БИБЛИОТЕКА
ИНОСТРАННЫХ
КНИГ ДЛЯ
ЭКОНОМИСТОВ
И СТАТИСТИКОВ

Г. ДЭВИД

МЕТОД ПАРНЫХ СРАВНЕНИЙ

БИБЛИОТЕЧКА
ИНОСТРАННЫХ КНИГ
ДЛЯ ЭКОНОМИСТОВ
И СТАТИСТИКОВ

THE METHOD OF PAIRED COMPARISONS

H. A. DAVID
University of North Carolina

Being NUMBER TWELVE of
GRIFFIN'S STATISTICAL
MONOGRAPHS & COURSES

EDITED BY
ALAN STUART, D. Sc. (Econ.)

Second Impression

CHARLES GRIFFIN & COMPANY LIMITED
LONDON

Г. ДЭВИД

МЕТОД ПАРНЫХ СРАВНЕНИЙ

Перевод с английского
Н. КОСМАРСКОЙ и Д. ШМЕРЛИНГА

Под редакцией Ю. АДЛЕРА

С приложением к русскому переводу

МОСКВА «СТАТИСТИКА» 1978

БИБЛИОТЕЧКА ИНОСТРАННЫХ КНИГ ДЛЯ ЭКОНОМИСТОВ И СТАТИСТИКОВ

Издательство «Статистика» выпускает на русском языке серию книг иностранных авторов по статистике, рассчитанных на круг читателей, нуждающихся в пополнении своих математических и статистических знаний. Задача книг — ознакомить статистиков и экономистов не на очень сложном материале с современными методами, которые за рубежом применяются в экономическом анализе и в различных хозяйственных расчетах.

Среди намеченных к выпуску как книги по общим вопросам статистики, так и книги, посвященные статистическому анализу в отдельных областях экономики. Издательство старается подбирать работы, не перегруженные сложными теоретическими изысканиями, но подводящие к применению таких изысканий на практике.

Вышли из печати книги:

1. М. Броуди. **О статистическом рассуждении.**
2. А. Бернстейн. **Справочник статистических решений.**
3. У. Дж. Рейхман. **Применение статистики.**
4. Х. Крыньский. **Математика для экономистов.**
5. С. Даймендд. **Мир вероятностей.**
6. А. Хьютсон. **Дисперсионный анализ.**
7. С. Лизер. **Эконометрические методы и задачи.**
8. Эм. Борель, Р. Дельтейль, Р. Юрон. **Вероятности, ошибки.**
9. **Статистические методы исследования корреляций в экономике.**
10. Л. Столерю. **Равновесие и экономический рост.**
11. Я. Окунь. **Факторный анализ.**
12. С. Сирл, У. Госман. **Матричная алгебра в экономике.**
13. Е. Грень. **Статистические игры и их применение.**
14. Д. Тёрнер. **Вероятность, статистика и исследование операций.**
15. Э. Кейн. **Экономическая статистика и эконометрия. Вып. 1.**
16. Э. Кейн. **Экономическая статистика и эконометрия. Вып. 2.**
17. Э. Колкот. **Проверка значимости.**

Подготавливаются к изданию:

Вайнберг Дж., Шумекер Дж. **Статистика. Логический подход.**
Кенуй М. Г. **Быстрые статистические вычисления.**

$Д \dfrac{10805^1\text{-}091}{008(01)\text{-}78} 38\text{-}78$

[1] Второй индекс 10803.

ПРЕДИСЛОВИЕ

Метод парных сравнений представляет интерес для исследователей разных направлений. Важный вклад в развитие этого метода был сделан не только статистиками, но и психологами, которые первыми начали его изучать, а также математиками и экономистами. Приложения этого метода интересуют еще более широкие круги специалистов.

В этой монографии делается попытка дать единую картину современного состояния предмета. Помимо проблем планирования и анализа эксперимента в книге обсуждается выбор подходящих вероятностных моделей, а также намечены возможные приложения. Особое значение придавалось простым, широко применяемым процедурам, основанным на минимуме конкретных допущений. Работая над книгой, как статистик я избегал специальных терминов из других областей. Читатель не найдет в книге упоминаний о психологических континуумах или упаковках, но если он знает, что это такое, то узнает их и путешествующими по книге инкогнито. В книге несколько сносок, отсылающих читателя к стандартным учебникам в надежде помочь не знакомому со статистикой читателю в преодолении возможных препятствий.

Я весьма благодарен доктору М. Дж. Кендэлу, вдохновившему меня на написание этой работы. Мне приятно выразить свою признательность доктору Д. Ф. Моррисону за внимательное чтение рукописи. Я благодарен Р. К. Бозе, Р. А. Брэдли, В. Г. Клэтуорси, Л. Р. Форду мл., С. М. Джонсону, Т. Г. Старксу, М. Е. Терри, Б. Дж. Травинскому, С. Юэре, а также редакторам соответствующих журналов за разрешение воспроизвести в книге таблицы и рисунки из работ этих авторов. Работа была написана при поддержке научно-исследовательского центра армии США, Дурхэм, Северная Каролина.

Г. А. Дэвид

I ГЛАВА ВЕРОЯТНОСТНЫЕ МОДЕЛИ

1.1. ВВЕДЕНИЕ

Shall I compare thee to a summer's day?*

В методе парных сравнений объекты предъявляются попарно одному или нескольким экспертам. Мы используем «объект» в качестве общего термина, которым можно заменять термины «вопрос», «обработка», «стимул» и им подобные. Основной элементарный экспериментальный акт — сравнение двух объектов A и B единственным экспертом, который в простейшей ситуации должен выбрать один из них. Мы будем говорить, что эксперт предпочитает данный объект, хотя выбор не обязательно будет выражать его предпочтение. В более общих случаях эксперт может провозгласить еще и равенство объектов или зафиксировать свои предпочтения на некоторой более тонкой шкале.

Сравнение A и B может выполняться всеми экспертами. Если же рассматривается более чем два объекта, то легко сделать так, чтобы каждый эксперт производил каждое возможное парное сравнение. Эту ситуацию можно назвать «сбалансированным экспериментом парных сравнений» и уподобить, говоря языком спорта, круговому турниру. Роль игроков в турнире аналогична роли объектов в эксперименте парных сравнений. Для t объектов и n экспертов число элементарных парных сравнений равно $n\binom{t}{2}$.

Метод парных сравнений первоначально применялся в случаях, когда сравниваемые объекты можно было сопоставить лишь субъективно, т. е. когда невозможно или невыгодно делать соответствующие измерения для того, чтобы решить, который из двух объектов предпочтительнее. Отсюда можно заключить, что парные сравнения широко используются психометриками. И действительно, метод в его первоначальной форме был введен Фехнером [41]** и после значительного развития стал популярным благодаря Тэрстоуну ([132], [133]).

*«Сравню ли с летним днем твои черты?» (Шекспир В. Сонеты. М., «Худож. лит.», 1963, сонет № 18). — *Прим. ред.*

** См. также Бардин К. В. [153]. — *Прим. пер.*

Наиболее распространенными приложениями были дегустация, изучение поведения потребителей, визуальная колориметрия, определение индивидуальных рейтингов и вообще изучение предпочтений и поведения при выборе. Конечно, есть и другие методы изучения предпочтений, но мы не будем вдаваться в подробное обсуждение их достоинств, в частности потому, что существует несколько обширных обзоров на эту тему ([62], [90], [86], [137]). Метод парных сравнений иногда служит тривиальной экспериментальной процедурой, как в опробовании бритв различных марок, где действие двух вариантов можно сравнивать на двух щеках одного мужчины. С другой стороны, при дегустации эксперту зачастую трудно справиться более чем с двумя «вкусами» и при добавлении третьего он может совершенно запутаться.

В иных случаях эксперты бывают способны сравнивать несколько объектов сразу. Если это легко сделать, простая ранжировка всех объектов может оказаться более предпочтительной. Однако когда различия между объектами невелики, желательно сравнивать каждую пару как можно более свободно от любых посторонних влияний, вызванных присутствием других объектов. Так, метод парных сравнений имеет некоторые преимущества, когда необходима ответственная экспертиза. Следует помнить, что ранжировка получается быстро только при вполне очевидных различиях; в противном случае процесс ранжирования практически требует многократного повторения попыток попарных сравнений «соседей», прежде чем будет достигнуто разумное упорядочение. В таких случаях ранжирование становится непрактичным, если объектов много. Вопрос о том, следует ли добиваться получения полностью удовлетворительного ранжирования, мы еще вкратце обсудим.

Определение степени предпочтения или числа очков для сравниваемых объектов — другая сторона дела, которую надо иметь в виду. Если несколько степеней предпочтения различимы, при разумном согласии экспертов такие методы упорядочения имеют то преимущество, что результаты можно обрабатывать почти как обычные измерения. А при их интерпретации можно, как правило, без серьезной ошибки пользоваться стандартными статистическими методами. Упорядочение менее успешно, когда различия малы и степени предпочтения трудно определить. В парных сравнениях тоже можно делать отметки на шкале, скажем, с 7 делениями ($s = 3, 2, 1, 0, -1, -2, -3$), читаемой следующим образом: «сильное предпочтение A», «предпочтение A», «слабое предпочтение A», «отсутствие предпочтения», «слабое предпочтение B» и т. д. Это компромисс, сохраняющий желательную точность сравнения между парами объектов. Так как процедура упорядочения по степеням предпочтения осуществляется индивидуально для каждого объекта, при ее выполнении трудно сохранить непротиворечивость, значение s — сравнительное значение. Конечно, успех зависит от того, существуют ли различия, достаточно ясные для лучшей классификации, чем просто установление предпочтения. Простейшая форма парных сравнений уменьшает зону возможной несогласованности между экспертами до минимума; даже допущение «ничьих» создает трудности, поскольку одни эксперты могут объявлять объекты равными более легко, чем другие.

1.2. ПОПАРНЫЕ СРАВНЕНИЯ ТРЕХ ОБЪЕКТОВ

Мы начнем со случая, когда сравниваются попарно три объекта A, B, C. В этой простой ситуации проявляются многие существенные черты экспериментов парных сравнений. Если не утверждается обратное, будет предполагаться, что каждое суждение состоит из простого предпочтения того или иного объекта. В частности, «ничьи» не допускаются.

Для каждого из трех сравнений (AB), (AC), (BC) возможны два исхода, так что есть 8 разных экспериментальных результатов. Из них 6 результатов типа

$$A \to B, \quad A \to C, \quad B \to C,$$

где стрелка означает «предпочтительнее, чем», удобно представить как тройки чисел[1] [2 1 0], которые указывают, что один объект (не обязательно A) «выигрывает» два раза, другой — один раз, а третий — ни разу. Два оставшихся результата

$$A \to B, \; B \to C, \; C \to A,$$

$$A \leftarrow B, \; B \leftarrow C, \; C \leftarrow A,$$

имеют представление [1 1 1] или $[1^3]$ и названы (см. [93]) *циклическими триадами**.

Ясно, что циклическая триада означает непоследовательность в суждениях части экспертов и ее простейшее объяснение состоит в том, что эксперт иногда гадает, объявляя свои предпочтения. Он может гадать из-за своей некомпетентности или потому, что на самом деле объекты весьма похожи; вероятность циклической триады равна 1/4, когда объекты идентичны. Но гадание — не единственное объяснение. Приемлемого упорядочения трех объектов может не быть, даже если они заметно различаются. Их важность может зависеть более чем от одного показателя, а попытка упорядочить их на линейной шкале может оказаться несколько искусственной. При этих обстоятельствах эксперт должен сконструировать некоторую функцию от соответствующих показателей и пользоваться ею как основанием для сравнения. Неудивительно, что при изучении сложных предпочтений эта функция неизвестна, да еще может изменяться от одного парного сравнения к другому, особенно когда различные пары объектов побуждают эксперта концентрировать свое внимание на различных признаках. Это последнее обстоятельство объясняет те ситуации, где некоторые циклические триады часто встречаются в повторных опытах. Аналогичные случаи обычны в турнирах, где в повторных встречах игроков или команд A, B, C возможно, что, как правило, $A \to B$, $B \to C$, но C побеждает A чаще, чем проигрывает. Фокус с камешком, ножницами и бумагой доставляет крайний пример.

Важное свойство метода парных сравнений — возможность проявления таких противоречий, а это путь к некоторым критериям,

---

[1] Мы несколько отходим от стандартных обозначений, записывая нули.

* Или циклическими тройками, нетранзитивными тройками. — *Прим. пер.*

описываемым в следующих двух главах. Стоит принять любые доступные меры для обеспечения независимости отдельных парных сравнений или чего-то близкого к ней. Это не так трудно в дегустации, где идентичность изучаемых предметов может быть скрыта, чему способствует и то, что в эксперименте обычно число сравнений больше трех. С другой стороны, в индивидуальном рейтинге есть реальная опасность, что из-за хорошей памяти эксперта его парные сравнения выродятся в ранжирование, если число сравниваемых людей не очень велико. Частичный выход из этого положения — выполнение одним экспертом только части из всех возможных парных сравнений. Это порождает интересные задачи планирования эксперимента (см. гл. 5). Если в некотором эксперименте не была достигнута приближенная независимость, получается ситуация промежуточная между прямым ранжированием и независимыми парными сравнениями. Тогда приходится вести анализ обоими методами. И только если результаты согласуются, заключение можно сделать с некоторой уверенностью.

1.3. ОСНОВНЫЕ МОДЕЛИ

Предшествующее обсуждение можно формализовать в виде ряда подходящих моделей, которые накладывают все более жесткие ограничения на вероятности предпочтений. Предположим, что t объектов $A_1, A_2, \ldots, A_t$ сравниваются попарно каждым из n экспертов и эксперт с номером γ делает r_γ повторных сравнений для каждой из $\binom{t}{2}$* возможностей. Пусть

$$x_{ij\gamma\delta},\ i, j = 1, 2, \ldots, t,\ i \neq j;\ \gamma = 1, 2, \ldots, n; \quad \delta = 1, 2, \ldots, r_\gamma$$

типичная случайная величина, принимающая значение 1 или 0 в зависимости от того, предпочитает ли эксперт γ A_i или A_j в δ-м сравнении двух объектов. Мы постулируем всюду, что все сравнения статистически независимы, так что $x_{ij\gamma\delta}$ взаимно независимы, если не считать того, что $x_{ij\gamma\delta} + x_{ji\gamma\delta} = 1$. Пусть**

$$\Pr(x_{ij\gamma\delta} = 1) = \pi_{ij\gamma\delta}. \tag{1.3.1}$$

Есть лишь такие ограничения на π:

$$0 \leqslant \pi_{ij\gamma\delta} \leqslant 1 \text{ и } \pi_{ji\gamma\delta} = 1 - \pi_{ij\gamma\delta}. \tag{1.3.2}$$

Представляют интерес очевидные частные случаи:

$$\pi_{ij\gamma\delta} = \pi_{ij\gamma} \text{ (нет эффекта от повторений)}, \tag{1.3.3}$$

$$\pi_{ij\gamma\delta} = \pi_{ij} \text{ (нет эффекта от повторений и от экспертов)}. \tag{1.3.4}$$

Если, как это бывает, эффект повторений незначителен, что не относится к различиям между экспертами, последняя модель все еще

---

* $\binom{t}{2} = C_t^2$. — *Прим. пер.*

** Pr (A) — вероятность события A. — *Прим. пер.*

слишком обща для описания единичного предпочтения ($\gamma = 1$). Возможный эффект, обусловленный порядком представления объектов в паре, игнорируется в этой книге (см., однако, 7.4). Если применима модель (1.3.4), объекты могут быть проранжированы в соответствии со значениями средних вероятностей предпочтения*

$$\pi_{i.} = \frac{1}{t-1} \sum_{j}{}' \pi_{ij}, \tag{1.3.5}$$

причем суммирование распространяется на все значения j, кроме $j = i$. Здесь из $\pi_{i.} \geqslant \pi_{j.}$ не следует $\pi_{ij} \geqslant \frac{1}{2}$; фактически π_{ij} может принимать все значения от 0 до 1. Другой способ ранжирования возможен, если для каждой тройки разных объектов A_i, A_j, A_k выполняются следующие условия *стохастической транзитивности*:

$$\pi_{ij} \geqslant \frac{1}{2}\,; \quad \pi_{jk} \geqslant \frac{1}{2} \text{ влечет } \pi_{ik} \geqslant \frac{1}{2}\,. \tag{1.3.6}$$

Условия (1.3.6) ведут к вполне определенному упорядочению каждой тройки и, следовательно, всех t объектов. Легко видеть, что эта ранжировка не обязательно совпадает с тем, что получается из (1.3.5).

Более строгой, чем (1.3.6), является модель *усиленной стохастической транзитивности***:

$$\pi_{ij} \geqslant \frac{1}{2}\,; \quad \pi_{jk} \geqslant \frac{1}{2} \text{ влечет } \pi_{ik} \geqslant \max(\pi_{ij}, \pi_{jk}). \tag{1.3.7}$$

Если к (1.3.7) мы добавим:

$$\pi_{ij} = \frac{1}{2} \text{ влечет, что } A_i \sim A_j \; (A_i \text{ эквивалентен } A_j), \tag{1.3.8}$$

то $\pi_{ik} = \pi_{jk}$ при $A_i \sim A_j$ и $\pi_{ik} = \pi_{ij}$, если $A_k \sim A_j$. Тогда объекты A_i, A_j, A_k можно представить точками на прямой, на которой вероятность предпочтения, соответствующая интервалу (A_k, A_i), будет

больше, чем вероятности, соответствующие подынтервалам (собственным) (A_k, A_j), (A_j, A_i), а (1.3.8) относится к совпадающим точкам. Это представление легко распространить на все t объектов на одной прямой. Равные расстояния не обязательно соотносятся с равными вероятностями, но их можно сделать равными подходящим растяжением прямой, что ведет к *линейной модели*.

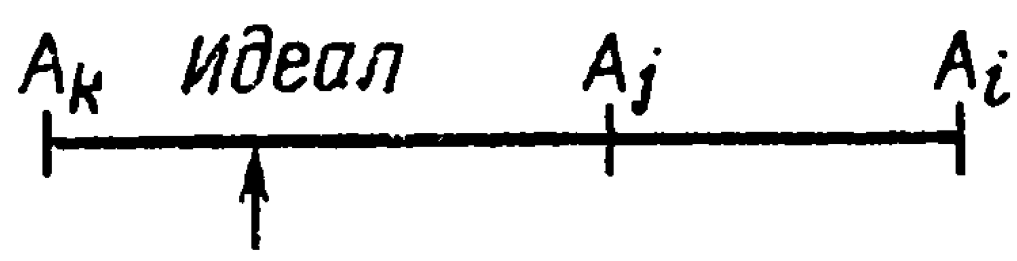
Интересный случай, когда можно ожидать стохастической, но не усиленной стохастической транзитивности, упоминается в [25]. Для

---

* То есть если $\pi_{i.} \geqslant \pi_{j.}$, то ранг A_i меньше ранга A_j или равен ему. — *Прим. пер.*

** Условия стохастической транзитивности обсуждаются в [319], [296], [272]. — *Прим. пер.*

иллюстрации этой, конечно, более общей ситуации, положим, что A_i, A_j, A_k — три серых объекта с уменьшающейся интенсивностью цвета. Эксперта просят дать предпочтение в каждом парном сравнении тому объекту, который, как он считает, лучше *представляет* серый цвет. Здесь кажется разумным предположить, что эксперт имеет некий идеал серого цвета и ранжирует другие оттенки по степени их близости к этому идеалу. Если относительное расположение объектов на шкале «серости» такое, как показано ниже,



то $\pi_{kj} > \frac{1}{2}$, $\pi_{ji} > \frac{1}{2}$, $\pi_{ki} > \frac{1}{2}$, но π_{ki} может получиться также меньше, чем π_{ji}, потому что сравнения по одну сторону от идеала более надежны, чем на противоположных сторонах. Это экспериментально подтверждено Кумбсом. Аналогичного эффекта можно опасаться, когда тремя объектами служат часы стоимостью 10 фунтов, 12 фунтов и еще незначительно большей (но неизвестной) стоимости (см. также [99]).

