеждаемся в том, что оно действительно имеет вид (2.2), причем

$$f_n = q^n/(\lambda n), \quad \lambda = \log p^{-1}; \quad (2.8)$$

$\{f_n\}$ называется *логарифмическим распределением* и широко используется статистиками.

е) Из разложений (8.9) и (8.10) гл. II легко получить, что при $q=1-p>p$ функции

$$f(s) = \sqrt{(q-p)\frac{q+ps}{q-ps}}, \quad g(s) = \frac{\sqrt{q-p}}{\sqrt{q^2-p^2s^2}} \quad (2.9)$$

удовлетворяют условию (2.6), и поэтому *как f, так и g являются безгранично делимыми вероятностными производящими функциями.* Интересно заметить, что

$$f(s) = g(s)(q+ps). \quad (2.10)$$

Мы получили здесь *разложение безгранично делимой функции в произведение двух производящих функций, из которых только одна безгранично делима.* Существование таких разложений поначалу воспринималось как большая неожиданность, и эта тема некоторое время оживленно обсуждалась. ►

Замечательное свойство обобщенного распределения Пуассона является поводом для некоторых любопытных умозаключений. Если мы введем для простоты обозначение $\lambda_i = \lambda f_i$, то производящую функцию h_t из (2.2) можно представить в виде произведения

$$h_t(s) = e^{\lambda_1 t(s-1)} e^{\lambda_2 t(s^2-1)} e^{\lambda_3 t(s^3-1)} \ldots \quad (2.11)$$

Здесь произведение может быть бесконечным, но это не имеет отношения к нашим рассуждениям, и мы можем считать, что лишь конечное число λ_i положительно. Первый сомножитель — это производящая функция обычного распределения Пуассона с математическим ожиданием $\lambda_1 t$. Второй сомножитель — это производящая функция удвоенной пуассоновской случайной величины, т. е. знакомую нам вероятность $e^{-\lambda_2 t}(\lambda_2 t)^n/n!$ имеет точка $2n$, а не n. Аналогично, k-й сомножитель соответствует распределению Пуассона, которое приписывает вероятности числам, кратным k. Таким образом, (2.11) дает новое представление $\mathbf{S_N}$ в виде суммы таких независимых случайных величин $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$, что $\mathbf{Y}_k$ принимает только значения $0, k, 2k, \ldots$ и имеет на них распределение Пуассона. Смысл соотношения (2.11) можно описать следующим образом. Пусть $\mathbf{N}_j$ — число величин $\mathbf{X}_1, \ldots, \mathbf{X_N}$, равных j. Тогда $\mathbf{N}=\mathbf{N}_1+\mathbf{N}_2+\ldots$, и (2.11) означает, что случайные величины $\mathbf{N}_k$ *взаимно независимы и имеют распределения Пуассона.*

Пример. ж) *Автомобильные катастрофы.* Будем рассматривать $\mathbf{X}_n$ как число автомобилей, попавших в n-ю катастрофу. При обычных предположениях о независимости $\mathbf{X}_n$ и о том, что число $\mathbf{N}$ катастроф имеет распределение Пуассона, находим, что общее число автомобилей, попавших в катастрофы, составляет $\mathbf{X}_1+\ldots+\mathbf{X_N}$ и имеет обобщенное распределение Пуассона (2.1). Мы можем теперь рассмотреть отдельно число $\mathbf{N}_k$ катастроф, в каждую из которых попало ровно k автомобилей. Согласно (2.11), случайные величины $\mathbf{N}_k$ взаимно независимы и имеют распределения Пуассона. Практические выводы из этого результата не нуждаются в комментариях. ▶

(Еще один вариант обобщенного распределения Пуассона будет приведен в примере гл. XVII, 9, а).)

2а. Процессы с независимыми приращениями

Интерес к предыдущим результатам увеличивает их тесная связь с одним важным классом случайных процессов. Эти процессы будут сейчас неформально описаны, хотя соответствующая теория лежит вне рамок этой книги.

Чтобы начать с простейшего примера, рассмотрим число вызовов, поступающих на телефонную станцию, как функцию времени. Процесс работы станции описывается заданием для каждого t числа $\mathbf{Z}(t)$ вызовов, поступивших между 0 и t. Если последовательные вызовы поступали в моменты $t_1, t_2, \ldots$, то $\mathbf{Z}(t)=0$ при $0 \leqslant t < t_1$, и вообще $\mathbf{Z}(t)=k$ при $t_k \leqslant t < t_{k+1}$. Обратно, любую неубывающую функцию, принимающую только значения $0, 1, 2, \ldots$, можно рассматривать как возможный вариант процесса работы телефонной станции. Поэтому вероятностная модель должна строиться на пространстве элементарных событий, точками которого являются функции $\mathbf{Z}(t)$ (а не последовательности, как в случае дискретных испытаний). Вероятности должны быть заданы способом, который позволял бы нам работать с такими сложными событиями, как событие, состоящее в том, что $\mathbf{Z}(t+1)-\mathbf{Z}(t)$ когда-нибудь превзойдет 17 или что $\mathbf{Z}(t)$ в некоторый момент времени превзойдет $at+b$ (последнее событие есть основной объект задачи о разорении в теории коллективного риска). В дальнейшем мы считаем само собой разумеющимся, что такое задание действительно возможно; наша цель — показать, что из простых и естественных предположений о природе процесса следует, что для каждого фиксированного t случайная величина $\mathbf{Z}(t)$ должна иметь обобщенное распределение Пуассона.

Аналогичные рассуждения справедливы для широкого круга реальных явлений. Случайная величина $\mathbf{Z}(t)$ может соответствовать не числу телефонных вызовов, а суммарной длительности (или стоимости) уже проведенных разговоров, числу автомобилей, попавших в катастрофы, суммарному ущербу, причиненному молниями, общему потреблению электроэнергии, суммарному количеству осадков и т. п. Оставаясь в рамках настоящей главы, мы должны предполагать, что случайные величины $\mathbf{Z}(t)$ принимают только целые неотрицательные значения, но теорию можно обобщить на произвольные случайные величины. Наше внимание будет сосредоточено на процессах, удовлетворяющих двум следующим основным условиям, которые во многих приложениях представляются естественными.

а) Процесс является *однородным по времени*, т. е. распределения приращений $\mathbf{Z}(t+h)-\mathbf{Z}(t)$ зависят лишь от длины интервала времени, но не от его положения [1]).

б) Приращения $\mathbf{Z}(t_2)-\mathbf{Z}(t_1)$ и $\mathbf{Z}(t_1)-\mathbf{Z}(t_0)$ на смежных интервалах времени взаимно независимы.

Результаты предыдущего раздела можно теперь переформулировать следующим образом: *если* существует процесс, удовлетворяющий условиям а) и б), то его приращения $\mathbf{Z}(t+h)-\mathbf{Z}(t)$ имеют обобщенные распределения Пуассона. В частности, если $\mathbf{Z}(t)$ имеет скачки только единичного размера, то эти случайные величины имеют обычные распределения Пуассона (см. формулу (2.11)).

Таким образом, мы охарактеризовали обычные и обобщенные распределения Пуассона присущими им свойствами; в отличие от рассуждений, проведенных в гл. VI, распределение Пуассона оказывается не приближением, а вполне самостоятельным распределением (можно сказать, проявлением некоторого естественного закона). Разумеется, теперь перед нами встает обратная задача: каждому ли семейству обобщенных распределений Пуассона соответствует некоторый случайный процесс? Ответ на этот вопрос утвердительный, однако (несколько неожиданно) оказывается, что двух наших условий не достаточно для *однозначного* определения процесса. Для однозначного определения представляющих интерес процессов необходимо усилить условие б), потребовав, чтобы n приращений, соответствующих конечному разбиению $t_0<t_1<\ldots<t_n$, были взаимно независимы при любом n. Это — определяющее свойство *процессов с независимыми приращениями*. Любое семейство обобщенных распределений Пуассона однозначно определяет процесс с независимыми приращениями, и поэтому никаких теоретических трудностей не возникает. Однако мы предполагали независимость только для двух интервалов. Такое ограниченное условие достаточно для определения вида распределений приращений, но можно построить довольно патологические процессы, удовлетворяющие указанному условию [2]). На этом примере видны трудности, присущие построению точной модели случайного процесса.

---

[1]) Это условие не столь ограничительно, как может показаться на первый взгляд. Например, вызовы на телефонную станцию в самый напряженный дневной час поступают чаще, чем от полуночи до часу ночи, поэтому процесс является неоднородным по времени. Однако по очевидным причинам связисты рассматривают в основном «час пик», а на этом отрезке времени процесс может считаться однородным. Практика показывает, что в течение часа пик поступление вызовов с удивительной точностью соответствует распределению Пуассона.

[2]) В таком процессе приращение $\mathbf{Z}(t_2)-\mathbf{Z}(t_1)$ не зависит ни от $\mathbf{Z}(t_3)-\mathbf{Z}(t_2)$, ни от $\mathbf{Z}(t_1)-\mathbf{Z}(t_0)$, но однозначно определяется последними двумя приращениями. См. Feller W., Non-Markovian processes with the semi-group property, Ann. Math. Statist., 30 (1959), 1252—1253.

§ 3. ПРИМЕРЫ ВЕТВЯЩИХСЯ ПРОЦЕССОВ

Мы опишем случайный процесс, который является упрощенной моделью ряда реальных процессов, а также показывает полезность производящих функций. Словесно этот процесс описывается следующим образом.

Мы рассматриваем частицы, которые могут порождать новые частицы того же вида. Одна начальная частица образует исходное, или нулевое, поколение. Каждая частица с вероятностью p_k ($k=0, 1, 2, \ldots$) *порождает ровно* k *новых частиц; непосредственные потомки частиц* n*-го поколения образуют* $(n+1)$*-е поколение. Частицы каждого поколения размножаются независимо одна от другой. Нас интересуют размеры последовательных поколений.*

Прежде чем дать строгое описание в терминах случайных величин, имеет смысл привести несколько примеров.

а) *Ядерные цепные реакции.* Это применение стало общеизвестным в связи с атомной бомбой [1]). Частицами являются нейтроны, которые случайным образом сталкиваются с другими частицами. Пусть p — вероятность того, что частица рано или поздно испытает столкновение, породив при этом m частиц; тогда $q=1-p$ — вероятность того, что у частицы не будет потомков, т. е. она останется инертной (удалится или поглотится каким-либо другим способом). В этой схеме единственно возможными значениями числа потомков являются 0 и m, а соответствующие вероятности равны p и q (т. е. $p_0=q$, $p_m=p$, $p_j=0$ для всех остальных j). В худшем случае первая частица останется инертной и процесс никогда не начнется. В лучшем случае в первом поколении будет m частиц, во втором m^2 и т. д. Если p близко к единице, то весьма вероятно, что число частиц будет быстро возрастать. В математической модели это число может расти неограниченно. На самом деле при очень большом числе частиц вероятности расщепления не могут оставаться постоянными и стохастической независимости тоже не может быть. Тем не менее для реальных цепных реакций математическое описание «неограниченно увеличивающегося числа частиц» можно интерпретировать как «взрыв».

б) *Выживание фамилий.* В этом случае (как часто в жизни) учитываются только потомки мужского пола; они играют роль частиц, и p_k — это вероятность того, что новорожденный мальчик будет отцом ровно k мальчиков. Наша схема содержит два искусственных упрощения. Рождаемость меняется от века к веку, и поэтому в действительности распределение $\{p_k\}$ изменяется от поколения к поколению. Более того, общность наследственности и ок-

---

[1]) Приведенное ниже описание следует работе Э. Шредингера (Schrödinger E., Probability problems in nuclear chemistry, Proceedings of the Royal Irish Academy, 51, sect. A, No. 1 (December 1945)). В этой работе пространственная однородность не предполагается.

ружающей среды обусловливают сходство братьев, что противоречит нашему предположению о стохастической независимости. Нашу модель можно уточнить, чтобы учесть эти замечания, но наиболее существенные ее черты изменить не удается. Мы найдем вероятность того, что в n-м поколении будет ровно k обладателей данной фамилии, в частности вероятность исчезновения фамилии. По-видимому, выживание фамилий — первая цепная реакция, изучавшаяся вероятностными методами. Впервые эту задачу рассматривал Ф. Гальтон в 1889 г.; подробности можно найти в книге А. Лотка [1]). Лотка указывает, что американские статистические данные хорошо соответствуют распределению $p_0=0{,}4825$, $p_k=(0{,}2126)\times\times(0{,}5893)^{k-1}$ $(k\geqslant1)$, которое (если отвлечься от первого члена), является геометрическим.

в) *Гены и мутации*. Каждый ген данного организма (см. гл. V, 5) может быть передан 0, 1, 2, ... непосредственным потомкам, и наша схема описывает этот процесс, пренебрегая, конечно, неоднородностями по популяции и по времени. Особенно полезна эта схема при изучении мутаций, т. е. изменений структуры гена. Случайная мутация приводит к образованию одного гена нового типа, который играет роль частицы нулевого поколения. Теория дает оценки для вероятностей сохранения и распространения мутантного гена. Чтобы пояснить основные идеи, рассмотрим (следуя Р. Э. Фишеру) растение кукурузы, которое является отцовским для 100 семян и материнским для стольких же семян. Если размер популяции не изменяется, то в среднем два из этих 200 семян разовьются в новые растения. Каждое зерно получает данный ген с вероятностью 1/2. Вероятность того, что мутантный ген будет иметься ровно у k новых растений, соответствует поэтому вероятности ровно k успехов в 200 испытаниях Бернулли с вероятностью успеха $p=1/200$, и представляется разумным предположить, что $\{p_k\}$ близко к распределению Пуассона со средним 1. Если мутантный ген обеспечивает какие-либо биологические преимущества, то получается распределение Пуассона со средним $\lambda>1$.

г) *Очереди* [2]). Теория ветвящихся процессов находит интересные применения в теории очередей. Например, покупатель, подошедший к свободному продавцу и сразу приступивший к выбору покупки, называется предком; его непосредственными потомками являются покупатели, пришедшие за время его обслуживания и образовавшие очередь. Этот процесс продолжается до тех пор, пока очередь не будет исчерпана. Мы подробнее рассмотрим этот процесс в примере 5, б), а его более интересный вариант — в примере 5, в). ▶

[1]) Lotka A., Théorie analytique des associations biologiques, v. 2, Actualités scientifiques et industrielles, No. 780 (1939), Paris, Hermann et Cie, 123—136.

[2]) Kendall D. G., Some problems in the theory of queues, J. Roy. Statist. Soc. (Series B), 13 (1951), 151—173; обсуждение с. 173—185.

§ 4. ВЕРОЯТНОСТИ ВЫРОЖДЕНИЯ ВЕТВЯЩИХСЯ ПРОЦЕССОВ

Обозначим через $\mathbf{Z}_n$ размер n-го поколения, а через P_n — производящую функцию его распределения. По предположению $\mathbf{Z}_0 = 1$ и

$$P_1(s) = P(s) = \sum_{k=0}^{\infty} p_k s^k. \qquad (4.1)$$

Можно разделить n-е поколение на $\mathbf{Z}_1$ семейств по их предкам в первом поколении. Это означает, что $\mathbf{Z}_n$ есть сумма $\mathbf{Z}_1$ случайных величин $\mathbf{Z}_n^{(k)}$, каждая из которых представляет собой размер потомства одной частицы первого поколения. По предположению каждая из величин $\mathbf{Z}_n^{(k)}$ имеет то же распределение вероятностей, что $\mathbf{Z}_{n-1}$, и (при фиксированном n) случайные величины $\mathbf{Z}_n^{(k)}$ взаимно независимы. Поэтому производящая функция P_n является сложной функцией:

$$P_n(s) = P(P_{n-1}(s)). \qquad (4.2)$$

Это равенство позволяет последовательно находить все производящие функции. Согласно (4.2), мы имеем $P_2(s) = P(P(s))$, $P_3(s) = P(P_2(s))$ и т. д. Эти вычисления не представляют трудностей, хотя явные выражения для P_n, как правило, найти нелегко. Тем не менее мы вскоре увидим, что из (4.2) можно получить важные следствия.

Пример. Предположим, что число непосредственных потомков имеет геометрическое распределение $\{qp^k\}$, где $p \neq q$. Тогда $P(s) = q/(1-ps)$, и, вычислив в явном виде P_2, P_3 и т. д. (и приложив известные усилия), мы приходим к общей формуле

$$P_n(s) = q \cdot \frac{p^n - q^n - (p^{n-1} - q^{n-1})\,ps}{p^{n+1} - q^{n+1} - (p^n - q^n)\,ps}. \qquad (4.3)$$

Легко проверить, что (4.3) действительно удовлетворяет (4.2).

Если $p=q$, то, полагая $p \to 1/2$, получаем

$$P_n(s) = [n-(n-1)s]/(n+1-ns). \qquad (4.4)$$

Заметим, что $P_n(0) \to q/p$, если $p > q$, но $P_n(0) \to 1$, если $p \leqslant q$. Теперь мы обсудим этот результат и найдем его аналог для произвольных распределений $\{p_k\}$. ▶

Самый первый вопрос, относящийся к нашему ветвящемуся процессу, состоит в том, будет процесс продолжаться бесконечно или же все потомство вымрет после конечного числа поколений. Положим

$$x_n = \mathbf{P}\{\mathbf{Z}_n = 0\} = P_n(0). \qquad (4.5)$$

Это вероятность того, что процесс окончится на n-м поколении или еще раньше. По определению $x_1 = p_0$, и из (4.2) следует, что

$$x_n = P(x_{n-1}). \qquad (4.6)$$

Ввиду тривиальности крайних случаев $p_0=0$ и $p_0=1$ мы далее будем предполагать, что $0<p_0<1$. Из монотонности P выводим, что $x_2=P(p_0)>P(0)=x_1$, и аналогично по индукции, что $x_1<x_2<x_3$. . . . Значит, эта последовательность имеет предел $x\leqslant 1$, и из (4.6) вытекает, что

$$x=P(x). \tag{4.7}$$

При $0\leqslant s\leqslant 1$ график $P(s)$ — *выпуклая вниз* кривая, начинающаяся в точке $(0, p_0)$ над биссектрисой первого квадранта и заканчивающаяся в точке $(1, 1)$ на этой биссектрисе. Таким образом, возможны только две ситуации.

Случай (i). Весь график находится выше биссектрисы. В этом случае $x=1$ — единственный корень уравнения (4.7), и поэтому $x_n\to 1$. Далее, в этом случае $1-P(s)\leqslant 1-s$ для всех s, и, устремляя s к 1, мы находим, что производная $P'(1)$ удовлетворяет неравенству $P'(1)\leqslant 1$.

Случай (ii). График P пересекает биссектрису в некоторой точке $\sigma<1$. Поскольку выпуклая кривая пересекает прямую линию не более чем в двух точках, в этом случае $P(s)>s$ при $s<\sigma$ и $P(s)<s$ при $\sigma<s<1$. Тогда $x_1=P(0)<P(\sigma)=\sigma$, и по индукции $x_n=P(x_{n-1})<<P(\sigma)=\sigma$. Значит, $x_n\to\sigma$, и, таким образом, $x=\sigma$. С другой стороны, по теореме о среднем значении между σ и 1 существует точка, в которой производная P' равна 1. Из монотонности этой производной следует, что $P'(1)>1$.

Итак, эти два случая характеризуются условиями $P'(1)\leqslant 1$ и $P'(1)>1$ соответственно. Но

$$\mu=P'(1)=\sum_{k=0}^{\infty} kp_k\leqslant\infty \tag{4.8}$$

есть *математическое ожидание числа непосредственных потомков*, т. е. нами доказана интересная теорема.

Теорема. *Если $\mu\leqslant 1$, то процесс вырождается с вероятностью* 1. *Если же $\mu>1$, то вероятность x_n того, что процесс вырождается в n-м поколении или раньше, стремится к единственному корню $x<1$ уравнения* (4.7).

Как правило, сходимость $x_n\to x$ является быстрой, и поэтому с большой вероятностью процесс либо довольно быстро вырождается, либо продолжается неограниченно. Математическое ожидание размера n-го поколения дается формулой $\mathbf{E}(\mathbf{Z}_n)=P_n'(1)$. Из (4.2) по правилу дифференцирования сложной функции мы получаем, что $P_n'(1)=P'(1)\,P_{n-1}'(1)=\mu\mathbf{E}(\mathbf{Z}_{n-1})$, и, следовательно[1]),

$$\mathbf{E}(\mathbf{Z}_n)=\mu^n. \tag{4.9}$$

[1]) Дальнейшие подробности можно найти в содержательной монографии Harris T. E., The theory of branching processes, Berlin, Springer, 1963. [Имеется перевод: Харрис Т. Е. Теория ветвящихся случайных процессов.— М.: Мир, 1966.] (См. также книгу Севастьянов Б. А. Ветвящиеся процессы.— М.: Наука, 1971.— *Перев.*)

Нет ничего удивительного в том, что процесс неизбежно вырождается при $\mu<1$, однако заранее не было ясно, что устойчивой ситуации не может быть даже при $\mu=1$[1]). Если $\mu>1$, то в соответствии с (4.9) следовало бы ожидать, что $\mathbf{Z}_n$ растет со скоростью геометрической прогрессии. До некоторой степени это верно; однако независимо от того, сколь велико μ, вероятность вырождения может быть отделена от 0. Легко проверить, что $P_n(s)\to x$ при любом $s<1$, а это значит, что коэффициенты при s, s^2, s^3 и т. д. стремятся к 0. Поэтому *с большой вероятностью после достаточно большого числа поколений либо потомков нет совсем, либо их очень много* (вероятности этих событий стремятся соответственно к x и $1-x$).

§ 5. ОБЩЕЕ ЧИСЛО ЧАСТИЦ В ВЕТВЯЩИХСЯ ПРОЦЕССАХ[2])

Рассмотрим теперь случайную величину

$$\mathbf{Y}_n=1+\mathbf{Z}_1+\ldots+\mathbf{Z}_n, \tag{5.1}$$

равную общему числу частиц в поколениях с номерами от 0 до n включительно. Полагая $n\to\infty$, мы получаем общее число частиц, которое может быть конечным или бесконечным. Очевидно, случайная величина $\mathbf{Y}_n$ корректно определена для каждого n; мы обозначим через R_n производящую функцию ее распределения вероятностей. Поскольку $\mathbf{Y}_1=1+\mathbf{Z}_1$, мы имеем $R_1(s)=sP(s)$. Рекуррентную формулу для R_n можно получить при помощи тех же рассуждений, что в предыдущем параграфе; единственное отличие состоит в том, что для получения $\mathbf{Y}_n$ мы должны прибавить к начальной частице сумму потомков $\mathbf{Z}_1$ частиц первого поколения. Следовательно,

$$R_n(s)=sP(R_{n-1}(s)). \tag{5.2}$$

По формуле (5.2) в принципе можно последовательно вычислять R_1, R_2, ..., но этим заниматься не стоит. Асимптотическое поведение R_n можно изучать при помощи тех же геометрических рассуждений, которые использовались в предыдущем параграфе при нахождении вероятности вырождения x.

Заметим прежде всего, что при любом $s<1$

$$R_2(s)=sP(R_1(s))<sP(s)=R_1(s), \tag{5.3}$$

и по индукции находим, что $R_n(s)<R_{n-1}(s)$. Таким образом, $R_n(s)$, монотонно убывая, стремится к пределу $\rho(s)$, который удовлетворяет уравнению

$$\rho(s)=sP(\rho(s)),\quad 0<s<1. \tag{5.4}$$

---

[1]) Если не обращать внимания на случай, когда число непосредственных потомков равно 1 с вероятностью 1.— *Прим. перев.*

[2]) Этот параграф написан под впечатлением работы И. Дж. Гуда (Good I. J., The number of individuals in a cascade process, Proc. Cambridge Philos. Soc., 45 (1949), 360—363).

Согласно теореме непрерывности гл. XI, 6, функция ρ как предел вероятностных производящих функций является производящей функцией такой последовательности ρ_k неотрицательных чисел, что $\sum \rho_k \leqslant 1$.

Из (5.4) следует, что при фиксированном $s<1$ значение $\rho(s)$ является корнем уравнения

$$t = sP(t). \tag{5.5}$$

Покажем, что этот корень является единственным. Для этого снова обозначим буквой x наименьший положительный [1]) корень уравнения $x=P(x)$ (так что $x \leqslant 1$). Заметим, что $y=sP(t)$ (при фиксированном s) является выпуклой вниз функцией t, и поэтому ее график пересекает прямую $y=t$ не более чем в двух точках. Но при $t=0$ правая часть (5.5) больше левой, тогда как при $t=x$ и при $t=1$ верно обратное неравенство; значит, (5.5) имеет ровно один корень между 0 и x и не имеет корней между x и 1. Таким образом, $\rho(s)$ как корень (5.5) определяется однозначно, и мы видим, далее, что $\rho(s)<x$. Но очевидно, $\rho(1)$ является корнем уравнения $t=P(t)$, и, поскольку x — наименьший корень этого уравнения, ясно, что $\rho(1)=x$. Иначе говоря, производящая функция ρ будет вероятностной тогда и только тогда, когда $x=1$. Мы можем подвести итоги этих рассуждений следующим образом.

Пусть ρ_k — вероятность того, что общее число частиц равно k.

а) $\sum \rho_k$ равна вероятности вырождения x (а $1-x$ есть вероятность того, что общее число частиц равно ∞).

б) Производящая функция $\rho(s)=\sum \rho_k s^k$ является единственным решением уравнения (5.5), и $\rho(s) \leqslant x$.

Мы уже знаем, что с вероятностью 1 *общее число частиц конечно, если* $\mu \leqslant 1$. Дифференцирование (5.4) теперь показывает, что *его математическое ожидание равно* $1/(1-\mu)$, *когда* $\mu<1$, *и бесконечно, когда* $\mu=1$.

Примеры. а) В примере 4,а) мы нашли, что $P(s)=q/(1-ps)$, и (5.5) сводится к квадратному уравнению $pt^2-t+qs=0$, из которого мы заключаем, что

$$\rho(s) = \left(1-\sqrt{1-4pqs}\right)/(2p). \tag{5.6}$$

(Эта производящая функция появлялась в связи с моментами первого достижения в гл. XI, 3.)

б) *Периоды занятости.* Займемся более подробным анализом задачи об очередях, упоминавшейся в примере 3, г). Предположим для простоты, что покупатели могут прибывать только по одному и только в целочисленные моменты времени. Мы предполагаем, что прибытия регулируются испытаниями Бернулли: в момент n поку-

[1]) При этом предполагается, что $P(0)>0$, так как при $P(0)=0$ случайная величина Y_n стремится к ∞ с вероятностью 1,— *Прим. перев.*

патель появляется с вероятностью p, а с вероятностью $q=1-p$ не появляется. Покупатель, который прибывает, когда продавец свободен, немедленно начинает обслуживаться; в противном случае он становится в очередь. Продавец работает без перерывов, пока в очереди есть покупатели, ожидающие обслуживания. Мы предполагаем, наконец, что последовательные времена обслуживания являются независимыми (целочисленными) случайными величинами с одним и тем же распределением $\{\beta_k\}$ и производящей функцией $\beta(s)=\sum\beta_k s^k$.

Допустим теперь, что покупатель появляется в нулевой момент времени и застает продавца свободным. Немедленно начинается обслуживание. Если оно имеет длительность n, то продавец освобождается в момент n при условии, что в моменты $1, 2, \ldots, n$ не появлялись новые покупатели. В противном случае обслуживание продолжается без перерыва. *Периодом занятости* мы называем длительность непрерывного обслуживания, начавшегося в нулевой момент времени. Мы покажем, как можно использовать теорию ветвящихся процессов при изучении длительности периода занятости.

Покупатель, появившийся в момент 0, начинает период занятости и будет называться предком. Первое поколение «частиц» состоит из покупателей, появившихся до момента окончания обслуживания предка или в этот момент. Если непосредственные потомки отсутствуют, то процесс вырождается. В противном случае непосредственные потомки последовательно обслуживаются, а появляющиеся за их времена обслуживания их непосредственные потомки присоединяются к очереди. Мы имеем здесь такой ветвящийся процесс, в котором *вероятность вырождения x равна вероятности того, что период занятости конечен, а множество всех «частиц» состоит из всех покупателей (включая предка), прибывших в течение периода занятости*. Необходимо отметить, что практически возможны только очереди с $x=1$.

Для применения наших результатов необходимо знать производящую функцию $P(s)$ числа непосредственных потомков. По определению это число определяется случайной суммой $\mathbf{X}_1+\ldots+\mathbf{X}_{\mathbf{N}}$, где $\mathbf{X}_j$ взаимно независимы и принимают значения 1 и 0 с вероятностями p и q, а $\mathbf{N}$ — время обслуживания предка. Поэтому в рассматриваемой ситуации $P(s)=\beta(ps+q)$, и, следовательно, $\mu=p\sigma$, где $\sigma=\beta'(1)$ — математическое ожидание длительности обслуживания. Значит, *период занятости с вероятностью 1 конечен только при $p\sigma\leqslant 1$. Математическое ожидание числа покупателей, прибывающих за время периода занятости, конечно только при $p\sigma<1$*. Иначе говоря, при $p\sigma=1$ неминуема давка, а если $p\sigma$ несущественно меньше 1, то длинные очереди должны быть обычным явлением.

в) *Длительность периода занятости.* В предыдущем примере изучалось число покупателей, прибывающих за время периода занятости, однако для практики большой интерес представляет длительность этого периода. Ее можно найти при помощи изящного приема [1]), состоящего в рассмотрении целочисленных моментов времени как частиц ветвящегося процесса. Будем говорить, что момент времени n не имеет потомков, если в этот момент не появляется покупатель. Если же покупатель появляется и его обслуживание начинается в момент $m \geqslant n$ и длится r единиц времени, то моменты времени $m+1, \ldots, m+r$ рассматриваются как непосредственные потомки момента n. Предположим, что в момент времени 0 продавец свободен. Несложное рассуждение показывает, что тогда либо покупатель не появляется и ветвящийся процесс не начинается вообще, либо покупатель появляется и участвующие в ветвящемся процессе частицы (моменты времени) образуют период занятости. Производящая функция числа непосредственных потомков имеет вид

$$P(s) = q + p\beta(s). \tag{5.7}$$

Корень x равен вероятности того, что период занятости конечен. Общее число частиц с вероятностью q равно 1 и с вероятностью p равно длительности периода занятости, начавшегося в момент 0. Очевидно, длительность периода занятости имеет производящую функцию $\beta(\rho(s))$. ▶

§ 6. ЗАДАЧИ

1. Распределение (1.1) случайной суммы $\mathbf{S_N}$ имеет среднее $\mathbf{E(N)\,E(X)}$ и дисперсию $\mathbf{E(N)\,Var(X)+Var(N)\,E^2(X)}$. Убедиться в этом: а) при помощи производящих функций, б) непосредственно из определения при помощи условных математических ожиданий.

2. *Отлов животных* (пример 1, б)). Доказать следующие утверждения. Если $\{g_n\}$ — геометрическое распределение, то искомое распределение тоже будет геометрическим. Если $\{g_n\}$ — логарифмическое распределение (см. (2.8)), то искомое распределение будет логарифмическим с добавочным членом.

3. Если случайная величина $\mathbf{N}$ имеет распределение Пуассона, то в $\mathbf{N}$ испытаниях Бернулли числа успехов и неудач стохастически независимы. Обобщить этот результат на полиномиальное распределение: а) непосредственно, б) используя производящие функции от нескольких переменных. (Ср. с примером гл. IX, 1, г).)

4. *Рандомизация*. Пусть $\mathbf{N}$ имеет распределение Пуассона с параметром λ, и пусть $\mathbf{N}$ шаров случайно размещаются по n ящикам. Показать без вычислений, что вероятность получить при этом ровно m пустых ящиков равна $\binom{n}{m} e^{-\lambda m/n}(1-e^{-\lambda/n})^{n-m}$.

5. *Продолжение* [2]). Показать, что если по n ящикам случайно размещается фиксированное число r шаров, то вероятность получить ровно m пустых ящиков равна коэффициенту при $e^{-\lambda}\lambda^r/r!$ в разложении приведенного выше выражения в ряд. а) Обсудить связь с производящими функциями моментов (задача 24 гл. XI, 7). б) Использовать этот результат для того, чтобы без всяких усилий получить соотношение (11.7) гл. II.

[1]) Этот прием предложил И. Дж. Гуд; см. материалы дискуссии, приведенные после статьи Кендалла, цитированной в примере 3, г).

[2]) Этот изящный способ получения разнообразных комбинаторных формул с помощью рандомизации параметра предложил Домб (Domb C., On the use of a random parameter in combinatorial problems, Proceedings Physical Society, Sec. A, 65 (1952), 305—309).

6. *Смеси вероятностных распределений*. Пусть $\{f_i\}$ и $\{g_i\}$ — два вероятностных распределения, $\alpha > 0$, $\beta > 0$, $\alpha+\beta=1$. Тогда $\{\alpha f_i+\beta g_i\}$ тоже является вероятностным распределением. Обсудить смысл этого преобразования и связь с урновыми схемами гл. V, 2. Обобщить на случай многих распределений. Показать, что такая смесь может быть обобщенным распределением Пуассона.

7. При помощи производящих функций показать, что в ветвящемся процессе $\mathrm{Var}(\mathbf{Z}_{n+1})=\mu\,\mathrm{Var}(\mathbf{Z}_n)+\mu^{2n}\sigma^2$. Используя условные математические ожидания, доказать эквивалентное соотношение $\mathrm{Var}(\mathbf{Z}_{n+1})=\mu^2\,\mathrm{Var}(\mathbf{Z}_n)+\mu^n\sigma^2$. Из каждого из этих соотношений вывести равенство $\mathrm{Var}(\mathbf{Z}_n)=\sigma^2(\mu^{2n-2}+\mu^{2n-3}+\ldots+\mu^{n-1})$.

8. *Продолжение*. Показать, что если $n > m$, то $\mathbf{E}(\mathbf{Z}_n\mathbf{Z}_m)=\mu^{n-m}\mathbf{E}(\mathbf{Z}_m^2)$.

9. *Продолжение*. Показать, что производящей функцией совместного распределения $(\mathbf{Z}_m, \mathbf{Z}_n)$ является $P_m(s_1P_{n-m}(s_2))$. Использовать этот факт для проверки утверждения из задачи 8.

10. Как изменится ветвящийся процесс, если каждая частица перед размножением может исчезнуть с фиксированной вероятностью p?

11. *Ветвящиеся процессы с двумя типами частиц*. Допустим, что каждая частица может иметь потомков двух типов; распределения чисел потомков частиц двух типов задаются производящими функциями $P_1(s_1, s_2)$ и $P_2(s_1, s_2)$. В этом случае существует две вероятности вырождения x, y в зависимости от типа предка. Показать, что пара (x, y) удовлетворяет уравнениям

$$x=P_1(x, y), \quad y=P_2(x, y). \tag{6.1}$$

Доказать, что эти уравнения имеют не более одного решения $0\leqslant x\leqslant 1$, $0\leqslant y\leqslant 1$, отличного от $(1, 1)$. Решение $(1, 1)$ единственно тогда и только тогда, когда $\mu_{11}\leqslant 1$, $\mu_{22}\leqslant 1$ и $(1-\mu_{11})(1-\mu_{22})\geqslant\mu_{12}\mu_{21}$, где $\mu_{ij}=\partial P_i(1, 1)/\partial s_j$.

ГЛАВА XIII РЕКУРРЕНТНЫЕ СОБЫТИЯ. ТЕОРИЯ ВОССТАНОВЛЕНИЯ

§ 1. НЕФОРМАЛЬНОЕ ВВЕДЕНИЕ И ПРИМЕРЫ

Мы будем изучать некоторые повторяющиеся, или рекуррентные, наборы исходов, связанные с последовательностями испытаний. Говоря нестрого, набор $\mathscr{E}$ годится для последующей теории, если после каждого появления набора $\mathscr{E}$ испытания начинаются заново в том смысле, что последовательность испытаний, следующих за появлением $\mathscr{E}$, образует копию всей последовательности. Промежутки времени между последовательными появлениями $\mathscr{E}$ — взаимно независимые одинаково распределенные случайные величины.

В простейшем частном случае $\mathscr{E}$ используется как обозначение «успеха» в последовательности испытаний Бернулли. Время ожидания первого успеха имеет геометрическое распределение; когда достигается первый успех, испытания начинаются заново и число испытаний между r-м и $(r+1)$-м успехами имеет то же самое геометрическое распределение. Время ожидания r-го успеха является суммой r независимых случайных величин (пример гл. IX, 3, в)). Эта ситуация повторяется, когда $\mathscr{E}$ обозначает «успех, за которым следует неудача»: появление набора *УН* восстанавливает исходное положение и время до появления следующего *УН* не зависит от исходов предыдущих испытаний. Приведем противоположный пример: пусть к группе людей последовательно присоединяются новые члены, и пусть $\mathscr{E}$ — событие «у двух человек в группе дни рождения совпадают». Такое событие $\mathscr{E}$ не является повторяющимся, потому что после первого его появления оно происходит постоянно. Если мы заменим определение на «$\mathscr{E}$ происходит, когда день рождения нового члена группы уже встречался в ней», то $\mathscr{E}$ может происходить любое число раз, но после появления $\mathscr{E}$ процесс *не* начинается заново. Это происходит потому, что с увеличением размера группы совпадения дней рождения становятся более вероятными, и поэтому большое время ожидания первого повторения дней рождения сочетается с более коротким временем ожидания второго повторения; таким образом, последовательные времена ожидания не являются независимыми и имеют неодинаковые распределения.

Большое значение теории рекуррентных событий объясняется тем, что часто такие события возникают в связи с различными последовательностями случайных величин (случайными процессами). Сложность законов, которым подчиняется последовательность случайных величин, может сделать невозможным ее исчерпывающий анализ, однако существование повторяющихся событий всегда позволяет указать характерные особенности последовательности, доказать существование некоторых пределов и т. п. Такой подход способствует упрощению и стандартизации многих исследований.

Мы рассмотрим теперь несколько типичных примеров; некоторые из них представляют и самостоятельный интерес. Первые примеры связаны с обычными испытаниями Бернулли, последние три касаются более сложных схем. В их описаниях мы используем такие термины, как «продавец» и «покупатель», но каждый раз приводится математическое определение последовательности случайных величин, которое является полным в том смысле, что оно однозначно определяет вероятности всех возможных событий. Часто оказывается, что теория рекуррентных событий позволяет получать содержательные результаты даже в тех случаях, когда основные вероятности не могут быть вычислены в явном виде.

Примеры. а) *Возвращение в начало.* Пусть для последовательности испытаний Бернулли $\mathcal{E}$ обозначает событие «суммарные числа успехов и неудач равны». Как и ранее, мы описываем исходы испытаний последовательностью $\mathbf{X}_1, \mathbf{X}_2, \ldots$ взаимно независимых случайных величин, принимающих значения 1 и -1 с вероятностями p и q соответственно. Как обычно полагаем

$$\mathbf{S}_0=0, \quad \mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n. \tag{1.1}$$

Тогда $\mathbf{S}_n$ — это разность между числом успехов и числом неудач, и событие $\mathcal{E}$ *происходит тогда и только тогда, когда* $\mathbf{S}_n=0$. Совершенно ясно, что наступление этого события восстанавливает исходную ситуацию в том смысле, что последующие частные суммы $\mathbf{S}_{n+1}, \mathbf{S}_{n+2}, \ldots$ образуют вероятностную копию всей последовательности $\mathbf{S}_1, \mathbf{S}_2, \ldots$. (Продолжение см. в примере 4, б).)

б) *Возвращение в начало по отрицательным значениям.* Усложним последний пример: будем считать, что $\mathcal{E}$ происходит при n-м испытании, если

$$\mathbf{S}_n=0, \text{ но } \mathbf{S}_1<0, \ldots, \mathbf{S}_{n-1}<0. \tag{1.2}$$

Очевидно, что и в этом случае наступление события $\mathcal{E}$ означает, что мы начинаем все сначала. (Продолжение см. в примере 4, в).)

в) Другим вариантом примера а) является событие $\mathcal{E}$, состоящее в том, что суммарное число успехов в λ раз больше суммарного числа неудач (где $\lambda>0$ — произвольное, но фиксированное число). Если $\mathcal{E}$ происходит при n-м испытании, то оно происходит еще раз при $(n+m)$-м испытании только тогда, когда число успехов в испытаниях с номерами от $n+1$ до $n+m$ ровно в λ раз больше числа неудач. Промежутки между последовательными появлениями $\mathcal{E}$ являются поэтому независимыми и одинаково распределенными. В качестве частного случая можно рассмотреть событие, состоящее в том, что при $6n$ бросаниях идеальной кости единица появится ровно n раз. (Продолжение см. в задачах 4 и 5.)

г) *Лестничные величины.* Сохраняя обозначения примера а), мы введем новое повторяющееся событие $\mathscr{E}$, *которое происходит при n-м испытании, если* $\mathbf{S}_n$ *больше всех предыдущих сумм.* т. е. если

$$\mathbf{S}_n > 0, \quad \mathbf{S}_n > \mathbf{S}_1, \ \ldots, \ \mathbf{S}_n > \mathbf{S}_{n-1}. \tag{1.3}$$

Если $\mathscr{E}$ происходит при n-м испытании, то процесс начинается заново в следующем смысле. При условии, что (1.3) выполнено, событие $\mathscr{E}$ происходит при $(n+m)$-м испытании тогда и только тогда, когда

$$\mathbf{S}_{n+m} > \mathbf{S}_n, \ \ldots, \ \mathbf{S}_{n+m} > \mathbf{S}_{n+m-1}. \tag{1.4}$$

Но разности $\mathbf{S}_{n+k} - \mathbf{S}_n$ — это частные суммы оставшейся последовательности $\mathbf{X}_{n+1}, \mathbf{X}_{n+2}, \ldots$, и поэтому новое появление $\mathscr{E}$ для этой оставшейся последовательности определяется точно так же, как событие $\mathscr{E}$ определялось для всей последовательности. Иначе говоря, при изучении события $\mathscr{E}$ все прошлое становится несущественным каждый раз, когда происходит $\mathscr{E}$. (Продолжение см. в примере 4, г).)

д) *Серии успехов в испытаниях Бернулли.* В предыдущих примерах определение события $\mathscr{E}$ не вызывало трудностей; теперь мы рассмотрим случай, когда применение теории рекуррентных событий становится возможным только при весьма осмотрительном выборе определения. Обычно слова «серия успехов длины r» используются для обозначения отрезка последовательности, состоящего ровно из r или не менее чем из r успехов. Ни одно из этих толкований не соответствует рекуррентному событию. В самом деле, если требуется наличие ровно r успехов, то успех в $(n+1)$-м испытании может разрушить серию, завершившуюся к n-му испытанию. С другой стороны, если требуется наличие не менее чем r успехов, то любая серия может продолжаться неограниченно, и ясно, что появление серии не восстанавливает исходного положения.

Классическая теория серий была довольно беспорядочной, и более систематический подход становится возможным, если определять серию длины r так, чтобы она стала рекуррентным событием. *Первая* серия длиной r определена однозначно, и мы условимся начинать наблюдения заново каждый раз, когда появляется серия. При таком соглашении последовательность *УУУ|УНУУУ|УУУ|Н* содержит три серии успехов длины 3 (появляющиеся при испытаниях с номерами 3, 8 и 11). Она содержит пять серий длины 2 (испытания с номерами 2, 4, 7, 9, 11). Формальное определение таково: *последовательность из n букв У и Н содержит столько У-серий длины r, сколько существует непересекающихся отрезков, состоящих в точности из r букв У каждый.* При таком соглашении мы говорим, что $\mathscr{E}$ происходит при n-м испытании, если к последовательности прибавляется новая серия длиной r. Это определяет рекуррентное событие и значительно упрощает теорию, не изменяя ее основных результатов. (Продолжение см. в § 7.)

е) *Продолжение: близкие события.* Очевидно, что рассуждения, проведенные в предыдущем примере, применимы к более общим событиям, например к появлению последовательности *УНУН*. Более интересным является то, что ограничение одним фиксированным набором вовсе не обязательно. Например, появление «двух успехов и трех неудач» определяет повторяющееся событие, как и появление «либо серии успехов длины r, либо серии неудач длины ρ». (Продолжение см. в § 8.)

ж) *Счетчики Гейгера.* Работу счетчиков, используемых при регистрации космических лучей и α-частиц, можно описать следующей упрощенной моделью[1]). Испытания Бернулли проводятся с постоянной скоростью. Счетчик предназначен для регистрации успехов, однако его механизм после каждой регистрации запирается ровно на $r-1$ испытаний. Иначе говоря, успех в n-м испытании регистрируется тогда и только тогда, когда в предыдущих $r-1$ испытаниях успехи не регистрировались. После этого счетчик запирается до тех пор, пока не закончатся испытания с номерами $n, \ldots, n+r-1$, и «освобождается» после $(n+r)$-го испытания, если оно приводит к неудаче. Последовательность на выходе счетчика представляет собой пример зависимых испытаний. За каждой регистрацией следует «мертвое время», однако в моменты, когда счетчик «свободен» (не заперт), состояния счетчика идентичны, и испытания начинаются заново. Обозначая буквой $\mathscr{E}$ событие «после испытания счетчик свободен», мы получаем типичное рекуррентное событие. (Продолжение см. в примере 4,д).)

з) *Простейшая модель очереди* строится по последовательности испытаний Бернулли и последовательности случайных величин $\mathbf{X}_1, \mathbf{X}_2, \ldots$, принимающих только целые положительные значения. Величины $\mathbf{X}_k$ имеют одно и то же распределение $\{\beta_m\}$, взаимно независимы и не зависят от испытаний Бернулли. Успех в n-м испытании мы интерпретируем как появление в момент времени n покупателя (или вызова на телефонной станции). Случайная величина $\mathbf{X}_k$ равна времени обслуживания k-го покупателя. В любой момент времени продавец либо «свободен», либо «занят», и процесс развивается по следующим правилам. Вначале (в момент времени 0) продавец свободен. Если покупатель появляется, когда продавец свободен, то обслуживание начинается немедленно, и после появления этого покупателя продавец занят в течение времени его обслуживания. Покупатели, которые приходят, когда продавец занят, образуют очередь. Продавец обслуживает покупателей непрерывно до тех пор, пока очередь не рассосется.

Эти правила определяют процесс однозначно, и по заданным

---

[1]) Это дискретный аналог так называемых счетчиков первого типа. Счетчики второго типа описаны в задаче 8. [Более подробную информацию о теории счетчиков можно найти в книге Кокс Д., Смит В., Теория восстановления.— М.: Советское радио, 1967.— *Перев.*]

варианту $\{У, Н, У, У, У, Н, Н, \ldots\}$ процесса прибытия покупателей и варианту $\{3, 1, 17, 2, \ldots\}$ последовательных времен обслуживания нетрудно найти размер очереди в любой момент времени и время ожидания k-го покупателя. Таким образом, в принципе мы могли бы вычислить все интересующие нас вероятности, однако найти практичные методы вычислений нелегко. Ясно, однако, что каждый раз, когда продавец свободен, ситуация является в точности такой же, как в нулевой момент времени. Поэтому в нашей терминологии условие «продавец свободен» задает рекуррентное событие. Мы увидим, что само существование таких рекуррентных событий имеет важные последствия; например, оно означает, что распределения вероятностей размера очереди в момент n, времени ожидания n-го покупателя и аналогичных случайных величин стремятся к некоторым пределам, когда $n \to \infty$ (теорема 5.2). Иначе говоря, существование рекуррентных событий позволяет доказывать существование стационарного режима и изучать его основные свойства.

и) *Обслуживание станков.* Представление об области применимости метода рекуррентных событий может дать вариант предыдущего примера, в котором появления покупателей определяются уже не испытаниями Бернулли. Интерпретируем «покупателей» как одинаковые станки, подверженные случайным поломкам, а «продавца» — как ремонтного рабочего. Мы сохраним предположения о порядке обслуживания и образования очереди, но введем новый случайный механизм «появления покупателей», т. е. поломок. Предположим, что общее число станков равно N, и рассмотрим два крайних случая.

i) Предположим сначала, что если станок исправен, то с фиксированной вероятностью p он может сломаться в следующий целочисленный момент времени; когда станок ломается, он заменяется точно таким же новым станком, и время обслуживания — это время, затрачиваемое на установку нового станка. Мы считаем, что станки независимы и что их поломки определяются N независимыми последовательностями испытаний Бернулли. Заметим, что чем больше станков находится в очереди, тем меньше станков исправно, и следовательно, длина очереди в любой момент времени влияет на вероятность новых поломок (или заявок на ремонт). В этом состоит существенное отличие от предыдущего примера, однако условие «продавец свободен» тем не менее определяет рекуррентное событие, поскольку независимо от момента его наступления оно означает наличие одной и той же ситуации.

ii) Допустим теперь, что каждый ремонт обладает последействием, а именно увеличивает вероятность последующих поломок. Это означает, что станки постепенно изнашиваются, и поэтому после поломки хотя бы одного станка повторение благоприятной исходной ситуации оказывается невозможным. В этом случае рекуррентных событий, облегчающих анализ, не существует. ►

§ 2. ОПРЕДЕЛЕНИЯ

Рассмотрим последовательность повторяющихся испытаний с возможными исходами E_j $(j=1, 2, \ldots)$. Испытания не обязаны быть независимыми (наиболее интересными оказываются применения к цепям Маркова). Как обычно, мы предполагаем, что испытания могут продолжаться неограниченно и что вероятности $\mathbf{P}\{E_{j_1}, E_{j_2}, \ldots, E_{j_n}\}$ определяются однозначно для всех конечных наборов. Пусть $\mathscr{E}$ — некоторое свойство конечных последовательностей, т. е. мы предполагаем, что для любого набора $(E_{j_1}, \ldots, E_{j_n})$ можно сказать, обладает он свойством $\mathscr{E}$ или нет. Будем понимать выражение «$\mathscr{E}$ происходит на n-м месте в (конечной или бесконечной) последовательности $E_{j_1}, E_{j_2}, \ldots$» как синоним слов «подпоследовательность $E_{j_1}, E_{j_2}, \ldots, E_{j_n}$ обладает свойством $\mathscr{E}$». Эта договоренность означает, что появление $\mathscr{E}$ при n-м испытании зависит только от исходов первых n испытаний. Подразумевается также, что, *говоря о «рекуррентном событии $\mathscr{E}$», мы на самом деле имеем в виду класс событий*, определенных условием «$\mathscr{E}$ происходит». Ясно, что $\mathscr{E}$ является скорее набором символов, чем событием. Здесь мы несколько вольно обращаемся с языком, так же как, например, при использовании термина «двумерная задача»: задача сама по себе не имеет размерности.

Определение 1. *Свойство $\mathscr{E}$ определяет рекуррентное событие, если:*

а) *для того чтобы $\mathscr{E}$ происходило на n-м и $(n+m)$-м местах последовательности $(E_{j_1}, E_{j_2}, \ldots, E_{j_{n+m}})$, необходимо и достаточно, чтобы $\mathscr{E}$ происходило на последнем месте каждой из двух подпоследовательностей $(E_{j_1}, E_{j_2}, \ldots, E_{j_n})$ и $(E_{j_{n+1}}, E_{j_{n+2}}, \ldots, E_{j_{n+m}})$;*

б) *если $\mathscr{E}$ происходит на n-м месте, то*

$$\mathbf{P}\{E_{j_1}, \ldots, E_{j_{n+m}}\} = \mathbf{P}\{E_{j_1}, \ldots, E_{j_n}\} \mathbf{P}\{E_{j_{n+1}}, \ldots, E_{j_{n+m}}\}.$$

Теперь приобретают очевидный смысл утверждения о том, что $\mathscr{E}$ *впервые* происходит в последовательности $(E_{j_1}, E_{j_2}, \ldots)$ на n-м месте, и т. п. Ясно также, что с каждым рекуррентным событием $\mathscr{E}$ связаны две последовательности чисел, определенные для $n=1, 2, \ldots$ следующим образом:

$$\begin{aligned} u_n &= \mathbf{P}\{\mathscr{E} \text{ происходит при } n\text{-м испытании}\}, \\ f_n &= \mathbf{P}\{\mathscr{E} \text{ впервые происходит при } n\text{-м испытании}\}. \end{aligned} \tag{2.1}$$

Для дальнейшего удобно положить

$$f_0 = 0, \quad u_0 = 1 \tag{2.2}$$

и ввести производящие функции

$$F(s) = \sum_{k=1}^{\infty} f_k s^k, \quad U(s) = \sum_{k=0}^{\infty} u_k s^k. \tag{2.3}$$

Заметим, что $\{u_k\}$ не является распределением вероятностей; более того, в типичных случаях $\sum u_k = \infty$. С другой стороны, события «$\mathscr{E}$ впервые происходит при n-м испытании» несовместны, и поэтому

$$f = F(1) = \sum_{n=1}^{\infty} f_n \leqslant 1. \qquad (2.4)$$

Ясно, что $1 - f$ следует интерпретировать как *вероятность того, что $\mathscr{E}$ ни разу не происходит в продолженной до бесконечности последовательности испытаний*. Если $f = 1$, то мы можем ввести случайную величину $\mathbf{T}$ с распределением

$$\mathbf{P}\{\mathbf{T} = n\} = f_n. \qquad (2.5)$$

Обозначение (2.5) мы будем использовать и при $f < 1$. В этой ситуации $\mathbf{T}$ *оказывается несобственной* (*или дефектной*) *случайной величиной, которая с вероятностью* $1 - f$ *не принимает никакого численного значения*. (Нам будет удобно приписывать $\mathbf{T}$ значение ∞, и должно быть ясно, что при этом не потребуется вводить новые правила.)

Время ожидания события $\mathscr{E}$, т. е. число испытаний до первого появления $\mathscr{E}$ (включая это испытание), является случайной величиной с распределением (2.5); однако эта случайная величина естественно определяется только на пространстве бесконечных последовательностей $(E_{j_1}, E_{j_2}, \ldots)$.

По определению рекуррентных событий вероятность того, что $\mathscr{E}$ впервые произойдет при k-м испытании и *во второй раз* при n-м испытании, равна $f_k f_{n-k}$. Поэтому вероятность $f_n^{(2)}$ того, что $\mathscr{E}$ происходит во второй раз при n-м испытании, равна

$$f_n^{(2)} = f_1 f_{n-1} + f_2 f_{n-2} + \ldots + f_{n-1} f_1. \qquad (2.6)$$

Правая часть — свертка $\{f_n\}$ с собой, и поэтому $\{f_n^{(2)}\}$ есть распределение вероятностей суммы двух независимых случайных величин, имеющих распределение (2.5). Вообще, если $f_n^{(r)}$ — вероятность того, что r-е появление $\mathscr{E}$ происходит при n-м испытании, то

$$f_n^{(r)} = f_1 f_{n-1}^{(r-1)} + f_2 f_{n-2}^{(r-1)} + \ldots + f_{n-1} f_1^{(r-1)}. \qquad (2.7)$$

Этот простой факт выражает следующая теорема.

Теорема. *Пусть* $f_n^{(r)}$ — *вероятность того, что* r*-е появление* $\mathscr{E}$ *происходит при* n*-м испытании. Тогда* $\{f_n^{(r)}\}$ — *распределение вероятностей суммы*

$$\mathbf{T}^{(r)} = \mathbf{T}_1 + \mathbf{T}_2 + \ldots + \mathbf{T}_r \qquad (2.8)$$

r *независимых случайных величин* $\mathbf{T}_1, \ldots, \mathbf{T}_r$, *имеющих распределение* (2.5). *Иначе говоря, при фиксированном* r *последовательность* $\{f_n^{(r)}\}$ *имеет производящую функцию* $F^r(s)$.

Отсюда следует, в частности, что

$$\sum_{n=1}^{\infty} f_n^{(r)} = F^r(1) = f^r. \tag{2.9}$$

Иначе говоря, *вероятность того, что $\mathscr{E}$ произойдет по меньшей мере r раз, равна f^r* (это можно было предвидеть). Введем теперь следующее определение.

Определение 2. *Рекуррентное событие $\mathscr{E}$ будем называть возвратным*[1]), *если $f=1$, и невозвратным, если $f<1$.*

Для невозвратного события $\mathscr{E}$ вероятность f^r того, что оно произойдет по меньшей мере r раз, стремится к нулю, тогда как для возвратного $\mathscr{E}$ эта вероятность остается равной единице. Это можно описать словесно: *с вероятностью единица возвратное событие $\mathscr{E}$ происходит бесконечно много раз, а невозвратное событие $\mathscr{E}$ происходит лишь конечное число раз.* (Это утверждение является не только описательным, но и формально правильным при интерпретации в терминах пространства элементарных событий, состоящего из бесконечных последовательностей $E_{j_1}, E_{j_2}, \ldots$.)

Нам потребуется еще одно определение. При испытаниях Бернулли возвращение в начало (см. пример 1, а)) может произойти только при испытании с *четным* номером. В таком случае $f_{2n+1} = u_{2n+1} = 0$ и производящие функции $F(s)$ и $U(s)$ являются степенными рядами не по s, а по s^2. Аналогично, если в примере 1, в) число λ целое, то $\mathscr{E}$ может произойти при n-м испытании только при условии, что n делится на $\lambda+1$. В таких случаях мы будем называть $\mathscr{E}$ периодическим. По существу периодические рекуррентные события лишь обозначениями отличаются от непериодических, но в каждой теореме приходится специально упоминать об исключительном периодическом случае. Таким образом, периодические рекуррентные события доставляют много неприятностей, не внося ничего интересного.

Определение 3. *Рекуррентное событие $\mathscr{E}$ называется периодическим, если существует такое целое число $\lambda>1$, что $\mathscr{E}$ может происходить только при испытаниях с номерами λ, 2λ, $3\lambda, \ldots$ (т. е. $u_n=0$, если n не делится на λ). Наибольшее λ, обладающее этим свойством, называется периодом $\mathscr{E}$.*

В заключение заметим, что на пространстве элементарных событий, состоящем из бесконечных последовательностей $E_{j_1}, E_{j_2}, \ldots$, число испытаний между $(r-1)$-м и r-м появлениями $\mathscr{E}$ является корректно определенной случайной величиной (может быть, несобственной) с тем же распределением, что $\mathbf{T}_r$. Иначе говоря, наши случайные величины $\mathbf{T}_r$ действительно соответствуют *про-*

[1]) В первом издании использовались термины «достоверное» (certain) и «недостоверное» (uncertain) события; введенные здесь термины предпочтительнее в применениях к цепям Маркова.

межуткам времени между последовательными появлениями $\mathscr{E}$ *(временам возвращения).* Мы определили $\mathbf{T}_r$ аналитически, чтобы не использовать такие пространства элементарных событий, которые в этом томе не рассматриваются; будем надеяться, однако, что вероятностная подоплека показана во всей ее интуитивной простоте. Понятие рекуррентных событий предназначено для сведения весьма общей ситуации к суммам независимых случайных величин. Обратно, *любое распределение вероятностей* $\{f_n\}$, $n=1, 2, \ldots$, *можно использовать для определения рекуррентного события.* Это утверждение доказывает следующий пример.

Пример. *Самовосстанавливающиеся устройства.* Рассмотрим электрическую лампочку, плавкий предохранитель или иной элемент с конечным сроком службы. Как только этот элемент отказывает, он заменяется новым элементом того же вида, который рано или поздно заменяется третьим элементом, и т. д. Допустим, что срок службы — случайная величина, принимающая только значения, кратные единичному отрезку времени (году, дню или секунде). Тогда каждая единица времени представляет собой испытание с возможными исходами «замена» или «нет замены». Последовательные замены можно истолковывать как рекуррентные события. Если f_n — вероятность того, что новый элемент проработает ровно n единиц времени, то $\{f_n\}$ — распределение времени возвращения. Если срок службы с вероятностью 1 конечен, то $\sum f_n=1$ и рекуррентное событие является возвратным. Как правило, известно, что срок службы не может превосходить фиксированного числа m; в таком случае производящая функция $F(s)$ — многочлен степени не выше m. В приложениях нужно знать вероятность u_n того, что в момент n происходит замена. Значение u_n можно вычислить с помощью (3.1) (см. ниже). Мы получили здесь класс рекуррентных событий, определяемых произвольным распределением $\{f_n\}$. Случай $f<1$ не исключается: $1-f$ — это вероятность бесконечного срока службы нашего элемента. ►

§ 3. ОСНОВНЫЕ СООТНОШЕНИЯ

Мы продолжаем пользоваться обозначениями (2.1) — (2.4) и собираемся изучить связь между $\{f_n\}$ и $\{u_n\}$. Вероятность того, что $\mathscr{E}$ происходит впервые при ν-м испытании, а затем снова при испытании с номером $n>\nu$, по определению равна $f_\nu u_{n-\nu}$. Вероятность того, что $\mathscr{E}$ происходит при n-м испытании впервые, есть $f_n=f_n u_0$. Поскольку эти случаи взаимно несовместны, мы имеем

$$u_n=f_1u_{n-1}+f_2u_{n-2}+\ldots+f_nu_0, \quad n\geqslant 1. \tag{3.1}$$

В выражении справа мы узнаем свертку $\{f_k\}*\{u_k\}$, производящая функция которой равна $F(s)\,U(s)$. В левой части мы обнаруживаем последовательность $\{u_n\}$ без члена u_0, так что ее производящая функция есть $U(s)-1$. Таким образом, $U(s)-1=F(s)U(s)$, и нами доказана следующая теорема.

Теорема 1. *Производящие функции* $\{u_n\}$ *и* $\{f_n\}$ *связаны соотношением*

$$U(s)=1/(1-F(s)). \tag{3.2}$$

Замечание. Правую часть (3.2) можно разложить в геометрический ряд $\sum F^r(s)$, сходящийся при $|s|<1$. Поскольку коэффициент $f_n^{(r)}$ при s^n в $F^r(s)$ равен вероятности r-го появления $\mathscr{E}$ при n-м испытании, соотношение (3.2) эквивалентно равенству

$$u_n=f_n^{(1)}+f_n^{(2)}+\dots, \tag{3.3}$$

выражающему тот очевидный факт, что если $\mathscr{E}$ происходит при n-м испытании, то до момента n оно происходило 0, 1, 2, ... или $n-1$ раз. (Очевидно, $f_n^{(r)}=0$ при $r>n$.)

Теорема 2. *Для того чтобы событие* $\mathscr{E}$ *было невозвратным, необходимо и достаточно, чтобы величина*

$$u=\sum_{j=0}^{\infty} u_j \tag{3.4}$$

была конечной. В этом случае вероятность того, что $\mathscr{E}$ *происходит хотя бы раз, равна*

$$f=(u-1)/u. \tag{3.5}$$

Замечание. Мы можем интерпретировать u_j как математическое ожидание случайной величины, которая равна 1 или 0 в зависимости от того, происходит или не происходит $\mathscr{E}$ при j-м испытании. Следовательно, $u_1+u_2+\ldots+u_n$ есть *математическое ожидание числа появлений* $\mathscr{E}$ *в* n *испытаниях, а* $u-1$ можно интерпретировать как математическое ожидание числа появлений $\mathscr{E}$ в бесконечной последовательности испытаний.

Доказательство. Поскольку коэффициенты u_k неотрицательны, ясно, что $U(s)$ монотонно возрастает при $s\to 1$ и что при любом N

$$\sum_{n=0}^{N} u_n \leqslant \lim_{s\to 1} U(s) \leqslant \sum_{n=0}^{\infty} u_n = u.$$

Так как $U(s)\to(1-f)^{-1}$, когда $f<1$, и $U(s)\to\infty$, когда $f=1$, то теорема доказана. ▶

Следующая теорема чрезвычайно важна [1]). Ее доказательство

[1]) В частных случаях ее утверждение легко доказывается (см. задачу 1) и давно известно. Огромное количество работ было посвящено ослаблению ее условий, но все были уверены в том, что какие-то ограничения необходимы. В полной общности теорема 3 была доказана Эрдёшем, Феллером и Поллардом (Erdős P.,

элементарно, но мы отложим его до конца главы, поскольку оно не помогает понять вероятностную основу теоремы [1]).

Теорема 3. *Пусть событие $\mathscr{E}$ возвратно и непериодично; обозначим через μ математическое ожидание времени возвращения $\mathbf{T}_\nu$, т. е.*

$$\mu = \sum jf_j = F'(1) \tag{3.6}$$

(допускается случай $\mu = \infty$). Тогда

$$u_n \to \mu^{-1} \quad при \quad n \to \infty \tag{3.7}$$

($u_n \to 0$, если математическое ожидание времени возвращения бесконечно).

От условия непериодичности $\mathscr{E}$ легко избавиться. Действительно, если $\mathscr{E}$ имеет период λ, то ряд $\sum f_n s^n$ содержит только степени s^λ. Будем называть степенной ряд правильным, если это не так для любого целого $\lambda > 1$. Теорему 3 можно тогда переформулировать следующим образом: *если F — правильная вероятностная производящая функция и U определяется соотношением* (3.2), *то $u_n \to 1/F'(1)$*. Далее, если $\mathscr{E}$ имеет период λ, то $F(s^{1/\lambda})$ — правильная вероятностная производящая функция, и поэтому коэффициенты $U(s^{1/\lambda})$ сходятся к $\lambda/F'(1)$. Таким образом, нами доказана следующая теорема.

Теорема 4. *Если $\mathscr{E}$ возвратно и имеет период λ, то*

$$u_{n\lambda} \to \lambda/\mu \tag{3.8}$$

и $u_k = 0$ при любом k, не делящемся на λ.

§ 4. ПРИМЕРЫ

а) *Успехи в испытаниях Бернулли.* Рассмотрим банальный пример, когда $\mathscr{E}$ обозначает «успех» в последовательности испытаний Бернулли. Тогда $u_n = p$ при $n \geqslant 1$, откуда

$$U(s) = \frac{1-qs}{1-s}, \quad и\ поэтому \quad F(s) = \frac{ps}{1-qs} \tag{4.1}$$

Feller W., Pollard H., A property of power series with positive coefficients, Bull. Amer. Math. Soc., 55 (1949), 201—204). После выхода в свет первого издания Чжун Кайлай заметил, что теорему можно вывести из результатов Колмогорова, описывающих асимптотическое поведение цепей Маркова. Многие известные математики доказывали различные обобщения этой теоремы для разных классов распределений вероятностей. Эти исследования способствовали развитию методов современной теории вероятностей . В конце концов оказалось, что аналог теоремы 3 справедлив для любых распределений вероятностей. Элементарное (но не простое) доказательство см. в гл. XI, 9 тома 2.

[1]) Вероятностное доказательство этой теоремы (с явной оценкой скорости сходимости при $\mu < \infty$) можно найти в работе Калашников В. В. Равномерная оценка скорости сходимости в теореме восстановления для дискретного времени.— Теория вероятн. и ее примен., 1977, т. XXII, № 1, 399—403.— *Прим. перев.*

в силу (3.2). В этом частном случае теорема 2 попросту подтверждает очевидный факт: промежутки времени между последовательными успехами имеют геометрическое распределение с математическим ожиданием $1/p$.

б) *Возвращения в начало (пример* 1,а)). При k-м испытании суммарные числа успехов и неудач могут быть равны только тогда, когда $k=2n$ четно, и в этом случае вероятность возвращения в начало есть

$$u_{2n}=\binom{2n}{n}p^n q^n=\binom{-1/2}{n}(-4pq)^n. \tag{4.2}$$

Из разложения бинома (формула (8.7) гл. II) следует поэтому, что

$$U(s)=1/\sqrt{1-4pqs^2}, \tag{4.3}$$

а отсюда и из (3.2) — что

$$F(s)=1-\sqrt{1-4pqs^2}. \tag{4.4}$$

Повторное использование разложения бинома приводит к явному выражению для f_{2n}. (Явные выражения для u_{2n} и f_{2n} при $p=1/2$ выведены комбинаторными методами в гл. III, 2—3; производящие функции U и F найдены другими способами в гл. XI, 3. Следует отметить, что лишь предложенный выше метод не требует изобретательности.)

При $s=1$ квадратный корень в (4.4) равен $|p-q|$, и поэтому

$$f=1-|p-q|. \tag{4.5}$$

Таким образом, *возвращение в начало представляет собой рекуррентное событие с периодом* 2, *которое невозвратно при* $p\neq q$ *и возвратно в симметричном случае* $p=q$. Вероятность по меньшей мере r возвращений в начало равна f^r.

Когда $p=q=1/2$, время ожидания первого возвращения в начало является собственной [1]) случайной величиной, но $F'(1)=\infty$, и поэтому *среднее время возвращения* μ *бесконечно*. (Это следует также из теоремы 4 и из того, что $u_n\to 0$.) Бесконечность среднего времени возвращения означает, что случайные флуктуации в длительной реализации игры с бросанием монеты существенно отличаются от обычной модели, описываемой нормальным распределением. Весьма парадоксальный характер этих флуктуаций обсуждался в гл. III.

в) *Возвращение в начало по отрицательным значениям*. В примере 1, б) возвращение в начало удовлетворяло условию, состоящему в том, что ни одна из предшествующих частных сумм $\mathbf{S}_j$ не положительна. Распределение времени возвращения для этого ре-

---

[1]) Случайная величина называется *собственной* (proper), если она с вероятностью 1 принимает конечные значения (см. формулы (2.4), (2.5)).— *Прим. перев.*

куррентного события определяется равенством

$$f_{2n}^{-}=\mathbf{P}\{S_{2n}=0,\ S_1<0, \ldots, S_{2n-1}<0\}, \tag{4.6}$$

и, конечно, $f_{2n-1}^{-}=0$. Кажется, что найти эти вероятности прямым рассуждением невозможно, однако их легко вычислить с помощью предыдущего примера. В самом деле, реализация последовательности $(\mathbf{X}_1, \ldots, \mathbf{X}_{2n})$, удовлетворяющая условию (4.6), содержит n положительных единиц и n отрицательных единиц и, значит, имеет ту же вероятность, что $(-\mathbf{X}_1, \ldots, -\mathbf{X}_{2n})$. Далее, *первое* возвращение в начало происходит либо по положительным, либо по отрицательным значениям, и мы заключаем, что эти две возможности имеют одинаковые вероятности. Следовательно, $f_{2n}^{-}=(1/2)f_{2n}$, где $\{f_n\}$ — распределение возвращений в начало, найденное в предыдущем примере. Поэтому производящая функция наших времен возвращения задается формулой

$$F^{-}(s)=1/2-(1/2)\sqrt{1-4pqs^2}, \tag{4.7}$$

и, значит,

$$U^{-}(s)=\frac{2}{1+\sqrt{1-4pqs^2}}=\frac{1-\sqrt{1-4pqs^2}}{2pqs^2}. \tag{4.8}$$

Событие $\mathscr{E}$ невозвратно, вероятность его наступления равна $1/2-(1/2)|p-q|$.

г) *Лестничные величины.* Первая положительная частная сумма может появиться при k-м испытании только тогда, когда $k=2n+1$ нечетно. Соответствующие вероятности мы запишем в виде

$$\varphi_{2n+1}=\mathbf{P}\{S_1<0, \ldots, S_{2n}=0,\ S_{2n+1}=1\}. \tag{4.9}$$

Таким образом, $\{\varphi_k\}$ — распределение рекуррентного события из примера 1, г). Далее, из условия (4.9) следует, что $\mathbf{X}_{2n+1}=1$ и что при $2n$-м испытании происходит рекуррентное событие из предыдущего примера. Значит, $\varphi_{2n+1}=p\cdot u_{2n}^{-}$. Поэтому в очевидных обозначениях

$$\Phi(s)=psU^{-}(s)=(1-\sqrt{1-4pqs^2})/(2qs). \tag{4.10}$$

Это производящая функция времени первого достижения, найденная в гл. XI (формула (3.6)). Явное выражение для φ_{2n+1} можно вывести из (4.10) с помощью разложения бинома (гл. II, формула (8.7)). Оно совпадает с выражением для φ_{2n+1}, найденным комбинаторными методами в теореме 2 гл. III, 7.

д) *Счетчики Гейгера.* В примере 1, ж) счетчик остается свободным, если в момент времени 1 не происходит регистрации. В противном случае он запирается и вновь «освобождается» в момент времени $r+1$, если в этот момент не появляется частица; счетчик «освобождается» в момент времени $2r+1$, если частица появляется в момент времени $r+1$, но не появляется в момент

$2r+1$, и т. д. Производящей функцией времени возвращения является поэтому

$$qs+qps^{r+1}+qp^2s^{2r+1}+\ldots=qs/(1-ps^r). \tag{4.11}$$

(См. также задачи 7—9.)

е) *Простейшая задача теории очередей* (пример 1, з)). В этом случае продавец остается свободным, если в момент времени 1 не появляется покупатель. Если же покупатель появляется, то начинается так называемый «период занятости», который заканчивается в момент, когда продавец впервые освобождается. Производящая функция $\rho(s)$ периода занятости выводилась в примере гл. XII, 5, в) при помощи методов теории ветвящихся процессов. Нетрудно вывести, что в нашем случае производящая функция времени возвращения имеет вид $qs+ps\rho(s)$, что согласуется с формулой (5.7) гл. XII.

ж) *Ничьи в играх с бросанием нескольких монет.* В заключение приведем простой пример, демонстрирующий возможность получить некоторые выводы, не зная явного вида производящих функций. Пусть $r\geqslant 2$ — произвольное целое число; рассмотрим последовательность одновременных независимых бросаний r монет. Пусть $\mathscr{E}$ обозначает рекуррентное событие, состоящее в том, что *все r монет находятся в одной и той же фазе* (т. е. что суммарные числа выпадения гербов одинаковы для всех r монет). Вероятность наступления этого события при n-м испытании есть

$$u_n=2^{-rn}\left[\binom{n}{0}^r+\binom{n}{1}^r+\ldots+\binom{n}{n}^r\right]. \tag{4.12}$$

В правой части мы узнаем члены биномиального распределения с $p=1/2$; применяя к последнему нормальное приближение, легко заключаем[1]), что при любом фиксированном r и $n\longrightarrow\infty$

$$u_n\sim[2/(\pi n)]^{r/2}\sum_j e^{-2rj^2/n} \tag{4.13}$$

---

[1]) Нормальное приближение $2^{-n}\binom{n}{k}\sim\left(\frac{2}{\pi n}\right)^{1/2}e^{-2(k-n/2)^2/n}$ справедливо, когда $n\longrightarrow\infty$ и $(k-n/2)^3/n^2\longrightarrow 0$ (см. гл. VII, 2). Из этого приближения следует, например, что

$$\sum_{k:\ (k-n/2)^{12}\leqslant n^7} 2^{-rn}\binom{n}{k}^r\sim\sum_{j:\ j^{12}\leqslant n^7}\left(\frac{2}{\pi n}\right)^{r/2}e^{-2rj^2/n},\qquad n\longrightarrow\infty.$$

Для доказательства (4.13) остается воспользоваться соотношением (4.14) и соотношениями

$$\sum_{k:\ (k-n/2)^{12}>n^7} 2^{-rn}\binom{n}{k}^r<n2^{-rn}\binom{n}{n/2+[n^{7/12}]}^r\sim n\left(\frac{2}{\pi n}\right)^{r/2}e^{-2n^{1/6}},$$

$$2\sqrt{\frac{r}{n}}\sum_{j:\ j^{12}>n^7}e^{-2rj^2/n}<2\int_{-\infty}^{-n^{7/12}+1}e^{-x^2/2}\,dx\longrightarrow 0,\qquad n\longrightarrow\infty.$$

— *Прим. перев.*

(суммирование проводится по всем целым числам от $-n/2$ до $n/2$). Но по определению интеграла

$$2\sqrt{r/n}\sum_j e^{-2rj^2/n} \to \int_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}, \tag{4.14}$$

и отсюда мы выводим, что

$$u_n \sim (1/\sqrt{r})\,[2/(\pi n)]^{(r-1)/2}. \tag{4.15}$$

Следовательно, $\sum u_n$ расходится, когда $r \leqslant 3$, но сходится, когда $r \geqslant 4$. Значит, $\mathscr{E}$ *возвратно при* $r \leqslant 3$ *и невозвратно при* $r \geqslant 4$. Поскольку $u_n \to 0$, при $r \leqslant 3$ среднее время возвращения бесконечно. (Ср. с задачами 2 и 3.) ▶

§ 5. РЕКУРРЕНТНЫЕ СОБЫТИЯ С ЗАПАЗДЫВАНИЕМ. ОБЩАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Введем теперь незначительное обобщение понятия рекуррентного события, столь очевидное, что о нем не стоило бы и упоминать, если бы не было удобно иметь для него специальный термин и перечень основных уравнений.