Линейная модель

Предположим, что объект A_i обладает истинной ценностью или «заслугой» V_i при оценке по некоторому признаку. Можно представить t истинных ценностей V_1, V_2, ..., V_t точками на шкале ценностей. *Наблюдаемая* (эмпирическая) ценность объекта A_i будет меняться от наблюдения к наблюдению и может быть представлена на некоторой шкале как непрерывная случайная величина y_i $(-\infty \leqslant y_i \leqslant \infty)$. В парном сравнении A_i с A_j первый из объектов будет предпочитаться второму, если $y_i > y_j$, второй — первому, если $y_i < y_j$. Если можно построить шкалу ценностей так, чтобы вероятность предпочтения

$$\pi_{ij} = \Pr(y_i - y_j > 0)$$

выражалась для всех i, j как $H(V_i - V_j)$, где $H(x)$ возрастает монотонно от $H(-\infty) = 0$ до $H(\infty) = 1$ и $H(-x) = 1 - H(x)$, то можно говорить, что вероятности предпочтений удовлетворяют условиям *линейной модели**. Очевидно, $H(x)$ — функция распределения случайной величины, симметричной относительно нуля.

Поскольку функция $H(x)$ делает π_{ij} зависящими лишь от разностей $V_i - V_j$, все π_{ij} могут быть выражены как функции $(t-1)$независимых разностей V_i с $\pi_{ij} \gtreqless \frac{1}{2}$, соответствующими $V_i \gtreqless V_j$. В силу этого ценности удобно представлять t точками на линейной шкале с произвольным началом отсчета, отсюда термин «линейная». Мы можем также говорить, что изучаемый *показатель* удовлетворяет линейной модели** или построен в соответствии с линейной шкалой, если ценность

* Модели типа Тэрстоуна.— *Прим. пер.*

** См. [197], [198], где обсуждается эта модель. — *Прим. пер.*

любого объекта (а не только t объектов A_i) может быть представлена на этой шкале.

Линейная модель — это обобщение модели Тэрстоуна—Мостеллера, в котором предполагается, что y_i — нормально распределенные случайные величины $N(V_i, \sigma^2)$, равно коррелированные с коэффициентом парной корреляции ρ. Это приводит к интегралу

$$\pi_{ij} = H(V_i - V_j) = \int\limits_{-(V'_i - V'_j)}^{\infty} z(x)\,dx, \tag{1.3.9}$$

где

$$V'_i = \frac{V_i}{[2\sigma^2(1-\rho)]^{\frac{1}{2}}} \quad \text{и} \quad z(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}.$$

Другой важный частный случай задается моделью Брэдли—Терри* [13], для которой

$$H(V_i - V_j) = \frac{1}{4} \int\limits_{-(V_i - V_j)}^{\infty} \operatorname{sech}^2 \frac{1}{2} y\,dy, \tag{1.3.10}$$

где в обозначениях Брэдли $V_i = \log \pi_i$ ($\pi_i \geqslant 0$, $\Sigma\pi_i = 1$); отсюда следует, что $\pi_{ij} = \pi_i/(\pi_i + \pi_j)$.

Бранк [18] высказал интересную точку зрения, что ценности («стоимости» в его терминологии) объектов играют роль некоторых аналогов «главных эффектов» в дисперсионном анализе. Предположение, что они определяют вероятности предпочтений, вполне аналогично гипотезе об отсутствии взаимодействий в дисперсионном анализе. Можно сказать, что величины V — это действительные ценности объектов в противоположность ситуации, примером которой служит (1.3.5), где ценности объектов детерминируются вероятностями предпочтений.

1.4. ОБОБЩЕНИЯ ДЛЯ ЗАДАЧИ МНОЖЕСТВЕННОГО ВЫБОРА

Предположим, что не два, а более объектов надо сопоставить в одном сравнении. Эксперта могут просить либо выразить только предпочтение для одного из объектов, либо построить полное ранжирование. Не являясь в строгом смысле моделью парных сравнений, такая ситуация множественного выбора допускает интересное обобщение. Упомянем здесь недавние попытки построить вероятностные модели в частных случаях. Например, читатель может обратиться к работе Льюса [98], который довольно подробно изучил обобщение модели Брэдли—Терри, и к [9], где весьма обстоятельно описаны логические соотношения для различных моделей парных и множественных сравнений. Льюс постулирует, что если M — *любое* подмножество

* См. [199], [200]. — *Прим. пер.*

t объектов, которое содержит A_i, то фактически вероятность предпочтения A_i будет равна:

$$\pi_{i\,|\,M} = \frac{\pi_i}{\sum_M \pi_j}\,. \tag{1.4.1}$$

Ясно, что от этой модели, намного более широкой, чем линейная модель Брэдли—Терри, можно ожидать пригодности к описанию лишь довольно частных ситуаций выбора. В действительности (1.4.1) означает, что если A_j всегда предпочитается A_i в парном сравнении двух объектов (так что $\pi_i = 0$) и если, хотя и редко, бывает, что A_i предпочитается A_k (так что $\pi_k = 0$), то A_k никогда не выбирается из этих трех объектов. Льюс считает это следствие линейности мешающим и предлагает устранить его ограничением (1.4.1) на ситуацию, в которой $\pi_{ij} \neq 0$. А если $\pi_{ij} = 0$, он не определяет $\pi_{i/M}$ и утверждает, что

$$\pi_{j\,|\,M} = \frac{\pi_j}{\sum_{M'} \pi_i}\,,$$

где M' есть множество M без объекта A_i. Это кажется весьма искусственной попыткой расширения области действия модели, так как модель (1.4.1) все еще применима, если $\pi_{ij} = \delta > 0$ для любого, как угодно малого, δ.

Модель (1.4.1) (не обобщенная) привлекательна, если имеется ограничивающее свойство «независимости несвязанных альтернатив», заключающееся в том, что если подмножество M содержит A_i и A_j, то*

$$\pi_{i\,|\,M}/\pi_{j\,|\,M} = \pi_i/\pi_j,$$

какие бы другие объекты ни входили в M.

Тройные сравнения

Проблема сравнения объектов в наборах по три привлекала и привлекает особое внимание. Пендергрhistóricoасс и Брэдли [115] изучали и пропагандировали модель Д. Р. Кокса, в соответствии с которой объекты A_i, A_j, A_k ранжируются в этом порядке с вероятностью

$$\frac{\pi_i^2 \pi_j}{\pi_i^2(\pi_j+\pi_k)+\pi_j^2(\pi_k+\pi_i)+\pi_k^2(\pi_i+\pi_j)}\,. \tag{1.4.2}$$

Другая модель, которую они кратко рассматривали, дает следующую вероятность:

$$\frac{\pi_i \pi_j}{(\pi_i+\pi_j+\pi_k)(\pi_j+\pi_k)}\,, \tag{1.4.3}$$

приводящую к «очень похожим результатам в приложениях». Для тройных (и множественных) сравнений кажется даже более важным,

---

* См. аксиому Льюса в книге [151]. — *Прим. пер.*

чем для парных сравнений*, дать эксперту ясные инструкции о том, как он должен выполнять ранжирование, если хочет получить состоятельные результаты. Если он сначала выбирает тот объект, который он считает лучшим из трех, а затем переходит к выбору лучшего из оставшихся двух, модель (1.4.3), вероятно, будет более пригодной. Модель (1.4.2) тесно связана с симметрической ситуацией, когда суждение явно или мысленно производится с помощью трех парных сравнений, которые должны быть состоятельны. Для достижения состоятельности может требоваться более чем три таких попарных суждения, и эта процедура, очевидно, является менее прямой, чем первая[1]. Ясно, что если модель Льюса выполняется, то (1.4.2) и (1.4.3) точны при сформулированных условиях, поэтому все сравнения независимы. Сошлемся также на [120].

1.5. ОБ ИСТОРИИ МЕТОДА ПАРНЫХ СРАВНЕНИЙ

В своем введении к книге Торгерсона «Theory and Methods of Scaling» [137] Галликсен говорит о «методе парных сравнений Фехнера». Научные идеи, высказываемые Фехнером, редко принимались безоговорочно, однако не может быть никакого сомнения в том, что истоки метода парных сравнений следует искать в до странного мало известном разделе знаменитой работы Фехнера «Elemente der Psychophysik» [41][2]. За несколько лет до появления книги Фехнер проводил обширные эксперименты с целью проверки закона Вебера, в особенности результатов, приведенных в его работе «Methode der richtigen und falschen Fälle» («Метод верных и неверных случаев»). Основой эксперимента было проводившееся многократно и в разных условиях определение того, какой из двух сосудов тяжелее, причем в действительности известные различия D в массе были малы (0,04 или 0,08 веса более легкого сосуда, масса которого изменялась от 300 до 3000 г). Поскольку истинный результат был известен, что типично для психофизики в противоположность психометрии, число верных и неверных решений можно было затабулировать. Первоначально Фехнер поручал подготовку сосудов ассистенту, так что он сам не знал, какой из сосудов тяжелее. Затем, в силу случайности, Фехнер отказался от этого метода и, что довольно странно, проводил сам все эксперименты, результаты которых, по его мнению, были достойны опубликования.

Аргументация Фехнера в пользу такой системы заключалась в том, что это позволяло ему полностью сконцентрировать внимание на выполнении стоящей перед ним задачи и получить наиболее удовлетворительные результаты. Техника проведения эксперимента, используемая Фехнером, представляет известный интерес, хотя некоторые детали мы вынуждены опустить. Начнем с того, что более легкий из двух сосудов стоял слева. Фехнер брал этот сосуд в течение 1 с (левой ру-

* То есть без циклических триад. — *Прим. пер.*

[1] Мэллоуз [103] положил эту идею в основу ненулевой модели ранжирования t объектов.

[2] Материалы симпозиума, посвященного столетию выхода в свет этой работы, были опубликованы в мартовском номере журнала «*Psychometrika*» за 1961 г.

где $1/h = \sigma\sqrt{2(1-\rho)}$ (в общем случае). Фехнер определял h как $(u_1 + u_2 + u_3 + u_4)/4D$, а S получал из

$$hS = \frac{1}{4}(u_1 - u_2 + u_3 - u_4).$$

Таким образом, S оказалось равно разнице в весе

$$\frac{u_1 - u_2 + u_3 - u_4}{u_1 + u_2 + u_3 + u_4} D$$

независимо от значения ρ.

Замечания

В последние годы был проявлен значительный интерес к математическим аспектам моделей предпочтения и поведения в условиях выбора. Особенно следует отметить работы [9], где содержатся ссылки на последующую литературу, а также [117].

Вместо вероятностей предпочтения можно работать с ожидаемыми значениями. Эти два подхода эквивалентны, когда каждое парное сравнение дает значения 1 или 0, но второй подход охватывает также случаи совпадающих рангов и другие методы определения значений (см. [18]).

УПРАЖНЕНИЯ

1.1. Докажите, что для модели стохастической транзитивности $1 \leqslant \pi_{ij} + \pi_{jk} + \pi_{ki} \leqslant 2$.

1.2. Покажите, что если

$$\pi_{ij} \geqslant \frac{1}{2}, \quad \pi_{jk} > \frac{1}{2}, \text{ но } \pi_{ik} < \frac{1}{2},$$

то вероятность наблюдения циклической триады (цикла) при сравнении A_i, A_j, A_k превышает 1/4.

[Указание: вероятность цикла выражается следующим образом (см. [9]):

$$\frac{1}{4} + q_i q_j + q_j q_k + q_k q_i,$$

где

$$q_i = \pi_{ij} - \frac{1}{2}, \quad q_j = \pi_{jk} - \frac{1}{2}, \quad q_k = \pi_{ki} - \frac{1}{2}.]$$

1.3. Докажите, что для модели стохастической транзитивности вероятность цикла равна самое большое 1/2 [9].

1.4. Покажите, что из условия (1.3.7) следует ранжирование по (1.3.5), однако условия (1.3.6) для этого недостаточно.

1.5. Пусть y_i ($i = 1, 2, \ldots, t$) — независимые нормально распределенные $N(V_i, \sigma_i^2)$ случайные величины. Покажите, что тогда вероятности предпочтения $\pi_{ij} = \Pr(y_i > y_j)$ не удовлетворяют линейной модели, кроме случая $\sigma_i^2 = \sigma^2$ (для всех i); действительно, не обязательно выполняется условие усиленной стохастической транзитивности. [Эта ситуация относится к случаям I—IV Тэрстоуна [133], причем допущение независимости не является решающим.]

1.6. Зададим для независимых случайных величин y_1, y_2, y_3 вероятности:

$$\Pr\left(y_1 = \frac{-5}{3}\right) = \frac{3}{8}, \quad \Pr(y_1 = 1) = \frac{5}{8},$$

$$\Pr(y_2 = -1) = \frac{5}{8}, \quad \Pr\left(y_2 = \frac{5}{3}\right) = \frac{3}{8},$$

$$\Pr\left(y_3 = \frac{-2}{3}\sqrt{30}\right) = \frac{1}{16}, \quad \Pr(y_3 = 0) = \frac{7}{8}, \quad \Pr\left(y_3 = \frac{2}{3}\sqrt{30}\right) = \frac{1}{16}.$$

Проверьте, что*

$$\mathscr{E}(y_1) = \mathscr{E}(y_2) = \mathscr{E}(y_3) = 0,$$

$$\operatorname{var} y_1 = \operatorname{var} y_2 = \operatorname{var} y_3 = \frac{5}{3},$$

и докажите, что

$$\pi_{21} = \pi_{32} = \pi_{13} = \frac{39}{64}.$$

[Штейнгауз и Трибула [131] установили, что $t = (\sqrt{5} - 1)/2$ — наибольшее значение t, для которого возможно равенство $\pi_{21} = \pi_{32} = \pi_{13} = t$. Этот результат показывает, между прочим, что характеристики, связанные с тремя объектами, не всегда можно представить независимыми случайными величинами. Сравните это с [140].]

1.7. Покажите, что для модели Брэдли—Терри вероятности циклов $A_i \to A_j \to A_k \to A_i$ и $A_i \leftarrow A_j \leftarrow A_k \leftarrow A_i$ равны. Для этого надо показать, что π_{ik} можно получить из π_{ij} и π_{jk} с помощью

$$\pi_{ik} = \frac{\pi_{ij}\pi_{jk}}{\pi_{ij}\pi_{jk} + \pi_{kj}\pi_{ji}}.$$

1.8. Покажите, что если $\pi_i > \frac{1}{2}(\pi_j + \pi_k)$, то вероятность выбора A_i, как лучшего из A_i, A_j, A_k, для модели (1.4.2) больше, чем для модели (1.4.3).

1.9. Положим, что три сравнения A_i, A_j, A_k независимы с вероятностями предпочтения, удовлетворяющими модели Брэдли—Терри. Если исходом эксперимента является цикл, ранжирование** выполнено случайно. Покажите, что A_i предпочитается с вероятностью

$$\frac{\pi_i^2(\pi_j + \pi_k) + \frac{2}{3}\pi_i\pi_j\pi_k}{(\pi_i + \pi_j)(\pi_i + \pi_k)(\pi_j + \pi_k)},$$

а A_i, A_j, A_k будут проранжированы в этом порядке с вероятностью

$$\frac{\pi_i^2\pi_j + \frac{1}{6}\pi_i\pi_j\pi_k}{(\pi_i + \pi_j)(\pi_i + \pi_k)(\pi_j + \pi_k)}.$$

---

* В отечественной литературе обычно $\mathscr{E}(\ )$ обозначают $M(\ )$, а $\operatorname{var} y_i$ — $\sigma^2\{y_i\}$ или $\sigma^2_{y_i}$, или $D^2\{y_i\}$, в книге $E(\cdot) = \mathscr{E}(\cdot)$. — *Прим. ред.*

** Ранжирование пар (A_i, A_j), (A_i, A_k), (A_j, A_k). — *Прим. пер.*

2 ГЛАВА КОМБИНАТОРНЫЕ МЕТОДЫ

2.1. ВВЕДЕНИЕ

Предположим, что эксперт выполнил все $\binom{t}{2}$ парных сравнений. Совпадения не допускаются, так что результаты можно представить в виде таблицы предпочтений с двумя входами, составленной из единиц и нулей и известной как турнирная таблица. Для $t = 5$ обычный исход эксперимента может быть следующим:

	A_1	A_2	A_3	A_4	A_5	Очки
A_1	—	0	1	1	1	3
A_2	1	—	1	0	0	2
A_3	0	0	—	0	0	0
A_4	0	1	1	—	1	3
A_5	0	1	1	0	—	2

Главная диагональ слева направо свободна и входы (места) ниже, строго говоря, избыточны. Видно, что A_1 предпочтительнее, чем A_3, A_4, A_5, но не чем A_2, и набрал *3 очка*[1]. В общем, число предпочтений, отданных A_i, обозначим a_i. Понятно, что

$$\sum_{i=1}^{t} a_i = \frac{1}{2} t (t-1) \tag{2.1.1}$$

и возможно всего $2^{\frac{1}{2}t(t-1)}$ различных таблиц предпочтений.

Альтернативное геометрическое представление, принадлежащее, как и большая часть этого параграфа, Кендэлу и Бэбингтону Смиту [93],

[1] Это использование термина «очки» (score) не должно приводить к путанице с его применением в гл. 1.

иногда более наглядно*. Для приведенного выше примера, который иллюстрирует общий случай, картинка имеет форму правильного пятиугольника, со всеми его связями (рис. 2.1). Направление стрелок указывает 10 предпочтений, так что a_i — число стрелок**, выходящих из вершины A_i.

Быть может, первые вопросы, на которые дает ответ таблица предпочтений или многоугольник предпочтений, это: 1) Последователен ли эксперт в своих суждениях? 2) Есть ли значимые различия между объектами?