Возможно, наилучшее неформальное описание рекуррентных событий с запаздыванием получится, если сказать, что они относятся к испытаниям, в которых мы «пропускаем начало и начинаем с середины». Время ожидания *первого* появления $\mathscr{E}$ имеет распределение $\{b_n\}$, отличающееся от распределения $\{f_n\}$ времени между последовательными появлениями $\mathscr{E}$. Теория переносится на этот случай без изменений, за исключением того, что испытания, следующие за каждым появлением $\mathscr{E}$, являются точными вероятностными копиями одной и той же последовательности случайных величин, которая не совпадает с исходной последовательностью.

Поскольку ситуация столь проста, мы воздержимся от формальностей и согласимся говорить о *рекуррентном событии* $\mathscr{E}$ *с запаздыванием, когда определение рекуррентных событий применимо лишь при игнорировании испытаний, предшествующих первому появлению* $\mathscr{E}$; *подразумевается, что время ожидания первого появления* $\mathscr{E}$ *есть случайная величина, не зависящая от последующих времен возвращения, но ее распределение* $\{b_n\}$ *может отличаться от общего распределения* $\{f_n\}$ *времен возвращения.*

Обозначим через v_n *вероятность появления* $\mathscr{E}$ *при* n*-м испытании.* При выводе выражения для v_n мы рассуждаем следующим образом. Допустим, что $\mathscr{E}$ происходит при испытании с номером $k<n$. По отношению к последующим испытаниям $\mathscr{E}$ оказывается обычным рекуррентным событием, и поэтому (условная) вероятность повторного появления $\mathscr{E}$ при n-м испытании равна u_{n-k}. Далее, если $\mathscr{E}$ происходит при n-м испытании, то либо это первое его появление, либо первое появление имело место при k-м испы-

тании с некоторым $k<n$. Суммируя по всем вариантам, получаем

$$v_n = b_n + b_{n-1}u_1 + b_{n-2}u_2 + \ldots + b_1 u_{n-1} + b_0 u_n. \qquad (5.1)$$

Таким образом, у нас есть возможность найти явное выражение для v_n. (Другое доказательство приводится в примере 10, а).) Соотношение (5.1) можно переписать в компактном виде уравнения свертки:

$$\{v_n\} = \{b_n\} * \{u_n\}. \qquad (5.2)$$

Это означает, что соответствующие производящие функции удовлетворяют тождеству

$$V(s) = B(s)\,U(s) = B(s)/(1 - F(s)). \qquad (5.3)$$

Пример. а) В испытаниях Бернулли, рассмотренных в примерах 4, а)—4, г), событие $\mathbf{S}_n = 1$ является рекуррентным событием с запаздыванием. Время ожидания его первого появления имеет производящую функцию Φ, заданную формулой (4.10); промежутки времени между последовательными появлениями события $\{\mathbf{S}_n = 1\}$ имеют производящую функцию F возвращений в начало (см. (4.4)). Таким образом, в этом случае $V = \Phi/(1 - F)$.

Легко показать, что асимптотическое поведение вероятностей v_n по существу такое же, как у u_n. Во избежание тривиальных оговорок предположим, что $\mathscr{E}$ непериодично [1]. Согласно § 3, в этом случае u_n стремится к конечному пределу и $\sum u_n < \infty$ тогда и только тогда, когда $\mathscr{E}$ невозвратно.

Теорема 1. *Если* $u_n \to \omega$, *то*

$$v_n \to b\omega, \quad \text{где } b = \sum b_k = B(1). \qquad (5.4)$$

Если $\sum u_n = u < \infty$, *то*

$$\sum v_n = bu. \qquad (5.5)$$

В частности, $v_n \to \mu^{-1}$, *если* $\mathscr{E}$ *возвратно.*

Доказательство. Положим $r_k = b_{k+1} + b_{k+2} + \ldots$. Поскольку $u_n \leqslant 1$, из (5.1) следует, что при $n > k$

$$b_0 u_n + \ldots + b_k u_{n-k} \leqslant v_n \leqslant b_0 u_n + \ldots + b_k u_{n-k} + r_k. \qquad (5.6)$$

Выберем k столь большим, чтобы $r_k < \varepsilon$. Тогда для достаточно больших n левая часть (5.6) больше $b\omega - 2\varepsilon$, а правая часть меньше $b\omega + 2\varepsilon$. Тем самым (5.4) доказано. Утверждение (5.5) можно доказывать, либо суммируя (5.1) по n, либо полагая $s = 1$ в (5.3). ►

Мы переходим теперь к предельной теореме, имеющей широкую область приложений. Допустим, что некоторая система имеет счет-

[1] Периодические рекуррентные события рассматриваются в теореме 2 § 10. Другое доказательство теоремы 1 приводится в примере 10, а).

ное множество возможных состояний E_0, $E_1, \ldots$ и что переходы из одного состояния в другое зависят от какого-то случайного механизма. Например, в простейшей модели очереди (пример 1,з)) мы говорим, что система находится в состоянии E_k, если в очереди находится k покупателей (вместе с покупателем, которого обслуживают). Чтобы задать состояние системы, включающей семнадцать продавцов, может потребоваться восемнадцать чисел, однако все мыслимые состояния могут быть упорядочены в последовательность E_0, $E_1, \ldots$. Нам не нужно думать о том, как это сделать наилучшим образом, потому что следующая теорема не дает конкретных методов оценки вероятностей. Она является чистой теоремой существования, показывающей, что стационарный режим существует в большинстве встречающихся на практике случаев. Это представляет теоретический интерес, но имеет также и практическое значение, поскольку, как правило, математическое исследование стационарного режима значительно проще, чем изучение зависящего от времени процесса.

Мы предполагаем, что для $n=1, 2, \ldots$ и для любого набора $(r_1, \ldots, r_n)$ определена вероятность того, что система в моменты времени $0, 1, \ldots, n-1$ проходит через состояния $E_{r_1}, \ldots, E_{r_n}$. Мы не делаем никаких предположений ни о взаимной зависимости этих событий, ни о вероятностях переходов из одного состояния в другое. Для простоты мы рассмотрим только *вероятности* $p_n^{(r)}$ *того, что в момент времени* n *система находится в состоянии* E_r. (Будет ясно, как теорема обобщается на пары состояний, тройки состояний и т. д.) Основное предположение состоит в том, что существует связанное с нашим процессом рекуррентное событие $\mathcal{E}$. Например, в модели очереди (пример 1, з)) таким рекуррентным событием является попадание в состояние E_0. Если бы в этом случае $\mathcal{E}$ было невозвратным, то была бы положительной вероятность того, что очередь никогда не окончится. Это значило бы, что раньше или позже мы столкнулись бы с бесконечной очередью, т. е. с очередью неограниченно возрастающего размера. Наша предельная теорема показывает, что такие системы на практике невозможны. Пример с очередью должен был пояснить роль условия возвратности события $\mathcal{E}$. (Условие непериодичности вводится только для того, чтобы избежать тривиальных оговорок.)

Теорема 2. *Допустим, что существует связанное с нашим процессом непериодическое возвратное рекуррентное событие* $\mathcal{E}$ *(возможно, с запаздыванием). Тогда при* $n \to \infty$

$$p_n^{(r)} \longrightarrow p^{(r)}, \tag{5.7}$$

причем

$$\sum p^{(r)} = 1, \tag{5.8}$$

если среднее время возвращения μ конечно, и $p^r=0$ в противном случае[1]).

Доказательство. Каждый раз, когда происходит $\mathscr{E}$, процесс начинается заново. Поэтому существует корректно определенная условная вероятность $g_n^{(r)}$ того, что если $\mathscr{E}$ происходит в некоторый момент времени, то состояние E_r появляется через n единиц времени, и это происходит *до* следующего наступления $\mathscr{E}$ (здесь $n=0, 1, \ldots$). Для рекуррентных событий с запаздыванием нам потребуется также вероятность $\gamma_n^{(r)}$ того, что E_r появляется в момент времени n *до* первого наступления $\mathscr{E}$. (Очевидно, $\gamma_n^{(r)}=g_n^{(r)}$, если $\mathscr{E}$ — рекуррентное событие без запаздывания.)

Классифицируем теперь способы появления E_r в момент n по моментам последнего появления $\mathscr{E}$ до момента n. Прежде всего, $\mathscr{E}$ может еще не произойти. Вероятность этого равна $\gamma_n^{(r)}$. В противном случае существует такое $k\leqslant n$, что $\mathscr{E}$ происходит в момент времени k и не происходит между k и n. Вероятность этого равна $v_k g_{n-k}^{(r)}$. Суммируя по всем этим взаимно исключающим случаям, находим

$$p_n^{(r)}=\gamma_n^{(r)}+g_{n-1}^{(r)}v_1+g_{n-2}^{(r)}v_2+\ldots+g_0^{(r)}v_n. \tag{5.9}$$

(Мы пользуемся здесь обозначениями теоремы 1. Для событий с запаздыванием $v_0=0$, для событий без запаздывания $v_k=u_k$ и $\gamma_n^{(r)}=g_n^{(r)}$.)

Соотношение (5.9) аналогично (5.1) и отличается лишь наличием члена $\gamma_n^{(r)}$ в правой части. Очевидно, эта величина не больше вероятности того, что $\mathscr{E}$ не произойдет до момента n, и, поскольку $\mathscr{E}$ возвратно, $\gamma_n^{(r)}\to 0$ при $n\to\infty$. К остальным слагаемым мы можем применить теорему 1, заменив u_k на v_k и b_k на $g_k^{(r)}$. Так как $\mathscr{E}$ возвратно, то $v_n\to\mu^{-1}$, и поэтому

$$p_n^{(r)}\to\mu^{-1}\sum_{k=0}^{\infty}g_k^{(r)}. \tag{5.10}$$

Это доказывает существование пределов (5.7). Чтобы доказать, что их сумма равна 1, заметим, что в любой момент времени система находится в каком-то состоянии, и, значит,

$$\sum_{r=0}^{\infty}g_n^{(r)}=g_n \tag{5.11}$$

есть вероятность того, что время возвращения не меньше n, т. е.

$$g_n=f_n+f_{n+1}+\ldots$$

[1]) Последнее утверждение теоремы, вообще говоря, неверно (кстати, автор его и не доказывает). Контрпримером является последовательность испытаний Бернулли с $p=q=1/2$ (процесс с двумя состояниями) и рекуррентное событие $\mathscr{E}$, состоящее в равенстве суммарных чисел успехов и неудач (см. пример 4, б)). — *Прим. перев.*

Таким образом,

$$\sum_{r=0}^{\infty} p^{(r)} = \mu^{-1} \sum_{n=0}^{\infty} g_n = 1 \qquad (5.12)$$

в силу формулы (1.8) гл. XI. ►

(Предельную теорему из примера 10, б) можно рассматривать как частный случай только что доказанной теоремы.)

§ 6. ЧИСЛО ПОЯВЛЕНИЙ $\mathscr{E}$

До сих пор мы изучали свойства рекуррентного события $\mathscr{E}$, относящиеся к промежуткам времени между его последовательными появлениями. Часто оказывается предпочтительнее считать заданным число n испытаний, а в качестве основной случайной величины выбрать *число* $\mathbf{N}_n$ *появлений* $\mathscr{E}$ в первых n испытаниях. Мы исследуем теперь асимптотическое поведение распределения $\mathbf{N}_n$ при больших n. Для простоты будем считать, что $\mathscr{E}$ — рекуррентное событие без запаздывания.

Пусть, как и в (2.8), $\mathbf{T}^{(r)}$ обозначает число испытаний до r-го появления $\mathscr{E}$, включая это последнее испытание. Распределения вероятностей $\mathbf{T}^{(r)}$ и $\mathbf{N}_n$ связаны очевидным тождеством

$$\mathbf{P}\{\mathbf{N}_n \geqslant r\} = \mathbf{P}\{\mathbf{T}^{(r)} \leqslant n\}. \qquad (6.1)$$

Начнем с простого случая, когда $\mathscr{E}$ возвратно и распределение $\{f_n\}$ его времени возвращения имеет конечные среднее μ и дисперсию σ^2. Так как $\mathbf{T}^{(r)}$ является суммой r независимых величин, то, согласно центральной предельной теореме из гл. X,1, для любого фиксированного x при $r \to \infty$

$$\mathbf{P}\{(\mathbf{T}^{(r)} - r\mu)/(\sigma\sqrt{r}) < x\} \to \mathfrak{N}(x), \qquad (6.2)$$

где $\mathfrak{N}(x)$ — функция нормального распределения. Устремим теперь n и r к бесконечности так, чтобы

$$(n - r\mu)/(\sigma\sqrt{r}) \to x; \qquad (6.3)$$

тогда из (6.1) и (6.2) будет следовать, что

$$\mathbf{P}\{\mathbf{N}_n \geqslant r\} \to \mathfrak{N}(x). \qquad (6.4)$$

Чтобы записать это соотношение в более привычном виде, мы введем *нормированную случайную величину*

$$\mathbf{N}_n^* = (\mu\mathbf{N}_n - n)\sqrt{\mu/(\sigma^2 n)}. \qquad (6.5)$$

Неравенство $\mathbf{N}_n \geqslant r$ эквивалентно неравенству

$$\mathbf{N}_n^* \geqslant [(r\mu - n)/(\sigma\sqrt{r})]\sqrt{r\mu/n} = -x\sqrt{r\mu/n}. \qquad (6.6)$$

Разделив (6.3) на r, получим, что $n/r \to \mu$, и поэтому правая часть (6.6) стремится к $-x$. Так как $\mathfrak{R}(-x) = 1 - \mathfrak{R}(x)$, то

$$\mathbf{P}\{\mathbf{N}_n^* \geqslant -x\} \to \mathfrak{R}(x), \quad или \quad \mathbf{P}\{\mathbf{N}_n^* < -x\} \to 1 - \mathfrak{R}(x), \tag{6.7}$$

и нами доказана следующая теорема.

Теорема. (Нормальное приближение.) *Если рекуррентное событие $\mathscr{E}$ возвратно и его время возвращения имеет конечные среднее μ и дисперсию σ^2, то число $\mathbf{T}^{(r)}$ испытаний до r-го появления $\mathscr{E}$ и число $\mathbf{N}_n$ появлений $\mathscr{E}$ в первых n испытаниях имеют асимптотически нормальные распределения, указанные в* (6.2) *и* (6.7).

Заметим, что в (6.7) центральная предельная теорема применяется к последовательности *зависимых* случайных величин $\mathbf{N}_n$. Полезность нашей теоремы будет продемонстрирована в следующем параграфе на примере применений к сериям успехов.

Соотношения (6.7) делают правдоподобными формулы

$$\mathbf{E}(\mathbf{N}_n) \sim n/\mu, \quad \operatorname{Var}(\mathbf{N}_n) \sim n\sigma^2/\mu^3, \tag{6.8}$$

где символ $\sim$ означает, что отношение левой части к правой стремится к 1. Чтобы доказать (6.8), мы заметим, что $\mathbf{N}_n$ есть сумма n таких (зависимых) случайных величин $\mathbf{Y}_k$, что $\mathbf{Y}_k$ равна единице или нулю в зависимости от того, происходит или не происходит $\mathscr{E}$ в k-м испытании. Тогда $\mathbf{E}(\mathbf{Y}_k) = u_k$ и

$$\mathbf{E}(\mathbf{N}_n) = u_1 + u_2 + \ldots + u_n. \tag{6.9}$$

Поскольку $u_n \to \mu^{-1}$, отсюда следует первое соотношение (6.8). Второе доказывается аналогичными рассуждениями (см. задачу 20).

К сожалению, удивительно много времен возвращения, встречающихся в различных случайных процессах и в приложениях, имеют *бесконечное математическое ожидание*. В таких случаях нормальное приближение заменяется более общими предельными теоремами совершенно другого характера [1]) и случайные флуктуации обладают неожиданными свойствами. Например, интуиция подсказывает, что $\mathbf{E}(\mathbf{N}_n)$ должно расти линейно по n, «поскольку в среднем при удвоении числа испытаний число появлений $\mathscr{E}$ тоже должно удвоиться». Однако *это не так*. Если математическое ожидание времени возвращения бесконечно, то $u_n \to 0$, и поэтому $\mathbf{E}(\mathbf{N}_n)/n \to 0$ в силу (6.9). Это означает, что в длинной последовательности испытаний появления $\mathscr{E}$ становятся все реже и реже, и это возможно только за счет того, что некоторые времена возвращения фантастически велики. Следующие два примера показывают, насколько резко может быть выражено это явление.

Примеры. а) Если $\mathscr{E}$ — возвращение в начало в игре с бросанием монеты (пример 4, б) с $p = 1/2$), то $u_{2n} \sim 1/\sqrt{\pi n}$ и (6.9) яв-

[1]) Feller W., Fluctuation theory of recurrent events, Trans. Amer. Math. Soc., 67 (1949), 98—119.

ляется приближением к интегралу от $1/\sqrt{\pi x}$; отсюда следует, что $\mathbf{E}(\mathbf{N}_{2n}) \sim 2\sqrt{n/\pi}$. Поэтому *среднее* время возвращения в интервале от 0 до n растет примерно как $\sqrt{n}$. Любопытные следствия этого подробно обсуждались в гл. III.

б) Возвращаясь к примеру 4, ж), рассмотрим повторяющиеся бросания $r=3$ монет и обозначим через $\mathscr{E}$ событие, состоящее в том, что все три монеты находятся в одной и той же фазе. Мы видели, что $\mathscr{E}$ — возвратное рекуррентное событие и что $u_n \sim 2/(\sqrt{3}\pi n)$. Поэтому $\mathbf{E}(\mathbf{N}_n)$ *растет примерно как* $\log n$, и, значит, *среднее* времен возвращения до момента n имеет фантастическую величину порядка $n/\log n$. ▶

§ 7*). ПРИЛОЖЕНИЯ К ТЕОРИИ СЕРИЙ УСПЕХОВ

В дальнейшем r будет обозначать фиксированное положительное целое число, а $\mathscr{E}$ — появление серии из r успехов в последовательности испытаний Бернулли. Важно, чтобы длина серии определялась так же, как в примере 1, д), поскольку в противном случае серии не являются рекуррентными событиями и вычисления усложняются. Как в (2.1) и (2.2), u_n — *это вероятность появления* $\mathscr{E}$ *при* n-*м испытании, а* f_n — *вероятность того, что первая серия длиной* r *появляется при* n-*м испытании.*

Вероятность того, что r испытаний с номерами $n, n-1, n-2, \ldots, n-r+1$ закончатся успехами, очевидно, равна p^r. Тогда $\mathscr{E}$ происходит при одном из этих r испытаний; вероятность того, что $\mathscr{E}$ произойдет при испытании с номером $n-k$ ($k=0, 1, \ldots, r-1$), а следующие k испытаний приведут к k успехам, равна $u_{n-k}p^k$. Поскольку эти r возможностей исключают друг друга, мы получаем рекуррентное соотношение [1])

$$u_n+u_{n-1}p+\ldots+u_{n-r+1}p^{r-1}=p^r, \tag{7.1}$$

справедливое при $n \geqslant r$. Очевидно,

$$u_1=u_2=\ldots=u_{r-1}=0, \quad u_0=1. \tag{7.2}$$

Умножая (7.1) на s^n и суммируя по $n=r, r+1, r+2, \ldots$, получаем в левой части

$$\{U(s)-1\}(1+ps+p^2s^2+\ldots+p^{r-1}s^{r-1}), \tag{7.3}$$

а в правой $p^r(s^r+s^{r+1}+\ldots)$. Эти два ряда являются геометрическими, и мы находим, что

$$\{U(s)-1\}(1-(ps)^r)/(1-ps)=p^rs^r/(1-s), \tag{7.4}$$

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

[1]) Классический подход состоит в выводе рекуррентного соотношения для f_n. Этот метод сложнее и не применим, например, к сериям произвольного вида или к событиям *УУННУУ*, к которым наш метод применим без изменений (ср. с примером 8, в)).

или

$$U(s) = (1 - s + qp^r s^{r+1})/[(1-s)(1-p^r s^r)]. \qquad (7.5)$$

Используя (3.2), находим теперь *производящую функцию времени возвращения*:

$$F(s) = \frac{p^r s^r (1-ps)}{1-s+qp^r s^{r+1}} = \frac{p^r s^r}{1-qs(1+ps+\ldots+p^{r-1}s^{r-1})}. \qquad (7.6)$$

Равенство $F(1)=1$ показывает, что в удлиняющейся последовательности испытаний число серий любой длины с вероятностью 1 увеличивается неограниченно. Среднее время возвращения μ можно было бы получить непосредственно из (7.1), поскольку мы знаем, что $u_n \to \mu^{-1}$. Так как нам потребуется и дисперсия, удобнее вычислить производные $F(s)$. Лучше всего при этом дифференцировать, предварительно умножив обе части (7.6) на знаменатель. Несложные вычисления показывают, что *среднее и дисперсия времен возвращения для серий длиной r равны*

$$\mu = \frac{1-p^r}{qp^r}, \quad \sigma^2 = \frac{1}{(qp^r)^2} - \frac{2r+1}{qp^r} - \frac{p}{q^2} \qquad (7.7)$$

соответственно. По теореме предыдущего параграфа при больших *n число $\mathbf{N}_n$ серий длиной r в n испытаниях имеет распределение, близкое к нормальному*, т. е. при фиксированных $\alpha < \beta$ вероятность того, что

$$n/\mu + \alpha\sigma\sqrt{n/\mu^3} < \mathbf{N}_n < n/\mu + \beta\sigma\sqrt{n/\mu^3}, \qquad (7.8)$$

стремится к $\mathfrak{N}(\beta) - \mathfrak{N}(\alpha)$. Впервые это было доказано Мизесом при помощи весьма утомительных вычислений. В табл. 1 приведено несколько типичных значений математических ожиданий времен возвращения.

Таблица 1

Математические ожидания времен возвращения для серий успехов, когда производится по одному испытанию в секунду

Длина серии r	$p=0,6$	$p=0,5$ (монета)	$p=1/6$ (кость)
5	30,7 с	1 мин	2,6 ч
10	6,9 мин	34,1 мин	28,0 мес
15	1,5 ч	18,2 ч	18 098 лет
20	19 ч	24,3 сут	140,7 млн. лет

Метод разложения на простые дроби из гл. XI, 4 позволяет получить превосходные приближенные формулы. Второе представление в (7.6) показывает, что знаменатель имеет единственный *положительный корень* $s=x$. Для любого действительного или комплексного числа s с $|s|\leqslant x$ имеем

$$|qs(1+ps+\ldots+p^{r-1}s^{r-1})|\leqslant qx(1+px+\ldots+p^{r-1}x^{r-1})=1, \quad (7.9)$$

где знак равенства возможен только тогда, когда аргументы всех слагаемых в левой части одинаковы, т. е. когда $s=x$. Следовательно, x по абсолютной величине меньше любого другого корня знаменателя в (7.6). Поэтому мы можем применить формулы (4.5) и (4.9) гл. XI, положив в них $s_1=x$, $U(s)=p^r s^r(1-ps)$ и $V(s)=1-s+qp^r\times\times s^{r+1}$. Используя равенство $V(x)=0$, находим

$$f_n\sim\frac{(x-1)(1-px)}{(r+1-rx)\,q}\,\frac{1}{x^{n+1}}. \quad (7.10)$$

Вероятность отсутствия серий в n испытаниях есть $q_n=f_{n+1}+ +f_{n+2}+f_{n+3}+\ldots$, и, суммируя геометрическую прогрессию в (7.10), получаем

$$q_n\sim\frac{1-px}{(r+1-rx)\,q}\,\frac{1}{x^{n+1}}. \quad (7.11)$$

Таким образом, мы доказали, что *вероятность отсутствия серий успехов длиной r в n испытаниях удовлетворяет соотношению* (7.11). Табл. 2 показывает, что правая часть (7.11) дает удивительно хорошие приближения даже для очень малых n, и точность аппроксимации быстро увеличивается с ростом n. Это иллюстрирует эффективность метода производящих функций и разложения на простые дроби.

Таблица 2

Вероятность отсутствия серий успехов длиной $r=2$ в n испытаниях с $p=1/2$

n	q_n точное	q_n по (7.11)	Погрешность
2	0,75	0,76631	0,0163
3	0,625	0,61996	0,0050
4	0,500	0,50156	0,0016
5	0,40625	0,40577	0,0005

Численные оценки. Имея в виду интересы читателя-прикладника, мы покажем, что вычисления, связанные с разложением на простейшие дроби, часто оказываются проще, чем может показаться с первого взгляда, и что можно получить хорошие оценки для погрешности.

В связи с асимптотической формулой (7.11) возникают две задачи: во-первых, оценить вклад $r-1$ отброшенных корней и, во-вторых, получить оценки для главного корня x.

Первое представление в (7.6) показывает, что все корни знаменателя функции $F(s)$ удовлетворяют уравнению

$$s=1+qp^r s^{r+1}, \quad (7.12)$$

однако (7.12) имеет посторонний корень $s=p^{-1}$. При положительных s график функции $f(s)=1+qp^rs^{r+1}$ выпукл вниз; он пересекает биссектрису $y=s$ в x и в p^{-1}, а между x и p^{-1} лежит *ниже* биссектрисы. Далее, $f'(p^{-1})=(r+1)q$. Если эта величина больше 1, то график $f(s)$ пересекает биссектрису при $s=p^{-1}$ снизу, и, следовательно, $p^{-1}>x$. Мы будем предполагать для определенности, что

$$(r+1)q>1; \tag{7.13}$$

в этом случае $x<p^{-1}$ и $f(s)<s$ при $x<s<p^{-1}$. Значит, для всех таких комплексных чисел s, что $x<|s|<p^{-1}$, мы имеем $|f(s)|\leqslant f(|s|)<|s|$, так что нули s_k не могут лежать в кольце $x<|s|<p^{-1}$. Поскольку x выбирался как корень, наименьший по абсолютной величине, это означает, что

$$|s_k|\geqslant p^{-1}, \quad \textit{когда } s_k\neq x. \tag{7.14}$$

Дифференцирование (7.12) показывает теперь, что все корни являются простыми.

Вклад в q_n каждого корня аналогичен вкладу (7.11) главного корня x, и поэтому $r-1$ членов, не учтенных в (7.11), имеют вид

$$A_k=\frac{ps_k-1}{rs_k-(r+1)}\,\frac{1}{qs_k^{n+1}}. \tag{7.15}$$

Нам потребуется оценка сверху для первой дроби в правой части. Чтобы получить ее, заметим, что для фиксированного $s>p^{-1}>(r+1)r^{-1}$

$$\left|\frac{pse^{i\theta}-1}{rse^{i\theta}-(r+1)}\right|\leqslant\frac{ps+1}{rs+r+1}; \tag{7.16}$$

действительно, величина в левой части, очевидно, принимает максимальное и минимальное значения при $\theta=0$ и $\theta=\pi$, и непосредственная подстановка показывает, что $\theta=0$ соответствует минимуму, а $\theta=\pi$ — максимуму. Тогда в силу (7.13) и (7.14)

$$|A_k|<\frac{2p^{n+1}}{(r+1+rp^{-1})\,q}<\frac{2p^{n+2}}{rq\,(1+p)}. \tag{7.17}$$

Мы приходим к выводу, что *погрешность в* (7.11), *возникшая из-за отбрасывания* $r-1$ *корней, отличных от* x, *по абсолютной величине меньше чем*

$$2\,(r-1)\,p^{n+2}/[rq\,(1+p)]. \tag{7.18}$$

Корень x легко вычислить при помощи (7.12) методом последовательных приближений, полагая $x_0=1$ и $x_{\nu+1}=f(x_\nu)$. Последовательность сходится к x монотонно, и каждый ее член дает оценку снизу для x, а любое значение s, для которого $s>f(s)$, дает оценку сверху. Нетрудно проверить, что

$$x=1+qp^r+(r+1)\,(qp^r)^2+\ldots\,. \tag{7.19}$$

§ 8*). СОБЫТИЯ БОЛЕЕ ОБЩЕГО ВИДА

Наш метод применим к более общим задачам, которые ранее казались значительно сложнее задач об обычных сериях.

Примеры. а) *Серии двух типов.* Пусть $\mathscr{E}$ означает событие «*либо серия успехов длины* r, *либо серия неудач длины* ρ» (см. пример 1, е)). Мы имеем дело с двумя рекуррентными событиями $\mathscr{E}_1$ и $\mathscr{E}_2$, где $\mathscr{E}_1$ — «серия успехов длины r», $\mathscr{E}_2$ — «серия неудач длины ρ», а $\mathscr{E}$ — событие «либо $\mathscr{E}_1$, либо $\mathscr{E}_2$». Событию $\mathscr{E}_1$ соответ-

*) Этот параграф посвящен специальным вопросам и может быть опущен при первом чтении.

ствует производящая функция (7.5), которая будет теперь обозначаться $U_1(s)$. Соответствующая производящая функция $U_2(s)$ для $\mathscr{E}_2$ получается из (7.5) перестановкой p и q и заменой r на ρ. Вероятность u_n того, что событие $\mathscr{E}$ произойдет при n-м испытании, есть сумма соответствующих вероятностей для $\mathscr{E}_1$ и $\mathscr{E}_2$, если исключить случай $u_0=1$. Следовательно,

$$U(s)=U_1(s)+U_2(s)-1. \tag{8.1}$$

Производящая функция $F(s)$ времени возвращения для $\mathscr{E}$ есть снова $F(s)=1-U^{-1}(s)$, или

$$F(s)=\frac{(1-ps)\,p^rs^r\,(1-q^\rho s^\rho)+(1-qs)\,q^\rho s^\rho\,(1-p^rs^r)}{1-s+qp^rs^{r+1}+pq^\rho s^{\rho+1}-p^rq^\rho s^{r+\rho}}. \tag{8.2}$$

Математическое ожидание времени возвращения находится дифференцированием:

$$\mu=(1-p^r)(1-q^\rho)/(qp^r+pq^\rho-p^rq^\rho). \tag{8.3}$$

При $\rho\to\infty$ это выражение стремится к математическому ожиданию времени возвращения для серии успехов, которое вычисляется по формуле (7.7).

б) В гл. VIII, 1 мы нашли вероятность x того, что *серия успехов длины r появится раньше серии неудач длины* ρ. Определим рекуррентные события $\mathscr{E}_1$ и $\mathscr{E}_2$ так же, как в примере а). Пусть x_n — вероятность того, что $\mathscr{E}_1$ впервые происходит при n-м испытании и до этого ни разу не происходило $\mathscr{E}_2$, а f_n — вероятность того, что $\mathscr{E}_1$ впервые происходит при n-м испытании (без дополнительных условий на $\mathscr{E}_2$). Определим y_n и g_n так же, как x_n и f_n соответственно, но с перестановкой $\mathscr{E}_1$ и $\mathscr{E}_2$.

Производящую функцию для f_n дает формула (7.6), а $G(s)$ получается при перестановке p и q и замене r на ρ. Для x_n и y_n мы имеем очевидные рекуррентные соотношения

$$\begin{aligned} x_n&=f_n-(y_1f_{n-1}+y_2f_{n-2}+\ldots+y_{n-1}f_1),\\ y_n&=g_n-(x_1g_{n-1}+x_2g_{n-2}+\ldots+x_{n-1}g_1). \end{aligned} \tag{8.4}$$

Они имеют вид сверток, и поэтому для соответствующих производящих функций мы имеем

$$\begin{aligned} X(s)&=F(s)-Y(s)F(s),\\ Y(s)&=G(s)-X(s)G(s). \end{aligned} \tag{8.5}$$

Из этих двух линейных уравнений получаем

$$X(s)=\frac{F(s)(1-G(s))}{1-F(s)G(s)},\quad Y(s)=\frac{G(s)(1-F(s))}{1-F(s)G(s)}. \tag{8.6}$$

Выражения для x_n и y_n можно найти, снова пользуясь методом разложения на простейшие дроби. При $s=1$ имеем $X(1)=\sum x_n=x$, т. е. вероятности появления $\mathscr{E}_1$ раньше $\mathscr{E}_2$. Как числитель, так и знаменатель обращаются в нуль, и значение $X(1)$ вычисляется по правилу Лопиталя дифференцированием числителя и знаменателя: $X(1)=G'(1)/(F'(1)+G'(1))$. Используя равенства $F'(1)=(1-p^r)/(qp^r)$ и $G'(1)=(1-q^\rho)/(pq^\rho)$ (см. 7.7)), мы получаем $X(1)$ в том же виде, что и в формуле (1.3) гл. VIII.

в) Рассмотрим рекуррентное событие *УУННУУ*. Повторяя рассуждения § 7, легко находим, что

$$p^4q^2=u_n+p^2q^2u_{n-4}+p^3q^2u_{n-5}. \tag{8.7}$$

Поскольку известно, что $u_n\to\mu^{-1}$, мы получаем для среднего времени возвращения равенство $\mu=p^{-4}q^{-2}+p^{-2}+p^{-1}$. При $p=q=1/2$ находим, что $\mu=70$, тогда как среднее время возвращения для серии успехов длины 6 равно 126. Это показывает, что вопреки ожиданиям *при бросаниях монеты имеется существенное различие между сериями гербов и другими наборами той же длины.* ▶

§ 9. ОТСУТСТВИЕ ПАМЯТИ ДЛЯ ВРЕМЕН ОЖИДАНИЯ С ГЕОМЕТРИЧЕСКИМ РАСПРЕДЕЛЕНИЕМ

Геометрическое распределение времени ожидания обладает интересным и важным свойством, отличающим его от других распределений. Рассмотрим последовательность испытаний Бернулли и обозначим через $\mathbf{T}$ число испытаний до первого успеха (включая это испытание). Тогда $\mathbf{P}\{\mathbf{T}>k\}=q^k$. Пусть нам известно, что в первых m испытаниях успехов не было; время ожидания $\mathbf{T}^{(m)}$ от m-й неудачи до первого успеха имеет то же самое распределение $\{q^k\}$ и не зависит от числа предшествовавших неудач. Иначе говоря, вероятность того, что ожидание успеха продлится еще по крайней мере k единиц времени, всегда равно исходной вероятности того, что полное время ожидания превосходит k.

Если время жизни атома или срок службы элемента устройства имеет геометрическое распределение, то *старение* отсутствует: в любой момент своей жизни атом имеет одну и ту же вероятность распада при следующем испытании. Радиоактивные атомы на самом деле обладают этим свойством (однако в случае непрерывного времени роль геометрического распределения играет показательное).

Обратно, если известно, что некоторое явление характеризуется полным отсутствием памяти или старения, то распределение вероятностей его длительности должно быть геометрическим или показательным. Типичным примером является хорошо известный вид телефонных разговоров, часто рассматриваемый как модель бессвязности и полной зависимости от сиюминутных импульсов: окончание разговора — это мгновенное случайное событие, не связанное с предшествующей беседой. Напротив, если мы знаем, что в последние пять минут трамвая не было, то это увеличивает нашу надежду на его скорое появление. При бросании монеты вероятность того, что после второго испытания суммарные числа успехов и неудач будут равны, составляет 1/2. Однако если эти числа различны, то вероятность их совпадения после двух дополнительных испытаний равна 1/4. Это примеры последействия.

Для строгой формулировки утверждения предположим, что время ожидания **T** принимает значения 0, 1, 2, ... с вероятностями p_0, p_1, p_2, ... Пусть распределение **T** обладает следующим свойством: *условная вероятность окончания времени ожидания при k-м испытании (при условии, что оно не окончилось ранее) равна p_0 (вероятности окончания при первом испытании). Мы утверждаем, что тогда $p_k=(1-p_0)^k p_0$, так что* **T** *имеет геометрическое распределение.*

Для доказательства снова введем «хвосты»

$$q_k=p_{k+1}+p_{k+2}+p_{k+3}+\ldots=\mathbf{P}\{\mathbf{T}>k\}.$$

Наше условие состоит в том, что $\mathbf{T}>k-1$, а его вероятность есть q_{k-1}. Условная вероятность события $\mathbf{T}=k$ равна поэтому p_k/q_{k-1}, и по предположению при любом $k\geqslant 1$

$$p_k/q_{k-1}=p_0. \tag{9.1}$$

Так как $p_k=q_{k-1}-q_k$, то (9.1) приводится к виду

$$q_k/q_{k-1}=1-p_0. \tag{9.2}$$

Поскольку $q_0=p_1+p_2+\ldots=1-p_0$, отсюда следует, что $q_k=(1-p_0)^{k+1}$, и поэтому $p_k=q_{k-1}-q_k=(1-p_0)^k p_0$, как и утверждалось. ▶

В теории случайных процессов описанное выше отсутствие памяти связывается с *марковским свойством*; мы вернемся к этому вопросу в гл. XV, 13.

§ 10. ТЕОРИЯ ВОССТАНОВЛЕНИЯ

Область применения уравнений свертки, которые служат основой теории рекуррентных событий, значительно шире, чем это могло показаться при чтении предыдущих параграфов. Поэтому мы сформулируем заново и в более общем виде аналитическую часть полученных в них результатов и опишем как типичные вероятностные методы теории восстановления, так и их применения к изучению разнообразных ситуаций.

Мы начинаем с двух произвольных [1]) последовательностей действительных чисел f_1, f_2, ... и b_0, b_1, ... Новую последователь-

[1]) Мы полагаем $f_0=0$. Из (10.1) ясно, что случай $0<f_0<1$ приводит лишь к изменению обозначений, т. е. к замене f_k на $f_k/(1-f_0)$ и b_k на $b_k/(1-f_0)$.

ность $v_0, v_1, \ldots$ можно задать уравнениями свертки

$$v_n = b_n + f_1 v_{n-1} + f_2 v_{n-2} + \ldots + f_n v_0. \qquad (10.1)$$

Они последовательно определяют $v_0, v_1, v_2, \ldots$, и поэтому v_n определяется однозначно в любом случае. Мы рассматриваем, однако, только последовательности, удовлетворяющие условиям [1])

$$f_n \geqslant 0, \quad f = \sum_{n=1}^{\infty} f_n < \infty; \quad b_n \geqslant 0, \quad b = \sum_{n=0}^{\infty} b_n < \infty. \quad (10.2)$$

В этом случае v_n неотрицательны и соответствующие производящие функции должны удовлетворять равенству

$$V(s) = B(s)/(1 - F(s)). \qquad (10.3)$$

Производящие функции F и B сходятся по меньшей мере при $0 \leqslant s < 1$, и поэтому (10.3) определяет степенной ряд, который сходится, если $F(s) < 1$. Соотношения (10.1) и (10.3) полностью эквивалентны. В § 3 мы рассматривали частный случай, когда $B(s) = 1$ ($v_n = u_n$ при любом n). В § 5 рассмотрена общая ситуация при условии $f \leqslant 1$. Чтобы обеспечить возможность применения к теории популяций, мы теперь будем допускать, что $f > 1$; оказывается, этот случай легко сводится к стандартному случаю $f = 1$.

Мы будем говорить, что *последовательность* $\{f_n\}$ *имеет период* $\lambda > 1$, если $f_n = 0$, когда n не делится на λ, и λ — максимальное целое число с таким свойством. Это эквивалентно следующему утверждению: $F(s) = F_1(s^\lambda)$ — степенной ряд по s^λ, но не по $s^{r\lambda}$ при любом $r > 1$. Снова положим

$$\mu = \sum n f_n \leqslant \infty \qquad (10.4)$$

и будем считать, что μ^{-1} нужно считать равным нулю, если $\mu = \infty$.

Теорема 1. (Теорема восстановления.) *Пусть выполняются условия* (10.2) *и* $\{f_n\}$ *непериодична. Тогда*

(i) *если* $f < 1$, *то* $v_n \to 0$ *и*

$$\sum_{n=0}^{\infty} v_n = b/(1-f); \qquad (10.5)$$

(ii) *если* $f = 1$, *то*

$$v_n \to b\mu^{-1}; \qquad (10.6)$$

(iii) *если* $f > 1$, *то существует единственный положительный корень уравнения* $F(\xi) = 1$ *и*

$$\xi^n v_n \to B(\xi)/(\xi F'(\xi)). \qquad (10.7)$$

---

[1]) Неотрицательность f_n существенна, а условие сходимости двух рядов введено только для удобства. Никаких общих выводов нельзя сделать, если $b = \infty$ *и* $f = \infty$. Соотношение (10.7) остается справедливым при $f = \infty$ (за исключением того, что в этом случае $F'(\xi)$ может быть равно бесконечности) и оказывается бессмысленным, если $b = \infty$ *и* $F'(\xi) = \infty$.

Очевидно, $\xi < 1$, и поэтому производная $F'(\xi)$ конечна; (10.7) доказывает, что последовательность $\{v_n\}$ асимптотически ведет себя как геометрическая прогрессия со знаменателем $\xi^{-1} > 1$.

Доказательство. Утверждения (i) и (ii) были доказаны в § 5. Для доказательства (iii) достаточно применить (ii) к последовательностям $\{f_n\xi^n\}$, $\{b_n\xi^n\}$ и $\{v_n\xi^n\}$ с производящими функциями $F(\xi s)$, $B(\xi s)$ и $V(\xi s)$ соответственно. ▶

Мы исключили периодические последовательности $\{f_n\}$, поскольку они не представляют особого интереса. Фактически они не дают ничего нового. В самом деле, если $\{b_n\}$ и $\{f_n\}$ имеют один и тот же период λ, то $B(s)$ и $F(s)$ — степенные ряды по s^λ, и поэтому то же верно для $V(s)$. К последовательностям $\{f_{n\lambda}\}$, $\{b_{n\lambda}\}$ и $\{v_{n\lambda}\}$ с производящими функциями $F(s^{1/\lambda})$, $B(s^{1/\lambda})$ и $V(s^{1/\lambda})$ применима теорема 1. В случае $F(1)=1$ отсюда следует, что $v_{n\lambda} \longrightarrow b\lambda/\mu$. Далее, произвольный степенной ряд B можно представить в виде линейной комбинации λ степенных рядов B_j, содержащих только степени s^λ:

$$B(s) = B_0(s) + sB_1(s) + \ldots + s^{\lambda-1}B_{\lambda-1}(s). \tag{10.8}$$

Подстановка этого разложения в (10.3) и применение только что сформулированного утверждения показывают, что справедлива следующая теорема.

Теорема 2. *Пусть выполняются условия* (10.2) *и* $\{f_n\}$ *имеет период* $\lambda > 1$. *Тогда*

(i) *если* $f < 1$, *то справедливо* (10.5);

(ii) *если* $f = 1$, *то для* $j = 0, 1, \ldots, \lambda - 1$ *при* $n \longrightarrow \infty$

$$u_{n\lambda+j} \longrightarrow \lambda B_j\ 1)/\mu; \tag{10.9}$$

(iii) *если* $f > 1$, *то для* $j = 0, 1, \ldots, \lambda - 1$ *при* $n \longrightarrow \infty$

$$\xi^{n\lambda} u_{n\lambda+j} \longrightarrow \lambda B_j(\xi)/(\xi\mu). \tag{10.10}$$

С помощью рассуждений, которые аналогичны проведенным для рекуррентных событий, можно показать, что некоторые вероятности, связанные с самыми разнообразными случайными процессами, удовлетворяют уравнению типа свертки (10.1). При этом многие важные предельные теоремы оказываются простыми следствиями теоремы 1. Такой подход, известный под названием *теории восстановления* [1]), теперь повсеместно вытеснил громоздкие старые методы. Полностью его преимущества раскрываются при применении к процессам с непрерывным временем, тем не менее первые два примера могут служить иллюстрацией. Другие примеры приводятся в задачах 8 и 9. Применение теоремы 1 к невероятностным предельным теоремам содержится в примере в). Последние два примера связаны с практическими приложениями.

Примеры. а) *Рекуррентные события с запаздыванием.* Выведем новым способом утверждение § 5 о рекуррентном событии $\mathscr{E}$ с запаздыванием в случае, когда $\{f_j\}$ — распределение времени

---

[1]) В оригинале renewal arguments. Многие авторы используют также термин «метод введения регенерирующего события». — *Прим. перев.*

возвращения, а $\{b_j\}$ — распределение момента первого появления $\mathscr{E}$. Пусть v_n означает вероятность того, что $\mathscr{E}$ происходит при n-м испытании. Покажем, что выполняется (10.1). Существуют две возможности появления $\mathscr{E}$ при n-м испытании. Появление $\mathscr{E}$ может быть первым, и вероятность этого равна b_n. В противном случае существует последнее появление $\mathscr{E}$ до n-го испытания, и поэтому существует такое число $j\ (1\leqslant j<n)$, что $\mathscr{E}$ происходило при j-м испытании, а в *следующий* раз — при n-м. Вероятность этого равна v_jf_{n-j}. Указанные возможности исключают друг друга, и, значит,

$$v_n=b_n+v_1f_{n-1}+v_2f_{n-2}+\ldots+v_{n-1}f_1, \tag{10.11}$$

что совпадает с (10.1). Значит, производящая функция V вычисляется по формуле (10.3), которая согласуется с результатами § 5. (Хотя результаты совпадают по форме, ход рассуждений был разным: в § 5 перечисление происходило в соответствии с первым появлением $\mathscr{E}$, а сейчас использовалось последнее появление. Обе процедуры применяются и в других задачах и иногда приводят к формально различным уравнениям.)

б) *Вероятности попадания.* Рассмотрим последовательность испытаний с обычным (без запаздывания) возвратным рекуррентным событием $\mathscr{E}$. Пусть $\nu\geqslant 0$ — целое число. Предположим, что мы начинаем наблюдать процесс только после ν-го испытания и что нас интересует время до следующего появления $\mathscr{E}$. Более формально, для $r=1,\ 2,\ \ldots$ определим $w_\nu(r)$ как вероятность того, что *первое появление $\mathscr{E}$ после ν-го испытания* происходит при $(\nu+r)$-м испытании. Таким образом, $w_0(r)=f_r$ и $w_\nu(0)=0$. (Величины $w_\nu(r)$ называются вероятностями попадания в соответствии с их интерпретацией в теории случайных блужданий. При других обстоятельствах естественнее говорить о распределении остаточного времени ожидания, начинающегося с ν-го испытания. Ср. с примером гл. XV, 2, л).)

Чтобы найти эти вероятности, мы используем стандартный метод теории восстановления следующим образом. Событие $\mathscr{E}$ может произойти в самый первый раз при $(\nu+r)$-м испытании. Вероятность этого равна $f_{\nu+r}$. В противном случае существует такое число $k\leqslant\nu$, что $\mathscr{E}$ происходит впервые при k-м испытании. Поведение процесса после k-го испытания является вероятностной копией поведения всего процесса, но ν-е испытание становится теперь $(\nu-k)$-м. Вероятность нашего события равна поэтому $f_kw_{\nu-k}(r)$, и, значит, при любом $r>0$

$$w_\nu(r)=f_{\nu+r}+\sum_{k=1}^{\nu}f_kw_{\nu-k}(r). \tag{10.12}$$

Это уравнение имеет стандартный вид (10.1) с $b_n=f_{n+r}$. Нас интересует не производящая функция, а асимптотическое поведение вероятностей попадания при очень больших ν. Найти его позволяет теорема 1. Положим

$$\rho_k = f_{k+1} + f_{k+2} + \ldots \tag{10.13}$$

и напомним (см. формулу (1.8) гл. XI), что математическое ожидание времени возвращения удовлетворяет равенству

$$\mu = \rho_1 + \rho_2 + \ldots . \tag{10.14}$$

Если $\mathscr{E}$ непериодично, то из теоремы 1 мы выводим, что при $\nu \to \infty$

$$w_\nu(r) \to \begin{cases} \rho_r/\mu, & \text{если } \mu < \infty, \\ 0, & \text{если } \mu = \infty. \end{cases} \tag{10.15}$$

Этот результат чрезвычайно интересен. В случае когда среднее время возвращения конечно, (10.14) означает, что $\{\rho_r/\mu\}$ — распределение вероятностей, и, значит, мы имеем предельную теорему обычного типа. Однако *при* $\mu = \infty$ *вероятность того, что время ожидания превысит любое данное число* r, *стремится к* 1. Иначе говоря, наши времена ожидания ведут себя значительно хуже, чем времена возвращения. Это неожиданное явление имеет важные следствия, которые детально обсуждаются в томе 2. (См. также задачу 10.)

в) *Повторное осреднение*. Следующая задача имеет аналитический характер и рассматривалась в различных контекстах значительно более сложными методами. Допустим, что $f_1 + \ldots + f_r = 1$ и $f_j \geqslant 0$. Для любых r чисел $v_1, \ldots, v_r$ назовем ***взвешенным средним*** величину $f_1 v_r + \ldots + f_r v_1$. Определим теперь бесконечную последовательность $v_1, v_2, \ldots$, которая начинается с заданных r чисел, полагая v_n равным взвешенному среднему предыдущих r членов. Иначе говоря, для $n > r$

$$v_n = f_1 v_{n-1} + \ldots + f_r v_{n-r}. \tag{10.16}$$

Так как последовательность $f_1, f_2, \ldots$ обрывается на r-м члене, эти уравнения имеют вид (10.1). Определим теперь b_k так, чтобы (10.1) выполнялось при всех n. Для этого нужно положить $b_0 = v_0 = 0$ и

$$b_k = v_k - f_1 v_{k-1} + \ldots - f_{k-1} v_1, \quad k \leqslant r. \tag{10.17}$$

(При $k > r$ по определению $b_k = 0$.) Без каких-либо вычислений из теоремы 1 следует, что при таком повторном осреднении v_n *стремится к конечному пределу*. Чтобы вычислить этот предел, мы должны найти $b = b_1 + \ldots + b_r$. Используя обозначение (10.13) для «хвостов» $\sum f_k$, нетрудно вывести из (10.17) и (10.6), что

$$v_n \to (v_1 \rho_{r-1} + \ldots + v_r \rho_0)/(f_1 + 2f_2 + \ldots + r f_r). \tag{10.18}$$

Например, если $r = 3$ и вычисляются *средние арифметические*, то $f_1 = f_2 = f_3 = 1/3$ и

$$v_n \to (1/6)\ (v_1 + 2v_2 + 3v_3). \tag{10.19}$$

Легкость, с которой мы получили этот результат, не должна скрывать того факта, что задача является трудной, если ее рассматривать вне настоящего контекста. (Другой подход можно найти в задаче 15 гл. XV,14.)

г) *Самовосстанавливающиеся устройства.* Вернемся к примеру из § 2, в котором элемент устройства, устанавливаемый в момент n, имеет срок службы с распределением $\{f_k\}$. Когда срок службы оканчивается, элемент немедленно заменяется новым элементом того же типа; таким образом, последовательные замены образуют возвратное рекуррентное событие в последовательности зависимых испытаний (исходами которых являются решения о том, нужно или нет производить замену).

Допустим теперь, что элемент, установленный в момент времени 0, имеет возраст k, а не является новым. Это влияет только на время ожидания, т. е. $\mathcal{E}$ оказывается рекуррентным событием *с запаздыванием.* Чтобы получить распределение $\{b_n\}$ первого времени ожидания, заметим, что b_n есть (условное) математическое ожидание того, что элемент откажет в возрасте $n+k$ при условии, что он уже достиг возраста k. Поэтому при $k \geqslant 1$

$$b_n = f_{n+k}/r_k, \quad \text{где } r_k = f_{k+1} + f_{k+2} + \ldots . \tag{10.20}$$

На практике нас интересует обычно не отдельный элемент, а все устройство (например, сеть уличных фонарей в городе). Предположим поэтому, что *исходное устройство* (в момент времени 0) *состоит из N элементов, из которых β_k имеют возраст k* (причем $\sum \beta_k = N$). Каждый элемент порождает последовательность потомков, которым может потребоваться замена в момент n. *Математическое ожидание v_n числа всех замен в момент времени n, очевидно, удовлетворяет основному уравнению* (10.1) *с*

$$b_n = \sum \beta_k f_{n+k}/r_k. \tag{10.21}$$

Мы имеем здесь первый пример, в котором v_n — *математическое ожидание*, а не вероятность; нам известно только, что $v_n \leqslant N$.

Несложные вычисления показывают, что $b = \sum b_n = N$, и поэтому, согласно теореме 1, величины $v_n \to N/\mu$, если замены непериодичны. Это означает существование *устойчивого предельного распределения возраста.* Действительно, для того чтобы элемент в момент времени n имел возраст k, необходимо и достаточно, чтобы он был установлен в момент $n-k$ и работал до возраста k. Поэтому математическое ожидание числа таких элементов равно $v_{n-k} r_k$ и стремится к $N r_k/\mu$ при $n \to \infty$. Иначе говоря, с течением времени *доля элементов устройства, имеющих возраст k, стремится к r_k/μ.* Таким образом, *предельное распределение возраста не зависит от начального распределения возраста* и определяется только вероятностями f_n. Аналогичное утверждение справедливо и при значительно более общих условиях. Численный пример приводится в табл. 3. При этом обнаруживается примечательный факт: стремление к пределу не является монотонным. (См. также задачи 16—18.)

Таблица 3

Изменение распределения возраста в устройстве, описанном в примере «г»

k \ n	0	1	2	3	4	5	6	7	∞
0	500	397	411,4	412	423,8	414,3	417,0	416,0	416,7
1	320	400	317,6	329,1	329,6	339,0	331,5	333,6	333,3
2	74	148	185	146,9	152,2	152,4	156,8	153,3	154,2
3	100	40	80	100	79,4	82,3	82,4	84,8	83,3
4	6	15	6	12	15	11,9	12,3	12,4	12,5

Столбцы дают распределение возраста в устройстве, состоящем из 1000 элементов, в моменты времени $n=0, 1, \ldots, 7$, а также предельное распределение. Предполагается, что распределение срока службы определяется вероятностями

$$f_1=0{,}20,\ f_2=0{,}43,\ f_3=0{,}17,\ f_4=0{,}17,\ f_5=0{,}03,$$

так что ни один элемент не может иметь возраст 5. Уравнение $1-F(s)=0$ имеет корни 1, $-5/3$, -5 и $\pm 2i$. Среднее время возвращения равно 2,40.

д) *Человеческие популяции.* В качестве примера, где $f=\sum f_n>1$, мы используем простейшую модель человеческой популяции. Она аналогична модели из предыдущего примера, однако размер популяции является теперь переменным, и рождения женщин играют роль замен. Новой особенностью является то, что женщина может иметь любое число дочерей, и поэтому ее потомство женского пола может исчезнуть, но может стать и многочисленным. Обозначим через f_n вероятность того, что новорожденная девочка доживет до возраста n и в этот момент у нее появится дочка. (Зависимостью от числа предыдущих детей и их возрастов мы пренебрегаем.) Тогда $f=\sum f_n$ — математическое ожидание числа дочерей, и поэтому в здоровой популяции $f>1$. Теорема 1 утверждает, что тогда размер популяции растет с примерно постоянной скоростью ξ и что распределение возраста в популяции стремится к пределу, как это описано в предыдущем примере [1]). Модель, по общему признанию, является грубой, но тем не менее представляет некоторый практический интерес: любопытную особенность предельного поведения ξ, конечно, нельзя было предугадать без соответствующего математического анализа.

[1]) Однако предельное распределение возраста вычисляется сложнее. Формулы для предельного распределения возраста частиц в ветвящихся процессах можно найти в книге Харрис Т. Е. Теория ветвящихся случайных процессов.— М.: Мир, 1966.— *Прим. перев.*

§ 11 *). ДОКАЗАТЕЛЬСТВО ОСНОВНОЙ ПРЕДЕЛЬНОЙ ТЕОРЕМЫ

В § 3 мы опустили доказательство теоремы 3, которую мы теперь сформулируем следующим образом. *Пусть* $f_1, f_2, \ldots$ — *такая последовательность неотрицательных чисел, что* $\sum f_n = 1$ *и* 1 *есть наибольший общий делитель тех* n, *для которых* $f_n > 0$. *Пусть* $u_0 = 1$ *и*

$$u_n = f_1 u_{n-1} + f_2 u_{n-2} + \ldots + f_n u_0, \quad n \geqslant 1. \tag{11.1}$$

Тогда

$$u_n \to \mu^{-1}, \quad \text{где} \quad \mu = \sum_{n=1}^{\infty} n f_n \tag{11.2}$$

(μ^{-1} *полагается равным нулю, когда* $\mu = \infty$).

Чтобы не прерывать изложение, мы предпошлем доказательству три известные леммы, которые часто используются вне рамок теории вероятностей.

Пусть A — множество всех целых чисел n, для которых $f_n > 0$; обозначим через A^+ множество всех возможных линейных комбинаций

$$p_1 a_1 + \ldots + p_r a_r \tag{11.3}$$

чисел $a_1, \ldots, a_r$ из A (коэффициенты p_j — положительные целые числа).

Лемма 1. *Существует такое число* N, *что* A^+ *содержит все целые числа* $n > N$.

Доказательство. Как было известно еще Евклиду, равенство единице наибольшего общего делителя чисел из A означает, что можно выбрать числа $a_1, \ldots, a_r$ из A и (не обязательно положительные) целые числа c_j так, что

$$c_1 a_1 + \ldots + c_r a_r = 1. \tag{11.4}$$

Положим $s = a_1 + \ldots + a_r$. *Любое* целое число n может быть единственным образом представлено в виде $n = xs + y$, где x и y — целые и $0 \leqslant y < s$. Поэтому

$$n = \sum_{k=1}^{r} (x + c_k y)\, a_k, \tag{11.5}$$

и все коэффициенты положительны, если x в y раз больше наибольшего из числа $|c_k|$. ▶

Лемма 2. (Принцип выбора.) *Допустим, что для любого целого* $\nu > 0$ *задана такая последовательность чисел* $z_1^{(\nu)}, z_2^{(\nu)}, \ldots$, *что* $0 \leqslant z_k^{(\nu)} \leqslant 1$. *Тогда существует такая последовательность* $\nu^{(1)}, \nu^{(2)}, \ldots \to \infty$, *что, когда* ν *пробегает ее,* $z_k^{(\nu)}$ *стремится к пределу при любом фиксированном* k.

*) Материал этого параграфа в дальнейшем не используется.

Доказательство[1]). Выберем возрастающую последовательность $\nu_1^{(1)}, \nu_2^{(1)}, \ldots$ так, чтобы, когда ν пробегало ее, $z_1^{(\nu)}$ стремилось бы к пределу z_1. Выберем из этой последовательности такую *подпоследовательность* $\nu_1^{(2)}, \nu_2^{(2)}, \ldots$, что, когда ν пробегает ее, $z_2^{(\nu)} \to z_2$. Продолжая таким же образом, мы построим для каждого n такую последовательность чисел $\nu_j^{(n)} \to \infty$, что, когда ν пробегает ее, $z_n^{(\nu)} \to z_n$, и любое $\nu_j^{(n)}$ является элементом предыдущей последовательности $\{\nu_j^{(n-1)}\}$. Положим теперь $\nu^{(r)} = \nu_r^{(r)}$. Пусть $r > n$. Все элементы $\nu^{(r)}$, кроме первых n, принадлежат последовательности $\nu_1^{(n)}, \nu_2^{(n)}, \ldots$, и поэтому $z_n^{(\nu)} \to z_n$, когда ν пробегает последовательность $\nu^{(1)}, \nu^{(2)}, \ldots$. ▶

Лемма 3. *Пусть* $\{w_n\}$ $(n = 0, \pm 1, \pm 2, \ldots)$ — *такая бесконечная в обе стороны последовательность, что* $0 \leqslant w_n \leqslant 1$ *и*

$$w_n = \sum_{k=1}^{\infty} f_k w_{n-k} \tag{11.6}$$

при любом n. *Если* $w_0 = 1$, *то* $w_n = 1$ *при всех* n.

Доказательство. Поскольку

$$w_0 = \sum_{k=1}^{\infty} f_k w_{-k} \leqslant \sum_{k=1}^{\infty} f_k = 1, \tag{11.7}$$

условие $w_0 = 1$ выполняется только тогда, когда эти два ряда почленно совпадают, а именно при любом k либо $f_k = 0$, либо $w_{-k} = 1$. Это означает, что $w_{-a} = 1$ для любого числа a из A. Но тогда рассуждения, проведенные при $n = 0$, применимы и к случаю $n = -a$, и мы заключаем, что $w_{-a-b} = 1$, если числа a и b принадлежат A. Продолжая по индукции, мы заключаем, что $w_{-m} = 1$ для любого числа m из A^+, и поэтому $w_{-m} = 1$ при любом $m > N$. Но это значит, что при $n = -N$ сумма в правой части (11.6) равна 1 и, таким образом, $w_{-N} = 1$. Полагая $n = -N + 1$, находим аналогично, что $w_{-N+1} = 1$, и, продолжая эти рассуждения, получаем по индукции, что $w_n = 1$ для всех n. ▶

Доказательство теоремы. Положим

$$\eta = \limsup_{n \to \infty} u_n. \tag{11.8}$$

В силу (11.1) очевидно, что $0 \leqslant \eta \leqslant 1$, и существует такая стремящаяся к бесконечности последовательность $r_1, r_2, \ldots$, что при $\nu \to \infty$

$$u_{r_\nu} \to \eta. \tag{11.9}$$

---

[1]) Это доказательство основано на так называемом *диагональном методе* Г. Кантора (1845—1918). Он давно стал стандартным, но во времена Кантора потрясал своей новизной.

Для каждого положительного целого числа ν введем бесконечную в обе стороны последовательность $\{u_n^{(\nu)}\}$ равенствами

$$u_n^{(\nu)} = \begin{cases} u_{r_\nu+n}, & \text{если } n \geqslant -r_\nu, \\ 0, & \text{если } n < -r_\nu. \end{cases} \tag{11.10}$$

Чтобы упростить изложение, лемма 2 была сформулирована для обычных последовательностей, однако она, очевидно, справедлива также и для бесконечных в обе стороны последовательностей. Поэтому можно найти возрастающую последовательность целых чисел $\nu_1, \nu_2, \ldots$, такую, что если индекс ν пробегает эту последовательность, то величина $u_n^{(\nu)}$ стремится к w_n для любого n. По построению $0 \leqslant w_n \leqslant \eta$ и $w_0 = \eta$. Далее, для любого ν и $n > -\nu$ определение (11.1) принимает вид

$$u_n^{(\nu)} = \sum_{k=1}^{\infty} f_k u_{n-k}^{(\nu)}, \tag{11.11}$$

и, переходя к пределу, мы получаем (11.6). По лемме 3 имеем $w_n = \eta$ при любом n.

Мы можем теперь завершить доказательство. Как и ранее, положим

$$\rho_k = f_{k+1} + f_{k+2} + \ldots, \tag{11.12}$$

так что $r_0 = 1$ и $\sum \rho_k = \mu$ (см. формулу (1.8) гл. XI). Суммируя соотношения (11.1) по $n = 1, 2, \ldots, N$ и приводя подобные члены, мы получаем тождество

$$\rho_0 u_N + \rho_1 u_{N-1} + \ldots + \rho_N u_0 = 1. \tag{11.13}$$

Используем это соотношение последовательно для $N = \nu_1, \nu_2, \ldots$. Когда N пробегает эту последовательность, $u_{N-k} \to w_{-k} = \eta$ при любом k. Если $\sum \rho_k = \infty$, то $\eta = 0$, и поэтому $u_n \to 0$, как и утверждалось. Если $\mu = \sum \rho_k < \infty$, то $\eta = \mu^{-1}$, и остается показать, что тогда $u_N \to \eta$ при любом способе стремления N к бесконечности. По определению верхнего предела мы имеем $u_{N-k} < \eta + \varepsilon$ при любом фиксированном k и достаточно большом N. Более того, $u_n \leqslant 1$ при всех n. Допустим теперь, что N стремится к бесконечности так, что $u_N \to \eta_0$. Из (11.13) легко вывести, что асимптотически

$$\rho_0 \eta_0 + (\rho_1 + \ldots + \rho_r)(\eta + \varepsilon) + \rho_{r+1} + \rho_{r+2} + \ldots \geqslant 1, \tag{11.14}$$

и поэтому

$$\rho_0(\eta_0 - \eta) + \mu(\eta + \varepsilon) \geqslant 1. \tag{11.15}$$

Но $\mu\eta = 1$ и $\eta_0 \leqslant \eta$ по определению η. Поскольку (11.15) выполняется для каждого $\varepsilon > 0$, отсюда следует, что $\eta_0 = \eta$, и, значит, $u_N \to \mu^{-1}$ при любом способе стремления N к бесконечности. ▶

§ 12. ЗАДАЧИ

1. Допустим, что $F(s)$ — многочлен. Доказать для этого случая все теоремы § 3, используя метод разложения на простейшие дроби из гл. XI, 4.

2. Пусть подбрасываются r монет и рекуррентное событие $\mathscr{E}$ состоит в том, что для каждой из r монет суммарные числа появлений гербов и решеток одинаковы. Возвратно или невозвратно событие $\mathscr{E}$? Для минимального r, при котором $\mathscr{E}$ невозвратно, найти вероятность того, что $\mathscr{E}$ произойдет хотя бы один раз.

3. Пусть событие $\mathscr{E}$ состоит в том, что в последовательности независимых бросаний идеальной кости суммарные числа появлений единиц, двоек, ..., шестерок одинаковы. Показать, что $\mathscr{E}$ — невозвратное (периодическое) рекуррентное событие, и найти вероятность f того, что $\mathscr{E}$ произойдет хотя бы один раз.

4. Пусть в последовательности испытаний Бернулли событие $\mathscr{E}$ происходит, когда суммарное число успехов в λ раз превосходит суммарное число неудач; здесь λ — целое положительное число. (См. пример 1, в).) Показать, что $\mathscr{E}$ возвратно тогда и только тогда, когда $p/q=\lambda$, т. е. $p=\lambda/(\lambda+1)$. *Указание*. Воспользоваться нормальным приближением.

5. Мы говорим, что в последовательности испытаний Бернулли происходит событие $\mathscr{E}$, если суммарное число успехов вдвое превосходит суммарное число неудач и это отношение никогда ранее не превышало 2. Показать, что $\mathscr{E}$ невозвратно и периодично. Показать, далее, что производящая функция определяется кубическим уравнением $F(s)=qs(U(s)ps)^2$. (*Указание*. $U(s)ps$ — производящая функция времени ожидания первого момента, когда число успехов будет *больше* удвоенного числа неудач.)

6. Пусть $\mathbf{X}_j$ — независимые целочисленные одинаково распределенные случайные величины. Допустим, что они принимают как положительные, так и отрицательные значения. Доказать, что событие, определяемое условиями $\mathbf{S}_n=0$, $\mathbf{S}_1\leqslant 0, \ldots, \mathbf{S}_{n-1}\leqslant 0$, является рекуррентным и невозвратным.

7. *Счетчики Гейгера*. (См. примеры 1, ж) и 4, д).) Обозначим через $\mathbf{N}_n$ и $\mathbf{Z}_n$ число появлений $\mathscr{E}$ и число регистраций до момента времени n включительно. Установить связь между этими величинами и найти асимптотические формулы для $\mathbf{E}(\mathbf{Z}_n)$ и $\operatorname{Var}(\mathbf{Z}_n)$.

8. В *счетчиках Гейгера второго типа* каждая появляющаяся частица (независимо от того, регистрируется она или нет) запирает счетчик ровно на r единиц времени (т. е. на $r-1$ испытаний, следующих за моментом ее появления). Длительность мертвого времени, следующего за регистрацией, является поэтому случайной величиной. Найти ее производящую функцию G. Для рекуррентного события $\mathscr{E}$, состоящего в том, что счетчик «свободен», выразить производящую функцию F его времени возвращения через G. Наконец, найти среднее время возвращения.

9. *Более общий тип счетчиков Гейгера*. Пусть, как и в задаче 8, каждая новая частица полностью уничтожает влияние предыдущих, но теперь мы будем предполагать, что время, на которое частица запирает счетчик, является случайной величиной с заданной производящей функцией $B(s)$. (В предыдущей задаче $B(s)=s^r$.) Решить задачу 8 при этих более общих условиях.

10. Для рекуррентного события $\mathscr{E}$ с запаздыванием вероятности v_n не зависят от n только в случае, когда производящая функция момента первого появления $\mathscr{E}$ имеет вид $B(s)=(1-F(s))/[\mu(1-s)]$, т. е. когда $b_n=(f_{n+1}+f_{n+2}+\ldots)/\mu$. Как это связано с предельной теоремой для вероятностей попадания в примере 10, б)?

11. Найти приближенное значение вероятности того, что при 10 000 бросаний монеты число серий гербов длины 3 лежит между 700 и 730.

12. Пусть $\mathscr{E}$ означает появление *ГРГ* в последовательности бросаний монеты. Обозначим через r_n вероятность того, что $\mathscr{E}$ не произойдет в n испытаниях. Найти производящую функцию и при помощи разложения на простейшие дроби получить асимптотическое разложение.

13. Показать, что в примере 8, б) математическое ожидание времени до появ-

ления серии успехов длины r или серии неудач длины ρ равно

$$\mu_1\mu_2/(\mu_1+\mu_2),$$

где μ_1 и μ_2— средние времена возвращения для серии успехов длины r и серии неудач длины ρ соответственно.

14. Возможные исходы A, B, C каждого испытания имеют вероятности α, β, γ $(\alpha+\beta+\gamma=1)$. Найти производящую функцию вероятностей того, что в n испытаниях отсутствуют серии длины r: а) из исходов A, б) из исходов A или B, в) любого вида.

15. *Продолжение.* Найти вероятность того, что первая A-серия длины r предшествует B-серии длины ρ и заканчивается при n-м испытании. (*Указание.* Производящая функция имеет тот же вид, что $X(s)$ в (8.6), причем F и G вычисляются по формулам (7.6) с заменой $p=1-q$ на α и β соответственно.)

16. *Самовосстанавливающиеся устройства.* В примере 10, г) найти предельное распределение возраста, предполагая, что распределение срока службы является геометрическим: $f_k=q^{k-1}p$.

17. *Продолжение.* Начальное *распределение возраста* $\{\beta_k\}$ *называется стационарным*, если оно не изменяется с течением времени. Показать (без вычислений), что это возможно только при $\beta_k=r_k/\mu$.

18. *Продолжение.* Обозначим через $w_k(n)$ математическое ожидание числа элементов, имеющих в момент времени n возраст k. Найти рекуррентные уравнения и вывести из них, что число элементов устройства остается постоянным. Показать далее, что математическое ожидание $w_0(n)$ удовлетворяет соотношению

$$w_0(n)=w_0(n-1)\,f_1/r_0+w_1(n-1)\,f_2/r_1+\ldots .$$

---

19. Пусть $\mathcal{E}$ — возвратное непериодическое рекуррентное событие. Допустим, что время возвращений имеет конечное среднее μ и дисперсию σ^2. Положим $q_n=f_{n+1}+f_{n+2}+\ldots$ и $r_n=q_{n+1}+q_{n+2}+\ldots$. Показать, что производящие функции $Q(s)$ и $R(s)$ сходятся при $s=1$. Доказать, что

$$u_0+\sum_{n=1}^{\infty}(u_n-1/\mu)\,s^n=R(s)/(\mu Q(s)) \tag{12.1}$$

и что поэтому

$$u_0+\sum_{n=1}^{\infty}(u_n-1/\mu)=(\sigma^2-\mu+\mu^2)/(2\mu^2). \tag{12.2}$$

20. Пусть $\mathcal{E}$ — возвратное рекуррентное событие и $\mathbf{N}_r$ — число появлений $\mathcal{E}$ в r испытаниях. Доказать, что

$$\mathbf{E}(\mathbf{N}_r^2)=u_1+\ldots+u_r+2\sum_{j=1}^{r-1}u_j\,(u_1+\ldots+u_{r-j}) \tag{12.3}$$

и что поэтому $\mathbf{E}(\mathbf{N}_r^2)$ является коэффициентом при s^r в

$$(F^2(s)+F(s))/[(1-s)\,(1-F(s))^2]. \tag{12.4}$$

(Заметить, что это можно сформулировать изящнее при помощи двойных производящих функций.)

21. Положим $q_{k,n}=\mathbf{P}\{\mathbf{N}_k=n\}$. Показать, что $q_{k,n}$ — это коэффициент при s^k в

$$F^n(s)\,(1-F(s))/(1-s). \tag{12.5}$$

Вывести отсюда, что $\mathbf{E}(\mathbf{N}_r)$ и $\mathbf{E}(\mathbf{N}_r^2)$ являются коэффициентами при s^r в

$$F(s)/[(1-s)\,(1-F(s))] \tag{12.6}$$

и в (12.4) соответственно.

22. Используя обозначения задачи 19, показать, что

$$\frac{F(s)}{(1-s)(1-F(s))} = -\frac{1}{1-s} + \frac{1}{\mu(1-s)^2} + \frac{R(s)}{\mu\{1-F(s)\}}. \tag{12.7}$$

Отсюда и из последней задачи вывести равенство

$$\mathrm{E}(\mathbf{N}_r) = r/\mu + (\sigma^2+\mu-\mu^2)/(2\mu^2) + \varepsilon_r, \tag{12.8}$$

где $\varepsilon_r \longrightarrow 0$.

23. *Продолжение.* Используя аналогичные рассуждения, показать, что

$$\mathrm{E}(\mathbf{N}_r^2) = r^2/\mu^2 + [(2\sigma^2+\mu-\mu^2)/\mu^3]\,r + \alpha_r, \tag{12.9}$$

где $\alpha_r/r \longrightarrow 0$. Поэтому

$$\mathrm{Var}(\mathbf{N}_r) \sim (\sigma^2/\mu^3)\,r. \tag{12.10}$$

(*Указание.* Представить разность выражений (12.4) и (12.7) в виде суммы трех дробей со знаменателями, кратными $(1-s)^k$, $k=1, 2, 3$.)

24. Пусть $q_{k,n}$ — вероятность того, что в k испытаниях Бернулли имеется ровно n серий успехов длины r. Используя решение задачи 21, показать, что производящая функция $Q_k(x) = \sum q_{k,n}x^n$ является коэффициентом при s^k в

$$(1-p^rs^r)/[1-s+qp^rs^{r+1}-(1-ps)\,p^rs^rx].$$

Показать далее, что наименьшим по абсолютной величине корнем знаменателя является $s_1 \approx 1+qp^r(1-x)$.

25. *Продолжение. Пуассоновское распределение длинных серий*[1]). Доказать, что если число k испытаний и длина r серии стремятся к бесконечности так, что $kqp^r \longrightarrow \lambda$, то вероятность получить ровно n серий длины r стремится к $e^{-\lambda}\lambda^n/n!$. *Указание.* Используя предыдущую задачу, показать, что для производящей функции справедлива асимптотическая формула $\{1+qp^r(1-x)\}^{-k} \sim e^{-\lambda(1-x)}$. Использовать *теорему непрерывности* гл. XI, 6.

[1]) Эта теорема была доказана Мизесом, однако наш метод значительно проще.

ГЛАВА XIV

СЛУЧАЙНОЕ БЛУЖДАНИЕ И ЗАДАЧИ О РАЗОРЕНИИ

§ 1. ОБЩИЕ ПОНЯТИЯ

Первая часть этой главы посвящена испытаниям Бернулли, и для упрощения и оживления формулировок мы здесь снова воспользуемся образным языком пари и теории случайных блужданий.

Обратимся к уже известному нам игроку, который выигрывает или проигрывает доллар с вероятностями p и q соответственно. Пусть его начальный капитал равен z, и пусть он играет против соперника с начальным капиталом $a-z$, так что их суммарный капитал равен a. Игра продолжается до тех пор, пока капитал нашего игрока либо не уменьшится до нуля, либо не возрастет до a, т. е. до тех пор, пока один из двух играющих не разорится. Мы интересуемся вероятностью разорения нашего игрока и распределением вероятностей для продолжительности игры. Это *классическая задача о разорении*.

Физические приложения и аналогии подсказывают более гибкую интерпретацию — при помощи понятия блуждающей точки или «частицы» на оси x. Эта частица выходит из *начального положения* z и совершает через равные промежутки времени единичные скачки в положительном или отрицательном направлении в зависимости от того, чем закончилось соответствующее испытание — успехом или неудачей. Положение частицы после n шагов представляет капитал игрока по завершении n-го испытания. Испытания прекращаются, когда частица впервые достигает либо точки $x=0$, либо точки $x=a$, и мы говорим в этом случае, что частица совершает *случайное блуждание с поглощающими экранами в точках* $x=0$ *и* $x=a$. Это случайное блуждание *ограничено* возможными положениями $1, 2, \ldots, a-1$; при отсутствии поглощающих экранов случайное блуждание называется *неограниченным*. Физики используют модель случайного блуждания в качестве грубого приближения для одномерной диффузии или броуновского движения, в котором физическая частица подвергается большому числу столкновений с молекулами, сообщающими ей случайное движение. Случай $p>q$ соответствует *сносу* вправо, когда толчки слева более вероятны; при $p=q=1/2$ случайное блуждание называется *симметричным*.

В предельном случае $a\to\infty$ мы получаем случайное блуждание на полубесконечной прямой: выходящая из $z>0$ частица совершает случайное блуждание до того момента, когда она впервые достигнет нуля. В этой формулировке мы узнаем *задачу о времени первого* *достижения*; она была решена элементарными методами в гл. III (по крайней мере для симметричного случая) и с использованием производящих функций в гл. XI,3. Мы встретимся с уже полученными ранее формулами, однако они будут выведены по-новому.

В этой главе мы воспользуемся методом *разностных уравнений*, которые являются аналогом дифференциальных уравнений теории диффузии. Эта аналогия естественным образом приводит к различным модификациям и обобщениям классической задачи о разорении; типичным и поучительным примером здесь является замена поглощающего экрана *отражающим* и *упругим* экранами. Чтобы описать отражающий экран, рассмотрим случайное блуждание в конечном интервале, определенное как и ранее, за тем исключением, что каждый раз, когда частица оказывается в точке $x=1$, она с вероятностью p перемещается в точку $x=2$ и с вероятностью q остается в точке $x=1$. В игровой терминологии это соответствует такому соглашению: когда игрок проигрывает свой последний доллар, этот доллар великодушно возвращается ему противником, так что игра может продолжаться. Физик представляет себе стенку, помещенную в точке $x=1/2$ оси x и обладающую тем свойством, что частица, движущаяся из точки $x=1$ в точку $x=0$, отражается от стенки и возвращается в точку $x=1$ вместо того, чтобы попасть в точку $x=0$.

Как поглощающий, так и отражающий экраны являются частными случаями так называемого упругого экрана. Мы определим *упругий экран в точке* $x=0$ *при помощи следующего правила: из точки* $x=1$ *частица с вероятностью* p *движется к точке* $x=2$, *с вероятностью* δq *остается в точке* $x=1$ *и с вероятностью* $(1-\delta)q$ *движется к точке* $x=0$ *и поглощается* (т. е. процесс прекращается). При $\delta=0$ мы имеем классическую задачу о разорении, или поглощающие экраны, при $\delta=1$ — отражающие экраны. Когда δ меняется от 0 до 1, мы получаем семейство промежуточных случаев. Чем больше δ, тем более вероятно, что процесс будет продолжаться, и процесс с двумя отражающими экранами никогда не сможет прекратиться.

Параграфы 2 и 3 посвящены элементарному обсуждению классической задачи о разорении и связанных с ней вопросов. Последующие три параграфа технически более сложны (и могут быть опущены); в § 4 и 5 мы выводим производящие функции и из них — явные выражения для распределения продолжительности игры и т. п. Параграф 6 содержит схему предельного перехода к уравнению диффузии (формальные решения этого уравнения являются предельными распределениями для случайного блуждания).

В § 7 изложение снова становится элементарным и посвящено *случайным блужданиям в двух и более измерениях*, где мы сталкиваемся с новыми явлениями. В § 8 рассматривается обобщение совершенно другого типа: одномерное случайное блуждание, при котором частица не обязательно перемещается единичными скачками, а может менять свое положение скачками произвольной величины, кратной единице. Такие обобщенные случайные блуждания привлекли широкий интерес в связи с развитой Вальдом теорией *последовательного анализа*.

Раздел задач содержит существенные дополнения к тексту и наброски альтернативных подходов. Можно надеяться, что сравнение используемых методов окажется весьма поучительным.

В заключение следует подчеркнуть, что каждое случайное блуждание представляет собой некоторую цепь Маркова, так что настоящая глава служит частично и введением в следующую, где некоторые задачи для случайных блужданий (например, задачи с упругими экранами) будут сформулированы иначе.

§ 2. КЛАССИЧЕСКАЯ ЗАДАЧА О РАЗОРЕНИИ

Рассмотрим задачу, поставленную в начале настоящей главы. Пусть q_z — вероятность окончательного разорения [1]) игрока, а p_z — вероятность того, что он в конце концов выиграет. В терминах теории случайного блуждания q_z и p_z суть вероятности того, что частица, выходящая из точки z, будет поглощена соответственно в 0 или в a. Мы покажем, что $p_z+q_z=1$, так что нам не нужно рассматривать возможность нескончаемой игры.

После первого испытания состояние игрока равно либо $z-1$, либо $z+1$, и поэтому

$$q_z=pq_{z+1}+qq_{z-1}, \tag{2.1}$$

если $1<z<a-1$. При $z=1$ первое же испытание может привести к разорению и (2.1) следует заменить на $q_1=pq_2+q$. Точно так же при $z=a-1$ первое испытание может закончиться победой, и поэтому $q_{a-1}=qq_{a-2}$. Чтобы объединить наши уравнения, положим

$$q_0=1, \quad q_a=0. \tag{2.2}$$

При этом соглашении вероятность q_z разорения удовлетворяет (2.1) для $z=1, 2, \ldots, a-1$.

Системы вида (2.1) известны под названием *разностных уравнений* [2]), а (2.2) представляют собой *граничные условия* на q_z. Мы выведем явное выражение для q_z при помощи *метода частных решений*, который будет использоваться и в более общих случаях.

Предположим сперва, что $p\neq q$. Легко проверить, что разностное уравнение (2.1) допускает два частных решения $q_z=1$ и $q_z=(q/p)^z$. Следовательно,

$$q_z=A+B(q/p)^z, \tag{2.3}$$

[1]) Строго говоря, вероятность разорения определена на пространстве элементарных исходов бесконечно продолжающихся игр, но мы можем работать и с пространством элементарных исходов для n испытаний. Вероятность разорения менее чем за n испытаний возрастает вместе с n и, следовательно, имеет предел. Мы назовем этот *предел* «вероятностью разорения». Все вероятности в этой главе можно интерпретировать таким образом, не обращаясь к бесконечным пространствам (ср. гл. VIII, 1).

[2]) Подробнее о разностных уравнениях см., например, Гельфонд А. О., Исчисление конечных разностей.— М.: Наука, 1967.— *Прим. перев.*

где A и B — произвольные константы, представляет собой формальное решение системы (2.1). Граничные условия (2.2) будут выполнены тогда и только тогда, когда A и B удовлетворяют двум линейным уравнениям $A+B=1$ и $A+B(q/p)^a=0$. Таким образом,

$$q_z=[(q/p)^a-(q/p)^z]/[(q/p)^a-1] \tag{2.4}$$

формально является решением разностного уравнения (2.1), удовлетворяющим граничным условиям (2.2). Для того чтобы доказать, что (2.4) является искомой вероятностью разорения, осталось показать, что это решение единственно, т. е. что *все* решения уравнения (2.1) имеют вид (2.3). Если нам теперь дано произвольное решение уравнения (2.1), то можно выбрать две константы A и B так, что (2.3) будет совпадать с ним при $z=0$ и $z=1$. Подставляя в (2.1) последовательно $z=1, 2, 3, \ldots$, мы можем найти по этим двум значениям все остальные. Стало быть, два решения, совпадающие при $z=0$ и $z=1$, тождественно равны, и, следовательно, каждое решение имеет вид (2.3).

При $p=q=1/2$ наше рассуждение не проходит, ибо тогда (2.4) бессмысленно, так как частные решения $q_z=1$ и $q_z=(q/p)^z$ в этом случае совпадают. Однако при $p=q=1/2$ у нас есть второе решение $q_z=z$, и поэтому $q_z=A+Bz$ дает нам зависящее от двух констант решение уравнения (2.1). Для того чтобы были удовлетворены граничные условия (2.2), мы должны положить $A=1$ и $A+Ba=0$. Следовательно,

$$q_z=1-z/a. \tag{2.5}$$

(То же численное значение можно формально получить из (2.4), если, пользуясь правилом Лопиталя, найти предел этого выражения при $p\to 1/2$.)

Таким образом, мы доказали, что искомая *вероятность разорения игрока дается формулой* (2.4) *при* $p\neq q$ *и формулой* (2.5) *при* $p=q=1/2$. Вероятность p_z выигрыша нашим игроком всей игры равна вероятности разорения его противника и, стало быть, получается из наших формул при замене p, q и z на q, p и $a-z$ соответственно. Легко видеть, что $p_z+q_z=1$, как уже говорилось выше.

Мы можем переформулировать наш результат следующим образом. *Пусть игрок с начальным капиталом z играет против бесконечно богатого соперника, всегда готового продолжать игру, тогда как сам игрок имеет право прервать игру по своему желанию. Игрок, согласно принятой им стратегии, играет по тех пор, пока он либо не потеряет свой капитал, либо не увеличит его до a (с чистым выигрышем $a-z$). Тогда q_z равно вероятности его проигрыша, а $1-q_z$ — вероятности его выигрыша.*

При такой системе окончательный выигрыш или проигрыш игрока будет случайной величиной $\mathbf{G}$, принимающей значения $a-z$ и $-z$ с вероятностями $1-q_z$ и q_z соответственно. Ожидаемый выигрыш равен

$$\mathbf{E}(\mathbf{G}) = a(1-q_z)-z. \tag{2.6}$$

Очевидно, $\mathbf{E}(\mathbf{G})=0$ тогда и только тогда, когда $p=q$. Это означает, что при описанной системе *«безобидная» игра остается безобидной, а «небезобидная» игра не может превратиться в безобидную.*

Из (2.5) мы видим, что в случае $p=q$ игрок с начальным капиталом $z=999$ имеет вероятность 0,999 выиграть доллар до того, как он проиграет весь свой капитал. При $q=0{,}6$, $p=0{,}4$ игра весьма неблагоприятна, но вероятность (2.4) выиграть доллар до потери капитала все еще будет около 2/3. Вообще, игрок с относительно большим начальным капиталом z имеет приличные шансы выиграть небольшую сумму $a-z$ до своего разорения [1]).

(Удивительное следствие нашего результата см. в задаче 4.)

Изучим теперь влияние *изменения ставок.* Изменение единицы с доллара на полдоллара эквивалентно удвоению начальных капиталов. Соответствующая вероятность разорения q_z^* получается из (2.4) с заменой z на $2z$ и a на $2a$:

$$q_z^* = \frac{(q/p)^{2a}-(q/p)^{2z}}{(q/p)^{2a}-1} = q_z \cdot \frac{(q/p)^a+(q/p)^z}{(q/p)^a+1}. \tag{2.7}$$

При $q>p$ последняя дробь больше единицы и $q_z^*>q_z$. Мы переформулируем это заключение следующим образом: *если ставки удваиваются, а начальные капиталы остаются неизменными, то вероятность разорения уменьшается для игрока, у которого вероятность успеха $p<1/2$, и возрастает для его противника (для которого игра выгодна [2]))*.

Предположим, например, что Петр имеет 90 долларов, а Павел 10 и что $p=0{,}45$, т. е. игра неблагоприятна для Петра. Если ставка в каждом испытании равна одному доллару, то табл. 1 показывает, что вероятность разорения Петра равна приблизительно 0,866. Если в этой же игре ставка равна 10 долларам, то вероятность разорения Петра падает ниже одной четвертой, а именно примерно до 0,210. Таким образом, влияние увеличения ставок выражено более резко, чем можно было бы ожидать.

---

[1]) Некий игрок из года в год ездил в Монте-Карло и всегда успешно покрывал расходы на свой отдых. Он твердо верил в магическую власть над случаем. В действительности же в его опыте нет ничего удивительного. Если предположить, что он начинал с суммой, в десять раз большей его окончательного выигрыша, то каждый год вероятность успеха составляла примерно 0,9. Вероятность непрерывающейся последовательности из десяти успехов равна приблизительно $(1-1/10)^{10}\approx e^{-1}\approx 0{,}37$. Таким образом, продолжительные успехи ни в коей мере не являются невероятными. Более того, в *одной* неудаче, по мнению игрока, конечно, были бы виноваты неосмотрительность или кратковременное недомогание.

[2]) Детальный анализ других возможных стратегий можно найти в (неэлементарной) книге Dubbins L. E., Savage L. J., How to gamble if you must (которая имеет более информативный подзаголовок: Inequalities for stochastic processes), New York, McGrow-Hill, 1965.

Таблица 1

К классической задаче о разорении игрока

p	q	z	a	Вероятность: разорения	Вероятность: успеха	Математическое ожидание: выигрыша	Математическое ожидание: продолжительности игры
0,5	0,5	9	10	0,1	0,9	0	9
0,5	0,5	90	100	0,1	0,9	0	900
0,5	0,5	900	1 000	0,1	0,9	0	90 000
0,5	0,5	950	1 000	0,05	0,95	0	47 500
0,5	0,5	8 000	10 000	0,2	0,8	0	16 000 000
0,45	0,55	9	10	0,210	0,790	−1,1	11
0,45	0,55	90	100	0,866	0,134	−76,6	765,6
0,45	0,55	99	100	0,182	0,818	−17,2	171,8
0,4	0,6	90	100	0,983	0,017	−88,3	441,3
0,4	0,6	99	100	0,333	0,667	−32,3	161,7

Начальный капитал равен z. Игра заканчивается разорением (проигрыш z) или обладанием капиталом a (выигрыш $a-z$).

Вообще, если при каждом испытании ставится k долларов, то вероятность разорения мы находим из (2.4), заменяя z на z/k и a на a/k; вероятность разорения убывает с ростом k. Поэтому в игре с постоянными ставками игрок будет минимизировать вероятность разорения, выбирая наибольшую ставку, совместимую с его намерением выиграть заранее фиксированную сумму. Эмпирическая обоснованность этого заключения подвергалась сомнению обычно теми людьми, которые утверждают, что каждое «невыгодное» пари неразумно. Если бы это было принято всерьез, то это означало бы конец страхового дела, ибо осторожный водитель, страхующийся на всякий случай, играет при этом, очевидно, в технически «невыгодную» игру. В действительности в теории вероятности не существует никакой теоремы, которая отбила бы у такого водителя охоту страховаться.

Предельный случай $a=\infty$ соответствует игре против бесконечно богатого соперника. Устремляя в (2.4) и (2.5) a к ∞, получаем

$$q_z=\begin{cases} 1, & \text{если } p\leqslant q, \\ (q/p)^z, & \text{если } p>q. \end{cases} \tag{2.8}$$

Мы интерпретируем q_z *как вероятность окончательного разорения игрока с начальным капиталом* z, *играющего против бесконечно* *богатого соперника* [1]). В терминологии теории случайных блужданий q_z есть вероятность того, что выходящая из $z>0$ частица когда-либо достигнет нуля. Более естественно перефразировать этот результат в следующем виде: *в случайном блуждании, начинающемся в нуле, вероятность когда-либо достичь положения* $z>0$ *равна* 1, *если* $p\geqslant q$, *и равна* $(p/q)^z$ *при* $p<q$.

§ 3. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ ПРОДОЛЖИТЕЛЬНОСТИ ИГРЫ

Вероятностное распределение продолжительности игры будет выведено в следующих параграфах. Однако ее математическое ожидание может быть получено гораздо более простым методом, который применяется столь широко, что стоит объяснить его сейчас (хотя для этого придется несколько повториться).

Мы по-прежнему рассматриваем классическую задачу о разорении, сформулированную в начале этой главы. Будем считать известным тот факт, что продолжительность игры имеет конечное математическое ожидание D_z (это будет строго доказано в следующем параграфе).

Если первое испытание заканчивается успехом, то игра продолжается, как если бы начальным положением было $z+1$. Стало быть, условное математическое ожидание продолжительности игры при условии первого успеха равно $D_{z+1}+1$. Это соображение показывает, что ожидаемая продолжительность D_z удовлетворяет разностному уравнению

$$D_z = pD_{z+1} + qD_{z-1} + 1, \quad 0 < z < a, \tag{3.1}$$

с граничными условиями

$$D_0 = 0, \quad D_a = 0. \tag{3.2}$$

Наличие члена 1 делает разностное уравнение (3.1) неоднородным. Если $p\neq q$, то $D_z=z/(q-p)$ будет формальным решением уравнения (3.1). Разность Δ_z любых двух решений (3.1) удовлетворяет однородным уравнениям $\Delta_z=p\Delta_{z+1}+q\Delta_{z-1}$, а мы уже знаем, что все решения этого уравнения имеют вид $A+B(q/p)^z$. Отсюда следует, что при $p\neq q$ все решения уравнения (3.1) представимы в виде

$$D_z = z/(q-p) + A + B(q/p)^z. \tag{3.3}$$

Для выполнения граничных условий (3.2) требуется, чтобы

$$A+B=0, \quad A+B(q/p)^a = -a/(q-p).$$

---

[1]) Легко видеть, что q_z представляет собой решение разностных уравнений (2.1), удовлетворяющее (теперь единственному) граничному условию $q_0=1$. При $p>q$ это решение не единственно. В действительности наш результат содержится в формуле (3.9) гл. XI и будет получен независимым образом (в усиленной форме) в § 4.

Разрешая эти равенства относительно A и B, находим, что

$$D_z = \frac{z}{q-p} - \frac{a}{q-p} \cdot \frac{1-(q/p)^z}{1-(q/p)^a}. \qquad (3.4)$$

Этот прием опять-таки непригоден при $q=p=1/2$. В этом случае мы заменяем $z/(q-p)$ на $-z^2$, которое является теперь решением (3.1). Отсюда вытекает, что если $p=q=1/2$, то все решения (3.1) имеют вид $D_z=-z^2+A+Bz$. Искомое решение D_z, удовлетворяющее граничным условиям (3.2), равно

$$D_z=z(a-z). \qquad (3.5)$$

Математическое ожидание продолжительности игры в классической задаче о разорении равно (3.4) *при* $p\neq q$ *и* (3.5) *при* $p=q=1/2$.

Следует отметить, что полученные нами значения для продолжительности игры значительно больше, чем мы обычно наивно ожидаем. Если два игрока, имеющие по 500 долларов каждый, бросают монету до тех пор, пока один из них не разорится, то средняя продолжительность игры равна 250 000 испытаниям. Если игрок имеет лишь один доллар, а его противник — 1000, то средняя продолжительность равна 1000 испытаниям. Дальнейшие примеры см. в табл. 1.

Как уже указывалось в конце предыдущего параграфа, мы можем перейти к пределу при $a \to \infty$ и рассмотреть игру против бесконечно богатого соперника. При $p>q$ игра может продолжаться бесконечно, и говорить о математическом ожидании ее продолжительности в этом случае бессмысленно. При $p<q$ мы получаем для математического ожидания продолжительности игры значение $z(q-p)^{-1}$, а при $p=q$ оно бесконечно. (Этот же результат был установлен в гл. XI,3 и будет независимо доказан в следующем параграфе.)

§ 4*). ПРОИЗВОДЯЩИЕ ФУНКЦИИ ДЛЯ ПРОДОЛЖИТЕЛЬНОСТИ ИГРЫ И ДЛЯ ВРЕМЕН ПЕРВОГО ДОСТИЖЕНИЯ

Мы воспользуемся методом производящих функций для изучения продолжительности игры в классической задаче о разорении, т. е. продолжительности ограниченного случайного блуждания с поглощающими экранами в точках $x=0$ и $x=a$. Начальным положением является z $(0<z<a)$. Пусть $u_{z,n}$ — вероятность того, что процесс окончится на n-м шаге на экране $x=0$ (разорение игрока при n-м испытании). После первого шага частица попадает или в точку $x=z+1$, или в точку $x=z-1$, и мы заключаем, что для $1<z<a-1$ и $n\geqslant1$

$$u_{z,\,n+1} = pu_{z+1,\,n} + qu_{z-1,\,n}. \qquad (4.1)$$

*) Этот параграф, как и связанный с ним § 5, может быть опущен при первом чтении.

Это — разностное уравнение, аналогичное (2.1), но зависящее уже от двух переменных z и n. Мы хотим (по аналогии с процедурой из § 2) определить граничные условия $u_{0,\,n}$, $u_{a,\,n}$ и $u_{z,\,0}$ так, чтобы (4.1) было также справедливо для $z=1$, $z=a-1$ и $n=0$. Для этой цели положим

$$u_{0,\,n}=u_{a,\,n}=0 \quad при \quad n\geqslant 1 \tag{4.2}$$

и

$$u_{0,\,0}=1, \quad u_{z,\,0}=0 \quad при \quad 0<z\leqslant a. \tag{4.3}$$

Тогда (4.1) справедливо для всех z при $0<z<a$ и всех $n\geqslant 0$.

Введем теперь производящие функции

$$U_z(s)=\sum_{n=0}^{\infty} u_{z,\,n}s^n. \tag{4.4}$$

Умножая (4.1) на s^{n+1} и суммируя по n, находим

$$U_z(s)=psU_{z+1}(s)+qsU_{z-1}(s), \quad 0<z<a; \tag{4.5}$$

граничные условия (4.2) и (4.3) приводят к равенствам

$$U_0(s)=1, \quad U_a(s)=0. \tag{4.6}$$

Система (4.5) состоит из разностных уравнений, аналогичных (2.1), а граничные условия (4.6) соответствуют (2.2). Новизна здесь заключается в том, что коэффициенты и неизвестные $U_z(s)$ зависят теперь от переменной s, однако для самого разностного уравнения s является просто произвольной постоянной. Если только нам удастся найти два частных решения системы (4.5), то мы снова сможем применять метод § 2. Естественно выяснить, существуют ли два решения $U_z(s)$ вида $U_z(s)=\lambda^z(s)$. Подставляя это выражение в (4.5), мы находим, что $\lambda(s)$ должно удовлетворять квадратному уравнению

$$\lambda(s)=ps\lambda^2(s)+qs, \tag{4.7}$$

имеющему два корня:

$$\lambda_1(s)=\frac{1+\sqrt{1-4pqs^2}}{2ps}, \quad \lambda_2(s)=\frac{1-\sqrt{1-4pqs^2}}{2ps} \tag{4.8}$$

(мы берем $0<s<1$ и положительный квадратный корень).

Теперь у нас есть два частных решения системы (4.5), и, как и в § 2, мы заключаем, что все ее решения имеют вид

$$U_z(s)=A(s)\,\lambda_1^z(s)+B(s)\,\lambda_2^z(s), \tag{4.9}$$

где $A(s)$ и $B(s)$ — произвольные функции. Чтобы удовлетворить граничным условиям (4.6), мы должны иметь $A(s)+B(s)=1$ и $A(s)\,\lambda_1^a(s)+B(s)\,\lambda_2^a(s)=0$, откуда

$$U_z(s)=\frac{\lambda_1^a(s)\,\lambda_2^z(s)-\lambda_1^z(s)\,\lambda_2^a(s)}{\lambda_1^a(s)-\lambda_2^a(s)}. \tag{4.10}$$

Пользуясь очевидным соотношением $\lambda_1(s)\lambda_2(s)=q/p$, это выражение можно упростить:

$$U_z(s)=\left(\frac{q}{p}\right)^z\frac{\lambda_1^{a-z}(s)-\lambda_2^{a-z}(s)}{\lambda_1^a(s)-\lambda_2^a(s)}. \tag{4.11}$$

Это и есть *искомая производящая функция вероятности разорения (поглощения в точке* $x=0$) *при* n-*м испытании.* Тот же метод показывает, что производящая функция для вероятности поглощения в точке $x=a$ дается формулой

$$\frac{\lambda_1^z(s)-\lambda_2^z(s)}{\lambda_1^a(s)-\lambda_2^a(s)}. \tag{4.12}$$

Производящая функция для *продолжительности игры* равна, конечно, сумме производящих функций (4.11) и (4.12).

Бесконечные интервалы и первые достижения

Проведенные выше рассуждения применимы равным образом и к случайным блужданиям на интервале $(0,\infty)$ с поглощающим экраном в точке $x=0$. Частица, выходящая из положения $z>0$, в конце концов поглощается в нуле, в противном случае блуждание продолжается вечно. Поглощение соответствует разорению игрока с начальным капиталом z, играющего против бесконечно богатого соперника. Производящая функция $U_z(s)$ вероятностей $u_{z,n}$ того, что поглощение происходит в точности при n-м испытании, вновь удовлетворяет разностным уравнениям (4.5) и, стало быть, имеет вид (4.9); однако если не выполняется условие $A(s)=0$, то это решение не ограничено на бесконечности. Единственное граничное условие теперь состоит в том, что $U_0(s)=1$, и поэтому $B(s)=1$, или

$$U_z(s)=\lambda_2^z(s) \tag{4.13}$$

(такой же результат можно получить, устремив в (4.11) a к ∞ и воспользовавшись тем, что $\lambda_1(s)\lambda_2(s)=q/p$).

Из (4.13) при $s=1$ следует, что поглощение неизбежно при $p\leqslant q$, а в противном случае имеет вероятность $(q/p)^z$. К такому же выводу мы пришли в § 2.

Рассматриваемое нами поглощение в нуле допускает и другую важную интерпретацию — как первое достижение в *неограниченном* случайном блуждании. Действительно, перенося начало координат в точку z, мы видим, что в начинающемся в нуле случайном блуждании на всей прямой $u_{z,n}$ *будет вероятностью того, что первое попадание в точку* $-z<0$ *произойдет при* n-м *испытании.* То, что соответствующая производящая функция (4.13) является z-й степенью λ_2, отражает тот очевидный факт, что время ожидания до первого прохождения через $-z$ является суммой z независимых времен ожидания между последовательными первыми прохождениями через $-1, -2, \ldots, -z$.

Явная формула для $u_{z,n}$ в частном случае $p=1/2$ была выведена элементарными методами в гл. III (формула (7.5)). Учитывая, что нужно совершить $(n+z)/2$ шагов влево и $(n-z)/2$ шагов вправо, легко заключить, что и в общем случае справедлива та же формула, за тем лишь исключением, что отдельные траектории имеют теперь вероятности $p^{(n-z)/2}q^{(n+z)/2}$, а не 2^{-n}. Таким образом,

$$u_{z,n}=\frac{z}{n}\binom{n}{(n+z)/2}p^{(n-z)/2}q^{(n+z)/2}, \tag{4.14}$$

где биномиальный коэффициент полагается равным нулю, если n и z — числа разной четности. (Относительно вывода этой формулы при помощи производящей функции см. конец § 3 гл. XI. Другой, внешне совершенно отличный вариант явной формулы содержится в задаче 13.)

§ 5*). ЯВНЫЕ ВЫРАЖЕНИЯ

Производящая функция U_z в (4.11) формально зависит от квадратного корня, однако в действительности она является рациональной функцией s. В самом деле, элементарное разложение бинома приводит знаменатель к виду

$$\lambda_1^a(s)-\lambda_2^a(s)=s^{-a}\sqrt{1-4pqs^2}P_a(s), \tag{5.1}$$

где P_a — многочлен четного порядка, равного $a-1$ при нечетном a и $a-2$ — при четном. Числитель имеет тот же вид, только вместо a стоит $a-z$. Таким образом, U_z является отношением двух многочленов, степени которых отличаются самое большее на 1. Следовательно, можно вывести явные выражения для вероятностей разорения при помощи описанного в гл. XI,4 метода разложения на простейшие дроби. Этот результат интересен из-за его связи с теорией диффузии, а его вывод является прекрасной демонстрацией техники, применяемой при практическом использовании простейших дробей.

Вычисления весьма упростятся, если мы воспользуемся вспомогательной переменной φ, определяемой соотношением

$$\cos\varphi=1/(2\sqrt{pq}\,s). \tag{5.2}$$

(Интервалу $0<s<1$ здесь соответствуют комплексные значения φ, однако это не оказывает никакого влияния на формальные выкладки.) Из (4.8) имеем

$$\lambda_1(s)=\sqrt{q/p}\,[\cos\varphi+i\sin\varphi]=\sqrt{q/p}\,e^{i\varphi}, \tag{5.3}$$

тогда как $\lambda_2(s)$ получается из этого выражения заменой i на $-i$. Поэтому

$$U_z(s)=(\sqrt{q/p})^z\sin[(a-z)\varphi]/\sin a\varphi. \tag{5.4}$$

*) См. примечание на с. 363.

Корни знаменателя $s_1, s_2, \ldots$ являются простыми, и, следовательно, существует разложение на простейшие дроби вида

$$(\sqrt{q/p})^z \frac{\sin(a-z)\varphi}{\sin a\varphi} = A + Bs + \frac{\rho_1}{s_1 - s} + \ldots + \frac{\rho_{a-1}}{s_{a-1} - s}. \quad (5.5)$$

В принципе мы должны рассматривать только те корни s_ν, которые не являются в то же время и корнями числителя, но если s_ν — корень числителя, то $U_z(s)$ будет непрерывна при $s = s_\nu$, и, значит, $\rho_\nu = 0$. Стало быть, такие взаимно сокращающиеся корни не дают никакого вклада в правую часть, и нет никакой необходимости рассматривать их отдельно.

Корни $s_1, \ldots, s_{a-1}$ соответствуют, очевидно, значениям $\varphi_\nu = \pi\nu/a$ при $\nu = 1, \ldots, a-1$, так что

$$s_\nu = 1/[2\sqrt{pq}\cos(\pi\nu/a)]. \quad (5.6)$$

Это выражение теряет смысл, когда $\nu = a/2$ и a четно, но в этом случае φ_ν будет также корнем числителя и этот корень должен быть отброшен. Будучи собственным, соответствующий член в окончательном выражении пропадает.

Чтобы вычислить ρ_ν, умножим обе части (5.5) на $s_\nu - s$ и устремим s к s_ν. Вспоминая, что $\sin a\varphi_\nu = 0$ и $\cos a\varphi_\nu = 1$, получаем

$$\rho_\nu = (\sqrt{q/p})^z \sin z\varphi_\nu \cdot \lim_{s \to s_\nu} [(s - s_\nu)/\sin a\varphi].$$

Последний предел определяется при помощи правила Лопиталя и дифференцирования (5.2), что дает

$$\rho_\nu = a^{-1} \cdot 2\sqrt{pq}\,(\sqrt{q/p})^z \sin z\varphi_\nu \cdot \sin\varphi_\nu \cdot s_\nu^2.$$

Из разложения правой части (5.5) в геометрический ряд мы получаем для $n > 1$

$$u_{z,n} = \sum_{\nu=1}^{a-1} \rho_\nu s_\nu^{-n-1} = a^{-1} 2\sqrt{pq}\,(\sqrt{q/p})^z \sum_{\nu=1}^{a-1} s_\nu^{-n+1} \cdot \sin\varphi_\nu \cdot \sin z\varphi_\nu$$

и окончательно

$$u_{z,n} = a^{-1} 2^n p^{(n-z)/2} q^{(n+z)/2} \sum_{\nu=1}^{a-1} \cos^{n-1}(\pi\nu/a) \sin(\pi\nu/a) \sin(\pi z\nu/a). \quad (5.7)$$

Итак, мы вывели явную формулу для *вероятности разорения при n-м испытании*. Она восходит к Лагранжу и многократно выводилась классиками [1]), однако продолжает переоткрываться и в совре-

[1]) Элементарный вывод, основанный на тригонометрической интерполяции, см. в работах Эллиса (Ellis R. E., Cambridge Math. J., 4 (1844); Ellis R. E., Collected works, Cambridge and London, 1863).

менной литературе. Интересно, что метод отражения приводит к другому явному выражению для $u_{z,n}$ — через биномиальные коэффициенты (задача 21). Другой метод вывода (5.7) будет описан в гл. XVI,3.

Переходя к пределу при $a \to \infty$, получаем вероятность того, что в игре против бесконечно богатого соперника игрок с начальным капиталом z будет разорен при n-м испытании. (См. задачу 13.)

Из вида суммы в (5.7) ясно, что члены, соответствующие $\nu=k$ и $\nu=a-k$, равны по абсолютной величине; они имеют один и тот же знак, когда n и z — числа одинаковой четности, и взаимно уничтожаются в противном случае. Следовательно, *при $n-z$ нечетном $u_{z,n}=0$, а при $n-z$ четном и $n>1$*

$$u_{z,\,n}=a^{-1}2^{n+1}p^{(n-z)/2}q^{(n+z)/2}\sum_{\nu<a/2}\cos^{n-1}(\pi\nu/a)\sin(\pi\nu/a)\sin(\pi z\nu/a), \tag{5.8}$$

причем суммирование ведется по всем положительным целым числам ν, меньшим $a/2$. Представление в таком виде более естественно, чем (5.7), поскольку коэффициенты $\cos(\pi\nu/a)$ образуют убывающую последовательность, и поэтому для больших n существенным является только первый член.

§ 6. СВЯЗЬ С ДИФФУЗИОННЫМИ ПРОЦЕССАМИ

Этот параграф посвящен неформальному обсуждению случайных блужданий, в которых длина δ отдельных шагов мала, но шаги столь быстро следуют друг за другом, что движение представляется практически непрерывным. Переход к пределу приводит к винеровскому процессу (броуновскому движению) и другим диффузионным процессам. Такая тесная связь между этими процессами и случайными блужданиями очень способствует пониманию обоих этих объектов [1]). Задачу здесь можно сформулировать и в математических, и в физических терминах.

Лучше всего начать с *неограниченного случайного блуждания*, начинающегося в нуле. В нем n-й шаг приводит частицу в положение $\mathbf{S}_n$, где $\mathbf{S}_n=\mathbf{X}_1+\ldots+\mathbf{X}_n$ — сумма n независимых случайных величин, каждая из которых принимает значения $+1$ и -1 с вероятностями p и q соответственно. Поэтому

$$\mathbf{E}(\mathbf{S}_n)=(p-q)n,\quad \operatorname{Var}(\mathbf{S}_n)=4pqn. \tag{6.1}$$

На рис. 4 в гл. III,6 представлены первые 10 000 шагов такого случайного блуждания с $p=q=1/2$; чтобы уместить его график на

[1]) Этот подход был плодотворным и исторически. Он был сполна использован (хотя и с эвристических позиций) Л. Башелье, работа которого вдохновила А. Н. Колмогорова на развитие формальных оснований теории марковских процессов. См., в частности, Bachelier L., Calcul des probabilités, Paris, Gauthier-Villars, 1912.

странице книги, пришлось выбрать подходящие масштабы для обеих осей. Теперь мы пойдем дальше и представим себе запечатлевший случайное блуждание кинофильм. Предположим, что он должен длиться 1000 с (почти 17 минут). Для того чтобы это соответствовало миллиону шагов, необходимо, чтобы случайное блуждание происходило со скоростью один шаг в миллисекунду, и это фиксирует шкалу времени. Какую же величину шага должны мы выбрать для блуждания, чтобы быть вполне уверенными, что запись его поместится на экране заданной высоты? Итак, мы рассматриваем вместо $\mathbf{S}_n$ величины $\delta\mathbf{S}_n$, где δ означает длину отдельного шага. Теперь

$$\mathrm{E}(\delta\mathbf{S}_n)=(p-q)\delta n, \quad \mathrm{Var}(\delta\mathbf{S}_n)=4pq\delta^2 n, \tag{6.2}$$

и из центральной предельной теоремы ясно, что наш фильм будет возможен только в том случае, если при $n=1\ 000\ 000$ обе величины [1]) (6.2) будут меньше высоты экрана. (Здесь подразумевается, что мы пользуемся фиксированной единицей измерения, скажем дюймами или футами, как для экрана, так и для величины отдельных шагов.) Но если $p\neq q$ и δn сравнимо с высотой экрана, то $\delta^2 n$ будет неотличимо от нуля и фильм покажет нам линейное движение без видимых случайных флуктуаций. Характер случайного блуждания может быть выявлен лишь тогда, когда $\delta^2 n$ имеет умеренную положительную величину, а это возможно только при $p-q$, сравнимом с δ.

Если бы вопрос был чисто математическим, мы заключили бы, что желаемое графическое представление невозможно, если только не выполняется условие $p=q$, однако с физической точки зрения ситуация выглядит совершенно иначе. В броуновском движении мы наблюдаем двигающиеся случайным образом взвешенные в жидкости частицы, и естественно возникает вопрос: можно ли интерпретировать это движение как результат огромного числа столкновений с более мелкими частицами в жидкости? Конечно, предполагать, что столкновения происходят через равные интервалы времени и что каждое столкновение вызывает перемещение, в точности равное $\pm\delta$,— это сверхупрощение. Во всяком случае, для начала мы будем считать, что столкновения управляются испытаниями Бернулли, и посмотрим, совместимо ли с этой картиной наблюдаемое движение частиц. Из реальных наблюдений мы находим среднее смещение c и дисперсию D за единичный интервал времени. Обозначим через r (неизвестное) число столкновений за единицу времени. Тогда мы приближенно должны иметь

$$(p-q)\delta r=c, \quad 4pq\delta^2 r=D. \tag{6.3}$$

В моделирующем эксперименте не наблюдалось бы никаких случайных флуктуаций, если бы оба условия (6.3) не выполнялись при

[1]) Точнее, первая из них и корень квадратный из второй.— *Прим. перев.*

$D>0$. Вполне мыслим эксперимент с $p=0{,}6$ и $\delta r=1$, однако в нем дисперсия была бы столь малой, что движение казалось бы детерминированным: группа частиц, сначала близких друг к другу, так и оставалась бы вместе, как если бы была твердым телом.

По существу те же рассуждения применимы и ко многим другим явлениям в физике, экономике, теории обучения, эволюционной теории и т. д., когда медленные флуктуации состояния системы интерпретируются как результат огромного числа последовательных малых изменений, вызванных случайными воздействиями. Простая модель случайного блуждания не представляется реалистичной ни в одном конкретном случае, но, к счастью, ситуация здесь подобна той, что имеет место в центральной предельной теореме. Характер отдельных изменений не будет играть никакой роли уже при удивительно слабых условиях, ибо наблюдаемый эффект будет зависеть лишь от их математических ожиданий и дисперсий [1]). При таких обстоятельствах естественно принять в качестве универсального прототипа простую модель случайного блуждания.

Итак, в качестве подготовки к более глубокому изучению различных стохастических процессов естественно рассмотреть случайные блуждания, в которых длина δ отдельных шагов мала, число r шагов за единицу времени велико, а $p-q$ мало, причем соотношение между ними таково, что справедливо (6.3) (где c и $D>0$ — заданные постоянные). Понятия «мало» и «велико» строго не определены и в практических применениях должны варьироваться [2]).

Аналитически проблема формулируется следующим образом. Каждому выбору δ, r и p соответствует конкретное случайное блуждание. *Спрашивается, что произойдет в пределе, когда* $\delta\to 0$, $r\to\infty$, *и* $p\to 1/2$ *так, что*

$$(p-q)\delta r\to c,\quad 4pq\delta^2 r\to D. \tag{6.4}$$

Тут возможны два подхода. Если мы располагаем явными выражениями для нужных нам вероятностей, то можем непосредственно перейти к пределу. Мы продемонстрируем этот метод, поскольку он проливает новый свет на нормальное приближение и предельные теоремы, выведенные в гл. III. Однако возможности этого метода ограничены, ибо он не годится для обобщений. Более

---

[1]) Это так называемый принцип инвариантности, впервые установленный М. Донскером. Подробное изложение этого вопроса см. в книге Биллингсли П. Сходимость вероятностных мер.— М.: Наука, 1977.— *Прим. перев.*

[2]) Число молекулярных толчков в единицу времени находится за пределом воображения. В противоположность этому в эволюционной теории рассматриваются небольшие изменения от одного поколения к другому, и разделяющее два поколения время по обыденным стандартам отнюдь не является малым. Число рассматриваемых поколений также не является фантастическим, но может составлять многие тысячи. Здесь дело в том, что этот процесс протекает так, что изменения кажутся практически непрерывными, и диффузионная модель с непрерывным временем предпочтительнее модели случайного блуждания.

плодотворным будет начать с разностных уравнений, описывающих случайные блуждания, и вывести из них предельные дифференциальные уравнения. Оказывается, что эти дифференциальные уравнения описывают вполне определенные стохастические процессы, зависящие от непрерывного временнóго параметра. То же справедливо и относительно различных очевидных обобщений этих дифференциальных уравнений, так что второй метод приводит к важному общему классу диффузионных процессов.

Чтобы описать прямой метод в простейшем случае, будем попрежнему обозначать через $\{\mathbf{S}_n\}$ обычное случайное блуждание с единичными шагами и положим

$$v_{k,\,n}=\mathbf{P}\,\{\mathbf{S}_n=k\}. \tag{6.5}$$

В нашем ускоренном блуждании n-й шаг произойдет в момент времени n/r, а положением частицы в этот момент будет $\mathbf{S}_n\delta=k\delta$. Мы интересуемся вероятностью найти частицу в данный момент t в окрестности заданной точки x и поэтому должны изучать асимптотическое поведение $v_{k,n}$, когда $k\to\infty$ и $n\to\infty$ таким образом, что $n/r\to t$ и $k\delta\to x$. Событие $\{\mathbf{S}_n=k\}$ требует, чтобы n и k были числами одинаковой четности, и происходит, когда ровно $(n+k)/2$ из первых n шагов направлены вправо. Стало быть, из теоремы Муавра — Лапласа мы заключаем, что при переходе к пределу

$$v_{k,\,n}\sim\frac{1}{\sqrt{2\pi npq}}\,e^{-[(n+k)/2-np]^2/(2npq)}=\frac{1}{\sqrt{2\pi npq}}\,e^{-[k-n(p-q)]^2/(8npq)}\sim$$
$$\sim\frac{2\delta}{\sqrt{2\pi Dt}}\,e^{-(x-ct)^2/(2Dt)}, \tag{6.6}$$

где знак $\sim$ означает, что отношение величин стремится к единице. Поскольку $v_{k,\,n}$ есть вероятность того, что $\mathbf{S}_n\delta$ находится между $k\delta$ и $(k+2)\delta$, а этот интервал имеет длину 2δ, мы можем сказать, что отношение $v_{k,n}/(2\delta)$ локально измеряет вероятность на единицу длины, т. е. является *плотностью* вероятности. Из второго соотношения (6.6) вытекает, что отношение $v_{k,\,n}/(2\delta)$ стремится к

$$v(t,\;x)=\left(1/\sqrt{2\pi Dt}\right)e^{-(x-ct)^2/(2Dt)}. \tag{6.7}$$

Отсюда следует, что суммы вероятностей $v_{k,\,n}$ можно аппроксимировать интегралами от $v(t,\;x)$, и наш результат может быть переформулирован следующим образом: при выбранном нами переходе к пределу

$$\mathbf{P}\,\{\alpha<\mathbf{S}_n\delta<\beta\}\to\left(1/\sqrt{2\pi Dt}\right)\int_\alpha^\beta e^{-(x-ct)^2/(2Dt)}\,dx. \tag{6.8}$$

Входящий сюда интеграл можно выразить через функцию нормального распределения $\mathfrak{N}$, и (6.8) в действительности является лишь вариантом записи предельной теоремы Муавра — Лапласа для биномиального распределения.

Более интересен подход, основанный на соответствующих разностных уравнениях. Если рассмотреть положение частицы при n-м и $(n+1)$-м испытаниях, то будет очевидно, что вероятности $v_{k,n}$ удовлетворяют разностным уравнениям

$$v_{k,\,n+1} = pv_{k-1,\,n} + qv_{k+1,\,n}. \tag{6.9}$$

Умножая это равенство на 2δ, мы видим (как следует из нашего предыдущего результата), что $v(t, x)$ должно быть *приближенным* решением разностного уравнения

$$v(t+r^{-1},\ x) = pv(t,\ x-\delta) + qv(t,\ x+\delta). \tag{6.10}$$

Поскольку v имеет непрерывные производные, мы можем разложить члены в этом равенстве в ряды Тейлора. Пользуясь приближением первого порядка в левой части и второго порядка — в правой, мы получаем (после сокращения главных членов)

$$\frac{\partial v(t,x)}{\partial t} = (q-p)\,\delta r\,\frac{\partial v(t,x)}{\partial x} + \frac{1}{2}\,\delta^2 r\,\frac{\partial^2 v(t,x)}{\partial x^2} + \ldots\,. \tag{6.11}$$

При переходе к пределу опущенные нами члены устремятся к нулю и (6.11) превратится в

$$\frac{\partial v(t,x)}{\partial t} = -c\,\frac{\partial v(t,x)}{\partial x} + \frac{1}{2}\,D\,\frac{\partial^2 v(t,x)}{\partial x^2}. \tag{6.12}$$

Это частный вид *уравнения диффузии*, известный также под названием *уравнения Фоккера — Планка*. Наши выкладки были исключительно формальными и эвристическими, однако не будет неожиданным, что функция v из (6.7) действительно удовлетворяет дифференциальному уравнению (6.12). Более того, можно показать, что (6.7) представляет собой единственное решение уравнения диффузии, обладающее очевидными свойствами, обусловленными вероятностной интерпретацией.

Уравнение диффузии (6.12) можно обобщить, допустив, что коэффициенты c и D зависят от x и t. Кроме того, у него есть очевидные аналоги в многомерном случае, и все эти обобщения могут быть выведены непосредственно из общих вероятностных постулатов. Этот вопрос будет обсуждаться в гл. X тома 2; здесь же мы должны довольствоваться краткими и эвристическими указаниями на связь между случайными блужданиями и общей теорией диффузии.

В качестве второго примера мы рассмотрим вероятности разорения $u_{z,\,n}$, обсуждавшиеся в предыдущих двух параграфах. Отвечающие им разностные уравнения (4.1) отличаются от (6.9) тем, что коэффициенты p и q в них меняются местами[1]). Формальные выкладки, указанные в (6.11), приводят теперь к уравнениям диффузии, получающимся из (6.12) при замене $-c$ на c. При нашей предельной процедуре мы переходим от вероятностей $u_{z,n}$ к удовлетворяющей модифицированному уравнению диффузии функции $u(t, \xi)$, которая по вероятностному смыслу подобна $u_{z,n}$: в диффузионном процессе, начинающемся в точке $\xi>0$, вероятность того, что частица попадет в нуль *до* момента достижения точки $\alpha > \xi$ и что это событие произойдет в интервале времени $t_1<t<t_2$, дается интегралом от $u(t, \xi)$ по этому интервалу.

Формальные вычисления выглядят следующим образом. Для $u_{z,n}$ у нас есть явное выражение (5.8). Поскольку z и n должны быть числами одинаковой четности, то $u_{z,n}$ соответствует интервалу между n/r и $(n+2)/r$, и мы должны вычислить предел отношения $u_{z,n}r/2$ при $r\to\infty$ и $\delta\to 0$ в соответствии с (6.4). Длина интервала a и начальное положение z должны быть подобраны так, чтобы получить в пределе α и ξ. Таким образом, $z \sim \xi/\delta$ и $a \sim \alpha/\delta$. Теперь легко получить пределы для отдельных множителей в (5.8).

Из первого соотношения (6.4) мы получаем, что $2p \sim 1+c\delta/D$ и $2q \sim 1-c\delta/D$; из второго соотношения (6.4) видим, что $\delta^2 r \to D$. Стало быть,

$$(4pq)^{n/2} (q/p)^{z/2} \sim (1-c^2\delta^2/D^2)^{rt/2} (1-2c\delta/D)^{\xi/(2\delta)} \sim e^{-c^2t/(2D)} \cdot e^{-c\xi/D}. \tag{6.13}$$

Аналогично для фиксированного ν

$$\cos^n (\nu\pi\delta/\alpha) \sim [1-\nu^2\pi^2\delta^2/(2\alpha^2)]^{tr} \sim e^{-\nu^2\pi^2 Dt/(2\alpha^2)}. \tag{6.14}$$

Наконец, $\sin \nu\pi\delta/\alpha \sim \nu\pi\delta/\alpha$. Подстановка полученных выражений в (5.8) формально приводит к результату

$$u(t, \xi) = \pi D\alpha^{-2} e^{-(ct+2\xi)c/(2D)} \sum_{\nu=1}^{\infty} \nu e^{-\nu^2\pi^2 Dt/(2\alpha^2)} \sin(\pi\xi\nu/\alpha). \tag{6.15}$$

(Поскольку этот ряд сходится равномерно, нетрудно показать законность наших формальных выкладок.) В физической теории диффузии формула (6.15) называется *формулой Фюрта для первого прохождения*. (Относительно предельного случая $\alpha=\infty$ см. задачу 14. Другую форму (6.15) см. в задаче 22.)

---

[1]) Причина этого состоит в том, что в $u_{z,n}$ переменная z означает *начальное* положение, тогда как вероятность $v_{k,n}$ относится к положению в текущий момент времени. В терминологии, которая будет введена в томе 2, вероятности, зависящие от начального положения, удовлетворяют *обратным* уравнениям, а остальные — *прямым* уравнениям (или уравнениям Фоккера — Планка). В физике последние иногда называются *уравнениями неразрывности*. С такой же ситуацией мы столкнемся в гл. XVII.

§ 7*). СЛУЧАЙНЫЕ БЛУЖДАНИЯ НА ПЛОСКОСТИ И В ПРОСТРАНСТВЕ

В двумерном случайном блуждании частица передвигается единичными шагами в четырех направлениях, параллельных осям x и y. Для выходящей из начала координат частицы возможными положениями будут все точки плоскости с целочисленными координатами. Каждое положение имеет четыре *соседних*. Аналогично в трех измерениях каждое положение имеет шесть соседних. Случайное блуждание определяется выбором соответствующих четырех или шести вероятностей. Для простоты мы будем рассматривать только *симметричный* случай, в котором все направления равновероятны. Сложность задач здесь намного выше, чем в одномерном случае, ибо теперь области, в которых происходит движение частицы, могут иметь произвольную форму и роль экранов играют границы сложного вида.

Мы начнем с интересной теоремы, принадлежащей Пойа [1]).

Теорема. *В одномерном и двумерном симметричных случайных блужданиях частица с вероятностью единица рано или поздно (а поэтому и бесконечное число раз) возвратится в свое начальное положение. Однако в трехмерном случае вероятность этого меньше единицы и равна примерно* 0,35. (Среднее число возвращений равно тогда $0{,}65\sum k(0{,}35)^k = 0{,}35/0{,}65 \approx 0{,}53$.)

Прежде чем доказывать эту теорему, приведем две другие ее формулировки, также принадлежащие Пойа. Прежде всего, почти очевидно, что из этой теоремы вытекает, что *в одномерном и в двумерном случаях с вероятностью* 1 *частица бесконечное число раз пройдет через каждое возможное положение*, однако в трехмерном случае это неверно. Таким образом, для двух измерений в известном смысле справедливо утверждение «все дороги ведут в Рим».

С другой стороны, рассмотрим *две* частицы, совершающие независимые симметричные случайные блуждания, причем перемещения их происходят одновременно. Встретятся ли они когда-нибудь? Чтобы упростить изложение, мы определим *расстояние* между двумя возможными положениями как наименьшее число шагов, ведущих из одного положения в другое. (Это расстояние равно сумме абсолютных величин разностей координат.) Если две частицы продвигаются на один шаг каждая, то расстояние между ними либо останется тем же, либо изменится на две единицы, и поэтому расстояние между частицами будет либо всегда четным, либо всегда нечетным. Во втором случае наши две частицы никогда не смогут занять одно и то же положение. В первом случае легко видеть, что веро-

*) Этот параграф посвящен специальному вопросу и может быть опущен при первом чтении.

[1]) Polya G., Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt in Strassennetz, Mathematische Annalen, 84 (1921), 149—160. Численное значение 0,35 было найдено в работе Мак-Кри и Уиппла (McCrea W. H., Whipple F. J. W., Random paths in two and three dimensions, Proceedings of the Royal Society of Edinburgh, 60 (1940), 281—298).

ятность их встречи на n-м шаге равна вероятности того, что первая частица за $2n$ шагов достигнет начального положения второй частицы.

Следовательно, наша теорема утверждает, что в двумерном (но не в трехмерном) случае две частицы наверняка бесконечное число раз будут занимать одно и то же положение. Если начальное расстояние между двумя частицами нечетно, то аналогичное рассуждение показывает, что они будут бесконечно много раз занимать соседние положения. Если назвать это встречей, то наша теорема утверждает, что *в одномерном и двумерном случаях две частицы с достоверностью встретятся бесконечное число раз, однако в трехмерном случае они с положительной вероятностью никогда не встретятся.*

Доказательство. Для одномерного случая теорема была доказана в примере гл. XIII, 4, б) при помощи метода рекуррентных событий. Доказательство для двумерного и трехмерного случаев проводится примерно так же. Пусть u_n — вероятность того, что n-е испытание приведет частицу в начальное положение. Согласно теореме 2 гл. XIII,3, нам надо доказать, что в двумерном случае ряд $\sum\limits_{n\geqslant 1} u_n$ расходится, тогда как в трехмерном случае $\sum\limits_{n\geqslant 1} u_n \approx 0,53$. В двумерном случае возвращение в начальное положение возможно только тогда, когда количества шагов в положительных направлениях осей x и y равны соответственно количествам шагов в отрицательных направлениях этих осей. Следовательно, $u_n=0$, если n нечетно, и (мы пользуемся полиномиальным распределением; см. формулу (9.2) гл. VI)

$$u_{2n}=\frac{1}{4^{2n}}\sum_{k=0}^{n}\frac{(2n)!}{k!\,k!\,(n-k)!\,(n-k)!}=\frac{1}{4^{2n}}\binom{2n}{n}\sum_{k=0}^{n}\binom{n}{k}^2. \qquad (7.1)$$

Согласно формуле (12.11) гл. II, правая часть равна $4^{-2n}\binom{2n}{n}^2$. Формула Стирлинга показывает теперь, что u_{2n} является величиной порядка $1/n$, так что ряд $\sum u_{2n}$, как и утверждалось, расходится.

Для трехмерного случая аналогично находим

$$u_{2n}=6^{-2n}\sum_{j,k}\frac{(2n)!}{j!\,j!\,k!\,k!\,(n-j-k)!\,(n-j-k)!}, \qquad (7.2)$$

где суммирование проводится по всем j и k, таким, что $j+k\leqslant n$. Легко убедиться, что

$$u_{2n}=\frac{1}{2^{2n}}\binom{2n}{n}\sum_{j,k}\left\{\frac{1}{3^n}\frac{n!}{j!\,k!\,(n-j-k)!}\right\}^2. \qquad (7.3)$$

Внутри скобок стоят члены триномиального распределения, и мы знаем, что в сумме они составляют единицу. Следовательно, сумма их квадратов не превосходит максимального из членов в скобках, а последнему соответствуют значения j и k, равные примерно $n/3$. Формула Стирлинга показывает, что этот максимум имеет порядок величины n^{-1}, и поэтому u_{2n} имеет величину $1/\sqrt{n^3}$, так что ряд $\sum u_{2n}$, как и утверждалось, сходится. ▶

Мы завершаем этот параграф еще одной задачей, которая обобщает понятие *поглощающего экрана*. Рассмотрим двумерный случай, в котором вместо интервала $0 \leqslant x \leqslant a$ у нас имеется плоская область D, т. е. набор точек с целочисленными координатами. Каждая точка имеет четыре соседних, однако у некоторых точек из D одна или более соседних точек лежат вне D. Такие точки образуют границу области D, а все остальные ее точки называются внутренними.

В одномерном случае границу образуют два экрана, и наша задача состояла в нахождении вероятности того, что, выходя из z, частица достигнет граничной точки $x=0$ до попадания в точку $x=a$. Теперь по аналогии мы интересуемся вероятностью того, что частица достигнет определенного участка границы до того момента, как она попадет в какую-либо граничную точку, не принадлежащую этому участку. Это означает, что мы разбиваем все граничные точки на два множества B' и B''. Если (x, y) является внутренней точкой, то мы ищем вероятность того, что выходящая из (x, y) частица достигнет какой-нибудь точки из B' ранее, чем любой точки из B''. В частности, если B' состоит из одной-единственной точки, то $u(x, y)$ будет вероятностью того, что частица рано или поздно будет поглощена в этой точке.

Пусть (x, y) — некоторая внутренняя точка. Первый шаг переводит частицу из (x, y) в одну из четырех соседних ей точек $(x \pm 1, y)$, $(x, y \pm 1)$, и если все четыре из них суть внутренние точки, то, очевидно,

$$u(x, y) = (1/4)[u(x+1, y) + u(x-1, y) + u(x, y+1) + u(x, y-1)]. \quad (7.4)$$

Это и есть разностное уравнение, которое появляется здесь вместо (2.1) (при $p=q=1/2$). Если $(x+1, y)$ является граничной точкой, то ее вклад $u(x+1, y)$ следует заменить на 1 и 0 в соответствии с тем, принадлежит ли $(x+1, y)$ множеству B' или B''. *Следовательно*, (7.4) *будет справедливо для всех внутренних точек, если мы будем считать, что* $u(\xi, \eta)=1$ *для граничной точки* (ξ, η) *из* B' *и* $u(\xi, \eta)=0$ *для граничной точки из* B''. *Это соглашение играет здесь роль граничных условий* (2.2).

Итак, мы получили систему линейных уравнений (7.4) для неизвестных $u(x, y)$; каждой внутренней точке соответствуют в ней одно неизвестное и одно уравнение. Эта система неоднородна, поскольку имеется хотя бы одна граничная точка (ξ, η) из B', привносящая в правую часть вклад 1/4. Если область D конечна, то уравнений у нас столько же, сколько и неизвестных, и известно, что такая система имеет единственное решение тогда и только тогда, когда соответствующая однородная система (с $u(\xi, \eta)=0$ для всех граничных точек) не имеет ненулевых решений. Но $u(x, y)$ является средним арифметическим четырех соседних значений $u(x\pm 1, y)$, $u(x, y\pm 1)$ и не может быть больше (а также меньше) всех этих значений. Иначе говоря, во внутренних точках $u(x, y)$ не имеет ни максимума, ни минимума в строгом смысле, и наибольшее и наименьшее ее значения достигаются в граничных точках. Следовательно, если все граничные значения равны нулю, то таковы же значения $u(x, y)$ и во всех внутренних точках, что доказывает существование и единственность решения (7.4). Поскольку граничные значения равны 0 или 1, все значения $u(x, y)$ лежат между 0 и 1, как и требуется для вероятностей. Как мы увидим из общей теоремы для бесконечных цепей Маркова, эти утверждения справедливы и для случая бесконечных областей [1]).

§ 8*). ОБОБЩЕННОЕ ОДНОМЕРНОЕ СЛУЧАЙНОЕ БЛУЖДАНИЕ (ПОСЛЕДОВАТЕЛЬНЫЙ АНАЛИЗ)

Вернемся теперь к одномерному случаю, отказавшись от ограничения, что частица движется единичными шагами. Вместо этого *частица на каждом шаге будет иметь вероятность p_k перейти из точки x в точку $x+k$*, где целое k может быть нулем, положительным или отрицательным числом. Мы будем рассматривать следующую *задачу о разорении. Частица выходит из точки z, такой, что $0<z<a$; мы ищем вероятность u_z того, что частица достигает какой-либо точки $x\leqslant 0$ раньше, чем любой из точек $x\geqslant a$.* Иначе говоря, после n-го испытания частица будет находиться в точке $x=z+\mathbf{X}_1+\mathbf{X}_2+\ldots\mathbf{X}_n$ на оси x, где $\{\mathbf{X}_k\}$ — взаимно независимые случайные величины с общим распределением $\{p_\nu\}$; процесс останавливается, когда впервые будет выполнено либо неравенство $\mathbf{X}_1+\ldots+\mathbf{X}_n\leqslant -z$, либо неравенство $\mathbf{X}_1+\ldots+\mathbf{X}_n\geqslant a-z$.

Эта задача привлекла широкий интерес в связи с *последовательным анализом*. Там $\mathbf{X}_k$ представляют некоторые характеристики выборок или наблюдений. Измерения производятся до тех пор, пока сумма $\mathbf{X}_1+\ldots+\mathbf{X}_k$ не выйдет за пределы двух заранее определенных границ (наших $-z$ и $a-z$). В первом случае эта процедура приводит (если следовать специальной терминологии) к тому, что гипотеза отклоняется, а во втором случае — к тому, что она принимается [2]).

---

[1]) Явные решения известны лишь для немногих случаев и имеют, как правило, очень сложный вид. Решения для случаев прямоугольных областей, бесконечных полос и т. д. можно найти в указанной в предыдущем примечании статье Мак-Кри и Уиппла.

*) Материал этого параграфа в дальнейшем не используется.

[2]) Общая теория последовательных статистических процедур была развита Абрахамом Вальдом во время второй мировой войны в связи с важными практическими задачами. Ее современное изложение можно найти в ряде курсов математической статистики. Описанная в примере схема Бартки относится к 1943 г. и яв-

Пример. а) В качестве иллюстрации возьмем предложенную Бартки многовыборочную схему контроля. Чтобы проверить партию изделий, производятся выборки объема N, которые и подвергаются экспертизе. Предполагается, что эти выборки стохастически независимы и что число дефектных изделий в каждой из них имеет одно и то же биномиальное распределение. Допускается наличие одного дефектного изделия на выборку, и мы положим $\mathbf{X}_k+1$ равным числу дефектных элементов k-й выборки. Тогда для $k\geqslant 0$

$$p_k=\binom{N}{k+1}p^{k+1}q^{N-k-1},$$

и $p_{-1}=q^N$, $p_x=0$ для $x<-1$.

Процедурное правило состоит в следующем. Извлекается первая выборка, и если она не содержит ни одного дефектного изделия, то принимается вся партия; если число дефектных образцов в ней превосходит a, то вся партия бракуется. В любом из этих случаев процесс прекращается. Однако если число z дефектных изделий лежит в интервале $1\leqslant z\leqslant a$, то процедура извлечения выборок продолжается описанным образом до тех пор, пока сумма величин $\mathbf{X}_k$ остается между 1 и a. Рано или поздно она либо обратится в нуль, и тогда партия принимается, либо станет $\geqslant a$, и тогда партия бракуется. ►

Предположим, не теряя общности, что шаги возможны как в положительном, так и в отрицательном направлениях. В противном случае мы имели бы либо $u_z=0$, либо $u_z=1$ для всех z. Вероятность разорения на *первом* шаге, очевидно, равна

$$r_z=p_{-z}+p_{-z-1}+p_{-z-2}+\ldots \tag{8.1}$$

(величина, которая может быть равна нулю). Случайное блуждание продолжается только в том случае, когда частица перемещается в положение x, $0<x<a$; вероятность скачка из z в x равна p_{x-z}, и вероятность последующего разорения составляет тогда u_x. Поэтому

$$u_z=\sum_{x=1}^{a-1}u_xp_{x-z}+r_z. \tag{8.2}$$

Мы снова имеем здесь $a-1$ линейных уравнений для $a-1$ неизвестных u_z. Эта система неоднородна, поскольку по крайней мере при $z=1$ вероятность r_1 отлична от нуля (ибо возможны шаги в отрицательном направлении). Чтобы доказать, что линейная система (8.2) имеет единственное решение, мы должны доказать, что соот-

---

ляется, по-видимому, самой первой последовательной процедурой, предложенной в литературе. [Подробнее об этом см. книги Вальд А. Последовательный анализ.— М.: Физматгиз, 1960; Ширяев А. Н. Статистический последовательный анализ.— 2-е изд., перераб.— М.: Наука, 1976.— *Перев.*]

ветствующая однородная система

$$u_z = \sum_{x=1}^{a-1} u_x p_{x-z} \tag{8.3}$$

не имеет ненулевых решений. Предположим (для сокращения числа появляющихся в доказательстве индексов), что $p_{-1} \neq 0$ (однако те же рассуждения применимы и при иных положительных членах с отрицательным индексом). Допустим теперь, что u_z удовлетворяет (8.3), и обозначим через M максимум значений u_z. Пусть $u_r = M$. Поскольку сумма всех коэффициентов p_{x-z} в (8.3) равна единице, то равенство это возможно при $z=r$ только тогда, когда те из u_x, которые действительно входят в правую часть (с положительными коэффициентами), равны M и если коэффициенты при них в сумме дают единицу. Следовательно, $u_{r-1}=M$, и, рассуждая таким же образом, получаем, что $u_{r-2}=u_{r-3}=\ldots=u_1=M$. Однако при $z=1$ сумма коэффициентов p_{x-r} в (8.3) меньше единицы, так что M должно быть равно нулю.

Отсюда следует, что (8.2) имеет единственное решение, так что наша задача вполне определена. Снова упростим запись, введя граничные условия

$$u_x = \begin{cases} 1, & \textit{если } x \leqslant 0, \\ 0, & \textit{если } x \geqslant a. \end{cases} \tag{8.4}$$

Тогда (8.2) можно переписать в виде

$$u_z = \sum u_x p_{x-z}, \tag{8.5}$$

и суммирование теперь ведется уже по всем x (слагаемые при $x \geqslant a$ не вносят никакого вклада в силу второго условия из (8.4); в силу первого условия слагаемые для $x \leqslant 0$ дают в сумме r_z).

При больших a непосредственное решение $a-1$ линейного уравнения громоздко, и лучше воспользоваться *методом частных решений*, аналогичным процедуре § 2. Он хорошо работает, когда распределение вероятностей $\{p_k\}$ содержит сравнительно немного положительных членов. Предположим, что отличны от нуля только p_k при $-\nu \leqslant k \leqslant \mu$, так что наибольшие возможные скачки в положительном и отрицательном направлениях суть μ и ν соответственно. *Характеристическое уравнение*

$$\sum p_k \sigma^k = 1 \tag{8.6}$$

эквивалентно алгебраическому уравнению степени $\nu+\mu$. Если σ — корень (8.6), то $u_z = \sigma^z$ будет формальным решением (8.5) для всех z, однако это решение не удовлетворяет граничным условиям (8.4). Если (8.6) имеет $\mu+\nu$ различных корней $\sigma_1, \sigma_2, \ldots$, то линейная комбинация

$$u_z = \sum A_k \sigma_k^z \tag{8.7}$$

снова будет формальным решением (8.5) при всех z, и мы должны подобрать постоянные A_k так, чтобы удовлетворялись граничные условия. Далее, при $0<z<a$ в (8.5) входят лишь значения x из интервала $-\nu+1\leqslant x\leqslant a+\mu-1$. Поэтому достаточно удовлетворить граничным условиям (8.4) только для $x=0, -1, -2, \ldots, -\nu+1$ и $x=a, a+1, \ldots, a+\mu-1$, так что всего мы имеем $\mu+\nu$ условий. Если σ_k — двойной корень (8.6), то мы теряем одну постоянную, но в этом случае легко видеть, что еще одним формальным решением будет $u_z=z\sigma_k^z$. В любом случае $\mu+\nu$ граничных условий определяют $\mu+\nu$ произвольных постоянных.

Пример. б) Предположим, что каждый отдельный шаг переводит частицу в одно из четырех ближайших положений и что $p_{-2}=p_{-1}=p_1=p_2=1/4$. Характеристическое уравнение (8.6) здесь имеет вид $\sigma^{-2}+\sigma^{-1}+\sigma+\sigma^2=4$. Положим $t=\sigma+\sigma^{-1}$: при такой подстановке наше уравнение превращается в $t^2+t=6$, корни которого суть $t=2$ и $t=-3$. Разрешая $t=\sigma+\sigma^{-1}$ относительно σ, находим четыре корня

$$\sigma_1=\sigma_2=1, \quad \sigma_3=\frac{-3+\sqrt{5}}{2}=\sigma_4^{-1}, \quad \sigma_4=\frac{-3-\sqrt{5}}{2}=\sigma_3^{-1}. \tag{8.8}$$

Поскольку σ_1 — двойной корень, общим решением уравнения (8.5) в нашей случае является

$$u_z=A_1+A_2z+A_3\sigma_3^z+A_4\sigma_4^z. \tag{8.9}$$

Граничные условия $u_0=u_{-1}=1$ и $u_a=u_{a+1}=0$ приводят к четырем линейным уравнениям для коэффициентов A_j и к окончательному решению

$$u_z=1-\frac{z}{a}+\frac{(2z-a)(\sigma_3^a-\sigma_4^a)-a(\sigma_3^{2z-a}-\sigma_4^{2z-a})}{a\{(a+2)(\sigma_3^a-\sigma_4^a)-a(\sigma_3^{a+2}-\sigma_4^{a+2})\}}. \tag{8.10}$$

▶

Численные приближения. Найти все корни уравнения (8.6) обычно бывает весьма затруднительно, однако для них можно получить вполне удовлетворительные приближения удивительно простым способом. Рассмотрим сперва случай, когда распределение вероятностей $\{p_k\}$ имеет нулевое среднее. Тогда характеристическое уравнение (8.6) имеет двойной корень в точке $\sigma=1$ и $A+Bz$ является формальным решением (8.5). Конечно, двух постоянных A и B недостаточно для того, чтобы удовлетворялись $\mu+\nu$ граничных условий (8.4). Однако если мы определим A и B так, чтобы $A+Bz$ обращалось в нуль при $z=a+\mu-1$ и было равно 1 при $z=0$, то $A+Bx\geqslant 1$ при $x\leqslant 0$ и $A+Bx\geqslant 0$ при $a\leqslant x<a+\mu$, так что $A+Bz$ будет удовлетворять граничным условиям (8.4), в которых знак $=$ заменен на $\geqslant$. Следовательно, разность $A+Bz-u_z$ будет формальным решением (8.5) с неотрицательными граничными значениями, и, стало быть, $A+Bz-u_z\geqslant 0$. Подобным же образом мы можем получить и нижнюю границу для u_z, определяя A и B так, чтобы $A+Bz$ обращалось в нуль при $z=a$ и в единицу при $z=-\nu+1$. Следовательно,

$$\frac{a-z}{a+\nu-1}\leqslant u_z\leqslant\frac{a+\mu-z-1}{a+\mu-1}. \tag{8.11}$$

Эта оценка очень точна, когда a велико по сравнению с $\mu+\nu$. [Конечно, приближение $u_z\approx(1-z/a)$ лучше, однако оно не дает точных границ для u_z.]

Рассмотрим теперь общий случай, когда среднее распределения $\{p_k\}$ отлично от нуля. Тогда характеристическое уравнение (8.6) имеет в точке $\sigma=1$ простой корень. Левая часть (8.6) стремится к ∞ при $\sigma\to 0$ и при $\sigma\to\infty$. При положительных σ кривая $y=\sum p_k\sigma^k$ непрерывна и выпукла вниз, и поскольку она пересекает прямую $y=1$ при $\sigma=1$, то она пересекает ее еще ровно в одной точке. Стало быть, характеристическое уравнение (8.6) имеет ровно два положительных корня, 1 и σ_1. Отсюда, как мы уже знаем, следует, что $A+B\sigma_1^z$ является формальным решением уравнения (8.5), и мы можем применить наши предыдущие рассуждения к этому решению вместо $A+Bz$. Мы находим, что в этом случае

$$\frac{\sigma_1^a-\sigma_1^z}{\sigma_1^a-\sigma_1^{-\nu+1}}\leqslant u_z\leqslant\frac{\sigma_1^{a+\mu-1}-\sigma_1^z}{\sigma_1^{a+\mu-1}-1}, \tag{8.12}$$

так что справедлива следующая теорема.

Теорема. *Решение нашей задачи о разорении удовлетворяет неравенствам* (8.11), *если* $\{p_k\}$ *имеет нулевое среднее, и неравенствам* (8.12) *в противном случае. Здесь* σ_1 *есть единственный отличный от нуля положительный корень* (8.6), *а* μ *и* $-\nu$ *равны соответственно наибольшему и наименьшему из индексов* k, *при которых* $p_k\neq 0$.

Пусть $m=\sum kp_k$ — *математическое ожидание выигрыша* в единичном испытании (или математическое ожидание длины одного шага). Из (8.6) легко видеть, что $\sigma_1>1$ при $m<0$ и $\sigma_1<1$ при $m>0$. Полагая $a\to\infty$, мы заключаем из нашей теоремы, что *в игре против бесконечно богатого соперника вероятность окончательного разорения равна единице тогда и только тогда, когда* $m\leqslant 0$.

Продолжительность игры можно исследовать аналогичными методами (ср. задачу 9).

§ 9. ЗАДАЧИ

Замечание. Задачи 1—4 относятся только к § 2 и не требуют никаких вычислений.

1. Найти вероятность того, что в случайном блуждании, начинающемся в нуле, точка $x=a>0$ будет достигаться раньше точки $x=-b<0$.

2. Доказать, что (в обозначениях § 2)

а) в случайном блуждании, начинающемся в нуле, вероятность достижения точки $x=a>0$ до возвращения в нуль равна $p(1-q_1)$;

б) в случайном блуждании, начинающемся в точке $x=a>0$, вероятность достижения нуля до возвращения в начальную точку равна qq_{a-1}.

3. Вывести из предыдущей задачи, что если $q\geqslant p$, то в начинающемся в нуле случайном блуждании число попаданий в точку $x=a>0$ до первого возвращения в нуль имеет геометрическое распределение со знаменателем $1-qq_{a-1}$. (Почему необходимо условие $q\geqslant p$?)

4. Используя результаты, полученные в двух предыдущих задачах, доказать следующую теорему[1]). *Математическое ожидание числа попаданий в точку* $x=a>0$ *до первого возвращения в нуль равно* $(p/q)^a$ *при* $p<q$ *и равно* 1 *при* $p=q$.

5. Рассмотреть задачу о разорении из § 2, 3 для случая видоизмененного

[1]) Поистине удивительные следствия этого результата проявляются лучше всего применительно к безобидным играм. Симметричная монета подбрасывается до тех пор, пока впервые не сравняются суммарные числа гербов и решеток. Игрок получает один цент каждый раз, когда суммарное число гербов будет превосходить суммарное число решеток ровно на m. *«Справедливая плата за вступление в игру» равна* 1 *центу независимо от* m.

Иное (элементарное) доказательство см. в задачах 1 и 2 гл. XII, 10 тома 2.

случайного блуждания, в котором частица либо совершает единичный шаг вправо или влево, либо остается на месте с вероятностями α, β и γ соответственно $(\alpha+\beta+\gamma=1)$. (В игровой терминологии это означает, что партия может окончиться вничью.)

6. Рассмотреть задачу о разорении из § 2, 3 для случая, когда в нуле помещен *упругий* экран (определение см. в § 1). Разностные уравнения для вероятности разорения (поглощения в нуле) и для средней продолжительности игры будут такими же, но уже с другими граничными условиями.

7. Частица при каждом шаге перемещается на две единицы направо или на одну единицу налево с вероятностями p и q соответственно $(p+q=1)$. Найти вероятность q_z того, что, выходя из точки $z>0$, частица когда-нибудь достигнет нуля. (Это задача о разорении при игре с бесконечно богатым противником.)

Указание. Аналог уравнения (2.1) приводит к кубическому уравнению, имеющему частное решение $q_z=1$ и два частных решения вида λ^z, где λ удовлетворяет квадратному уравнению.

8. *Продолжение*[1]. Показать, что q_1 равно вероятности того, что в последовательности испытаний Бернулли суммарное число неудач когда-нибудь будет вдвое больше суммарного числа успехов.

(При $p=q$ эта вероятность равна $(\sqrt{5}-1)/2$.)