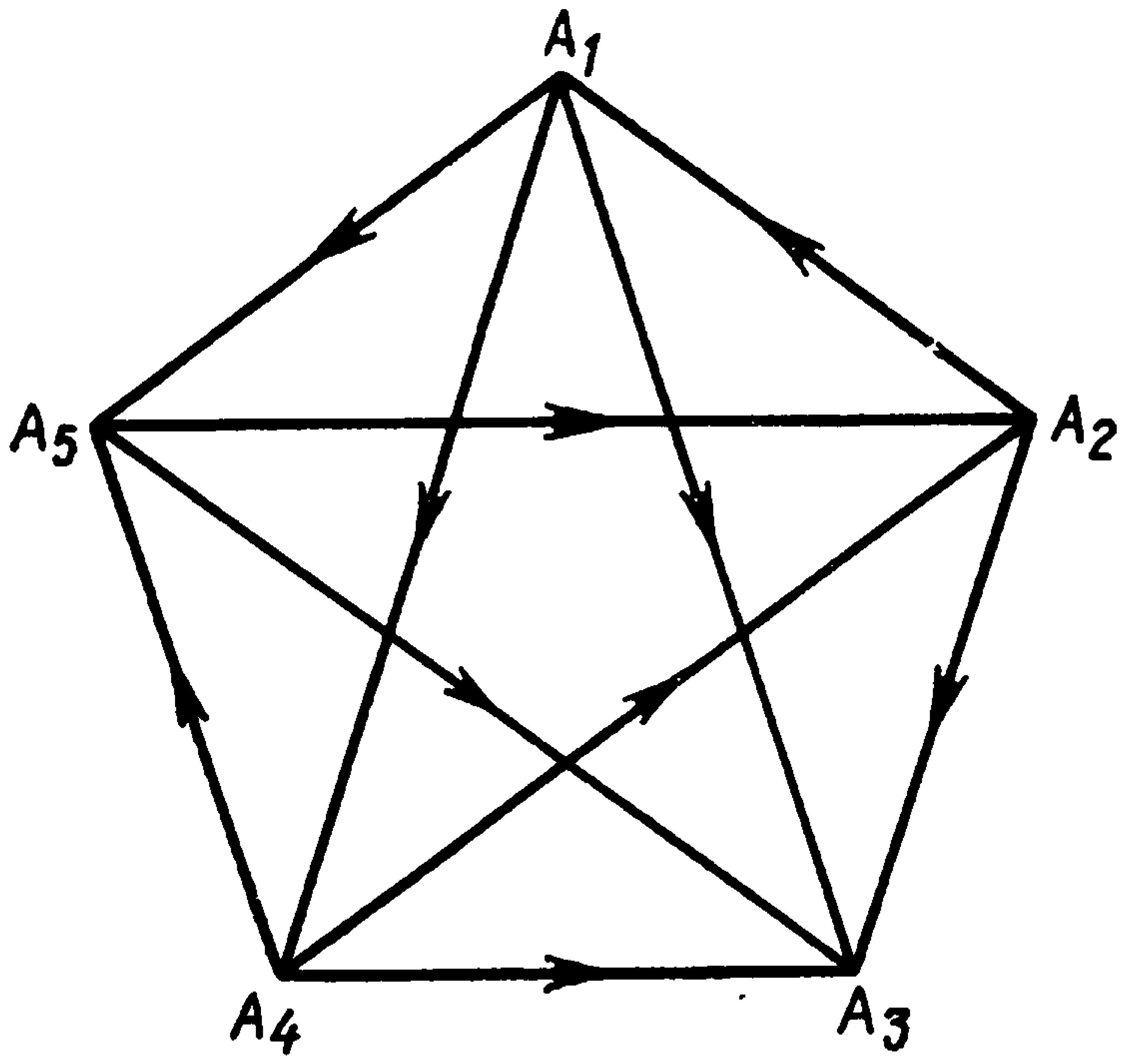


Рис. 2.1. Многоугольник предпочтений

Важно понимать, что эти вопросы тесно связаны; если нет различия между объектами, неразумно ожидать от эксперта состоятельности, и, наоборот, ему легко быть состоятельным, если различия велики. Мы будем обращаться к нулевой гипотезе, согласно которой суждения выносятся случайно ($\pi_{ij} = \frac{1}{2}$ для всех i, j; $i \neq j$). Ее разумно называть «гипотезой случайности».

Для группы из трех объектов результат непоследовательного (несостоятельного) эксперта есть циклическая триада; для большой группы множество суждений можно считать тем более последовательным, чем меньше циклических триад оно содержит. Видно, что из 10 триад нашего примера A_1, A_4, A_2 и A_1, A_5, A_2 — циклические триады. Ясно, что прямое перечисление неудобно, если t не очень мало. Однако число циклов c связано с числом очков соотношением:

$$c = \frac{t}{24}(t^2-1) - \frac{1}{2}T, \tag{2.1.2}$$

где $T = \Sigma\left(a_i - \bar{a}\right)^2$ и $\bar{a} = \Sigma a_i/t = \frac{1}{2}(t-1)$.

Для доказательства этого мы покажем сначала, что если изменить направление некоторого предпочтения (на противоположное) и число циклов возрастет на p, то T уменьшится на $2p$, и наоборот. Рассмотрим множество предпочтений, включающих $A_1 \to A_2$. Если X — объект, отличный от A_1 и A_2, должно быть $a_1 - 1$ предпочтений типа $A_1 \to X$ и a_2 — типа $A_2 \to X$. Только в триадах, содержащих A_1 и A_2,

* В теории графов такое представление называется турниром (полным ориентированным графом), см. [211] и [287]. — *Прим. пер.*

** Число a_i называется полустепенью исхода вершины. — *Прим. пер.*

произошла смена $A_1 \to A_2$ на $A_2 \to A_1$. Имеется четыре типа таких триад:

$A_1 \to X \leftarrow A_2$, скажем x штук,
$A_1 \leftarrow X \to A_2$,
$A_1 \to X \to A_2$, которых должно быть $a_1 - 1 - x$,
$A_1 \leftarrow X \leftarrow A_2$, которых должно быть $a_2 - x$.

Когда предпочтение $A_1 \to A_2$ меняет направление, первые две триады остаются нециклическими. Триада третьего типа становится циклической, триада четвертого типа перестает быть таковой. Прирост числа циклических триад

$$(a_1 - 1 - x) - (a_2 - x) = a_1 - a_2 - 1 = p.$$

Так как

$$T = \Sigma a_i^2 - \frac{1}{4} t (t-1)^2,$$

соответствующее *уменьшение* T есть

$$a_1^2 - (a_1 - 1)^2 + a_2^2 - (a_2 + 1)^2 = 2(a_1 - a_2 - 1) = 2p.$$

Более общо: если в результате инверсии некоторого числа предпочтений c увеличивается на p, то T должно уменьшиться на $2p$; при инверсии предпочтений влияние на c и T накапливается одновременно. Поэтому ясно, что

$$c = k - \frac{1}{2} T,$$

где k — константа, зависящая только от t.

Для определения k заметим, что при $c = 0$ возникает полная согласованность. В этом случае $(a_1, a_2, \ldots, a_t)$ — просто перестановка $(0, 1, 2, \ldots, t-1)$, что дает

$$\Sigma a_i^2 = \frac{1}{6} (t-1) t (2t-1),$$

откуда

$$k = \frac{1}{24} t (t^2 - 1).$$

Это доказывает (2.1.2).

Кендэл и Бэбингтон Смит определили *коэффициент совместимости* как*

$$\zeta = \begin{cases} 1 - \dfrac{24c}{t(t^2-1)}, & \text{при нечетном } t; \\ 1 - \dfrac{24c}{t(t^2-4)}, & \text{при четном } t. \end{cases}$$

* Число циклических триад удобнее вычислять по формуле $c = \left(\frac{1}{12}\right) t (t-1)(2t-1) - \sum\limits_{i}^{t} a_i^2/2$, чем по (2.1.2) (см., например, [90]). — *Прим. пер.*

Если и только если $\zeta=1$, не существует несовместимостей в конфигурации предпочтений, которая, следовательно, немедленно порождает ранжировку *. По мере того как ζ уменьшается до нуля, несовместимость, измеряемая числом циклических триад, возрастает (см. упр. 2.1).

Поскольку T есть линейная функция от ζ, обе они, по существу, эквивалентны. Хотя T не дает непосредственной меры совместимости в суждениях экспертов, представляется, что можно легко обобщить результаты на ситуации, где каждое парное сравнение повторяется n раз. Таким образом, если a_i теперь обозначает число очков i-го объекта для эксперимента с повторениями, мы можем просто определить T_n как общую сумму квадратов очков[1]:

$$T_n = \Sigma\,(a_i - \bar{a})^2, \tag{2.1.3}$$

так что $T = T_1$. В основном по этой причине мы будем использовать далее T-статистику. Большие значения приведут нас к отклонению нулевой гипотезы о случайности H_0. Для проверки значимости мы покажем в § 2.3, что для удобства стандартизированное T_n распределено при H_0 приближенно как χ^2 с $t-1$ степенями свободы.

Циклические триады представляют собой весьма фундаментальный тип непоследовательности во мнениях экспертов. Нет необходимости рассматривать отдельно циклы, включающие больше чем три объекта, поскольку их можно разбить на циклические триады**. Однако если

---

* Турнир без циклов называется транзитивным. Транзитивностью называется следующее свойство: из $A_i \to A_j$, $A_j \to A_k$ следует $A_i \to A_k$ для всех $i \neq j$, $i \neq k$, $j \neq k$.

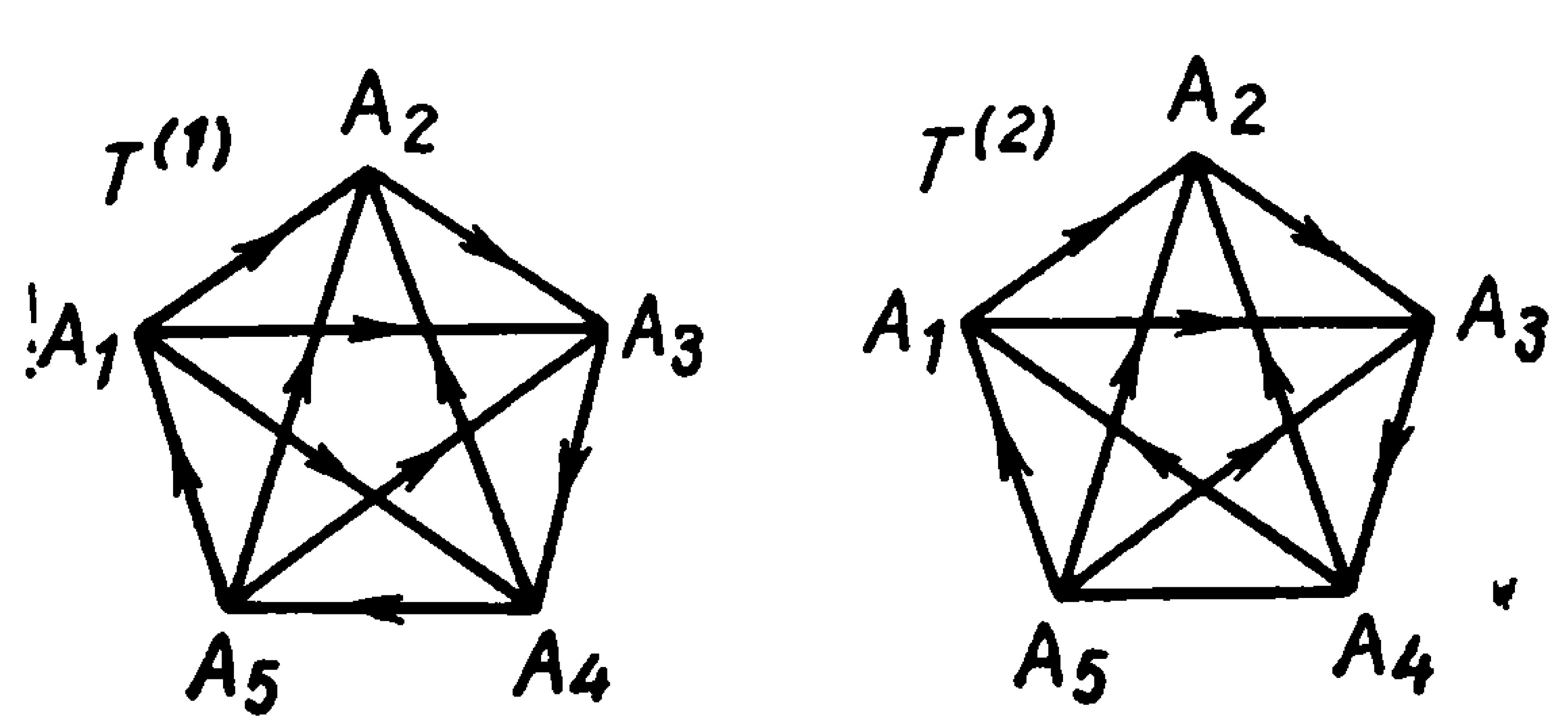
Для матрицы

$$\alpha_{i,\,j} = \begin{cases} 1 & A_i \to A_j \\ 0 & A_i \leftarrow A_j \end{cases}$$

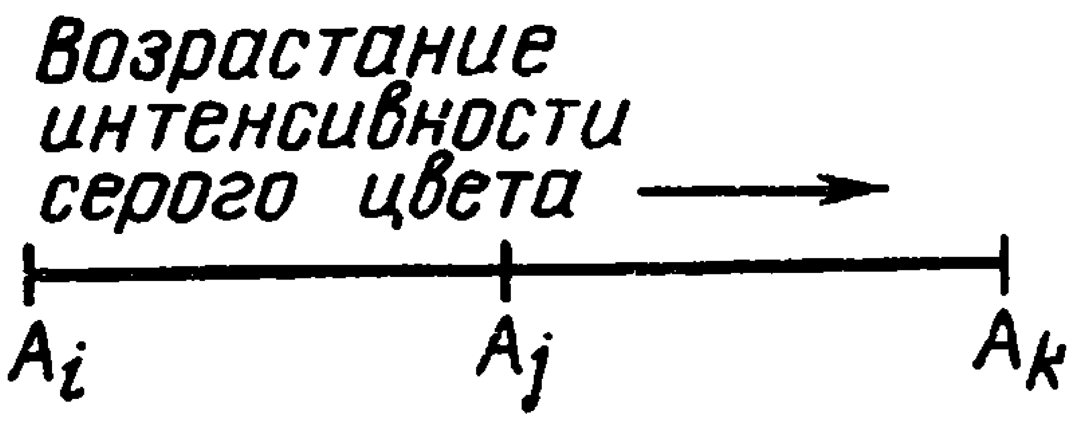
оно формулируется так: из $\alpha_{ij} = 1$, $\alpha_{jk} = 1$ следует $\alpha_{ik} = 1$, для $1 \leqslant i$, j, $k \leqslant t$, $i \neq j$, $i \neq k$, $j \neq k$. — *Прим. пер.*

[1] Это фактически частный случай статистики, определенной Дарбином [34] для использования в экспериментах по ранжированию.

** Это высказывание не вполне обоснованно. Можно привести пример двух турниров, в которых число циклических троек одинаково, а число циклических четверок различно. Вот он:



Выпишем векторы очков (полустепеней исхода) $a^{(1)} = (3, 1, 1, 2, 3)$, $a^{(2)} = (2, 1, 1, 3, 3)$, отсюда $\zeta^{(1)} = \zeta^{(2)}$. В турнире $T^{(1)}$ две циклические четверки, а в $T^{(2)}$ — три. Таким образом, количество циклических троек не полностью хасуществует предварительное (априорное) упорядочение объектов по отношению к заданному параметру (например, интенсивности серого цвета в примере Кумбса, § 1.3), может возникнуть другой тип несовместимости во мнениях экспертов. В ситуации, иллюстрируемой ниже, результаты $A_i \to A_j$ и $A_k \to A_j$ несовместимы (хотя они не допускают возникновения циклической триады); для $A_i \to A_j$ предполагается, что идеал I находится слева от A_j, а для $A_k \to A_j$ предполагается, что I находится справа от A_j. Джерард и Шапиро



[51] назвали этот тип несовместимости в суждениях *сепарацией*. Они показали, что число сепараций S в попарном сравнении t объектов может быть получено из таблицы предпочтений с помощью формулы

$$S = \sum_{i=1}^{t} C_i (t - i - R_i), \tag{2.1.4}$$

где C_i, R_i — суммы по i-му столбцу и i-й строке элементов выше главной диагонали*.

Таким образом, в примере, помещенном в начале этого раздела, $S=4$. Доказательство (2.1.4) следует немедленно с учетом того, что C_i и $t-i-R_i$ — это числа случаев предпочтения объектов, лежащих соответственно слева и справа от A_i, объекту A_i.

рактеризует непоследовательность во мнениях эксперта. Учитывая это, можно ввести коэффициент

$$\zeta_{(4)} = 1 - c_{(4)}(T_t) / \max_{\{T_t\}} c_{(4)},$$

где $c_{(4)}(T_t)$ — число циклических четверок в турнире T_t, максимум берется по всем турнирам с t вершинами:

$$\max_{\{T_t\}} c_{(4)} = \begin{cases} \frac{1}{48} t(t^2-1)(t-3) & \text{для нечетных } t, \\ \frac{1}{48} t(t^2-4)(t-3) & \text{для четных } t. \end{cases}$$

Однако распространение этого подхода на циклы с k вершинами, $k > 4$, связано с серьезными затруднениями, так как задача поиска максимального числа k-циклов, $k > 4$, еще не решена. По этим вопросам см. [152]. Кроме того, см. работы [219], [210], [218]. — *Прим. пер.*

* $C_j = \sum_{t \leq i < j} \alpha_{ij}$, где суммирование ведется по i; а $R_i = \sum_{i < j \leq t} \alpha_{ij}$, где сумма берется по j, $\alpha_{ij} = \begin{cases} 1 & A_i \to A_j, \\ 0 & A_i \leftarrow A_j. \end{cases}$ — *Прим. пер.*

До сих пор мы рассматривали случай, когда работал один эксперт. Если же работают n экспертов и каждый из них выполняет $\binom{t}{2}$ сравнений, результаты также можно представить в виде таблицы (матрицы) предпочтений с элементами α_{ij}, равными числу случаев, когда A_i предпочтительнее, чем A_j. Тогда $\alpha_{ji} = n - \alpha_{ij}$. Если мнения экспертов полностью совпадают, одна половина α_{ij} равна n, а другая—нулю. Заметим, что согласие экспертов может быть полным, даже если в их суждениях есть несовместимости; наоборот, отсутствие согласия не влечет за собой несовместимости.

Пусть

$$\Sigma = \sum_{i \neq j} \binom{\alpha_{ij}}{2},$$

где суммирование ведется по $t\,(t-1)$ членам. Σ — это сумма числа случаев совпадения мнений между парами экспертов. Кендэл и Бэбингтон Смит определили коэффициент согласия u как

$$u = \frac{2\Sigma}{\binom{n}{2}\binom{t}{2}} - 1. \tag{2.1.5}$$

Если существует полная согласованность экспертов, и только в этом случае, $u = 1$. Чем дальше мы отклоняемся от этого случая, что измеряется числом совпадений между парами экспертов, тем меньше становится u*. Минимальное число совпадений будет тогда, когда каждое α_{ij} равно $\frac{1}{2}n$, при n — четном, или $\frac{1}{2}(n \pm 1)$, при n — нечетном. Соответственно n будет равно $-1/(n-1)$ или $-1/n$.

По-другому Σ можно выразить так:

$$\Sigma = \frac{1}{2}\left[\sum_{i \neq j} \alpha_{ij}^2 - n\binom{t}{2}\right] = \sum_{i<j} \alpha_{ij}^2 - n\sum_{i<j} \alpha_{ij} + \binom{n}{2}\binom{t}{2}.$$

---

* Для случая, когда определяется согласие мнений n экспертов, каждый из которых ранжирует t объектов, можно также применять коэффициент u. В этом случае он равен:

$$u = \sum_{1 \leqslant l < q \leqslant n} \tau_{lq}, \tag{I}$$

где τ_{lq} — коэффициент ранговой корреляции Кендэла (см. [90]) l-го и q-го экспертов. Отсюда и из работ [212] и [178] ясно, что (I) и

$$\sum_{1 \leqslant l < q \leqslant n} d_{lq}, \tag{II}$$

где d_{lq} — расстояние между ранжировками l-го и q-го эксперта (см. [175]), эквивалентные статистики. Об этом, а также о распределении u для «ранговой» ситуации см. [39] и [205], где дано обобщение (I) на случай совпадающих рангов. А. Орлов [195] независимо получил распределение статистики (II) другими методами. — *Прим. пер.*

причем последняя форма записи удобней для вычислений, если объекты упорядочены по возрастанию числа очков, так что α_{ij} будут в большинстве своем относительно невелики.

Проверка значимости u возможна при нуль-гипотезе о том, что предпочтения всех экспертов случайны. Это можно сделать [90] с помощью таблиц распределения Σ для комбинаций (t, n) при малых t и n, а когда они велики, можно полагать, что

$$\frac{4}{n-2}\left[\Sigma-\frac{1}{2}\binom{t}{2}\binom{n}{2}\frac{n-3}{n-2}\right]$$

распределено приближенно как χ^2 с числом степеней свободы, равным

$$\binom{t}{2}\frac{n(n-1)}{(n-2)^2}.$$

2.2. ОСНОВЫ ТЕОРИИ РАСПРЕДЕЛЕНИЙ

Как мы видели, результаты эксперимента парных сравнений можно суммировать в виде вектора очков $(a_1, a_2, \ldots, a_t)$, где $\Sigma a_i = \frac{1}{2}t(t-1)$. Для многих целей нумерации объектов последовательность, в которой расположены очки, неважна, и $(a_1, a_2, \ldots, a_t)$ можно заменить размещением $[x_1^{r_1}x_2^{r_2}\ldots x_m^{r_m}]$, где $x_1, \ldots, x_m$ — просто величины $a_1, \ldots, a_t$, переставленные в убывающем порядке, а r_1 — число случаев, когда наибольшее значение есть x_1 и т. д. Отсюда следует, что

$$\sum_{u=1}^{m} r_u = t, \quad \sum_{u=1}^{m} r_u x_u = \frac{1}{2}t(t-1).$$

Распределение вероятностей допустимых размещений служит важной основой, так как из него можно получить распределения всех статистик критериев, таких, как T, которые являются функциями очков. Когда гипотеза случайности верна, частоты различных размещений дают ту же информацию, что и распределение, так как их можно превратить в вероятности делением на $2^{\frac{1}{2}t(t-1)}$.