9. В задаче § 8 для обобщенного случайного блуждания по аналогии с (8.1) положим $\rho_z=p_{a-z}+p_{a+1-z}+p_{a+2-z}+\ldots$ и обозначим через $d_{z,n}$ вероятность того, что игра продлится ровно n шагов. Показать, что при $n\geqslant 1$

$$d_{z,\,n+1}=\sum_{x=1}^{a-1} d_{x,\,n}p_{x-z},$$

где $d_{z,1}=r_z+\rho_z$. Исходя из этого, доказать, что производящая функция $d_z(\sigma)=\sum d_{z,\,n}\sigma^n$ является решением системы линейных уравнений

$$\sigma^{-1}d_z(\sigma)-\sum_{x=1}^{a-1} d_x(\sigma)\,p_{x-z}=r_z+\rho_z.$$

Дифференцируя, вывести отсюда, что средняя продолжительность e_z игры будет решением системы уравнений

$$e_z-\sum_{x=1}^{a-1} e_x p_{x-z}=1.$$

10. Пусть $w_{z,\,n}(x)$ — вероятность того, что в случайном блуждании с начальным положением z и *поглощающими* экранами в точках $x=0$ и $x=a$ n-й шаг приводит частицу в точку x. Найти разностные уравнения и граничные условия, которым удовлетворяет $w_{z,\,n}(x)$.

11. *Продолжение.* Видоизменить граничные условия на случай двух *отражающих* экранов (т. е. упругих экранов с $\delta=1$).

12. Пусть при симметричном случайном блуждании $(p=q)$ частица может находиться в точках $1, 2, \ldots, a-1$. В нуле расположен поглощающий экран, а на другом конце интервала — отражающий. Найти производящую функцию для времени ожидания поглощения.

---

[1]) Эта задача была сформулирована Д. Дж. Ньюменом. То, что ее решение является простым следствием предыдущей задачи, заметил У. Во. Читатель может попробовать воспользоваться аналогичным подходом для более общих задач, в которых множитель 2 заменяется каким-то другим рациональным числом. Решение, основанное на иных идеях, предложил Дж. С. Фрэйм; см. Frame J. S., Solution to problem 4864, Amer. Math. Monthly, 67 (1960), 700—702.

13. *Другой вариант записи вероятностей первого достижения.* В явной формуле (5.7) для вероятностей разорения устремим a к ∞. Показать, что в результате получится

$$u_{z,\,n}=2^n p^{(n-z)/2} q^{(n+z)/2} \int\limits_0^1 \cos^{n-1}\pi x\cdot\sin\pi x\cdot\sin\pi x z\cdot dx.$$

Следовательно, эта формула должна быть эквивалентна (4.14). Проверить это, показав, что удовлетворяются соответствующие разностные уравнения и граничные условия.

14. *Продолжение. Моменты первого достижения в диффузии.* Показать, что при описанном в § 6 предельном переходе из последней формулы получается выражение

$$(z/\sqrt{2\pi D t^3})\, e^{-(z+ct)^2/(2Dt)}$$

для плотности вероятности для времени ожидания поглощения в нуле в диффузии, начинающейся из точки $z>0$. При $p=q$ этот результат эквивалентен предельной теореме 3 из гл. III, 7.

Замечание. В следующих ниже задачах $v_{x,\,n}$ означает вероятность (6.1) того, что в начинающемся в нуле неограниченном случайном блуждании частица на n-м шаге попадает в положение x. Принцип отражения (гл. III, 1) приводит к иной интерпретации.

15. *Метод отражения*[1]). Пусть $p=q=1/2$. Обозначим через $u_{z,\,n}(x)$ вероятность того, что в случайном блуждании на $(0, \infty)$ с поглощающим экраном в нуле и начальным положением $z>0$ n-й шаг приведет частицу в положение $x>0$. Показать, что $u_{z,\,n}(x)=v_{x-z,\,n}-v_{x+z,\,n}$. (*Указание.* Показать, что разностное уравнение, соответствующее (4.1), и подходящие граничные условия в этом случае удовлетворяются.)

16. *Продолжение.* Доказать, что если в точке $x=0$ помещен *отражающий* экран, то

$$u_{z,\,n}(x)=v_{x-z,\,n}+v_{x+z-1,\,n}.$$

17. *Продолжение.* Доказать, что если случайное блуждание ограничено интервалом $(0, a)$ и оба экрана являются *поглощающими*, то

$$u_{z,\,n}(x)=\sum_k \{v_{x-z-2ka,\,n}-v_{x+z-2ka,\,n}\}, \tag{9.1}$$

где суммирование производится по всем k, положительным и отрицательным (лишь конечное число слагаемых отлично от нуля). Доказать, что если оба экрана являются *отражающими*, то уравнение (9.1) остается в силе при замене *минуса* на *плюс* и $x+z$ на $x+z-1$.

18. *Распределение максимумов.* Пусть $\mathbf{M}_n$ — максимальная абсцисса частицы за первые n шагов симметричного неограниченного случайного блуждания,

---

[1]) Задачи 15—17 являются примерами применения *метода отражения*. Член $v_{x-z,\,n}$ соответствует частице в неограниченном случайном блуждании, а $v_{x+z,\,n}$ — ее «отражению». В (9.1) используются отраженные точки, выходящие из различных положений, полученных повторными отражениями от обеих границ. В задачах 20 и 21 при помощи производящих функций мы получаем общий результат для несимметричного случайного блуждания. В теории дифференциальных уравнений метод отражений всегда связывают с именем Кельвина. В вероятностной литературе эквивалентный принцип приписывается Д. Андрэ. (См. примечание к лемме в гл. III, 1).

начинающегося в нуле. Используя результат задачи 15, показать, что

$$\mathbf{P}\{\mathbf{M}_n = z\} = v_{z,\,n} + v_{z+1,\,n}. \tag{9.2}$$

19. Пусть $V_x(s) = \sum v_{x,\,n} s^n$ (см. замечание перед задачей 15). Доказать, что $V_x(s) = V_0(s)\,\lambda_2^{-x}(s)$ при $x \leqslant 0$ и $V_x(s) = V_0(s)\,\lambda_1^{-x}(s)$ при $x \geqslant 0$, где $\lambda_1(s)$ и $\lambda_2(s)$ определены формулами (4.8). Более того, $V_0(s) = (1 - 4pqs^2)^{-1/2}$.

Замечание. Эти соотношения следуют *непосредственно* из того факта, что $\lambda_1(s)$ и $\lambda_2(s)$ являются производящими функциями времен первого достижения, как это было установлено в конце § 4.

20. Пусть $u_{z,\,n}(x)$ — вероятность того, что в случайном блуждании в $(0, \infty)$ с поглощающим экраном в нуле и начальным положением z n-й шаг приведет частицу в положение x, и пусть

$$U_z(s;\,x) = \sum_{n=0}^{\infty} u_{z,\,n}(x)\,s^n. \tag{9.3}$$

Используя результат задачи 19, показать, что $U_z(s;x) = V_{x-z}(s) - \lambda_2^z(s) V_x(s)$. Получить отсюда, что

$$u_{z,\,n}(x) = v_{x-z,\,n} - (q/p)^z \cdot v_{x+z,\,n}. \tag{9.4}$$

Сравнить с результатом задачи 15 и вывести (9.4) из последнего при помощи комбинаторных методов.

21. *Другая формула для вероятности разорения* (5.7). Разложив (4.11) в геометрический ряд, доказать, что

$$u_{z,\,n} = \sum_{k=0}^{\infty} (p/q)^{ka} w_{2ka+z,\,n} - \sum_{k=1}^{\infty} (p/q)^{ka-z} w_{2ka-z,\,n},$$

где $w_{z,\,n}$ — вероятность (4.14) первого достижения.

22. Показать, что если к выражению для $u_{z,\,n}$, приведенному в предыдущей задаче, применить предельный переход из § 6, то плотность вероятности для времени поглощения будет равна[1])

$$(1/\sqrt{2\pi D t^3})\, e^{-(ct+2\xi)c/(2D)} \sum_{k=-\infty}^{\infty} (\xi + 2k\alpha)\, e^{-(\xi+2k\alpha)^2/(2Dt)}.$$

(*Указание.* Использовать нормальное приближение для биномиального распределения.)

23. *Метод восстановления для задачи о разорении*[2]). Пусть в случайном блуждании с двумя поглощающими экранами $u_{z,\,n}$ и $u^*_{z,\,n}$ — вероятности поглощения на левом и на правом экранах соответственно. Посредством надлежащей интерпретации доказать справедливость двух следующих уравнений:

$$V_{-z}(s) = U_z(s)\,V_0(s) + U_z^*(s)\,V_{-a}(s),$$
$$V_{a-z}(s) = U_z(s)\,V_a(s) + U_z^*(s)\,V_0(s).$$

Вывести (4.11), разрешив эту систему относительно $U_z(s)$.

24. Пусть $u_{z,\,n}(x)$ — вероятность того, что выходящая из точки z частица на n-м шаге попадет в точку x, не коснувшись предварительно поглощающих экранов. Показать, что для соответствующей производящей функции $U_z(s;\,x)=$

---

[1]) Совпадение этой новой формулы с предельной формой (6.15) — известный факт теории тэта-функций. См. формулу (5.8) гл. XIX тома 2.

[2]) Задачи 23—25 содержат новый и независимый вывод основных результатов, относящихся к одномерным случайным блужданиям.

$=\sum u_{z,\,n}(x)\,s^n$ мы имеем (используя обозначения задачи 23) равенство

$$U_z(s;x)=V_{x-z}(s)-U_z(s)\,V_x(s)-U_z^*(s)\,V_{x-a}(s).$$

(Здесь не требуется никаких вычислений.)

25. *Продолжение.* Доказать, что производящую функцию $U_z(s;x)$ из предыдущей задачи можно получить в виде $U_z(s;x)=V_{x-z}(s)-A\lambda_1^z(s)-B\lambda_2^z(s)$, где постоянные определяются так, чтобы при $z=0$ и $z=a$ выполнялись граничные условия $U_z(s;x)=0$. (В случае *отражающих экранов* граничные условия имеют вид $U_0(s;x)=U_1(s;x)$ и $U_a(s;x)=U_{a-1}(s;x)$.)

26. Доказать формулу

$$v_{x,\,n}=(2\pi)^{-1}2^n p^{(n+x)/2}q^{(n-x)/2}\int_{-\pi}^{\pi}\cos^n t\cdot\cos tx\cdot dt,$$

показав, что удовлетворяется соответствующее разностное уравнение. Вывести отсюда, что

$$V_x(s)=(2\pi)^{-1}(p/q)^{x/2}\int_{-\pi}^{\pi}\frac{\cos tx}{1-2\sqrt{pq}\cdot s\cdot\cos t}\,dt.$$

27. Показать, что в трехмерном симметричном случайном блуждании частица с вероятностью единица бесконечное число раз пересечет любую фиксированную прямую $x=m$, $y=n$. (*Указание.* Ср. с задачей 5.)

28. В двумерном симметричном случайном блуждании, начинающемся в нуле, вероятность того, что n-й шаг приведет частицу в точку (x,y), равна

$$(2\pi)^{-2}2^{-n}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}(\cos\alpha+\cos\beta)^n\cdot\cos x\alpha\cdot\cos y\beta\cdot d\alpha\,d\beta.$$

Проверить эту формулу и найти ее аналог для трехмерного случая. (*Указание.* Проверить, что это выражение удовлетворяет соответствующему разностному уравнению.)

29. В двумерном симметричном случайном блуждании обозначим через $\mathbf{D}_n^2=x^2+y^2$ квадрат расстояния частицы от начала координат в момент времени n. Доказать, что $\mathbf{E}(\mathbf{D}_n^2)=n$. (*Указание.* Вычислить $\mathbf{E}(\mathbf{D}_{n-1}^2-\mathbf{D}_n^2)$.)

30. Доказать, что в симметричном случайном блуждании в d-мерном пространстве частица с вероятностью 1 будет бесконечное число раз возвращаться в положения, которые уже были ею заняты ранее. (*Указание.* При каждом шаге вероятность попадания в новое положение не превышает $(2d-1)/(2d)$.)

31. Показать, что описанный в § 8 метод применим также к производящей функции $U_z(s)$ для времени ожидания разорения.

§ 1. ОПРЕДЕЛЕНИЕ

До сих пор мы занимались главным образом независимыми испытаниями, которые можно описать следующим образом. Задано множество возможных исходов $E_1, E_2, \ldots$ (в конечном или бесконечном числе), и каждому из них соотнесена некоторая вероятность p_k; вероятности последовательностей исходов определяются по правилу умножения: $\mathbf{P}\{(E_{j_0}, E_{j_1}, \ldots, E_{j_n})\} = p_{j_0}p_{j_1}\ldots p_{j_n}$. В теории цепей Маркова мы рассматриваем простейшее обобщение этой схемы, которое состоит в том, что для любого испытания допускается зависимость его от непосредственно предшествующего ему испытания (и только от него). С исходом E_k не связана более фиксированная вероятность p_k, но зато каждой паре (E_j, E_k) теперь соответствует *условная вероятность* p_{jk}; при условии, что E_j появился в некотором испытании, вероятность появления E_k в следующем испытании равна p_{jk}. Помимо p_{jk} нам должны быть заданы вероятности a_k исходов E_k в *начальном* испытании. Чтобы p_{jk} имели приписанный им смысл, вероятности последовательностей исходов, соответствующих двум, трем или четырем испытаниям, должны быть определены равенствами

$$\mathbf{P}\{(E_j, E_k)\} = a_j p_{jk}, \quad \mathbf{P}\{(E_j, E_k, E_r)\} = a_j p_{jk} p_{kr},$$
$$\mathbf{P}\{(E_j, E_k, E_r, E_s)\} = a_j p_{jk} p_{kr} p_{rs},$$

и вообще

$$\mathbf{P}\{(E_{j_0}, E_{j_1}, \ldots, E_{j_n})\} = a_{j_0} p_{j_0 j_1} p_{j_1 j_2} \ldots p_{j_{n-2} j_{n-1}} p_{j_{n-1} j_n}. \quad (1.1)$$

Здесь начальному испытанию присвоен номер нуль, так что испытание номер один является вторым. (Это соглашение удобно и — без специального на то указания — уже использовалось в предыдущей главе.)

Некоторые процессы, рассматривавшиеся нами в предыдущих главах, являются цепями Маркова, однако в конкретных случаях часто бывает лучше воспользоваться другими обозначениями и способами описания. Основные результаты настоящей главы касаются существования некоторых пределов и равновесных распределений; эти результаты, конечно же, не зависят от обозначений и применимы ко всем цепям Маркова.

Примеры. а) *Случайные блуждания.* Случайное блуждание на прямой является цепью Маркова, однако в этом случае возможные положения естественно представить в виде бесконечной в обе стороны последовательности ..., —2, —1, 0, 1, 2, ... При таком порядке переходы будут возможны только между соседними положениями, т. е. $p_{jk}=0$, если $k\neq j\pm 1$. Чтобы воспользоваться нашими нынешними обозначениями, нам пришлось бы расположить целые числа в простую последовательность, скажем 0, 1, —1, 2, —2, . . . , и это привело бы к громоздким формулам для вероятностей p_{jk}. То же замечание справедливо и в отношении случайных блужданий в пространствах высшей размерности: для практических вычислений лучше обозначать точки значениями их координат, а для теоретических целей можно пользоваться символикой настоящей главы.

б) *Ветвящиеся процессы.* Вместо того чтобы говорить, что n-е испытание окончилось исходом E_k, в гл. XII, 3 мы говорили, что в n-м поколении имеется k частиц. Мы рассматривали там обычную цепь Маркова, у которой переходная вероятность p_{jk} равна коэффициенту при s^k в разложении j-й степени $p^j(s)$ заданной производящей функции.

в) *Урновые модели.* Очевидно, что некоторые урновые модели из гл. V, 2 представляют собой цепи Маркова. И обратно, как мы сейчас увидим, каждая цепь Маркова эквивалентна некоторой урновой модели.

Пусть каждый из имеющихся у состояний цепи индексов представлен отдельной урной и каждая урна содержит шары с метками E_1, E_2, ... Содержимое урн остается фиксированным, но меняется от урны к урне; вероятность извлечь из j-й урны шар с меткой E_k равна p_{jk}. В *начальном*, или нулевом, испытании в соответствии с распределением вероятностей $\{a_i\}$ выбирается урна. Из этой урны случайным образом извлекается шар, и если он помечен E_j, то следующее извлечение производится из j-й урны и т. д. Очевидно, что при такой процедуре вероятность выборки $(E_{j_0}, \ldots, E_{j_n})$ дается формулой (1.1). Мы видим, что понятие цепи Маркова не является более общим, чем понятие урновой модели, однако новая символика окажется более практичной и более интуитивно ясной. ►

Если a_k — вероятность появления E_k в начальном (или нулевом) испытании, то мы должны иметь $a_k\geqslant 0$ и $\sum a_k=1$. Более того, после появления E_j непременно должно произойти одно из E_k, и поэтому необходимо, чтобы для всех j и k

$$p_{j1}+p_{j2}+\ldots=1, \qquad p_{jk}\geqslant 0. \tag{1.2}$$

Теперь мы покажем, что для любых чисел a_k и p_{jk}, удовлетворяющих этим условиям, формула (1.1) является допустимым определением вероятностей в пространстве элементарных событий, соответствующем $n+1$ испытаниям. Поскольку числа, определяемые по (1.1), неотрицательны, нам нужно доказать только то, что в сумме они составляют единицу. Фиксируем первые $j_0, j_1, \ldots, j_{n-1}$ и сложим числа (1.1) при всех возможных j_n. Используя (1.2) при $j=j_{n-1}$, мы немедленно убеждаемся, что эта сумма равна $a_{j_0}p_{j_0j_1}\ldots p_{j_{n-2}j_{n-1}}$. Таким образом, сумма всех чисел (1.1) не зависит от n, и, поскольку $\sum a_{j_0}=1$, сумма эта равна единице для всех n.

Определение (1.1) формально зависит от числа испытаний, однако наше рассуждение доказывает взаимную согласованность определений (1.1) при всех n. Например, чтобы получить вероятность события «первые два испытания окончились исходами (E_j, E_k)», мы должны фиксировать $j_0=j$ и $j_1=k$ и сложить вероятности (1.1) при всех возможных $j_2, j_3, \ldots, j_n$. Только что мы показали, что эта сумма равна a_jp_{jk} и, таким образом, не зависит от n. Это означает, что обычно нет необходимости говорить о полном числе испытаний; событие $(E_{j_0}, \ldots, E_{j_r})$ имеет одну и ту же вероятность во всех пространствах элементарных событий для более чем r испытаний. В связи с независимыми испытаниями неоднократно отмечалось, что с математической точки зрения наиболее удовлетворительным будет ввести только одно пространство элементарных событий — неограниченных последовательностей испытаний — и рассматривать результат конечного числа испытаний как начальный отрезок бесконечной последовательности. Это утверждение справедливо также и для цепей Маркова. К сожалению, пространства бесконечных последовательностей испытаний выводят изложение за пределы теории дискретных вероятностей, которой мы ограничились в этом томе.

Суммируя сказанное выше, получаем в качестве отправного пункта следующее определение.

Определение. *Последовательность испытаний с возможными исходами $E_1, E_2, \ldots$ называется цепью Маркова* [1], *если вероятности последовательностей исходов определяются формулой* (1.1) *через распределение вероятностей $\{a_k\}$ для E_k в начальном (или нулевом) испытании и через фиксированные условные вероятности p_{jk} появления E_k при условии, что в предыдущем испытании появился E_j.*

Для приложений цепей Маркова удобнее несколько видоизмененная терминология. Возможные исходы E_k обычно называются возможными *состояниями системы*; вместо того чтобы говорить, что n-е испытание окончилось появлением E_k, говорят, что n-й *шаг приводит к состоянию* E_k или что система попадает в E_k на n-м шаге. Наконец, p_{jk} называется *вероятностью перехода из* E_j *в* E_k. Как обычно, мы считаем, что испытания происходят через равные ин-

---

[1]) Это нестандартная терминология. Здесь мы рассматриваем только частный класс цепей Маркова, и, строго говоря, здесь и в последующих параграфах термин «цепь Маркова» должен всякий раз уточняться добавлением слов «со стационарными переходными вероятностями». На самом деле общий тип цепей Маркова рассматривается редко. Он будет определен в § 13, где марковское свойство будет обсуждаться в связи с общими стохастическими процессами. Там же читатель найдет примеры зависимых испытаний, не образующих цепи Маркова.

тервалы времени, так что номер шага служит временным параметром.

Вероятности перехода p_{jk} будут расположены в *матрицу переходных вероятностей*

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{bmatrix}, \tag{1.3}$$

где первый индекс означает номер строки, а второй — номер столбца. Ясно, что P — квадратная матрица с неотрицательными элементами и единичными суммами по строкам. Такая матрица (конечная или бесконечная) называется *стохастической матрицей. Любая стохастическая матрица может служить матрицей переходных вероятностей; вместе с нашим начальным распределением* $\{a_k\}$ *она полностью определяет цепь Маркова с состояниями* E_1, $E_2, \ldots$.

В некоторых частных случаях бывает удобно нумеровать состояния, начиная с 0, а не с 1. Тогда к матрице P следует добавить нулевые строку и столбец.

Историческое замечание. Многие задачи, рассматривавшиеся в классической литературе при помощи урновых моделей, представляются теперь при помощи цепей Маркова частного вида, однако первоначальные методы исследования были совершенно другими. Более того, ряд урновых моделей носят совсем иной характер, ибо в них имеет место эффект последействия, и эта существенная разница в должной мере не была понята. В действительности эта путаница сохранялась долгое время после первооткрывающей работы Маркова. А. А. Марков заложил основания теории конечных цепей Маркова, однако конкретные приложения ограничивались главным образом гасованием карт и лингвистическими задачами. Теоретическое изучение проводилось обычно алгебраическими методами, близкими к описанным в следующей главе. Основы этого подхода изложены в монографии М. Фреше [1]).

Теория цепей с бесконечным числом состояний была введена А. Н. Колмогоровым [2]). Этот новый подход, изложенный в первом издании настоящей книги, сделал эту теорию доступной более широкому кругу специалистов и привлек внимание к множеству ее возможных приложений. С тех пор цепи Маркова стали стандартным элементом теории вероятностей и знакомым инструментом во многих приложениях. Более современные теоретические разработки упоминаются в замечаниях в § 11 и 12.

---

[1]) Fréchet M., Recherches théoriques modernes sur le calcul des probabilités, t. 2 (Théorie des événements en chaine dans le cas d'un nombre fini d'états possibles), Paris, 1938.

[2]) Kolmogoroff A., Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen, Математический сборник, н. с., 1 (1936), 607—610. Эта статья не содержит доказательств. Полное изложение см. в работе Колмогоров А. Н., Цепи Маркова со счетным числом состояний, Бюлл. МГУ (А), 1 : 3 (1937), 1—16.

§ 2. ПОЯСНИТЕЛЬНЫЕ ПРИМЕРЫ

(Приложения к классической задаче о тасовании карт см. в § 10.)

а) Когда у цепи есть только два возможных состояния E_1 и E_2, матрица переходных вероятностей с необходимостью имеет вид

$$P = \begin{bmatrix} 1-p & p \\ \alpha & 1-\alpha \end{bmatrix}.$$

Подобная цепь могла бы быть реализована в следующем мысленном эксперименте. Частица движется вдоль оси x таким образом, что абсолютная величина ее скорости остается постоянной, но направление движения может меняться на противоположное. Говорят, что система находится в состоянии E_1, если частица движется направо, и в состоянии E_2, если она движется налево. Тогда p — вероятность поворота, когда частица движется направо, а α — вероятность поворота при движении налево. (Подробный анализ этой цепи см. в примере гл. XVI, 2,а).)

б) *Случайное блуждание с поглощающими экранами.* Пусть возможными состояниями будут $E_0, E_1, \ldots, E_\rho$; рассмотрим матрицу переходных вероятностей

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ q & 0 & p & 0 & \ldots & 0 & 0 & 0 \\ 0 & q & 0 & p & \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot & \cdot & \\ 0 & 0 & 0 & 0 & \ldots & q & 0 & p \\ 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 1 \end{bmatrix}.$$

Из каждого «внутреннего» состояния $E_1, \ldots, E_{\rho-1}$ возможны переходы в правое и левое соседние состояния (с вероятностями $p_{i,\,i+1}=p$ и $p_{i,\,i-1}=q$). Однако ни из E_0, ни из E_ρ невозможны переходы в какое-либо иное состояние; система может переходить из одного состояния в другое, но коль скоро будет достигнуто E_0 или E_ρ, система останется неизменной навсегда. Ясно, что эта цепь Маркова только терминологией отличается от модели случайного блуждания с поглощающими экранами в точках 0 и ρ, обсуждавшейся в предыдущей главе. Там случайное блуждание начиналось в фиксированной точке z рассматриваемого интервала. В терминологии цепей Маркова это равносильно выбору такого начального распределения, что $a_z=1$ (и, следовательно, $a_x=0$ при $x\neq z$). Случайно выбранному начальному состоянию соответствует начальное распределение $a_k=1/(\rho+1)$.

в) *Отражающие экраны.* Интересный вариант предыдущего примера представляет собой цепь с возможными состояниями

$E_1, \ldots, E_\rho$ и переходными вероятностями

$$P=\begin{bmatrix} q & p & 0 & 0 & \ldots & 0 & 0 & 0 \\ q & 0 & p & 0 & \ldots & 0 & 0 & 0 \\ 0 & q & 0 & p & \ldots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \ldots & q & 0 & p \\ 0 & 0 & 0 & 0 & \ldots & 0 & q & p \end{bmatrix}.$$

Эту цепь можно интерпретировать на языке азартных игр, рассматривая двух игроков, ведущих игру с единичными ставками и с соглашением, что каждый раз, когда один из игроков проигрывает свой последний доллар, тот немедленно возвращается ему его противником, так что игра может продолжаться бесконечно. Мы предполагаем, что игроки вместе имеют $\rho+1$ долларов, и говорим, что система находится в состоянии E_k, если их капиталы равны k и $\rho-k+1$ соответственно. Тогда переходные вероятности даются нашей матрицей P. В терминологии, введенной в гл. XIV, 1, наша цепь представляет собой случайное блуждание с отражающими экранами в точках $1/2$ и $\rho+1/2$. Аналогичным образом можно рассматривать случайные блуждания с упругими экранами. Подробный анализ цепи для случая отражающих экранов читатель найдет в гл. XVI, 3 (см. также пример 7, в)).

г) *Циклические случайные блуждания.* Пусть возможными состояниями снова будут $E_1, E_2, \ldots, E_\rho$, но теперь мы упорядочим их циклическим образом, так что для E_ρ соседними состояниями будут $E_{\rho-1}$ и E_1. Если, как и ранее, система всегда переходит либо в правое, либо в левое соседнее состояние, то строки матрицы P будут такими же, как в примере б), за тем исключением, что первая строка будет $(0, p, 0, 0, \ldots, 0, q)$, а последняя $(p, 0, 0, 0, \ldots, 0, q, 0)$.

В более общем случае мы можем допустить переходы между любыми двумя состояниями. Пусть $q_0, q_1, \ldots, q_{\rho-1}$ — соответственно вероятности остаться на месте или передвинуться на $1, 2, \ldots, \rho-1$ единиц вправо (причем переход на k единиц вправо — то же самое, что переход на $\rho-k$ единиц влево). Тогда P будет циклической матрицей

$$P=\begin{bmatrix} q_0 & q_1 & q_2 & \cdots & q_{\rho-2} & q_{\rho-1} \\ q_{\rho-1} & q_0 & q_1 & \cdots & q_{\rho-3} & q_{\rho-2} \\ q_{\rho-2} & q_{\rho-1} & q_0 & \cdots & q_{\rho-4} & q_{\rho-3} \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ q_1 & q_2 & q_3 & \cdots & q_{\rho-1} & q_0 \end{bmatrix}.$$

Анализ этой цепи см. в примере гл. XVI, 2, г).

д) *Модель Эренфестов для диффузии.* Снова рассмотрим цепь с $\rho+1$ состоянием $E_0, E_1, \ldots, E_\rho$, переходы в которой возможны только в правое и левое соседние состояния; на этот раз мы положим $p_{j,\,j+1}=1-j/\rho$ и $p_{j,\,j-1}=j/\rho$, так что

$$P=\begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ \rho^{-1} & 0 & 1-\rho^{-1} & 0 & \ldots & 0 & 0 \\ 0 & 2\rho^{-1} & 0 & 1-2\rho^{-1} & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \ldots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \ldots & 0 & \rho^{-1} \\ 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}.$$

Эта цепь имеет две интересные физические интерпретации. При обсуждении различных задач о возвращении в статистической механике П. и Т. Эренфесты [1]) описали мысленный урновый эксперимент, в котором ρ молекул распределены по двум сосудам A и B. При каждом испытании случайным образом выбирается одна молекула и перемещается из своего сосуда в другой. Состояние системы определяется числом молекул в A. Предположим, что в некоторый момент времени в сосуде A находится ровно k молекул. При следующем испытании система переходит в E_{k-1} или E_{k+1} в зависимости от того, выбрана ли молекула из A или из B; соответствующие вероятности суть k/ρ и $(\rho-k)/\rho$, и, стало быть, наша цепь действительно описывает эксперимент Эренфестов. Однако нашу цепь можно интерпретировать и как *диффузию при наличии центральной силы*, т. е. как случайное блуждание, в котором вероятность шага вправо меняется вместе с состоянием. Из $x=j$ переход частицы вправо (влево) будет более вероятен, если $j<\rho/2$ ($j>\rho/2$); это означает, что частица имеет тенденцию двигаться к точке $x=\rho/2$, что соответствует упругой силе притяжения, возрастающей прямо пропорционально расстоянию. (Модель Эренфестов была описана в примере гл. V, 2,в); см. также пример 7, г) и задачу 12.)

е) *Модель Бернулли — Лапласа для диффузии* [2]). Модель, похожая на модель Эренфестов, была предложена Д. Бернулли в качестве вероятностного аналога течения двух несжимаемых жид-

[1]) Ehrenfest P., Ehrenfest T., Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem, Physikalische Zeitschrift, 8 (1907), 311—314; Ming Chen Wang, Uhlenbeck G. E., On the theory of the Brownian motion II, Reviews of Modern Physics, 17 (1945), 323—342. Более полное обсуждение см. Kac M., Random walk and the theory of Brownian motion, Amer. Math. Monthly, 54 (1947), 369—391. Эти авторы не упоминали цепи Маркова, однако Кац использовал методы, близко связанные с описываемыми в следующей главе. См. также Friedman B., A simple urn model, Communications on Pure and Applied Mathematics, 2 (1949), 59—70.

[2]) В виде урновой модели эту задачу рассматривал Даниил Бернулли в 1769 г., критиковал Мальфатти в 1782 г. и анализировал Лаплас в 1812 г. См. Todhunter I., A history of the mathematical theory of probability, Cambridge, 1865.

костей между двумя сосудами. На этот раз мы имеем всего 2ρ частиц, среди которых ρ черных и ρ белых. Поскольку предполагается, что частицы представляют несжимаемые жидкости, их плотности не должны меняться, так что число ρ частиц в каждой урне остается постоянным. Мы говорим, что система находится в состоянии E_k ($k=0, 1, \ldots, \rho$), если первая урна содержит k белых частиц. (Это означает, что она содержит $\rho-k$ черных частиц, тогда как вторая урна содержит $\rho-k$ белых и k черных частиц.) При каждом испытании из каждой урны выбирается по одной частице, и эти две частицы меняются местами. Переходные вероятности даются тогда формулами

$$p_{j,\,j-1}=\left(\frac{j}{\rho}\right)^2,\quad p_{j,\,j+1}=\left(\frac{\rho-j}{\rho}\right)^2,\quad p_{jj}=2\,\frac{j\,(\rho-j)}{\rho^2} \tag{2.1}$$

и $p_{jk}=0$, если $|j-k|>1$ (здесь $j=0, \ldots, \rho$). (О стационарном распределении см. в примере 7, д); обобщение этой модели см. в задаче 10.)

ж) *Случайные размещения шаров.* Рассмотрим последовательность независимых испытаний, в каждом из которых шар случайным образом помещается в один из заданных ρ ящиков (или урн). Мы скажем, что система находится в состоянии E_k, если заняты ровно k ящиков. Это определяет цепь Маркова с набором состояний $E_0, \ldots, E_\rho$ и переходными вероятностями, такими, что

$$p_{jj}=j/\rho,\qquad p_{j,\,j+1}=(\rho-j)/\rho, \tag{2.2}$$

и, конечно, $p_{jk}=0$ для всех других комбинаций j и k. Если сначала все ящики были пустыми, то распределение $\{a_k\}$ определяется равенствами $a_0=1$ и $a_k=0$ при $k>0$. (Дальнейший анализ этой цепи см. в примере гл. XVI, 2, д). Случайные размещения шаров рассматривались нами с других точек зрения в гл. II, 5 и гл. IV, 2.)

з) *Пример из клеточной генетики* [1]. Цепь Маркова с набором состояний $E_0, \ldots, E_N$ и переходными вероятностями

$$p_{jk}=\binom{2j}{k}\binom{2N-2j}{N-k}\binom{2N}{N}^{-1} \tag{2.3}$$

появляется в биологической задаче, которую упрощенно можно сформулировать следующим образом. Каждая клетка некоторого организма содержит N частиц, причем одни из них относятся к типу A, а другие — к типу B. Говорят, что клетка находится в состоянии E_j, если она содержит ровно j частиц типа A. Дочерние клетки образуются в результате клеточного деления, но перед делением каждая из частиц реплицируется (т. е. удваивается); дочерняя клетка

[1] Schensted I. V., Model of subnuclear segregation in the macronucleus of ciliates, The Amer. Naturalist, 92 (1958), 161—170. Этот автор существенно использует методы гл. XVI, но не упоминает цепи Маркова. Наша формулировка задачи математически эквивалентна рассматривавшейся в этой работе, но биологически сверхупрощена.

наследует N частиц, выбранных случайным образом из $2j$ частиц типа A и $2N-2j$ частиц типа B, имевшихся в родительской клетке. Вероятность того, что дочерняя клетка будет находится в состоянии E_k, дается тогда гипергеометрическим распределением (2.3).

В примере 8, б) будет показано, что *после достаточно большого числа поколений вся популяция будет состоять (и останется состоящей) из клеток, находящихся в одном из чистых состояний E_0 или E_N*; вероятности того, что отдельная линия потомков придет к тому или иному чистому состоянию, суть соответственно $1-j/N$ и j/N (E_j есть начальное состояние).

и) *Примеры из популяционной генетики* [1]. Рассмотрим последовательные поколения в популяции (такой, как растения на кукурузном поле), величина которой сохраняется постоянной в результате отбора N особей в каждом поколении. Отдельный ген, который может относиться к типу A или a, имеет в популяции $2N$ представителей; если в n-м поколении A встречается j раз, то a встречается $2N-j$ раз. В этом случае мы говорим, что популяция находится в состоянии E_j $(0 \leqslant j \leqslant 2N)$. Если предположить, что скрещивание происходит случайным образом, то состав следующего поколения определяется $2N$ испытаниями Бернулли, в которых вероятность появления гена A будет равна $j/(2N)$. Стало быть, мы имеем цепь Маркова с

$$p_{jk} = \binom{2N}{k} \left(\frac{j}{2N}\right)^k \left(1-\frac{j}{2N}\right)^{2N-k}. \tag{2.4}$$

В состояниях E_0 и E_{2N} все гены относятся к одному типу, и уход из этих состояний невозможен. (Такие состояния называются гомозиготными.) В примере 8, б) будет показано, что *в конце концов вся популяция окажется в одном из гомозиготных состояний E_0 или E_{2N}*. Если популяция находится вначале в состоянии E_j, то соответствующие вероятности суть $1-j/(2N)$ и $j/(2N)$.

Эту модель можно модифицировать с тем, чтобы учесть возможные мутации и селективные преимущества генов.

к) *Задача о скрещивании*. В так называемом братско-сестринском скрещивании скрещиваются две особи, и среди их прямых потомков случайным образом выбираются две особи разного пола. Они вновь скрещиваются, и процесс этот продолжается бесконечно. Имея три генотипа AA, Aa, aa для каждого родителя, мы должны различать шесть комбинаций родителей, которые мы пометим следующим образом: $E_1 = AA \times AA$, $E_2 = AA \times Aa$, $E_3 = Aa \times Aa$, $E_4 = Aa \times aa$, $E_5 = aa \times aa$, $E_6 = AA \times aa$. Используя правила из гл. V. 5, легко видеть, что матрица переходных вероятностей имеет в этом слу-

[1] Эту задачу исследовали различными методами Р. Э. Фишер и С. Райт. Формулировка в терминах цепей Маркова дается в статье Malécot G., Sur un problème de probabilités en chaine que pose la génétique, Comptes rendus de l'Académie des Sciences, 219 (1944), 379—381.

чае вид

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 & 0 & 0 \\ 1/16 & 1/4 & 1/4 & 1/4 & 1/16 & 1/8 \\ 0 & 0 & 1/4 & 1/2 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

(Обсуждение этой цепи продолжается в задаче 4; подробный ее анализ дан в примере гл. XVI, 4, б).)

л) *Рекуррентные события и остаточные времена ожидания.* Мы неоднократно будем использовать цепь с набором состояний E_0, $E_1, \ldots$ и переходными вероятностями

$$P = \begin{bmatrix} f_1 & f_2 & f_3 & f_4 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdots \end{bmatrix};$$

вероятности f_k произвольны, за тем лишь исключением, что в сумме они должны давать единицу. Чтобы наглядно представить себе этот процесс, предположим, что начальным состоянием будет E_0. Если первый шаг приводит в E_{k-1}, то система обязана пройти последовательно через состояния E_{k-2}, $E_{k-3}, \ldots$ и на k-м шаге вернуться в E_0, откуда процесс начнется сначала. Таким образом, последовательные возвращения в E_0 представляют из себя возвратное рекуррентное событие $\mathscr{E}$ с распределением $\{f_k\}$ для времен возвращения. Состояние системы в любой момент времени определяется временем ожидания *следующего* прохождения через E_0.

В большинстве конкретных реализаций рекуррентных событий время ожидания следующего осуществления события зависит от будущего, и, значит, наша цепь Маркова не имеет никакого практического значения. Однако эта цепь имеет смысл в том случае, когда можно представить себе, что одновременно с каждым осуществлением события $\mathscr{E}$ производится случайный эксперимент, исход которого определяет величину следующего времени ожидания. Такие ситуации встречаются на практике, хотя они и составляют скорее исключение, чем правило. Например, в теории самовосстанавливающихся устройств (пример гл. XIII, 10, г)) иногда предполагается, что срок службы вновь установленного элемента зависит от выбора этого элемента, но вполне определен, коль скоро выбор уже сделан. С другой стороны, в теории массового обслуживания (в очередях к продавцу или на телефонных линиях) последовательные моменты начала обслуживания отдельных клиентов обычно соответствуют рекуррентным событиям. Предположим теперь, что имеется много типов клиентов, и для каждого из этих типов требуется обслуживание известной продолжительности. Тогда время ожидания между двумя последовательными моментами начала обслуживания определяется единственным образом с того момента, когда начинается обслуживание соответствующего клиента. (См. пример 7, ж).)

м) *Другая цепь, связанная с рекуррентными событиями.* Рассмотрим цепь с набором возможных состояний E_0, E_1, . . . и переходными вероятностями

$$P=\begin{bmatrix} q_1 & p_1 & 0 & 0 & 0 & \dots \\ q_2 & 0 & p_2 & 0 & 0 & \dots \\ q_3 & 0 & 0 & p_3 & 0 & \dots \\ q_4 & 0 & 0 & 0 & p_4 & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots \end{bmatrix},$$

где $p_k+q_k=1$. Для наглядности мы можем интерпретировать состояние E_k как представляющее «возраст» системы. По достижении системой возраста k процесс старения с вероятностью p_{k+1} продолжается, а с вероятностью q_{k+1} система «омолаживается», и процесс начинается заново — с нулевого возраста. Последовательные прохождения через состояние E_0 здесь снова представляют рекуррентное событие, и вероятность того, что время возвращения равно k, дается произведением $p_1p_2 \ldots p_{k-1}q_k$. Можно подобрать $\{p_k\}$ так, чтобы получить заданное распределение $\{f_k\}$ для времен возвращения; достаточно положить $q_1=f_1$, затем $q_2=f_2/p_1$ и т. д. В общем виде

$$p_k=\frac{1-f_1-\ldots-f_k}{1-f_1-\ldots-f_{k-1}}. \tag{2.5}$$

Таким образом, произвольное рекуррентное событие $\mathcal{E}$ с распределением $\{f_k\}$ времен возвращения соответствует цепи Маркова с матрицей P, определяемой вероятностями (2.5). После n-го испытания система окажется в состоянии E_k тогда и только тогда, когда последним испытанием, при котором произошло событие $\mathcal{E}$, было испытание с номером $n-k$ (здесь $k=0, 1, \ldots$). Номер этого состояния часто называется «затраченным временем ожидания». (Обсуждение продолжается в примерах 5, б), 7, е) и 8, д).)

н) *Серии успехов.* В качестве частного случая предыдущего примера рассмотрим последовательность испытаний Бернулли и условимся, что при n-м испытании система будет в состоянии E_k, если последняя *неудача* наблюдалась при испытании с номером $n-k$. Здесь $k=0, 1, \ldots$, и считается, что нулевое испытание привело к неудаче. Иначе говоря, индекс k равен длине непрерывавшейся последовательности успехов, оканчивающейся n-м испытанием. Переходные вероятности здесь те же, что и в предыдущем примере с $p_k=p$ и $q_k=q$ при всех k.

§ 3. ВЕРОЯТНОСТИ ПЕРЕХОДА ЗА НЕСКОЛЬКО ШАГОВ

Мы обозначим через $p_{jk}^{(n)}$ *вероятность перехода из E_j в E_k ровно за n шагов*. Иначе говоря, $p_{jk}^{(n)}$ есть условная вероятность попадания в E_k на n-м шаге при условии, что начальным состоянием было E_j; она равна сумме вероятностей всех путей $E_jE_{j_1}\ldots$ $\ldots E_{j_{n-1}}E_k$ длины n, начинающихся в E_j и оканчивающихся в E_k. В частности, $p_{jk}^{(1)}=p_{jk}$ и

$$p_{jk}^{(2)}=\sum_{\nu} p_{j\nu}p_{\nu k}. \tag{3.1}$$

По индукции мы получаем *общую рекуррентную формулу*

$$p_{jk}^{(n+1)}=\sum_{\nu} p_{j\nu}p_{\nu k}^{(n)}; \tag{3.2}$$

дальнейшая индукция по m приводит к *основному тождеству*

$$p_{jk}^{(m+n)}=\sum_{\nu} p_{j\nu}^{(m)}p_{\nu k}^{(n)} \tag{3.3}$$

(которое является частным случаем уравнения Колмогорова — Чэпмена). Оно отражает тот простой факт, что первые m шагов приводят из E_j в некоторое промежуточное состояние E_ν и что вероятность последующего перехода из E_ν в E_k не зависит от того, каким образом было достигнуто E_ν[1]).

Так же как и в случае p_{jk}, образовавших матрицу P, *мы расположим $p_{jk}^{(n)}$ в матрицу, которую обозначим P^n*. Тогда (3.2) утверждает, что для того, чтобы получить элемент $p_{jk}^{(n+1)}$ матрицы P^{n+1}, мы должны умножить элементы j-й строки P на соответствующие элементы k-го столбца P^n и сложить все полученные произведения. Эта операция называется умножением матриц P и P^n и выражается символически равенством $P^{n+1}=PP^n$. Данное определение позволяет назвать P^n n-й степенью P; уравнение (3.3) выражает известный закон $P^{m+n}=P^mP^n$.

Для того чтобы (3.3) было справедливо для всех $n\geqslant 0$, мы определим $p_{jk}^{(0)}$, положив $p_{jj}^{(0)}=1$ и $p_{jk}^{(0)}=0$ при $j\neq k$, что вполне естественно.

Примеры. а) *Независимые испытания.* Обычно бывает трудно получить явные выражения для вероятностей перехода за несколь-

[1]) Последнее свойство является характерным для марковских процессов, которые будут определены в § 13. Долгое время предполагалось, что (3.3) можно использовать для определения цепей Маркова, однако это неожиданно оказалось неверным (см. пример 13,е)).

ко шагов, однако, к счастью, они не представляют особого интереса. Как важное, хотя и тривиальное исключение, мы отметим частный случай независимых испытаний. Этот случай имеет место тогда, когда все строки P тождественно совпадают с данным распределением вероятностей, и ясно без вычислений, что отсюда следует равенство $P^n = P$ при всех n.

б) *Серии успехов.* В примере 2, н) легко видеть (либо из рекуррентной формулы (3.2), либо из самого определения процесса), что

$$p_{jk}^{(n)} = \begin{cases} qp^k & \text{при } k=0,\ 1,\ \ldots,\ n-1, \\ p^n & \text{при } k=n+j, \\ 0 & \text{в остальных случаях.} \end{cases}$$

В этом случае ясно, что P^n сходится к матрице, такой, что все элементы в ее столбце с номером k равны qp^k. ▶

Безусловные вероятности

Пусть снова a_j означает вероятность состояния E_j в начальном (нулевом) испытании. Тогда (безусловная) вероятность попадания в E_k на n-м шаге равна

$$a_k^{(n)} = \sum_j a_j p_{jk}^{(n)}. \tag{3.4}$$

Обычно мы считаем, что процесс начинается из фиксированного состояния E_i, т. е. полагаем $a_i = 1$. В этом случае $a_k^{(n)} = p_{ik}^{(n)}$.

Интуитивно мы чувствуем, что влияние начального состояния должно постепенно ослабевать, так что при больших n распределение (3.4) должно быть почти независимым от начального распределения $\{a_j\}$. Так оно и будет, если (как в последнем примере) $p_{jk}^{(n)}$ сходится к не зависящему от j пределу, т. е. если P^n сходится к матрице с одинаковыми строками. Мы увидим, что обычно это действительно так, хотя нам и придется еще принимать в расчет досадные исключения, обусловленные периодичностью.

§ 4. ЗАМЫКАНИЯ И ЗАМКНУТЫЕ МНОЖЕСТВА

Будем говорить, что E_k *достижимо из* E_j, *если существует такое* $n \geqslant 0$, *что* $p_{jk}^{(n)} > 0$ (т. е. если имеется положительная вероятность попасть в E_k из E_j, включая случай $E_k = E_j$). Например, в неограниченном случайном блуждании каждое состояние достижимо из любого другого состояния, однако из поглощающего экрана не достижимо никакое другое состояние.

Определение. *Множество состояний C называется замкнутым, если никакое состояние вне C не достижимо ни из одного состояния E_j из C. Для произвольного множества C замыканием C называется наименьшее замкнутое множество, содержащее C.*

Если одно состояние E_k образует замкнутое множество, то оно называется поглощающим.

Цепь Маркова называется неприводимой, если не существует замкнутых множеств, отличных от множества всех состояний.

Ясно, что C замкнуто тогда и только тогда, когда $p_{jk}=0$ при всех j и k, таких, что E_j принадлежит C, а E_k не принадлежит C, ибо в этом случае мы видим из (3.2), что $p_{jk}^{(n)}=0$ при всех n. Таким образом, имеет место очевидная теорема.

Теорема. *Если вычеркнуть в матрицах P^n все строки и столбцы, соответствующие состояниям, не принадлежащим замкнутому множеству C, то останутся стохастические матрицы, для которых вновь будут справедливы фундаментальные соотношения* (3.2) *и* (3.3).

Это означает, что мы имеем цепь Маркова, определенную на C, и эту *подцепь можно изучать независимо от всех остальных состояний.*

Состояние E_k является поглощающим тогда и только тогда, когда $p_{kk}=1$; в этом случае матрица из последней теоремы сводится к одному-единственному элементу. Вообще, ясно, что все состояния E_k, достижимые из данного состояния E_j, образуют замкнутое множество. (Поскольку замыкание E_j не может быть меньше этого множества, то оно с ним совпадает.) Неприводимая цепь не содержит собственных замкнутых подмножеств, так что мы имеем следующий простой, но полезный критерий.

Критерий. *Цепь неприводима тогда и только тогда, когда каждое ее состояние достижимо из любого другого состояния.*

Примеры. а) Для того чтобы найти все замкнутые множества, достаточно знать, какие p_{jk} равны нулю и какие положительны. В соответствии с этим мы воспользуемся символом $*$ для обозначения положительных элементов и рассмотрим типичную матрицу, скажем

$$P=\begin{bmatrix} 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \\ 0 & * & * & 0 & * & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & * & 0 \\ * & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 & * & * & 0 & 0 \\ 0 & 0 & * & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & * & 0 & 0 & 0 & 0 & * \end{bmatrix}.$$

Занумеруем состояния числами от 1 до 9. В пятой строке $*$ стоит только на пятом месте, и, стало быть, $p_{55}=1$: состояние E_5 является *поглощающим.* Третья и восьмая строки содержат лишь по одному положительному элементу каждая, и ясно, что E_3 и E_8 образуют *замкнутое* множество. Из E_1 возможны переходы в E_4 и E_9, а из них — только в E_1, E_4, E_9. Следовательно, три состояния E_1, E_4, E_9 образуют другое *замкнутое* множество.

Из E_2 возможны непосредственные переходы в него самого и в E_3, E_5 и E_8. Пара (E_3, E_8) образует замкнутое множество, тогда как E_5 есть поглощающее состояние; поэтому замыкание E_2 состоит из множества E_2, E_3, E_5, E_8. Легко видеть, что замыкания оставшихся состояний E_6 и E_7 состоят из всех девяти состояний.

Внешний вид нашей матрицы и процедуру выделения замкнутых множеств можно упростить, перенумеровав состояния в следующем порядке:

$$E_5 E_3 E_8 E_1 E_4 E_9 E_2 E_6 E_7.$$

Тогда замкнутые множества будут содержать лишь соседние состояния и структура цепи будет понятна с первого взгляда на новую матрицу.

б) В матрице из примера 2, к) состояния E_1 и E_5 являются поглощающими, и других замкнутых множеств не существует.

в) В генетическом примере 2, и) состояния E_0 и E_{2N} являются поглощающими. При $0<j<2N$ замыкание E_j содержит все состояния. В примере 2, з) состояния E_0 и E_N являются поглощающими.

Рассмотрим цепь с набором состояний $E_1, \ldots, E_\rho$, такую, что $E_1, \ldots, E_r$ образуют замкнутое множество ($r<\rho$). Тогда ($r\times r$)-подматрица, стоящая в левом верхнем углу P, тоже будет стохастической, и мы можем представить P в виде блочной матрицы

$$P = \begin{bmatrix} Q & 0 \\ U & V \end{bmatrix}. \tag{4.1}$$

Матрица в верхнем правом углу имеет r строк и $\rho-r$ столбцов и состоит только из нулевых элементов. Аналогично U означает матрицу с $\rho-r$ строками и r столбцами, тогда как V есть квадратная матрица. Мы будем пользоваться символическим блочным разбиением (4.1) и в том случае, когда замкнутое множество C и его дополнение C' содержат бесконечно много состояний; это разбиение просто указывает на группировку состояний и на тот факт, что $p_{jk}=0$, если E_j принадлежит C, а E_k — его дополнению C'. Из рекуррентной формулы (3.2) очевидно, что вероятности перехода за несколько шагов допускают аналогичное разбиение:

$$P^n = \begin{bmatrix} Q^n & 0 \\ U_n & V^n \end{bmatrix}. \tag{4.2}$$

Сейчас мы не интересуемся видом элементов матрицы U_n, стоящей в левом нижнем углу. Главное здесь в том, что из (4.2) становятся ясны три очевидных, но важных факта. Во-первых, $p_{jk}^{(n)}=0$, когда $E_j \in C$, но $E_k \in C'$. Во-вторых, появление степени Q^n показывает, что когда оба состояния E_j и E_k принадлежат C, то переходные вероятности $p_{jk}^{(n)}$ получаются из рекуррентной формулы (3.2) при суммировании, ограниченном только состояниями из замкнутого множества C. Наконец, наличие V^n указывает, что последнее утверждение остается справедливым при замене C на его дополнение C'. Следовательно, можно будет упростить дальнейшее изучение цепей Маркова, рассматривая отдельно состояния, принадлежащие замкнутому множеству C, и состояния, принадлежащие дополнению C'.

Заметим, что мы не предполагали, что Q неприводима. Если C разбивается на несколько замкнутых подмножеств, то Q допускает дальнейшее разбиение. Существуют цепи с бесконечным числом замкнутых подмножеств.

Пример. г) Как уже упоминалось выше, случайное блуждание на плоскости представляет собой специальную цепь Маркова, хотя упорядочивание состояний в простую последовательность было бы неудобным для практических целей. Предположим теперь, что мы видоизменяем случайное блуждание, считая, что по достижении оси x частица будет продолжать случайное блуждание вдоль по этой оси, уже не покидая ее. Тогда точки оси x образуют бесконечное замкнутое множество. С другой стороны, если мы условимся, что по достижении оси x частица навсегда остается в точке первого попадания на эту ось, то каждая точка оси x превратится в поглощающее состояние. ▶

§ 5. КЛАССИФИКАЦИЯ СОСТОЯНИЙ

В процессе с начальным состоянием E_j последовательные возвращения в E_j представляют собой рекуррентное событие, тогда как последовательные прохождения через какое-либо иное состояние будут представлять собой уже рекуррентное событие с запаздыванием (определенное в гл. XIII,5). Стало быть, теория цепей Маркова сводится к одновременному изучению многих рекуррентных событий. Общая теория рекуррентных событий применима к ней без каких-либо модификаций, однако, чтобы избежать чрезмерного количества ссылок на гл. XIII, мы заново сформулируем основные определения. Таким образом, настоящая глава будет по существу самостоятельной и независимой от гл. XIII, за тем лишь исключением, что трудное доказательство сходимости (5.8) не будет повторено во всей полноте.

Состояния цепи Маркова будут классифицированы независимым образом с двух точек зрения. Классификация по свойствам возвратности и невозвратности является фундаментальной, тогда как классификация по свойствам периодичности и непериодичности касается технических деталей. Периодичность связана с некоторыми неудобствами, ибо требует постоянных тривиальных замечаний; начинающий читатель должен концентрировать свое внимание на цепях, не имеющих периодических состояний. Все определения в этом параграфе связаны только с матрицей переходных вероятностей и не зависят от начального распределения $\{a_j\}$.

Определение 1. *Состояние E_j имеет период $t > 1$, если $p_{jj}^{(n)} = 0$, когда n не является кратным t, и t — наибольшее целое число, обладающее этим свойством. Состояние E_j является непериодическим, если такого $t > 1$ не существует*[1].

Для изучения периодического состояния E_j достаточно рассмотреть нашу цепь при испытаниях с номерами t, $2t$, $3t$, ... Таким образом мы получим новую цепь Маркова с переходными вероятностями $p_{ik}^{(t)}$, и в этой новой цепи E_j будет уже непериодическим состоянием. При помощи этого приема результаты для непериодических состояний могут быть перенесены на периодический случай. Детали мы обсудим в § 9 и (за исключением следующего примера) сосредоточим теперь наше внимание на непериодических цепях.

Пример. а) В неограниченном случайном блуждании все состояния имеют период 2. В случайном блуждании с поглощающими экранами в точках 0 и ρ (пример 2, б)) внутренние состояния имеют период 2, однако поглощающие состояния E_0 и E_ρ будут, конечно же, непериодическими. Если хотя бы один из экранов сделать отражающим (пример 2, в)), то все состояния станут непериодическими. ▶

Обозначения. *На всем протяжении этой главы $f_{jk}^{(n)}$ означает вероятность того, что в начинающемся из E_j процессе первое попадание в E_k произойдет на n-м шаге. Мы положим $f_{jk}^{(0)} = 0$ и*

$$f_{jk} = \sum_{n=1}^{\infty} f_{jk}^{(n)}, \tag{5.1}$$

$$\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)}. \tag{5.2}$$

Очевидно, f_{jk} есть вероятность того, что, выходя из E_j, система когда-нибудь пройдет через E_k. Поэтому $f_{jk} \leqslant 1$. Когда $f_{jk} = 1$, последовательность $\{f_{jk}^{(n)}\}$ есть собственное распределение вероятностей, и мы будем называть его *распределением времени первого достижения* E_k. В частности, $\{f_{jj}^{(n)}\}$ будет распределением для *времен возвращения в* E_j. Определение (5.2) имеет смысл

[1]) Состояние E_j, возвращение в которое невозможно (т. е. для которого $p_{jj}^{(n)} = 0$ при всех $n > 0$), будет считаться непериодическим.

только тогда, когда $f_{jj}=1$, т. е. когда возвращение в E_j достоверно. В этом случае $\mu_j \leqslant \infty$ будет *средним временем возвращения в* E_j.

Для наших целей не потребуется действительного вычисления вероятностей $f_{jk}^{(n)}$, однако для ясности мы укажем, как можно определить $f_{jk}^{(n)}$ (при помощи стандартных для теории восстановления рассуждений). Если первое достижение E_k осуществляется при ν-м испытании ($1 \leqslant \nu \leqslant n-1$), то (условная) вероятность попадания в E_k при n-м испытании будет равна $p_{kk}^{(n-\nu)}$. Вспоминая наше соглашение $p_{kk}^{(0)}=1$, мы заключаем, что

$$p_{jk}^{(n)} = \sum_{\nu=1}^{n} f_{jk}^{(\nu)} p_{kk}^{(n-\nu)}. \tag{5.3}$$

Полагая последовательно $n=1, 2, \ldots$, мы получим рекуррентно $f_{jk}^{(1)}, f_{jk}^{(2)}, \ldots$. И обратно, если для пары j, k известны $f_{jk}^{(n)}$, то (5.3) определяет все переходные вероятности $p_{jk}^{(n)}$.

Для любого состояния E_j прежде всего встает вопрос о том, является ли достоверным возвращение в него. Если оно достоверно, то возникает вопрос: конечно или нет среднее время возвращения μ_j? Следующее определение согласуется с терминологией гл. XIII.

Определение 2. *Состояние E_j возвратно, если $f_{jj}=1$, и невозвратно, если $f_{jj}<1$.*

Возвратное состояние E_j называется нулевым, если для него среднее время возвращения $\mu_j=\infty$.

Это определение применимо и к периодическим состояниям. Все возвратные состояния оно подразделяет на нулевые и ненулевые состояния. Последние представляют особый интерес, и, поскольку обычно мы фокусируем наше внимание на непериодических состояниях, нам будет удобно воспользоваться термином «эргодическое состояние» для непериодических возвратных ненулевых состояний [1]). Итак, мы вводим следующее определение.

Определение 3. *Непериодическое возвратное состояние E_j, у которого $\mu_j<\infty$, называется эргодическим.*

В следующей теореме приведены выраженные через переходные вероятности $p_{jj}^{(n)}$ условия того, что E_j является состоянием того или иного типа. Содержащиеся в ней критерии весьма важны, хотя обычно слишком трудны для практического их использования. Бо-

[1]) К сожалению, эта терминология не является общепринятой. В терминологии Колмогорова невозвратные состояния называются *«несущественными»*, однако в этой главе мы намереваемся показать, что именно невозвратные состояния часто представляют теоретический и практический интерес. (Это мнение поддерживается и современной теорией потенциала.) Эргодические состояния иногда называются *«положительными»*, термин «эргодическое» иногда используется вместо нашего «возвратное».

лее удобные критерии можно найти в § 7 и 8, однако, к сожалению, простого универсального критерия не существует.

Теорема. (i) *E_j является невозвратным состоянием тогда и только тогда, когда*

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} < \infty. \tag{5.4}$$

В этом случае при всех i

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty. \tag{5.5}$$

(ii) *E_j является (возвратным) нулевым состоянием тогда и только тогда, когда*

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty, \text{ но } p_{jj}^{(n)} \to 0 \tag{5.6}$$

при $n \to \infty$. В этом случае при всех i

$$p_{ij}^{(n)} \to 0. \tag{5.7}$$

(iii) *Непериодическое (возвратное) состояние E_j является эргодическим тогда и только тогда, когда $\mu_j < \infty$. В этом случае при $n \to \infty$*

$$p_{ij}^{(n)} \to f_{ij}\mu_j^{-1}. \tag{5.8}$$

Следствие. *Если E_j — непериодическое состояние, то $p_{ij}^{(n)}$ стремится либо к нулю, либо к пределу в* (5.8).

Доказательство. Утверждение (5.4) содержится в теореме 2 гл. XIII, 3. Утверждение (5.5) является непосредственным следствием (5.4) и (5.3), однако оно содержится также в теореме 1 гл. XIII, 5.

Для непериодического возвратного состояния E_j теорема 3 гл. XIII, 3 утверждает, что $p_{jj}^{(n)} \to \mu_j^{-1}$, где правую часть надо считать равной нулю, если $\mu_j = \infty$. Утверждения (5.7) и (5.8) непосредственно следуют из этого соотношения и (5.3) или же из теоремы 1 гл. XIII, 5.

Пусть E_j возвратно и $\mu_j = \infty$. По теореме 4 гл. XIII, 3 в этом случае $p_{jj}^{(n)} \to 0$, откуда снова следует (5.7). ▶

Примеры. б) Рассмотрим состояние E_0 в цепи из примера 2, м). Своеобразная структура матрицы переходных вероятностей показывает, что первое возвращение в E_0 при n-м испытании возможно лишь при наличии последовательности переходов

$$E_0 \to E_1 \to E_2 \to \ldots \to E_{n-1} \to E_0,$$

и поэтому при $n > 1$

$$f_{00}^{(n)} = p_1 p_2 \ldots p_{n-1} q_n, \tag{5.9}$$

а $f_{00}^{(1)}=q_1$. В частном случае, когда p_k определяются формулой (2.5), это сводится к $f_{00}^{(n)}=f_n$. Поэтому E_0 невозвратно, если $\sum f_n<1$. Для возвратного E_0 среднее время возвращения μ_0 совпадает со средним распределения $\{f_n\}$. Наконец, если E_0 имеет период t, то $f_n=0$, за исключением тех n, которые кратны t. Короче говоря, как и следовало ожидать, при любых обстоятельствах E_0 имеет тот же тип, что и рекуррентное событие $\mathscr{E}$, связанное с нашей цепью Маркова.

в) Если в примере 4, а) система покинет состояние E_2, то возвращение в это состояние станет невозможным, и поэтому E_2 невозвратно. Небольшое усовершенствование этого рассуждения показывает, что состояния E_6 и E_7 также невозвратны. Из теоремы 4 § 6 следует, что все остальные состояния цепи эргодичны.

§ 6. НЕПРИВОДИМЫЕ ЦЕПИ. РАЗЛОЖЕНИЯ

Для краткости мы будем говорить, что *два состояния однотипны*, если они одинаково характеризуются во всех классификациях, введенных в предыдущем параграфе. Другими словами, два однотипных состояния либо имеют один и тот же период, либо оба непериодичны; или оба невозвратны, или оба возвратны; в последнем случае средние времена возвращения у обоих либо бесконечны, либо конечны.

Полезность нашей классификации обусловлена в значительной степени тем фактом, что для всех практических целей всегда можно ограничиться рассмотрением состояний какого-то одного типа. Следующая теорема показывает, что для неприводимых цепей это справедливо всегда.

Теорема 1. *В неприводимой цепи все состояния однотипны.*

Доказательство. Пусть E_j и E_k — два произвольных состояния неприводимой цепи. В силу критерия из § 4 каждое состояние достижимо из любого другого состояния, и поэтому существуют такие целые числа r и s, что $p_{jk}^{(r)}=\alpha>0$ и $p_{kj}^{(s)}=\beta>0$. Очевидно,

$$p_{jj}^{(n+r+s)}\geqslant p_{jk}^{(r)}p_{kk}^{(n)}p_{kj}^{(s)}=\alpha\beta p_{kk}^{(n)}. \tag{6.1}$$

Здесь j, k, r и s фиксированы, тогда как n произвольно. Если E_j невозвратно, то левая часть (6.1) является членом сходящегося ряда, и, стало быть, то же справедливо и для $p_{kk}^{(n)}$. Более того, если $p_{jj}^{(n)}\to 0$, то и $p_{kk}^{(n)}\to 0$. Аналогичные утверждения будут иметь место и в случае, когда j и k поменяются ролями, и поэтому либо E_j и E_k оба невозвратны, либо оба возвратны; если одно из них является нулевым, то таким же будет и другое состояние.

Предположим, наконец, что E_j имеет период t. При $n=0$ правая часть (6.1) положительна, и, следовательно, $r+s$ кратно t. Но тогда левая часть (6.1) при n, не кратных t, будет обращаться в нуль, и поэтому E_k имеет период, который кратен t. Меняя j и k ролями, мы видим, что эти состояния имеют одинаковый период. ▶

Важность теоремы 1 отчасти выявляет следующая теорема.

Теорема 2. *Для любого возвратного состояния E_j существует единственное неприводимое замкнутое множество C, содержащее E_j и такое, что для каждой пары E_i, E_k состояний из C*

$$f_{ik}=1 \quad и \quad f_{ki}=1. \tag{6.2}$$

Иначе говоря: выходя из произвольного состояния E_i из C, система с достоверностью пройдет через каждое состояние из множества C; по определению замыкания выход из C невозможен.

Доказательство. Пусть E_k — достижимое из E_j состояние. Тогда очевидно, что в E_k можно попасть, не возвращаясь предварительно в E_j; вероятность этого события мы обозначим через α. Но, попав в E_k, мы никогда не вернемся в E_j с вероятностью $1-f_{kj}$. Стало быть, вероятность того, что выходящая из E_j система никогда не вернется в E_j, не меньше $\alpha(1-f_{kj})$. Однако для возвратного E_j вероятность невозвращения равна нулю, и поэтому $f_{kj}=1$ для любого E_k, достижимого из E_j.

Обозначим через C совокупность всех достижимых из E_j состояний. Если E_i и E_k принадлежат C, то, как мы видели, E_j достижимо из E_k, и, следовательно, E_i также достижимо из E_k. Таким образом, каждое состояние из C достижимо из любого другого состояния из C, и поэтому в силу критерия из § 4 C неприводимо. Отсюда следует, что все состояния из C возвратны, и поэтому роль E_j в первой части рассуждения может играть любое E_i из этого множества. Это означает, что $f_{ki}=1$ для всех E_k из C, и тем самым справедливо (6.2). ▶

Из этой теоремы вытекает, что замыкание возвратного состояния неприводимо. Для невозвратного состояния это не обязательно так.

Пример. Предположим, что $p_{jk}=0$ при $k\leqslant j$, но $p_{j,\,j+1}>0$. Здесь имеют место только переходы в состояния с бо́льшими номерами, и поэтому невозможно возвращение ни в одно состояние. Каждое E_j невозвратно, и замыкание E_j состоит из состояний E_j, E_{j+1}, E_{j+2}, . . ., однако и само это замыкание содержит замкнутое подмножество, получающееся при вычеркивании E_j. Отсюда следует, что неприводимых множеств у этой цепи не существует. ▶

Из последней теоремы вытекает, в частности, что из возвратного состояния не достигается ни одно невозвратное состояние. Если же цепь содержит оба типа состояний, то это означает, что матрица P допускает символическое разбиение вида (4.1), где матрица Q соответствует возвратным состояниям. Излишне говорить, что матрица Q вновь может быть разложимой. Но каждое возвратное состояние принадлежит единственному *неприводимому* подмножеству, и между этими подмножествами невозможны никакие переходы. Сказанное резюмирует следующая теорема.

Теорема 3. *Состояния цепи Маркова единственным образом могут быть разбиты на неперекрывающиеся множества $T, C_1, C_2, \ldots$, такие, что*

(i) *T состоит из всех невозвратных состояний;*

(ii) *если E_j принадлежит C_ν, то $f_{jk}=1$ для всех E_k из C_ν, тогда как $f_{jk}=0$ для всех E_k, не принадлежащих C_ν.*

Отсюда вытекает, что C_ν неприводимо и содержит только однотипные возвратные состояния. Приведенный выше пример показывает, что все состояния могут быть невозвратными, а пример 4, г) доказывает, что множеств C_ν может быть бесконечно много.

Следующую теорему мы выведем как простое следствие из теоремы 2, однако она может быть доказана и другими простыми способами (см. задачи 18—20).

Теорема 4. *В конечной цепи не существует нулевых состояний, и все состояния ее не могут быть невозвратными.*

Доказательство. Для каждой строки матрицы P^n сумма элементов равна единице, и так как строки эти состоят из фиксированного числа элементов, то невозможно, чтобы $p_{jk}^{(n)} \to 0$ для всех j, k. Поэтому не все состояния невозвратны. Но возвратное состояние принадлежит некоторому неприводимому множеству C. Все состояния из C однотипны. Поэтому тот факт, что C содержит возвратное состояние и хотя бы одно ненулевое состояние, и будет означать, что в C нет ни одного нулевого состояния. ▶

§ 7. ИНВАРИАНТНЫЕ РАСПРЕДЕЛЕНИЯ

Поскольку каждое возвратное состояние принадлежит некоторому неприводимому множеству, асимптотическое поведение которого можно изучать независимо от остальных состояний, мы сейчас займемся исключительно неприводимыми цепями. Все состояния такой цепи однотипны, и мы начнем с простейшего случая, а именно с цепей с конечными средними временами возвращения μ_j. Чтобы избежать тривиальных переформулировок, мы отложим обсуждение периодических цепей до § 9. Иначе говоря, мы рассмотрим теперь цепи, состояния которых непериодичны и возвратны с конечными временами возвращения. Такие цепи называются эргодическими (определение 5.3).

Теорема. *В неприводимой эргодической цепи существуют не зависящие от начального состояния j пределы*

$$u_k = \lim_{n\to\infty} p_{jk}^{(n)}. \tag{7.1}$$

Более того, $u_k > 0$,

$$\sum u_k = 1, \tag{7.2}$$

и [1])

$$u_j = \sum_i u_i p_{ij}. \tag{7.3}$$

Обратно, предположим, что цепь неприводима и непериодична и что существуют числа $u_k \geqslant 0$, удовлетворяющие (7.2) и (7.3). Тогда цепь эргодична, и для u_k справедливы соотношения (7.1) и

$$u_k = 1/\mu_k, \tag{7.4}$$

где μ_k — среднее время возвращения в E_k.

Доказательство. (i) Предположим, что цепь неприводима и эргодична, и определим u_k по формуле (7.4). Теорема 2 § 6 гарантирует, что $f_{ij}=1$ для каждой пары состояний, и поэтому утверждение (7.1) сводится к (5.8). Далее,

$$p_{ik}^{(n+1)} = \sum_j p_{ij}^{(n)} p_{jk}. \tag{7.5}$$

При $n \to \infty$ левая часть сходится к u_k, тогда как общий член справа стремится к $u_j p_{jk}$. Взяв справа лишь конечное число членов, мы заключаем, что

$$u_k \geqslant \sum_j u_j p_{jk}. \tag{7.6}$$

Для фиксированных i и n левые части в (7.5) дают в сумме единицу, и поэтому

$$s = \sum u_k \leqslant 1. \tag{7.7}$$

Суммируя по k в (7.6), мы получаем соотношение $s \geqslant s$, знак неравенства в котором невозможен. Следовательно, в (7.6) при всех k справедливо равенство, и тем самым первая часть теоремы доказана.

(ii) Допустим теперь, что $u_k \geqslant 0$ и выполняются равенства (7.2) и (7.3). По индукции

$$u_k = \sum_i u_i p_{ik}^{(n)} \tag{7.8}$$

для всех $n > 1$. Поскольку цепь предполагается неприводимой, все состояния однотипны. Если бы они были невозвратными или нулевыми, то правая часть (7.8) стремилась бы к 0 при $n \to \infty$, а этого при всех k быть не может, ибо u_k в сумме составляют единицу. Поскольку случай периодических цепей мы исключили, то это означает, что цепь эргодична, так что применима первая часть теоремы.

---

[1]) Если представить $\{u_j\}$ как вектор-строку, то (7.3) можно будет записать в матричной форме: $u = uP$.

Поэтому, устремляя n к ∞, получаем

$$u_k = \sum_i u_i \mu_k^{-1}. \tag{7.9}$$

Следовательно, распределение вероятностей $\{u_k\}$ пропорционально распределению $\{\mu_k^{-1}\}$, и поэтому, как утверждалось, $u_k = \mu_k^{-1}$. ►

Чтобы оценить все значение этой теоремы, рассмотрим эволюцию процесса от начального распределения $\{a_j\}$. Вероятность попадания при n-м шаге в состояние E_k равна

$$a_k^{(n)} = \sum_j a_j p_{jk}^{(n)} \tag{7.10}$$

(см. (3.4)). Поэтому в силу (7.1) при $n \longrightarrow \infty$

$$a_k^{(n)} \longrightarrow u_k. \tag{7.11}$$

Иначе говоря, каким бы ни было начальное распределение, вероятность состояния E_k стремится к u_k. С другой стороны, если начальное распределение равно $\{u_k\}$ (т. е. если $a_k = u_k$), то из (7.3) вытекает, что $a_k^{(1)} = u_k$ и — по индукции — что $a_k^{(n)} = u_k$ при всех n. Таким образом, удовлетворяющее (7.3) начальное распределение «увековечивается» на все времена. По этой причине оно называется инвариантным.

Определение. *Распределение вероятностей $\{u_k\}$, удовлетворяющее условию (7.3), называется инвариантным или стационарным (для данной цепи Маркова).*

Теперь главную часть предыдущей теоремы можно переформулировать следующим образом.

Неприводимая непериодичная цепь Маркова обладает инвариантным распределением вероятностей $\{u_k\}$ тогда и только тогда, когда она эргодична. В этом случае $u_k > 0$ при всех k и безусловные вероятности $a_k^{(n)}$ стремятся к u_k независимо от начального распределения.

Физический смысл стационарности станет понятным, если мы представим себе большое число одновременно протекающих процессов. Конкретнее, рассмотрим N частиц, независимо совершающих случайные блуждания одного типа. Среднее число частиц, находящихся на n-м шаге в состоянии E_k, равно величине $Na_k^{(n)}$, которая стремится к Nu_k при $n \to \infty$. После достаточно длительного промежутка времени распределение станет приблизительно инвариантным, и физик в такой ситуации сказал бы, что он наблюдает частицы в равновесии. Поэтому распределение $\{u_k\}$ называется также *равновесным распределением*. К сожалению, этот термин отвлекает внимание от того важного обстоятельства, что относится он к так называемому *макроскопическому равновесию*, т. е.

равновесию, поддерживаемому большим числом переходов в противоположных направлениях. Отдельная же частица не проявляет никакой тенденции к равновесию, и для этого индивидуального процесса никаких следствий из нашей предельной теоремы не проистекает.

Типичным в этом отношении является симметричное случайное блуждание, рассматривавшееся в гл. III. Если большое число частиц независимо совершают такие случайные блуждания, начавшиеся в нуле, то в любой момент времени приблизительно половина из них будет справа, а другая половина слева от нуля. Однако это не означает, что большинство частиц проводят половину своего времени на положительной стороне. Напротив, законы арксинуса показывают, что большинство частиц проводят непропорционально большую часть своего времени на одной стороне от нуля, и в этом смысле *большинство не является представителем ансамбля*. Этот случай является крайним в том смысле, что средние времена возвращения в нем бесконечны. В эргодических цепях случайные флуктуации более умеренны, но практически они будут носить тот же характер, если времена возвращения будут иметь очень большие (или бесконечные) дисперсии. Должным образом осознав статистическую природу «тенденции к равновесию», можно было бы избежать многих пространных рассуждений и ложных заключений.

В предыдущей теореме мы предполагали цепь неприводимой и непериодичной, и теперь уместно спросить, в какой степени существенны эти предположения. Внимательное чтение доказательства показывает, что на самом деле мы доказали больше, чем утверждалось в теореме. В частности, мы по ходу дела получили следующий критерий, применимый к произвольным цепям (включая периодические и приводимые цепи).

Критерий. *Если цепь обладает инвариантным распределением вероятностей* $\{u_k\}$, *то* $u_k=0$ *для каждого* E_k, *являющегося либо невозвратным, либо возвратным нулевым состоянием.*

Иначе говоря, из $u_k>0$ вытекает, что E_k возвратно и имеет конечное среднее время возвращения, но может быть и периодическим.