Один из возможных подходов к задаче перечисления — рассмотрение производящей функции

$$G(t) = \prod_{i<j}(b_i + b_j), \tag{2.2.1}$$

произведение $\frac{1}{2}t(t-1)$ членов которой соответствует сделанным сравнениям. Разложение $G(t)$ содержит $2^{\frac{1}{2}t(t-1)}$ членов, показатели степеней которых дают возможные наборы очков, например

$$G(3) = (b_1+b_2)(b_1+b_3)(b_2+b_3) = \Sigma b_1^2 b_2 + 2b_1b_2b_3, \tag{2.2.2}$$

где складываются 6 членов. Таким образом, есть 6 исходов типа* [2 1 0] и 2 исхода типа [1^3]. $G(t)$ симметрична относительно b_i, b_j, $i \neq j$, $1 \leqslant i$, $j \leqslant t$ и разлагается как (2.2.2) в сумму одночленных симметричных функций. В табл. 2.1 приведены наборы очков и их частоты для $t \leqslant 6$. Таблица была построена и ее можно продолжить (см. [28]), замечая, что

$$\text{(a)} \quad G(t+1) = G(t) \prod_{i=1}^{t} (b_{t+1} + b_i) =$$

$$= G(t)\left[b_{t+1}^{t} + (1)\, b_{t+1}^{t-1} + (1^2)\, b_{t+1}^{t-2} + \ldots + (1^t)\right],$$

где $(1^s) = \Sigma b_1 b_2 \ldots b_s$ $(s = 1, 2, \ldots t)$;

(b) наборы

$$\left[x_1^{r_1} x_2^{r_2} \ldots x_m^{r_m}\right] \text{ и } \left[(t-1-x_m)^{r_m} \ldots (t-1-x_2)^{r_2} (t-1-x_1)^{r_1}\right]$$

имеют одинаковые частоты, так как одно из другого можно получить заменой побед на поражения.

Таблица 2.1

Наборы очков и их частоты в эксперименте простого парного сравнения t объектов

(из [28], с разрешения издателя)

$t=3$		$t=6$	
[210]	6	[543210]	720
1^3	2	5431^3	240
Итого	8	542^30	240
		542^21^2	720
$t=4$		53^310	240
[3210]	24	53^22^20	720
31^3	8	53^221^2	1440
2^30	8	532^31	1680
2^21^2	24	52^5	144
Итого	64	4^3210	240
		4^31^3	80
$t=5$		4^23^210	720
[43210]	120	4^232^20	1440
431^3	40	4^2321^2	2880
42^30	40	4^22^31	1680
42^21^2	120	43^220	1680
3^310	40	43^31^2	1680
3^22^20	120	43^22^21	8640
3^221^2	240	432^4	2400
32^31	280	3^32^3	2640
2^5	24	3^421	2400
Итого	1024	3^50	144
		Итого	32768

* Именно $G(3) = \sum_{i<j} (b_i^2 b_j + b_i b_j^2) + 2b_1b_2b_3$ и исходы суть: [2 1 0], [2 0 1], [0 2 1], [1 2 0], [1 0 2], [0 1 2] и 2 раза [1 1 1]. — *Прим. пер.*

Возможно, заслуживают внимания три обобщения (2.2.1), хотя и без них легко получить решение задач перечисления.

1. Если для всех $i \neq j$ произведено не одно, а n_{ij} сравнений A_i и A_j, соответствующая производящая функция

$$\prod_{i<j} (b_i + b_j)^{n_{ij}}. \tag{2.2.3}$$

2. Если $A_i \to A_j$ с вероятностью π_{ij} для всех $i \neq j$, то $G(t)$ можно записать в виде

$$\prod_{i<j} (b_i \pi_{ij} + b_j \pi_{ji}), \tag{2.2.4}$$

который порождает вероятности различных наборов.

3. Предположим, что вместо двухточечной шкалы (1 и 0) оценивание ведется по $(r+1)$-точечной шкале $\{1, (r-1)/r, \ldots, 1/r, 0\}$ с общей суммой очков, в любом отдельном сравнении все же равной единице. Тогда

$$\prod_{i<j} \left(b_i + b_i^{(r-1)/r} b_j^{1/r} + b_i^{(r-2)/r} b_j^{2/r} + \ldots + b_j\right)$$

порождает все возможные исходы. Это включает случай, когда разрешены ничьи так же, как и прямые предпочтения ($r = 2$). Первое обобщение представляет особый интерес для $n_{ij} = n$ при всех $i \neq j$, которому соответствует сбалансированный эксперимент парных сравнений с n повторениями. Производящая функция для этого случая есть просто $[G(t)]^n$, и таблицы наборов (размещений) и их частот для параметров (t, n) можно построить из наборов и частот для $(t, 1)$ в табл. 2.1. Травинский [138] табулировал все случаи до (3; 20), (4; 7) и (5; 3) в дополнение к $n = 1$, $t \leqslant 8$. Очень похожие таблицы были построены ранее Брэдли и Терри [17] и Брэдли [15].

2.3. РАСПРЕДЕЛЕНИЕ ОЧКОВ И ФУНКЦИЙ ОТ НИХ

В этом разделе мы получим распределение различных функций от очков в сбалансированном эксперименте с t объектами и n повторениями (репликами), если верна нуль-гипотеза. Результаты будут использованы в гл. 3 для построения ряда критериев значимости при проверке эквивалентности t объектов.

Сами очки

Так как при гипотезе H_0 о случайности объект A_i с вероятностью $\frac{1}{2}$ превосходит в каждом из $n(t-1)$ сравнений любой из $t-1$ объектов, очки a_i распределены биномиально $b\left(\frac{1}{2}, n(t-1)\right)$.

Совместное распределение очков

Рассмотрим сначала случай $n = 1$. Для нахождения совместного распределения двух значений, которые без потери общности можно положить равными a_1 и a_2, заметим, что можно различать два случая, соответствующих $A_1 \to A_2$ и $A_1 \leftarrow A_2$. Если $A_1 \to A_2$, то A_1, A_2 должны предпочитаться соответственно $a_1 - 1$ и a_2 объектам из оставшихся $t - 2$ штук и т. д. Таким образом,

$$f(a_1, a_2) = \frac{1}{2^{2t-3}}\left[\binom{t-2}{a_1-1}\binom{t-2}{a_2} + \binom{t-2}{a_2}\binom{t-2}{a_2-1}\right]. \qquad (2.3.1)$$

Это соображение можно распространить на совместное распределение $a_1, a_2, \ldots, a_s$ $(s < t)$, рассматривая $2^{\frac{1}{2}s(s-1)}$ исходов сравнений только s объектов. Если очки в этом частном эксперименте равны a'_1, $a'_2, \ldots, a'_s$, окончательные результаты (когда в эксперимент включены все t объектов) будут $a_1, a_2, \ldots, a_s$ с вероятностью

$$f(a_1, a_2, \ldots, a_s \mid a'_1, a'_2, \ldots, a'_s) =$$
$$= \frac{1}{2^{s(t-s)}}\binom{t-s}{a_1-a'_1}\binom{t-s}{a_2-a'_2}\cdots\binom{t-s}{a_s-a'_s}, \qquad (2.3.2)$$

с учетом соглашения

$$\binom{n}{r} = 0 \text{ при } r < 0 \text{ или } r > n.$$

Отсюда следует, что

$$f(a_1, a_2, \ldots, a_s) = \frac{1}{2^{\frac{1}{2}s(2t-s-1)}} \Sigma \prod_{r=1}^{s}\binom{t-s}{a_r - a'_r}, \qquad (2.3.3)$$

где суммирование ведется по всем исходам частного эксперимента, совместимым с окончательными данными. Для $n \geqslant 1$ (2.3.3) обобщается до

$$f(a_1, a_2, \ldots, a_s) = \frac{1}{2^{\frac{1}{2}ns(2t-s-1)}} \Sigma \prod_{r=1}^{s}\binom{n(t-s)}{a_r - a'_r}. \qquad (2.3.4)$$

Альтернативный подход дан в § 2.4.

Разность двух значений

Распределение $f(d)$ случайной величины $d = a_1 - a_2$ можно, конечно, получить из совместного распределения a_1 и a_2. Распределение $f(d)$ сходится к нормальному при возрастании n или t. Это следует из того (упражнение 2.4), что характеристическая функция $\varphi(u)$ случайной величины $d/\sqrt{\frac{1}{2}nt}$, равная

$$\varphi(u) = \left[\cos\frac{1}{2}u\sqrt{(2/nt)}\right]^{2n(t-2)}\left[\cos u\sqrt{(2/nt)}\right]^n, \qquad (2.3.5)$$

сходится к $e^{-\frac{1}{2}u^2}$ с ростом n или t.

Найдем теперь точное выражение для

$$P_{tnm} = \Pr(|a_1 - a_2| \geqslant m \mid H_0),$$

где m — положительное целое число. Мы имеем

$P_{tnm} = 2\Pr(a_1 - a_2 \geqslant m \mid H_0) = 2 \sum_{r=m}^{n(t-1)} \sum_{p=0}^{n}$ [Pr (во всех сравнениях между объектами A_1 и A_2 A_1 предпочитается в $(2p-n)$ раз больше, чем A_2) $\times$ Pr (A_1 предпочитается в $(r-2p+n)$ раз больше, чем A_2, в сравнениях с другими $(t-2)$ объектами)] $=$

$$= 2\sum_{r=m}^{n(t-1)} \sum_{p=0}^{n} \binom{n}{p} 2^{-n} \sum_{q=r-2p+n}^{n(t-2)} \binom{n(t-2)}{q} \binom{n(t-2)}{q-r+2p-n} 2^{-2n(t-2)} =$$

$$= 2^{3n-2nt+1} \sum_{r=m}^{n(t-1)} \sum_{p=0}^{n} \binom{n}{p} \sum_{q=r-2p+n}^{n(t-2)} \binom{n(t-2)}{q} \binom{n(t-2)}{q-r+2p-n}. \tag{2.3.6}$$

Равенство коэффициентов x^{r+2p-n} в разложении обеих сторон тождества

$$(1+x)^{n(t-2)}(1+x^{-1})^{n(t-2)} \equiv (1+x)^{2n(t-2)} x^{-n(t-2)}$$

дает

$$\sum_{q=r-2p+n}^{n(t-2)} \binom{n(t-2)}{q} \binom{n(t-2)}{q-r+2p-n} = \binom{2n(t-2)}{n(t-3)-r+2p}. \tag{2.3.7}$$

Подставляя (2.3.7) в (2.3.6), мы получаем

$$P_{tnm} = 2^{3n-2nt+1} \sum_{r=m}^{n(t-1)} \sum_{p=0}^{n} \binom{n}{p} \binom{2n(t-2)}{n(t-3)-r+2p}.$$

Наибольшее значение x_1

Для t и n, не больших, чем в исходных таблицах § 2.2, легко построить распределение вероятностей x_1. Обычно оно требуется лишь для больших значений x_1. Поэтому мы опишем метод, который удобен в нашем случае, и применим также для больших значений t и n.

Пусть E_i — событие $a_i \geqslant M$ $(0 \leqslant M \leqslant n(t-1))$. При H_0 ясно, что*

$$\Pr(E_{i_1} E_{i_2} \ldots E_{i_s}) = \Pr(E_1 E_2 \ldots E_s).$$

---

* $H_0 : \pi_{ij} = \frac{1}{2}$ для всех $i \neq j$ и отсюда

$$\Pr(A_{i_1} A_{i_2} \ldots A_{i_t}) = \Pr(A_1, A_2, \ldots, A_t),$$

где $(A_{i_1}, A_{i_2}, \ldots A_{i_t})$ — ранжировка (упорядочение) объектов, такая, что $R(A_{i_1}) < R(A_{i_2}) < \ldots < R(A_{i_t})$. Следовательно, можно перенумеровать объекты, не изменив вероятностей любого события, такого, что $\Pr(A) = f(\pi_{ij})$, $i \neq j$. — *Прим. пер.*

Тогда по принципу включения и исключения [1] получаем

$$\Pr(x_1 \geqslant M) = \Pr\left(\sum_{i=1}^{t} E_i\right) = \sum_{s=1}^{t} (-1)^{s-1} \binom{t}{s} \Pr(E_1 E_2 \ldots E_s). \tag{2.3.8}$$

Для больших M последние члены в сумме будут в основном нули и лишь несколько значений могут оказаться большими. В частности, только наибольшее значение может превышать $n(t-1) - \frac{1}{2}n$, так что для $M > n(t-1) - \frac{1}{2}n$

$$\Pr(x_1 \geqslant M) = \binom{t}{1} \Pr(E_1) = t \cdot 2^{-n(t-1)} \sum_{k=M}^{n(t-1)} \binom{n(t-1)}{k}. \tag{2.3.9}$$

Точная оценка вероятности $\Pr(x_1 \geqslant M)$ значительно более трудоемка для малых значений M, однако часто используется простое приближение. Предположим, что M достаточно велико, так что $\Pr(E_1) < 1/t$; тогда можно применить неравенства Бонферрони*, после чего получаем

$$t\Pr(E_1) - \binom{t}{2} \Pr(E_1 E_2) \leqslant \Pr(x_1 \geqslant M) \leqslant t \Pr(E_1).$$

Так как сумма очков постоянна, мы имеем

$$\Pr(E_1 \mid E_2) \leqslant \Pr(E_1),$$

и поэтому **

$$\Pr(E_1 E_2) \leqslant [\Pr(E_1)]^2.$$

Отсюда следует, что

$$t\Pr(E_1) - \binom{t}{2} [\Pr(E_1)]^2 \leqslant \Pr(x_1 \geqslant M) \leqslant t \Pr(E_1). \tag{2.3.10}$$

Поскольку $\Pr(E_1)$ можно вычислить прямо или по таблицам биномиального распределения, легко получить границы для $\Pr(x_1 \geqslant M)$. Так, если $t\Pr(E_1) = \beta$, имеем

$$\beta - \frac{(t-1)}{2t}\beta^2 \leqslant \Pr(x_1 \geqslant M) \leqslant \beta,$$

и расстояние между двумя границами уменьшается при уменьшении t или β.

---

[1] См., например, [43, т. 1, § 4.1].

* Неравенство Бонферрони см. в [43, т. 1, § 4.3, § 4.5, задача 17, гл. V]. — *Прим. пер.*

** $P(E_1E_2) = P(E_1|E_2)\,P(E_2) \leqslant P(E_1)\,P(E_2) \leqslant [P(E_1)]^2$,
так как $P(E_2) \leqslant P(E_1)$,
$P(a_2 \geqslant M) \leqslant P(a_1 \geqslant M)$, если A_1 — объект с наибольшим значением x_1. — *Прим. пер.*

Так как значение a_i распределено биномиально

$$b\left(\frac{1}{2},\quad n(t-1)\right)\ \text{при}\ H_0,$$

$$d_i=2\left[a_i-\frac{1}{2}n(t-1)\right]\Big/\sqrt{nt} \tag{2.3.11}$$

имеет нулевое среднее и дисперсию $\sigma^2 = (t-1)/t$. Зная также, что $\Sigma d_i = 0$ и что d_i симметрично коррелированы с коэффициентом корреляции*, скажем, ρ, мы имеем

$$\operatorname{var}(\Sigma d_i) = t\sigma^2 + 2\binom{t}{2}\rho\sigma^2 = 0,$$

так что $\rho\sigma^2 = -1/t^2$. Матрица ковариации Σd_i случайного вектора $\mathrm{d} = (d_1, d_2, \ldots, d_t)$ поэтому есть $(t \times t)$-матрица с элементами $(t-1)/t$ на главной диагонали и $-1/t$ вне ее. Из многомерной центральной предельной теоремы (см., например, [3]) следует, что асимптотическое распределение d при $n \to \infty$ нормальное $N(0, \Sigma d)$. Кроме того, определитель

$$\begin{vmatrix} \frac{t-1}{t}-\lambda & -\frac{1}{t} & & -\frac{1}{t} \\ -\frac{1}{t} & \frac{t-1}{t}-\lambda & & \\ \cdots & & \cdots & \\ -\frac{1}{t} & -\frac{1}{t} & \frac{t-1}{t} & -\lambda \end{vmatrix} = -\lambda(1-\lambda)^{t-1},$$

так что Σd имеет $(t-1)$ характеристических корней, равных 1, а остальные равны нулю. Как показано у Крамера [27, § 24.5], вышеупомянутый результат достаточен, чтобы доказать, что

$$D_n = \Sigma d_i^2 = 4T_n/(nt)$$

имеет предельное χ^2-распределение с $(t-1)$ степенями свободы при $n \to \infty$.

Действительно, можно было ожидать, что D_n будет приближенно распределено как χ^2_{t-1}, что верно, когда $n(t-1)$ не слишком мало; d_i приближенно нормальны $N(0, (t-1)/t)$, и Σd_i^2 можно приблизить, применив ортогональное преобразование в $\sum_{j=1}^{t-1} y_j^2$, где y_j некоррелированы и имеют приближенно нормированное нормальное распределение.

Аппроксимацию χ^2-распределением для D_n предложил Дарбин [34] и проверил Старкс [128] сравнением с точными результатами, полу-

---

* Из гипотезы $H_0 : \pi_{ij} = 1/2$ для всех $i \neq j$, поэтому a зависит от a_j только через $\alpha_{ij} = \sum_f \xi_{ij}^{(f)}$, $\xi_{ij}^{(f)} = \begin{cases} 1 & \Pr=\frac{1}{2} \\ 0 & \Pr=\frac{1}{2} \end{cases}$ для любых $i \neq j$, $d_i = ka_i - c$, $\rho(d_i, d_j) = \rho(a_i, a_j)$. — *Прим. пер.*

ченными из исходных таблиц, которые приведены в 2.2. Это приближение оказывается в разумной степени удовлетворительным в окрестности верхней 10-процентной точки и выше и может быть с некоторым доверием использовано за пределами табл. 1 приложения, которая построена по исходным таблицам [138]. Удивительно, что это простое приближение, в общем вполне эффективно около верхних процентных точек, как и более сложные приближения, которые предлагались.

Размах результатов

Из предыдущего мы можем ожидать, что распределение размаха* величин d_i будет асимптотически такое же, как распределение размаха W_t для t независимых нормальных переменных с дисперсией

$$\sigma^2 (1 - \rho) = (t - 1)/t + 1/t = 1$$

(сравните с [75]). Функцию распределения можно получить из [114, табл. 23]**. Следовательно, $\Pr (W_t \geqslant 2R/\sqrt{nt})$ — легко достижимое приближение для $\Pr$ (размах $a_i \geqslant R$) и его точность будет улучшаться по мере роста n $(t - 1)$. Эмпирически установлено, что лучшее приближение получается, с учетом поправки на непрерывность (половины обычного значения), для приближающей функции $\Pr \left(W_t \geqslant \left(2R - \frac{1}{2}\right)/\sqrt{nt}\right)$.

2.4. ТЕОРИЯ РАСПРЕДЕЛЕНИЯ В СЛУЧАЕ НЕНУЛЕВОЙ ГИПОТЕЗЫ

Сначала, имея в виду применения в гл. 6, мы рассмотрим ненулевой случай, по существу, модель (1.3.4), для которой

$$\Pr (x_{ij\gamma} = 1) = \pi_{ij} \ (i, j = 1, 2, \ldots, t;\ i \neq j;\ \gamma = 1, 2, \ldots, n).$$

Совместное распределение очков

Пусть α_{pr} $(p > r)$ — число случаев, в которых A_p предпочтительнее, чем A_r в n сравнениях. Совместное распределение этих независимых переменных дается выражением

$$f(\alpha_{t,\, t-1}, \alpha_{t,\, t-2}, \ldots, \alpha_{21}) = \prod_{p > r}^{t} \binom{n}{\alpha_{pr}} \pi_{pr}^{\alpha_{pr}} \pi_{rp}^{n - \alpha_{pr}}. \tag{2.4.1}$$

Результаты можно записать в виде

$$\begin{aligned} a_t &= \alpha_{t1} + \alpha_{t2} + \ldots + \alpha_{t,t-1}, \\ a_{t-1} &= \alpha_{t-1,1} + \alpha_{t-1,2} + \ldots \\ &\ldots + \alpha_{t-1,t-2} + (n - \alpha_{t,t-1}), \\ a_1 &= (n - \alpha_{21}) + (n - \alpha_{31}) + \ldots + (n - \alpha_{t1}). \end{aligned} \tag{2.4.2}$$

---

* Размахом называется разность $d_{(t)} - d_{(1)}$, где $d_{(1)} \leqslant \ldots \leqslant d_{(t)}$ — члены вариационного ряда из $d_1, \ldots, d_t$. — *Прим. пер.*

** Или из книги [158, табл. 3.8]. — *Прим. пер.*

Средние, дисперсии и ковариации значений

Если $a_{i\gamma}$ — число очков, полученных объектом A_i в γ-м повторении (реплике), мы сразу же имеем

$$\mathscr{E}(a_{i\gamma})=\sum_{i}^{t}{}'\,\pi_{ij},\quad \operatorname{var}(a_{i\gamma})=\sum_{j}^{t}{}'\,\pi_{ij}\,\pi_{ji}.$$

Для нахождения η ковариаций $a_{i\gamma}$ и $a_{j\gamma}$ ($i\neq j$) рассмотрим дисперсию $a_{i\gamma}-a_{j\gamma}$, для удобства полагая $i=1$ и $j=t$. Тогда

$$\operatorname{var}(a_{1\gamma}-a_{t\gamma})=\operatorname{var}\left(\sum_{k=2}^{t}x_{1k\gamma}-\sum_{k=1}^{t-1}x_{tk\gamma}\right)=$$

$$=\operatorname{var}\left[\sum_{k=2}^{t-1}(x_{1k\gamma}-x_{tk\gamma})+2x_{1t\gamma}-1\right]=\sum_{k=2}^{t-1}(\pi_{1k}\,\pi_{k1}+\pi_{tk}\,\pi_{kt})+$$

$$+4\pi_{1t}\,\pi_{t1}=\operatorname{var}(a_{1\gamma})+\operatorname{var}(a_{2\gamma})+2\pi_{1t}\,\pi_{t1}.$$

Отсюда $\operatorname{cov}(a_{1\gamma},\ a_{t\gamma})=-\pi_{1t}\pi_{t1}$.

Из независимости повторных сравнений следует, что

$$\operatorname{var}(a_1)=n\sum_{k=2}^{t}\pi_{1k}\,\pi_{k1},\quad \operatorname{cov}(a_1,\ a_t)=-n\pi_{1t}\,\pi_{t1}. \qquad (2.4.6)$$

Разности значений

Особый интерес для последующей работы представляет распределение вектора разностей

$$\mathbf{b}'=(b_1,\ b_2,\ \ldots,\ b_{t-1}),\ \text{где}\ b_i=a_i-a_t\quad (i=1,\ 2,\ \ldots,\ t-1)$$

Случайная величина b_1 имеет среднее

$$\beta_1=n\left(\sum_{k=2}^{t}\pi_{1k}-\sum_{k=1}^{t-1}\pi_{tk}\right) \qquad (2.4.7)$$

и дисперсию

$$\sigma_{b_1b_1}=n\left[\sum_{k=2}^{t-1}(\pi_{1k}\,\pi_{k1}+\pi_{tk}\,\pi_{kt})+4\pi_{1t}\,\pi_{t1}\right]. \qquad (2.4.8)$$

Так же получается ковариация b_1 и b_2

$$\sigma_{b_1b_2}=n\left[\sum_{k=1}^{t-1}\pi_{tk}\,\pi_{kt}+(\pi_{1t}\,\pi_{t1}+\pi_{2t}\,\pi_{t2}-\pi_{12}\,\pi_{21})\right].$$

Теперь

$$b_i=\sum_{\gamma=1}^{n}(a_{i\gamma}-a_{t\gamma})=\sum_{\gamma}b_{i\gamma}$$

и $b_{i\gamma}$ имеют для любого заданного γ средние, дисперсии и ковариации β_1/n, $\sigma_{b_1b_1}/n$ и $\sigma_{b_1b_2}/n$ соответственно. Из независимости реплик (повторов) и многомерной центральной предельной теоремы (например, [3]) следует, что предельное распределение при $n\to\infty$ вектора $\frac{1}{\sqrt{n}}(\mathbf{b}-\boldsymbol{\beta})$ будет многомерным нормальным $N(\mathbf{0},\boldsymbol{\Sigma})$, где $\boldsymbol{\Sigma}$—матрица $\frac{1}{n}(\sigma_{b_ib_j})$.

Замечания

Предположим, что в двух множествах попарных сравнений* пяти объектов $A_1, A_2, \ldots, A_5$ мы имеем $A_i \to A_j$ для $i < j$, за исключением того, что в первом множестве $A_3 \to A_1$ и во втором множестве $A_5 \to A_1$. В обоих случаях есть единственная несовместимость по сравнению с ранжировкой $A_1 \to A_2 \to \ldots \to A_4 \to A_5$**. С другой стороны, эти множества имеют возрастание своих исходов экспериментов $[3^3 10]$, $[3^2 2 1^2]$ с различным числом циклических триад, а именно $c = 1,3$. Распределение числа несовместимостей с ближайшей ранжировкой было изучено Слейтером [127], который пропагандировал использование несовместимостей, отличных от циклических триад, как критерия для гипотезы H_0 о случайности. Он доказывает, что каждой несовместимости должен приписываться равный вес, пока H_0 не отвергнута. Представляется, что эта точка зрения пренебрегает тем, что более разумной альтернативой H_0 против случайности (отличной от ляпсусов и случайных ошибок) служит предположение, что непоследовательность $A_3 \to A_1$ более правдоподобна***, чем $A_5 \to A_1$. Поэтому первое множество сравнений более согласовано с H_a, чем второе, менее совместимое с H_0, что отражено его более низким c-значением.

УПРАЖНЕНИЯ

2.1. Покажите, что коэффициент совместимости в мнениях ζ равен нулю, когда число циклических триад c максимально [93].

2.2. Покажите, что, если объекты $A_1, A_2, \ldots, A_t$ эквивалентны, математическое ожидание и дисперсия c даются формулами:

$$\mu = \frac{1}{4}\binom{t}{3}, \quad \sigma^2 = \frac{3}{16}\binom{t}{3} \quad (\text{см. [106]}).$$

2.3. Докажите (2.3.4).

2.4. Пусть d — разность числа очков объектов A_1 и A_2 ($d = a_1 - a_2$) при их сравнениях между собой и с объектами $A_3, A_4, \ldots, A_t$. Покажите, что в предположении истинности гипотезы H_0 о случайности распределение вероятности и характеристическая функция d равны соответственно:

$$p(d) = \frac{1}{2^{2t-3}}\left[\binom{2t-4}{t-d-1} + \binom{2t-4}{t-d-3}\right],$$

$$d = -(t-1), \ -(t-2), \ldots, (t-1),$$

$$\varphi(u) = \left(\cos \frac{1}{2} u\right)^{2t-4} \cos u.$$

Это доказывает (2.3.5).

---

* То есть в двух турнирах. — *Прим. пер.*

** То есть различие между 1, 2 турнирами и транзитивным турниром в одной дуге, которая ориентирована по-разному. Слейтер [127] изучал распределение статистики $i_{\min}$, равной числу несовместимостей (по-разному ориентированных дуг) у турнира T_t парных сравнений $A_1, \ldots, A_t$ и транзитивного турнира, обладающего следующим свойством: число переориентировки дуг в T_t для получения исконного транзитивного турнира минимально по всем транзитивным турнирам с t вершинами. Подход Слейтера прямо связан с подходом Кемени—Снелла [175] (см. обзор [216]) и существенно развит в работах [314], [303], [315], [260], [261], [279], [271], [272] и многих других (см. обзоры [256] и [281]). — *Прим. пер.*

*** В упорядочении $A_1 \to A_2 \to A_3 \to A_4 \to A_5$ число объектов между A_1 и A_3 меньше, чем между A_1 и A_4. — *Прим. пер.*

2.5. Для $n = 1$ докажите, что при H_0

$$\Pr(x_1 \geqslant t-2) = \frac{t^2}{2^{t-1}} - \frac{\binom{t}{2}(t-1)}{2^{2t-4}} + \frac{\binom{t}{3}}{2^{3t-7}}.$$

2.6. Покажите, что при H_0 x_1 асимптотически ($n \to \infty$) распределено как

$$\frac{1}{2} n(t-1) + \frac{1}{2}(u_{\max} - \bar{u})\sqrt{nt},$$

где $u_{\max}$, $\bar{u}$ соответственно наибольшее и среднее из t нормированных нормальных случайных величин $N(0, 1)$.

2.7. Для множества из $(t-1)$ взаимно ортогональных контрастов

$$Q_k = \sum_{i=1}^{t} L_{i(k)} d_i \quad (k=1, 2,\dots, t-1),$$

где $\sum_{i=1}^{t} L_{i(k)} = 0$ и d_i определено в (2.3.11), покажите, что дисперсия Q_k равна:

$$S_k = \sum_{i=1}^{t} L_{i(k)}^2$$

и что*

$$D_n = \sum_{k=1}^{t-1} Q_k^2/S_k.$$

2.8. Покажите, что в ненулевом случае коэффициент согласия u имеет среднее и дисперсию

$$\mathscr{E}(u) = 1 - \frac{8}{t(t-1)} \sum_{i<j} \pi_{ij}\pi_{ji},$$

$$\operatorname{var}(u) = \frac{64}{n(n-1)t^2(t-1)^2} \sum_{i<j} \{(n-1)[\pi_{ij}\pi_{ji} - 4\pi_{ij}^2\pi_{ji}^2] + 2\pi_{ij}^2\pi_{ji}^2\} \text{ (см. [39])}.$$

2.9. Докажите, что если сумма целых чисел $a_1, a_2, \dots, a_t$ равна $\frac{1}{2}t(t-1)$, то они представляют возможные исходы простого эксперимента парных сравнений тогда и только тогда, когда

$$\sum_r a_i \geqslant \frac{1}{2} r(r-1) \text{ для } r = 1, 2,\dots, t-1,$$

где $\sum_r a_i$ — сумма по любым из r величин a_i (см. [96], [2]).

---

* $D_n = \Sigma d_i^2$ (см. 2.3). — *Прим. пер.*

3 ГЛАВА ПРОВЕРКА ЗНАЧИМОСТИ

3.1. ВВЕДЕНИЕ

Эта глава описывает различные критерии значимости для результатов a_i сбалансированного эксперимента парных сравнений. Многие из них аналогичны широко известным критериям для разностей средних по обработкам (опытам) в дисперсионном анализе. Соответствующая теория распределений в основном была построена в 2.3. Здесь же применение критериев будет показано на числовом примере. Даются также необходимые нестандартные таблицы.

Читатель, желающий перейти от приводимых проверок значимости к доверительным интервалам, может сделать это обычным способом.

3.2. КРИТЕРИЙ ДЛЯ ОСОБОГО ОБЪЕКТА

Иногда какой-то объект, скажем, A_1, априори особенно интересен в эксперименте парных сравнений. В таком случае у экспериментатора может возникнуть желание проверить частную нуль-гипотезу:

$$H_0' : \pi_{1.} = \frac{1}{2},$$

что A_1 — в точности средний объект, против альтернативы

$$H_a : \pi_{1.} > \frac{1}{2},$$

что он лучше, чем средний. Когда верна гипотеза H_0 о случайности сравнений

$$H_0 : \pi_{ij} = \frac{1}{2}; \; i, j = i, \ldots, t, \; i \neq j$$

очки a_1 объекта A_1 имеют биномиальное распределение $b\left(\frac{1}{2}, n(t-1)\right)$ и к проверке H_0 против H_a можно применить обычный биномиальный критерий. О гипотезе H_0' заметим, что для ее выполнения не требуется равенства всех π_{1j} между собой. В этом случае a_1 имеет обобщенное биномиальное распределение с тем же математическим ожиданием, что и при H_0; но с дисперсией меньшей, чем при H_0, и равной (сравните [91, § 5.10])

$$n \sum_{i=2}^{t} \left(\pi_{1i} - \frac{1}{2}\right)^2 .$$

Это уменьшение дисперсии предполагает, и это можно строго показать [80], что биномиальный критерий для H_0' против H_α будет консервативным критерием. Это означает, что если уровень значимости биномиального критерия равен α при выполнении H_0, то при H_0' он во всяком случае не превышает α. В этом смысле биномиальный критерий, пригодный в первом приближении для проверки H_0 против H_α, является безопасным критерием для проверки H_0' против H_α. Есть веские указания на то, что аналогичные свойства присущи и критериям из следующих параграфов.

Конечно, биномиальность распределения можно также использовать для проверки H_0' против альтернатив $\pi_{1.} < \frac{1}{2}$ и $\pi_{1.} \neq \frac{1}{2}$. В действительности консервативность также сохраняется, когда проверяется более общая нуль-гипотеза

$$\pi_{1.} = \pi_0 \quad (0 < \pi_0 < 1)$$

и a_1 полагается биномиально распределенным с $b\,(\pi_0,\ n\,(t-1))$. Часто нормальная аппроксимация биномиального распределения дает критерий достаточной точности.

3.3. ПРОВЕРКА ЭКВИВАЛЕНТНОСТИ ДВУХ ОСОБЫХ ОБЪЕКТОВ

Предположим, что перед осуществлением эксперимента с t объектами возник вопрос о существовании различий между двумя объектами, скажем, A_1 и A_2. После эксперимента мы можем проверить $H_0' : \pi_{1.} = \pi_{2.}$ против $H_a : \pi_{1.} \neq \pi_{2.}$. Распределение критерия для выражения $|a_1 - a_2|$ рассматривалось в 2.3 при условии истинности гипотезы случайности H_0. Следовательно, для проверки H_0' против H_a мы используем следующую *процедуру проверки*:

1) выбираем желаемый уровень значимости α;

2) находим m_c, наименьшее целое значение числа m, для которого $P_{tnm} = \Pr(|a_1 - a_2| \geqslant m \mid H_0)$ не превосходит α;

3) принимаем H_a, если $|a_1 - a_2|$ больше или равно m_c.

Для проверки H_0' против односторонней альтернативы $\pi_{1.} > \pi_{2.}$ надо применить рассмотренную выше процедуру проверки с $\frac{1}{2} P_{tnm}$ вместо P_{tnm} на втором этапе и без знака модуля на третьем этапе. Критические значения m для одно- и двусторонних критериев при $\alpha = 0{,}01$ или $0{,}05$ приведены в приложении (табл. 2).

3.4. КРИТЕРИЙ ДЛЯ НАИБОЛЬШЕГО ЗНАЧЕНИЯ («ПОБЕДИТЕЛЯ»)

После проведения эксперимента парных сравнений часто хочется знать, является ли объект с наибольшим значением x_1 существенно лучшим, чем средний. Назовем этот объект A_μ и соответствующую среднюю вероятность предпочтения — $\pi_{\mu.}$. Тогда нужная нулевая гипотеза

$$H_0': \pi_{\mu.} = \frac{1}{2}$$

должна проверяться против альтернативы

$$H_a: \pi_{\mu.} > \frac{1}{2}.$$

Рассмотрение распределения наибольшего значения x_1 в § 2.3 приводит к следующей *процедуре проверки*:

1) выбираем желаемый уровень значимости α;

2) находим в таблице кумулятивной функции биномиального распределения целое число M_β, такое, что*

$$t \Pr (a_1 \geqslant M_\beta \mid H_0) = \beta \leqslant \alpha < t \Pr (a_1 \geqslant M_\beta - 1 \mid H_0);$$

3) если наибольшее значение в эксперименте равно или больше M_β, принимаем гипотезу о том, что соответствующий объект лучше, чем средний.

С очевидными модификациями процедуру можно использовать для проверки гипотезы о том, будет ли объект с наименьшим значением хуже среднего.

Ясно, что критерии, рассматриваемые в данном параграфе, по существу, отличаются от предшествующих критериев тем, что нет априорного выделения каких-либо особых объектов. Следует также отметить, что частная нуль-гипотеза H_0' в данном случае фактически эквивалентна гипотезе случайности H_0, если исходная вероятностная модель линейна. Однако без такой линейности $\pi_{\mu.}$ может быть равно 1/2 и без равенства $\pi_{\mu i}$ между собой.

3.5. ОБЩИЙ КРИТЕРИЙ ЭКВИВАЛЕНТНОСТИ

Мы переходим к рассмотрению критерия, который аналогичен F-критерию для средних по обработкам (опытам) в дисперсионном анализе. Такой общий критерий эквивалентности объектов можно построить на стандартизированной сумме квадратов значений D_n из § 2.3. Нуль-гипотеза для критерия —

$$H_0': \pi_{i.} = \frac{1}{2} \quad \text{для всех } i;$$

---

* Таблицы функции биномиального распределения см. в [158], где $n = 5/5/30$ и $\alpha = 0{,}01$; $0{,}02/0{,}02/0{,}10/0{,}10/0{,}50$, или в [196], где $n = 2/1/25$. Для больших n рекомендуется пользоваться аппроксимациями биномиального распределения нормальным распределением, см. [157]. — *Прим. пер.*

альтернативная гипотеза гласит, что не все $\pi_{i.}$ равны между собой. H_0' эквивалентна H_0, если модель линейна.

Процедура проверки:

1) выбираем желаемый уровень значимости α;

2а) при небольших экспериментах используем табл. 1 (приложение) для нахождения критического значения суммы квадратов $\Sigma\, \alpha_i^2$;

2б) для больших экспериментов сравниваем

$$D_n = 4\left[\sum_{i=1}^{t} a_i^2 - \frac{1}{4}\, tn^2 (t-1)^2\right] \Big/ nt$$

с верхней критической точкой уровня α χ^2-распределения с $(t-1)$ степенями свободы;

3) отклоняем H_0', если наблюдаемое значение Σa_i^2 или D_n больше соответствующего критического значения или равно ему.

3.6. МЕТОД НАИМЕНЬШЕЙ ЗНАЧИМОЙ РАЗНОСТИ

Как следует из названия, этот метод аналогичен методу парных t-критериев, применяемому после получения значимо большого F-критерия в общей проверке равенства средних по обработкам. В дисперсионном анализе это старейший из многочисленных критериев множественных сравнений [42, гл. II.1], целью которых является ответ на вопрос: какие способы обработки (опыты) отличаются друг от друга? Все, что нужно сделать в случае парных сравнений, — это использовать критические значения Σa_i^2 или D_n в последовательном применении двусторонних критериев для пар объектов, как в 3.3.

Процедура проверки

Этапы (1)—(3) аналогичны рассмотренным в 3.5;

4) если на этапе (3) не найдено значимого различия между объектами, принимаем гипотезу H_0' и на этом заканчиваем проверку; в противном случае находим критическое значение m_c двустороннего критерия для двух особых объектов, как в 3.3, и считаем, что объекты значимо различаются, если каждая пара значений отличается на m_c или больше.