Доказательство. Мы видели, что из стационарности распределения $\{u_k\}$ следует соотношение (7.8). Если E_k — невозвратное или нулевое состояние, то $p_{jk}^{(n)} \to 0$ при всех j, и поэтому, как и утверждалось, $u_k=0$. ▶

Что касается периодических цепей, то, забегая вперед, упомянем результат, доказанный в § 9 и заключающийся в том, что *для каждой неприводимой цепи, состояния которой имеют конечные средние времена возвращения, существует единственное инвариантное распределение вероятностей* $\{u_k\}$. Случай периодических цепей был исключен в настоящем параграфе лишь потому, что простые предельные соотношения (7.1) и (7.11) принимают для него менее привлекательный вид, который только отвлекает от существа дела, в действительности на него не влияя.

Примеры. а) Цепи с несколькими неприводимыми компонентами могут допускать несколько стационарных решений. Банальный, но типичный пример этого — случайное блуждание с двумя поглощающими состояниями E_0 и E_ρ (см. пример 2, б)). Для него каждое распределение вероятностей вида $(\alpha, 0, 0, \ldots, 0, 1-\alpha)$, приписывающее положительные веса только состояниям E_0 и E_ρ, является стационарным.

б) По данной матрице переходных вероятностей не всегда легко определить, существует ли инвариантное распределение $\{u_k\}$. Замечательное исключение составляет случай, когда

$$p_{jk}=0 \quad при \quad |k-j|>1, \tag{7.12}$$

т. е. когда все ненулевые элементы матрицы лежат на главной диагонали или на непосредственно примыкающей к ней прямой. Если состояния перенумерованы, начиная с 0, то соотношения (7.3) здесь примут вид

$$\begin{aligned} u_0&=p_{00}u_0+p_{10}u_1, \\ u_1&=p_{01}u_0+p_{11}u_1+p_{21}u_2 \end{aligned} \tag{7.13}$$

и т. д. Чтобы избежать тривиальных оговорок, предположим, что $p_{j,\,j+1}>0$ и $p_{j,\,j-1}>0$ при всех $j>0$; относительно диагональных элементов p_{jj} не предполагается ничего. Уравнения (7.13) могут быть последовательно разрешены относительно $u_1, u_2, \ldots$. Вспоминая, что для каждой строки матрицы P сумма элементов равна единице, мы находим

$$u_1=\frac{p_{01}}{p_{10}}u_0, \quad u_2=\frac{p_{01}p_{12}}{p_{10}p_{21}}u_0, \quad u_3=\frac{p_{01}p_{12}p_{23}}{p_{10}p_{21}p_{32}}u_0 \tag{7.14}$$

и т. д. Получаемая в результате (конечная или бесконечная) последовательность $u_0, u_1, \ldots$ представляет собой единственное решение (7.13). Чтобы превратить ее в распределение вероятностей, надо выбрать нормирующий множитель u_0 так, чтобы $\sum u_k=1$. Такой выбор возможен тогда и только тогда, когда

$$\sum \frac{p_{01}p_{12}p_{23}\cdots p_{k-1,\,k}}{p_{10}p_{21}p_{32}\cdots p_{k,\,k-1}}<\infty. \tag{7.15}$$

Таким образом, это есть необходимое и достаточное условие для существования инвариантного распределения вероятностей; если же последнее существует, то оно с необходимостью единственно. (Если (7.15) не выполняется, то (7.14) является так называемой инвариантной мерой; см. § 11.)

В примере 8,г) мы выведем аналогичный критерий для проверки возвратности состояний. Следующие три примера иллюстрируют применимость нашего критерия.

в) *Отражающие экраны.* Пример 2, в) (с $\rho \leqslant \infty$) является частным случаем предыдущего примера с $p_{j,j+1} = p$ при всех $j < \rho$ и $p_{j,j-1} = q$ при всех $j > 1$. Когда число состояний конечно, существует инвариантное распределение с u_k, пропорциональными $(p/q)^k$. Если же число состояний бесконечно, то для сходимости (7.15) требуется, чтобы $p < q$, и в этом случае $u_k = (1 - p/q)(p/q)^k$. Из общей теории случайных блужданий ясно, что при $p > q$ состояния цепи будут невозвратными, а при $p = q$ — возвратными нулевыми. Это будет вытекать также из критерия в примере 8,г).

г) *Модель Эренфестов для диффузии.* Для матрицы из примера 2,д) решение (7.14) сводится к

$$u_k = \binom{\rho}{k} u_0, \quad k = 0, \ \ldots, \ \rho. \tag{7.16}$$

Биномиальные коэффициенты являются членами разложения бинома Ньютона $(1+1)^\rho$, и поэтому, чтобы получить распределение вероятностей, мы должны положить $u_0 = 2^{-\rho}$. Эта цепь имеет период 2, состояния имеют конечные средние времена возвращения, и *инвариантным является биномиальное распределение с* $p = 1/2$.

Данный результат можно интерпретировать следующим образом: каково бы ни было начальное число молекул в первом сосуде, через достаточно долгое время вероятность обнаружения в нем k молекул будет примерно такой же, как если бы ρ молекул были распределены по сосудам случайным образом, причем для каждой молекулы вероятность находиться в первом сосуде равнялась бы 1/2. Это типичный пример того, как наши результаты приобретают физический смысл.

При больших ρ нормальное приближение для биномиального распределения показывает, что, коль скоро приблизительно установится предельное распределение, мы практически наверняка найдем примерно половину молекул в каждом из сосудов. Для физика число $\rho = 10^6$, конечно, является малым. Но даже при $\rho = 10^6$ молекул вероятность обнаружить более 505 000 молекул в одном сосуде (флуктуация плотности примерно на 1 процент) имеет величину порядка 10^{-23}. При $\rho = 10^8$ флуктуация плотности на одну тысячную будет иметь такую же пренебрежимо малую вероятность. Верно, что система будет изредка попадать в весьма невероятные состояния, однако времена возвращения для них фантастически велики по сравнению с временами возвращения в состояния, близкие к равновесию. Физическая необратимость проявляется в том факте, что когда бы система ни оказалась в состоянии, далеком от равновесия, для нее гораздо правдоподобнее изменения в сторону равновесия, чем в противоположном направлении.

д) *Модель Бернулли — Лапласа для диффузии.* Для матрицы с элементами (2.1) получаем из (7.14), что

$$u_k = \binom{\rho}{k}^2 u_0, \quad k = 0, \ \ldots, \ \rho. \tag{7.17}$$

Сумма квадратов биномиальных коэффициентов равна $\binom{2\rho}{\rho}$ (см. формулу (12.11) гл. II), и, следовательно,

$$u_k = \binom{\rho}{k}^2 \binom{2\rho}{\rho}^{-1} \tag{7.18}$$

представляет собой *инвариантное распределение*. Это гипергеометрическое распределение (см. гл. II, 6). Таким образом, в состоянии равновесия распределение частиц по цвету в каждой урне такое же, как если бы содержащиеся в ней ρ частиц были выбраны случайным образом из набора из ρ черных и ρ белых частиц.

е) В примере 2, м) соотношения для инвариантного распределения вероятностей примут вид

$$u_k = p_k u_{k-1}, \quad k=1,\ 2,\ \ldots, \tag{7.19 а}$$

$$u_0 = q_1 u_0 + q_2 u_1 + q_3 u_2 + \ldots . \tag{7.19 б}$$

Из (7.19а) получаем

$$u_k = p_1 \ldots p_k u_0, \tag{7.20}$$

и теперь легко видеть, что первые k членов в правой части (7.19б) в сумме составляют $u_0 - u_k$. Поэтому (7.19б) автоматически выполняется, если $u_k \to 0$, и *инвариантное распределение вероятностей существует тогда и только тогда, когда*

$$\sum_k p_1 p_2 \ldots p_k < \infty. \tag{7.21}$$

(См. также примеры 8, д) и 11, в).)

ж) *Рекуррентные события*. В примере 2,л) условия для инвариантного распределения вероятностей сводятся к

$$u_k = u_{k+1} + f_{k+1} u_0, \quad k=0,\ 1,\ \ldots . \tag{7.22}$$

Суммируя по $k=0,\ 1, \ldots$, получаем

$$u_n = r_n u_0, \quad \textit{где}\ r_n = f_{n+1} + f_{n+2} + \ldots . \tag{7.23}$$

Но $r_0 + r_1 + \ldots = \mu$ равно математическому ожиданию распределения $\{f_k\}$. Итак, *инвариантное распределение вероятностей дается формулой $u_n = r_n/\mu$, если $\mu < \infty$*; при $\mu = \infty$ такого распределения не существует.

Напомним, что наша цепь Маркова связана с рекуррентным событием $\mathscr{E}$ с распределением $\{f_k\}$ времени возвращения. В частном случае $p_k = r_k / r_{k-1}$ цепь из предыдущего примера связана с тем же рекуррентным событием $\mathscr{E}$, и в этом случае (7.20) и (7.23) эквивалентны. Следовательно, *инвариантные распределения здесь совпадают*. На языке теории массового обслуживания следовало бы сказать, что *распределения затраченного времени ожидания и оста-*

точного времени ожидания сходятся к одному и тому же распределению, а именно к $\{r_n/\mu\}$.

Основные предельные теоремы для цепей Маркова мы вывели из теории рекуррентных событий. Теперь мы видим, что и, наоборот, рекуррентные события могут рассматриваться как специальные цепи Маркова. (См. также пример 11,г).)

з) *Дважды стохастические матрицы*. Стохастическая матрица P называется дважды стохастической, если не только суммы по строкам, но и суммы по столбцам равны единице. Если такая цепь содержит лишь конечное число a состояний, то инвариантным распределением будет $u_k=a^{-1}$. Это означает, что при макроскопическом равновесии все состояния равновероятны. ▶

§ 8. НЕВОЗВРАТНЫЕ СОСТОЯНИЯ

Мы видели в § 6, что возвратные состояния любой цепи Маркова могут быть разбиты на неперекрывающиеся замкнутые неприводимые множества $C_1, C_2, \ldots$. В общем случае существует также непустой класс T невозвратных состояний. Когда система выходит из невозвратного состояния, возникают две возможности: либо система в конце концов попадает в одно из замкнутых множеств C_ν и останется в нем навсегда, либо система останется навсегда во множестве T невозвратных состояний. Наша главная задача состоит в определении соответствующих вероятностей. Решение ее даст нам критерий, позволяющий решить, каким является данное состояние — возвратным или невозвратным.

Примеры. а) *Мартингалы*. Цепь называется мартингалом, если для каждого j математическое ожидание распределения вероятностей $\{p_{jk}\}$ равно j, т. е. если [1])

$$\sum_k p_{jk}k = j. \tag{8.1}$$

Рассмотрим конечную цепь с набором состояний $E_0, \ldots, E_a$. Полагая в (8.1) $j=0$ и $j=a$, мы видим, что $p_{00}=p_{aa}=1$, и поэтому E_0 и E_a суть поглощающие состояния. Чтобы избежать тривиальных оговорок, предположим, что цепь не содержит других замкнутых множеств. Отсюда следует, что внутренние состояния $E_1, \ldots, E_{a-1}$ невозвратны, и поэтому процесс рано или поздно закончится либо в E_0, либо в E_a. Из (8.1) выводим по индукции, что при всех n

$$\sum_{k=0}^{a} p_{ik}^{(n)}k = i. \tag{8.2}$$

[1]) Очевидно, если отождествить E_j с целым числом j, то такая цепь будет мартингалом в смысле обычного определения (см. том 2). — *Прим. перев.*

Но $p_{ik}^{(n)} \to 0$ для каждого невозвратного состояния E_k, и поэтому из (8.2) вытекает, что при всех $i > 0$

$$p_{ia}^{(n)} \to i/a. \tag{8.3}$$

Иначе говоря, если процесс начинается в E_i, то *вероятности окончательного поглощения в* E_0 *и* E_a *равны* $1-i/a$ *и* i/a *соответственно.*

б) *Частные случаи.* Рассмотренные в примерах из генетики 2, з) и 2, и) цепи имеют вид, обсуждавшийся в предыдущем примере, с $a=N$ и $a=2N$ соответственно. Стало быть, при условии, что начальным состоянием является E_i, вероятность окончательного прекращения процесса в E_0 равна $1-i/a$.

в) Рассмотрим цепь с набором состояний $E_0, E_1, \ldots$, такую, что E_0 — поглощающее состояние, тогда как из других состояний E_j возможны переходы только в правое соседнее состояние E_{j+1} и в E_0. Для $j \geqslant 1$ положим

$$p_{j0}=\varepsilon_j, \quad p_{j,j+1}=1-\varepsilon_j, \tag{8.4}$$

где $\varepsilon_j > 0$. При начальном состоянии E_j вероятность непоглощения в E_0 в первых n испытаниях равна

$$(1-\varepsilon_j)(1-\varepsilon_{j+1}) \ldots (1-\varepsilon_{j+n-1}). \tag{8.5}$$

Это произведение убывает с ростом n и, следовательно, стремится к некоторому пределу λ_j. Отсюда мы выводим, что вероятность окончательного поглощения равна $1-\lambda_j$, тогда как с вероятностью λ_j система навсегда останется в множестве невозвратных состояний. Для положительности λ_j необходимо и достаточно, чтобы $\sum \varepsilon_k < \infty$. ►

Основой для изучения невозвратных состояний является подматрица матрицы P, получающаяся вычеркиванием всех строк и столбцов, соответствующих возвратным состояниям, и содержащая лишь те элементы p_{jk}, для которых и E_j, и E_k невозвратны. Суммы элементов строк этой подматрицы уже не равны единице, и здесь удобно ввести следующее определение.

Определение. *Квадратная матрица Q с элементами q_{ik} называется субстохастической, если $q_{ik} \geqslant 0$ и суммы элементов каждой строки* $\leqslant 1$.

Каждая стохастическая матрица является субстохастической в смысле этого определения, и, наоборот, каждая субстохастическая матрица может быть дополнена до стохастической добавлением поглощающего состояния E_0. (Иначе говоря, мы добавляем сверху строку 1, 0, 0, . . ., а слева — столбец, элементы которого равны дефектам строк [1]) матрицы Q.) Поэтому очевидно, что все сказанное

[1]) Дефектом строки называется разность между единицей и суммой элементов этой строки.— *Прим. перев.*

относительно стохастических матриц применимо без существенных изменений и к субстохастическим матрицам. В частности, рекуррентное соотношение (3.2) определяет n-ю степень Q^n как матрицу с элементами

$$q_{ik}^{(n+1)}=\sum_{\nu} q_{i\nu} q_{\nu k}^{(n)}. \tag{8.6}$$

Обозначим через $\sigma_i^{(n)}$ сумму элементов i-й строки Q^n. Тогда для $n \geqslant 1$

$$\sigma_i^{(n+1)}=\sum_{\nu} q_{i\nu}\sigma_{\nu}^{(n)}, \tag{8.7}$$

и это соотношение остается справедливым и для $n=0$, если положить $\sigma_{\nu}^{(0)}=1$ при всех ν. Тот факт, что Q является субстохастической, означает, что $\sigma_i^{(1)} \leqslant \sigma_i^{(0)}$, и из (8.7) выводим по индукции, что $\sigma_i^{(n+1)} \leqslant \sigma_i^{(n)}$. Стало быть, для фиксированного i последовательность $\{\sigma_i^{(n)}\}$ монотонно убывает к пределу $\sigma_i \geqslant 0$, и ясно, что

$$\sigma_i=\sum_{\nu} q_{i\nu}\sigma_{\nu}. \tag{8.8}$$

Вся теория невозвратных состояний основывается на решениях этой системы уравнений. В некоторых случаях ненулевых решений у (8.8) не существует (т. е. $\sigma_i=0$ для всех i). В других случаях может существовать бесконечно много линейно независимых решений, т. е. различных последовательностей чисел x_i, удовлетворяющих системе

$$x_i=\sum_{\nu} q_{i\nu} x_{\nu}. \tag{8.9}$$

Первая наша задача состоит в том, чтобы охарактеризовать частное решение $\{\sigma_i\}$. Мы заинтересованы только в тех решениях $\{x_i\}$, для которых $0 \leqslant x_i \leqslant 1$ при всех i. Это можно переписать в виде $0 \leqslant x_i \leqslant \sigma_i^{(0)}$; сравнивая (8.9) и (8.7), мы видим, что по индукции $x_i \leqslant \sigma_i^{(n)}$ при всех n, и поэтому

$$0 \leqslant x_i \leqslant 1 \text{ влечет } x_i \leqslant \sigma_i \leqslant 1. \tag{8.10}$$

Решение $\{\sigma_i\}$ мы будем называть *максимальным*, однако следует помнить, что во многих случаях $\sigma_i=0$ при всех i. Наш результат резюмирует следующая лемма.

Лемма. *Для субстохастической матрицы Q линейная система (8.9) обладает максимальным решением $\{\sigma_i\}$ со свойством (8.10). Эти σ_i являются пределами сумм элементов строк матриц Q^n.*

Отождествим теперь Q с подматрицей матрицы P, составленной только из тех элементов p_{jk}, для которых E_j и E_k невозвратны.

Линейную систему (8.9) можно записать в виде

$$x_i = \sum_T p_{i\nu} x_\nu, \quad E_i \in T, \tag{8.11}$$

где суммирование распространяется только на те ν, при которых E_ν принадлежит множеству T невозвратных состояний. При таком отождествлении $\sigma_i^{(n)}$ будет вероятностью того, что за первые n испытаний с начальным состоянием E_i не произойдет ни одного перехода в какое-либо возвратное состояние. Следовательно, предел σ_i равен вероятности того, что ни один такой переход не произойдет никогда. Таким образом, доказана следующая теорема.

Теорема 1. *Вероятности x_i того, что система с начальным состоянием E_i навсегда останется среди невозвратных состояний, даются максимальным решением системы* (8.11).

Такие же рассуждения дают нам следующий критерий.

Критерий. *В неприводимой*[1] *цепи Маркова с набором состояний $E_0, E_1, \ldots$ состояние E_0 является возвратным тогда и только тогда, когда линейная система*

$$x_i = \sum_{\nu=1}^{\infty} p_{i\nu} x_\nu, \quad i \geqslant 1, \tag{8.12}$$

не допускает никаких решений, удовлетворяющих $0 < x_i \leqslant 1$, за исключением $x_i = 0$ при всех i.

Доказательство. Отождествим матрицу Q из леммы с подматрицей матрицы P, получающейся вычеркиванием строки и столбца, соответствующих E_0. Рассуждения, использовавшиеся нами в теореме 1, показывают, что σ_i есть вероятность того, что (при начальном состоянии E_i) система навсегда останется среди состояний $E_1, E_2, \ldots$. Но если E_0 возвратно, то вероятность f_{i0} достижения E_0 равна 1, и, следовательно, $\sigma_i = 0$ при всех i. ▶

Примеры. г) Как и в примере 7, б), рассмотрим цепь с набором состояний $E_0, E_1, \ldots$, такую, что

$$p_{jk} = 0, \text{ когда } |k-j| > 1. \tag{8.13}$$

Чтобы избежать тривиальных оговорок, мы предположим, что $p_{j,\,j+1} \neq 0$ и $p_{j,\,j-1} \neq 0$. Эта цепь неприводима, потому что каждое состояние достижимо из любого другого состояния. Поэтому все состояния однотипны, и нам достаточно проверить характер E_0.

[1] Неприводимость предполагается только для того, чтобы избежать усложнения обозначений. Никакого ограничения она здесь не представляет, поскольку нам достаточно рассмотреть замыкание E_0. Критерий применим также к периодическим цепям.

Уравнения (8.12) сводятся к рекуррентной системе

$$\begin{aligned} x_1 &= p_{11}x_1 + p_{12}x_2, \\ p_{j,j-1}(x_j - x_{j-1}) &= p_{j,j+1}(x_{j+1} - x_j), \quad j \geqslant 2. \end{aligned} \tag{8.14}$$

Поэтому

$$x_j - x_{j+1} = \frac{p_{21}p_{32}\ldots p_{j,\,j-1}}{p_{23}p_{34}\ldots p_{j,\,j+1}}(x_1 - x_2). \tag{8.15}$$

Так как $p_{10} > 0$, то $x_2 - x_1 > 0$, и поэтому ограниченное неотрицательное решение $\{x_j\}$ существует тогда и только тогда, когда

$$\sum \frac{p_{21}\ldots p_{j,\,j-1}}{p_{23}\ldots p_{j,\,j+1}} < \infty. \tag{8.16}$$

Наша цепь возвратна тогда и только тогда, когда этот ряд расходится. В частном случае случайных блужданий мы имеем $p_{j,\,j+1} = p$ и $p_{j,\,j-1} = q$ при всех $j > 1$, и мы вновь видим, что все состояния возвратны тогда и только тогда, когда $p \leqslant q$.

(Цепь из этого примера можно интерпретировать как случайное блуждание на прямой с вероятностями, меняющимися от точки к точке.)

д) Для матрицы из примера 2, м) уравнения (8.12) сводятся к

$$x_j = p_{j+1}x_{j+1}, \tag{8.17}$$

и здесь ограниченное положительное решение существует тогда и только тогда, когда бесконечное произведение $p_1p_2\ldots$ сходится. Если наша цепь связана с рекуррентным событием $\mathscr{E}$, то p_k даются формулой (2.5), и это произведение сходится тогда и только тогда, когда $\sum f_j < \infty$. Таким образом (как и можно было бы предвидеть), и цепь, и событие $\mathscr{E}$ либо оба возвратны, либо оба невозвратны. ►

Чтобы ответить на последний вопрос, поставленный в начале этого параграфа, снова обозначим через T класс невозвратных состояний; пусть C — произвольное *замкнутое* множество возвратных состояний. (Мы не требуем, чтобы C было неприводимым.) Обозначим через y_i *вероятность окончательного поглощения в* C при условии, что начальным состоянием было E_i. Мы намереваемся показать, что y_i удовлетворяют системе неоднородных уравнений

$$y_i = \sum_T p_{i\nu}y_\nu + \sum_C p_{i\nu}, \quad E_i \in T, \tag{8.18}$$

где суммы берутся по тем ν, для которых $E_\nu \in T$ и $E_\nu \in C$ соответственно. Система (8.18) может допускать несколько независимых решений, однако следующее доказательство покажет, что среди них существует *минимальное* решение, определяемое очевидным образом по аналогии с (8.10).

Теорема 2. *Вероятности y_i окончательного поглощения замкнутым возвратным множеством C даются минимальным неотрицательным решением системы* (8.18).

Доказательство. Обозначим через $y_i^{(n)}$ вероятности поглощения множеством C на n-м шаге или до него. Тогда ясно, что при $n \geqslant 1$

$$y_i^{(n+1)} = \sum_T p_{i\nu} y_\nu^{(n)} + \sum_C p_{i\nu}, \tag{8.19}$$

и если мы положим $y_\nu^{(0)} = 0$ для всех ν, то (8.19) будет справедливо и при $n = 0$. Для фиксированного i последовательность $\{y_i^{(n)}\}$ не убывает, но остается ограниченной единицей. Пределы этих последовательностей, очевидно, удовлетворяют (8.18). Обратно, если $\{y_i\}$ — произвольное неотрицательное решение системы (8.18), то $y_i \geqslant y_i^{(1)}$ поскольку вторая сумма в (8.18) равна $y_i^{(1)}$. По индукции $y_i \geqslant y_i^{(n)}$ при всех n, так что пределы $y_i^{(n)}$ действительно являются минимальным решением. ►

Иллюстрацию к этой теореме см. в примере в).

§ 9 *). ПЕРИОДИЧЕСКИЕ ЦЕПИ

Ни трудностей, ни неожиданных новых свойств в случае периодических цепей не появляется. Эти цепи были исключены из формулировки главной теоремы в § 7 лишь потому, что они представляют вторичный интерес, а описание их оказывается непропорционально многословным. Содержащееся в настоящем параграфе обсуждение этого случая проводится скорее ради полноты изложения, чем из-за представляемого им интереса.

Простейшим примером цепи с периодом 3 является цепь с тремя состояниями, в которой возможны только переходы $E_1 \to E_2 \to E_3 \to E_1$. В этом случае

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad P^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad P^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Мы покажем теперь, что этот пример во многих отношениях типичен.

Рассмотрим неприводимую цепь с конечным или бесконечным числом состояний $E_1, E_2, \ldots$. По теореме 6.1 все состояния имеют один и тот же период t (мы предполагаем, что $t > 1$). Поскольку в неприводимой цепи каждое состояние достижимо из любого другого состояния, то для каждого состояния E_k существуют два целых числа a и b, такие, что $p_{1k}^{(a)} > 0$ и $p_{k1}^{(b)} > 0$. Но $p_{11}^{(a+b)} \geqslant p_{1k}^{(a)} p_{k1}^{(b)}$,

*) Результаты этого параграфа не будут использоваться в дальнейшем.

и поэтому сумма $a+b$ должна быть кратной периоду t. Фиксируя b, мы заключаем, что каждое целое a, для которого $p_{1k}^{(a)}>0$, имеет вид $\alpha+\nu t$, где α—фиксированное целое, $0\leqslant\alpha<t$. Целое число α является характеристикой состояния E_k, и поэтому множество всех состояний может быть разбито на t непересекающихся классов $G_0, \ldots, G_{t-1}$ так, что

$$\text{если } E_k\in G_\alpha, \quad \text{то } p_{1k}^{(n)}=0 \quad \text{при } n\neq\alpha+\nu t. \tag{9.1}$$

Мы будем считать, что классы $G_0, \ldots, G_{t-1}$ упорядочены циклическим образом, так что G_{t-1} является левым соседом G_0.

Теперь очевидно, что за *один шаг* переход возможен только в состояние из правого соседнего класса, и, следовательно, каждый путь из t шагов всегда приводит в состояние из исходного класса. Отсюда вытекает, что в цепи Маркова с переходной матрицей P^t каждый класс G_α будет замкнутым множеством[1]). Это множество будет неприводимо, потому что в исходной цепи каждое состояние достижимо из любого другого, и внутри одного класса необходимое для этого число шагов с необходимостью должно делиться на t. Таким образом, нами доказана следующая теорема.

Теорема. *В неприводимой цепи с периодом t состояния могут быть разбиты на t непересекающихся классов $G_0, \ldots, G_{t-1}$, таких, что справедливо* (9.1) *и переходы за один шаг всегда приводят в состояние из правого соседнего класса (в частности, из G_{t-1} в G_0). В цепи с матрицей P^t каждый класс G_α соответствует неприводимому замкнутому множеству.*

Теперь, используя эту теорему, легко описать асимптотическое поведение переходных вероятностей $p_{jk}^{(n)}$. Мы знаем, что $p_{jk}^{(n)}\to 0$, если E_k является либо невозвратным, либо возвратным нулевым состоянием, а также что все состояния однотипны (§ 6). Стало быть, нам надо рассмотреть лишь случай, когда каждое состояние E_k имеет конечное среднее время возвращения μ_k. В цепи с матрицей P^t состояние E_k имеет среднее время возвращения μ_k/t, и по отношению к этой цепи каждый класс G_α эргодичен. Поэтому, если E_j принадлежит G_α, то

$$\lim_{n\to\infty} p_{jk}^{(nt)}=\begin{cases} t/\mu_k, & \text{если } E_k\in G_\alpha, \\ 0 & \text{в противном случае,}\end{cases} \tag{9.2}$$

---

[1]) При $t=3$ имеется три класса, и в символической блочной записи, введенной в § 4, матрица P принимает вид

$$\begin{pmatrix} 0 & A & 0 \\ 0 & 0 & B \\ C & 0 & 0 \end{pmatrix},$$

где A—матрица вероятностей переходов из G_0 в G_1, и т. д.

и веса t/μ_k определяют распределение вероятностей на состояниях из класса G_α (см. теорему § 7). Поскольку всего у нас t таких классов, то *числа* $u_k = 1/\mu_k$ *определяют распределение вероятностей на множестве всех состояний*, как это было и в случае непериодических цепей. Мы покажем, что это распределение инвариантно. Для этой цели нам нужны соотношения, соответствующие (9.2), когда показатель степени не делится на период t.

Мы отправимся от фундаментального соотношения

$$p_{jk}^{(nt+\beta)} = \sum_\nu p_{j\nu}^{(\beta)} p_{\nu k}^{(nt)}. \tag{9.3}$$

Множитель $p_{j\nu}^{(\beta)}$ обращается в нуль, за исключением тех случаев когда E_ν принадлежит $G_{\alpha+\beta}$. (Если $\alpha+\beta \geqslant t$, то под $G_{\alpha+\beta}$ следует понимать $G_{\alpha+\beta-t}$.) В этих случаях $p_{\nu k}^{(nt)}$ обращается в нуль, если E_k не лежит в $G_{\alpha+\beta}$, и, следовательно, для фиксированных β и E_j из G_α

$$\lim_{n\to\infty} p_{jk}^{(nt+\beta)} = \begin{cases} t/\mu_k, & \textit{если } E_k \in G_{\alpha+\beta}, \\ 0 & \textit{в противном случае.} \end{cases} \tag{9.4}$$

Перепишем теперь (9.3) в виде

$$p_{ik}^{(nt+1)} = \sum_\nu p_{i\nu}^{(nt)} p_{\nu k}. \tag{9.5}$$

Рассмотрим произвольное состояние E_k; пусть G_ρ — тот класс, которому оно принадлежит. Тогда $p_{\nu k}=0$, если $E_\nu \notin G_{\rho-1}$, и поэтому обе части равенства (9.5) обращаются в нуль, если $E_i \notin G_{\rho-1}$. В этом случае $p_{ik}^{(nt+1)} \to tu_k$, откуда

$$u_k = \sum_\nu u_\nu p_{\nu k}. \tag{9.6}$$

Поскольку E_k — произвольное состояние, то тем самым мы доказали, что *распределение вероятностей* $\{u_k\}$ *инвариантно*.

§ 10. ПРИМЕНЕНИЕ К ТАСОВАНИЮ КАРТ

Колоду из N карт, занумерованных числами $1, 2, \ldots, N$, можно упорядочить $N!$ различными способами, и каждый из них представляет возможное состояние этой системы. Каждая отдельная операция тасования колоды осуществляет переход от имеющегося состояния в некоторое другое. Например, «снимание» заменит порядок $(1, 2, \ldots, N)$ на один из N циклически эквивалентных порядков $(r, r+1, \ldots, N, 1, 2, \ldots, r-1)$. Та же операция, примененная к обратному порядку $(N, N-1, \ldots, 1)$, приведет к состоянию $(N-r+1, N-r, \ldots, 1, N, N-1, \ldots, N-r+2)$. Иначе говоря, каждую отдельную операцию тасования колоды мы представляем себе как преобразование $E_j \to E_k$. Если повторять *в точности* одну и ту же операцию, то система пройдет (начиная с данного состояния E_j) через вполне определенную последовательность состояний, и в результате после конечного числа шагов восстановится первоначальный порядок. С этих пор эта последовательность состояний будет периодически повторяться. Для большинства операций этот период будет довольно мал, и *ни в одном случае* с помощью такой процедуры не могут быть получены все состояния [1]). Например, точная «врезка» сменила бы в колоде из $2m$ карт порядок $(1, \ldots, 2m)$ на $(1, m+1, 2, m+2, \ldots, m, 2m)$. В колоде из 6 карт начальный порядок восстанавливает [illegible] кратное применение этой операции. При десяти картах начальный порядок [illegible]вится вновь после шести операций, так что повторное применение э[illegible]ой операции к колоде из десяти карт может привести только к шести [illegible] $10! = 3\,628\,800$ возможных порядков расположения карт.

На практике игрок может пожелать варьировать опера[illegible] сования, и, конечно, разного рода отклонения будут вы[illegible]цию [illegible] случая. Предположим, что мы можем учесть при[illegible]ваны игрой влияние случайных отклонений, считая, что [illegible]вычки игрока и операция имеет определенную вероятност[illegible] каждая отдельная Ни в каких допущениях относительно [illegible] (возможно, нулевую). роятностей мы не нуждаемся, но [illegible]численных значений этих веигрока не зависят от прошло[illegible]будем предполагать, что действия Отсюда вытекает, что [illegible] испыт[illegible]го и что он не знает порядка карт [2]). [illegible] независим[illegible] последовательные операции соответствуют [illegible]ниям с фиксированными вероятностями; для самой колоды [illegible] карт мы тогда имеем цепь Маркова.

Покаже[illegible]м теперь, что матрица переходных вероятностей P является дважды стохастической (пример 7, з)). Действительно, если операция изменяет состояние (порядок карт) E_j на E_k, то существует другое состояние E_r, которое этой операцией изменяется на E_j. Это означает, что элементы j-го столбца матрицы P совпадают с элементами ее j-й строки с той лишь разницей, что располагаются они там в другом порядке. Поэтому все суммы по столбцам равны единице.

Отсюда следует, что ни одно из состояний не может быть невозвратным. *Если цепь неприводима и непериодична, то в пределе все состояния станут равновероятными.* Иначе говоря, годится *любой* способ тасования, если только он приводит к неприводимой и непериодичной цепи. Можно с уверенностью полагать, что обычно так оно и есть. Предположим, однако, что колода содержит четное число карт и наша процедура состоит в разделении их на две равные

---

[1]) В терминологии теории групп это равносильно утверждению, что группа перестановок не является циклической и поэтому не может порождаться одной операцией.

[2]) Это предположение соответствует обычной ситуации при игре в бридж. Легко придумать более сложную технику тасования, при которой операции будут зависеть от предшествующих действий и окончательный результат не будет цепью Маркова (см. пример 13, д)).

части и в тасовании последних отдельно друг от друга любым методом. Если эти две части складываются затем вместе в их первоначальном порядке, то полученная цепь Маркова будет приводимой (поскольку не из каждого состояния достижимо любое другое состояние). Если же эти части складываются в обратном порядке, то цепь будет иметь период 2. Теоретически могут возникнуть обе эти возможности, но практически они едва ли осуществимы, поскольку из-за влияния случая операции не могут выполняться абсолютно точно.

Итак, вполне можно ожидать, что продолжительное тасование приведет к полной «случайности» и уничтожит все следы первоначального порядка. Следует отметить, однако, что необходимое для этой цели число операций чрезвычайно велико [1]).

§ 11 *). ИНВАРИАНТНЫЕ МЕРЫ. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ДЛЯ ОТНОШЕНИЙ

В этом параграфе мы рассмотрим неприводимую цепь с возвратными нулевыми состояниями. Наша главная цель — вывести аналоги для результатов, полученных в § 7 для цепей, в которых состояния имеют конечные средние времена возвращения. Характерным свойством таких цепей является существование инвариантного (или стационарного) распределения вероятностей, определяемого соотношением

$$u_k = \sum_{\nu} u_\nu p_{\nu k}. \tag{11.1}$$

Мы знаем, что в случае, когда средние времена возвращения бесконечны, такое инвариантное распределение вероятностей не существует, но мы покажем, что линейная система (11.1) все же допускает положительное решение $\{u_k\}$, такое, что $\sum u_k = \infty$. Такое решение $\{u_k\}$ называется *инвариантной (или стационарной) мерой*. Если цепь неприводима и возвратна, то инвариантная мера единственна с точностью до произвольной нормирующей постоянной.

Примеры. а) Предположим, что матрица переходных вероятностей P дважды стохастическая, т. е. суммы и по столбцам, и по стро-

---

[1]) По поводу анализа невероятно плохих результатов тасования в записях экспериментов по экстрасенсорному восприятию см. Feller W., Statistical aspects of ESP, Journal of Parapsychology, 4 (1940), 271—298. Гринвуд и Стюарт в своей поразительной статье (Greenwood J. A., Stuart C. E., A review of Dr. Feller's critique, там же, 299—319) пытаются показать, что эти результаты обусловлены случаем, однако и их вычисления, и их эксперименты имеют определенный оттенок сверхъестественности.

*) Этот и следующий параграфы посвящены вопросам, играющим важную роль в современных исследованиях, однако полученные в них результаты не будут использоваться в настоящей книге.

кам равны единице. Тогда справедливо (11.1) при $u_k=1$ для всех k. Чтобы выразить этот факт, говорят, что *равномерная мера инвариантна*.

б) *Случайные блуждания*. Интересным частным случаем цепи из предыдущего примера является неограниченное случайное блуждание на прямой. Занумеруем состояния в их естественном порядке от $-\infty$ до ∞. Это не позволит нам представить переходные вероятности в обычной форме матрицы, однако необходимые изменения обозначения очевидны. Если переходы в правое и левое соседние состояния имеют вероятности p и q соответственно, то система (11.1) принимает вид

$$u_k=pu_{k-1}+qu_{k+1}, \quad -\infty<k<\infty.$$

Состояния возвратны только при $p=q=1/2$, и в этом случае $u_k=1$ является единственным положительным решением. Это решение сохраняется и при $p\neq q$, однако оно уже не будет единственным; вторым неотрицательным решением будет $u_k=(p/q)^k$. Этот пример доказывает, что инвариантная мера может существовать и у невозвратных цепей, но уже не обязана быть единственной. Мы вернемся к этому интересному вопросу в следующем параграфе.

Инвариантную меру $\{u_j\}$ можно интерпретировать на интуитивном уровне, если рассмотреть одновременно бесконечно много процессов с одной и той же матрицей P переходных вероятностей. Для каждого j определим случайную величину $\mathbf{N}_j$, имеющую распределение Пуассона со средним u_j, и рассмотрим $\mathbf{N}_j$ независимых процессов, начинающихся из E_j. Мы делаем это для всех состояний одновременно, предполагая, что все процессы взаимно независимы. Нетрудно показать, что в каждый данный момент времени в любом заданном состоянии E_k с вероятностью единица может быть лишь конечное число процессов. Стало быть, число процессов, находящихся на n-м шаге в состоянии E_k, является случайной величиной $\mathbf{X}_k^{(n)}$, и *инвариантность* $\{u_k\}$ *влечет равенство* $\mathbf{E}\{\mathbf{X}_k^{(n)}\}=u_k$ *для всех* n (см. задачу 29).

в) В примере 7, е) мы обнаружили, что инвариантное распределение вероятностей существует только тогда, когда сходится ряд (7.21). В случае когда он расходится, (7.20) все еще будет инвариантной мерой, если $u_k\to 0$, а это равносильно тому, что $p_1p_2\ldots$ $\ldots p_k\to 0$. Инвариантной меры не существует, когда произведение $p_1\ldots p_k$ остается отделенным от 0, например когда $p_k=1-(k+1)^{-2}$. В этом случае цепь невозвратна.

г) В примере 7, ж) соотношения (7.23) определяют инвариантную меру даже тогда, когда $\mu=\infty$. ►

В эргодических цепях вероятности $p_{jk}^{(n)}$ стремятся к значению u_k инвариантного распределения вероятностей. Для возвратных нулевых цепей мы докажем слабый вариант этого результата, а именно что для всех E_α и E_β при $N \to \infty$

$$\left(\sum_{n=0}^{N} p_{\alpha i}^{(n)}\right)\left(\sum_{n=0}^{N} p_{\beta j}^{(n)}\right)^{-1} \to u_i/u_j. \tag{11.2}$$

Суммы в левой части представляют собой математические ожидания числа прохождений через E_i и E_j за первые N испытаний. Грубо говоря, (11.2) утверждает, что эти математические ожидания асимптотически независимы от начальных состояний E_α и E_β и находятся в той же пропорции, что и соответствующие значения инвариантных мер. Таким образом, характерные факты здесь те же, что и в случае эргодических цепей, хотя ситуация и более запутана. С другой стороны, периодические цепи не требуют теперь специального рассмотрения. (В действительности (11.2) справедливо для *всех* [неприводимых.— *Перев.*] возвратных цепей. Для эргодической цепи числитель в левой части равен $\sim Nu_i$.)

Соотношения вида (11.2) называются *предельными теоремами для отношений*. Мы выведем (11.2) из более сильного результата, который до недавнего времени рассматривался как более сложное уточнение. Наши доказательства основываются на рассмотрении только тех путей, которые не проходят через некоторое выделенное состояние E_r. Следуя Чжун Кайлаю, мы назовем это запрещенное состояние *табу*, а вероятности перехода в него — *табу-вероятностями*.

Определение. *Пусть E_r — произвольное фиксированное состояние. Для $E_k \neq E_r$ и $n \geq 1$ мы обозначим через ${}_rp_{jk}^{(n)}$ вероятность того, что начинающийся из E_j процесс попадет на n-м шаге в состояние E_k, не заходя по пути в E_r.*

Здесь E_j может совпадать с E_r. Мы распространим это определение естественным образом на случаи $E_k = E_r$ и $n = 0$, полагая

$${}_rp_{jr}^{(n)} = 0, \quad n \geq 1, \tag{11.3}$$

и

$${}_rp_{jk}^{(0)} = \begin{cases} 1, & \text{если } E_j = E_k, \\ 0 & \text{в противном случае.} \end{cases} \tag{11.4}$$

В аналитической записи мы имеем для $n \geq 0$ и $E_k \neq E_r$

$${}_rp_{jk}^{(n+1)} = \sum_\nu {}_rp_{j\nu}^{(n)} p_{\nu k}. \tag{11.5}$$

В действительности при $n=0$ сумма в правой части сводится к одному члену, а именно к p_{jk}. При $n \geq 1$ соответствующий $\nu = r$ член обращается в нуль в силу (11.3), так что (11.5) эквивалентно первоначальному определению.

Введение табу на состояние E_r равносильно рассмотрению исходного марковского процесса лишь до первого достижения состояния E_r. В неприводимой возвратной цепи состояние E_r достигается с вероятностью единица из любого начального состояния E_j. Отсюда следует, что в цепи с табу E_r последовательные прохождения через начальное состояние E_j составляют *невозвратное* рекуррентное событие, а прохождения через любое другое состояние $E_k \neq E_r$ составляют невозвратное рекуррентное событие с запаздыванием. Поэтому в силу основной теоремы 2 гл. XIII, 3 для $E_k \neq E_r$

$$\sum_{n=0}^{\infty} {}_r p_{jk}^{(n)} = {}_r\pi_{jk} < \infty. \tag{11.6}$$

При $E_k = E_r$ слагаемые с $n \geqslant 1$ обращаются в нуль и сумма сводится к 1 при $j=r$ и к 0 при $j \neq r$.

Теперь мы в состоянии доказать существование инвариантной меры, т. е. чисел u_k, удовлетворяющих (11.1). В доказательстве теоремы 2 это использоваться не будет.

Теорема 1. *Если цепь неприводима и возвратна, то последовательность чисел*

$$u_k = {}_r\pi_{rk} \tag{11.7}$$

представляет собой инвариантную меру; более того, $u_k > 0$ для всех k, а $u_r = 1$.

Обратно, если $u_k \geqslant 0$ при всех k и справедливо (11.1), *то существует постоянная λ, такая, что $u_k = \lambda \cdot {}_r\pi_{rk}$.*

Состояние E_r здесь произвольно, однако из единственности вытекает, что получаемые при разных r последовательности $\{u_k\}$ отличаются лишь множителями пропорциональности. Заметим, что теорема и ее доказательство охватывают также случай цепей с конечными средними временами возвращения.

Доказательство. Если $k \neq r$, то, суммируя (11.5) при $j=r$ по $n=0, 1, \ldots$, получаем

$${}_r\pi_{rk} = \sum_{\nu} {}_r\pi_{r\nu} p_{\nu k}, \tag{11.8}$$

так что числа (11.5) удовлетворяют уравнениям (11.1) хотя бы при $k \neq r$. При $j=k=r$ ясно, что

$$\sum_{\nu} {}_r p_{r\nu}^{(n)} p_{\nu r} = f_{rr}^{(n+1)} \tag{11.9}$$

есть вероятность того, что (в исходной цепи) первое возвращение в E_r произойдет на $(n+1)$-м шаге. Поскольку цепь неприводима и возвратна, в сумме эти вероятности составляют единицу. Поэтому, суммируя (11.9) по $n=0, 1, \ldots$, мы получаем

$$\sum_{\nu} {}_r\pi_{r\nu} p_{\nu r} = 1. \tag{11.10}$$

Но по определению ${}_r\pi_{rr}=1$, так что (11.8) справедливо и при $k=r$. Следовательно, (11.7) действительно представляет собой инвариантную меру.

Рассмотрим, далее, произвольную неотрицательную инвариантную меру $\{u_k\}$. Из определения (11.1) ясно, что если $u_k=0$ при некотором k, то $u_\nu=0$ для всех ν, таких, что $p_{\nu k}>0$. По индукции отсюда следует, что $u_\nu=0$ при каждом ν, таком, что E_k достижимо из E_ν. Поскольку цепь неприводима, отсюда вытекает, что $u_\nu=0$ для всех ν. Поэтому инвариантная мера строго положительна (или тождественно равна нулю).

Стало быть, в обратном утверждении теоремы мы можем считать, что заданная инвариантная мера нормирована условием $u_r=1$ для некоторого заранее выбранного r. Тогда

$$u_k=p_{kr}+\sum_{j\neq r}u_jp_{jk}. \tag{11.11}$$

Предположим, что $k\neq r$. Подставим под знаком суммы вместо чисел u_j соответствующие им выражения из соотношения (11.1) и в полученной двойной сумме снова выделим содержащий u_r член. В результате имеем

$$u_k=p_{rk}+{}_rp_{rk}^{(2)}+\sum_{\nu\neq r}u_\nu\cdot{}_rp_{\nu k}^{(2)}. \tag{11.12}$$

Продолжая в том же духе, мы получим для каждого n

$$u_k=p_{rk}+{}_rp_{rk}^{(2)}+\ldots+{}_rp_{rk}^{(n)}+\sum_{\nu\neq r}u_\nu\cdot{}_rp_{\nu k}^{(n)}. \tag{11.13}$$

Устремляя n к ∞, мы видим, что $u_k\geqslant{}_r\pi_{rk}$. Отсюда следует, что $\{u_k-{}_r\pi_{rk}\}$ определяет [неотрицательную.— *Перев.*] инвариантную меру, обращающуюся в нуль при $k=r$. Но такая мера равна нулю тождественно, и тем самым справедливо (11.7). ►

Вскоре мы увидим, что уточнением предельной теоремы для отношений является следующая теорема.

Теорема 2. *В неприводимой возвратной цепи для всех* N

$$0\leqslant\sum_{n=0}^{N}p_{kk}^{(n)}-\sum_{n=0}^{N}p_{\alpha k}^{(n)}\leqslant{}_\alpha\pi_{kk} \tag{11.14}$$

и

$$-1\leqslant(1/{}_j\pi_{ii})\sum_{n=0}^{N}p_{ii}^{(n)}-(1/{}_i\pi_{jj})\sum_{n=0}^{N}p_{jj}^{(n)}\leqslant 1. \tag{11.15}$$

Доказательство (11.14). Рассмотрим первое попадание в E_k; ясно, что при $\alpha\neq k$

$$p_{\alpha k}^{(n)}=\sum_{\nu=1}^{n}f_{\alpha k}^{(\nu)}p_{kk}^{(n-\nu)}. \tag{11.16}$$

(Это то же самое, что (5.3).) Суммируя по n, получаем

$$\sum_{n=0}^{N} p_{\alpha k}^{(n)} \leqslant \sum_{n=0}^{N} p_{kk}^{(n)} \cdot \sum_{\nu=1}^{\infty} f_{\alpha k}^{(\nu)} = \sum_{n=0}^{N} p_{kk}^{(n)}, \tag{11.17}$$

что доказывает первое неравенство в (11.14).

Заметим теперь, что при выходе из E_k возвращение в E_k [на n-м шаге.— *Перев.*] произойдет либо без промежуточного прохождения через E_α, либо первое попадание в E_α произойдет на ν-м шаге, $1 \leqslant \nu \leqslant n$. Это означает, что

$$p_{kk}^{(n)} = {}_\alpha p_{kk}^{(n)} + \sum_{\nu=1}^{n} f_{k\alpha}^{(\nu)} p_{\alpha k}^{(n-\nu)}. \tag{11.18}$$

Суммирование по n приводит ко второму неравенству в (11.14).

Доказательство (11.15). В силу очевидной симметрии i и j достаточно доказать второе неравенство. Мы отправимся от тождества

$$p_{ii}^{(n)} = {}_j p_{ii}^{(n)} + \sum_{\nu=1}^{n-1} p_{ij}^{(n-\nu)} \cdot {}_j p_{ji}^{(\nu)}, \tag{11.19}$$

которое выражает тот факт, что возвращение из E_i в E_i происходит либо без промежуточного прохождения через E_j, либо *последнее* попадание в E_j имеет место на $(n-\nu)$-м шаге и следующие ν шагов приводят из E_j в E_i без возвращения в E_j. Суммируя по n, в силу (11.14) получаем

$$\sum_{n=0}^{N} p_{ii}^{(n)} \leqslant {}_j\pi_{ii} + {}_j\pi_{ji} \sum_{n=0}^{N} p_{ij}^{(n)} \leqslant {}_j\pi_{ii} + {}_j\pi_{ji} \sum_{n=0}^{N} p_{jj}^{(n)}. \tag{11.20}$$

Чтобы привести это неравенство к симметричному виду (11.15), достаточно заметить, что

$$ {}_j\pi_{ji} = {}_j\pi_{ii} / {}_i\pi_{jj}. \tag{11.21}$$

Действительно, по аналогии с (11.16)

$$ {}_j\pi_{ji} = {}_j f_{ji} \cdot {}_j\pi_{ii}, \tag{11.22}$$

где ${}_j f_{ji}$ — вероятность достижения E_i из E_j без предварительного возвращения в E_j. Альтернативой для этого события является возвращение в E_j до достижения E_i, и, следовательно,

$$ {}_j f_{ji} = 1 - {}_i f_{jj} = 1 / {}_i\pi_{jj}. \tag{11.23}$$

(Последнее уравнение является основным тождеством для невозвратного рекуррентного события, состоящего в возвращении в E_j без промежуточного прохождения через E_i.) Подставляя (11.23) в (11.22), получаем утверждение (11.21), и доказательство теоремы этим завершается. ►

Из соотношения (11.21) вытекает интересное следствие.

Следствие 1. *Если* $\{u_k\}$ — *инвариантная мера, то*

$$_j\pi_{ii}/_i\pi_{jj}=u_i/u_j. \qquad (11.24)$$

Доказательство. Инвариантная мера определяется с точностью до постоянного множителя, и поэтому правая часть в (11.24) определена единственным образом. Стало быть, мы можем предположить, что $\{u_k\}$ есть инвариантная мера, определенная в (11.7), когда запрещенное состояние E_r отождествляется с E_j. Но тогда $u_j=1$ и $_j\pi_{ji}=u_i$, и поэтому (11.21) сводится к (11.24). ▶

Следствие 2. (Предельная теорема для отношений.) *Для неприводимой возвратной цепи справедливо предельное соотношение* (11.2).

Доказательство. Суммы из теоремы 2 стремятся к ∞ при $N\to\infty$. Стало быть, отношение двух сумм в (11.14) стремится к единице, и поэтому достаточно доказать (11.2) для частного случая $\alpha=i$ и $\beta=j$. Однако при таком выборе (11.2) является непосредственным следствием (11.15) и (11.24). ▶

Существование инвариантной меры для возвратной цепи впервые доказал Дерман [1]). Существование предела в (11.2) установил Дёблин [2]). Табу-вероятности в качестве мощного инструмента теории цепей Маркова ввел Чжун Кайлай [3]); более детально они изучаются в его фундаментальном труде [4]). Ограниченность частных сумм $\sum\limits_{n=0}^{N}\left(p_{kk}^{(n)}-p_{ii}^{(n)}\right)$ доказал Орей, рассматривавший также вопрос о сходимости [5]).

§ 12 *). ОБРАЩЕННЫЕ ЦЕПИ. ГРАНИЦЫ

При изучении эволюции некоторой системы мы обычно интересуемся вероятностями возможных будущих событий, однако иногда бывает необходимо изучать и прошлое. В специальном случае цепи Маркова мы можем поинтересоваться, чему равна (условная) вероятность того, что в определенный момент в прошлом система

---

[1]) Derman C., A solution to a set of fundamental equations in Markov chains, Proc. Amer. Math. Soc., 5 (1954), 332—334.

[2]) Doeblin W., Sur deux problèmes de A. Kolmogoroff concernant les chaines dénombrables, Bull. Soc. Math. de France, 66 (1938), 1—11.

[3]) Chung K. L., Contribution to the theory of Markov chains, I, Journ. Research, Nat. Bureau of Standards, 50 (1953), 203—208; II, Trans. Amer. Math. Soc., 76 (1954), 397—419.

[4]) Chung K. L., Markov chains with stationary transition probabilities, Berlin, Springer, 1960. [Имеется перевод: Чжун Кай-лай. Однородные цепи Маркова.— М.: Мир, 1964.]

[5]) Orey S., Sums arising in the theory of Markov chains, Proc. Amer. Math. Soc., 12 (1961), 847—856.

*) См. примечание на с. 423.

была в состоянии E_i при условии, что в настоящем ее состоянием является E_j.

Рассмотрим сперва цепь, имеющую строго положительное инвариантное распределение вероятностей $\{u_k\}$, т. е. предположим, что $u_k>0$ и $\sum u_k=1$, где

$$u_k=\sum_{\nu} u_\nu p_{\nu k}. \tag{12.1}$$

(Напомним, что по теореме из § 7 инвариантное распределение вероятностей в неприводимой цепи автоматически будет строго положительным.)

Если начальное распределение совпадает с $\{u_k\}$, то в любой момент времени вероятность того, что система находится в е… ой нии E_i, будет равна u_i. При условии, что это событие ос… остоялось, условная вероятность того, что n единиц времени … назад система была в состоянии E_j, равна

$$q_{ij}^{(n)}=u_j p_{ji}^{(n)}/u_i. \tag{12.2}$$

При $n=1$ получаем

$$q_{ij}=u_j p_{ji}/u_i. \tag{12.3}$$

В силу (12.1) ясно, что q_{ij} являются элементами *стохастической* матрицы Q. Более того, вероятности $q_{ij}^{(n)}$ являются просто элементами n-й степени Q^n этой матрицы (иначе говоря, $q_{ij}^{(n)}$ можно вычислить по q_{ij} точно так же, как $p_{ji}^{(n)}$ вычисляются по p_{ji}). Теперь очевидно, что *изучение прошлой эволюции нашей цепи Маркова сводится к изучению цепи Маркова с переходными вероятностями* q_{ij}. Безусловные вероятности для новой цепи совпадают, конечно же, с инвариантным распределением вероятностей $\{u_k\}$. Вероятности q_{ij} называются *обращенными вероятностями* (по отношению к исходной цепи), и процедура, приводящая от одной цепи к другой, называется *обращением времени*. В частном случае, когда $q_{ij}=p_{ij}$, говорят, что цепь *обратима*; вероятностные соотношения для такой цепи симметричны по времени.

Мы знаем, что неприводимая цепь обладает инвариантным распределением вероятностей только тогда, когда ее состояния имеют конечные средние времена возвращения. Если состояния цепи возвратные нулевые, то существует инвариантная мера, единственная с точностью до постоянного множителя. Для невозвратных цепей возможны все варианты: некоторые цепи не имеют инвариантных мер, другие имеют их бесконечно много (см. примеры 11,б) и 11,в)). При этих обстоятельствах заслуживает внимания тот факт, что преобразование (12.3) *определяет стохастическую матрицу Q, если $\{u_k\}$ является строго положительной инвариантной мерой.* Степени матрицы Q даются формулой (12.2). В этом смысле *каждая строго положительная инвариантная мера определяет обращенную цепь Маркова.* К сожалению, новые переходные вероятности q_{ij}

нельзя интерпретировать непосредственно как условные вероятности для исходного процесса [1]).

Из (12.3) ясно, что $\{u_j\}$ является инвариантной мерой и для обращенной цепи. Более того, из (12.2) следует, что ряды $\sum_n q_{ii}^{(n)}$ и $\sum_n p_{ii}^{(n)}$ оба сходятся или оба расходятся. Отсюда следует, что *состояния обеих цепей имеют один тип*: если одна из цепей невозвратна, либо возвратна, то такова же и другая.

Примеры. а) Инвариантное распределение вероятностей, соответствующее *модели Эренфестов* (пример 2, д)), было найдено в (7.16). Простое вычисление показывает, что *модель Эренфестов обратима* в том смысле, что $q_{ij}=p_{ij}$.

б) В примере 11,б) мы нашли инвариантные меры, соответствующие случайному блужданию на прямой, в котором переходы в правую и левую соседние точки имеют вероятности p и q соответственно. Если мы выберем меру с $u_k=1$ при $k=0, \pm1, \pm2, \ldots$, то получим $q_{ij}=p_{ji}$ и тем самым придем к случайному блужданию, в котором p и q поменялись ролями. С другой стороны, инвариантная мера с $u_k=(p/q)^k$ приводит к обращенному блужданию, совпадающему с исходным.

в) В примерах 2, л) и 2, м) мы ввели две цепи Маркова, связанные с рекуррентным событием $\mathscr{E}$. В примере 7,ж) мы видели, что для возвратного $\mathscr{E}$ с конечным средним временем возвращения μ эти две цепи имеют одно и то же инвариантное распределение вероятностей, определенное в (7.23). При $\mu=\infty$ эти соотношения определяют общую для обеих цепей инвариантную меру (см. примеры 11,в) и 11,г)). Простое вычисление показывает теперь, что *эти две цепи получаются одна из другой при обращении времени*. Это и не удивительно, поскольку цепь из примера 2,л) связана со временем ожидания следующего осуществления события $\mathscr{E}$, тогда как пример 2,м) относится ко времени, прошедшему с момента последнего осуществления этого события.

Рассмотрим теперь произвольную неприводимую невозвратную цепь с инвариантной мерой $\{u_k\}$. Определяющие инвариантную меру уравнения (12.1) могут допускать и другие решения, и вопрос о единственности для этой системы тесно связан с вопросом о единственности для сопряженной системы линейных уравнений [2])

$$\xi_i=\sum_\nu p_{i\nu}\xi_\nu, \tag{12.4}$$

[1]) Для конкретной интерпретации q_{ij} необходимо рассмотреть бесконечно много одновременно протекающих процессов, как было указано в примере 11, б).

[2]) Если через ξ обозначить вектор-столбец с компонентами ξ_i, то система (12.4) будет эквивалентна векторно-матричному уравнению $\xi=P\xi$. Система (12.1) соответствует уравнению $u=uP$, где u — вектор-строка.

сыгравшей важную роль в § 8. Эта система допускает тривиальное решение $\xi_i = c$ при всех i. Каждое неотрицательное решение автоматически будет строго положительным. (В самом деле, из $\xi_i = 0$ вытекало бы, что $\xi_\nu = 0$ при всех ν, таких, что $p_{i\nu} > 0$. Это, в свою очередь, означало бы, что $\xi_\nu = 0$ при $p_{i\nu}^{(2)} > 0$, и вообще $\xi_\nu = 0$ для каждого достижимого из E_i состояния E_ν. Поэтому $\xi_\nu = 0$ для всех ν, ибо цепь неприводима.) Если $\{\xi_i\}$ — непостоянное решение (12.4), то (12.3) показывает, что

$$v_i = u_i \xi_i \tag{12.5}$$

определяет инвариантную меру для обращенной цепи с матрицей Q. И обратно, если обозначить такую меру через $\{v_i\}$, то (12.5) будет определять положительное решение системы (12.4). Иначе говоря, *положительные решения* (12.4) *состоят во взаимно однозначном соответствии с инвариантными мерами для обращенной цепи*[1]) *с матрицей* Q.

В современной теории цепей Маркова и потенциалов положительные решения $\{\xi_i\}$ и $\{u_k\}$ используются для определения границ. Описание того, как это делается, выходит за рамки нашей книги, однако следующие примеры могут дать некоторое понятие о том, что называется *границей-выходом* (exit boundary).

Примеры. а) Рассмотрим случайное блуждание на бесконечной прямой, такое, что из состояния $j \neq 0$ частица с вероятностью p перемещается на единичный шаг в направлении от начала координат и с вероятностью q — на единичный шаг к началу координат. Из нуля частица перемещается с равными вероятностями в $+1$ или в -1. Мы предположим, что $p > q$.

В этой цепи Маркова состояния занумерованы числами от $-\infty$ до $+\infty$ и уравнения (12.4) принимают вид

$$\begin{aligned} \xi_i &= p\xi_{i+1} + q\xi_{i-1}, \qquad & i > 0, \\ \xi_0 &= (1/2)\,\xi_1 + (1/2)\,\xi_{-1}, & \\ \xi_i &= q\xi_{i+1} + p\xi_{i-1}, & i < 0. \end{aligned} \tag{12.6}$$

---

[1]) Для неприводимой возвратной цепи инвариантная мера единственна с точностью до постоянного множителя. Поскольку цепи с матрицами P и Q однотипны, то нами доказана следующая теорема.

Теорема. *Для неприводимой возвратной цепи единственным неотрицательным решением системы* (12.4) *является* $\xi_i = \text{const}$.

Это можно доказать также почти дословным повторением последней части доказательства теоремы 1 § 11. В самом деле, по индукции мы находим, что для произвольных i, r и n

$$\xi_i = \left[f_{ir}^{(1)} + \ldots + f_{ir}^{(n)}\right]\xi_r + \sum_\nu {}_r p_{i\nu}^{(n)}\,\xi_\nu.$$

Для возвратной цепи выражение внутри скобок стремится к 1, тогда как ряд сходится к 0. Следовательно, $\xi_i = \xi_r$, как и утверждалось.

Положим

$$\eta_i=\begin{cases}1-(1/2)\,(q/p)^i & \text{при } i\geqslant 0,\\ (1/2)\,(q/p)^{-i} & \text{при } i\leqslant 0.\end{cases} \tag{12.7}$$

Легко видеть, что $\xi_i=\eta_i$ и $\xi_i=1-\eta_i$ определяют два[1]) нетривиальных решения системы (12.6). Отсюда следует, что наша цепь невозвратна, и поэтому положение частицы с необходимостью устремится либо к $+\infty$, либо к $-\infty$.

К этому заключению можно прийти и непосредственно из теории случайных блужданий. В самом деле, из гл. XIV, 2 мы знаем, что если частица выходит из положения $i>0$, то вероятность того, что она когда-либо достигнет нуля, равна $(q/p)^i$. В силу симметрии выходящая из нуля частица имеет равные вероятности сноса в $+\infty$ и в $-\infty$, и поэтому вероятность окончательного сноса в частицы в $-\infty$ равна $(1/2(q/p)^i$. Мы заключаем, что η_i *есть вероятность того, что выходящую из произвольного положения i частицу в конце концов снесет в* $+\infty$. Снос в $-\infty$ имеет вероятность $1-\eta_i$. В современной теории эта ситуация была бы описана посредством введения «точек границы-выхода» $+\infty$ и $-\infty$.

б) Простота предыдущего примера несколько обманчива, и поэтому может быть полезно иметь пример *границы, состоящей из бесконечного числа точек*. Рассмотрим для этой цели следующее случайное блуждание в плоскости x, y. Координата x совершает обычное случайное блуждание, в котором шаги $+1$ и -1 имеют вероятности p и $q<p$. Координата y остается фиксированной, за исключением тех моментов, когда координата x обращается в нуль, и в эти моменты координата y уменьшается на 1. Точнее когда $j\neq 0$, возможны только переходы $(j, k)\to(j+1, k)$ и $(j-1, k)$, и они имеют вероятности p и $q<p$ соответственно. Из положения $(0, k)$ частица с вероятностью p перемещается в $(1, k-1)$ и с вероятностью q — в $(-1, k-1)$.

Из теории случайных блужданий мы знаем, что координата x обязана стремиться к $+\infty$ и что (с вероятностью единица) она лишь конечное число раз пройдет через 0. Отсюда следует, что (исключая событие нулевой вероятности) координата y изменится лишь конечное число раз. Это означает, что после конечного числа изменений координаты y частица навсегда останется на прямой $y=r$. В этом смысле существует бесконечно много «путей бегства в бесконечность», и для каждого начального положения (j, k) мы можем вычислить[2]) вероятность $\xi_{j,k}^{(r)}$ того, что частица в конце концов

[1]) Наиболее общее решение системы (12.6) дается формулой $\xi_i=A+B\eta_i$, где A и B — произвольные постоянные. Действительно, эти постоянные могут быть выбраны так, чтобы получить заданные значения ξ_1 и ξ_{-1}, а из (12.6) очевидно, что значения ξ_1 и ξ_{-1} определяют все ξ_i единственным образом.

[2]) Явное выражение для $\xi_{j,k}^{(r)}$ может быть получено из относящихся к одномерным случайным блужданиям результатов гл. XIV,2. Для начального положения $i\leqslant 0$ вероятность того, что частица побывает в нуле ровно $\rho>0$

обоснуется на прямой $y=r$. Легко видеть, что для фиксированного r вероятности $\xi_{j,k}^{(r)}$ представляют собой решение системы, соответствующей (12.4), и что наиболее общее решение такой системы является линейной комбинацией этих частных решений. Более того, частное решение $\xi_{j,k}^{(r)}$ характеризуется тем интуитивно очевидным «граничным условием», что $\xi_{j,k}^{(r)} \to 0$ при $j \to \infty$, за исключением случая $k=r$, в котором $\xi_{j,r}^{(r)} \to 1$. ▶

Эти примеры типичны в следующем смысле. Для данной неприводимой невозвратной цепи Маркова всегда можно определить «границу», такую, что с вероятностью единица состояние системы стремится к некоторой точке этой границы. Если нам дано множество Γ на границе, то мы можем поинтересоваться, какова вероятность η_i того, что, выходя из начального состояния E_i, система устремится к точке из Γ. Назовем $\{\eta_i\}$ *вероятностями поглощения в* Γ. Оказывается, что такие вероятности поглощения всегда являются решениями линейной системы (12.4), и, наоборот, все ограниченные решения (12.4) представляют собой линейные комбинации вероятностей поглощения. Более того, вероятности $\{\eta_i\}$ поглощения в Γ даются единственным решением системы (12.4), которое имеет граничное значение 1 на Γ и граничное значение 0 на дополнении Γ до границы. Мы теперь можем образовать новую стохастическую матрицу $\hat{P}$ с элементами

$$\hat{p}_{ik} = p_{ik}\eta_k/\eta_i. \tag{12.8}$$

Это условная вероятность перехода из E_i в E_k при условии, что система в конечном счете устремится к одной из точек Γ. Марковский процесс с матрицей $\hat{P}$ может быть описан как условный процесс, получающийся из исходного процесса при условии окончательного поглощения в Γ. Поскольку будущая эволюция никогда не может быть известна заранее, такой переход на первый взгляд кажется бессмысленным. Тем не менее он является мощным аналитическим инструментом и даже имеет реальный смысл для процессов, которые протекают уже очень длительное время.

Граница может быть определена также и для матрицы Q, получаемой при обращении времени. Стало быть, в общем случае существуют две различные границы, соответствующие данной цепи. Они

раз, равна $(2q)^{\rho-1}(p-q)$; при $i \geqslant 0$ эта вероятность равна $(q/p)^i(2q)^{\rho-1}(p-q)$. Вероятность никогда не попасть в нуль равна 0 при $i \leqslant 0$ и $1-(q/p)^i$ при $i \geqslant 0$. Отсюда легко получить, что при $i \leqslant 0$

$$\xi_{i,k}^{(r)} = (2q)^{k-r-1}(p-q), \quad k > r,$$

тогда как при $i > 0$

$$\xi_{i,k}^{(r)} = (q/p)^i (2q)^{k-r-1}(p-q), \quad k > r,$$
$$\xi_{i,r}^{(r)} = 1-(q/p)^i$$

и, конечно же, $\xi_{i,r}^{(k)} = 0$ при $k < r$.

называются соответственно *гра*[illegible]*ница-вход* (entrance boundary) и *граница-выход*. Грубо говоря, [illegible] первая из них [illegible] прошлому, а вторая [illegible] к [illegible] относится к отдаленному [illegible] отдаленному будущему.

Обращенные [illegible] системы (12 [illegible] цепи впервые рассмотрел А. Н. Колмогоров [1]). Роль решений вход и [illegible] была подчеркнута в предыдущих изданиях этой книги. Границу-[illegible] границу-выход ввел В. Феллер [2]). Предложенное им построение удовлетворительно, когда существует лишь конечное число граничных точек, а в общем случае, как указал Дж. Л. Дуб [3]), проще принять построение, введенное в теории гармонических функций Р. С. Мартином. Величины (12.8) были введены В. Феллером [2]); аналогичное преобразование в теории классических гармонических функций было определено примерно в то же время М. Брело [4]).

§ 13. ОБЩИЙ МАРКОВСКИЙ ПРОЦЕСС

В приложениях обычно бывает удобно описывать цепи Маркова в терминах случайных величин. Это может быть сделано с помощью простой замены в предыдущих параграфах символа E_k на целое число k. Состояние системы в момент времени n будет тогда случайной величиной $\mathbf{X}^{(n)}$, принимающей значение k с вероятностью $a_k^{(n)}$; совместное распределение $\mathbf{X}^{(n)}$ и $\mathbf{X}^{(n+1)}$ дается формулой $\mathbf{P}\{\mathbf{X}^{(n)}=j, \mathbf{X}^{(n+1)}=k\}=a_j^{(n)}p_{jk}$, а совместное распределение $(\mathbf{X}^{(0)}, \ldots, \mathbf{X}^{(n)})$ — формулой (1.1). Можно также (иногда это предпочтительнее) приписывать состоянию E_k численное значение e_k, отличное от k. При такой записи цепь Маркова становится специальным стохастическим процессом [5]), или, иначе говоря, последовательностью (зависимых) случайных величин [6]) $(\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \ldots)$. Верхний индекс n играет роль времени. В гл. XVII мы получим некоторое представление о более общих стохастических процессах, в которых временно́й параметр может меняться непрерывно. Термин «марковский процесс» применяется к весьма обширному и важному классу стохастических процессов (как с дискретным, так и с непрерывным временем). Даже в дискретном случае существуют более общие марковские процессы, чем те простые цепи, которые изучались нами до сих пор. Поэтому будет полезно дать определение марковского свойства, указать спе-

---

[1]) Kolmogoroff A., Zur Theorie der Markoffschen Ketten, Mathematische Annalen, 112 (1935), 155—160.

[2]) Feller W., Boundaries induced by positive matrices, Trans. Amer. Math. Soc., 83 (1956), 19—54.

[3]) Doob J. L., Discrete potential theory and boundaries, J. Math. Mechanics, 8 (1959), 433—458.

[4]) Brelot M., Le problème de Dirichlet. Axiomatique et frontière de Martin, J. Math. Pures Appl., 35 (1956), 297—335.

[5]) Термины «стохастический процесс» и «случайный процесс» являются синонимами и охватывают практически всю теорию вероятностей от бросания монеты до гармонического анализа. На практике термин «стохастический процесс» используется главным образом тогда, когда рассматривается случай непрерывного времени.

[6]) Эта формулировка относится к бесконечному произведению пространств, но в действительности мы имеем дело лишь с совместными распределениями конечных наборов случайных величин.

циальное условие, характеризующее наши цепи Маркова, и, наконец, привести несколько примеров немарковских процессов.

По идее марковский процесс является вероятностным аналогом процессов классической механики, где будущая эволюция полностью определяется состоянием в настоящий момент и не зависит от того, каким образом это состояние было достигнуто. Эти процессы существенно отличаются от процессов с последействием (или с наследственностью [1])), таких, которые встречаются, например, в теории пластичности, где вся предыдущая история системы влияет на ее будущее. В стохастических процессах будущее не определяется единственным образом, однако у нас по крайней мере имеются вероятностные соотношения, позволяющие нам делать предсказания. Для изучавшихся в этой главе цепей Маркова ясно, что вероятностные соотношения, относящиеся к будущему, зависят лишь от состояния в настоящий момент, но не от того, каким образом это состояние возникло в прошлом. Иначе говоря, если две независимые системы с одинаковыми переходными вероятностями оказываются в одном и том же состоянии, то все вероятности, относящиеся к их будущей эволюции, совпадают. Это довольно неопределенное описание формализуется следующим определением.

Определение. *Последовательность дискретных случайных величин является марковским процессом, если для каждого конечного набора целых чисел* $n_1 < n_2 < \ldots < n_r < n$ *соответствующее совместное распределение* $(\mathbf{X}^{(n_1)}, \mathbf{X}^{(n_2)}, \ldots, \mathbf{X}^{(n_r)}, \mathbf{X}^{(n)})$ *определено таким образом, что условная вероятность соотношения* $\mathbf{X}^{(n)} = x$ *при условии* $\mathbf{X}^{(n_1)} = x_1, \ldots, \mathbf{X}^{(n_r)} = x_r$ *совпадает с условной вероятностью события* $\mathbf{X}^{(n)} = x$ *при единственном условии* $\mathbf{X}^{(n_r)} = x_r$. *Здесь* $x_1, \ldots, x_r$ *суть произвольные числа, для которых упомянутые условия имеют положительные вероятности.*

Проще говоря, в определении сказано, что если дано настоящее состояние x_r, то никакие дополнительные сведения относительно состояний системы в прошлом не могут изменить (условную) вероятность состояния x в некоторый будущий момент времени.

Изучавшиеся до сих пор в этой главе цепи Маркова являются, очевидно, марковскими процессами, однако они обладают тем дополнительным свойством, что для них *переходные вероятности* $p_{jk} = \mathbf{P}\{\mathbf{X}^{(m+1)} = k \mid \mathbf{X}^{(m)} = j\}$ *не зависят от* m. Более общие переходные вероятности

$$p_{jk}^{(n-m)} = \mathbf{P}\{\mathbf{X}^{(n)} = k \mid \mathbf{X}^{(m)} = j\}, \quad m < n, \tag{13.1}$$

зависят только от разности $n - m$. Такие переходные вероятности называются *стационарными (или однородными по времени)*. Для общей целочисленной цепи Маркова правая часть (13.1) зависит от m и n. Мы обозначим ее через $p_{jk}(m, n)$, так что $p_{jk}(n, n+1)$ опреде-

---

[1]) В оригинале hereditary processes.— *Прим. перев.*

ляют вероятности перехода за один шаг. Для вероятности пути $(j_0, j_1, \ldots, j_n)$ мы получаем теперь вместо (1.1) выражение

$$a_{j_0}^{(0)} p_{j_0 j_1}(0, 1)\, p_{j_1 j_2}(1, 2) \ldots p_{j_{n-1} j_n}(n-1, n). \qquad (13.2)$$

Соответствующим обобщением (3.3) будет, очевидно, тождество

$$p_{jk}(m, n) = \sum_{\nu} p_{j\nu}(m, r)\, p_{\nu k}(r, n), \qquad (13.3)$$

справедливое для всех r, таких, что $m<r<n$. Это тождество следует непосредственно из определения марковского процесса, а также из (13.2); оно называется *уравнением Колмогорова — Чэпмена*. (Переходные вероятности $p_{jk}(m, n)$ определены и для немарковских дискретных процессов, однако для них множитель $p_{\nu k}(r, n)$ в (13.3) должен быть заменен выражением, зависящим не только от ν и k, но и от j.)

Изучавшиеся в этой главе цепи Маркова представляют общий однородный по времени дискретный марковский процесс. Мы не будем останавливаться на неоднородных марковских процессах. Следующие примеры могут быть полезными для понимания марковского свойства и проиллюстрируют ситуации, в которых уравнение Колмогорова — Чэпмена (13.3) не имеет места.

Примеры немарковских процессов.

а) *Урновая схема Пойа* (пример гл. V,2, в)). Пусть $\mathbf{X}^{(n)}$ равно 1 или 0 в зависимости от того, какой шар будет вынут при n-м извлечении—черный или красный. Последовательность $\{\mathbf{X}^{(n)}\}$ *не является* марковским процессом. Например,

$$\mathbf{P}\{\mathbf{X}^{(3)} = 1 \mid \mathbf{X}^{(2)} = 1\} = (b+c)/(b+r+c),$$

но

$$\mathbf{P}\{\mathbf{X}^{(3)} = 1 \mid \mathbf{X}^{(2)} = 1,\ \mathbf{X}^{(1)} = 1\} = (b+2c)/(b+r+2c)$$

(см. задачи 19 и 20 гл. V). С другой стороны, если $\mathbf{Y}^{(n)}$—число черных шаров в урне в момент времени n, то $\{\mathbf{Y}^{(n)}\}$ является обычной цепью Маркова с постоянными переходными вероятностями.