3.7. КРИТЕРИЙ МНОЖЕСТВЕННОГО СРАВНЕНИЯ * ДЛЯ РАЗМАХОВ

Хорошо известно (см., например, [33]), что применение значимого общего F-критерия с последующим набором t-критериев может вести к высокой вероятности объявления некоторых объектов существенно

* См. [214, гл. 3] и [297]. — *Прим. ред.*

различными, когда на самом деле это не так. С другой стороны, метод, предложенный Тьюки и рассматриваемый нами в этом параграфе, страдает, возможно, противоположным недостатком, а именно затрудняет выявление имеющихся различий между объектами. Однако этот подход имеет ту привлекательную черту, что пары объектов можно выбрать для сравнения после изучения результатов эксперимента, и вероятность того, что будут сделаны *какие-либо* неверные выводы о различиях между объектами, можно контролировать на выбранном уровне значимости α.

Для экспериментов по парным сравнениям процедура основана на размахе значений, полученных для t объектов. Применяя теоретические положения из 2.3 к методу Тьюки, мы получаем следующую *процедуру проверки*:

1) выбираем желаемый уровень значимости α;

2) находим положительное целое число $R_{\beta(\alpha)}$, такое, что

$$\Pr[\text{range } a_i \geqslant R_{\beta(\alpha)}] = \beta \leqslant \alpha < \Pr[\text{range } a_i \geqslant R_{\beta(\alpha)} - 1], \quad (3.7.1)$$

где вероятности вычисляются при условии выполнения нуль-гипотезы случайности H_0:

а) если эксперимент мал, используем для получения $R_{\beta(\alpha)}$ табл. 3 приложения[1];

б) если эксперимент слишком велик, чтобы применять метод (а), находим верхнюю критическую точку уровня α, скажем, $W_{t,\alpha}$, для распределения W_t [114, табл. 22][2] решаем уравнение $W_{t,\alpha} = (2R^* - \frac{1}{2})(nt)^{-\frac{1}{2}}$ относительно R^*. Если R^+ — наименьшее целое число, равное или большее, чем R^*, превышает $n(t-1) - \frac{1}{2}n$, воспользуемся уравнением (3.7.1) для получения β и $R_{\beta(\alpha)}$; в противном случае полагаем $R_{\beta(\alpha)} = R^+$;

3) считаем любую парную разность, равную или большую, чем $R_{\beta(\alpha)}$, значимой.

3.8. МЕТОД СУЖДЕНИЯ О КОНТРАСТАХ ЗНАЧЕНИЙ

Контрасты значений были определены в упражнении 2.7. В дисперсионном анализе, как было показано Шеффе [123], F-критерий служит основой критерия множественного сравнения для контрастов средних по вариантам опыта. Аналогично можно использовать D-статистику для рассмотрения некоторого, фактически любого, числа контрастов значений, контролируя вероятность ошибочного объявления значимости на выбранном уровне α. Логика этой процедуры такая же, как и для критерия множественных сравнений размахов. В действительности процедура проверки, приводимая ниже, может быть прямо применена к разностям значений, которые на самом деле являются контрастами. Однако, если нас интересуют только разности, предпочтителен предыдущий метод, который менее консервативен.

---

[1] Эта таблица также рассматривается в [129].

[2] См. также [158]. — *Прим. пер.*

Процедура проверки

Этапы (1)—(3) такие же, как в 3.5;

4) если на этапе (3) не обнаружено значимых различий между объектами, принимаем H_0' и на этом заканчиваем проверку; в противном случае для любого выбранного контраста $\sum_{i=1}^{t} L_i d_i$ $(\Sigma L_i = 0)$ вычисляем:

$$\text{а)}\quad Q^2 = \left(\sum_{i=1}^{t} L_i d_i\right)^2 = 4\,(\Sigma L_i\, a_i)^2/(nt);$$

$$\text{б)}\quad S = \Sigma L_i^2;$$

$$\text{в)}\quad SD_{n,c};$$

5) если $Q^2 \geqslant SD_{n,c}$, объявляем контраст значимо отличающимся от нуля.

3.9. ПРИМЕР

Для иллюстрации представленных в данной главе процедур проверки они применены к результатам эксперимента парных сравнений, проведенного с пятью сортами копирки и описанного Флекенштейн и др. [47]. Эксперимент и анализ данных были первоначально проведены в соответствии с методом, описанным Шеффе [122]. Для целей нашего анализа авторская семиточечная шкала была сжата, пренебрегая степенью предпочтения, а для совпадающих рангов предпочтения назначались случайно для одного из двух объектов каждой такой пары. Число случаев, когда каждый сорт бумаги предпочитался в результате 30 повторений эксперимента, следующее:

$$a_1 = 66;\ a_2 = 51;\ a_3 = 89;\ a_4 = 24;\ a_5 = 70.$$

Таким образом, $t = 5$, $n = 30$ и мы произвольно выбираем всегда 5%-ный уровень значимости. Конечно, на практике нет необходимости применять все критерии для одного и того же набора данных.

Критерий для данного особого сорта бумаги

Предположим, что до проведения эксперимента с копиркой бумага № 5 была рекомендована в качестве лучшей из-за невысокой стоимости и предполагаемого хорошего качества. В таком случае экспериментатор может быть заинтересован в выяснении того, является ли бумага № 5 лучшей, чем среднее из пяти сортов. Проверка H_0': $\pi_{5.} = \frac{1}{2}$ против H_a: $\pi_{5.} > \frac{1}{2}$ производится следующим образом:

1) уровень значимости 5%;

2) так как $n\,(t-1) = 30 \times 4 = 120$, мы можем применить нормальную аппроксимацию для нахождения критического значения (a_c) для значения выделенного объекта (бумаги № 5):

$$a_c = 1{,}64\,[n\,(t-1)/4]^{\frac{1}{2}} + \frac{1}{2}\,n\,(t-1) + \frac{1}{2} = 69{,}48;$$

3) $a_c = 70$;

4) $a_5 > a_c$ (действительный уровень значимости для a_5 равен 0,0412 при H_0'). Наш вывод заключается в том, что сорт № 5 лучше, чем среднее из пяти сортов.

Проверка эквивалентности для двух особых сортов бумаги

Если бы сорт 2 был дороже, чем сорт 4, то можно было бы предположить, что возникает потребность выяснить в ходе эксперимента, действительно ли сорт 2 превосходит сорт 4 или нет. Односторонний критерий $H_0 : \pi_{4.} = \pi_{2.}$ против $H_a : \pi_{2.} > \pi_{4.}$ таков:

1) уровень значимости 5%;

2) $1{,}64\,(nt/2)^{1/2} + 0{,}5 = 1{,}64\sqrt{75} + 0{,}5 = 14{,}7$, $m_c = 15$;

3) $a_2 - a_4 = 51 - 24 = 27 > m_c$. Мы принимаем альтернативную гипотезу о том, что бумага сорта 2 превосходит бумагу сорта 4.

Критерий для наибольшего значения

Естественно, что экспериментатору интересно знать, является ли наибольшее значение x_1, которое в этом примере принадлежит сорту 3, действительно значимо большим, чем среднее, т. е. $\pi_{\mu.} > \frac{1}{2}$. Проверка нулевой гипотезы против этой альтернативы проходит следующим образом:

1) уровень значимости 5%;

2) из таблицы биномиального распределения (например, см. [77]*) находим при $n = 120$ и $p = \frac{1}{2}$, что

$$\Pr\,(a_i \geqslant 74\,|\,H_0') = 0{,}033/5$$

и

$$\Pr\,(a_i \geqslant 73\,|\,H_0') > 0{,}05/5.$$

Поэтому $M_\beta = 74$ и $\beta = 0{,}033$;

3) $x_1 = a_3 = 89 > M_\beta$. Наш вывод заключается в том, что бумага сорта 3 значимо лучше, чем среднее.

Общий критерий эквивалентности

1) уровень значимости 5%;

2) $D_{n,c} \doteq \chi^2_{4,0;\,0,05} = 9{,}488$;

3) $$D_n = 4\left[\Sigma a_i^2 - \frac{1}{4}\,tn^2\,(t-1)^2\right] \Big/ (nt) =$$
$$= 4\,[20354 - 18000]/150 = 62{,}77.$$

Так как $D_n > D_{n,c}$, между значениями существует значимое различие.

* В этих таблицах число независимых испытаний 1/1/50/2/100/10/200/20/500/50/1000. — *Прим. пер.*

Метод наименьшей значимой разности

Этапы (1)—(3), как описано выше;

4) $1{,}96\left(\frac{1}{2}nt\right)^{\frac{1}{2}}+\frac{1}{2}=1{,}96\sqrt{75}+\frac{1}{2}=17{,}47,$

критическое значение $m_c = 18$, полученное упорядочение

$$\begin{array}{ccccc} a_4 & a_2 & a_1 & a_5 & a_3 \\ 24 & 51 & 66 & 70 & 89. \end{array}$$

Любые два сорта бумаги, которые не подчеркнуты одной и той же чертой, могут рассматриваться как существенно различающиеся.

Критерий множественного сравнения для размаха

1) уровень значимости 5%;

2) $W_{5,0;\,0,05} = 3{,}86$ (получено из [114, табл. 22])*,

$$R^* = \frac{1}{2} W_{5,0;\,0,05}\sqrt{nt}+\frac{1}{4}=3{,}86\sqrt{\frac{150}{4}}+\frac{1}{4}=23{,}887,$$

$$R^+ = 24 < n(t-1)-\frac{1}{2}n = 105,$$

$$\beta \doteq \Pr\left[W_t \geqslant \left(2R-\frac{1}{2}\right)(nt)^{-\frac{1}{2}}\right]=\Pr\left[W_5 \geqslant 3{,}878\right]=0{,}048$$

(из [114, табл. 23])*;

3)
$$\begin{array}{ccccc} a_4 & a_2 & a_1 & a_5 & a_3 \\ 24 & 51 & 66 & 70 & 89. \end{array}$$

Любые два сорта, которые не подчеркнуты одной и той же чертой на предыдущем шаге, могут рассматриваться как существенно различающиеся.

Методы контрастов

Этапы (1)—(3), как для общего критерия эквивалентности;

4а) деление на группы можно произвести в этом случае вычислением квадратов четырех контрастов:

$$Q_1^2 = 4\,(a_3 - a_2)^2/nt = 4\,(89 - 51)^2/150 = 38{,}5,$$
$$Q_2^2 = 4\,(a_3 - a_1)^2/nt = 4\,(89 - 66)^2/150 = 14{,}1,$$
$$Q_3^2 = 4\,(a_5 - a_2)^2/nt = 4\,(70 - 51)^2/150 = 9{,}6,$$
$$Q_4^2 = 4\,(a_2 - a_4)^2/nt = 4\,(51 - 24)^2/150 = 19{,}4;$$

* См. [158, 3.8б и 3.8а]. — *Прим. пер.*

4б) для каждого контраста

$$S = \sum_{i=1}^{2} L_i^2 = 1^2 + 1^2 = 2;$$

4в) $SD_{n,c} = 2\,(9{,}488) = 18{,}976$;

5) так как $Q_1^2 > Q_4^2 > SD_{n,c} > Q_2^2 > Q_3^2$, результирующее упорядочение в обычной интерпретации следующее:

a_4	a_2	a_1	a_5	a_3
24	51	66	70	89

Более общие контрасты могут быть легко проверены, как показано в 3.8.

УПРАЖНЕНИЯ

3.1. В примере из 3.9 определите методом контрастов, значимо ли отличается a_5 от среднего a_1 и a_2 на уровне 5%.

3.2. Покажите, что критерий множественного сравнения размахов может быть использован для объявления значимости на уровне α любого контраста $\sum_{i=1}^{t} c_i a_i$ $(\Sigma c_i = 0)$, который равен или больше $\frac{1}{2} R_{\beta(\alpha)} \Sigma\, |c_i|$, где $R_{\beta(\alpha)}$ было определено в (3.7.1). (Этот метод является аналогом процедуры Тьюки для контрастов. См. также [123].)

4 ГЛАВА ЛИНЕЙНЫЕ МОДЕЛИ

4.1. ОБЩИЙ ПОДХОД

Линейная модель была определена в § 1.3. Прежде чем детально рассмотреть некоторые ее важные модификации, мы остановимся на общем подходе, которым пользовался Ноэзер [112]. Первой решенной им задачей было оценивание ценностей V_i. ($i = 1, 2, \ldots, t$). Так как начало координат на линейной шкале выбирается произвольно, удобно ввести ограничение

$$\sum_{i=1}^{t} V_i = 0, \qquad (4.1.1)$$

которое допускает единственную оценку V_i (другая возможность состоит в том, чтобы положить $V_{(1)}$ — наименьшую ценность — равной нулю, но (4.1.1) имеет некоторые преимущества благодаря симметрии). Пусть

$$\delta_{ij} = V_i - V_j \; (j = 1, 2, \ldots, t;\; j \neq i), \qquad (4.1.2)$$

так что вероятность предпочтения π_{ij} равна:

$$\pi_{ij} = H(\delta_{ij}), \qquad (4.1.3)$$

а H — функция распределения, соответствующая выбранной линейной модели. Суммируя по всем j, отличным от $j = i$, мы имеем

$$\Sigma' \delta_{ij} = (t-1) V_i - \Sigma' V_j$$

или

$$V_i = \frac{1}{t} \sum_j{}' \delta_{ij}. \qquad (4.1.4)$$

Это приводит к следующему методу нахождения оценок v_i для V_i в сбалансированном эксперименте парных сравнений с n повторами. Пусть $p_{ij} = \alpha_{ij}/n$ — доля предпочтений объекта A_i объекту A_j. Подобно тому, как в (4.1.3) определяется δ_{ij}, определим d_{ij}:

$$H(d_{ij}) = p_{ij}, \qquad (4.1.5)$$

тогда*

$$d_{ji} = -d_{ij}. \qquad (4.1.6)$$

Трудности возникают тотчас же, если $p_{ij} = 0$ или $p_{ij} = 1$ *и* если H — функция распределения неограниченной случайной величины; при получении конечной величины для d_{ij} имеется возможность заменить наблюдаемые значения p_{ij} (0 или 1) на $1/(2n)$ или $1 - 1/(2n)$ соответственно.

Вообще говоря, невозможно удовлетворить соотношениям $v_i - v_j = d_{ij}$, отвечающим (4.1.2), так как уравнений больше, чем неизвестных. Однако эти соотношения могут выполняться «в среднем» в смысле (4.1.4), в предположении

$$v_i = \frac{1}{t} \sum_j{}' d_{ij}. \qquad (4.1.7)$$

Оценки v_i, полученные этим способом, есть оценки по методу (невзвешенных) наименьших квадратов, т. е. они минимизируют

$$S = \sum_{\substack{i, j \\ i \neq j}} (d_{ij} - V_i + V_j)^2 \qquad (4.1.8)$$

для

$$\frac{\partial S}{\partial V_i} = -2 \sum_j{}' (d_{ij} - V_i + V_j) + 2 \sum_j{}' (d_{ji} - V_j + V_i) =$$

$$= -4 \sum_j{}' (d_{ij} - \delta_{ij}) \, \{\text{при } (4.1.2),\ (4.1.6)\} =$$

$$= -4 \left(\sum_j{}' d_{ij} - tV_i \right) \{\text{при } (4.1.4)\},$$

что дает (4.1.7) как решение уравнений $\partial S / \partial V_i = 0$.

Понятно, что возможны другие оценки V_i, но простые оценки (4.1.7) имеют те преимущества, что метод вычисления один и тот же при любой функции распределения H. Это свойство, вообще говоря, не сохраняется для случая оценок, основанных на методах максимального

* Пусть $^{-}F(x)$ — функция распределения случайной величины $-\xi$, а $F(x)$ — функция распределения случайной величины ξ. Тогда в точках непрерывности (см. [43, т. 2, с. 186]):

$$^{-}F(x) = 1 - F(-x).$$

Для симметричного распределения $F(x) = {}^{-}F(x)$ и потому

$$F(x) = 1 - F(-x).$$

В нашем случае, так как H — симметричное распределение,

$$H(d_{ij}) = 1 - H(-d_{ij}). \qquad (A)$$

Кроме того,

$$H(d_{ij}) + H(d_{ji}) = 1. \qquad (B)$$

Поскольку H — взаимнооднозначная функция, $d_{ji} = -d_{ij}$. — *Прим. пер.*

правдоподобия, минимума хи-квадрат (χ^2) и т. д. Для рассмотренного выше способа нужна формула или таблицы функции H, или лучше H^{-1}, чтобы по данному p_{ij} получать d_{ij} из (4.1.5) как

$$d_{ij} = H^{-1}(p_{ij}).$$

Модель Тэрстоуна—Мостеллера

Здесь d_{ij} и, следовательно, v_i немедленно получаются из таблиц, дающих нормированную нормальную случайную величину, превышаемую с вероятностью $q_{ij} = 1 - p_{ij}$ (например*, [114, табл. 4] или [46, табл. IX]).

Еще проще использование ранкитов, как предложено Блиссом и др. [8], превратившими p_{ij} в ожидаемое значение $(n - \alpha_{ij} + 1)$-го наибольшего члена случайной выборки объема $n+1$ из совокупности нормированных нормальных случайных величин. Эти ранкиты (или нормализованные «очки») приближенно равны $H^{-1}(p_{ij})$ и могут быть найдены сразу для $n < 50$ из табл. XX [46] или табл. 28 [114]. Теперь для значений $p_{ij} = 0$ или 1 нет трудностей. Более подробные таблицы даны в [74]**.

Модель Брэдли — Терри

Из (1.3.10) находим, что

$$p_{ij} \equiv H(d_{ij}) = \frac{1}{2}\left[1 + \tanh\left(\frac{1}{2} d_{ij}\right)\right]$$

и, следовательно, что

$$d_{ij} = \log(p_{ij}/q_{ij})$$

— преобразование, которое упрощается благодаря табл. XI в [46]. В обычной записи мы поэтому имеем

$$v_i = \log \prod_j{}' [\alpha_{ij}/(n - \alpha_{ij})]. \tag{4.1.9}$$

Однако надо заметить, что v_i оценивают параметры, скажем, V_i', для которых

$$\Sigma V_i' = \Sigma \log \pi_i' = 0,$$

т. е. параметры π_i' не удовлетворяют ограничению $\Sigma \pi_i' = 1$, которое использовалось Брэдли и Терри, но

$$\pi_1' \pi_2' \dots \pi_t = 1.$$

---

* См. [158, табл. 1.3]. — *Прим. пер.*

** В [196, § 7, с. 151—154] приведена таблица для $n + 1 \leqslant 50$. — *Прим. пер.*

Параметры π_i' надо лишь нормировать для получения π_i с помощью формулы

$$\pi_i = \pi_i'/\Sigma\pi_i',$$

которая показывает, что

$$\pi_{ij} \equiv \frac{\pi_i}{\pi_i+\pi_j} = \frac{\pi_i'}{\pi_i'+\pi_j'}.$$

Равномерное распределение

Этот важный случай возникает, если мы положим $H(x) = x + \frac{1}{2}$.