б) *Суммы высших порядков.* Пусть $\mathbf{Y}_0, \mathbf{Y}_1, \ldots$—взаимно независимые случайные величины, и пусть $\mathbf{S}_n = \mathbf{Y}_0 + \ldots + \mathbf{Y}_n$. Разность $\mathbf{S}_n - \mathbf{S}_m$ (при $m < n$) зависит лишь от $\mathbf{Y}_{m+1}, \ldots, \mathbf{Y}_n$, и поэтому легко видеть, что последовательность $\{\mathbf{S}_n\}$ является марковским процессом. Определим теперь новую последовательность случайных величин $\mathbf{U}_n$, полагая

$$\mathbf{U}_n = \mathbf{S}_0 + \mathbf{S}_1 + \ldots + \mathbf{S}_n = \mathbf{Y}_n + 2\mathbf{Y}_{n-1} + 3\mathbf{Y}_{n-2} + \ldots + (n-1)\mathbf{Y}_0.$$

Последовательность $\{\mathbf{U}_n\}$ образует стохастический процесс, вероятностные соотношения для которого могут быть в принципе выражены через распределения величин $\mathbf{Y}_k$. Но этот процесс в общем случае уже не является марковским, поскольку нет никакой причины, по которой, например, $\mathbf{P}\{\mathbf{U}_n = 0 \mid \mathbf{U}_{n-1} = a\}$ должна была бы равняться $\mathbf{P}\{\mathbf{U}_n=0 \mid \mathbf{U}_{n-1}=a,\ \mathbf{U}_{n-2}=b\}$; зная $\mathbf{U}_{n-1}$ и $\mathbf{U}_{n-2}$, можно сделать лучшие предсказания, чем зная одно лишь $\mathbf{U}_{n-1}$.

В случае непрерывного временно́го параметра суммирование здесь заменяется интегрированием. В теории диффузии $\mathbf{Y}_n$ играют роль ускорений; тогда $\mathbf{S}_n$ — скорости, а $\mathbf{U}_n$ — положения. Если могут быть измерены только положения, то мы вынуждены будем изучать немарковский процесс, хотя он и определяется косвенным образом через марковский процесс.

в) *Скользящие средние.* Пусть снова $\{\mathbf{Y}_n\}$ — последовательность взаимно независимых случайных величин. Скользящие средние порядка r определяются как $\mathbf{X}^{(n)}=(\mathbf{Y}_n+\mathbf{Y}_{n+1}+\ldots+\mathbf{Y}_{n+r-1})/r$. Легко видеть, что $\{\mathbf{X}^{(n)}\}$ не является марковским процессом. Процессы этого типа встречаются во многих приложениях (см. задачу 25).

г) *Задача об уличном движении.* Чтобы получить эмпирический пример немарковского процесса, Р. Фюрт [1]) провел наблюдения над числом пешеходов на определенном участке улицы. Идеализированная математическая модель этого процесса может быть получена следующим образом. Для простоты мы предположим, что все пешеходы имеют одну и ту же скорость v, и рассмотрим только пешеходов, движущихся в одном направлении. Разобьем ось x на отрезки I_1, $I_2, \ldots$ фиксированной длины d и будем регулярно наблюдать за расположением пешеходов в моменты, отстоящие друг от друга на d/v единиц времени. Определим случайную величину $\mathbf{Y}_k$ как число пешеходов, находившихся первоначально в I_k. При n-м наблюдении эти пешеходы будут обнаружены в I_{k-n}, тогда как интервал I_k будет содержать $\mathbf{Y}_{k+n}$ пешеходов. Стало быть, полное число пешеходов в интервале $0<x<Nd$ будет тогда равно $\mathbf{X}^{(n)}=\mathbf{Y}_{n+1}+\ldots+\mathbf{X}_{n+N}$, и поэтому наш процесс является по существу процессом скользящего среднего. Простейшая модель для случайных величин $\mathbf{Y}_k$ представляется испытаниями Бернулли. В пределе при $d\to 0$ они приводят к непрерывной модели, в которой биномиальное распределение заменяется распределением Пуассона.

д) *Суперпозиция марковских процессов (сложное тасование).* Существует много технических устройств (таких, как группы селекторов в телефонных коммутаторах, счетчики, фильтры), действие которых может быть описано как суперпозиция двух марковских процессов, результат которой уже не является марковским. Некоторое представление о таких механизмах можно получить, изучая следующий метод тасования карт.

Пусть помимо основной колоды из N карт у нас есть такая же вспомогательная колода, и эта вспомогательная колода тасуется обычным образом. Если карты в ней расположатся в порядке $(a_1, a_2, \ldots, a_N)$, то мы переставим карты в основной колоде так, чтобы

---

[1]) Fürth R., Schwankungserscheinungen in der Physik, Sammlung Vieweg, Braunschweig, 1920, 17ff. Результаты наблюдений появились в Physikalische Zeitschrift, 19 (1918), 20 (1919).

первая, вторая, ..., N-я карта переместились на места с номерами $a_1, a_2, \ldots, a_N$. Таким образом, тасование вспомогательной колоды косвенно определяет последовательные порядки карт в основной колоде. Эти последние образуют *стохастический процесс немарковского типа*. Для доказательства этого достаточно показать, что два последовательных порядка карт в основной колоде дают в общем случае больше информации о будущем, чем один последний порядок. Мы покажем это для простого частного случая.

Пусть $N=4$; предположим, что вспомогательная колода находится вначале в порядке (2431). Предположим, кроме того, что операция тасования всегда представляет собой только «снимание», т. е. порядок (a_1, a_2, a_3, a_4) меняется на один из трех следующих: (a_2, a_3, a_4, a_1), (a_3, a_4, a_1, a_2), (a_4, a_1, a_2, a_3); каждой из этих трех возможностей мы припишем вероятность 1/3. При этих соглашениях вспомогательная колода в любой момент времени будет иметь один из четырех порядков (2431), (4312), (3124), (1243). С другой стороны, уже непродолжительное экспериментирование покажет, что основная колода постепенно пройдет через все 24 возможных порядка и что каждый из них будет появляться в комбинации с каждым из четырех возможных порядков вспомогательной колоды. Это означает, что порядок (1234) основной колоды появится бесконечное число раз и что за ним всегда будет следовать один из четырех порядков (4132), (3421), (2314), (1243). Но вспомогательная колода не может оставаться в одном и том же порядке, и, следовательно, основная колода не может дважды подряд подвергнуться одной и той же перестановке. Стало быть, если при испытаниях с номерами $n-1$ и n имели место соответственно порядки карт (1234) и (1243), то в следующем испытании порядок (1234) будет невозможен. Таким образом, два последовательных наблюдения дают больше информации, чем одно-единственное.

е) *Немарковский процесс, удовлетворяющий уравнению Колмогорова—Чэпмена.* Тождество (3.3) было выведено из того предположения, что переход из E_ν в E_k не зависит от того способа, каким было достигнуто состояние E_ν. Поэтому сначала представлялось интуитивно ясным, что ни один немарковский процесс не должен удовлетворять этому тождеству; это предположение, казалось, подтверждалось тем фактом, что вероятности перехода за n шагов для такого процесса должны удовлетворять целому множеству любопытных тождеств. Тем не менее оказалось, что существуют исключения (хотя бы в теории). Действительно, в гл. IX,1 мы встретились с бесконечной последовательностью попарно независимых одинаково распределенных случайных величин, принимающих значения 1, 2 и 3 с вероятностью 1/3 каждое. Таким образом, мы имели процесс с возможными состояниями 1, 2, 3, такой, что $p_{jk}=1/3$ для всех комбинаций j и k. Стало быть, тождество (3.3) выполнялось тривиальным образом при $p_{jk}^{(n)}=1/3$. Тем не менее этот процесс не является марковским. Чтобы убедиться в этом, предположим, что первый шаг приводит систему в состояние 2. Тогда переход в 3 на следующем шаге будет возможен тогда и только тогда, когда начальным состоянием было 1. Таким образом, переходы, следующие за первым шагом, зависят не только от настоящего, но и от начального состояния. (Модификации различного вида можно найти в замечании в гл. IX,1 и в примечании к нему.)

§ 14. ЗАДАЧИ

1. Рассмотрим последовательность испытаний Бернулли и скажем, что в момент времени n наблюдается состояние E_1, если испытания с номерами $n-1$ и n привели к результату $УУ$. Аналогично E_2, E_3, E_4 означают исходы $УН$, $НУ$, $НН$. Найти матрицу P и все ее степени. Обобщить схему.

2. Классифицировать состояния четырех цепей, матрицы P которых имеют приведенные ниже строки. В каждом случае найти $P^{(2)}$ и асимптотическое поведение $p_{jk}^{(n)}$:

а) (0, 1/2, 1/2), (1/2, 0, 1/2), (1/2, 1/2, 0);
б) (0, 0, 0, 1), (0, 0, 0, 1), (1/2, 1/2, 0, 0), (0, 0, 1, 0);
в) (1/2, 0, 1/2, 0, 0), (1/4, 1/2, 1/4, 0, 0), (1/2, 0, 1/2, 0, 0), (0, 0, 0, 1/2, 1/2), (0, 0, 0, 1/2, 1/2);
г) (0, 1/2, 1/2, 0, 0, 0), (0, 0, 0, 1/3, 1/3, 1/3), (0, 0, 0, 1/3, 1/3, 1/3), (1, 0, 0, 0, 0, 0), (1, 0, 0, 0, 0, 0), (1, 0, 0, 0, 0, 0).

3. Рассмотрим бросания правильной кости и условимся говорить, что в момент n система находится в состоянии E_j, если j — наибольшее из чисел, выпавших в первых n бросаниях. Найти матрицу P^n и убедиться в том, что (3.3) выполняется.

4. В примере 2,к) найти вероятности (поглощения) x_k и y_k того, что, выходя из E_k, система закончит эволюцию соответственно в E_1 или в E_5 ($k=2$, 3, 4, 6). (Решить эту задачу, исходя из основных определений и не обращаясь к § 8.)

5. Рассмотреть пример гл. 1, 5,б) как цепь Маркова. Вычислить вероятность выигрыша для каждого игрока.

6. Пусть E_0 — поглощающее состояние (т. е. $p_{00}=1$). Пусть $p_{jj}=p$ и $p_{j,j-1}=q$ при $j>0$, где $p+q=1$. Найти вероятность $f_{j0}^{(n)}$ того, что поглощение в E_0 произойдет в точности на n-м шаге. Найти также математическое ожидание этого распределения.

7. Первая строка матрицы P равна $v_0, v_1, \ldots$. При $j>0$ имеем (как и в предыдущей задаче) $p_{jj}=p$ и $p_{j,j-1}=q$. Найти распределение времени возвращения в E_0.

8. Пусть $p_{j,j+2}=v_j$ и $p_{j0}=1-v_j$ при $j=0, 1, \ldots$. Выяснить характер состояний.

9. *Два отражающих экрана.* Цепь с набором состояний 1, 2, ..., ρ имеет матрицу, первая и последняя строки которой суть $(q, p, 0, \ldots, 0)$ и $(0, \ldots, 0, q, p)$. Во всех остальных строках $p_{k,k+1}=p$, $p_{k,k-1}=q$. Найти стационарное распределение. Может ли эта цепь быть периодической?

10. Обобщить *модель Бернулли — Лапласа для диффузии* (пример 2,е)), предполагая, что имеются $b\geqslant\rho$ черных и $w=2\rho-b$ белых частиц. Число частиц в каждой урне остается равным ρ.

11. Цепь с набором состояний E_0, E_1, ... имеет переходные вероятности

$$p_{jk}=e^{-\lambda}\sum_{\nu=0}^{j}\binom{j}{\nu}p^{\nu}q^{j-\nu}\frac{\lambda^{k-\nu}}{(k-\nu)!},$$

где слагаемые с $\nu>k$ следует считать равными нулю. Показать, что

$$p_{jk}^{(n)}\longrightarrow e^{-\lambda/q}(\lambda/q)^k/k!.$$

Замечание. Эта цепь встречается в статистической механике[1]) и может быть интерпретирована следующим образом. Состояние системы определяется числом

[1]) Chandrasekhar S. Stochastic problems in physics and astronomy, Rew. of Modern Physics, 15(1943), 1—89; p. 45. [Имеется перевод: Чандрасекар С. Стохастические процессы в физике и астрономии.— М.: ИЛ, 1947, с. 1—126, в частности с. 82.]

частиц в некоторой области пространства. В течение каждого единичного интервала времени каждая частица может с вероятностью q покинуть эту область, причем все частицы стохастически независимы. Кроме того, в эту область могут попадать новые частицы, и вероятность появления [за единицу времени.— *Перев.*] r новых частиц дается выражением Пуассона $e^{-\lambda}\lambda^r/r!$. Тогда стационарным распределением будет распределение Пуассона с параметром λ/q.

12. *Модель Эренфестов.* Пусть в примере 2,д) в первом сосуде сначала было j молекул, и пусть $\mathbf{X}^{(n)} = 2k - \rho$, если на n-м шаге система находилась в состоянии k (так что $\mathbf{X}^{(n)}$ есть разность числа молекул в двух наших сосудах). Пусть $e_n = \mathrm{E}(\mathbf{X}^{(n)})$. Доказать, что $e_{n+1} = (\rho-2)\, e_n/\rho$, откуда $e_n = (1-2/\rho)^n (2j-\rho)$. (Заметим, что $e_n \to 0$ при $n \to \infty$.)

13. Рассмотреть модель из задачи о счетчике (пример гл. XIII, 1, ж)) как цепь Маркова.

14. *Случайное блуждание на плоскости с отражающими экранами.* Рассмотрим симметричное случайное блуждание в ограниченной области плоскости. Ее граница является отражающей в том смысле, что каждый раз, когда при неограниченном случайном блуждании частица должна была бы покинуть эту область, она бывает вынуждена вернуться в предыдущее положение. Показать, что если каждая точка области достижима из любой другой точки, то существует стационарное распределение, и что $u_k = 1/a$, где a — число положений в области. (Если область не ограничена, то состояния являются возвратными нулевыми и $u_k = 1$ определяет инвариантную меру.)

15. *Повторное осреднение.* Пусть $\{x_1, x_2, \ldots\}$ — ограниченная последовательность чисел, и пусть P — матрица эргодической цепи. Доказать, что $\sum\limits_j p_{ij}^{(n)} x_j \to \sum u_j x_j$. Показать, что процедура повторного осреднения из примера гл. XIII, 10,в) является частным случаем этой задачи.

16. В теории *массового обслуживания* встречается матрица переходных вероятностей

$$\begin{bmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & p_0 & p_1 & \cdots \end{bmatrix},$$

где $\{p_k\}$ — некоторое распределение вероятностей. Используя производящие функции, выяснить характер состояний. Найти производящую функцию стационарного распределения, если оно существует.

17. *Время ожидания поглощения.* Пусть $\mathbf{Y}_j$ — момент времени, когда выходящая из невозвратного состояния E_j система *впервые* попадает в возвратное состояние. Предполагая, что вероятность остаться навсегда в множестве невозвратных состояний равна нулю, доказать, что $d_j = \mathrm{E}(\mathbf{Y}_j)$ однозначно определяется как решение системы линейных уравнений

$$d_j = \sum_T p_{j\nu} d_\nu + 1,$$

где суммирование проводится по всем ν, таким, что E_ν невозвратно. Однако d_ν не обязательно конечны.

18. Если число состояний $a < \infty$ и если E_k достижимо из E_j, то оно достижимо не более чем за $a-1$ шагов ($j \neq k$).

19. Пусть цепь содержит a состояний, и пусть E_j — возвратное состояние. Существует число $q < 1$, такое, что при $n \geqslant a$ вероятность того, что время возвращения в E_j превосходит n, меньше чем q^n. (*Указание.* Использовать результат задачи 18.)

20. В конечной цепи состояние E_j невозвратно тогда и только тогда, когда существует E_k, такое, что E_k достижимо из E_j, а E_j недостижимо из E_k. (Для бесконечных цепей это неверно, как показывают случайные блуждания.)

21. Неприводимая цепь, для которой положителен хотя бы *один* диагональный элемент p_{jj}, не может быть периодической.

22. Конечная неприводимая цепь непериодична тогда и только тогда, когда существует n, такое, что $p_{jk}^{(n)} > 0$ при всех j и k.

23. Пусть a состояний $(x_1, \ldots, x_a)$, некоторой цепи дают решение системы линейных уравнений $x_j = \sum p_{j\nu} x_\nu$. Доказать, что а) если $x_j \leqslant 1$ при всех j, то состояния, для которых $x_r = 1$, образуют замкнутое множество; б) если E_j и E_k принадлежат одному неприводимому множеству, то $x_j = x_k$; в) в неприводимой цепи решение $\{x_j\}$ сводится к константе. *Указание.* Рассмотреть сужение этих уравнений на замкнутом множестве.

24. *Продолжение.* Если $(x_1, \ldots, x_a)$ есть (комплекснозначное) решение системы $x_j = s \sum p_{j\nu} x_\nu$ при $|s| = 1$ и $s \neq 1$, то существует целое число $t > 1$, такое, что $s^t = 1$. Если цепь неприводима, то наименьшее такое целое число является периодом цепи.

Указание. Предположить без потери общности, что $x_1 = 1 \geqslant |x_\nu|$. Рассмотреть последовательно состояния, которые достигаются за 1, 2, ... шагов.

25. *Скользящие средние.* Пусть $\{\mathbf{Y}_k\}$ — последовательность взаимно независимых случайных величин, каждая из которых принимает значения ± 1 с вероятностями 1/2, и пусть $\mathbf{X}^{(n)} = (\mathbf{Y}_n + \mathbf{Y}_{n+1})/2$. Найти переходные вероятности

$$p_{jk}(m, n) = \mathbf{P}\{\mathbf{X}^{(n)} = k \mid \mathbf{X}^{(m)} = j\},$$

где $m < n$ и $j, k = -1, 0, 1$. Вывести отсюда, что $\{\mathbf{X}^{(n)}\}$ не является марковским процессом и что (13.3) не выполняется.

26. Рассмотрим последовательность испытаний Бернулли и скажем, что в момент времени n наблюдается состояние E_1, если испытания с номерами $n-1$ и n привели к успеху; в противном случае система находится в состоянии E_2. Найти вероятности перехода за n шагов и установить немарковский характер этого процесса.

Замечание. Этот процесс получается из цепи задачи 1 при объединении трех состояний в одно. Такая *группировка* применима к любой цепи Маркова и нарушает марковское свойство. Процессы такого типа изучал Харрис[1]).

27. *Смесь цепей Маркова.* Пусть заданы две цепи Маркова с одним и тем же числом состояний и матрицами P_1 и P_2. Новый процесс определяется некоторым начальным распределением и матрицей вероятностей перехода за n шагов $(1/2)P_1^n + (1/2)P_2^n$. Установить немарковский характер этого процесса и его связь с урновыми моделями гл. V, 2.

28. Пусть $\mathbf{N}$ — случайная величина, имеющая распределение Пуассона со средним λ. Рассмотрим $\mathbf{N}$ независимых марковских процессов, начинающихся из E_0 и имеющих одну и ту же матрицу P. Обозначим через $\mathbf{Z}_k^{(n)}$ число этих процессов, находящихся после n шагов в состоянии E_k. Показать, что $\mathbf{Z}_k^{(n)}$ имеет распределение Пуассона со средним $\lambda p_{0k}^{(n)}$.

Указание. Использовать результат примера гл. XII, 1, б).

29. Используя результат предыдущей задачи, показать, что случайная величина $\mathbf{X}_k^{(n)}$ из примера 11, б) имеет распределение Пуассона со средним $\sum_j u_j p_{jk}^{(n)} = u_k$.

---

[1]) Harris T. E., On chains of infinite order, Pacific Journal of Mathematics, 5 (1955), Supplement 1, 707—724.

АЛГЕБРАИЧЕСКАЯ ТРАКТОВКА КОНЕЧНЫХ ЦЕПЕЙ МАРКОВА

В этой главе мы рассматриваем цепь Маркова с конечным числом состояний $E_1, \ldots, E_\rho$ и заданной матрицей переходных вероятностей p_{jk}. Наша главная цель состоит в выводе явных формул для вероятностей $p_{jk}^{(n)}$ перехода за n шагов. Нам не потребуются результаты предыдущей главы, за исключением общих понятий и обозначений из § 3.

Мы воспользуемся методом производящих функций и получим желаемые результаты при помощи разложения на простые дроби, описанного в гл. XI, 4. Наши результаты можно также получить непосредственно из теории приведения матриц к каноническому виду (которая, в свою очередь, может быть выведена из наших результатов). Кроме того, для *конечных* цепей из результатов настоящей главы следуют эргодические свойства, доказанные в гл. XV. Однако для простоты мы несколько ограничим общность и будем игнорировать исключительные случаи, которые усложняют общую теорию и едва ли встречаются в практических примерах.

Общий метод намечен в основных чертах в § 1 и иллюстрируется в § 2 и 3. В § 4 особое внимание уделяется невозвратным состояниям и вероятностям поглощения. В § 5 наша теория применяется для нахождения дисперсий времен возвращения в состояния E_j.

§ 1. ОБЩАЯ ТЕОРИЯ

Введем для фиксированных j и k производящую функцию[1])

$$P_{jk}(s) = \sum_{n=0}^{\infty} p_{jk}^{(n)} s^n. \tag{1.1}$$

Умножая на sp_{ij} и суммируя по $j = 1, \ldots, \rho$, получаем

$$s \sum_{j=1}^{\rho} p_{ij} P_{jk}(s) = P_{ik}(s) - p_{ik}^{(0)}. \tag{1.2}$$

---

*) Эта глава посвящена специальному вопросу и может быть опущена. [Подробное изложение близких вопросов можно найти в книге Романовский В. И. Дискретные цепи Маркова.— М.— Л.: Гостехиздат, 1949.— *Перев.*]

[1]) Напомним, что $p_{jk}^{(0)}$ равняется 0 при $j \neq k$ и 1 при $j = k$. (Такие $p_{jk}^{(0)}$ известны под названием символов Кронекера.)

Это означает, что при фиксированных k и s величины $z_j = P_{jk}(s)$ удовлетворяют системе линейных уравнений вида

$$z_i - s\sum_{j=1}^{\rho} p_{ij} z_j = b_i. \tag{1.3}$$

Решения z_j системы (1.3), очевидно, являются рациональными функциями от s с общим знаменателем $D(s)$ — детерминантом этой системы. Для согласования со стандартными обозначениями из линейной алгебры положим $s = t^{-1}$. Тогда $t^{\rho} D(t^{-1})$ есть многочлен степени ρ (называемый характеристическим многочленом матрицы P переходных вероятностей p_{jk}). Его корни $t_1, \ldots, t_\rho$ называются *характеристическими числами* (или собственными значениями) матрицы P.

Введем теперь *упрощающие предположения о том, что характеристические числа просты* (отличны друг от друга) *и отличны от нуля*[1]). Это представляет некоторое ограничение общности, однако наша теория будет все же охватывать большинство представляющих интерес случаев.

Как уже говорилось, для фиксированных k и ρ величины $P_{jk}(s)$ суть рациональные функции от s с общим знаменателем $D(s)$. Корни многочлена $D(s)$ равны обратным величинам не обращающихся в нуль характеристических чисел t_ν. Поэтому из результатов гл. XI,4 следует, что существуют постоянные $b_{jk}^{(\nu)}$, такие, что[2])

$$P_{jk}(s) = b_{jk}^{(1)}/(1 - st_1) + \ldots + b_{jk}^{(\rho)}/(1 - st_\rho). \tag{1.4}$$

Разлагая дроби в геометрические ряды, мы получаем эквивалентные соотношения

$$p_{jk}^{(n)} = b_{jk}^{(1)} t_1^n + \ldots + b_{jk}^{(\rho)} t_\rho^n, \tag{1.5}$$

справедливые для всех целых $n \geqslant 0$. Покажем теперь, что коэффициенты $b_{jk}^{(\nu)}$ однозначно определяются как решения некоторых систем линейных уравнений. Величину $p_{ik}^{(n+1)}$ можно получить, заменяя в (1.5)[3]) n на $n+1$, но ее можно получить также, умножая (1.5) на p_{ij} и суммируя по $j = 1, \ldots, \rho$. Приравнивая эти два выражения, мы получаем тождество вида

$$C_1 t_1^n + \ldots + C_\rho t_\rho^n = 0, \tag{1.6}$$

---

[1]) Условие $t_r \neq 0$ вскоре будет отброшено. В примере 4, б) проводится численный анализ цепи с кратными корнями.

[2]) Теоретически мы должны были бы опустить те корни t_r, которые совпадают с корнями числителя. Однако для таких корней мы положим $b_{jk}^{(\nu)} = 0$, и поэтому (1.4) и (1.5) остаются верными при любых обстоятельствах.

[3]) При $j = i$.— *Прим. перев.*

справедливое для всех n. Это, очевидно, невозможно, если все коэффициенты не обращаются в нуль, и мы заключаем, что

$$\sum_{j=1}^{\rho} p_{ij} b_{jk}^{(\nu)} = t_\nu b_{ik}^{(\nu)} \tag{1.7}$$

для всех комбинаций i, k и ν. Умножая (1.5) на p_{kr} и суммируя по k, аналогичным образом получаем

$$\sum_{k=1}^{\rho} b_{jk}^{(\nu)} p_{kr} = t_\nu b_{jr}^{(\nu)}. \tag{1.8}$$

Рассмотрим $(\rho \times \rho)$-матрицу $b^{(\nu)}$ с элементами $b_{jk}^{(\nu)}$. В силу соотношений[1]) (1.7) ее k-й столбец представляет решение системы ρ линейных уравнений

$$\sum_{j=1}^{\rho} p_{ij} x_j - t x_i = 0 \tag{1.9}$$

при $t = t_\nu$; аналогично в силу (1.8) ее j-я строка удовлетворяет системе

$$\sum_{k=1}^{\rho} y_k p_{kr} - t y_r = 0 \tag{1.10}$$

при $t = t_\nu$. Система (1.10) получается из (1.9), когда строки и столбцы меняются местами, и поэтому детерминанты этих систем одинаковы. Детерминант системы (1.9) обращается в нуль только тогда, когда t совпадает с одним из характеристических значений $t_1, \ldots, t_\rho$. Иначе говоря, системы (1.9) и (1.10) допускают нетривиальные решения тогда и только тогда, когда $t = t_\nu$ для некоторого ν. Обозначим пару соответствующих решений через $(x_1^{(\nu)}, \ldots, x_\rho^{(\nu)})$ и $(y_1^{(\nu)}, \ldots, y_\rho^{(\nu)})$. Они определены с точностью до постоянного множителя, так что

$$b_{jk}^{(\nu)} = c^{(\nu)} x_j^{(\nu)} y_k^{(\nu)}, \tag{1.11}$$

где $c^{(\nu)}$ — постоянная (не зависящая от j и k). Чтобы найти эту неизвестную постоянную, заметим, что из (1.9) по индукции вытекает, что

$$\sum_{j=1}^{\rho} p_{ij}^{(n)} x_j = t^n x_i \tag{1.12}$$

при всех n. Воспользуемся этим соотношением при $t = t_\lambda$, где λ — произвольное целое число, лежащее между 1 и ρ. Подставляя

[1]) Системы (1.7) и (1.8) можно записать в компактной векторно-матричной форме $Pb^{(\nu)} = t_\nu b^{(\nu)}$ и $b^{(\nu)}P = t_\nu b^{(\nu)}$.

вместо $p_{ij}^{(n)}$ выражения (1.5), находим, что

$$t_\lambda^n x_i = t_1^n c^{(1)} x_i^{(1)} \sum_{k=1}^{\rho} y_k^{(1)} x_k^{(\lambda)} + \ldots + t_\rho^n c^{(\rho)} x_i^{(\rho)} \sum_{k=1}^{\rho} y_k^{(\rho)} x_k^{(\lambda)}. \quad (1.13)$$

Это тождество вида (1.6), которое может выполняться лишь тогда, когда все коэффициенты обращаются в нуль. Приравнивая коэффициенты при t_λ^n в обеих частях, получаем окончательно[1])

$$c^{(\lambda)} \sum_{k=1}^{\rho} y_k^{(\lambda)} x_k^{(\lambda)} = 1. \quad (1.14)$$

Это соотношение определяет коэффициент $b_{jk}^{(\lambda)}$ в (1.11). Правда, $x_j^{(\lambda)}$ и $y_k^{(\lambda)}$ определяются только с точностью до постоянного множителя, но замена $x_j^{(\lambda)}$ на $Ax_j^{(\lambda)}$ и $y_k^{(\lambda)}$ на $By_k^{(\lambda)}$ меняет $c^{(\lambda)}$ на $c^{(\lambda)}/(AB)$, и коэффициент $b_{jk}^{(\lambda)}$ остается неизменным.

Резюмируем этот результат следующим образом. Системы линейных уравнений (1.9) и (1.10) допускают нетривиальные решения не более чем для ρ различных значений t (одних и тех же для обеих систем). Предположим, что существует ровно ρ таких значений $t_1, \ldots, t_\rho$ и что все они отличны от 0. *Выберем для каждого t_λ ненулевое решение $(x_1^{(\lambda)}, \ldots, x_\rho^{(\lambda)})$ системы* (1.9) *и ненулевое решение $(y_1^{(\lambda)}, \ldots, y_\rho^{(\lambda)})$ системы* (1.10). *Тогда при $c^{(\lambda)}$, определяемых в* (1.14), *имеем для* $n = 0, 1, \ldots$

$$p_{jk}^{(n)} = \sum_{\lambda=1}^{\rho} c^{(\lambda)} x_j^{(\lambda)} y_k^{(\lambda)} t_\lambda^n. \quad (1.15)$$

Таким образом, мы нашли явное выражение для всех переходных вероятностей[2]).

Предположение о том, что все характеристические числа различны, удовлетворяется в большинстве практических случаев, за исключением случая разложимых цепей, который требует лишь небольших изменений в рассуждениях (см. § 4). Однако нередко одним из характеристических чисел оказывается нуль. Положим в этом случае $t_\rho = 0$. Новизна состоит здесь в том, что детерминант $D(s)$

---

[1]) Обращение в нуль остальных коэффициентов влечет равенство $\sum_{k=1}^{\rho} y_k^{(\lambda)} x_k^{(\nu)} = 0$ при $\lambda \neq \nu$.

[2]) В векторно-матричной форме окончательная формула (1.15) становится более изящной. Пусть $X^{(\lambda)}$ — вектор-*столбец* (или $(\rho \times 1)$-матрица) с элементами $x_j^{(\lambda)}$, а $Y^{(\lambda)}$ — вектор-*строка* (или $(1 \times \rho)$-матрица) с элементами $y_k^{(\lambda)}$. Тогда (1.15) принимает вид

$$P^n = \sum_{\lambda=1}^{\rho} c^{(\lambda)} X^{(\lambda)} Y^{(\lambda)} t_\lambda^n,$$

а $c^{(\lambda)}$ определяется скалярным уравнением $c^{(\lambda)} Y^{(\lambda)} X^{(\lambda)} = 1$.

системы (1.3) имеет только $\rho-1$ корней $t_1^{-1}, \ldots, t_{\rho-1}^{-1}$, и поэтому производящая функция $P_{jk}(s)$ является отношением двух многочленов степени $\rho-1$. Для разложения на простейшие дроби требуется, чтобы степень числителя была меньше степени знаменателя, и, чтобы добиться этого, мы должны сперва вычесть из $P_{jk}(s)$ подходящую постоянную. Таким образом мы получим для $P_{jk}(s)$ разложение на простые дроби, отличающееся от (1.4) тем, что последний член в нем заменяется постоянной. Из (1.15) ясно, что такая замена влияет на правую часть этого равенства лишь при $n=0$. Иначе говоря, *явное представление* (1.15) *для* $p_{jk}^{(n)}$ *остается справедливым при* $n \geqslant 1$, *даже если* $t_\rho = 0$ (при условии, что корни $t_1, \ldots, t_{\rho-1}$ различны и отличны от нуля).

Левая часть (1.15) может оставаться ограниченной при всех n только тогда, когда $|t_\lambda| \leqslant 1$ для всех λ. При $t=1$ уравнения (1.9) имеют решение $x_j=1$, так что одно характеристическое число равно 1. Не ограничивая общности, мы можем положить $t_1=1$. Если цепь непериодична, то для всех других характеристических чисел мы имеем $|t_\lambda|<1$, и из (1.15) видно, что при $n \to \infty$

$$p_{jk}^{(n)} \to c^{(1)} y_k^{(1)}. \tag{1.16}$$

Иначе говоря, *инвариантное распределение вероятностей можно характеризовать как решение системы* (1.10) *при* $t=1$.

2. ПРИМЕРЫ

а) Рассмотрим сначала цепь, имеющую только два состояния. Матрица переходных вероятностей принимает простой вид

$$P=\begin{pmatrix} 1-p & p \\ \alpha & 1-\alpha \end{pmatrix},$$

где $0<p<1$ и $0<\alpha<1$. Вычисления здесь тривиальны, поскольку рассматриваются лишь системы из двух уравнений. Характеристические числа суть $t_1=1$ и $t_2=1-\alpha-p$. Явное представление (1.15) для $p_{jk}^{(n)}$ можно записать в матричной форме:

$$P^n=\frac{1}{\alpha+p}\begin{pmatrix} \alpha & p \\ \alpha & p \end{pmatrix}+\frac{(1-\alpha-p)^n}{\alpha+p}\begin{pmatrix} p & -p \\ -\alpha & \alpha \end{pmatrix}$$

(где множители, общие для всех четырех элементов матриц, вынесены как коэффициенты при матрицах). Эта формула справедлива для $n \geqslant 0$.

б) Пусть

$$P=\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2.1}$$

(это матрица из задачи 2, б гл. XV, 14). Система (1.9) сводится к

$$x_4 = tx_1, \quad x_4 = tx_2, \quad (1/2)(x_1 + x_2) = tx_3, \quad x_3 = tx_4. \tag{2.2}$$

Значению $t=0$ здесь соответствует решение $(1, -1, 0, 0)$, однако мы видели, что для явного представления $p_{jk}^{(n)}$ при $n \geqslant 1$ корень 0 не требуется. Стандартная процедура последовательного исключения переменных показывает, что остальные корни удовлетворяют кубическому уравнению $t^3 = 1$. Если для сокращения записи положить

$$\theta = e^{2\pi i/3} = \cos(2\pi/3) + i \sin(2\pi/3) \tag{2.3}$$

(где $i^2 = -1$), то этими тремя корнями будут $t_1 = 1$, $t_2 = \theta$ и $t_3 = \theta^2$ (или, что то же самое, $t_3 = \theta^{-1}$). Теперь мы должны решить при этих значениях t системы (1.9) и (1.10). Поскольку постоянный множитель остается произвольным, мы можем положить $x_1^{(\nu)} = y_1^{(\nu)} = 1$. Тогда решения в окончательном явном представлении будут совпадать соответственно с первыми столбцами и первыми строками трех матриц:

$$P^n = \frac{1}{6}\begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix} + \frac{\theta^n}{6}\begin{bmatrix} 1 & 1 & 2\theta & 2\theta^2 \\ 1 & 1 & 2\theta & 2\theta^2 \\ \theta^2 & \theta^2 & 2 & 2\theta \\ \theta & \theta & 2\theta^2 & 2 \end{bmatrix} +$$

$$+ \frac{\theta^{2n}}{6}\begin{bmatrix} 1 & 1 & 2\theta^2 & 2\theta \\ 1 & 1 & 2\theta^2 & 2\theta \\ \theta & \theta & 2 & 2\theta^2 \\ \theta^2 & \theta^2 & 2\theta & 2 \end{bmatrix}. \tag{2.4}$$

Поскольку мы отбросили корень $t=0$, эта формула справедлива только при $n \geqslant 1$.

Из (2.4) очевидно, что наша цепь имеет период 3. Чтобы найти асимптотическое поведение P^n, заметим, что $1+\theta+\theta^2=0$. Используя этот факт, легко проверить, что при $n \to \infty$ по числам вида $n = 3k$ строки матрицы P^n будут стремиться к $(1/2, 1/2, 0, 0)$. Для $n = 3k+1$ и $n = 3k+2$ соответствующие пределы суть $(0, 0, 0, 1)$ и $(0, 0, 1, 0)$. Отсюда следует, что инвариантное распределение вероятностей дается вектором $(1/6, 1/6, 1/3, 1/3)$.

в) Пусть $p+q=1$ и

$$P = \begin{bmatrix} 0 & p & 0 & q \\ q & 0 & p & 0 \\ 0 & q & 0 & p \\ p & 0 & q & 0 \end{bmatrix}. \tag{2.5}$$

Эта цепь представляет собой частный случай цепи из следующего примера, но из-за своей простоты рассматривается отдельно. Легко видеть, что система (1.9) сводится к двум линейным уравнениям для двух неизвестных x_1+x_3 и x_2+x_4, и, следовательно, четыре характеристических числа здесь суть

$$t_1=1,\ t_2=-1,\ t_3=i(q-p),\ t_4=-i(q-p). \tag{2.6}$$

Соответствующими решениями будут $(1, 1, 1, 1)$, $(-1, 1, -1, 1)$, $(-i, -1, i, 1)$ и $(i, -1, -i, 1)$. (Заметим, что они имеют вид $(\theta, \theta^2, \theta^3, \theta^4)$, где θ — корень четвертой степени из единицы.) Система (1.10) отличается от (1.9) лишь тем, что p и q в ней меняются местами, и поэтому без дальнейших вычислений мы получаем

$$p_{jk}^{(n)}=(1/4)\{1+(q-p)^n i^{j-k-n}\}\{1+(-1)^{k+j-n}\}. \tag{2.7}$$

г) В *общем циклическом случайном блуждании* из примера гл. XV, 2, г) первая строка матрицы P имеет вид $q_0, \ldots, q_{\rho-1}$, а остальные строки получаются из нее циклическими перестановками. В предыдущем примере было показано, что в частном случае $\rho=4$ $x_j^{(\nu)}$ и $y_k^{(\nu)}$ могут быть представлены как степени корней четвертой степени из единицы. Поэтому естественно попытаться найти здесь аналогичное представление через корни ρ-й степени из единицы, а именно через величины

$$\theta=e^{2i\pi/\rho}. \tag{2.8}$$

Все корни ρ-й степени из единицы суть $1, \theta, \theta^2, \ldots, \theta^{\rho-1}$. Положим для $r=0, \ldots, \rho-1$

$$t_r=\sum_{\nu=0}^{\rho-1} q_\nu \theta^{\nu r}. \tag{2.9}$$

Легко проверить, что при $t=t_r$ системы (1.9) и (1.10) имеют решения

$$x_j^{(r)}=\theta^r, \quad y_k^{(r)}=\theta^{-rk} \tag{2.10}$$

и что для соответствующих коэффициентов $c^{(r)}$ мы во всех случаях имеем $c^{(r)}=1/\rho$. Таким образом, окончательно[1])

$$p_{jk}^{(n)}=\rho^{-1}\sum_{r=0}^{\rho-1} \theta^{r(j-k)} t_r^n. \tag{2.11}$$

[1]) При $n=0$ правая часть (2.11) определена только тогда, когда ни одно из t_r не обращается в нуль. В действительности мы доказали (2.11) для $n \geqslant 1$, предполагая, что все корни t_r различны, а это отнюдь не всегда верно в настоящей ситуации. Например, если $q_k=\rho^{-1}$ при всех k, то $t_0=1$, а $t_1=\ldots=t_{\rho-1}=0$. Но даже в этом крайнем случае (2.11) остается справедливым при всех j, k и $n \geqslant 1$. К счастью, (2.11) нетрудно проверить непосредственной индукцией по n. В частности, для $n=1$ множитель при q_ν

д) *Задача о размещении.* Пример гл. XV, 2, ж) показывает, что классическая задача о размещении может изучаться методом цепей Маркова. Система находится в состоянии j, если имеется j занятых и $\rho-j$ пустых ящиков. Если это состояние является начальным и по ящикам случайным образом размещается еще n дополнительных шаров, то $p_{jk}^{(n)}$ есть вероятность того, что после этого будет k занятых и $\rho-k$ пустых ящиков (так что $p_{jk}^{(n)}=0$ при $k<j$). Для $j=0$ эта вероятность получается по формуле (11.7) гл. II. Теперь мы выведем формулу для $p_{jk}^{(n)}$, обобщив таким образом результаты гл. II.

Поскольку $p_{jj}=j/\rho$ и $p_{j,\,j+1}=(\rho-j)/\rho$, система (1.9) сводится к

$$(\rho t-j)\,x_j=(\rho-j)\,x_{j+1}. \tag{2.12}$$

При $t=1$ отсюда вытекает, что $x_j=1$ для всех j. Когда $t\neq 1$, то с необходимостью $x_\rho=0$, и поэтому существует такой индекс r, что $x_{r+1}=0$, но $x_r\neq 0$; из (2.12) следует, что $\rho t=r$. Стало быть, характеристические числа здесь суть

$$t_r=r/\rho, \quad r=1,\ \ldots,\ \rho. \tag{2.13}$$

Соответствующие решения (2.12) даются формулой

$$x_j^{(r)}=\binom{r}{j}\binom{\rho}{j}^{-1}, \tag{2.14}$$

так что $x_j^{(r)}=0$ при $j>r$. Для $t=t_r$ система (1.10) сводится к

$$(r-j)\,y_j^{(r)}=(\rho-j+1)\,y_{j-1}^{(r)} \tag{2.15}$$

и имеет решение

$$y_j^{(r)}=\binom{\rho-r}{j-r}(-1)^{j-r}, \tag{2.16}$$

где, конечно же, $y_j^{(r)}=0$ при $j<r$. Поскольку $x_j^{(r)}=0$ при $j>r$ и $y_j^{(r)}=0$ при $j<r$, мы имеем

$$1/c^{(r)}=x_r^{(r)}y_r^{(r)}=\binom{\rho}{r},$$

и, следовательно,

$$p_{jk}^{(n)}=\sum_{r=j}^{k}\left(\frac{r}{\rho}\right)^n\binom{\rho}{r}\binom{r}{j}\binom{\rho-r}{k-r}(-1)^{k-r}\binom{\rho}{j}^{-1}. \tag{2.17}$$

в (2.11) сводится к

$$\rho^{-1}\sum_{r=0}^{\rho-1}\theta^{r(j-k+\nu)}.$$

Сумма здесь равна нулю, за исключением случаев $j-k+\nu=0$ и $j-k+\nu=\rho$, а в этих случаях каждое слагаемое равно единице. Следовательно, $p_{jk}^{(1)}$ сводится к q_{k-j}, если $k\geqslant j$, и к $q_{\rho+k-j}$, если $k<j$, а это и есть заданная матрица (p_{jk}).

Если выразить биномиальные коэффициенты через факториалы, то эта формула упростится:

$$p_{jk}^{(n)}=\binom{\rho-j}{\rho-k}\sum_{\nu=0}^{k-j}\left(\frac{\nu+j}{\rho}\right)^n(-1)^{k-j-\nu}\binom{k-j}{\nu} \tag{2.18}$$

и $p_{jk}^{(n)}=0$ при $k<j$. ▶

(Численную иллюстрацию см. в примере 4, б).)

§ 3. СЛУЧАЙНОЕ БЛУЖДАНИЕ С ОТРАЖАЮЩИМИ ЭКРАНАМИ

Теперь применение цепей Маркова будет продемонстрировано на примере полного обсуждения случайного блуждания с состояниями $1, 2, \ldots, \rho$ и двумя отражающими экранами[1]). Матрица P для него приведена в примере гл. XV, 2, в). Для $2\leqslant k\leqslant\rho-1$ имеем $p_{k,\,k+1}=p$ и $p_{k,\,k-1}=q$; первая и последняя строки суть $(q, p, 0, \ldots, 0)$, $(0, \ldots, 0, q, p)$.

Для удобства сравнения с выводами гл. XIV мы теперь откажемся от переменной $t=s^{-1}$ и запишем характеристические числа в виде s_r^{-1} (вместо t_r); будет удобно пронумеровать их от 0 до $\rho-1$. Для переменной s линейная система (1.9) запишется в виде

$$\begin{aligned} &x_1=s\,(qx_1+px_2),\\ &x_j=s\,(qx_{j-1}+px_{j+1}), \quad j=2,\ 3,\ \ldots,\ \rho-1,\\ &x_\rho=s\,(qx_{\rho-1}+px_\rho). \end{aligned} \tag{3.1}$$

Эта система допускает решение $x_j=1$, соответствующее корню $s=1$. Чтобы найти все остальные решения, применим метод частных решений (которым мы уже пользовались для решения аналогичных уравнений в гл. XIV, 4). Среднее уравнение в (3.1) удовлетворяется при $x_j=\lambda^j$, если λ — корень квадратного уравнения $\lambda=qs+\lambda^2ps$. Два корня этого уравнения суть

$$\lambda_1(s)=\frac{1+\sqrt{1-4pqs^2}}{2ps}, \quad \lambda_2(s)=\frac{1-\sqrt{1-4pqs^2}}{2ps}, \tag{3.2}$$

и поэтому общим решением среднего из уравнений (3.1) является

$$x_j=A(s)\,\lambda_1^j(s)+B(s)\,\lambda_2^j(s), \tag{3.3}$$

где $A(s)$ и $B(s)$ произвольны. Первое и последнее из уравнений (3.1) будут иметь решение (3.3) тогда и только тогда, когда $x_0=x_1$ и $x_\rho=x_{\rho+1}$. Для этого требуется, чтобы $A(s)$ и $B(s)$ удов-

[1]) Часть последующего изложения является повторением теории из гл. XIV. Там приведенное в тексте квадратное уравнение встречается под номером (4.7); выражения (3.2) для $\lambda_1(s)$ и $\lambda_2(s)$ даются в (4.8), а общее решение (3.3) появляется в виде (4.9). Два эти метода связаны друг с другом, однако во многих случаях детали вычислений оказываются совершенно различными.

летворяли условиям

$$A(s)\{1-\lambda_1(s)\}+B(s)\{1-\lambda_2(s)\}=0,$$
$$A(s)\lambda_1^\rho(s)\{1-\lambda_1(s)\}+B(s)\lambda_2^\rho(s)\{1-\lambda_2(s)\}=0. \tag{3.4}$$

И обратно, если эти два уравнения справедливы для некоторого значения s, то (3.3) представляет собой решение линейной системы (3.1), и это решение тождественно обращается в нуль только тогда, когда $\lambda_1(s)=\lambda_2(s)$. Стало быть, наша задача состоит в том, чтобы найти те значения s, для которых

$$\lambda_1^\rho(s)=\lambda_2^\rho(s), \quad \text{но} \quad \lambda_1(s)\neq\lambda_2(s). \tag{3.5}$$

Поскольку $\lambda_1(s)\lambda_2(s)=q/p$, то из первого соотношения вытекает, что $\lambda_1(s)\sqrt{p/q}$ должно быть корнем (2ρ)-й степени из единицы, т. е. мы должны иметь

$$\lambda_1(s)=\sqrt{q/p}\,e^{i\pi r/\rho}, \tag{3.6}$$

где r — целое число, $0\leqslant r<2\rho$. Из определения (3.2) легко видеть, что (3.6) имеет место только при $s=s_r$, где

$$s_r^{-1}=2\sqrt{pq}\cos\pi r/\rho. \tag{3.7}$$

Для значения $s=s_\rho$ второе условие (3.5) нарушается; более того, $s_r=s_{2\rho-r}$, и поэтому ρ различных характеристических чисел даются формулой (3.7) при $r=0, 1, \ldots, \rho-1$.

Решая (3.4) при $s=s_r$ и подставляя результат в (3.3), мы получаем

$$x_j^{(r)}=(q/p)^{j/2}\sin(\pi rj/\rho)-(q/p)^{(j+1)/2}\sin[\pi r(j-1)/\rho] \tag{3.8}$$

для $r=1, \ldots, \rho-1$, тогда как для $r=0$

$$x_j^{(0)}=1. \tag{3.9}$$

Сопряженная система (1.10) сводится к

$$\begin{aligned} y_1&=sq(y_1+y_2),\\ y_k&=s(py_{k-1}+qy_{k+1}), \quad k=2, \ldots, \rho-1,\\ y_\rho&=sp(y_{\rho-1}+y_\rho). \end{aligned} \tag{3.10}$$

Среднее уравнение здесь такое же, как в (3.1), но с переставленными p и q, и поэтому его общее решение получается из (3.3) с переставленными p и q. Первое и последнее уравнения могут быть удовлетворены, если $s=s_r$, и простые вычисления показывают, что для $r=1, 2, \ldots, \rho-1$ решением (3.10) является

$$y_k^{(r)}=(p/q)^{k/2}\sin(\pi rk/\rho)-(p/q)^{(k-1)/2}\sin[\pi r(k-1)/\rho]. \tag{3.11}$$

Для $s_0=1$ мы аналогично получаем

$$y_k^{(0)}=(p/q)^k. \tag{3.12}$$

Остается найти коэффициенты $c^{(r)}$, определяемые соотношениями

$$c^{(r)} \sum_{k=0}^{\rho-1} x_k^{(r)} y_k^{(r)} = 1. \tag{3.13}$$

При $r=0$ k-й член этой суммы равняется $(p/q)^k$, и поэтому

$$c^{(0)} = (q/p)\cdot[(p/q)-1]/[(p/q)^{\rho}-1], \tag{3.14}$$

за исключением случая $p=q$, когда $c_0 = 1/\rho$. При $r \geqslant 1$ элементарные, хотя и трудоемкие, вычисления[1]) приводят к выражению

$$c^{(r)} = (2p/\rho)\{1-2\sqrt{pq}\cos(\pi r/\rho)\}^{-1}. \tag{3.15}$$

Поэтому общее представление (1.15) для переходных вероятностей высших порядков приводит к окончательному результату[2])

$$p_{jk}^{(n)} = \frac{(p/q)-1}{(p/q)^{\rho}-1}\left(\frac{p}{q}\right)^{k-1} + \frac{2p}{\rho}\sum_{r=1}^{\rho-1} \frac{x_j^{(r)} y_k^{(r)}[2\sqrt{pq}\cos(\pi r/\rho)]^n}{1-2\sqrt{pq}\cos(\pi r/\rho)}, \tag{3.16}$$

где $x_j^{(r)}$ и $y_k^{(r)}$ определены формулами (3.8) и (3.11). При $p=q$ первый член в правой части следует считать равным $1/\rho$.

§ 4. НЕВОЗВРАТНЫЕ СОСТОЯНИЯ; ВЕРОЯТНОСТИ ПОГЛОЩЕНИЯ

Теорема § 1 была выведена в предположении, что корни t_1, t_2, ... различны. Наличие кратных корней не требует существенных изменений, однако мы обсудим лишь особо важный частный случай. Корень $t_1=1$ является кратным, когда цепь содержит две или несколько замкнутых подцепей; эта ситуация часто встречается в задачах, связанных с вероятностями поглощения. Легко приспособить метод § 1 к этому случаю. Для краткости и ясности мы объясним эту процедуру на примерах, в которых обнаружатся главные черты общего случая.

---

[1]) Эти вычисления значительно упростятся, если воспользоваться комплексной записью и тождеством $\sin v = [e^{iv}-e^{-iv}]/(2i)$. Сумма в (3.13) сводится к линейной комбинации (с комплексными коэффициентами) сумм вида

$$\sum_{j=0}^{\rho-1} e^{2j\pi i m/\rho},$$

где $m=0$ или $m=\pm 1$. В первом случае эта сумма равна ρ, во втором — нулю, откуда (3.15) следует тривиальным образом.

[2]) Аналогичные формулы в случае одного отражающего и одного поглощающего экрана см. в статье Каца (Kac M., Random walk and the theory of Brownian motion, Amer. Math. Monthly, 54 (1947), 369—391). Определение отражающего экрана модифицировано там таким образом, что частица может достичь 0; когда это происходит, следующий шаг переводит ее в 1. Явные формулы оказываются в этом случае более сложными. Статья Каца содержит также формулы для $p_{jk}^{(n)}$ в модели Эренфестов (пример гл. XV, 2, д)).

Примеры. а) Рассмотрим матрицу переходных вероятностей

$$P=\begin{bmatrix} 1/3 & 2/3 & 0 & 0 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 1/5 & 4/5 & 0 & 0 \\ 1/4 & 0 & 1/4 & 0 & 1/4 & 1/4 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}. \tag{4.1}$$

Ясно, что E_1 и E_2 образуют замкнутое множество (т. е. из них невозможен переход ни в одно из остальных четырех состояний; ср. гл. XV, 4). Аналогично E_3 и E_4 образуют другое замкнутое множество. Наконец, E_5 и E_6 суть невозвратные состояния. После конечного числа шагов система перейдет в одно из двух замкнутых множеств и останется там навсегда.

Матрица P имеет вид блочной матрицы

$$P=\begin{bmatrix} A & 0 & 0 \\ 0 & B & 0 \\ U & V & T \end{bmatrix}, \tag{4.2}$$

где каждая буква обозначает (2×2)-матрицу, а каждый нуль — матрицу из четырех нулей. Например, A имеет строки (1/3, 2/3) и (2/3, 1/3); это матрица переходных вероятностей, соответствующая цепи, образованной двумя состояниями E_1 и E_2. Эту матрицу можно изучать отдельно, а степени A^n могут быть получены из примера 2, а) при $p=\alpha=2/3$. Когда степени P^2, P^3, ... будут вычислены, окажется, что на первые две строки остальные четыре никак не влияют. Точнее, P^n имеет вид

$$P^n=\begin{bmatrix} A^n & 0 & 0 \\ 0 & B^n & 0 \\ U_n & V_n & T^n \end{bmatrix}, \tag{4.3}$$

где A^n, B^n, T^n суть n-е степени A, B и T соответственно и могут быть вычислены[1]) методом § 1 (ср. пример 2, а), где выполнены все вычисления). Вместо шести уравнений с шестью неизвестными перед нами теперь только системы из двух уравнений с двумя неизвестными каждая.

Следует отметить, что матрицы U_n и V_n в (4.3) не являются степенями U и V и не могут быть получены тем же простым способом, что и A^n, B^n и T^n. Однако при вычислении P^2, P^3, ...

[1]) Суммы по строкам в T не равны единице, так что T не является стохастической матрицей (это субстохастическая матрица в смысле определения гл. XV, 8). Метод § 1 применим без изменений, за исключением того, что $t=1$ не является более корнем (так что $T^n \longrightarrow 0$).

третий и четвертый столбцы никак не влияют на остальные четыре столбца. Иначе говоря, вычеркнув в P^n соответствующие E_3 и E_4 строки и столбцы, мы получим матрицу

$$\begin{pmatrix} A^n & 0 \\ U_n & T^n \end{pmatrix}, \tag{4.4}$$

которая является n-й степенью соответствующей подматрицы матрицы P, т. е. матрицы

$$\begin{pmatrix} A & 0 \\ U & T \end{pmatrix} = \begin{bmatrix} 1/3 & 2/3 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 \\ 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}. \tag{4.5}$$

Стало быть, матрица (4.4) может быть вычислена методом § 1, который в настоящем случае значительно упрощается. Матрица V_n может быть получена аналогичным образом.

Обычно явный вид матриц U_n и V_n интересен лишь постольку, поскольку они связаны с *вероятностями поглощения*. Если система выходит, скажем, из E_5, то какова будет *вероятность* λ *того, что она в конце концов попадет в замкнутое множество, образованное состояниями* E_1 *и* E_2 (а не в другое замкнутое множество)? Чему равна *вероятность* λ_n *того, что это произойдет в точности на* n*-м шаге?* Ясно, что $p_{51}^{(n)}+p_{52}^{(n)}$ есть вероятность того, что рассматриваемое событие произойдет не позже чем на n-м шаге, т. е.

$$p_{51}^{(n)}+p_{52}^{(n)}=\lambda_1+\lambda_2+\ldots+\lambda_n.$$

Устремляя n к ∞, получаем λ. Предпочтительный способ вычисления λ_n состоит в следующем. На $(n-1)$-м шаге система должна попасть в состояние, отличное от E_1 и E_2, т. е. либо в E_5, либо в E_6 (поскольку из E_3 или из E_4 переход в E_1 или E_2 невозможен). Затем n-й шаг переводит систему в E_1 или в E_2. Следовательно,

$$\lambda_n=p_{55}^{(n-1)}(p_{51}+p_{52})+p_{56}^{(n-1)}(p_{61}+p_{62})=(1/4)\,p_{55}^{(n-1)}+(1/3)\,p_{56}^{(n-1)}.$$

Отметим, что λ_n вполне определяется элементами T^{n-2}, а эту матрицу легко вычислить. В настоящем случае

$$p_{55}^{(n)}=p_{56}^{(n)}=(1/4)\,(5/12)^{n-1}, \quad \text{и, следовательно,} \quad \lambda_n=(7/48)\,(5/12)^{n-2}.$$

б) *Братско-сестринское скрещивание.* Мы завершим этот параграф численным анализом цепи из примера гл. XV, 2, к). Суть последующих рассуждений состоит в том, чтобы показать, что каноническое представление

$$p_{jk}^{(n)}=\sum_{r=1}^{6} t_r^n c^{(r)} x_j^{(r)} y_k^{(r)} \tag{4.6}$$

остается справедливым, хотя $t=1$ является *двойным корнем* характеристического уравнения.

Система (1.9) линейных уравнений принимает вид

$$\begin{gathered} x_1=tx_1, \quad (1/4)\,x_1+(1/2)\,x_2+(1/4)\,x_3=tx_2, \\ (1/16)\,x_1+(1/4)\,x_2+(1/4)\,x_3+(1/4)\,x_4+(1/16)\,x_5+(1/8)\,x_6=tx_3, \\ (1/4)\,x_3+(1/2)\,x_4+(1/4)\,x_5=tx_4, \quad x_5=tx_5, \quad x_3=tx_6, \end{gathered} \tag{4.7}$$

и эти уравнения выявляют форму заданной матрицы. Из первого и пятого уравнений ясно, что $x_1=x_5=0$, если не выполнено $t=1$. Стало быть, при $t\neq 1$ система на самом деле сводится к четырем уравнениям с четырьмя неизвестными, и стандартная процедура исключения переменных приводит к уравнению четвертой степени для t в качестве условия совместности этих четырех уравнений. Поскольку всего у нашей матрицы шесть собственных значений, отсюда следует, что $t=1$ является двойным корнем. Нетрудно проверить, что эти шесть собственных значений суть [1])

$$t_1=t_2=1, \quad t_3=1/2, \quad t_4=1/4, \quad t_5=1/4+\sqrt{5}/4, \qquad t_6=1/4-\sqrt{5}/4. \tag{4.8}$$

Соответствующие решения $(x_1^{(r)}, \ldots, x_6^{(r)})$ системы (4.7) могут быть выбраны следующим образом:

$$\begin{gathered} (1,\,3/4,\,1/2,\,1/4,\,0,\,1/2), \quad (0,\,1/4,\,1/2,\,3/4,\,1,\,1/2), (0,\,1,\,0,\,-1,\,0,\,0), \\ (0,\,1,\,-1,\,1,\,0,\,-4), \quad (0,\,1,\,-1+\sqrt{5},\,1,\,0,\,6-2\sqrt{5}), \\ (0,\,1,\,-1-\sqrt{5},\,1,\,0,\,6+2\sqrt{5}). \end{gathered} \tag{4.9}$$

Следующая задача состоит в нахождении соответствующих решений $(y_1^{(r)}, \ldots, y_6^{(r)})$ системы, получаемой из (4.7) при перестановке местами строк и столбцов [в матрице коэффициентов.— *Перев.*] Для $r\geqslant 3$ это решение определено с точностью до постоянного множителя, однако для двойного корня $t_1=t_2=1$ мы должны выбирать среди бесконечного числа решений вида $(a, 0, 0, 0, b, 0)$. Подходящий выбор становится очевидным из вида искомого представления (4.6). В самом деле, из (4.9) ясно, что $x_1^{(r)}=0$, за исключением случая $r=1$, и поэтому из (4.6) имеем $p_{1k}^{(n)}=c^{(1)}y_k^{(1)}$ для всех k и n. Но E_1 является поглощающим состоянием, и, очевидно, $p_{1k}^{(n)}=0$ при всех $k\neq 1$. Отсюда следует, что для $r=1$ мы должны выбрать решение вида $(a,\ 0,\ 0,\ 0,\ 0,\ 0)$, и по той же причине соответствующим $r=2$ решением будет $(0,\,0,\,0,\,0,\,b,\,0)$.

Решения, соответствующие остальным характеристическим числам, находятся легко. (Те из них, которые были выбраны для наших вычислений, представляются вторыми строками приводимых ниже матриц.) Затем из (1.14) определяются нормирующие постоянные

---

[1]) Корень $t_3=1/2$ легко обнаружить, поскольку он соответствует простому решению $x_2=-x_4=1$ и $x_1=x_3=x_5=x_6=0$. Для остальных корней получается несложное кубическое уравнение.

$c^{(r)}$, и таким образом мы получаем все величины, входящие в представление (4.6)

Для записи окончательного результата матрицы, соответствующие $r=1$ и $r=2$, были объединены в одну. Более того, соответствующие $r=5$ и $r=6$ элементы $c^{(r)}x_j^{(r)}y_k^{(r)}$ имеют вид $a \pm b\sqrt{5}$. По техническим причинам и для ясности было необходимо перегруппировать вносимый ими вклад и представить его в виде $a[t_5^n+t_6^n]$ и $b\sqrt{5}[t_5^n-t_6^n]$:

$$P^n=\begin{vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{3}{4} & 0 & 0 & 0 & \frac{1}{4} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{4} & 0 & 0 & 0 & \frac{3}{4} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \end{vmatrix} + \frac{2^{-n}}{4}\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

$$+\frac{4^{-n}}{20}\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -4 & 4 & -1 & -2 \\ 1 & -4 & 4 & -4 & 1 & 2 \\ -1 & 4 & -4 & 4 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & -16 & 16 & -16 & 4 & 8 \end{vmatrix} + \frac{t_5^n+t_6^n}{40}\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -9 & 6 & 4 & 6 & -9 & 2 \\ -11 & 4 & 16 & 4 & -11 & -2 \\ -9 & 6 & 4 & 6 & -9 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -14 & 16 & -16 & 16 & -14 & 12 \end{vmatrix}$$

$$+\frac{t_5^n-t_6^n}{40}\sqrt{5}\begin{vmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -4 & 2 & 4 & 2 & -4 & 0 \\ -5 & 4 & 0 & 4 & -5 & 2 \\ -4 & 2 & 4 & 2 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -6 & 0 & 16 & 0 & 6 & -4 \end{vmatrix}.$$

Легко проверить, что эта формула справедлива при $n=0$. С другой стороны, из структуры правой части (4.6) ясно, что если (4.6) имеет место для некоторого n, то оно справедливо и для $n+1$. Таким способом истинность (4.6) может быть установлена и без обращения к общей теории § 1.

§ 5. ПРИЛОЖЕНИЕ К ВРЕМЕНАМ ВОЗВРАЩЕНИЯ

В задаче 19 гл. XIII,12 показывается, как выразить среднее μ и дисперсию σ^2 времени возвращения рекуррентного события $\mathscr{E}$ через вероятности u_n осуществления события $\mathscr{E}$ при n-м испытании. Если $\mathscr{E}$ не является периодическим, то

$$u_n \to 1/\mu \quad \text{и} \quad \sum_{n=0}^{\infty}(u_n-1/\mu)=(\sigma^2-\mu+\mu^2)/(2\mu^2) \tag{5.1}$$

при условии, что σ^2 конечна.

Если отождествить $\mathscr{E}$ с возвратным состоянием E_j, то $u_n=p_{jj}^{(n)}$ (и $u_0=1$). В конечной цепи Маркова все времена возвращения имеют конечные дисперсии (см. задачу 19 гл. XV, 14), так что (5.1) применимо. Предположим, что E_j не является периодическим и применима формула (1.5). Тогда $t_1 = 1$ и $|t_r| < 1$ при $r = 2, 3, \ldots$, так что $p_{jj}^{(n)} \to b_{jj}^{(1)} = \mu_j^{-1}$. Члену $u_n - \mu^{-1}$ суммы в (5.1) соответствует

$$p_{jj}^{(n)} - 1/\mu_j = \sum_{r=2}^{\rho} b_{jj}^{(r)} t_r^n. \tag{5.2}$$

Эта формула справедлива при $n \geqslant 1$; суммируя геометрическую прогрессию со знаменателем t_r, находим

$$\sum_{n=1}^{\infty} (p_{jj}^{(n)} - 1/\mu_j) = \sum_{r=2}^{\rho} b_{jj}^{(r)} t_r/(1 - t_r). \tag{5.3}$$

Подставляя это выражение в (5.1), мы находим, что *если E_j — непериодическое возвратное состояние, то среднее время возвращения в него равно $\mu_j = 1/b_{jj}^{(1)}$, а дисперсия этого времени возвращения равна*

$$\sigma_j^2 = \mu_j - \mu_j^2 + 2\mu_j^2 \sum_{r=2}^{\rho} b_{jj}^{(r)} t_r/(1 - t_r) \tag{5.4}$$

при условии, что формула (1.5) применима и $t_1 = 1$. Случай периодических состояний и наличие двойных корней потребуют лишь очевидных изменений.

ГЛАВА XVII *) ПРОСТЕЙШИЕ СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ С НЕПРЕРЫВНЫМ ВРЕМЕНЕМ

§ 1. ОБЩИЕ ПОНЯТИЯ. МАРКОВСКИЕ ПРОЦЕССЫ

Цепи Маркова, изучавшиеся нами в предыдущих главах, могут быть описаны очень грубо как стохастические процессы, в которых будущее зависит лишь от настоящего состояния, но не от прошлой истории или того способа, которым было достигнуто настоящее состояние. Эти процессы имеют только счетное множество значений (состояний) E_1, E_2,... и зависят от дискретного временно́го параметра, т. е. изменения могут происходить лишь в фиксированные моменты времени $t=0, 1, \ldots$. В настоящей главе мы рассмотрим такие явления, как телефонные вызовы, радиоактивный распад и расщепление хромосом, в которых изменения могут происходить в любой момент времени. С математической точки зрения мы будем иметь дело со стохастическими процессами со счетным множеством состояний, но зависящими уже от непрерывного временно́го параметра. В рамках дискретных вероятностей полное описание таких процессов невозможно, и мы на самом деле не в состоянии формально определить интересующий нас класс марковских процессов.

Действительно, чтобы описать прошлую историю процесса, мы должны указать те моменты времени, в которые происходили изменения, а для этого потребуются вероятности на континууме. Выражение «будущее развитие не зависит от прошлой истории» имеет очевидное интуитивное значение (по крайней мере по аналогии с дискретными цепями Маркова), но для формального определения потребуется понятие условной вероятности, лежащее за пределами этой книги. Однако многие задачи, связанные с такими процессами, можно изучать отдельно при помощи довольно элементарных методов, если принять на веру, что эти процессы действительно существуют. Так мы и будем теперь поступать.

Переходной вероятности $p_{jk}^{(n)}$ для цепей Маркова теперь соответствует переходная вероятность $P_{jk}(t)$, а именно условная вероятность состояния E_k в момент $t+s$ при условии, что в момент $s < t+s$ система находилась в состоянии E_j. Как показывает обозначение, предполагается, что эта вероятность зависит только от продолжительности t временно́го интервала, но не от его положения на оси времени. Такие переходные вероятности называются

---

*) Эта глава почти независима от гл. X—XVI. Относительно использования термина «стохастический процесс» см. примечание 5 на с. 435.

стационарными или однородными по времени. (В § 9, однако, будут рассматриваться неоднородные процессы.) Аналогом основных соотношений (3.3) гл. XV является *уравнение Колмогорова—Чэпмена*

$$P_{ik}(\tau+t)=\sum_j P_{ij}(\tau)\,P_{jk}(t), \tag{1.1}$$

которое основано на следующем рассуждении. Предположим, что в момент времени 0 система находится в состоянии E_i. Тогда j-й член в правой части представляет вероятность сложного события, состоящего в том, что система в момент времени τ находится в состоянии E_j, а в более поздний момент $\tau+t$ — в состоянии E_k. Но переход из состояния E_i в момент времени 0 в состояние E_k в момент $\tau+t$ с необходимостью происходит через некоторое промежуточное состояние E_j в момент времени τ, и, суммируя по всем возможным E_j, мы видим, что (1.1) должно выполняться для произвольных (фиксированных) $\tau>0$ и $t>0$.

В этой главе мы будем изучать решения основного уравнения (1.1). Будет показано, что простые постулаты, приспособленные к конкретным ситуациям, приводят к системам дифференциальных уравнений для $P_{jk}(t)$ и что из этих дифференциальных уравнений, даже не решая их, можно получить интересные результаты. И эти результаты имеют смысл, потому что наши решения действительно являются переходными вероятностями марковского процесса, который однозначно определяется этими вероятностями и начальным положением в момент времени 0. Этот интуитивно очевидный факт [1]) мы примем без доказательства.

Для фиксированных j и t переходные вероятности $P_{jk}(t)$ определяют обычное дискретное распределение вероятностей. Оно зависит от непрерывного параметра t, однако мы уже встречались с многими семействами распределений, зависящих от непрерывного параметра. Технически рассуждения последующих параграфов остаются в рамках дискретных вероятностей, но это искусственное ограничение является для многих целей слишком строгим. Этот момент может проиллюстрировать распределение Пуассона $\{e^{-\lambda t}(\lambda t)^n/n!\}$. Нулевой член $e^{-\lambda t}$ этого распределения можно интерпретировать как вероятность того, что за интервал времени фиксированной длины t не поступило ни одного телефонного вызова. Но тогда $e^{-\lambda t}$ будет также вероятностью того, что время ожидания первого вызова превышает t, и поэтому мы косвенно имеем дело с непрерывным распределением вероятностей на оси времени. Мы вернемся к этому вопросу в § 6.

[1]) Стоит отметить, однако, что могут существовать (довольно патологические) немарковские процессы с теми же переходными вероятностями. Этот вопрос подробно обсуждался в разд. 2а гл. XII в связи с процессами с независимыми приращениями (которые являются частным классом марковских процессов). См. также § 9, в частности примечание на с. 485—486.

§ 2. ПУАССОНОВСКИЙ ПРОЦЕСС

Пуассоновский процесс можно рассматривать с различных точек зрения, и здесь мы рассмотрим его в качестве прототипа всех процессов из этой главы. Последующий вывод распределения Пуассона наилучшим образом подходит для наших обобщений, однако он никоим образом не является лучшим и в других контекстах. Его следует сравнить с элементарным выводом в гл. VI, 6 и с разд. 2а гл. XII, где пуассоновский процесс фигурировал как простейший процесс с независимыми приращениями.

В качестве эмпирических предпосылок возьмем такие случайные события, как распад частиц, поступающие телефонные вызовы, расщепление хромосом под действием вредной радиации. Предполагается, что все наблюдаемые события однотипны, и мы интересуемся полным числом $\mathbf{Z}(t)$ событий, происшедших в течение произвольного интервала времени длины t. Каждое событие представляется точкой на оси времени, и поэтому мы в действительности рассматриваем некоторые случайные размещения точек на прямой. Лежащие в основе нашей математической модели физические предположения состоят в том, что силы и воздействия, управляющие процессом, остаются постоянными, так что вероятность любого отдельного события одна и та же для всех интервалов времени продолжительности t и не зависит от прошлого развития процесса. Математически это означает, что наш процесс является однородным по времени марковским процессом в смысле, описанном в предыдущем параграфе. Как уже выше говорилось, мы не стремимся к полной теории таких процессов, а удовольствуемся выводом основных вероятностей

$$P_n(t) = \mathbf{P}\{\mathbf{Z}(t) = n\}. \tag{2.1}$$

Они могут быть выведены из простых постулатов без обращения к более глубоким теоретическим соображениям.

Чтобы ввести понятия, подходящие и для других процессов из этой главы, мы выберем начало отсчета времени и будем говорить, что в *момент времени* $t>0$ *система находится в состоянии* E_n, если между 0 и t произошло ровно n скачков [функции $\mathbf{Z}(t)$.— *Перев.*]. Тогда $P_n(t)$ равняется вероятности состояния E_n в момент t, однако $P_n(t)$ может быть также описана как вероятность перехода из произвольного состояния E_j в произвольный момент времени s в состояние E_{j+n} к моменту $s+t$. Теперь наше нестрогое описание процесса мы преобразуем в свойства вероятностей $P_n(t)$.

Разобьем временно́й интервал единичной длины на N подынтервалов длины $h=N^{-1}$. Вероятность скачка внутри любого из этих подынтервалов равна $1 - P_0(h)$, и поэтому математическое ожидание числа интервалов, содержащих скачки, равно $h^{-1}[1 - P_0(h)]$. Интуитивно представляется, что при $h \to 0$ это число должно стремиться к математическому ожиданию числа скачков внутри произвольного интервала времени единичной длины, и поэтому естественно предположить [1]), что существует число $\lambda>0$, такое, что

$$h^{-1}[1-P_0(h)] \to \lambda. \tag{2.2}$$

Физическая картина процесса требует также, чтобы скачок обязательно приводил из состояния E_j в соседнее состояние E_{j+1}, и отсюда вытекает, что математическое ожидание числа подынтервалов (длины h), содержащих более чем один скачок, должно стремиться к 0. Поэтому мы должны предположить, что при $h \to 0$

$$h^{-1}[1-P_0(h)-P_1(h)] \to 0. \tag{2.3}$$

Чтобы окончательно сформулировать постулаты, запишем (2.2) в виде $P_0(h)=1-\lambda h+o(h)$, где (как обычно) $o(h)$ обозначает величину, по порядку меньшую чем h. (Точнее говоря, $o(h)$ означает такую величину, что $h^{-1}o(h) \to 0$ при $h \to 0$.) С учетом этого (2.3) эквивалентно соотношению $P_1(h)=\lambda h+o(h)$. Сформулируем теперь следующие постулаты.

Постулаты пуассоновского процесса. *Процесс начинается в момент времени* 0 *в состоянии* E_0. (i) *Непосредственный переход из состояния* E_j *возможен только в состояние* E_{j+1}. (ii) *Каково бы ни было состояние* E_j *процесса в момент времени* t, *(условная) вероятность скачка внутри последующего короткого интервала времени между* t *и* $t+h$ *равна* $\lambda h+o(h)$, *тогда как (условная) вероятность наличия в нем более чем одного скачка есть* $o(h)$.

Как было объяснено в предыдущем параграфе, эти условия слабее нашего исходного предположения об отсутствии влияния прошлой истории процесса на его будущую эволюцию. С другой стороны, наши постулаты носят чисто аналитический характер, и их достаточно, чтобы показать, что мы должны иметь

$$P_n(t)=[(\lambda t)^n/n!]\,e^{-\lambda t}. \tag{2.4}$$

Для доказательства этого возьмем сперва $n\geqslant 1$ и рассмотрим событие, состоящее в том, что в момент времени $t+h$ система находится в состоянии E_n. Вероятность этого события равна $P_n(t+h)$, и осуществиться оно может тремя взаимоисключающими способами. Во-первых, в момент времени t система может находиться в состоянии E_n, и между t и $t+h$ не произойдет ни одного скачка. Вероятность этой возможности равна [2])

$$P_n(t)P_0(h)=P_n(t)[1-\lambda h]+o(h).$$

---

[1]) Предположение (2.2) вводится в основном потому, что его легко обобщить на другие процессы. В настоящем случае более естественным было бы заметить, что $P_0(t)$ должно удовлетворять функциональному уравнению $P_0(t+\tau)=P_0(t)\,P_0(\tau)$, из которого вытекает (2.2) (см. § 6).