Тогда $d_{ij} = p_{ij} - \frac{1}{2}$ и

$$v_i = \frac{1}{t}\sum_j{}' \left(p_{ij} - \frac{1}{2}\right) = \frac{1}{tn}\left[a_i - \frac{1}{2}n(t-1)\right], \qquad (4.1.10)$$

т. е. v_i — линейные функции очков (строчных сумм) и ранжирование объектов по значениям a_i то же, что и по v_i. Взяв математические ожидания, мы получим[1]

$$V_i = \frac{t-1}{t}\left(\pi_{i.} - \frac{1}{2}\right).$$

Таким образом, с точки зрения оценивания ценности V_i предположение о равномерном распределении эквивалентно работе с очками, как в предыдущих двух главах. Конечно, предыдущие предположения более сильные и определяют π_{ij} как функции от $\pi_{i.}$, поскольку

$$\pi_{ij} \equiv H(V_i - V_j) = V_i - V_j + \frac{1}{2} = \frac{t-1}{t}(\pi_{i.} - \pi_{j.}) + \frac{1}{2}. \quad (4.1.11)$$

Ноэзер утверждает, что в той степени, в какой все функции $H(x)$ приближенно линейны в окрестности $x = 0$, можно ожидать, что для множества V_i, которые не слишком отличаются друг от друга, оценки V_i будут приближенно эквивалентны оценкам, полученным для равномерного распределения, независимо от того, используется ли фактически функция $H(x)$. Действительно, если n не слишком мало, мы имеем приближенно

$$p_{ij} - \pi_{ij} = H(d_{ij}) - H(\delta_{ij}) \doteq (d_{ij} - \delta_{ij})\, h(\delta_{ij}),$$

где $h(x) = H'(x)$. Отсюда следует, что, если V_i не слишком различаются, хорошим приближением будет обычно

$$d_{ij} \doteq \left(p_{ij} - \frac{1}{2}\right) \Big/ h(0)$$

[1] Здесь $\pi_{i.} = \frac{1}{t-1}\sum_j{}' \pi_{ij}$, а не $1/t\sum_{j=1}^{t} \pi_{ij}$ ($\pi_{ii} = 1/2$), как обозначено в [112].

но веса W_{ij} неизвестны, так что необходима трудоемкая итеративная процедура (см. [86, с. 66]*). Чтобы обойти эти трудности, выполняется тригонометрическое преобразование. Полагая**

$$d_{ij} = 2 \sin^{-1} \sqrt{p_{ij}} - \frac{1}{2}\pi = \sin^{-1}(2p_{ij} - 1), \qquad (4.2.1)$$

мы имеем*** приближенно для больших n

$$\operatorname{var}(d_{ij}) = 1/n.$$

Преобразование (4.2.1) удобно выполнять с помощью табл. XII [70]****. Если надо, то можно еще улучшить свойство преобразования — стабилизировать дисперсию для малых n заменой p_{ij} на

$$(\alpha_{ij} + 3/8)/(n + 3/4).$$

Заметим, что (4.2.1) соответствует

$$p_{ij} = H(d_{ij}) = \frac{1}{2}(1 + \sin d_{ij}). \qquad (4.2.2)$$

Поэтому мы могли работать с тригонометрическим преобразованием в § 4.1. Новой чертой будет лишь то, что невзвешенная сумма квадратов S здесь приближенно равна взвешенной сумме. Более того, для всей области, кроме площадей нижнего и верхнего 5%-ных «хвостов», распределение тригонометрического преобразования полностью аналогично нормальному с точностью до масштаба.

4.3. МОДЕЛЬ БРЭДЛИ—ТЕРРИ — ОЦЕНИВАНИЕ

Брэдли и Терри [17] использовали метод максимального правдоподобия для оценивания t «баллов» π_i ($\pi_i \geqslant 0$, $\Sigma\pi_i = 1$), появившихся в их модели:

$$\pi_{ij} = \pi_i/(\pi_i + \pi_j) \qquad (i, j = 1, 2, \ldots, t;\ i \neq j).$$

С учетом независимости всех сравнений, вероятность наблюдения α_{ij} предпочтений объекта A_i в его сравнениях с A_j равна

$$\binom{n}{\alpha_{ij}} \left(\frac{\pi_i}{\pi_i + \pi_j}\right)^{\alpha_{ij}} \left(\frac{\pi_j}{\pi_i + \pi_j}\right)^{n - \alpha_{ij}} \qquad (\alpha_{ij} = 0, 1, \ldots, n).$$

---

* См. также [239]. — *Прим. пер.*

** Обозначение $\sin^{-1}$ соответствует принятому у нас arcsin. — *Прим. пер.*

*** $\sin^{-1} x = x$, тогда $\sin^{-1}(2p_{ij} - 1) \doteq 2p_{ij} - 1$,

$$D \sin^{-1}(2p_{ij} - 1) \doteq 4Dp_{ij} = \frac{4}{n} D\alpha_{ij} = \frac{4\pi_{ij}\pi_{ji}}{n} = \frac{1}{n},$$

так как $\pi_{ij}\pi_{ji} \approx \left(\frac{1}{2}\right)^2$. — *Прим. пер.*

**** Преобразование $\varphi = 2 \arcsin \sqrt{p}$ табулировано для $p = 0{,}0001$ (0,0001), 0,0045 (0,0005), 0,005 (0,001), 0,995 (0,0001) 0,9998 в книге Оуэна [196, табл. 9.9]. *Прим. пер.*

Функция правдоподобия L равна произведению таких вероятностей для всех $\binom{t}{2}$ независимых пар и может быть выражена через суммы очков a_i как

$$L = C\frac{\prod_{i=1}^{t} \pi_i^{a_i}}{\prod_{i<j} (\pi_i+\pi_j)^n}, \qquad (4.3.1)$$

где $C = \prod_{i<j} \binom{n}{\alpha_{ij}}$ не зависит от π_i. Отсюда следует, что a_i — достаточные статистики для π_i. Дифференцируя $\log L$ по π_i, мы получаем оценки максимального правдоподобия (м. п.-оценки) p_i для π_i из всех $t-1$ уравнений

$$\frac{a_i}{p_i} - n\sum_j{}' \frac{1}{p_i+p_j} = 0, \qquad (4.3.2)$$

взятых вместе с

$$\Sigma p_i = 1.$$

Мы видим, что p_i — функции от a_i и не зависят от индивидуальных α_{ij} в отличие от оценок (4.1.9). Для поиска p_i удобно записать (4.3.2) в форме

$$p_i = \frac{a_i}{n\sum_j{}' (p_i+p_j)^{-1}}. \qquad (4.3.3)$$

Тогда, начиная с исходного набора решений $(p_1^{(0)}, p_2^{(0)}, \ldots, p_t^{(0)})$, мы можем получить $p_i^{(1)}$ $(i = 1, 2, \ldots, t)$ из

$$p_i^{(1)} = \frac{a_i}{n}\left[\frac{1}{p_i^{(0)}+p_1^{(1)}} + \ldots + \frac{1}{p_i^{(0)}+p_{i-1}^{(1)}} + \right.$$
$$\left. + \frac{1}{p_i^{(0)}+p_{i+1}^{(0)}} + \ldots + \frac{1}{p_i^{(0)}+p_t^{(0)}}\right]^{-1}, \qquad (4.3.4)$$

продолжая итерационный процесс до тех пор, пока согласие между $p_i^{(r+1)}$ и $p_i^{(r)}$ не будет достаточно хорошим. Используя этот метод, Брэдли и Терри [17] и Брэдли [15] табулировали p_i с двумя десятичными знаками для всех возможных исходов экспериментов размером (t, n), не большим, чем (3; 10), (4; 8) и (5; 5)[1]. Дикстра [35] заметил, что итерационная процедура сходится медленно, поэтому важно найти хорошие начальные значения до вычислений по (4.3.4) в случае, когда размер эксперимента не охватывается таблицами. С этой целью он советовал для начала взять все p_j $(j \neq i)$ равными, так что

$$p_j = (1 - p_i)/(t - 1),$$

---

[1] Если $A_i \rightarrow A_j$, эти авторы назначали ранг 1 объекту A_i и ранг 2 объекту A_j. Суммарный ранг Σr_i соотносится с суммами очков a_i так: $\Sigma r_i = 2n(t-1) - a_i$.

получая p_i из (4.3.3) в виде

$$p_i = a_i/[n(t-1)^2 - a_i(t-2)]. \qquad (4.3.5)$$

Это приближение к м. п.-оценкам π_i может быть еще улучшено заменой p_i в (4.3.5) на $p_i - k_i$, где k_i — эмпирическая поправка, предложенная Дикстрой и равная

$$k_i = \frac{[(t-1)R - S^2]/t + a_i[n(t-1)-a_i]}{(t-1)R - S^2 + \Sigma a_i[n(t-1)-a_i]}\left[\left(\sum_i p_i - 1\right)\right], \qquad (4.3.6)$$

где

$$R = \sum a_i^2 \text{ и } S = \frac{1}{2}nt(t-1).$$

Заметим, что сумма p_i из (4.3.5), вообще говоря, не равняется единице. Приближения $p_i^{(r)}$ также не составляют в сумме единицы для $r > 0$, но нормировка их на каждом шаге не нужна, так как мы на самом деле работаем с их отношениями, а не с действительными значениями (сравните с [48]).

Форд заметил интересный факт: ранжировка, получаемая по м. п.-оценкам, та же, что и получаемая по суммам очков; если $a_i > a_j$, то из (4.3.2) следует:

$$0 < n\left(\sum_k{}' \frac{p_i}{p_i+p_k} - \sum_k{}' \frac{p_j}{p_j+p_k}\right) = n(p_i - p_j)\sum_k \frac{p_k}{(p_i+p_k)(p_j+p_k)},$$

таким образом

$$p_i > p_j.$$

Неравное число повторений

Если число сравнений A_i с A_j не постоянно, а равно n_{ij}, последнее утверждение может, конечно, не выполняться. Однако м. п-оценивание p_i из-за этого изменяется мало: нам нужно просто заменить (4.3.3) на

$$p_i = \frac{a_i}{\sum_j{}' n_{ij}(p_i+p_j)^{-1}}.$$

Сходимость итерационного процесса была установлена Фордом в общем случае при условии выполнения следующего допущения: «В каждом возможном разбиении объектов на 2 непустых подмножества некоторый объект второго множества предпочитается по крайней мере однажды некоторому объекту из первого множества».

Выполнение этого предположения весьма правдоподобно. Если оно нарушается, то существуют два подмножества S_1 и S_2 вообще без сравнений между множествами или же со всеми сравнениями в пользу S_1. В первом случае, очевидно, можно не ранжировать объекты из S_1 вместе с любым объектом из S_2. В последнем случае, который встречается и в сбалансированном эксперименте с равными n_{ij}, наибольшие значения π_i для объектов S_2 должны оказаться равными нулю, так как иначе нам пришлось бы увеличить функцию правдоподобия

$$L = \frac{C \prod\limits_{i=1}^{t} \pi_i^{a_i}}{\prod\limits_{i<j} (\pi_i + \pi_j)^{n_{ij}}}, \tag{4.3.7}$$

умножая эти π_i на общий множитель, меньший чем 1, и оставшиеся π_i — на общий множитель, больший чем 1. Эти множители можно выбрать так, чтобы в сумме они давали 1.

Сомножители L, включающие i, j, которые оба принадлежат либо S_1, либо S_2, остаются неизменными при этой операции. Однако множители, содержащие i из S_1 и j из S_2, возрастают, из-за чего возрастает и L. Но если π_i для объектов из S_2 должны быть все равны нулю, мы теряем надежду сравнить отдельные объекты S_2 в этом процессе. Конечно, мы можем обработать эти объекты совсем отдельно от объектов из S_1.

Как указывает Дикстра [37], результаты для неравных n_{ij} полезны также в планируемом эксперименте, где выполняется лишь часть всех возможных сравнений для того, чтобы уменьшить его объем. В этом случае $n_{ij} = 1$ или 0 в соответствии с тем, сравнивается ли A_i с A_j или нет. Есть и другие ситуации в планировании, когда n_{ij} не равны (см. гл. 5).

Точность p_i

Мы ограничимся изложением свойств оценок p_i в больших выборках для случая равных n_{ij}, хотя эти результаты могут быть обобщены и на случай неравных повторений [37]. Пусть

$$\lambda_{ij} = \frac{1}{\pi_i} \sum_j{}' \pi_j (\pi_i + \pi_j)^{-2} \qquad (i = 1, 2, \ldots, t),$$

$$\lambda_{ij} = -(\pi_i + \pi_j)^{-2} \qquad (j = 1, 2, \ldots, t;\ j \neq i).$$

Опираясь на стандартную теорию и учитывая ограничение $\Sigma\pi_i = 1$, Брэдли [16] показал, что $\sqrt{n}\,(p_1 - \pi_1), \ldots, \sqrt{n}\,(p_t - \pi_t)$ имеет, при $n \to \infty$, вырожденное многомерное нормальное распределение размерности $t - 1$ в пространстве размерности t с нулевым математическим ожиданием и ковариационной матрицей σ_{ij}, определенной равенством

$$\sigma_{ij} = \frac{\left\{\begin{matrix}\text{алгебраическое}\\ \text{дополнение } \lambda_{ij} \text{ в}\end{matrix} \begin{vmatrix} (\lambda_{ij}) & (1) \\ (1)' & 0 \end{vmatrix}\right\}}{\begin{vmatrix} (\lambda_{ij}) & (1) \\ (1)' & 0 \end{vmatrix}},$$

где $(1)'$, (1) обозначают соответственно вектор-строку и вектор-столбец из единиц.

4.4. МОДЕЛЬ БРЭДЛИ—ТЕРРИ — ПРОВЕРКИ ГИПОТЕЗ

В случае, когда вероятности парных сравнений описываются моделью Брэдли—Терри, нулевая гипотеза эквивалентности t объектов A_i превращается в

$$H_0 : \pi_i = 1/t \qquad (i = 1, 2, \ldots, t).$$

Из (4.3.1) соответствующая функция правдоподобия при равных n_{ij} есть

$$C \cdot 2^{-\frac{1}{2} nt(t-1)}.$$

Критерий для проверки H_0 против общей альтернативы о неравенстве π_i можно построить на отношении правдоподобия

$$\lambda = \frac{\prod\limits_{i<j} (p_i + p_j)^n}{2^{\frac{1}{2} nt(t-1)} \prod\limits_{i=1}^{t} p_i^{a_i}}. \qquad (4.4.1)$$

Брэдли и Терри [17] использовали вместо него монотонную функцию от λ, а именно

$$B_1 = n \sum_{i<j} \log_{10} (p_i + p_j) - \sum a_i \log_{10} p_i. \qquad (4.4.2)$$

После того как найдены м. п.-оценки p_i, можно вычислить B_1. Если это сделано для всех возможных исходов эксперимента, то можно построить распределение B_1 из известных вероятностей индивидуальных исходов, точно так же, как в § 2.3 было получено распределение $T_n = \Sigma (a_i - \bar{a})^2$. Положение иллюстрируется для случая $t = 4$, $n = 2$ в табл. 4.1. В наших предыдущих обозначениях первый вход таблицы есть размещение [6420]. Прочерки говорят о том, что соответствующие p_i неопределенны по причинам, указанным при обсуждении (4.3.7). Столбец P дает вероятность, с которой оценка B_1 больше или равна табличному значению, когда H_0 верна. Последний столбец, который мы добавили, дает значение T_2 (это также сумма квадратов Σr_i). Хотя упорядочение ранговых сумм согласуется с возрастающими значениями B_1, оно соответствует в данном случае и строго убывающим значениям T_2. Таким образом, две статистики вполне эквивалентны *как критерии* H_0. Это не всегда так, но «разногласие», если оно и есть, невелико. Конечно, охват таблиц Брэдли—Терри определяет местонахождение лишь определенных упорядочений рангов (таких, как 7, 7, 11, 11 в табл. 4.1), для которых сразу указывается соответствующее значение $P = 0{,}0290$. Для значений n за пределами таблиц, когда $t \leqslant 5$, и для всех меньших (t, n) комбинаций с $t > 5$ можно считать

$$-2 \log \lambda = nt(t-1) \log 2 - 2B_1 \log 10 \qquad (4.4.3)$$

Таблица 4.1

Возможные суммы рангов в случае $t=4$, $n=2$ и соответствующие p_i, B_1 и P
(из [17], с разрешения авторов и издателя)

Σr_1	Σr_2	Σr_3	Σr_4	p_1	p_2	p_3	p_4	B_1	P	T_2
6	8	10	12	1	—	—	—	0	0,0058	20
6	8	11	11	1	—	—	—	0,602	0,0232	18
6	9	9	12	1	—	—	—			
7	7	10	12	0,50	0,50	—	—			
7	7	11	11	0,50	0,50	—	—	1,204	0,0290	16
6	9	10	11	1	—	—	—	1,498	0,0994	14
7	8	9	12	0,59	0,28	0,13	—			
6	10	10	10	1	—	—	—	1,806	0,1190	12
8	8	8	12	0,33	0,33	0,33	—			
7	8	10	11	0,60	0,29	0,08	0,04	2,359	0,2245	10
7	9	9	11	0,62	0,17	0,17	0,05	2,631	0,3065	8
7	9	10	10	0,62	0,18	0,10	0,10	2,898	0,5643	6
8	8	9	11	0,37	0,37	0,21	0,06	3,158	0,6639	4
8	8	10	10	0,38	0,38	0,12	0,12			
8	9	9	10	0,40	0,23	0,23	0,14	3,389	0,9627	2
9	9	9	9	0,25	0,25	0,25	0,25	3,612	1,0000	0

распределенным приблизительно как χ^2 с $t-1$ степенями свободы при верной нуль-гипотезе. С другой стороны, использование статистики T_n из 2.1 ведет к очень похожим результатам при проверке H_0. Асимптотическая эквивалентность (при выполнении модели Брэдли и Терри) критериев, основанных на B_1 и T_n, которая была фактически показана Брэдли [16], для ненулевого случая так же справедлива, как и для нулевого. Он рассматривал альтернативную гипотезу

$$H_a : \pi_i = \frac{1}{t} + \frac{\delta_{in}}{\sqrt{n}} \qquad \left(i = 1, 2, \ldots, t; \ \sum_i \delta_{in} = 0\right),$$

где δ_{in} — последовательность констант, сходящихся к δ_i при $n \to \infty$. Как обычно, при построении предельной функции мощности H_a — это множество альтернатив, приближающихся к H_0 с ростом n. Предельное распределение при $n \to \infty$ статистики — $2\log\lambda$:

$$D_n = \frac{4}{nt} \sum (a_i - \bar{a})^2$$

есть нецентральное[1] χ^2 с $t-1$ степенями свободы и параметром нецентральности

$$\lambda' = \frac{1}{4} t^3 \sum \delta_i^2 .$$

[1] Общую теорию критериев отношения правдоподобия и нецентральных распределений см. в [92, гл. 24].

Таким образом, для больших n адекватное приближение для мощности получается в предположении о нецентральном χ^2-распределении с λ', равным

$$\frac{1}{4} nt^3 \sum (\pi_i - 1/t)^2.$$

Для данных значений π_i возможно поэтому применить нецентральное χ^2-распределение в обычной манере, для того чтобы получить наименьшее n, гарантирующее с предписанной вероятностью β, что H_0 будет отвергаться критерием с уровнем значимости α.

Пример

Пусть $\alpha = 0{,}05$, $\beta = 0{,}90$, $t = 4$, $\pi_1 = 0{,}4$, $\pi_2 = 0{,}3$, $\pi_3 = 0{,}2$, $\pi_4 = 0{,}1$.

Мы находим $\lambda' = 0{,}8\ n$. Заглянув в таблицы [113]* с $\alpha = 0{,}05$, $\beta = 0{,}90$, $\nu_1 = t - 1 = 3$, $\nu_2 = \infty$, мы видим, что $\varphi \equiv \sqrt{\lambda'/t} = 1{,}88$. Следовательно, наименьшая достаточная величина n есть 18.

Брэдли применил этот подход для сравнения асимптотической функции мощности этого критерия и другого, который он называет мультибиномиальным критерием. Последний является объединением $\frac{1}{2} t (t - 1)$ независимых биномиальных критериев, что можно сделать по таблице предпочтений с помощью статистики

$$4n \sum_{i<j} \left(p_{ij} - \frac{1}{2}\right)^2,$$

которая при H_a имеет асимптотическое нецентральное χ^2-распределение с $\frac{1}{2} t\ (t - 1)$ степенями свободы и параметром нецентральности λ'. Можно предсказать, что мультибиномиальный критерий, очевидно, здесь плох. Однако этот критерий интересен в тех случаях, когда неприменима линейная модель.