[2]) В этом равенстве используются не входящие в постулаты пуассоновского процесса условия независимости и однородности по времени изменений процесса на непересекающихся интервалах времени. Легко убедиться в том, что для вывода (2.4) в действительности достаточно сформулированных постулатов; ср. замечание, сделанное после формулировки этих постулатов.— *Прим. перев.*

Вторая возможность состоит в том, что в момент времени t система находится в состоянии E_{n-1} и между t и $t+h$ происходит в точности один скачок. Вероятность этого равна $P_{n-1}(t)\cdot\lambda h+o(h)$. Любое другое состояние в момент t потребует [1]) более одного скачка в интервале между t и $t+h$, и вероятность подобного события есть $o(h)$. Следовательно, мы должны иметь

$$P_n(t+h)=P_n(t)(1-\lambda h)+P_{n-1}(t)\lambda h+o(h), \tag{2.5}$$

а это соотношение можно переписать в виде

$$[P_n(t+h)-P_n(t)]/h=-\lambda P_n(t)+\lambda P_{n-1}(t)+o(h)/h. \tag{2.6}$$

При $h\to 0$ последний член стремится к нулю; следовательно, предел [2]) левой части существует и равен

$$P_n'(t)=-\lambda P_n(t)+\lambda P_{n-1}(t),\quad n\geqslant 1. \tag{2.7}$$

При $n=0$ вторая и третья из упомянутых выше возможностей не возникают, и поэтому (2.5) следует заменить на

$$P_0(t+h)=P_0(t)(1-\lambda h)+o(h), \tag{2.8}$$

что приводит к

$$P_0'(t)=-\lambda P_0(t). \tag{2.9}$$

Отсюда и из $P_0(0)=1$ получаем $P_0(t)=e^{-\lambda t}$. Подставляя это значение $P_0(t)$ в (2.7) при $n=1$, мы получаем обыкновенное дифференциальное уравнение для $P_1(t)$. Поскольку $P_1(0)=0$, мы легко находим, что $P_1(t)=\lambda te^{-\lambda t}$, а это полностью согласуется с (2.4). Продолжая таким же образом, мы последовательно находим все члены (2.4).

§ 3. ПРОЦЕСС ЧИСТОГО РАЗМНОЖЕНИЯ

Простейшее обобщение пуассоновского процесса получается при предположении, что вероятности скачков могут зависеть от текущего состояния системы. Это приводит нас к следующим требованиям.

Постулаты. (i) *Непосредственный переход из состояния E_j возможен только в состояние E_{j+1}.* (ii) *Если в момент времени t систе-*

---

[1]) Для достижения E_n. Имеется в виду любое другое состояние E_k при $k<n-1$ (при $k>n$ переход из E_k в E_n невозможен).— *Прим. перев.*

[2]) Поскольку мы рассматриваем лишь положительные h, то $P_n'(t)$ в (2.7) следует понимать как правостороннюю производную. В действительности же это обычная двусторонняя производная. В самом деле, член $o(h)$ в (2.5) не зависит от t и поэтому остается неизменным при замене t на $t-h$. Таким образом, из (2.5) вытекает непрерывность, а из (2.6) — дифференцируемость в обычном смысле. Это замечание применимо на протяжении всей главы и повторяться не будет.

ма находится в состоянии E_n, то (условная) вероятность одного скачка в последующем коротком интервале времени между t и $t+h$ равна $\lambda_n h+o(h)$, тогда как (условная) вероятность более чем одного скачка в этом интервале есть $o(h)$.

Отличительная черта этого предположения заключается в том, что время, которое система проводит в любом конкретном состоянии, не играет никакой роли; возможны внезапные изменения состояния, однако, пока система находится в одном состоянии, она не стареет.

Пусть $P_n(t)$ снова будет вероятностью того, что в момент времени t система находится в состоянии E_n. Эти функции $P_n(t)$ удовлетворяют системе дифференциальных уравнений, которую можно вывести при помощи рассуждений предыдущего параграфа с тем лишь изменением, что (2.5) заменяется на

$$P_n(t+h)=P_n(t)(1-\lambda_n h)+P_{n-1}(t)\lambda_{n-1}h+o(h). \tag{3.1}$$

Таким образом мы получим *основную систему дифференциальных уравнений*

$$\begin{aligned} P_n'(t) &= -\lambda_n P_n(t)+\lambda_{n-1}P_{n-1}(t), \quad n\geqslant 1,\\ P_0'(t) &= -\lambda_0 P_0(t). \end{aligned} \tag{3.2}$$

В пуассоновском процессе было естественно предполагать, что в момент времени 0 система выходит из начального состояния E_0. Теперь мы можем допустить более общий случай, когда система выходит из произвольного начального состояния E_i. Тогда получаем, что [1])

$$P_i(0)=1, \quad P_n(0)=0 \text{ при } n\neq i. \tag{3.3}$$

Эти *начальные условия* единственным образом определяют решение $\{P_n(t)\}$ системы (3.2). (В частности, $P_0(t)=P_1(t)=\ldots=P_{i-1}(t)=0$.) Явные формулы для $P_n(t)$ выводились независимо многими авторами, однако для нас они не представляют интереса. Легко проверить, что для произвольных заданных λ_n система $\{P_n(t)\}$ обладает всеми требуемыми свойствами, за исключением того, что при некоторых условиях $\sum P_n(t)<1$. Это явление будет обсуждаться в § 4.

Примеры. а) *Радиоактивный распад.* В результате испускания частиц или γ-лучей радиоактивный атом, скажем урана, может превратиться в атом другого вида. Каждый вид представляет собой возможное состояние, и, когда процесс протекает, мы получаем последовательность переходов $E_0\to E_1\to E_2\to\ldots\to E_m$. Согласно принятым физическим теориям, вероятность перехода $E_n\to E_{n+1}$ остается неизменной, пока атом находится в состоянии E_n, и эта гипотеза находит выражение в нашем исходном предположении. Стало быть, этот процесс описывается дифференциальными уравнениями (3.2)

[1]) Следует отметить, что здесь $P_n(t)$ — та же вероятность, что и переходная вероятность $P_{in}(t)$ в § 1.

(факт, хорошо известный физикам). Если E_m — конечное состояние, из которого невозможны никакие другие переходы, то $\lambda_m=0$ и система (3.2) обрывается при $n=m$. (При $n>m$ мы автоматически получаем $P_n(t)=0$.)

б) *Процесс Юла.* Рассмотрим совокупность, элементы которой могут (путем деления или каким-либо другим способом) порождать новых членов, но не могут умереть (исчезнуть). Предположим, что на протяжении любого короткого интервала времени длины h каждый элемент совокупности с вероятностью $\lambda h+o(h)$ производит новый элемент; постоянная λ определяет скорость разрастания совокупности. Если между ее элементами нет никакого взаимодействия и в момент времени t объем совокупности был равен n, то вероятность того, что ее увеличение произойдет в некоторый момент времени между t и $t+h$, будет равна $n\lambda h+o(h)$. Поэтому вероятность $P_n(t)$ того, что в совокупности насчитывается ровно n элементов, удовлетворяет уравнениям (3.2) с $\lambda_n=n\lambda$, т. е. уравнениям

$$\begin{aligned} &P_n'(t)=-n\lambda P_n(t)+(n-1)\lambda P_{n-1}(t), \quad n\geqslant 1, \\ &P_0'(t)=0. \end{aligned} \tag{3.4}$$

Обозначим начальный объем совокупности через i. Тогда имеют место начальные условия (3.3), и легко проверить, что при $n\geqslant i>0$

$$P_n(t)=\binom{n-1}{n-i}e^{-i\lambda t}(1-e^{-\lambda t})^{n-i} \tag{3.5}$$

и, конечно, $P_n(t)=0$ при $n<i$ и всех t. Используя обозначение (8.1) гл. VI для отрицательного биномиального распределения, мы можем переписать (3.5) в виде $P_n(t)=f(n-i;\ i,\ e^{-\lambda t})$. Отсюда следует (ср. пример гл. IX, 3, в)), что объем совокупности в момент времени t является суммой i независимых случайных величин, каждая из которых имеет распределение, получающееся из (3.5) при замене i на 1. Эти i величин представляют собой потомства исходных элементов нашей совокупности.

Процесс такого типа впервые исследовал Юл [1]) в связи с матема-

---

[1]) Yule G. U., A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S., Philosophical Transactions of the Royal Society, London, Ser. B, 213 (1924), 21—87. Юл не вводил дифференциальных уравнений (3.4), а получил $P_n(t)$ предельным переходом, подобным тому, который использовался в гл. VI, 5.

Гораздо более общие и более гибкие модели того же типа были предложены и использованы для анализа эпидемий и роста популяций в написанной без особых претензий, но чрезвычайно интересной статье подполковника М'Кендрика (M'Kendrick A. G., Applications of mathematics to medical problems, Proceedings Edinburgh Mathematical Society, 44 (1925), 1—34). К сожалению, эта замечательная работа прошла практически незамеченной. В частности, она была неизвестна автору настоящей книги, когда он вводил различные стохастические модели для роста популяции в статье Feller W., Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung, Acta Biotheoretica, 5 (1939), 11—40.

тической теорией эволюции. Популяция состоит из видов в пределах одного рода, и появление нового вида обусловливается мутациями. Предположение о том, что каждый вид имеет одну и ту же вероятность породить новый вид, пренебрегает разницей в размерах видов. Поскольку мы пренебрегли также возможностью вымирания вида, можно ожидать, что (3.5) даст лишь грубое приближение.

Фэрри[1]) использовал ту же модель для описания процесса, связанного с космическими лучами, однако приближение снова было довольно грубым. Дифференциальные уравнения (3.4) применимы, строго говоря, к совокупностям частиц, которые могут делиться, образуя точные копии самих себя, при условии, конечно, что между частицами нет никаких взаимодействий. ▶

§ 4*). РАСХОДЯЩИЙСЯ ПРОЦЕСС РАЗМНОЖЕНИЯ

Решение $\{P_n(t)\}$ бесконечной системы дифференциальных уравнений (3.2), удовлетворяющее начальным условиям (3.3), может быть найдено индуктивно, начиная с $P_i(t)=e^{-\lambda_i t}$. Поэтому распределение $\{P_n(t)\}$ определено единственным образом. Из известных формул для решения линейных дифференциальных уравнений следует также, что $P_n(t)\geqslant 0$. Остается открытым лишь вопрос о том, является ли $\{P_n(t)\}$ настоящим распределением вероятностей, т. е. будет ли выполняться условие

$$\sum P_n(t)=1 \tag{4.1}$$

при всех t. Мы увидим, что это не всегда так: если коэффициенты λ_n быстро возрастают, то может случиться, что

$$\sum P_n(t)<1. \tag{4.2}$$

Когда эта возможность была обнаружена, она показалась настораживающей, однако ее легко объяснить. Левую часть (4.2) можно интерпретировать как вероятность того, что на протяжении интервала времени величины t имело место лишь *конечное число* скачков. Следовательно, разность правой и левой частей (4.2) отвечает возможности бесконечного числа скачков, или своего рода взрыва. Чтобы лучше понять это явление, сравним нашу вероятностную модель роста с обычным детерминистским подходом.

Величина λ_n в (3.2) может быть названа средней скоростью роста совокупности объема n. Например, в частном случае (3.4) имеем $\lambda_n=n\lambda$, так что здесь средняя скорость роста пропорциональна фактическому размеру популяции. Если рост не подвержен случайным флуктуациям и увеличивается со скоростью, пропорциональной

[1]) Furry W. H., On fluctuation phenomena in the passage of high-energy electrons through lead, Physical Reviews, 52 (1937), 569.

* Этот параграф посвящен специальному вопросу и может быть опущен.

мгновенному размеру популяции $x(t)$, то последний меняется в соответствии с детерминистским дифференциальным уравнением

$$dx(t)/dt=\lambda x(t), \tag{4.3}$$

из которого следует, что

$$x(t)=ie^{\lambda t}, \tag{4.4}$$

где $i=x(0)$ есть начальный объем совокупности. Легко видеть, что математическое ожидание $\sum nP_n(t)$ распределения (3.5) совпадает с $x(t)$, так что $x(t)$ описывает не только детерминированный процесс роста, но также и средний объем совокупности в примере 3, б).

Рассмотрим теперь детерминированный процесс роста, в котором скорость роста увеличивается быстрее объема совокупности. Скорости роста, пропорциональной $x^2(t)$, соответствует дифференциальное уравнение

$$dx(t)/dt=\lambda x^2(t), \tag{4.5}$$

решение которого имеет вид

$$x(t)=i/(1-\lambda it). \tag{4.6}$$

Заметим, что $x(t)$ неограниченно возрастает при $t\to 1/(\lambda i)$. Иначе говоря, из предположения о том, что скорость роста увеличивается как квадрат объема совокупности, вытекает бесконечный рост за конечный интервал времени. Аналогично, если λ_n в (3.4) возрастают слишком быстро, то с конечной вероятностью внутри конечного интервала времени имеет место бесконечное число изменений. Точный ответ о том, каковы должны быть условия, при которых происходит такой неограниченный рост, дает следующая теорема.

Теорема. *Для того чтобы* $\sum P_n(t)=1$ *при всех* t, *необходимо и достаточно, чтобы ряд* $\sum \lambda_n^{-1}$ *расходился* [1]).

Доказательство. Положим

$$S_k(t)=P_0(t)+\ldots+P_k(t). \tag{4.7}$$

В силу очевидной монотонности существует предел

$$\mu(t)=\lim_{k\to\infty}[1-S_k(t)]. \tag{4.8}$$

Суммируя дифференциальные уравнения (3.2) по $n=0, \ldots, k$, мы получаем

$$S_k'(t)=-\lambda_k P_k(t). \tag{4.9}$$

---

[1]) Нетрудно видеть, что неравенство $\sum P_n(t)<1$ либо выполняется для всех $t>0$, либо не выполняется ни для одного $t>0$; см. задачу 22.

С учетом начальных условий (3.3) отсюда вытекает, что при $k \geqslant i$

$$1-S_k(t)=\lambda_k \int_0^t P_k(\tau)\,d\tau. \tag{4.10}$$

В силу (4.8) величина в левой части лежит между μ и 1; следовательно,

$$\lambda_k^{-1}\mu(t) \leqslant \int_0^t P_k(s)\,ds \leqslant \lambda_k^{-1}. \tag{4.11}$$

Суммируя по $k=i, \ldots, n$, мы получаем для $n \geqslant i$

$$\mu(t)\,[\lambda_i^{-1}+\ldots+\lambda_n^{-1}] \leqslant \int_0^t S_n(s)\,ds \leqslant \lambda_i^{-1}+\ldots+\lambda_n^{-1}. \tag{4.12}$$

Если $\sum \lambda_n^{-1} < \infty$, то величина в самой правой части неравенства остается ограниченной при $n \to \infty$, и поэтому невозможно, чтобы подынтегральное выражение стремилось к 1 при всех s. Наоборот, если $\sum \lambda_n^{-1} = \infty$, то из первого неравенства мы заключаем, что $\mu(t)=0$ при всех t, и в силу (4.8) отсюда вытекает, что $S_n(t) \to 1$, что и утверждалось. ▶

При вероятностной интерпретации этот критерий представляется весьма разумным. Система проводит некоторое время в начальном состоянии E_0, переходит из него в E_1, остается там на время, переходит в E_2 и т. д. Вероятность $P_0(t)$ того, что время пребывания в E_0 превосходит t, получается из (3.2) равной $P_0(t)=e^{-\lambda_0 t}$. Это время пребывания $\mathbf{T}_0$ является случайной величиной, однако область ее значений есть положительная ось t, и поэтому эта величина формально не укладывается в рамки этой книги. Однако, поскольку переход от геометрического распределения к показательному тривиален, мы можем без особого вреда для себя немного отойти от строгого изложения. Приближение $\mathbf{T}_0$ дискретной случайной величиной с геометрическим распределением показывает, что среднее время пребывания в E_0 естественно определить, как

$$\mathbf{E}(\mathbf{T}_0)=\int_0^\infty te^{-\lambda_0 t}\lambda_0\,dt=\lambda_0^{-1}. \tag{4.13}$$

В тот момент времени, когда система попадает в E_j, состояние E_j берет на себя роль начального состояния, и поэтому такое же заключение применимо и ко времени пребывания $\mathbf{T}_j$ в E_j: *среднее время пребывания в E_j равно* $\mathbf{E}(\mathbf{T}_j)=\lambda_j^{-1}$. Отсюда следует, что $\lambda_0^{-1}+\lambda_1^{-1}+\ldots+\lambda_n^{-1}$ есть средняя продолжительность времени, требующегося системе для того, чтобы пройти через $E_0, E_1, \ldots, E_n$, и мы можем переформулировать критерий из § 4 следующим образом.

Для того чтобы $\sum P_n(t)=1$ для всех t, необходимо и достаточно, чтобы

$$\sum \mathbf{E}(\mathbf{T}_j)=\sum \lambda_j^{-1}=\infty; \tag{4.14}$$

т. е. суммарная средняя продолжительность времени, проведенного в $E_0, E_1, E_2, \ldots$, должна быть бесконечной. Разумеется, $L_0(t)=1-\sum P_n(t)$ есть вероятность того, что система прошла через все состояния до момента времени t.

При такой интерпретации возможность неравенства (4.2) становится понятной. Если среднее время пребывания в E_j равно 2^{-j}, то вероятность того, что система пройдет через все состояния за время $1+2^{-1}+2^{-2}+\ldots=2$, должна быть положительной. Аналогично частица, двигающаяся вдоль оси x с экспоненциально возрастающей скоростью, пройдет всю ось за конечное время.

(Мы вернемся к расходящемуся процессу размножения в примере 9, б).)

§ 5. ПРОЦЕСС РАЗМНОЖЕНИЯ И ГИБЕЛИ

Процесс чистого размножения из § 3 дает удовлетворительное описание радиоактивных превращений, однако он не может служить реалистической моделью для изменений объема совокупности, члены которой могут умирать (или исчезать). Это наводит на мысль об обобщении нашей модели, допускающем переходы из E_n не только в ближайшее сверху состояние E_{n+1}, но и в ближайшее снизу состояние E_{n-1}. (Процессы более общего вида будут определены в § 9.)

Таким образом, мы начнем со следующих постулатов.

Постулаты. *Изменения системы осуществляются только путем переходов из состояний в ближайшие к ним соседние состояния (из E_n в E_{n+1} или в E_{n-1} при $n\geqslant 1$, а из E_0 — только в E_1). Если в момент времени t система находится в состоянии E_n, то вероятность того, что между t и $t+h$ произойдет переход $E_n\to E_{n+1}$, равна $\lambda_n h+o(h)$, а вероятность перехода $E_n\to E_{n-1}$ (если $n\geqslant 1$) равна $\mu_n h+o(h)$. Вероятность более чем одного изменения на протяжении $(t,\ t+h)$ есть $o(h)$.*

Легко приспособить метод § 2 для вывода дифференциальных уравнений для вероятностей $P_n(t)$ того, что система находится в состоянии E_n. Чтобы вычислить $P_n(t+h)$, заметим, что состояние E_n в момент $t+h$ возможно лишь при выполнении одного из следующих условий: 1) в момент t система находится в E_n и между t и $t+h$ не происходит никаких изменений; 2) в момент t система находится в E_{n-1} и происходит переход в E_n; 3) в момент t система находится в E_{n+1} и происходит переход в E_n; 4) между t и $t+h$ происходит два или несколько переходов. По предположению вероятность последнего события есть $o(h)$. Первые три возможности взаимно исключаются, и их вероятности складываются; поэтому

$$P_n(t+h)=P_n(t)\{1-\lambda_n h-\mu_n h\}+\\ +\lambda_{n-1}hP_{n-1}(t)+\mu_{n+1}hP_{n+1}(t)+o(h). \quad (5.1)$$

Перенося член $P_n(t)$ в левую часть и деля обе части уравнения на h, получаем в левой части разностное отношение для $P_n(t)$, и в пределе при $h\to 0$ приходим к

$$P_n'(t)=-(\lambda_n+\mu_n)P_n(t)+\lambda_{n-1}P_{n-1}(t)+\mu_{n+1}P_{n+1}(t). \quad (5.2)$$

Это уравнение справедливо при всех $n \geqslant 1$. Для $n=0$ аналогичным образом получаем

$$P_0'(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t). \tag{5.3}$$

Если начальным состоянием является E_i, то *начальными условиями* будут

$$P_i(0)=1, \quad P_n(0)=0 \text{ при } n \neq i. \tag{5.4}$$

Таким образом, мы видим, что процесс размножения и гибели зависит от бесконечной системы дифференциальных уравнений (5.2)—(5.3) с начальными условиями (5.4). Вопрос о существовании и единственности решения в этом случае отнюдь не тривиален. В процессе чистого размножения система дифференциальных уравнений (3.2) также была бесконечной, но она имела вид рекуррентных соотношений: $P_0(t)$ определялась первым уравнением, а $P_n(t)$ могла быть вычислена по $P_{n-1}(t)$. Новая система (5.2) имеет иной вид, и все $P_n(t)$ должны находиться одновременно. Здесь (как и в некоторых других случаях в этой главе) мы сформулируем свойства решений без доказательства [1]).

Для произвольных заданных коэффициентов $\lambda_n \geqslant 0$, $\mu_n \geqslant 0$ *всегда существует положительное решение* $\{P_n(t)\}$ *системы* (5.2)—(5.4), *такое, что* $\sum P_n(t) \leqslant 1$. *Если коэффициенты ограничены* (*или возрастают достаточно медленно*), *то это решение единственно и удовлетворяет условию регулярности* $\sum P_n(t) = 1$. Однако можно выбрать коэффициенты таким образом, что $\sum P_n(t) < 1$ и будет существовать бесконечно много решений. В последнем случае мы сталкиваемся с явлением, аналогичным изучавшемуся в предыдущем параграфе для процесса чистого размножения. Эта ситуация представляет значительный теоретический интерес, однако читатель может без опасения считать, что во всех практически интересных случаях условия единственности выполнены; в этом случае автоматически $\sum P_n(t) = 1$ (см. § 9).

При $\lambda_0 = 0$ переход $E_0 \to E_1$ невозможен. В терминологии цепей Маркова E_0 является *поглощающим состоянием*, выход из которого невозможен; коль скоро система окажется в E_0, она останется там навсегда. Из (5.3) следует, что в этом случае $P_0'(t) \geqslant 0$, так что $P_0(t)$ монотонно возрастает. Предел ее $P_0(\infty)$ есть вероятность *окончательного поглощения*.

[1]) Простейшее доказательство существования и критерий единственности получается как частный случай из общей теории, развитой автором (см. § 9). Недавно привлекли широкое внимание такие решения уравнений для процесса размножения и гибели, что $\sum P_n(t) < 1$. По этому вопросу см. Lederman W., Reuter G. E., Spectral theory for the differential equations of simple birth and death processes, Philos. Trans. Roy. Soc., London, Ser. A, 246 (1954), 387—391; Karlin S., McGregor J. L., The differential equations of birth-and-death processes and the Stieltjes moment problem, Trans. Amer. Math. Soc., 85 (1957), 489—546; Karlin S., McGregor J. L., The classification of birth and death processes, ibid, 86 (1957), 366—400. См. также Feller W., The birth and death processes as diffusion processes, Journal de Mathématiques Pures et Appliquées, 38(1959), 301—345.

Можно показать (либо используя явный вид решений, либо из общих эргодических теорем для марковских процессов), что в любом случае *пределы*

$$\lim_{t \to \infty} P_n(t) = p_n \tag{5.5}$$

существуют и не зависят от начальных условий (5.4); они удовлетворяют системе линейных уравнений, которая получается из (5.2)—(5.3) при замене производных в левой части нулями.

Соотношения (5.5) напоминают предельные теоремы, выведенные в гл. XV, 7 для обычных цепей Маркова, и это не просто формальное сходство. Эти соотношения становятся интуитивно почти очевидными при сравнении нашего процесса с простой цепью Маркова с переходными вероятностями

$$p_{n,\,n+1}=\lambda_n/(\lambda_n+\mu_n), \qquad p_{n,\,n-1}=\mu_n/(\lambda_n+\mu_n). \tag{5.6}$$

В этой цепи единственными возможными переходами являются $E_n \to E_{n+1}$ и $E_n \to E_{n-1}$, и они имеют те же условные вероятности, что и в нашем процессе; разница между этой цепью и нашим процессом заключается в том, что в последнем изменения могут происходить в произвольные моменты времени, так что число переходов в интервале времени длины t является случайной величиной. Однако при больших t это число заведомо будет большим, и поэтому весьма правдоподобно, что при $t \to \infty$ вероятности $P_n(t)$ будут вести себя так же, как соответствующие вероятности для простой цепи.

Если простая цепь с переходными вероятностями (5.6) невозвратна, то при всех n мы имеем $p_n=0$; если эта цепь эргодична, то p_n определяют стационарное распределение вероятностей. В этом случае (5.5) обычно интерпретируется как «тенденция к устойчивому положению», и это двусмысленное наименование вызвало много путаницы. Следует понимать, что, за исключением того случая, когда E_0 есть поглощающее состояние, случайные флуктуации продолжаются вечно, не ослабевая, и (5.5) показывает лишь, что в конце концов влияние начальных условий исчезает. Сделанные в гл. XV, 7 замечания о статистическом равновесии применимы здесь без изменений.

Главной областью приложений процессов размножения и гибели являются задачи о времени ожидания, задачи о телефонных линиях и т. д.; см. § 6 и 7.

Примеры. а) *Линейный рост.* Предположим, что элементы совокупности могут делиться или умирать. Для любого живого элемента вероятность разделиться на два на протяжении любого короткого интервала времени длины h равна $\lambda h+o(h)$, тогда как соответствующая вероятность гибели равна $\mu h+o(h)$. Здесь λ и μ— две постоянные, характеризующие совокупность. Если между ее элементами нет никакого взаимодействия, то мы приходим к процессу размножения и гибели с $\lambda_n = n\lambda$, $\mu_n = n\mu$. Основные дифференциальные уравнения принимают вид

$$\begin{aligned} &P_0'(t) = \mu P_1(t), \\ &P_n'(t) = -(\lambda+\mu)\, nP_n(t) + \lambda(n-1)\, P_{n-1}(t) + \mu(n+1)\, P_{n+1}(t). \end{aligned} \tag{5.7}$$

Можно найти явные решения[1]) (см. задачи 11—14), однако мы не будем обсуждать этот аспект. Пределы (5.5) здесь существуют и удовлетворяют (5.7) при $P_n'(t) = 0$. Из первого уравнения находим $p_1 = 0$, а из второго уравнения по индукции находим, что $p_n = 0$ при всех $n \geqslant 1$. Если $p_0 = 1$, то мы можем сказать, что вероятность окончательного вымирания равна 1. Если $p_0 < 1$, то из соотношений $p_1 = p_2 = \ldots = 0$ вытекает, что с вероятностью $1 - p_0$ совокупность будет безгранично расти; в конце концов она либо вымрет, либо будет бесконечно возрастать. Чтобы найти вероятность p_0 вымирания, сравним наш процесс с соответствующей цепью Маркова. В нашем случае переходные вероятности (5.6) не зависят от n, и стало быть, мы имеем обычное случайное блуждание, в котором шаги направо и налево имеют вероятности $p = \lambda/(\lambda+\mu)$ и $q = \mu/(\lambda+\mu)$ соответственно. Состояние E_0 является поглощающим. Из классической задачи о разорении (см. гл. XIV, 2) мы знаем, что вероятность вымирания равна 1 при $p \leqslant q$ и равна $(q/p)^i$, если $q < p$ и i есть начальное состояние. Отсюда мы заключаем, что *в нашем процессе вероятность* $p_0 = \lim P_0(t)$ *окончательного вымирания равна* 1 *при* $\lambda \leqslant \mu$ *и равна* $(\mu/\lambda)^i$ *при* $\lambda > \mu$. (Это легко проверить при помощи явного решения; см. задачи 11—14.)

Как и во многих подобных случаях, явный вид решения системы (5.7) довольно сложен, и поэтому хорошо было бы найти среднее и дисперсию распределения $\{P_n(t)\}$ непосредственно из дифференциальных уравнений. Для среднего имеем

$$M(t) = \sum_{n=1}^{\infty} nP_n(t). \tag{5.8}$$

Мы опустим формальное доказательство того, что $M(t)$ конечно, а следующие формальные операции обоснованы (и то, и другое легко следует из вида решения, приведенного в задаче 12). Умножая второе из уравнений (5.7) на n и суммируя по $n = 1, 2, \ldots$, мы обнаруживаем, что члены, содержащие n^2, взаимно уничтожаются,

---

[1]) Систематический подход здесь состоит в выводе уравнения в частных производных для производящей функции $\sum P_n(t)\, s^n$. Более общий процесс, в котором допускается зависимость коэффициентов λ и μ в (5.7) от времени, детально обсуждается в работе Kendall D. G., The generalized "birth and death" process, Ann. Math. Statist., 19 (1948), 1—15. См. также статью того же автора Stochastic processes and population growth, Journal of the Royal Statistical Society, B, 11 (1949), 230—265, где теория обобщается так, что она учитывает распределение возраста элементов в биологических популяциях.

и получаем

$$M'(t) = \lambda \sum (n-1) P_{n-1}(t) - \mu \sum (n+1) P_{n+1}(t) = \\ = (\lambda - \mu) M(t). \qquad (5.9)$$

Это дифференциальное уравнение для $M(t)$. Начальный объем совокупности равен i, и поэтому $M(0)=i$. Стало быть,

$$M(t) = ie^{(\lambda-\mu)t}. \qquad (5.10)$$

Мы видим, что среднее стремится к 0 при $\lambda<\mu$ и к бесконечности при $\lambda>\mu$. Дисперсию распределения $\{P_n(t)\}$ можно вычислить аналогичным образом (см. задачу 14).

б) *Очередь в случае одного канала.* В простейшем случае постоянных коэффициентов $\lambda_n=\lambda$ и $\mu_n=\mu$ процесс размножения и гибели сводится к частному случаю примера 7, б) (задачи об очередях) при $a=1$.

§ 6. ПОКАЗАТЕЛЬНЫЕ ВРЕМЕНА ОБСЛУЖИВАНИЯ [1]

Основная область приложения процессов размножения и гибели связана с расчетом числа телефонных линий [2] и различными типами очередей к телефонам, прилавкам или машинам. Этот тип задач может изучаться на различных уровнях математической абстракции. Метод процесса размножения и гибели является простейшим подходом, однако эта модель основана на математическом упрощении, известном как *предположение о показательных временах обслуживания.* Мы начнем с обсуждения этого основного предположения.

Рассмотрим для конкретности телефонный разговор и предположим, что его продолжительность обязательно равна целому числу секунд. Мы рассматриваем продолжительность разговора как случайную величину $\mathbf{X}$ и считаем известным ее распределение вероятностей $p_n=\mathbf{P}\{\mathbf{X}=n\}$. Тогда телефонная линия представляет собой физическую систему с двумя возможными состояниями «занято» (E_0) и «свободно» (E_1). Когда линия занята, вероятность изменения состояния в течение следующей секунды зависит от того, как долго уже идет разговор. Иначе говоря, прошлое влияет на будущее, и поэтому наш процесс не является марковским (см. гл. XV, 13). Это обстоятельство является источником трудностей, однако, к счастью, существует простое исключение, подробно обсуждавшееся в гл. XIII, 9.

Представим себе, что решение о том, продолжать или нет разговор, принимается каждую секунду бросанием несимметричной монеты. Иначе говоря, со скоростью одно испытание в секунду производится последовательность испытаний Бернулли, которая продол-

---

[1] В оригинале exponential holding times.— *Прим. перев.*

[2] В оригинале trunking in telephone engineering.— *Прим. перев.*

жается до первого успеха. Когда этот первый успех произойдет, разговор окончится. В этом случае общая продолжительность разговора, или «время обслуживания», имеет геометрическое распределение $p_n = q^{n-1}p$. Когда линия занята, вероятность того, что она останется занятой более одной секунды, равна q, а вероятность перехода $E_0 \to E_1$ на следующем шаге равна p. Теперь эти вероятности не зависят от того, как долго была занята линия.

Когда использование дискретного временно́го параметра оказывается нежелательным, приходится работать с непрерывными случайными величинами. Тогда роль геометрического распределения для времен ожидания берет на себя *показательное распределение*. Это единственное распределение, имеющее марковский характер, т. е. наделенное полной потерей памяти. Иначе говоря, вероятность того, что разговор, происходящий в момент времени x, продлится до $x+h$, не зависит от продолжительности предыдущего разговора тогда и только тогда, когда вероятность того, что разговор продлится более t единиц времени, равна $e^{-\lambda t}$. Мы уже встречались с этим «показательным временем обслуживания» как с нулевым членом распределения Пуассона (2.4), т. е. как с временем ожидания первого изменения.

Метод процесса размножения и гибели применим только тогда, когда изучаемые переходные вероятности не зависят от прошлого; для задач о числе занятых линий и об очередях это означает, что все времена обслуживания должны быть показательными. С практической точки зрения это предположение может на первый взгляд показаться довольно искусственным, однако опыт показывает, что оно неплохо описывает реальные явления. В частности, многочисленные измерения показали, что телефонные разговоры внутри города [1]) следуют показательному закону с поразительной степенью точности. Та же ситуация превалирует и в случае других времен обслуживания (например, продолжительности ремонта машин).

Остается охарактеризовать так называемый входной поток (поступающие вызовы, поломки станков и т. д.). Мы предположим, что вероятность поступления вызова в течение любого интервала времени длины h равна λh плюс пренебрежимо малые члены и что вероятностью более чем одного вызова в этом интервале в пределе можно пренебречь. Согласно результатам § 2, это означает, что число поступивших [до момента времени t.— *Перев.*] вызовов имеет распределение Пуассона со средним λt. Мы будем описывать эту ситуацию, говоря, что *входной поток имеет пуассоновский тип с интенсивностью* λ.

Легко проверить описанное свойство показательных времен обслуживания. Обозначим через $u(t)$ вероятность того, что разговор продлится не менее t единиц времени. Вероятность $u(t+s)$ того, что разговор, начавшийся в момент 0, окон-

[1]) Для междугородных разговоров единицей измерения обычно служат три минуты, и поэтому времена обслуживания часто являются кратными трем минутам. При таких обстоятельствах показательное распределение неприменимо.

чится после момента $t+s$, равна вероятности того, что он продлится более t единиц времени, умноженной на условную вероятность того, что разговор продлится еще s дополнительных единиц времени при условии, что его длина превышает t. Если продолжительность предыдущего разговора не оказывает никакого влияния, то последняя условная вероятность должна равняться $u(s)$, т. е. мы должны иметь

$$u(t+s)=u(t)\,u(s). \tag{6.1}$$

Для доказательства сформулированного выше характеристического свойства показательных времен обслуживания было бы достаточно показать, что *монотонные* решения этого функционального уравнения с необходимостью имеют вид $e^{-\lambda t}$. Мы докажем несколько более сильный результат, представляющий самостоятельный интерес [1]).

Теорема. *Пусть u — решение уравнения* (6.1), *определенное при $t>0$ и ограниченное на некотором интервале. Тогда либо $u(t)=0$ при всех t, либо $u(t)=e^{-\lambda t}$, где λ — некоторая постоянная.*

Доказательство. Ясно, что

$$u(a)=u^2(a/2). \tag{6.2}$$

Предположим сперва, что $u(a)=0$ для некоторого значения a. Из (6.2) мы заключаем по индукции, что $u(2^{-n}a)=0$ для всех целых n, а из (6.1) ясно, что из $u(s)=0$ вытекает равенство $u(t)=0$ при всех $t>s$. Поэтому из $u(a)=0$ следует, что u тождественно обращается в нуль. Поскольку (6.2) очевидным образом исключает отрицательные значения u, то остается рассмотреть лишь строго положительные решения уравнения (6.1).

Положим $e^{-\lambda}=u(1)$ и $v(t)=e^{\lambda t}u(t)$. Тогда

$$v(t+s)=v(t)\,v(s)\quad и\quad v(1)=1. \tag{6.3}$$

Мы должны доказать, что отсюда вытекает равенство $v(t)=1$ при всех t. Очевидно, что для произвольных положительных целых чисел m и n

$$v(m/n)=v^m(1/n)=\sqrt[n]{v^m(1)}=1, \tag{6.4}$$

и поэтому $v(s)=1$ для всех рациональных s. Кроме того, если $v(a)=c$, то $v(na)=c^n$ для любого положительного или отрицательного целого числа n. Отсюда следует, что если v принимает некоторое значение $c\neq 1$, то оно принимает также произвольно большие значения. Однако из (6.3) при $t+s=\tau$ видно, что $v(\tau-s)=v(\tau)$ для всех рациональных s. Следовательно, если значение A принимается в некоторой точке τ, то то же самое значение принимается на каждом сколь угодно малом интервале. Стало быть, ограниченность u на каком-либо заданном интервале не допускает значений v, отличных от единицы. ►

§ 7. ОЧЕРЕДИ И ЗАДАЧИ ОБСЛУЖИВАНИЯ

а) *Простейшая задача о телефонных линиях* [2]). Предположим, что имеется бесконечно много линий или каналов и что вероятность окончания разговора между моментами времени t и $t+h$ равна $\mu h+$

---

[1]) Уравнение (6.1) является лишь логарифмическим вариантом записи известного уравнения Гамеля $f(t+s)=f(t)+f(s)$. Мы доказываем, что его решения либо имеют вид at, либо неограниченны на *каждом* интервале. (Известно, что ни одно такое решение не является функцией Бэра, т. е. ни одно такое решение не может быть получено при помощи разложения в ряд или какого-либо иного предельного процесса, отправляющегося от непрерывных функций.)

[2]) Palm C., Intensitätsschwankungen in Fernsprechverkehr, Ericsson Technics (Stockholm), 44 (1943), 1—189, в частности, с. 57. Очереди и задачи обслуживания

$+o(h)$ (показательное время обслуживания). Поступающие вызовы образуют поток пуассоновского типа с параметром λ. Система находится в состоянии E_n, если занято n линий.

Предполагается, конечно, что продолжительности разговоров взаимно независимы. Если занято n линий, то вероятность того, что одна из них освободится в течение времени h, равна $n\mu h+o(h)$. Вероятность прекращения за это время двух или нескольких разговоров является, очевидно, величиной порядка h^2, и поэтому ей можно пренебречь. Вероятность поступления нового вызова равна $\lambda h+$ $+o(h)$. Вероятность нескольких вызовов или поступления вызова и окончания разговора тоже есть $o(h)$. Таким образом, в обозначениях: § 5

$$\lambda_n=\lambda, \quad \mu_n=n\mu. \tag{7.1}$$

Основные дифференциальные уравнения (5.2) — (5.3) принимают вид

$$\begin{aligned} P_0'(t) &= -\lambda P_0(t)+\mu P_1(t), \\ P_n'(t) &= -(\lambda+n\mu)P_n(t)+\lambda P_{n-1}(t)+(n+1)\mu P_{n+1}(t) \end{aligned} \tag{7.2}$$

при $n\geqslant 1$. Явные решения могут быть получены путем вывода дифференциального уравнения в частных производных для производящей функции (см. задачу 15). Мы определим лишь величины $p_n=$ $=\lim P_n(t)$ из (5.5). Они удовлетворяют уравнениям

$$\begin{aligned} \lambda p_0 &= \mu p_1, \\ (\lambda+n\mu)p_n &= \lambda p_{n-1}+(n+1)\mu p_{n+1}. \end{aligned} \tag{7.3}$$

По индукции находим $p_n=p_0(\lambda/\mu)^n/n!$, и отсюда

$$p_n=e^{-\lambda/\mu}(\lambda/\mu)^n/n!. \tag{7.4}$$

Таким образом, *предельное распределение является распределением Пуассона с параметром λ/μ. Оно не зависит от начального состояния.*

Легко найти среднее $M(t)=\sum nP_n(t)$. Умножая n-е уравнение (7.2) на n, суммируя и учитывая, что $P_n(t)$ в сумме дают единицу, мы получаем

$$M'(t)=\lambda-\mu M(t). \tag{7.5}$$

Если начальным является состояние E_i, то $M(0)=i$ и

$$M(t)=(\lambda/\mu)(1-e^{-\mu t})+ie^{-\mu t}. \tag{7.6}$$

Читатель может проверить, что в частном случае $i=0$ величины $P_n(t)$ даются в точности распределением Пуассона со средним $M(t)$.

для телефонных станций изучались задолго до появления теории стохастических процессов и оказали стимулирующее влияние на развитие этой теории. В частности, замечательная работа Пальма оказалась полезной на протяжении многих лет. Самые ранние работы в этой области принадлежат А. К. Эрлангу (1878—1929); см. Brockmeyer E., Halström H. L., Jensen Arne, The life and works of A. K. Erlang, Transactions of the Danish Academy Technical Sciences, No. 2, Copenhagen, 1948. Ценная основополагающая работа была независимо проделана Фраем, книга которого (Fry T. C., Probability and its engineering uses, New York, Van Nostrand, 1928) много сделала для развития приложений теории вероятностей в технике.

б) *Очереди в случае конечного числа каналов* [1]). Видоизменим теперь последний пример, чтобы получить более реалистичную модель. Предположения здесь те же самые, за исключением того, что *число a линий или каналов конечно. Если все a каналов заняты, то каждый новый вызов встает в очередь и ожидает до тех пор, пока не освободится какой-либо канал.* Это означает, что все линии имеют *общую* очередь.

Слово «линия» может быть заменено на *стойку* в почтовом отделении, а «разговор» — на *обслуживание*. Практически мы рассматриваем общую задачу об очереди в предположении, что клиент должен ожидать лишь тогда, когда все a каналов заняты.

Мы говорим, что *система находится в состоянии E_n, если ровно n клиентов либо обслуживаются, либо стоят в очереди.* Такая очередь существует только при $n>a$, и тогда в ней стоят $n-a$ клиентов.

До тех пор пока хотя бы один канал свободен, ситуация здесь точно такая же, как и в предыдущем примере. Однако если система находится в состоянии E_n при $n>a$, то идет только a разговоров, и поэтому $\mu_n=a\mu$ при $n\geqslant a$. Стало быть, основная система дифференциальных уравнений состоит из уравнений (7.2) для $n<a$ и из уравнений

$$P_n'(t)=-(\lambda+a\mu)P_n(t)+\lambda P_{n-1}(t)+a\mu P_{n+1}(t) \tag{7.7}$$

для $n\geqslant a$.

В частном случае одного канала ($a=1$) эти уравнения сводятся к уравнениям для процесса размножения и гибели с коэффициентами, не зависящими от n.

Пределы $p_n=\lim P_n(t)$ удовлетворяют (7.3) при $n<a$ и

$$(\lambda+a\mu)p_n=\lambda p_{n-1}+a\mu p_{n+1} \tag{7.8}$$

при $n\geqslant a$. Последовательно находим

$$p_n=p_0(\lambda/\mu)^n/n!,\qquad n\leqslant a, \tag{7.9}$$

$$p_n=p_0(\lambda/\mu)^n/(a!a^{n-a}),\qquad n\geqslant a. \tag{7.10}$$

Ряд $\sum(p_n/p_0)$ сходится только тогда, когда

$$\lambda/\mu<a. \tag{7.11}$$

Следовательно, предельное распределение $\{p_k\}$ не может существовать при $\lambda\geqslant a\mu$. *В этом случае $p_n=0$ для всех n, и это означает, что очередь будет безгранично возрастать.* С другой стороны, если выполнено (7.11), то мы можем определить p_0 так, чтобы $\sum p_n=1$. С помощью явных выражений для $P_n(t)$ можно показать, что получаемые таким образом p_n действительно представляют собой *предельное распределение* для $P_n(t)$. Табл. 1 дает численную иллюстрацию для $a=3$, $\lambda/\mu=2$.

---

[1]) Колмогоров А. Н. О проблеме ожидания.— Математический сборник, 1931, 38 : 1—2, 101—106. Относительно аналогичных процессов см. задачи 6—8 и 20,

Таблица 1

Предельные вероятности в случае $a=3$ каналов и $\lambda/\mu=2$

n	0	1	2	3	4	5	6	7
Число занятых каналов	0	1	2	3	3	3	3	3
Число ожидающих клиентов	0	0	0	0	1	2	3	4
p_n	0,1111	0,2222	0,2222	0,1481	0,09888	0,0658	0,0439	0,0293

в) *Обслуживание станков* [1]). Для ориентировки мы начнем с простейшего случая и обобщим его в следующем примере. Задача состоит в следующем.

Рассмотрим станки-автоматы, которые обычно не требуют ухода за исключением того, что они могут сломаться и потребовать обслуживания. Время, требующееся для обслуживания станка, вновь считается случайной величиной с показательным распределением. Иначе говоря, станок характеризуется двумя постоянными λ и μ со следующими свойствами. Если в момент времени t станок находится в рабочем состоянии, то вероятность того, что он потребует обслуживания до момента $t+h$, равняется λh плюс члены, пренебрежимые в пределе при $h\to 0$. Наоборот, когда станок находится на обслуживании, вероятность того, что время обслуживания окончится до момента $t+h$ и станок вернется в рабочее состояние, равняется $\mu h+o(h)$. Для производительного станка λ должно быть сравнительно малым, а μ — сравнительно большим. Отношение λ/μ называется *коэффициентом обслуживания*.

Мы предполагаем, что *m независимо работающих станков с одними и теми же параметрами λ и μ обслуживаются одним рабочим*. Сломавшийся станок обслуживается немедленно, если рабочий не занят обслуживанием другого станка, а в последнем случае образуется очередь. Мы говорим, что *система находится в состоянии E_n, если не работает n станков*. При $1\leqslant n\leqslant m$ это означает, что один станок находится на обслуживании, а $n-1$ — в очереди; в состоянии E_0 все станки работают, и рабочий не занят.

Переход $E_n\to E_{n+1}$ вызывается поломкой одного из $m-n$ работающих станков, тогда как переход $E_n\to E_{n-1}$ происходит тогда, когда обслуживаемый станок возвращается в рабочее состояние. Следовательно, мы имеем процесс размножения и гибели с коэффи-

[1]) Примеры в) и г), включающие численные иллюстрации, взяты из статьи Пальма «Распределение числа рабочих при обслуживании автоматов» (на шведском языке) из Industritidningen Norden, 75 (1947), 75—80, 90—94, 119—123. Пальм приводит таблицы и графики для наиболее экономичного числа рабочих.

циентами

$$\lambda_n=(m-n)\lambda,\quad \mu_0=0,\quad \mu_1=\mu_2=\ldots\mu_m=\mu. \tag{7.12}$$

При $1\leqslant n\leqslant m-1$ основные дифференциальные уравнения (5.2) принимают вид

$$P'_n(t)=-\{(m-n)\lambda+\mu\}P_n(t)+ +(m-n+1)\lambda P_{n-1}(t)+\mu P_{n+1}(t), \tag{7.13}$$

в то время как для крайних состояний $n=0$ и $n=m$

$$P'_0(t)=-m\lambda P_0(t)+\mu P_1(t),$$
$$P'_m(t)=-\mu P_m(t)+\lambda P_{m-1}(t). \tag{7.13a}$$

Мы пришли к конечной системе дифференциальных уравнений, которая может быть решена обычными методами. Пределы $p_n = \lim P_n(t)$ определяются из уравнений

$$m\lambda p_0=\mu p_1,$$
$$\{(m-n)\lambda+\mu\}p_n=(m-n+1)\lambda p_{n-1}+\mu p_{n+1}, \tag{7.14}$$
$$\mu p_m=\lambda p_{m-1},$$

которые можно привести к рекуррентному виду:

$$(m-n)\lambda p_n=\mu p_{n+1}. \tag{7.15}$$

Подставляя последовательно $n=m-1,\ m-2,\ \ldots,\ 1,\ 0$, получаем

$$p_{m-k}=(1/k!)\,(\mu/\lambda)^k\,p_m.$$

Оставшаяся неизвестной постоянная p_m может быть найдена из того условия, что p_j в сумме дают единицу. Результат известен как *формула Эрланга*:

$$p_m=\{1+(1/1!)\,(\mu/\lambda)^1+\ldots+(1/m!)\,(\mu/\lambda)^m\}^{-1}. \tag{7.16}$$

Типичные численные значения приведены в табл. 2.

Таблица 2

Формула Эрланга
Вероятности p_n для случая
$\lambda/\mu=0{,}1,\ m=6$

n	Число станков в очереди	p_n	n	Число станков в очереди	p_n
0	0	0,4845	4	3	0,0175
1	0	0,2907	5	4	0,0035
2	1	0,1454	6	5	0,0003
3	2	0,0582			

Вероятность p_0 можно интерпретировать как вероятность того, что рабочий будет свободен (в примере из табл. 2 он должен быть свободен примерно половину всего времени). *Среднее число станков в очереди равно*

$$w = \sum_{k=1}^{m} (k-1) p_k = \sum_{k=1}^{m} kp_k - (1-p_0). \tag{7.17}$$

Эту величину можно вычислить суммированием соотношений (7.15) при $n=0, 1, \ldots, m$. Используя тот факт, что p_n в сумме составляют единицу, мы получаем

$$m\lambda - \lambda w - \lambda(1-p_0) = \mu(1-p_0),$$

или

$$w = m - [(\lambda+\mu)/\lambda](1-p_0). \tag{7.18}$$

В примере из табл. 2 мы имеем $w=6\cdot 0{,}0549$. Таким образом, 0,0549 есть средний вклад одного станка в очередь.

г) *Продолжение: несколько рабочих.* Мы не будем менять основные предположения предыдущей задачи, за исключением того, что *теперь m станков обслуживаются r рабочими* ($r<m$). Таким образом, при $n\leqslant r$ состояние E_n означает, что $r-n$ рабочих не занято, n станков обслуживаются и в очереди на ремонт нет ни одного станка. При $n>r$ состояние E_n означает, что r станков обслуживаются и $n-r$ станков стоят в очереди. Мы можем воспользоваться здесь рассуждениями из предыдущего примера с тем исключением, что (7.12) должно быть, очевидно, заменено на

$$\begin{aligned} &\lambda_0 = m\lambda, && \mu_0 = 0, \\ &\lambda_n = (m-n)\lambda, && \mu_n = n\mu, \quad 1 \leqslant n \leqslant r, \\ &\lambda_n = (m-n)\lambda, && \mu_n = r\mu, \quad r \leqslant n \leqslant m. \end{aligned} \tag{7.19}$$

Мы не будем выписывать основную систему дифференциальных уравнений, а ограничимся лишь уравнениями для предельных вероятностей p_n. При $1\leqslant n<r$

$$\{(m-n)\lambda + n\mu\} p_n = (m-n+1)\lambda p_{n-1} + (n+1)\mu p_{n+1}, \tag{7.20а}$$

а при $r \leqslant n \leqslant m$

$$\{(m-n)\lambda + r\mu\} p_n = (m-n+1)\lambda p_{n-1} + r\mu p_{n+1}. \tag{7.20б}$$

При $n=0$, очевидно, $m\lambda p_0 = \mu p_1$. Это равенство определяет отношение p_1/p_0, и из (7.20а) мы получаем по индукции, что при $n<r$

$$(n+1)\mu p_{n+1} = (m-n)\lambda p_n; \tag{7.21}$$

наконец, при $n \geqslant r$ мы видим из (7.20б), что

$$r\mu p_{n+1} = (m-n)\lambda p_n. \tag{7.22}$$

Эти уравнения позволяют последовательно вычислить все отношения p_n/p_0. Наконец, p_0 определяется из условия $\sum p_k = 1$. Значения, приведенные в табл. 3, получены именно таким способом.

Таблица 3

Вероятности p_n для случая $\lambda/\mu = 0,1$, $m = 20$, $r = 3$

n	Число обслуживаемых станков	Число станков в очереди	Число незанятых рабочих	p_n
0	0	0	3	0,13625
1	1	0	2	0,27250
2	2	0	1	0,25888
3	3	0	0	0,15533
4	3	1	0	0,08802
5	3	2	0	0,04694
6	3	3	0	0,02347
7	3	4	0	0,01095
8	3	5	0	0,00475
9	3	6	0	0,00190
10	3	7	0	0,00070
11	3	8	0	0,00023
12	3	9	0	0,00007

При сравнении табл. 2 и 3 выясняются удивительные факты. Эти таблицы относятся к одним и тем же станкам ($\lambda/\mu = 0,1$), однако во втором случае мы имеем $m = 20$ станков и $r = 3$ рабочих. Число станков на одного рабочего возросло с 6 до $6^2/_3$, и все же станки обслуживаются более эффективно. Определим *коэффициент простоя для станков*

$$\frac{w}{m} = \frac{\text{Среднее число станков в очереди}}{\text{Общее число станков}} \tag{7.23}$$

и коэффициент простоя для рабочих

$$\frac{\rho}{r} = \frac{\text{Среднее число незанятых рабочих}}{\text{Общее число рабочих}}. \tag{7.24}$$

Для практических целей мы можем отождествить вероятности $P_n(t)$ с их пределами p_n. Тогда в случае табл. 3 мы имеем $w = p_4 + 2p_5 + 3p_6 + \ldots + 17p_{20}$ и $\rho = 3p_0 + 2p_1 + p_2$. Табл. 4 убе-

Таблица 4

Сравнение эффективности двух систем, рассмотренных в примерах «в» и «г»

	«в»	«г»
Число станков	6	20
Число рабочих	1	3
Число станков на одного рабочего	6	$6^2/_3$
Коэффициент простоя для рабочих	0,4845	0,4042
Коэффициент простоя для станков	0,0549	0,01694

дительно доказывает, что для наших конкретных станков (с $\lambda/\mu=0{,}1$) *иметь трех рабочих на двадцать станков намного экономичнее, чем одного рабочего на шесть станков.*

д) *Задача о снабжении энергией*[1]. Одна электрическая цепь снабжает энергией a сварщиков, которые используют эту энергию лишь время от времени. Если в момент времени t сварщик использует энергию, то вероятность того, что он перестанет пользоваться ею до момента $t+h$, равна $\mu h+o(h)$; если в момент времени t электроэнергия ему не требуется, то вероятность того, что она потребуется ему до момента $t+h$, равна $\lambda h+o(h)$. Сварщики работают независимо друг от друга.

Мы говорим, что система находится в состоянии E_n, если электроэнергию используют n сварщиков. Таким образом, мы имеем лишь конечное число состояний $E_0, \ldots, E_a$.

Если система находится в состоянии E_n, то $a-n$ сварщиков не потребляют электроэнергию, и вероятность возникновения на протяжении интервала времени длины h новой потребности в энергии равна $(a-n)\lambda h+o(h)$; с другой стороны, вероятность того, что один из n сварщиков перестанет пользоваться энергией, равна $n\mu h+$ $+o(h)$. Следовательно, мы имеем процесс размножения и гибели с

$$\lambda_n=(a-n)\lambda, \quad \mu_n=n\mu, \quad 0\leqslant n\leqslant a. \tag{7.25}$$

Основные дифференциальные уравнения принимают вид

$$\begin{aligned} P_0'(t) &= -a\lambda P_0(t)+\mu P_1(t), \\ P_n'(t) &= -\{n\mu+(a-n)\lambda\}P_n(t)+(n+1)\mu P_{n+1}(t)+ \\ &\quad +(a-n+1)\lambda P_{n-1}(t), \\ P_a' &= -a\mu P_a(t)+\lambda P_{a-1}(t). \end{aligned} \tag{7.26}$$

(Здесь $1\leqslant n\leqslant a-1$.) Легко проверить, что *предельные вероятности даются биномиальным распределением*

$$p_n=\binom{a}{n}\left(\frac{\lambda}{\lambda+\mu}\right)^n\left(\frac{\mu}{\lambda+\mu}\right)^{a-n}, \tag{7.27}$$

результат, который можно было бы здесь интуитивно ожидать (явные выражения для $P_n(t)$ приведены в задаче 17).

§ 8. ОБРАТНЫЕ (ОБРАЩЕННЫЕ В ПРОШЛОЕ) УРАВНЕНИЯ

В предыдущих параграфах мы изучали вероятности $P_n(t)$ того, что система в момент времени t находится в состоянии E_n. Это обозначение удобно, однако оно может и ввести в заблуждение, так

---

[1]) Этот пример был подсказан задачей, изучавшейся (хотя и неудачно) в работе Adler H. A., Miller K. W., A new approach to probability problems in electrical engineering, Transactions of the American Institute of Electrical Engineers, 65 (1946), 630—632.

как в нем не указывается начальное состояние E_i системы в момент времени нуль. Для дальнейшего развития нашей теории предпочтительнее вернуться к обозначениям § 1 и воспользоваться *переходными вероятностями*. В соответствии с этим мы обозначим через $P_{in}(t)$ (условную) вероятность состояния E_n в момент времени $t+s$ при условии, что в момент s система находилась в состоянии E_i. Мы будем по-прежнему обозначать через $P_n(t)$ (безусловную) вероятность состояния E_n в момент времени t. Если задано начальное состояние E_i, то безусловная вероятность $P_n(t)$ совпадает с $P_{in}(t)$, однако, когда начальное состояние выбирается в соответствии с распределением вероятностей $\{a_i\}$, мы имеем

$$P_n(t)=\sum_i a_i P_{in}(t). \tag{8.1}$$

Для рассматривавшихся до сих пор частных процессов мы показали, что при фиксированном i переходные вероятности $P_{in}(t)$ *удовлетворяют основным дифференциальным уравнениям* (3.2) *и* (5.2). Индекс i появляется только в начальных условиях, а именно

$$P_{in}(0)=\begin{cases} 1 & \textit{при } n=i, \\ 0 & \textit{в противном случае}. \end{cases} \tag{8.2}$$

Чтобы подготовиться к теории более общих процессов, мы покажем сейчас, что эти же переходные вероятности удовлетворяют также и второй системе дифференциальных уравнений. Для объяснения основных идей начнем с процесса чистого размножения из § 3. При выводе дифференциальных уравнений (3.2) мы продолжали интервал времени $(0,\ t)$ до $(0,\ t+h)$ и рассматривали возможные изменения на протяжении короткого промежутка времени $(t,\ t+h)$. Но мы могли таким же образом продолжать интервал $(0,\ t)$ и в направлении прошлого и рассматривать изменения на протяжении $(-h,\ 0)$. Таким образом мы получаем новую систему дифференциальных уравнений, в которой остается фиксированным n (вместо i). Действительно, переход из E_i в момент времени $-h$ в состояние E_n в момент времени t может произойти тремя взаимоисключающими способами: 1) между $-h$ и 0 нет никаких скачков, и система переходит в E_n из состояния E_i в момент времени 0; 2) между $-h$ и 0 происходит ровно один скачок, и система переходит из E_{i+1} в момент времени 0 в состояние E_n в момент t; 3) между $-h$ и 0 происходит более одного скачка. Вероятность первой возможности равна $1-\lambda_i h+o(h)$, второй $\lambda_i h+o(h)$, тогда как третья возможность имеет вероятность $o(h)$. Как и в § 2 и 3, мы заключаем, что

$$P_{in}(t+h)=P_{in}(t)(1-\lambda_i h)+P_{i+1,\,n}(t)\lambda_i h+o(h). \tag{8.3}$$

Следовательно, при $i\geqslant 0$ новая основная система принимает вид

$$P'_{in}(t)=-\lambda_i P_{in}(t)+\lambda_i P_{i+1,\,n}(t). \tag{8.4}$$

Эти уравнения называются *обратными уравнениями*, а уравнения (3.2) называются для различия *прямыми уравнениями*. Начальные условия имеют вид (8.2). (Интуитивно следовало бы ожидать, что

$$P_{in}(t)=0 \quad \text{при} \quad n<i, \tag{8.5}$$

однако существуют патологические исключения; см. пример 9, б).)

В случае процесса размножения и гибели основные *прямые уравнения* (при фиксированном i) имеют вид (5.2) — (5.3). То же рассуждение, которое привело к (8.4), приводит теперь к соответствующим *обратным уравнениям*

$$P'_{in}(t)=-(\lambda_i+\mu_i)P_{in}(t)+\lambda_i P_{i+1,\,n}(t)+\mu_i P_{i-1,\,n}(t). \tag{8.6}$$

Должно быть ясно, что прямые и обратные уравнения не являются независимыми друг от друга: решение обратных уравнений с начальными условиями (8.2) автоматически удовлетворяет прямым уравнениям, за исключением тех редких случаев, когда решение не единственно.

Пример. *Пуассоновский процесс.* В § 2 мы интерпретировали выражение Пуассона (2.4) как вероятность того, что на протяжении произвольного интервала времени длины t поступило ровно n вызовов. Будем говорить, что в момент времени t система находится в состоянии E_n, если за интервал времени от 0 до t поступило в точности n вызовов. Переход из E_i в момент времени t_1 в состояние E_n в момент t_2 означает, что между t_1 и t_2 поступило $n-i$ вызовов. Это возможно лишь при $n\geqslant i$, и поэтому для переходных вероятностей пуассоновского процесса мы имеем

$$\begin{aligned} P_{in}(t) &= e^{-\lambda t}(\lambda t)^{n-i}/(n-i)!, && \text{если } n\geqslant i, \\ P_{in}(t) &= 0, && \text{если } n<i. \end{aligned} \tag{8.7}$$

Они удовлетворяют прямым уравнениям

$$P'_{in}(t)=-\lambda P_{in}(t)+\lambda P_{i,\,n-1}(t), \tag{8.8}$$

а также обратным уравнениям

$$P'_{in}(t)=-\lambda P_{in}(t)+\lambda P_{i+1,\,n}(t). \tag{8.9}$$

▶

§ 9. ПРОЦЕССЫ ОБЩЕГО ВИДА

До сих пор наша теория ограничивалась процессами, в которых непосредственные переходы из состояния E_n возможны только в соседние состояния E_{n+1} и E_{n-1}. Кроме того, эти процессы были однородными по времени, т. е. переходные вероятности $P_{in}(t)$ были одними и теми же для всех интервалов времени длины t. Теперь мы рассмотрим процессы более общего вида, в которых оба этих предположения опущены.

Как и в теории обычных цепей Маркова, мы допустим непосредственные переходы из любого состояния E_i в любое состояние E_n. Переходные вероятности могут меняться с течением времени. Это вынуждает указывать для любого интервала времени оба его конца, а не только его длину. В соответствии с этим мы будем обозначать через $P_{in}(\tau, t)$ *условную вероятность того, что система в момент времени t находится в состоянии E_n при условии, что в предшествующий ему момент τ состоянием системы было E_i*. Символ $P_{in}(\tau, t)$ утрачивает смысл, если не выполняется условие $\tau<t$. Если процесс однороден по времени, то $P_{in}(\tau, t)$ зависит лишь от разности $t-\tau$, и вместо $P_{in}(\tau, \tau+t)$ (которое не зависит тогда от τ) мы можем писать $P_{in}(t)$.

В § 1 мы видели, что переходные вероятности однородного по времени марковского процесса удовлетворяют *уравнению Колмогорова — Чэпмена*

$$P_{in}(s+t)=\sum_{\nu} P_{i\nu}(s)\, P_{\nu n}(t). \tag{9.1a}$$

Для неоднородных процессов аналогичное тождество имеет вид

$$P_{in}(\tau, t)=\sum_{\nu} P_{i\nu}(\tau, s)\, P_{\nu n}(s, t) \tag{9.1б}$$

и справедливо при $\tau<s<t$. Это соотношение выражает тот факт, что переход из состояния E_i в момент времени τ в состояние E_n в момент t происходит через некоторое состояние E_ν в промежуточный момент времени s, а для марковских процессов вероятность $P_{\nu n}(s, t)$ перехода из E_ν в E_n не зависит от предшествующего состояния E_i. Стало быть, переходные вероятности марковского процесса со счетным множеством состояний являются решениями уравнения Колмогорова—Чэпмена (9.1б), удовлетворяющими дополнительным условиям

$$P_{ik}(\tau, t)\geqslant 0, \quad \sum_k P_{ik}(\tau, t)=1. \tag{9.2}$$

Мы примем без доказательства тот факт, что и, обратно, каждое такое решение представляет собой переходные вероятности некоторого марковского процесса [1]). Отсюда следует, что основная задача

[1]) Понятие марковского процесса включает требование, чтобы при заданном состоянии E_ν в момент времени s развитие процесса до момента s не оказывало бы никакого влияния на его будущую эволюцию. Как было отмечено в § 1, уравнение Колмогорова—Чэпмена отражает это требование лишь частично, поскольку в него входят лишь один момент $\tau<s$ и один момент $t>s$. На остававшийся долгое время нерешенным вопрос о существовании немарковских процессов, переходные вероятности которых удовлетворяют (9.1), получен теперь утвердительный ответ; простейший из известных таких процессов однороден по времени и имеет лишь три возможных состояния E_j; см. Feller W., Ann. Math. Statist., 30 (1959), 1252—1253. [См. также пример

теории марковских процессов состоит в нахождении всех решений уравнения Колмогорова — Чэпмена, удовлетворяющих условиям (9.2).

Главная цель настоящего параграфа состоит в том, чтобы показать, что постулаты процесса размножения и гибели допускают естественное обобщение, разрешающее произвольные непосредственные переходы $E_i \to E_j$. Из этих постулатов мы выведем две системы обыкновенных дифференциальных уравнений, которые будут названы соответственно прямыми и обратными уравнениями. При обычных обстоятельствах каждая из двух этих систем определяет переходные вероятности единственным образом. Прямые уравнения с вероятностной точки зрения являются более естественными, однако, как ни странно, при их выводе требуются более сильные и менее интуитивно ясные предположения.

В однородном по времени процессе размножения и гибели из § 5 исходные постулаты относились к поведению переходных вероятностей $P_{jk}(h)$ при малых h; там просто требовалось существование в нуле производных P'_{jk}. В случае неоднородных процессов мы наложим те же условия на $P_{jk}(t, t+x)$, рассматриваемые как функции от x. Эти производные будут иметь аналогичную вероятностную интерпретацию, однако теперь они будут функциями от t.

Предположение 1. *Каждому состоянию E_n соответствует непрерывная функция $c_n(t) \geqslant 0$, такая, что при $h \to 0$*

$$[1 - P_{nn}(t, t+h)]/h \to c_n(t). \tag{9.3}$$

Предположение 2. *Каждой паре состояний E_j, E_k при $j \neq k$ соответствуют переходные вероятности $p_{jk}(t)$ (зависящие от времени), такие, что при $h \to 0$*

$$P_{jk}(t, t+h)/h \to c_j(t)\, p_{jk}(t), \quad j \neq k. \tag{9.4}$$

Величины $p_{jk}(t)$ непрерывны по t, и для любых фиксированных t и j

$$\sum_k p_{jk}(t) = 1, \quad p_{jj}(t) = 0. \tag{9.5}$$

Вероятностная интерпретация (9.3) очевидна: если в момент времени t система находится в состоянии E_n, то вероятность того, что между t и $t+h$ произойдет изменение, равна $c_n(t)h + o(h)$. Коэффициент $p_{jk}(t)$ можно интерпретировать как условную вероятность того, что *если* между t и $t+h$ система меняет свое состояние E_j, то

гл. XV, 13, е). — *Перев.*] Подобные процессы, однако, имеют довольно патологический характер, и их существование не противоречит тому утверждению, что *каждое* решение уравнения Колмогорова — Чэпмена, удовлетворяющее (9.2), соответствует (единственным образом) некоторому марковскому процессу.

это изменение переводит систему из E_j в E_k. В процессе размножения и гибели $c_n(t)=\lambda_n+\mu_n$,

$$p_{j,\ j+1}(t)=\lambda_j/(\lambda_j+\mu_j), \qquad p_{j,\ j-1}(t)=\mu_j/(\lambda_j+\mu_j) \tag{9.6}$$

и $p_{jk}(t)=0$ для всех других комбинаций j и k. Для каждого фиксированного t величины $p_{jk}(t)$ можно интерпретировать как переходные вероятности цепи Маркова.

Двух приведенных предположений достаточно для вывода системы обратных уравнений для $P_{jk}(\tau,\ t)$, однако для вывода прямых уравнений потребуется дополнительное предположение.

Предположение 3. *При фиксированном k переход к пределу в* (9.4) *равномерен по j.*

Необходимость этого предположения представляет значительный теоретический интерес и будет обсуждаться нами несколько позже.

Перейдем теперь к выводу дифференциальных уравнений для $P_{ik}(\tau,\ t)$ как функций от t и k (прямые уравнения). Из (9.1) имеем

$$P_{ik}(\tau,\ t+h)=\sum_j P_{ij}(\tau,\ t)\,P_{jk}(t,\ t+h). \tag{9.7}$$

Подставив вместо $P_{kk}(t,\ t+h)$ в правой части его выражение из (9.3), мы получим

$$\frac{P_{ik}(\tau,\ t+h)-P_{ik}(\tau,\ t)}{h}=$$
$$=-c_k(t)\,P_{ik}(\tau,\ t)+h^{-1}\sum_{j\neq k}P_{ij}(\tau,\ t)\,P_{jk}(t,\ t+h)+\ldots, \tag{9.8}$$

где опущенные члены стремятся к 0 вместе с h, а суммирование проводится по всем j, за исключением $j=k$. Применим теперь (9.4) к членам этой суммы. Поскольку (по предположению 3) переход к пределу равномерен по j, правая часть имеет предел. Следовательно, левая часть также имеет предел, а это означает, что $P_{ik}(\tau,\ t)$ имеет частную производную по t и

$$\partial P_{ik}(\tau,\ t)/\partial t=-c_k(t)\,P_{ik}(\tau,\ t)+\sum_j P_{ij}(\tau,\ t)\,c_j(t)\,p_{jk}(t). \tag{9.9}$$

Это основная система прямых дифференциальных уравнений. Здесь i и τ фиксированы, так что мы имеем (несмотря на формальное появление частной производной) систему *обыкновенных дифференциальных уравнений* [1]) для функций $P_{ik}(\tau,\ t)$. Параметры i и τ входят

---

[1]) В стандартной форме имеющей вид

$$x_k'(t)=-c_k(t)\,x_k(t)+\sum_j x_j(t)\,c_j(t)\,p_{jk}(t).$$

только в *начальное условие*

$$P_{ik}(\tau, \tau) = \begin{cases} 1 & \text{при } k = i, \\ 0 & \text{в противном случае.} \end{cases} \tag{9.10}$$

Обратимся теперь к обратным уравнениям. В них постоянными остаются k и t, так что переходные вероятности $P_{ik}(\tau, t)$ рассматриваются как функции от начальных данных E_i и τ. При формулировке наших исходных предположений эти начальные переменные были фиксированы, однако для вывода обратных уравнений предпочтительнее сформулировать те же условия по отношению к интервалу времени от $t-h$ до t. Иначе говоря, естественнее начать со следующей *альтернативной формы условий* (9.3) *и* (9.4):

$$[1 - P_{nn}(t-h, t)]/h \to c_n(t), \tag{9.3a}$$

$$P_{jk}(t-h, t)/h \to c_j(t)\, p_{jk}(t), \quad j \neq k. \tag{9.4a}$$

Нетрудно доказать эквивалентность этих двух наборов условий (или представить их в единой форме), но мы удовлетворимся тем, что начнем с этой альтернативной формы. Замечательная черта последующего вывода состоит в том, что нам не понадобится никакого аналога предположения 3.

Согласно уравнению Колмогорова—Чэпмена (9.1б),

$$P_{ik}(\tau - h, t) = \sum_{\nu} P_{i\nu}(\tau - h, \tau)\, P_{\nu k}(\tau, t), \tag{9.11}$$

и, используя (9.3а) при $n = i$, мы получаем

$$[P_{ik}(\tau - h, t) - P_{ik}(\tau, t)]/h = \\ = -c_i(\tau)\, P_{ik}(\tau, t) + h^{-1} \sum_{\nu \neq i} P_{i\nu}(\tau - h, \tau)\, P_{\nu k}(\tau, t) + \dots \tag{9.12}$$

Здесь $h^{-1} P_{i\nu}(\tau - h, \tau) \to c_i(\tau)\, p_{i\nu}(\tau)$, и переход к пределу в сумме в правой части (9.12) всегда равномерен. Действительно, если $N > i$, то

$$0 \leqslant h^{-1} \sum_{\nu = N+1}^{\infty} P_{i\nu}(\tau - h, \tau)\, P_{\nu k}(\tau, t) \leqslant h^{-1} \sum_{\nu = N+1}^{\infty} P_{i\nu}(\tau - h, \tau) = \\ = h^{-1} \left\{ 1 - \sum_{\nu = 0}^{N} P_{i\nu}(\tau - h, \tau) \right\} \to c_i(\tau) \left\{ 1 - \sum_{\nu = 0}^{N} p_{i\nu}(\tau) \right\}. \tag{9.13}$$

В силу условия (9.5) правая часть здесь может быть сделана сколь угодно малой за счет выбора достаточно большого N. Отсюда следует, что в (9.12) возможен почленный переход к пределу, и мы получаем

$$\partial P_{ik}(\tau, t)/\partial \tau = c_i(\tau)\, P_{ik}(\tau, t) - c_i(\tau) \sum_{\nu} p_{i\nu}(\tau)\, P_{\nu k}(\tau, t). \tag{9.14}$$

Это и есть основные *обратные дифференциальные уравнения*. Здесь k и t фигурируют как фиксированные параметры, и поэтому (9.14) представляет собой систему обыкновенных дифференциальных уравнений. Параметры k и t входят лишь в *начальные условия*

$$P_{ik}(t,\ t)=\begin{cases}1 & \text{при } i=k,\\ 0 & \text{в противном случае.}\end{cases} \tag{9.15}$$

Пример. а) *Обобщенный пуассоновский процесс.* Рассмотрим случай, когда *все* $c_i(t)$ *равны одной и той же постоянной*, $c_i(t)=\lambda$, *а* p_{jk} *не зависят от* t. В этом случае p_{jk} являются переходными вероятностями обыкновенной цепи Маркова, и (как и в гл. XV) ее вероятности перехода за несколько шагов мы обозначим через $p_{jk}^{(n)}$.

Из $c_i(t)=\lambda$ следует, что вероятность осуществления перехода за время между t и $t+h$ не зависит от состояния системы и равна $\lambda h+o(h)$. Отсюда вытекает, что число переходов между τ и t имеет распределение Пуассона с параметром $\lambda(t-\tau)$. При условии, что произошло ровно n переходов, (условная) вероятность перехода из j в k равна $p_{jk}^{(n)}$. Следовательно,

$$P_{ik}(\tau,\ t)=e^{-\lambda(t-\tau)}\sum_{n=0}^{\infty}[\lambda^n(t-\tau)^n/n!]\,p_{ik}^{(n)} \tag{9.16}$$

(где, как обычно, $p_{jj}^{(0)}=1$ и $p_{jk}^{(0)}=0$ при $j\neq k$). Легко проверить, что (9.16) действительно является решением обеих систем (9.9) и (9.14) дифференциальных уравнений и удовлетворяет нужным граничным условиям.

В частности, если

$$p_{jk}=0 \text{ при } k<j,\quad p_{jk}=f_{k-j} \text{ при } k\geqslant j, \tag{9.17}$$

то (9.16) сводится к *обобщенному распределению Пуассона* (см. гл. XII, 2). ►

Обе наши системы дифференциальных уравнений впервые были выведены А. Н. Колмогоровым в работе, в которой были развиты основания теории марковских процессов [1]). Предполагая, что последовательность коэффициентов $c_n(t)$ остается ограниченной при каждом t, В. Феллер показал, что существует единственное общее для обеих систем решение $\{P_{jk}(\tau,\ t)\}$ и что это решение удовлетворяет уравнению Колмогорова — Чэпмена (9.1б), а также условиям (9.2). Более того, в этом случае ни у одной из двух наших систем нет других решений, и поэтому эти системы по существу эквивалентны. Однако конкретные задачи вскоре привели уравнениям с неограниченными последовательностями $\{c_n\}$, и, как показано в § 4, в таких случаях мы иногда сталкиваемся с неожиданными решения-

[1]) Kolmogoroff A., Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung, Mathematische Annalen, 104 (1931), 415—458.

ми, для которых в соотношении

$$\sum_k P_{jk}(\tau,\ t) \leqslant 1 \tag{9.18}$$

имеет место строгое неравенство. Было показано [1]) (без каких-либо ограничений на коэффициенты $c_n(t)$), что всегда существует *минимальное решение* $\{P_{jk}(\tau,\ t)\}$, удовлетворяющее обеим системам дифференциальных уравнений, а также уравнению Колмогорова — Чэпмена (9.16) и соотношению (9.18). Это решение называется минимальным потому, что

$$\overline{P}_{jk}(\tau,\ t) \geqslant P_{jk}(\tau,\ t), \tag{9.19}$$

если $\overline{P}_{jk}(\tau,\ t)$ удовлетворяет либо обратным, либо прямым дифференциальным уравнениям (вместе с банальными начальными условиями (9.10)). Когда минимальное решение удовлетворяет (9.18) со знаком равенства при всех t, это означает, что ни у обратных, ни у прямых уравнений не может быть никаких имеющих вероятностный смысл решений, кроме $P_{jk}(\tau,\ t)$. Иначе говоря, когда минимальное решение не имеет дефекта, наш процесс однозначно определяется любой из двух этих систем уравнений. Как уже говорилось выше, это имеет место тогда, когда для каждого фиксированного t коэффициенты $c_n(t)$ остаются ограниченными.

Ситуация становится совершенно другой, когда минимальное решение дефектно, т. е. когда в (9.18) для некоторого (а потому и для любого) t имеет место знак неравенства. В этом случае существует бесконечное число настоящих переходных вероятностей, удовлетворяющих обратным уравнениям и уравнению Колмогорова — Чэпмена, и, следовательно, существует бесконечно много марковских процессов, удовлетворяющих предположениям 1 и 2, лежащим в основе обратных уравнений. Некоторые из них могут удовлетворять также и прямым уравнениям, однако в остальных случаях решение прямых уравнений единственно [2]).

---

[1]) Feller W., On the integro-differential equations of purely discontinuous Markoff processes, Trans. Amer. Math. Soc., 48 (1940), 488—515. В этой статье рассматриваются более общие пространства состояний, однако счетные пространства состояний упоминаются в ней как наиболее интересный частный случай. Это было не замечено последующими авторами, которые дали более сложные и менее полные доказательства. Минимальное решение в однородном по времени случае получается в гл. XIV, 7 тома 2 при помощи преобразований Лапласа. Более полное изложение см. в работе Feller W., On boundaries and lateral conditions for the Kolmogorov differential equations, Ann. Math., 65 (1957), 527—570.

[2]) Напомним, что только предположения 1 и 2 имеют вероятностный смысл, тогда как предположение 3 носит чисто аналитический характер и было введено лишь для удобства. Оно не является естественным в том смысле, что даже не все решения прямых уравнений удовлетворяют наложенному условию равномерности. Таким образом, обратные уравнения выражают имеющие вероятностный смысл условия и приводят к интересной теории, однако о прямых уравнениях этого сказать нельзя. Это объясняет, почему вся теория марковских процессов должна

Пример. б) *Процесс размножения.* Дифференциальные уравнения (3.2) для однородного по времени процесса размножения имели вид

$$x_0'(t) = -\lambda_0 x_0(t), \quad x_k'(t) = -\lambda_k x_k(t) + \lambda_{k-1} x_{k-1}(t). \tag{9.20}$$

Это прямые уравнения. Поскольку они образуют рекуррентную систему, их решение однозначно определяется начальными значениями при $t=0$. Стало быть, для переходных вероятностей мы последовательно получаем $P_{ik}(t)=0$ при всех $k<i$,

$$P_{ii}(t) = e^{-\lambda_i t}, \quad P_{i,\,i+1}(t) = \frac{\lambda_i}{\lambda_i - \lambda_{i+1}}\left(e^{-\lambda_{i+1}t} - e^{-\lambda_i t}\right), \tag{9.21}$$

и, наконец, при $k > i$

$$P_{ik}(t) = \lambda_{k-1}\int_0^t e^{-\lambda_k s} P_{i,\,k-1}(t-s)\,ds. \tag{9.22}$$

Для того чтобы увидеть, что эти переходные вероятности удовлетворяют уравнению Колмогорова — Чэпмена (9.1а), достаточно заметить, что при фиксированных i и s обе части этого уравнения представляют собой решения дифференциальных уравнений (9.20), имеющие одинаковые начальные значения.

Обратные уравнения были выведены в (8.4) и имеют вид

$$y_i'(t) = -\lambda_i y_i(t) + \lambda_i y_{i+1}(t). \tag{9.23}$$

Мы должны показать, что этому уравнению удовлетворяют $P_{ik}(t)$ при фиксированном k. При $k<i$ это тривиально, ибо в этом случае все три члена в (9.23) обращаются в нуль. Используя (9.21), видим, что это утверждение справедливо также и при $k-i=0$ и $k-i=1$. Теперь мы можем продолжать доказательство по индукции, используя тот факт, что при $k>i+1$

$$P_{ik}'(t) = \lambda_{k-1}\int_0^t e^{-\lambda_k s} P_{i,\,k-1}'(t-s)\,ds. \tag{9.24}$$

Допустим, что $P_{ik}(t)$ удовлетворяет (9.23) при $k-i\leqslant n$. При $k=i+1+n$ мы можем выразить подынтегральное выражение в (9.24), используя правую часть (9.23), и получим в результате, что (9.23) справедливо и при $k-i=n+1$.

Таким образом, мы доказали, что система переходных вероятностей $P_{ik}(t)$ однозначно определяется прямыми уравнениями и что эти вероятности удовлетворяют обратным уравнениям, а также уравнению Колмогорова — Чэпмена.

основываться на обратных уравнениях (или — в абстрактной формулировке — на полугруппах преобразований функций, а не на вероятностных мерах).

Обратные уравнения (9.23) могут иметь и другие решения. Упоминавшееся выше свойство минимальности (9.19) наших переходных вероятностей можно переформулировать следующим образом. Для произвольных неотрицательных решений уравнений (9.23) имеем:

$$\textit{если } y_i(0) = P_{ik}(0), \quad \textit{то } y_i(t) \geqslant P_{ik}(t) \tag{9.25}$$

при всех $t > 0$. Здесь k произвольно, но фиксировано. Это утверждение тривиально для $k < i$, поскольку в этом случае правые части в (9.25) обращаются в нуль. Если y_{i+1} задано, то решение y_i уравнения (9.23) может быть представлено в явном виде как интеграл, аналогичный (9.22), и теперь справедливость (9.25) доказывается рекуррентным образом при $i = k, k-1, \dots$.

Предположим теперь, что $\sum \lambda_k^{-1} < \infty$. В § 4 было показано, что в этом случае величины

$$L_i(t) = 1 - \sum_{k=0}^{\infty} P_{ik}(t) \tag{9.26}$$

не обращаются тождественно в нуль. Очевидно, $L_i(t)$ можно интерпретировать как вероятность того, что при выходе из E_i «бесконечность» достигается до момента времени t. Ясно также, что L_i являются решениями дифференциальных уравнений (9.23) с начальными условиями $L_i(0) = 0$. Рассмотрим теперь произвольные неотрицательные функции A_k и определим

$$\overline{P}_{ik}(t) = P_{ik}(t) + \int_0^t L_i(t-s)\, A_k(s)\, ds. \tag{9.27}$$

Легко проверить, что при фиксированном k *функции* $\overline{P}_{ik}(t)$ *удовлетворяют обратным дифференциальным уравнениям и* $\overline{P}_{ik}(0) = P_{ik}(0)$. Встает вопрос о том, можно ли определить $A_k(t)$ таким образом, чтобы $\overline{P}_{ik}(t)$ были переходными вероятностями, удовлетворяющими уравнению Колмогорова — Чэпмена. Ответ на этот вопрос утвердительный. Мы воздержимся от доказательства этого утверждения, но приведем его вероятностную интерпретацию.

Вероятности $P_{ik}(t)$ определяют так называемый *процесс с поглощающим экраном: когда система достигает бесконечности, процесс останавливается.* Дуб [1]) впервые исследовал *процесс с возвращением*, в котором по достижении бесконечности система мгновенно возвращается в E_0 (или в некоторое другое заранее заданное состояние), и процесс начинается заново. В таком процессе система может перейти из E_0 в E_5 либо за пять, либо за бесконечное число шагов, совершая один или несколько проходов от E_0 до «бесконечности».

---

[1]) Doob J. L., Markoff chains — denumerable case, Trans. Amer. Math. Soc., 58 (1945), 455—473.

Переходные вероятности этого процесса имеют вид (9.27). Они *удовлетворяют обратным уравнениям* (8.4) *или* (9.23), *но не удовлетворяют прямым уравнениям* (9.24) *или* (8.5). ▶

Этим объясняется, почему при выводе прямых уравнений мы были вынуждены ввести странное на первый взгляд предположение 3, которое не является необходимым для обратных уравнений: интуитивно понятные и имеющие простой вероятностный смысл предположения 1—2 совместимы с процессами с возвращением, для которых прямые уравнения (9.24) не выполняются. Иначе говоря, если мы будем исходить из предположений 1—2, то *обратные уравнения Колмогорова будут выполнены, однако к прямым уравнениям нужно будет прибавить еще один член* [1]).

Конечно, пример с процессом чистого размножения слишком банален, чтобы быть действительно интересным, но описанная здесь ситуация типична и для наиболее общего случая уравнений Колмогорова. Однако встречаются и два новых явления. Во-первых, у процесса размножения существует только один путь ухода на «бесконечность», или, в абстрактной терминологии, одна *граничная* точка. В противоположность этому у общих процессов могут быть границы сложной топологической структуры. Во-вторых, в процессе размножения движение направлено к границе, так как возможны лишь переходы $E_n \to E_{n+1}$. Но можно построить процессы и другого типа; например, направление может быть изменено на противоположное, и мы получим процесс, в котором будут возможны только переходы $E_{n+1} \to E_n$. Такой процесс может *начинаться* на границе, вместо того чтобы оканчиваться на ней. В процессе размножения и гибели, как и в одномерной диффузии, переходы возможны в обоих направлениях. Оказывается, что в этом случае существуют процессы, аналогичные процессам с упругими и отражающими экранами из теории диффузии, однако их описание вывело бы нас за рамки этой книги.

§ 10. ЗАДАЧИ

1. Пусть в процессе чистого размножения, определенного уравнением (3.2), коэффициенты $\lambda_n > 0$ при всех n. Доказать, что для каждого фиксированного $n \geqslant 1$ функция $P_n(t)$ сперва возрастает, а затем убывает до 0. Если t_n — точка ее максимума, то $t_1 < t_2 < t_3 < \ldots$. *Указание.* Воспользоваться индукцией; продифференцировать (3.2).

2. *Продолжение.* Показать, что если $\sum \lambda_n^{-1} = \infty$, то $t_n \longrightarrow \infty$. *Указание.* Если $t_n \longrightarrow \tau$, то при фиксированном $t > \tau$ последовательность $\lambda_n P_n(t)$ возрастает. Использовать (4.10).

3. *Процесс Юла.* Вывести среднее и дисперсию распределения, определенного уравнениями (3.4). (Использовать только эти дифференциальные уравнения, а не явный вид решения (3.5).)

4. *Процесс чистой гибели.* Вывести дифференциальные уравнения процесса

[1]) Дальнейшие подробности см. в гл. XIV, 8 тома 2.

типа Юла с переходами только из E_n в E_{n-1}. Найти распределение $P_n(t)$, его среднее и дисперсию, предполагая, что начальным состоянием является E_i.

5. *Стоянки автомашин.* На автостоянку с числом мест N прибывают автомобили, образующие поток пуассоновского типа с интенсивностью λ, до тех пор, пока имеются свободные места. Времена стоянки автомобилей имеют пуассоновское распределение (как и времена обслуживания в § 7). Вывести соответствующие дифференциальные уравнения для вероятностей $P_n(t)$ того, что ровно n мест оказываются занятыми.

6. *Различные дисциплины обслуживания.* Рассмотрим очередь к одному каналу, подчиняющуюся правилам, приведенным в примере 7, б). На этот раз мы будем рассматривать процесс исключительно с точки зрения мистера Смита, чей вызов поступает в момент времени 0. Его время ожидания зависит от дисциплины обслуживания, а именно от того, в каком порядке обслуживаются ожидающие вызовы. Наибольший интерес представляют следующие дисциплины:

а) *последним прибыл — последним обслуживается*, т. е. вызовы обслуживаются в порядке поступления;

б) *случайный порядок*, т. е. члены очереди имеют равные вероятности быть обслуженными в первую очередь;

в) *последним прибыл — первым обслуживается*, т. е. вызовы обслуживаются в порядке, обратном порядку поступления [1]).

Удобно перенумеровать состояния, начиная с -1. На протяжении времени действительного обслуживания мистера Смита система находится в состоянии E_0, а по истечении этого времени она переходит в E_{-1}, где и остается навсегда. При $n \geqslant 1$ система находится в состоянии E_n, если вызов мистера Смита все еще находится в очереди вместе с $n-1$ другими вызовами, которые будут (или могут быть) обслужены до него. (Обслуживаемый в этот момент вызов в очередь не включается.) Обозначим $P_n(t)$ вероятность E_n в момент времени t. Доказать, что во всех трех случаях

$$P'_{-1}(t) = \mu P_0(t).$$

Кроме того,

а) при дисциплине *последним прибыл — последним обслуживается*

$$P'_n(t) = -\mu P_n(t) + \mu P_{n-1}(t), \quad n \geqslant 0;$$

б) при дисциплине *со случайным порядком* для $n \geqslant 2$

$$P'_n(t) = -(\lambda+\mu) P_n(t) + [n\mu/(n+1)] P_{n+1}(t) + \lambda P_{n-1}(t),$$
$$P'_1(t) = -(\lambda+\mu) P_1(t) + (1/2)\mu P_2(t),$$
$$P'_0(t) = -\mu P_0(t) + \mu P_1(t) + (1/2)\mu P_2(t) + (1/3)\mu P_3(t) + \ldots;$$

в) при дисциплине *последним прибыл — первым обслуживается* для $n \geqslant 2$

$$P'_n(t) = -(\lambda+\mu) P_n(t) + \mu P_{n+1}(t) + \lambda P_{n-1}(t),$$
$$P'_1(t) = -(\lambda+\mu) P_1(t) + \mu P_2(t),$$
$$P'_0(t) = -\mu P_0(t) + \mu P_1(t).$$

(См. также задачу 20.)

7. *Продолжение.* Предположим, что действует дисциплина обслуживания последним прибыл — последним обслуживается (случай а)), и что $P_r(0) = 1$. Показать, что

$$P_k(t) = [(\mu t)^{r-k}/(r-k)!]\, e^{-\mu t}, \quad 0 \leqslant k \leqslant r.$$

[1]) Эта дисциплина имеет смысл в обрабатывающих информацию машинах, когда последняя информация (или наблюдение) имеет наибольший вес. Она была предложена в работе Vaulot E., Delais d'attente des appels téléphoniques dans l'ordre inverse de leur arrivée, Comptes Rendues, Académie des Sciences, Paris, 238 (1954), 1188—1189.

8. *Продолжение.* Обобщить задачу 6 на случай a каналов.

9. *Процесс Пойа* [1]). Это нестационарный процесс чистого размножения, у которого λ_n зависят от времени:

$$\lambda_n(t) = (1+an)/(1+at). \tag{10.1}$$

Показать, что решение, удовлетворяющее начальному условию $P_0(0)=1$, имеет вид

$$P_0(t) = (1+at)^{-1/a},$$

$$P_n(t) = \frac{(1+a)(1+2a)\dots\{1+(n-1)a\}}{n!} t^n (1+at)^{-n-1/a}. \tag{10.2}$$

Показать (при помощи дифференциальных уравнений), что среднее и дисперсия здесь равны t и $t(1+at)$ соответственно.

10. *Продолжение.* Процесс Пойа можно получить предельным переходом из урновой схемы Пойа (пример гл. V, 2, в)). Если состояние системы определяется числом красных шаров, то вероятность перехода $E_k \to E_{k+1}$ при $(n+1)$-м извлечении равна

$$p_{k,n} = \frac{r+kc}{r+b+nc} = \frac{p+k\gamma}{1+n\gamma}, \tag{10.3}$$

где $p=r/(r+b)$, $\gamma = c/(r+b)$.

Пусть, как и при переходе от испытаний Бернулли к распределению Пуассона, интервалы между извлечениями шаров составляют h единиц времени, и пусть $h\to 0$, $n\to\infty$ так, что $np\to t$, $n\gamma\to at$. Показать, что в пределе (10.3) приводит к (10.1). Показать также, что распределение Пойа (2.3) гл. V переходит в (10.2).

11. *Линейный рост.* Если в процессе, определенном уравнениями (5.7), $\lambda=\mu$ и $P_1(0)=1$, то

$$P_0(t) = \lambda t/(1+\lambda t), \qquad P_n(t) = (\lambda t)^{n-1}/(1+\lambda t)^{n+1}. \tag{10.4}$$

Вероятность окончательного вымирания равна 1.

12. *Продолжение.* Взяв в качестве пробного решения системы (5.7) функции вида $P_n(t) = A(t)B^n(t)$, доказать, что решение с $P_1(0)=1$ имеет вид

$$P_0(t) = \mu B(t), \qquad P_n(t) = \{1-\lambda B(t)\}\{1-\mu B(t)\}\{\lambda B(t)\}^{n-1} \tag{10.5}$$

при

$$B(t) = (1-e^{(\lambda-\mu)t})/(\mu-\lambda e^{(\lambda-\mu)t}) \tag{10.6}$$

13. *Продолжение.* Доказать, что производящая функция $P(s,t)=\sum P_n(t)s^n$ удовлетворяет уравнению в частных производных

$$\partial P/\partial t = \{\mu-(\lambda+\mu)s+\lambda s^2\}\,\partial P/\partial s. \tag{10.7}$$

14. *Продолжение.* Пусть $M_2(t)=\sum n^2P_n(t)$ и $M(t)=\sum nP_n(t)$ (как и в § 5). Показать, что

$$M_2'(t) = 2(\lambda-\mu)M_2(t) + (\lambda+\mu)M(t). \tag{10.8}$$

Вывести отсюда, что при $\lambda>\mu$ *дисперсия* распределения $\{P_n(t)\}$ дается формулой

$$e^{2(\lambda-\mu)t}\{1-e^{(\mu-\lambda)t}\}(\lambda+\mu)/(\lambda-\mu). \tag{10.9}$$

15. Доказать, что производящая функция $P(s,t)=\sum P_n(t)s^n$ для про-

[1]) Lundberg O., On random processes and their applications to sickness and accident statistics, Uppsala, 1940.

цесса (7.2) удовлетворяет уравнению в частных производных

$$\partial P/\partial t = (1-s)\{-\lambda P + \mu \partial P/\partial s\}, \tag{10.10}$$

решением которого является

$$P(s, t) = e^{-\lambda(1-s)\left(1-e^{-\mu t}\right)/\mu}\{1-(1-s)e^{-\mu t}\}^i.$$

При $i=0$ это распределение Пуассона с параметром $\lambda(1-e^{-\mu t})/\mu$. *При $t \to \infty$ распределение $\{P_n(t)\}$ стремится к распределению Пуассона с параметром λ/μ.*

16. Доказать, что производящая функция $P(s) = \sum p_n s^n$ для стационарного состояния процесса, определенного уравнениями (7.26), удовлетворяет уравнению в частных производных

$$(\mu + \lambda s)\,\partial P/\partial s = a\lambda P, \tag{10.11}$$

которое имеет решение $P = \{(\mu+\lambda s)/(\lambda+\mu)\}^a$.

17. Рассмотрим для дифференциальных уравнений (7.26) пробное решение вида

$$P_n(t) = \binom{a}{n} A^n (1-A)^{a-n}.$$

Доказать, что эти функции являются решением тогда и только тогда, когда

$$A = [\lambda/(\lambda+\mu)]\,(1-e^{-(\lambda+\mu)t}).$$

18. Пусть в простейшей задаче о телефонных линиях (пример 7, а)) $Q_n(t)$ — вероятность, того, что, выходя из E_n, система достигнет состояния E_0 до момента времени t. Доказать справедливость дифференциальных уравнений

$$\begin{aligned} Q_n'(t) &= -(\lambda+n\mu)\,Q_n(t) + \lambda Q_{n+1}(t) + n\mu Q_{n-1}(t), \quad n \geqslant 2, \\ Q_1'(t) &= -(\lambda+\mu)\,Q_1(t) + \lambda Q_2(t) + \mu \end{aligned} \tag{10.12}$$

с начальными условиями $Q_n(0) = 0$.

19. *Продолжение.* Рассмотреть ту же задачу для процесса, определяемого произвольной системой прямых уравнений. Показать, что тогда $Q_n(t)$ удовлетворяют соответствующим *обратным уравнениям* (при фиксированном k), в которых $P_{0k}(t)$ заменено на 1.

20. Показать, что дифференциальные уравнения из задачи 6 по существу представляют собой прямые уравнения для переходных вероятностей. Вывести соответствующие обратные уравнения.

21. Допустим, что решение хотя бы одной из двух систем (прямых и обратных) уравнений единственно. Доказать, что переходные вероятности, удовлетворяющие этой системе, удовлетворяют уравнению Колмогорова — Чэпмена (1.1). *Указание.* Показать, что выражения, стоящие в обеих частях этого уравнения, удовлетворяют одной и той же системе дифференциальных уравнений с одинаковыми начальными условиями.

22. Пусть $P_{ik}(t)$ удовлетворяют уравнению Колмогорова — Чэпмена (1.1). Предполагая, что $P_{ik}(t) > 0$ и $S_i(t) = \sum\limits_k P_{ik}(t) \leqslant 1$, доказать, что либо $S_i(t) = 1$ при всех t, либо $S_i(t) < 1$ при всех t.

23. *Эргодические свойства.* Рассмотрим стационарный процесс с конечным числом состояний, т. е. предположим, что система дифференциальных уравнений (9.9) конечна и что коэффициенты c_j и p_{jk} суть постоянные величины. Доказать, что решениями здесь являются линейные комбинации показательных членов $e^{\lambda(t-\tau)}$, где действительная часть λ отрицательна, если $\lambda \neq 0$. Вывести отсюда, что асимптотическое поведение переходных вероятностей такое же, как и в случае *конечных* цепей Маркова, за тем исключением, что периодический случай теперь невозможен.

ОТВЕТЫ К ЗАДАЧАМ

Глава I

1. а) 3/5; б) 3/5; в) 3/10.
2. События S_1, S_2, $S_1 \cup S_2$ и S_1S_2 содержат 12, 12, 18 и 6 точек соответственно.
4. Пространство элементарных событий содержит две точки $ГГ$ и $РР$ с вероятностями 1/4, две точки $ГРР$ и $РГГ$ с вероятностями 1/8 и вообще по две точки с вероятностями 2^{-n} при $n \geqslant 2$. Сумма этих вероятностей равна единице, так что нет необходимости рассматривать возможность бесконечной серии бросаний. Искомые вероятности равны 15/16 и 2/3 соответственно.
9. $\mathbf{P}\{AB\} = 1/6$, $\mathbf{P}\{A \cup B\} = 23/36$, $\mathbf{P}\{AB'\} = 1/3$.
12. $x = 0$ в случае событий а), б) и ж); $x = 1$ в случае событий д) и е); $x = 2$ в случае события г); $x = 4$ в случае события в).
15. а) A; б) AB; в) $B \cup (AC)$.
16. Правильными являются соотношения в), г), д), е), з), и), л) и м). Соотношение а) бессмысленно, за исключением случая $C \subset B$. Вообще говоря, оно неверно даже и в этом случае, но будет правильным при дополнительных условиях $C \subset B$, $AC = 0$. Соотношение б) правильно в случае $C \supset AB$. Соотношение ж) должно выглядеть так: $(A \cup B) - A = A'B$. Наконец, правильным вариантом к) является л).
17. а) $AB'C'$; б) ABC'; в) ABC; г) $A \cup B \cup C$; д) $AB \cup AC \cup BC$; е) $AB'C' \cup A'BC' \cup A'B'C$; ж) $ABC' \cup AB'C \cup A'BC = (AB \cup AC \cup BC) - ABC$; з) $A'B'C'$; и) $(ABC)'$.
18. $A \cup B \cup C = A \cup (B - AB) \cup \{C - C(A \cup B)\} = A \cup BA' \cup CA'B'$.

Глава II

1. а) 26^3; б) $26^2 + 26^3 = 18\,252$; в) $26^2 + 26^3 + 26^4$. В городе с 20 000 жителями или некоторые люди имеют одинаковые инициалы, или по меньшей мере 1748 людей имеют более трех инициалов.
2. $2(2^{10} - 1) = 2046$.
3. $\binom{n}{2} + n = \dfrac{n(n+1)}{2}$. 4. а) $\dfrac{1}{n}$; б) $\dfrac{1}{n(n-1)}$.
5. $q_A = (5/6)^6$, $q_B = (5/6)^{12} + 12(5/6)^{11} \cdot (1/6)$.
6. а) $p_1 = 0{,}01$, $p_2 = 0{,}27$, $p_3 = 0{,}72$;
   б) $p_1 = 0{,}001$, $p_2 = 0{,}063$, $p_3 = 0{,}432$, $p_4 = 0{,}504$.
7. $p_r = (10)_r\, 10^{-r}$. Например, $p_3 = 0{,}72$, $p_{10} = 0{,}00036288$. Формула Стирлинга дает $p_{10} = 0{,}0003598\ldots$.
8. а) $(9/10)^k$; б) $(9/10)^k$; в) $(8/10)^k$; г) $2(9/10)^k - (8/10)^k$; д) AB и $A \cup B$.
9. $\binom{n}{2} n!\, n^{-n}$. 10. $9\binom{12}{8}^{-1} = \dfrac{1}{55}$.
11. Вероятность ровно r испытаний равна $(n-1)_{r-1}/(n)_r = n^{-1}$.
12. а) $[1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2n-1)]^{-1} = 2^n n!/(2n)!$;
    б) $n!\,[1 \cdot 3 \cdot \ldots \cdot (2n-1)]^{-1} = 2^n \binom{2n}{n}^{-1}$.

13. В предположении случайности вероятность того, что все двенадцать штрафов придутся на вторники или четверти, равна $(2/7)^{12}=0{,}0000003\ldots$. Имеется лишь $\binom{7}{2}=21$ пара дней, так что вероятность остается крайне малой даже для любых двух дней. Следовательно, разумно предположить, что полиция действует по системе.

14. В предположении случайности вероятность события равна $(6/7)^{12}$, или приблизительно 1/6. Никакого уверенного заключения сделать нельзя.

15. $(90)_{10}/(100)_{10}=0{,}330476\ldots$.

16. $25!\,(5!)^{-5}5^{-25}=0{,}00209\ldots$.

17. $2\,(n-2)_r\,(n-r-1)!/n!=2\,(n-r-1)/[n\,(n-1)]$.

18. а) 1/216; б) 83/3888.

19. Вероятности равны $1-(5/6)^4=0{,}517747\ldots$ и $1-(35/36)^{24}=0{,}491404\ldots$.

20. а) $(n-N)_r/(n)_r$; б) $(1-N/n)^r$. При $r=N=3$ вероятности равны а) $0{,}911812\ldots$; б) $0{,}912673\ldots$. При $r=N=10$ они равны а) 0,330476; б) $0{,}348678\ldots$.

21. а) $(1-N/n)^{r-1}$; б) $(n)_{Nr}/((n)_N)^r$.

22. $(1-2/n)^{2^r-2}$; для медианы приблизительно $2^{r+1}=0{,}7n$.

23. В предположении случайности вероятности того, что три или четыре тарелки разбиты а) одной девочкой, б) самой младшей девочкой, равны соответственно $13/64\approx 0{,}2$ и $13/256\approx 0{,}05$.

24. а) $12!/12^{12}=0{,}000054$; б) $\binom{12}{2}(2^6-2)\,12^{-6}=0{,}00137$.

25. $\dfrac{30!}{2^6 6^6}\binom{12}{6}12^{-30}\approx 0{,}00035\ldots$.

26. а) $\binom{n}{2r}2^{2r}\binom{2n}{2r}^{-1}$; б) $n\binom{n-1}{2r-2}2^{2r-2}\binom{2n}{2r}^{-1}$;

в) $\binom{n}{2}\binom{n-2}{2r-4}2^{2r-4}\binom{2n}{2r}^{-1}$.

27. $\binom{N-3}{r-1}\binom{N-1}{r-1}^{-1}$.

28. $p=\binom{2N}{N}^2\binom{4N}{2N}^{-1}\approx\sqrt{2/(N\pi)}$.

29. $p=\binom{4}{k}\binom{48}{13-k}\binom{39}{13}\binom{26}{13}\binom{52}{13}^{-1}\binom{39}{13}^{-1}\binom{26}{13}^{-1}=$

$=\binom{4}{k}\binom{48}{13-k}\binom{52}{13}^{-1}$.

30. См. задачу 29. Вероятность равна

$$\binom{13}{m}\binom{39}{13-m}\binom{13-m}{n}\binom{26+m}{13-n}\binom{52}{13}^{-1}\binom{39}{13}^{-1}.$$

31. $\binom{4}{k}\binom{48}{26-k}\binom{52}{26}^{-1}$.

32. $\binom{13}{a}\binom{39}{13-a}\binom{13-a}{b}\binom{26+a}{13-b}\binom{13-a-b}{c}\binom{13+a+b}{13-c}\times$

$\times\binom{52}{13}^{-1}\binom{39}{13}^{-1}\binom{26}{13}^{-1}$.

33. а) $24p\,(5, 4, 3, 1)$; б) $4p\,(4, 4, 4, 1)$; в) $12p\,(4, 4, 3, 2)$.

34. $\binom{13}{a}\binom{13}{b}\binom{13}{c}\binom{13}{d}\binom{52}{13}^{-1}$. (См. задачу 33, где найдена вероятность того, что у одного игрока имеется a карт одной масти, b — другой и т. д.)

35. $p_0(r)=(52-r)_4/(52)_4$; $p_1(r)=4r(52-r)_3/(52)_4$;
$p_2(r)=6r(r-1)(52-r)_2/(52)_4$;
$p_3(r)=4r(r-1)(r-2)(52-r)/(52)_4$; $p_4(r)=(r)_4/(52)_4$.

36. Вероятности того, что времена ожидания первого, ... четвертого туза *превосходят* r, равны $w_1(r)=p_0(r)$; $w_2(r)=p_0(r)+p_1(r)$; $w_3(r)=p_0(r)+p_1(r)+p_2(r)$; $w_4(r)=1-p_4(r)$. Далее, $f_i(r)=w_i(r-1)-w_i(r)$. Медианы равны 9, 20, 33, 44.

37. а) $\binom{4}{k}\binom{4-k}{k}\binom{48}{r-k}\binom{48-r+k}{r-k}\binom{52}{r}^{-1}\binom{52-r}{r}^{-1}$ при $k\leqslant 2$;

б) $\left\{\binom{4}{k}\binom{48}{r-k}\binom{52}{r}^{-1}\right\}^2$ при $k\leqslant 4$.

39. $\binom{r_1+n-1}{r_1}\binom{r_2+n-1}{r_2}$. **40.** $\binom{r_1+5}{5}(r_2+1)$.

41. $(r_1+r_2+r_3)!/(r_1!\,r_2!\,r_3!)$. **42.** $(49)_4/(52)_4$.

43.
$$
\begin{aligned}
&\mathbf{P}\{(7)\} &&=10\cdot 10^{-7} &&=0{,}000001,\\
&\mathbf{P}\{(6,1)\} &&=\frac{10!}{8!\,1!\,1!}\cdot\frac{7!}{1!\,6!}\cdot 10^{-7} &&=0{,}000063,\\
&\mathbf{P}\{(5,2)\} &&=\frac{10!}{8!\,1!\,1!}\cdot\frac{7!}{2!\,5!}\cdot 10^{-7} &&=0{,}000189,\\
&\mathbf{P}\{(5,1,1)\} &&=\frac{10!}{7!\,2!\,1!}\cdot\frac{7!}{1!\,1!\,5!}\cdot 10^{-7} &&=0{,}001512,\\
&\mathbf{P}\{(4,3)\} &&=\frac{10!}{8!\,1!\,1!}\cdot\frac{7!}{3!\,4!}\cdot 10^{-7} &&=0{,}000315,\\
&\mathbf{P}\{(4,2,1)\} &&=\frac{10!}{7!\,1!\,1!\,1!}\cdot\frac{7!}{1!\,2!\,4!}\cdot 10^{-7} &&=0{,}007560,\\
&\mathbf{P}\{(4,1,1,1)\} &&=\frac{10!}{6!\,3!\,1!}\cdot\frac{7!}{1!\,1!\,1!\,4!}\cdot 10^{-7} &&=0{,}017640,\\
&\mathbf{P}\{(3,3,1)\} &&=\frac{10!}{7!\,2!\,1!}\cdot\frac{7!}{1!\,3!\,3!}\cdot 10^{-7} &&=0{,}005040,\\
&\mathbf{P}\{(3,2,2)\} &&=\frac{10!}{7!\,2!\,1!}\cdot\frac{7!}{2!\,2!\,3!}\cdot 10^{-7} &&=0{,}007560,\\
&\mathbf{P}\{(3,2,1,1)\} &&=\frac{10!}{6!\,2!\,1!\,1!}\cdot\frac{7!}{1!\,1!\,2!\,3!}\cdot 10^{-7} &&=0{,}105840,\\
&\mathbf{P}\{(3,1,1,1,1)\} &&=\frac{10!}{5!\,4!\,1!}\cdot\frac{7!}{1!\,1!\,1!\,1!\,3!}\cdot 10^{-7} &&=0{,}105840,\\
&\mathbf{P}\{(2,2,2,1)\} &&=\frac{10!}{6!\,3!\,1!}\cdot\frac{7!}{1!\,2!\,2!\,2!}\cdot 10^{-7} &&=0{,}052920,\\
&\mathbf{P}\{(2,2,1,1,1)\} &&=\frac{10!}{5!\,3!\,2!}\cdot\frac{7!}{1!\,1!\,1!\,2!\,2!}\cdot 10^{-7} &&=0{,}317520,\\
&\mathbf{P}\{(2,1,1,1,1,1)\} &&=\frac{10!}{4!\,5!\,1!}\cdot\frac{7!}{1!\,1!\,1!\,1!\,1!\,2!}\cdot 10^{-7} &&=0{,}317520,\\
&\mathbf{P}\{(1,1,1,1,1,1,1)\} &&=\frac{10!}{3!\,7!}\cdot 7!\cdot 10^{-7} &&=0{,}060480.
\end{aligned}
$$

44. Пусть S, D, T, Q означает, что в какой-то день родились один, два, три и четыре человека соответственно. Тогда

$$
\begin{aligned}
&\mathbf{P}\{22S\} &&=\frac{365!}{343!}\cdot 365^{-22} &&=0{,}52430,\\
&\mathbf{P}\{20S+1D\} &&=\frac{365!}{1!\,344!}\cdot\frac{22!}{20!\,2!}\cdot 365^{-22} &&=0{,}35208,\\
&\mathbf{P}\{18S+2D\} &&=\frac{365!}{2!\,345!}\cdot\frac{22!}{18!\,2!\,2!}\cdot 365^{-22} &&=0{,}09695,
\end{aligned}
$$

$$\mathbf{P}\{16S+3D\} = \frac{365!}{3!\,346!}\cdot\frac{22!}{16!\,2!\,2!\,2!}\cdot 365^{-22} = 0{,}01429,$$

$$\mathbf{P}\{19S+1T\} = \frac{365!}{345!}\cdot\frac{22!}{19!\,3!}\cdot 365^{-22} = 0{,}00680,$$

$$\mathbf{P}\{17S+1D+1T\} = \frac{365!}{346!}\cdot\frac{22!}{17!\,2!\,3!}\cdot 365^{-22} = 0{,}00336,$$

$$\mathbf{P}\{14S+4D\} = \frac{365!}{347!}\cdot\frac{22!}{14!\,2!\,2!\,2!\,2!}\cdot 365^{-22} = 0{,}00124,$$

$$\mathbf{P}\{15S+2D+1T\} = \frac{365!}{347!}\cdot\frac{22!}{15!\,2!\,2!\,3!}\cdot 365^{-22} = 0{,}00066,$$

$$\mathbf{P}\{18S+1Q\} = \frac{365!}{346!}\cdot\frac{22!}{18!\,4!}\cdot 365^{-22} = 0{,}00009.$$

45. Пусть $q=\binom{52}{5}=2\,598\,960$. Вероятности равны
а) $4/q$; б) $13\cdot 12\cdot 4\cdot q^{-1}=1/4165$; в) $13\cdot 12\cdot 4\cdot 6\cdot q^{-1}=6/4165$; г) $9\cdot 4^5\cdot q^{-1}=$ $=768/216580$; д) $13\cdot\binom{12}{2}\cdot 4\cdot 4^2\cdot q^{-1}=88/4165$; е) $\binom{13}{2}\cdot 11\cdot 6\cdot 6\cdot 4\cdot q^{-1}=$ $=198/4165$; ж) $13\cdot\binom{12}{3}\cdot 6\cdot 4^3\cdot q^{-1}=1760/4165$.

Глава IV

1. 99/323. **2.** 0,21... . **3.** 1/4. **4.** $7/2^6$. **5.** 1/81 и $31/6^6$.

6. Если A_k — событие, состоящее в том, что $(k,\ k)$ не появится, то из (1.5) следует, что

$$1-p_r=6\left(\frac{35}{36}\right)^r-\binom{6}{2}\left(\frac{34}{36}\right)^r+\binom{6}{3}\left(\frac{33}{36}\right)^r-\binom{6}{4}\left(\frac{32}{36}\right)^r+6\left(\frac{31}{36}\right)^r-\left(\frac{30}{36}\right)^r.$$

7. Положим $p^{-1}=\binom{52}{13}$. Тогда $S_1=13\binom{48}{9}p$; $S_2=\binom{13}{2}\binom{44}{5}p$; $S_3=$ $=40\binom{13}{3}p$. Приближенные численные значения: $P_{[0]}=0{,}09658$; $P_{[1]}=$ $=0{,}0341$; $P_{[2]}=0{,}0001$.

8. $u_r=\sum\limits_{k=0}^{N}(-1)^k\binom{N}{k}\left(1-\frac{k}{n}\right)^r$.

9. $u_r=\sum\limits_{k=0}^{N}(-1)^k\binom{N}{k}\frac{(n-k)_r}{(n)_r}$. Для доказательства того, что эти два результата согласуются, использовать формулу (12.18) гл. II.

10. Общий член имеет вид $a_{1k_1}a_{2k_2}\ldots a_{Nk_N}$, где $(k_1,\ \ldots,\ k_N)$ — перестановка $(1,\ 2,\ \ldots,\ N)$. Для диагонального элемента $k_\nu=\nu$.

12. $u_r=\sum\limits_{k=0}^{n}(-1)^k\binom{n}{k}\frac{(ns-ks)_r}{(ns)_r}$.

14. Заметить, что по определению $u_r=0$ при $r<n$ и $u_n=n!\,s^n/(ns)_n$.

15. $u_r-u_{r-1}=\sum\limits_{k=1}^{n}(-1)^{k-1}\binom{n-1}{k-1}\frac{(ns-ks)_{r-1}}{(ns-1)_{r-1}}$.

Предел равен $\sum_{k=0}^{n-1}(-1)^k\binom{n-1}{k}\left(1-\frac{k+1}{n}\right)^{r-1}$.

16. $\binom{N}{2}^{-r}\binom{N}{m}\sum_{k=2}^{m}(-1)^{m-k}\binom{m}{k}\binom{k}{2}^r$.

17. Воспользоваться тем, что $\binom{52}{5}S_k=\binom{4}{k}\binom{52-13k}{5}$. Приближенные численные значения: $P_{[0]}=0{,}264$, $P_{[1]}=0{,}588$, $P_{[2]}=0{,}146$, $P_{[3]}=0{,}002$.

18. Воспользоваться тем, что $\binom{52}{13}S_k=\binom{4}{k}\binom{52-2k}{13-2k}$. Приближенные численные значения: $P_{[0]}=0{,}780217$, $P_{[1]}=0{,}204606$, $P_{[2]}=0{,}014845$, $P_{[3]}=0{,}000330$, $P_{[4]}=0{,}000002$.

19. $m!\,N!\,u_m=\sum_{k=0}^{N-m}(-1)^k(N-m-k)!/k!$.

20. См. формулу в ответе к задаче 21 при $r=2$.

21. $(rN)!\,x=$

$$=\binom{N}{2}r^2(rN-2)!-\binom{N}{3}r^3(rN-3)!+-\ldots+(-1)^N r^N(rN-N)!.$$

24. $P_{[m]}=\binom{n}{m}\binom{n+r-1}{r}^{-1}\sum_{k=0}^{n-m}(-1)^k\binom{n-m}{k}\binom{n-m+r-1-k}{r}$.

25. Воспользоваться формулами (12.16) и (12.4) гл. II.

26. Положить $U_N=A_1\cup\ldots\cup A_N$ и заметить, что $U_{N+1}=U_N\cup A_{N+1}$ и $U_NA_{N+1}=(A_1A_{N+1})\cup\ldots\cup(A_NA_{N+1})$.

Глава V

1. $1-(5)_3/(6)_3=1/2$. 2. $p=1-10\cdot5^9/(6^{10}-5^{10})=0{,}61\ldots$.

3. а) $\binom{35}{13}\binom{39}{13}^{-1}=0{,}182\ldots$. Вероятность ровно одного туза равна $4\binom{35}{12}\binom{39}{13}^{-1}=0{,}411\ldots$; б) приближенно $1-0{,}182-0{,}411=0{,}407$.

4. а) $2\binom{23}{10}\binom{26}{13}^{-1}=\frac{11}{50}$; б) $2\binom{23}{12}\binom{26}{13}^{-1}=\frac{13}{50}$.

6. 125/345; 140/345; 80/345. 7. 20/21. 9. $(5/6)^2$. 10. $1-(5/6)^2$.

12. $p/(2-p)$. 13. б) 3/5; в) $2^n(1+2^n)^{-1}$.

14. г) Положим $a_n=x_n-4/7$, $b_n=y_n-1/7$, $c_n=z_n-2/7$; тогда

$$|a_n|+|b_n|+|c_n|=(1/2)\{|a_{n+1}|+|b_{n+1}|+|c_{n+1}|\}.$$

Следовательно, $|a_n|+|b_n|+|c_n|$ возрастает со скоростью геометрической прогрессии.

15. $p=(1-p_1)(1-p_2)\ldots(1-p_n)$.

16. Использовать неравенство $1-x<e^{-x}$ при $0<x<1$ или разложение в ряд Тейлора функции $\log(1-x)$; см. формулу (12.26) гл. II.

18. $(b+c)/(b+c+r)$.

19. Предположим, что утверждение справедливо для n-го извлечения при любых b, r и c. Рассматривая две возможности при *первом* извлечении, находим, что вероятность черного шара в $(n+1)$-м испытании равна

$$\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}+\frac{r}{b+r}\cdot\frac{b}{b+r+c}=\frac{b}{b+r}.$$

20. В предыдущей задаче устанавливается, что утверждение верно при $m=1$ и любом n. Для применения индукции рассмотреть две возможности в первом испытании.

23. Воспользоваться формулой (12.9) гл. II.

24. Биномиальный коэффициент в правой части является пределом первого множителя в числителе (8.2). Заметить, что

$$\binom{-1/\gamma}{n} \sim \binom{-1/\gamma}{n_2}(1+\rho)^{n_1}.$$

26. В силу (5.2) $2v=2p(1-p)\leqslant 1/2$.

28. а) u^2; б) $u^2+uv+v^2/4$; в) $u^2+(25uv+9v^2+vw+2uw)/16$.

33. $p_{11}=p_{32}=2p_{21}=p$, $p_{12}=p_{33}=2p_{23}=q$, $p_{13}=p_{31}=0$, $p_{22}=1/2$.

Глава VI

1. 5/16. **2.** Вероятность равна 0,02804... . **3.** $(0,9)^x\leqslant 0,1$, $x\geqslant 22$.

4. $q^x\leqslant 1/2$ и $(1-4p)^x\leqslant 1/2$, где $p=\binom{48}{9}\binom{52}{13}^{-1}$. Следовательно, $x\geqslant 263$ и $x\geqslant 66$ соответственно.

5. $1-(0,8)^{10}-2(0,8)^9=0,6242...$.

6. $\{1-(0,8)^{10}-2(0,8)^9\}/\{1-(0,8)^{10}\}=0,6993...$.

7. $\binom{26}{2}\binom{26}{11}\binom{52}{13}^{-1}=0,003954...$ и $\binom{13}{2}\frac{1}{2^{13}}=0,00952...$.

8. $\binom{12}{2}\{6^{-6}-2\cdot 12^{-6}\}$.

9. Истинные значения: 0,6651..., 0,40187... и 0,2009...; пуассоновские приближения: $1-e^{-1}=0,6321...$, 0,3679... и 0,1839... .

10. $e^{-2}\sum\limits_{4}^{\infty}2^k/k!=0,143...$. **11.** $e^{-1}\sum\limits_{3}^{\infty}1/k!=0,080...$.

12. $e^{-x/100}\leqslant 0,05$, или $x\geqslant 300$.

13. $e^{-1}=0,3679...$, $1-2\cdot e^{-1}=0,264...$.

14. $e^{-x}\leqslant 0,01$, $x\geqslant 5$. **15.** $1/p=649\,740$.

16. $1-p^n$, где $p=p(0;\lambda)+...+p(k;\lambda)$.

18. q^3 при $k=0$; pq^3 при $k=1, 2, 3$ и pq^3-pq^6 при $k=4$.

19. $\sum\limits_{k=0}^{n}\binom{n}{k}^2 2^{-2n}=\binom{2n}{n}2^{-2n}\approx 1/\sqrt{\pi n}$ для больших n.

20. $\sum\limits_{k=a}^{a+b-1}\binom{a+b-1}{k}p^k q^{a+b-1-k}$. Этот результат можно записать также в виде $p^a\sum\limits_{k=0}^{b-1}\binom{a+k-1}{k}q^k$, где k-е слагаемое равно вероятности того, что a-й успех имеет место непосредственно после $k\leqslant b-1$ неудач.

21. $x_r=\binom{2N-1-r}{N-r}\cdot 2^{-2N+r+1}$.

22. а) $x=\sum\limits_{r=1}^{N}x_r 2^{-r-1}=|2^{-2N}\sum\limits_{r=1}^{N}\binom{2N-1-r}{N-1}$; б) Воспользоваться формулой (12.6) гл. II.

23. $k_i\approx np_i$, $k_{12}\approx np_{12}$; следовательно, $n\approx k_1k_2/k_{12}$.

24. $\binom{n}{n_1} \cdot \binom{n-s_1}{n_2} \dots \binom{n-s_{r-1}}{n_r} \cdot q^{s_r} p^{(rn-s_1-\dots-s_r)}$, где $s_l = n_1 + \dots + n_l$.
25. $p = p_1q_2(p_1q_2 + p_2q_1)^{-1}$.
31. Из разложения логарифма в ряд Тейлора следует, что

$$b(0;\ n,\ p) = q^n = (1-\lambda/n)^n < e^{-\lambda} = p(0;\ \lambda).$$

Члены каждого распределения дают в сумме единицу, поэтому невозможно, чтобы *все* члены одного распределения были больше соответствующих членов другого.

32. Только конечное число членов распределения Пуассона превосходят ε, а остальные члены оказываются больше соответствующих членов биномиального распределения.

Глава VII

1. Действовать так же, как в § 1. 2. Воспользоваться формулой (1.7).
3. $\mathfrak{N}(-32/30) = 0{,}143\dots$. 4. 0,99. 5. 511. 6. 66 400.
7. Конечно. Из неравенств гл. VI следует, что вероятность превышения восьмикратного стандартного отклонения крайне мала.
8. $(2\pi n)^{-1}\{p_1p_2(1-p_1-p_2)\}^{-1/2}$.

Глава VIII

1. $\beta = 21$.
2. $x = pu + qv + rw$, где u, v, w — решения системы

$$u = p^{\alpha-1} + (qv + rw)(1 - p^{\alpha-1})/(1-p),$$
$$v = (pu + rw)(1 - q^{\beta-1})/(1-q),$$
$$w = pu + qv + rw = x.$$

3. $u = p^{\alpha-1} + (qv + rw)(1 - p^{\alpha-1})/(1-p)$,
$v = (pu + rw)(1 - q^{\beta-1})/(1-q)$,
$w = (pu + qv)(1 - r^{\gamma-1})/(1-r)$.
4. Заметить, что $\mathbf{P}\{A_n\} < (2p)^n$, но

$$\mathbf{P}\{A_n\} > 1 - (1-p^n)^{2^n/(2n)} > 1 - e^{(-2p)^n/(2n)}.$$

Если $p = 1/2$, то последняя величина имеет порядок $1/(2n)$, если $p > 1/2$, то $\mathbf{P}\{A_n\}$ даже не стремится к нулю.

Глава IX

1. Возможны комбинации (0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (2, 0), (2, 1), (3, 0). Их вероятности равны 0,047539, 0,108883, 0,017850, 0,156364, 0,214197, 0,321295, 0,026775, 0,107098.
2. а) Совместное распределение задается (6×6)-матрицей. Главная диагональ состоит из элементов q, $2q$, ..., $6q$, где $q = 1/36$. По одну сторону от главной диагонали все элементы равны нулю, по другую все они равны q. б) $\mathbf{E}(\mathbf{X}) = 7/2$, $\operatorname{Var}(\mathbf{X}) = 35/12$, $\mathbf{E}(\mathbf{Y}) = 161/36$, $\operatorname{Var}(\mathbf{Y}) = 2555/1296$, $\operatorname{Cov}(\mathbf{X}, \mathbf{Y}) = 105/72$.
3. Строки совместного распределения равны (перед всеми строками подразумевается множитель 32^{-1})

для **X**, **Y**: (1, 0, 0, 0, 0, 0), (0, 5, 4, 3, 2, 1), (0, 0, 6, 6, 3, 0), (0, 0, 0, 1, 0, 0);
для **X**, **Z**: (1, 0, 0, 0, 0, 0), (0, 5, 6, 1, 0, 0), (0, 0, 4, 6, 1, 0), (0, 0, 0, 3, 2, 0), (0, 0, 0, 0, 2, 0), (0, 0, 0, 0, 0, 1);

для $\mathbf{Y}$, $\mathbf{Z}$: (1, 0, 0, 0), (0, 5, 6, 1), (0, 4, 7, 0), (0, 3, 2, 0), (0, 2, 0, 0), (0, 1, 0, 0);
распределение $\mathbf{X}+\mathbf{Y}$: (1, 0, 5, 4, 9, 8, 5), значения $\mathbf{X}+\mathbf{Y}$ меняются от 0 до 6;
распределение $\mathbf{XY}$: (1, 5, 4, 3, 8, 1, 6, 0, 3, 1), значения $\mathbf{XY}$ меняются от 0 до 9.
$\mathbf{E}(\mathbf{X})=5/2$, $\mathbf{E}(\mathbf{Y})=3/2$, $\mathbf{E}(\mathbf{Z})=31/16$, $\mathrm{Var}(\mathbf{X})=5/4$, $\mathrm{Var}(\mathbf{Y})=3/8$, $\mathrm{Var}(\mathbf{Z})==303/256$.

4. а) $p/(1+q)$; б) $1/(1+q+q^2)$; в) $1/(1+q)^2$.

8. Распределение $\mathbf{V}_n$ задается формулой (3.5), для $\mathbf{U}_n$ оно получается по симметрии.

9. а) $\mathbf{P}\{\mathbf{X}\leqslant r,\ \mathbf{Y}\geqslant s\}=N^{-n}(r-s+1)^n$ для $r\geqslant s$;

$$\mathbf{P}\{\mathbf{X}=r,\ \mathbf{Y}=s\}=\begin{cases}N^{-n}\{(r-s+1)^n-2(r-s)^n+(r-s-1)^n\}, & \text{если } r>s,\\ N^{-n}, & \text{если } r=s.\end{cases}$$

б) $x=\dfrac{r^{n-2}-(r-1)^{n-2}}{r^n-(r-1)^n}$, если $j<r$ и $k<r$;

$x=\dfrac{r^{n-2}}{r^n-(r-1)^n}$, если $j\leqslant r$ и $k=r$ или $j=r$ и $k\leqslant r$;

$x=0$, если $j>r$ или $k>r$.

в) $\sigma^2\approx nN^2/[(n+1)^2(n+2)]$.

10. Вероятность того, что будет n двойных бросаний, равна $2pq(p^2+q^2)^{n-1}$. Математическое ожидание равно $1/(2pq)$.

12. $\mathbf{P}\{\mathbf{N}=n,\ \mathbf{K}=k\}=\binom{n}{k}p^{n-k}(qq')^k\cdot qp'$ для $k\leqslant n$;

$\mathbf{P}\{\mathbf{N}=n\}=(1-qp')^n qp'$;

$\mathbf{P}\{\mathbf{K}=k\}=(qq')^k qp'\sum\binom{-k-1}{\nu}(-p)^\nu=p'q'^k$.

13. $$\mathbf{E}\left(\frac{\mathbf{K}}{\mathbf{N}+1}\right)=\sum_{k,\,n}kp_{k,\,n}/(n+1)=q^2p'q'\sum_{n=1}^{\infty}\left(1-\frac{1}{n+1}\right)(p+qq')^{n-1}=$$
$$=qq'/(1-qp')-q^2p'q'/(1-qp')^2\log[1/(qp')];$$
$$\mathbf{E}(\mathbf{K})=q'/p';\quad \mathbf{E}(\mathbf{N})=(1-qp')/(qp');\quad \mathrm{Cov}(\mathbf{K},\ \mathbf{N})=q'/(qp'^2),$$
$$\rho(\mathbf{K},\ \mathbf{N})=\sqrt{q'/(1-qp')}.$$

14. а) $p_k=p^kq+q^kp$; $\mathbf{E}(\mathbf{X})=pq^{-1}+qp^{-1}$; $\mathrm{Var}(\mathbf{X})=pq^{-2}+qp^{-2}-2$.
б) $q_k=p^2q^{k-1}+q^2p^{k-1}$; $\mathbf{P}\{\mathbf{X}=m,\ \mathbf{Y}=n\}=p^{m+1}q^n+q^{m+1}p^n$ с $m,\ n\geqslant 1$; $\mathbf{E}(\mathbf{Y})=2$; $\sigma^2=2(pq^{-1}+qp^{-1}-1)$.

17. $\binom{n}{k}364^{n-k}365^{1-n}$.

18. а) $365\{1-364^n\cdot 365^{-n}-n364^{n-1}\cdot 365^{-n}\}$; б) $n\geqslant 28$.

19. а) $\mu=n$, $\sigma^2=(n-1)n$; б) $\mu=(n+1)/2$, $\sigma^2=(n^2-1)/12$.

20. $\mathbf{E}(\mathbf{X})=np_1$; $\mathrm{Var}(\mathbf{X})=np_1(1-p_1)$; $\mathrm{Cov}(\mathbf{X},\ \mathbf{Y})=-np_1p_2$.

21. $-n/36$. Это частный случай задачи 20.

25. $\mathbf{E}(\mathbf{Y}_r)=\displaystyle\sum_{k=1}^{r}\frac{N}{r-k+1}$; $\mathrm{Var}(\mathbf{Y}_r)=\displaystyle\sum_{k=1}^{r}\frac{N(N-r+k-1)}{(r-k+1)^2}$.

26. а) $1-q^k$; б) $\mathbf{E}(\mathbf{X})=N\{1-q^k+k^{-1}\}$; в) $d\mathbf{E}(\mathbf{X})/dk=0$.

27. $\sum(1-p_j)^n$. Положить $\mathbf{X}_j=1$ или 0 в соответствии с тем, представлен или нет j-й класс.

28. $\mathbf{E}(\mathbf{X})=\dfrac{r_1(r_2+1)}{r_1+r_2}$; $\mathrm{Var}(\mathbf{X})=\dfrac{r_1r_2(r_1-1)(r_2+1)}{(r_1+r_2-1)(r_1+r_2)^2}$.

30. $\mathrm{E}(\mathbf{S}_n)=\frac{nb}{b+r}$; $\mathrm{Var}(\mathbf{S}_n)=\frac{nbr\,(b+r+nc)}{(b+r)^2\,(b+r+c)}$.

33. $$\mathrm{E}\left(\frac{r}{\mathbf{X}}\right)=r\sum_{k=r}^{\infty}k^{-1}\binom{k-1}{r-1}p^rq^{k-r}=$$
$$=\sum_{k=1}^{r-1}(-1)^{k-1}\frac{r}{r-k}\left(\frac{p}{q}\right)^k+\left(\frac{-p}{q}\right)^r r\log p.$$

Чтобы вывести последнюю формулу из первой, положить $f(q)=r\sum k^{-1}\binom{k-1}{r-1}q^k$. Используя формулу (12.4) гл. II, найти $f'(q)=rq^{r-1}(1-q)^{-r}$. Утверждение теперь доказывается повторным интегрированием по частям.

Глава XI

1. $sP(s)$ и $P(s^2)$.
2. а) $(1-s)^{-1}P(s)$; б) $(1-s)^{-1}sP(s)$; в) $\{1-sP(s)\}/(1-s)$; г) $p_0s^{-1}+\{1-s^{-1}P(s)\}/(1-s)$; д) $(1/2)\{P(\sqrt{s})+P(-\sqrt{s})\}$.
3. $U(s)=pqs^2/[(1-ps)(1-qs)]$. Математическое ожидание равно $1/(pq)$; дисперсия равна $(1-3pq)/(p^2q^2)$.
6. Производящая функция удовлетворяет квадратному уравнению $A(s)=A^2(s)+s$. Поэтому $A(s)=(1/2)-(1/2)\sqrt{1-4s}$ и $a_n=n^{-1}\binom{2n-2}{n-1}$.
10. а) $\Phi^r(s)F^k(s)\,|\,p-q\,|$;
 б) $\Phi^r(s)\,[1+F(s)+\ldots+F^k(s)]\,|\,p-q\,|$.
11. а) $(q/p)^r\,\Phi^{2r}(s)$; б) $(q/p)^r\,\Phi^{2r}(s)\,U(s)$.
12. Используя производящую функцию случайной величины $\mathbf{X}_\nu$, имеющей геометрическое распределение, сразу получаем
$$P_r(s)=s^r\cdot\frac{N-1}{N-s}\,\frac{N-2}{N-2s}\cdots\frac{N-r+1}{N-(r-1)s}.$$
13. $P_r(s)\{N-(r-1)s\}=P_{r-1}(s)(N-r-1)s$.
14. $$P_r(s)=\frac{s}{N-(N-1)s}\,\frac{2s}{N-(N-2)s}\cdots\frac{rs}{N-(N-r)s}.$$
15. $\mathbf{S}_r$ является суммой r независимых случайных величин с общим геометрическим распределением. Поэтому
$$P_r(s)=\left(\frac{q}{1-ps}\right)^r,\quad p_{r,k}=q^rp^k\binom{r+k-1}{k}.$$
16. а) $$\mathbf{P}\{\mathbf{R}=r\}=\sum_{k=0}^{\nu-1}\mathbf{P}\{\mathbf{S}_{r-1}=k\}\,\mathbf{P}\{\mathbf{X}_r\geqslant\nu-k\}=$$
$$=\sum_{k=0}^{\nu-1}q^{r-1}p^k\binom{r+k-2}{k}p^{\nu-k}=p^\nu q^{r-1}\binom{r+\nu-2}{\nu-1},$$
$$\mathrm{E}(\mathbf{R})=1+q^\nu/p,\quad \mathrm{Var}(\mathbf{R})=\nu q/p^2;$$
 б) $$(p_1p_2)^N\sum_{\nu=1}^{\infty}\binom{N+\nu-2}{\nu-1}^2(q_1q_2)^{\nu-1}.$$

17. Учесть, что
$$1+s+\ldots+s^{ab-1}=(1+s+\ldots+s^{a-1})\,(1+s^a+s^{2a}+\ldots+s^{(b-1)\,a}).$$

21. $u_n=q^n+\sum\limits_{k=3}^{n}\binom{k-1}{2}p^3q^{k-3}u_{n-k}$, где $u_0=1$, $u_1=q$, $u_2=q^2$, $u_3=p^3+q^3$. Используя тот факт, что это рекуррентное соотношение является соотношением типа свертки, получаем
$$U(s)=\frac{1}{1-qs}+\frac{(ps)^3}{(1-qs)^3}\,U(s).$$

22. $u_n=pw_{n-1}+qu_{n-1}$, $v_n=pu_{n-1}+qv_{n-1}$, $w_n=pv_{n-1}+qw_{n-1}$. Поэтому $U(s)-1=psW(s)+qsU(s)$, $V(s)=psU(s)+qsV(s)$, $W(s)=psV(s)+qsW(s)$.

Глава XIII

1. Достаточно показать, что если $F(s)=1$ и $s\neq 1$, то $|s|\geqslant 1$, причем равенство $|s|=1$ возможно только в периодическом случае.

2. $u_{2n}\left\{\binom{2n}{n}2^{-2n}\right\}^r\sim\dfrac{1}{\sqrt{(\pi n)^r}}$. Поэтому $\mathscr{E}$ возвратно только при $r=2$. При $r=3$ численное интегрирование методом касательных дает [1])
$$u-1=\sum_{n=1}^{\infty}u_{2n}\approx(1/\sqrt{\pi^3})\int\limits_{1/2}^{\infty}\left[1/\sqrt{x^3}\right]dx=\sqrt{(2/\pi)^3}\approx 1/2.$$

3. $u_{6n}\sim\sqrt{6/(2\pi n)^5}$. Поэтому $u-1\approx\sqrt{6/(2\pi)^5}\int\limits_{1/2}^{\infty}x^{-5/2}\,dx$. Значит [2]), $u\approx$ $\approx 1{,}047$ и $f\approx 0{,}045$.

4. $u_{(\lambda+1)\,n}=\binom{(\lambda+1)\,n}{n}p^{\lambda n}q^n$. Верхняя грань отношений последовательных членов меньше 1, если $p\neq\lambda/(\lambda+1)$. (Это утверждение следует также из закона больших чисел.)

6. Из оценки $\sum f_i+\mathbf{P}\{\mathbf{X}_1>0\}\leqslant 1$ вывести, что $f<1$, если $\mathbf{P}\{\mathbf{X}_1>0\}\neq 0$. Если $\mathbf{P}\{\mathbf{X}_1>0\}=0$, то все $\mathbf{X}_i\leqslant 0$ и $\mathscr{E}$ либо происходит при первом испытании, либо не происходит никогда.

7. $\mathbf{Z}_n=\min\{m\colon m\geqslant(n-\mathbf{N}_n)/r,\ m$ целое$\}$. Далее, $\mathbf{E}(\mathbf{Z}_n)\sim np/(q+pr)$, $\operatorname{Var}\mathbf{Z}_n\sim npq\,(q+pr)^{-3}$.

8. $G(s)=\dfrac{(1-qs)\,q^rs^r}{1-s+pq^rs^{r+1}}$, $F(s)=qs+psG(s)$, $\mu=q^{-r}$.

9. $G(s)=\dfrac{(1-qs)\,B(qs)}{1-s+psB(qs)}$, $F(s)$ та же, что в задаче 8.

11. $\mathbf{N}_n^*\approx(\mathbf{N}_n-714{,}3)/22{,}75$; $\mathfrak{N}(2/3)-\mathfrak{N}(-2/3)\approx 1/2$.

12. $r_0=r_1=r_2=1$, $r_n=r_{n-1}-(1/4)\,r_{n-2}+(1/8)\,r_{n-3}$,
$$R(s)=(8+2s^2)\,(8-8s+2s^2-s^3)^{-1},\qquad r_n\sim 1{,}444248\,(1{,}139680)^{n-1}.$$

14. Если a_n — вероятность появления A-серии длины r при n-м испытании, то $A(s)$ определяется формулой (7.5) с заменой p на α и q на $1-\alpha$,

---

[1]) Подынтегральное выражение плохо приближает первые члены суммы, дающие в нее основной вклад. Более громоздкие вычисления показывают, что $u-1\sim 0{,}393$, и поэтому $f=1-u^{-1}\approx 0{,}282$. — *Прим. перев.*

[2]) См. примечание к предыдущему ответу. Более точные значения: $u\approx 1{,}0225$ и $f\approx 0{,}0220$. — *Прим. перев.*

Пусть $B(s)$ и $C(s)$ — соответствующие функции для B- и C-серий. Искомыми производящими функциями являются $[(1-s)U(s)]^{-1}$, где в случае а) $U(s)=A(s)$, в случае б) $U(s)=A(s)+B(s)-1$, в случае в) $U(s)=A(s)+B(s)+C(s)-2$.

15. Скомбинировать методы, использованные в примере 8, б) и в решении задачи 14.
16. Математическое ожидание для возраста k равно Npq^k.
18. $w_k(n)=v_{n-k}r_k$, если $n>k$, и $w_k(n)=\beta_{k-n}r_k/r_{k-n}$, если $n\leqslant k$.
19. Учесть, что $1-F(s)=(1-s)Q(s)$ и $\mu-Q(s)=(1-s)R(s)$, причем $Q(1)=\mu$, $2R(1)=\sigma^2-\mu+\mu^2$. Степенной ряд для $Q^{-1}(s)=\sum(u_n-u_{n-1})s^n$ при $s=1$ сходится.

Глава XIV

1. $((q/p)^b-1)/((q/p)^{a+b}-1)$ при $p\neq q$ и $b/(a+b)$ при $p=q$.
3. При $q<p$ число попаданий является дефектной случайной величиной.
4. Математическое ожидание числа попаданий равно $p(1-q_1)/(qq_6-1)=(p/q)^a$.
5. Вероятность разорения по-прежнему задается формулой (2.4) при $p=\alpha(1-\gamma)^{-1}$, $q=\beta(1-\gamma)^{-1}$. Ожидаемая продолжительность игры равна $D_z(1-\gamma)^{-1}$, где D_z определяется по (3.4) или (3.5).
6. Граничные условия (2.2) заменяются условиями $q_0-\delta q_1=1-\delta$, $q_a=0$. Решению (2.4) соответствует решение
$$q_z=\frac{\{(q/p)^a-(q/p)^z\}(1-\delta)}{(q/p)^a(1-\delta)+\delta q/p-1}.$$
Граничные условия (3.2) превращаются в $D_0=\delta D_1$, $D_a=0$.
7. Соотношению (2.1) здесь соответствует уравнение $q_z=pq_{z+2}+qq_{z-1}$, и $q_z=\lambda^z$ будет частным решением последнего, если $\lambda=p\lambda^3+q$, т. е. если $\lambda=1$ или $\lambda^2+\lambda=qp^{-1}$. Вероятность разорения равна
$$q_z=\begin{cases}1, & \text{если } q\geqslant 2p,\\ \left(\sqrt{1/4+q/p}-1/2\right)^z, & \text{если } q\leqslant 2p.\end{cases}$$
10. $w_{z,n+1}(x)=pw_{z+1,n}(x)+qw_{z-1,n}(x)$ с граничными условиями 1) $w_{0,n}(x)=w_{a,n}(x)=0$, 2) $w_{z,0}(x)=0$ при $z\neq x$ и $w_{x,0}(x)=1$.
11. Заменить 1) на $w_{0,n}(x)=w_{1,n}(x)$ и $w_{a,n}(x)=w_{a-1,n}(x)$.
12. Граничное условие: $u_{a,n}=u_{a-1,n}$. Производящая функция:
$$\frac{\lambda_1^z(s)\lambda_2^{a-1/2}(s)+\lambda_2^z(s)\lambda_1^{a-1/2}(s)}{\lambda_1^{a-1/2}(s)+\lambda_2^{a-1/2}(s)}=\frac{\lambda_1^{a-z-1/2}(s)+\lambda_2^{a-z-1/2}(s)}{\lambda_1^{a-1/2}(s)+\lambda_2^{a-1/2}(s)}.$$
18. $\mathbf{P}\{\mathbf{M}_n<z\}=\sum_{x=1}^{\infty}(v_{x-z,n}-v_{x+z,n})$, $\mathbf{P}\{\mathbf{M}_n=z\}=\mathbf{P}\{\mathbf{M}_n<z+1\}-\mathbf{P}\{\mathbf{M}_n<z\}$.
19. Первое достижение x должно произойти в момент $k\leqslant n$, и за последующие $n-k$ шагов частица возвращается в x.
31. Соотношение (8.2) заменяется на
$$U_z(s)=s\sum_{x=1}^{a-1}U_x(s)\,p_{x-z}+sr_z.$$
Характеристическое уравнение имеет вид $s\sum p_k\sigma^k=1$.

Глава XV

1. P имеет строки $(p, q, 0, 0)$, $(0, 0, p, q)$, $(p, q, 0, 0)$ и $(0, 0, p, q)$. При $n>1$ строки таковы: (p^2, pq, pq, q^2).

2. а) Цепь неприводима и эргодична; $p_{jk}^{(n)} \longrightarrow 1/3$ при всех j, k. (Заметим, что P дважды стохастическая.)
б) Цепь имеет период 3, класс G_0 содержит E_1 и E_2; состояние E_4 образует класс G_1, а E_3 — класс G_2. Имеем $u_1 = u_2 = 1/2$, $u_3 = u_4 = 1$.
в) Состояния E_1 и E_3 образуют замкнутое множество S_1, а E_4 и E_5 — другое замкнутое множество S_2, тогда как E_2 невозвратно. Матрицы, соответствующие замкнутым множествам, суть (2×2)-матрицы с элементами $1/2$. Следовательно, $p_{jk}^{(n)} \longrightarrow 1/2$, если E_j и E_k принадлежат одному и тому же S_r; $p_{j2}^{(n)} \longrightarrow 0$; наконец, $p_{2k}^{(n)} \longrightarrow 1/2$, если $k = 1$, 3, и $p_{2k}^{(n)} \longrightarrow 0$, если $k = 2$, 4, 5.
г) Цепь имеет период 3. Полагая $a = (0, 0, 0, 1/3, 1/3, 1/3)$, $b = (1, 0, 0, 0, 0, 0)$, $c = (0, 1/2, 1/2, 0, 0, 0)$, мы находим, что строки матриц $P^2 = P^5 = \dots$ суть a, b, b, c, c, c, строки $P^3 = P^6 = \dots$ суть b, c, c, a, a, a, а строки $P = P^4 = \dots$ суть c, a, a, b, b, b.

3. $p_{jj}^{(n)} = (j/6)^n$, $p_{jk}^{(n)} = (k/6)^n - ((k-1)/6)^n$ при $k > j$ и $p_{jk}^{(n)} = 0$ при $k < j$.

4. $x_k = (3/4, 1/2, 1/4, 1/2)$, $y_k = (1/4, 1/2, 3/4, 1/2)$.

6. При $n \geqslant j$

$$f_{j0}^{(n)} = \binom{n-1}{j-1} p^{n-j} q^j = \binom{i}{n-j} (-p)^j q^j.$$

Производящая функция $(qs)^j (1 - ps)^{-j}$. Математическое ожидание равно j/q.

7. $f_{00}^{(n)} = \sum\limits_{k-1}^{n=1} v_k \binom{n-2}{k-1} p^{n-1-k} q^k$ при $n > 1$.

8. Состояния с четными номерами образуют неприводимое замкнутое множество. Вероятность возвращения в E_0 не более чем за n шагов равна

$$1 - v_0 + v_0(1 - v_2) + v_0 v_2 (1 - v_4) + \dots + v_0 v_2 \dots v_{2n-2}(1 - v_{2n}) = \\ = 1 - v_0 v_2 v_4 \dots v_{2n}.$$

Поэтому состояния с четными номерами возвратны тогда и только тогда, когда последнее произведение стремится к 0. Вероятность того, что выходящая из E_{2r+1} система останется навсегда среди состояний с нечетными номерами (невозвратных состояний), равна $v_{2r+1} v_{2r+3} \dots$.

9. $u_r = [1 - p/q]\,(p/q)^{r-1}\,[1 - (p/q)^\rho]^{-1}$.

10. Возможные состояния $E_0, \dots, E_w$. При $j > 0$

$$p_{j,\,j-1} = j(\rho - w + j)\rho^{-2}, \quad p_{j,\,j+1} = (\rho - j)(w - j)\rho^{-2},$$
$$p_{jj} = j(w - j)\rho^{-2} + (\rho - j)(\rho - w + j)\rho^{-2},$$
$$u_k = \binom{w}{k}\binom{b}{\rho - k}\binom{2\rho}{\rho}^{-1}.$$

13.
$$P = \begin{bmatrix} q & p & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \\ q & p & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

14. Заметим, что матрица дважды стохастическая; использовать пример 7, з).

15. Положить $p_{k,\,k+1} = 1$ при $k = 1, \dots, N-1$ и $p_{Nk} = p_k$.

16. $\sum u_j p_{jk} = u_k$; тогда $U(s) = u_0(1 - s)P(s)\{P(s) - s\}^{-1}$. Для эргодичности необходимо и достаточно, чтобы $\mu = P'(1) < 1$. По правилу Лопиталя $U(1) = u_0(1 - \mu)$, откуда $u_0 = (1 - \mu)^{-1}$.

25. Если $n \geqslant m - 2$, то величины $\mathbf{X}^{(m)}$ и $\mathbf{X}^{(n)}$ независимы и, следовательно, все три строки матрицы $p_{jk}^{(m, n)}$ совпадают с распределением $\mathbf{X}^{(n)}$, а именно с (1/4, 1/2, 1/4). При $n=m+1$ строки матрицы суть (1/2, 1/2, 0), (1/4, 1/2, 1/4), (0, 1/2, 1/2).

Глава XVII

3. $\mathbf{E}(\mathbf{X}) = ie^{\lambda t}$, $\operatorname{Var}(\mathbf{X}) = ie^{\lambda t}(e^{\lambda t}-1)$.
4. $P_n' = -\lambda n P_n + \lambda(n+1) P_{n+1}$,
$$P_n(t) = \binom{i}{n} e^{-i\lambda t}(e^{\lambda t}-1)^{i-n} \quad (n \leqslant i),$$
$$\mathbf{E}(\mathbf{X}) = ie^{-\lambda t}, \quad \operatorname{Var}(\mathbf{X}) = ie^{-\lambda t}(1-e^{-\lambda t}).$$
6. $P_n'(t) = -(\lambda + n\mu) P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t)$ при $n \leqslant N-1$ и $P_N'(t) = -N\mu P_N(t) + \lambda P_{N-1}(t)$.
19. Обычный метод решения линейных дифференциальных уравнений приводит к системе линейных уравнений.

ИМЕННОЙ УКАЗАТЕЛЬ

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ОГЛАВЛЕНИЕ

Вильям Феллер

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ И ЕЕ ПРИЛОЖЕНИЯ

В 2-х томах

Том 1

Ст. научный редактор Г. М. Ильичева
Мл. научный редактор Т. А. Денисова
Художник Е. И. Волков
Художественный редактор В. И. Шаповалов
Технический редактор Н. И. Борисова
Корректор Е. Г. Литвак

И Б № 3127

Сдано в набор 26.09.83. Подписано к печати 18.04.84. Формат $60 \times 90^{1}/_{16}$. Бумага типографская № 2. Объем 16,50 бум. л. Усл. кр.-отт. 33,00. Гарнитура литературная. Печать высокая. Усл. печ. л. 33,00. Уч.-изд. л. 33,28. Изд. № 1/1766. Тираж 40 000 экз. Заказ № 221. Цена 2 р. 60 к.

ИЗДАТЕЛЬСТВО «МИР»

129820, Москва, И-110, ГСП, 1-й Рижский пер., 2

Набрано и сматрицировано в ордена Октябрьской Революции и ордена Трудового Красного Знамени Первой Образцовой типографии имени А. А. Жданова Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. Москва, М-54, Валовая, 28

Отпечатано в Ленинградской типографии № 2 головном предприятии ордена Трудового Красного Знамени Ленинградского объединения «Техническая книга» им. Евгении Соколовой Союзполиграфпрома при Государственном комитете СССР по делам издательств, полиграфии и книжной торговли. 198052, г. Ленинград, Л-52, Измайловский проспект, 29.