Таким образом, точка зрения о существенной близости всех линейных моделей в окрестности нуль-гипотезы опять (сравните 4.1) приводит к заключению, что простая D_n-статистика дает удовлетворительный критерий H_0 для большой выборки.

Асимптотическую функцию мощности D_n-критерия можно сравнить также с ее аналогом в дисперсионном анализе — F-критерием для сбалансированного неполноблочного эксперимента с двумя *измерениями*

---

* См. [158, с. 97 и табл. 4.10, с. 330], которая дает при 3 степенях свободы $Q = 0{,}05$, $\beta = \rho = 0{,}9$, $\lambda' = a = 14{,}172$, $n = \lambda'/0{,}8 = 17{,}72$. — *Прим. пер.*

на блок. При допущениях линейной модели ван Элтерен и Ноэзер [40] получили асимптотическую относительную эффективность*

$$8\sigma^2 \left[\int h^2(x)\,dx\right]^2,$$

которая не зависит от t и уменьшается до $2/\pi$ при нормальной плотности $h(x)$ с дисперсией σ^2.

Сочетание экспериментов

При использовании критерия (4.4.3) предполагается, что те же самые параметры π_i относятся ко всему эксперименту с n повторениями. В более общем случае могут существовать g однородных групп, причем в u-группе проводится n_u повторений ($\Sigma n_u = n$), для каждой группы рейтинги равны $\pi_i^{(u)} (i = 1, 2, \ldots, t; u = 1, 2, \ldots, g)$. Внутри такой группы применимы вышеприведенные процедуры, оценивание $\pi_i^{(u)}$ обеспечено так же, как и $B_1^{(u)}$, аналогично статистике B_1. Если $\lambda^{(c)}$ есть отношение правдоподобия для объединенного эксперимента, мы имеем вместо (4.4.3)

$$-2\log \lambda^{(c)} = nt(t-1)\log 2 - 2B_1^{(c)} \log 10, \qquad (4.4.4)$$

где

$$B_1^{(c)} = \sum_{u=1}^{g} B_1^{(u)} \log 10.$$

При нулевой гипотезе о полной случайности ($\pi_i^{(u)} = 1/t$ для всех i, u) — $2\log \lambda^{(c)}$ асимптотически распределена как χ^2 с $g(t-1)$ степенями свободы. В частном случае зависимости от некоторой априорной информации относительно возможных группировок разностей используются либо (4.4.4), либо (4.4.3).

Если интересна не полная случайность, а скорее отсутствие групповых различий ($\pi_i^{(u)} = \pi_i$ для всех u), то критерий для больших выборок получается с помощью выражения

$$-2\log(\lambda^{(c)}/\lambda) = -2(B_1^{(c)} - B_1)\log 10 \qquad (4.4.5)$$

табличной χ^2-распределенной случайной величины с $(g-1)(t-1)$ степенями свободы. Кажутся целесообразными некоторые предосторожности, так как распределение $-2\log(\lambda^{(c)}/\lambda)$ в конечной выборке зависит от мешающих параметров π_i.

Читатель отсылается к работам [17] и [145], в которых рассмотрены численные примеры и приводится дальнейшее обсуждение.

Эксперименты, где объектами были опыты в факторном эксперименте, изучались Абельсоном и Брэдли [1].

* Асимптотической эффективностью критериев T и S называется предел (если он существует и не зависит от α и β) отношения n_S/n_T, где n_T и n_S — числа наблюдений, требуемые T- и S-критериями соответственно для получения одной и той же мощности β при одних и тех же альтернативах, см., например, [181, с. 326] или [92, гл. 25]. — *Прим. пер.*

4.5. КРИТЕРИИ СОГЛАСИЯ МОДЕЛЕЙ

В практической ситуации редко бывает очевидно, которая из предшествующих линейных моделей — и бесконечного числа других, которые можно рассмотреть, — действительно пригодна. К счастью, есть множество общих соображений в пользу того, что шкалы предпочтения, конструируемые в соответствии с различными законами, стремятся в основном к согласию, если применима *какая-либо* линейная модель. Эмпирическое изучение этого вопроса см., например, в [85] и [110]. Вместо прямых сравнений шкал можно исследовать, насколько хорошо оценены определенные точки на шкале, т. е. v_i, которые служат для воспроизведения наблюдаемых частот предпочтения p_{ij}. Так, Тэрстоун [134] «восстановил» оценки p'_{ij} вероятностей предпочтения π_{ij} благодаря

$$p'_{ij} = H(v_i - v_j), \tag{4.5.1}$$

где H — функция распределения, используемая для получения v_i, и сравнивал p'_{ij} с p_{ij}. Эта идея полезна и широко применяется, хотя она не обеспечивает формального критерия того, как проверить удовлетворительность восстановления. Мостеллер [108c] взял ее как основание для аппроксимации критерия согласия χ^2 в модели Тэрстоуна—Мостеллера. Теоретические вопросы восстановления рассмотрены у Ноэзера [112].

Критерий согласия Мостеллера

Тригонометрическое преобразование (4.2.1) порождает случайные величины d_{ij}, которые для больших n приближенно нормальны с дисперсией $1/n$. Если соответственно мы определим

$$d'_{ij} = \sin^{-1}(2p'_{ij} - 1),$$

то[1]

$$X^2 = n \sum_{i<j} (d_{ij} - d'_{ij})^2,$$

как можно ожидать, имеет приближенное распределение χ^2, если проверяемая модель истинна. Поскольку d'_{ij} получено не вполне строгим методом, число степеней свободы несколько неопределенно, однако мы не ошибемся серьезно, положив число степеней свободы равным

$$\frac{1}{2}t(t-1) - (t-1) = \frac{1}{2}(t-1)(t-2).$$

Метод легко распространяется не только на нормальный случай, но и на другие линейные модели. Однако, как указал Мостеллер, есть тенденция к получению согласия, лучшего, чем оно есть на самом деле. Возможное объяснение состоит в том, что величины π_{ij} не постоянны

---

[1] Это идентично критерию Мостеллера. Углы d_{ij}, d'_{ij} измеряются в радианах.

во всех n повторениях эксперимента. Хорошо известно (сравните 3.1), что это приводит к уменьшению дисперсий p_{ij} в сравнении с постоянными π_{ij} и потому к уменьшению X^2. Другое объяснение предлагается в 4.6.

Критерий согласия Брэдли

Критерий для модели Брэдли—Терри можно получить прямо из метода максимума правдоподобия [14]. Для гипотезы

$$H_0 : \pi_{ij} = \pi_i/(\pi_i + \pi_j) \quad (i, j = 1, 2, \ldots, t; \; i \neq j)$$

функция правдоподобия дается (4.3.1), в то время как для альтернативы

$H_1 : \pi_{ij} \neq \pi_i/(\pi_i + \pi_j)$ для некоторых i и j она равна

$$C \prod_{i<j} \pi_{ij}^{\alpha_{ij}} \pi_{ji}^{\alpha_{ji}}.$$

Так как p_{ij} — м.п.-оценка π_{ij} при гипотезе H_1, критерий отношения правдоподобия для проверки H_0 против H_1 равен

$$\lambda = \frac{\prod_{i=1}^{t} p_i^{a_i}}{\prod_{i<j} (p_i + p_j)^n} \cdot \frac{1}{\prod_{i<j} p_{ij}^{\alpha_{ij}} p_{ji}^{\alpha_{ji}}}$$

или

$$-2 \log \lambda = 2 \left[\sum_{i \neq j} \alpha_{ij} \log p_{ij} - \Sigma a_i \log p_i + n \sum_{i<j} \log (p_i + p_j) \right] =$$

$$= 2 \left[\sum_{i \neq j} \alpha_{ij} \log p_{ij} + B_1 \log 10 \right]$$

с B_1, как в (4.4.2). H_0 отвергается на уровне значимости α, если $-2 \log \lambda$ превышает верхнее α%-ное значение χ^2-распределения с $\frac{1}{2}(t-1)(t-2)$ степенями свободы. Этот критерий, кажется, встретится с теми же практическими трудностями, что и критерий Мостеллера, который легко принимает модель.

4.6. ТРЕХКОМПОНЕНТНАЯ МОДЕЛЬ БОКА

Обсуждавшиеся до сих пор линейные модели не позволяют учитывать возможные различия между экспертами. Бок [11] предложил, чтобы наблюдаемая ценность объекта A_i в его сравнении с A_j экспертом γ была представлена в виде

$$y_{i\gamma(j)} = V_i + w_{i\gamma} + z_{i\gamma(j)} \tag{4.6.1}$$

$$(i, j = 1, 2, \ldots, t; \; i \neq j; \quad \gamma = 1, 2, \ldots, n),$$

где скобки, в которые взято j, указывают, что этот нижний индекс служит просто обозначением. Малые буквы обозначают случайные величины, $z_{i\gamma(j)}$ — взаимно независимые $N(0, \delta^2)$- величины. Особая черта этой модели заключается в $w_{i\gamma}$ компонентах, которые характеризуют определенные объекты и экспертов, независимо от выборки экспертов. Для одного и того же эксперта они коррелированы, $w_{i\gamma}$ и $w_{j\gamma}$ имеют двумерное нормальное распределение $N(0, 0, \zeta^2, \zeta^2, \rho)$. Это допущение равносильно (обычно $\rho \geqslant 0$) представлению $w_{i\gamma}$ в виде суммы нормально распределенного эффекта от экспертов и нормально распределенного эффекта взаимодействия. «Смешанная» природа модели отражает особый интерес к объектам, но эксперты рассматриваются лишь как выборка из бесконечной совокупности экспертов. Если мы положим $d_{ij\gamma} = y_{i\gamma(j)} - y_{j\gamma(i)}$, то заметим, что оно нормально распределено с математическим ожиданием $V_i - V_j$ и дисперсией $2\zeta^2(1-\rho) + 2\delta^2$, откуда следует, что модель линейна. Полагая дисперсию равной единице, мы получаем

$$\pi_{ij} = \Pr(d_{ij\gamma} > 0) = H(V_i - V_j),$$

где H — функция распределения единичной нормальной переменной, как в модели Тэрстоуна—Мостеллера. Может показаться, что это одна и та же модель, если сделана стандартизация. Однако в данном случае сравнения различны для общего объекта и выполняются не независимо для одного и того же эксперта с ковариацией

$$\operatorname{cov}(d_{ij\gamma}, d_{ik\gamma}) = \zeta^2(1-\rho) \quad (i \neq j \neq k).$$

В терминах характеристических случайных переменных $x_{ij\gamma}$ из § 1.3 мы имеем, например,

$$\Pr(x_{ij\gamma} = 1) = \pi_{ij}, \quad \Pr(x_{ij\gamma} = 0) = \pi_{ji}$$

и

$$\Pr(x_{ij\gamma} = 1, \ x_{ik\gamma} = 1) = \pi_{ij,ik}.$$

Поскольку $p_{ij} = \sum_{\gamma} x_{ij\gamma}/n$,

то

$$\mathscr{E}(p_{ij}) = \pi_{ij}, \quad \operatorname{var}(p_{ij}) = \pi_{ij}\pi_{ji}/n$$

и

$$\operatorname{cov}(p_{ij}, p_{ik}) = (\pi_{ij,ik} - \pi_{ij}\pi_{ik})/n.$$

Исходя из соображений, рассмотренных в 4.2, Бок применил тригонометрическое преобразование (4.2.1) к p_{ij}. Хотя это преобразование приближенно стабилизирует дисперсию для больших n, оно, как все такие преобразования, оставляет корреляцию неизменной. Она равна:

$$\operatorname{corr.}(d_{ij}, d_{ik}) \doteq \operatorname{corr.}(p_{ij}, p_{ik}) = \frac{\pi_{ij,ik} - \pi_{ij}\pi_{ik}}{\sqrt{\pi_{ij}\pi_{ji}\pi_{ik}\pi_{ki}}}.$$

Для того чтобы получить решение, с которым можно работать в трехкомпонентной модели, Бок предположил, что эта корреляция (ρ', скажем) одна и та же для всех i, j, k, — допущение, которое, как мы говорили, не будет хорошим в общем случае, но в случае, когда все π_{ij} близки к 1/2, приемлемо. Это так близко к нулевой гипотезе, что даже для больших n требуется чувствительный критерий.

Подчиняясь различным введенным приближениям, мы выразим d_{ij} в виде

$$d_{ij} = \alpha_i + \beta_j + e_{ij} \quad (i,\ j = 1,\ 2,\ \ldots,\ t), \tag{4.6.2}$$

где α_i — ценность объекта A_i на преобразованной шкале, $\beta_j = -\alpha_j$. Удобно положить $d_{ii} = 0$. Тогда случайная ошибка e_{ij} имеет следующие свойства:

$$e_{ij} = -e_{ji},\quad \mathscr{E}(e_{ij}) = 0,\quad \mathscr{E}(e_{ij}^2) = \sigma^2 = 1/n \tag{4.6.3}$$

и для i, j, k, l, не равных между собой

$$\mathscr{E}(e_{ij}e_{ik}) = \rho'\sigma^2,\quad \mathscr{E}(e_{ij}e_{kl}) = 0. \tag{4.6.4}$$

Если

$$d_i = \sum_{j=1}^{t} d_{ij},$$

то сумма квадратов между строками матрицы (d_{ij}) есть

$$S_\alpha = \Sigma d_i^2/t$$

и равна сумме квадратов между столбцами S_β. Из (4.6.2) мы имеем

$$S_\alpha = \frac{1}{t}\sum_i \left(t\alpha_i + \sum_j e_{ij}\right)^2,$$

так что с учетом (4.6.3) и (4.6.4)

$$\mathscr{E}(S_\alpha) = t\Sigma\alpha_i^2 + (t-1)\,\sigma^2 + (t-1)\,(t-2)\,\rho'\sigma^2.$$

Аналогично остаточная сумма квадратов

$$S_R = \Sigma d_{ij}^2 - S_\alpha - S_\beta$$

имеет математическое ожидание

$$(t-1)\,(t-2)\,\sigma^2 - 2\,(t-1)\,(t-2)\,\rho'\sigma^2.$$

Эти результаты могут быть сведены в таблицу дисперсионного анализа.

Источник вариации	Степени свободы	Сумма квадратов	Ожидаемый средний квадрат
Между объектами	$t-1$	$S_\alpha = \frac{1}{t}\,\Sigma d_i^2$	$(t-1)[1+(t-2)\,\rho']/n + t\Sigma\alpha_i^2$
Остаток	$\frac{1}{2}\,(t-1)(t-2)$	$\frac{1}{2}S_R = \frac{1}{2}S_T - S_\alpha$	$(t-1)(t-2)(1-2\rho')/(2n)$
Общий	$\frac{1}{2}\,t\,(t-1)$	$\frac{1}{2}\,S_T = \sum_{i<j} d_{ij}^2$	

За исключением приближения нормального распределения угловым, критерий согласия Мостеллера эквивалентен приближению nS_R распределением χ^2 с $\frac{1}{2}(t-1)(t-2)$ степенями свободы. Но для $\rho' > 0$ мы видим, что S_R имеет малое математическое ожидание. Модель Бока, таким образом, может использоваться для учета отклонений критерия Мостеллера.

Также можно заметить из таблицы, что при проверке эквивалентности объектов, пользуясь $n\Sigma d_i^2/t$ как χ^2 -распределенной случайной величиной с $(t-1)$ степенями свободы, мы можем отвергать гипотезу так же легко и при $\rho' > 0$.

Замечания

Наше рассмотрение важной модели Тэрстоуна—Мостеллера и других «случаев» (см. упражнение 1.5) отличается от того, которое было проведено Тэрстоуном [132] и вкратце обобщено Джонсом и Боком [86] и Торгерсоном [137]. Более специальные ссылки, видимо, можно извлечь из обсуждения случая неполных данных у Галликсена [63] и в работах Мостеллера [108b], Гибсона [52], Берроса и Гибсона [19].

5 ГЛАВА ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

5.1. ПОЛНОСТЬЮ СБАЛАНСИРОВАННЫЕ ПЛАНЫ

На языке планирования эксперимента то, что мы называем сбалансированным экспериментом парных сравнений, известно как «сбалансированный неполноблочный план» (BIBD) с блоком размером два. Задача планирования, конечно, остается одной и той же, имеем ли мы для каждого блока выражение предпочтения, как в парном сравнении, или два различных значения, как в обычной экспериментальной ситуации. Есть также тесная связь между сбалансированным экспериментом парных сравнений и круговым турниром, используемым во многих видах спорта, когда каждый из игроков (или команд) играет с каждым один или несколько раз. Планирование полностью сбалансированного эксперимента не представляет трудностей, однако может быть интересен, в частности, простой подход, данный Крейчиком [94]. Его процедуру можно рассмотреть на примере $t = 7$. Напишем числа 1—7 последовательно в семь строк по четыре $(= \frac{1}{2}(t+1))$. Получается четыре столбца цифр, напечатанных жирным шрифтом. Затем продолжим запись цифр в обратном порядке, начиная с правого нижнего угла и располагая при этом числа в три столбца, как показано цифрами, напечатанными обычным шрифтом.

(8) **1**	7 **2**	6 **3**	5 **4**
5 (8)	4 **6**	3 **7**	2 **1**
(8) **2**	1 **3**	7 **4**	6 **5**
6 (8)	5 **7**	4 **1**	3 **2**
(8) **3**	2 **4**	1 **5**	7 **6**
7 (8)	6 **1**	5 **2**	4 **3**
(8) **4**	3 **5**	2 **6**	1 **7**

Результат — расположение семи игроков в круговом турнире, требующем 7 туров. Игроки из первого столбца свободны от игры. Заметим, что каждый игрок появляется первым в трех из его 6 игр, что соответствует проведению 3 игр дома или белыми фигурами и т. д. С точки зрения эксперимента парных сравнений мы имеем, исключая объекты первого столбца, план, в котором любой эффект порядка представления (объектов) сбалансирован и учтены другие предосторожности. Этот процесс возможен для любых нечетных t. Если t четное, скажем, $t=8$, мы просто ставим восьмого игрока в пару с игроком, свободным от игры, записывая 8 поочередно то слева, то справа от первого столбца, как показано. Влияние порядка представления в этом случае сбалансировано лишь приближенно; полный баланс будет достигнут повторением всего эксперимента с изменением порядка в каждой паре на противоположный и строками, которые перерандомизированы. Другой простой подход см. в [50].

Метод, который дополнительно дает наибольшее возможное разделение между парами, имеющими общий объект, был дан Россом [121]. Для нечетных t выпишем пары следующим образом:

12	$t3$	$t-1,4$	$t-2,5$	...	$\frac{1}{2}(t+5), \frac{1}{2}(t+1)$
$1, \frac{1}{2}(t+3)$	23	$t4$	$t-1,5$	...	$\frac{1}{2}(t+5), \frac{1}{2}(t+3)$
13	24	$t5$	$t-1,6$	...	$\frac{1}{2}(t+7), \frac{1}{2}(t+3)$
$1, \frac{1}{2}(t+5)$	34	25	$t6$	...	$\frac{1}{2}(t+7), \frac{1}{2}(t+5)$
		...			
$1, \frac{1}{2}(t+1)$	$\frac{1}{2}(t-1), \frac{1}{2}(t+3)$	$\frac{1}{2}(t-3), \frac{1}{2}(t+5)$	$\frac{1}{2}(t-5), \frac{1}{2}(t+7)$	...	$2, t-1$
$1t$	$\frac{1}{2}(t+1), \frac{1}{2}(t+3)$	$\frac{1}{2}(t-1), \frac{1}{2}(t+5)$	$\frac{1}{2}(t-3), \frac{1}{2}(t+7)$	...	$2, t$

Для $t=7$ мы имеем после уравновешивания следующий порядок представления: 12, 73, 64, 51, 32, 47, 56, 13, 24, 75, 61, 43, 52, 67, 14, 35, 26, 71, 45, 36, 27. Случай $t=6$ может быть получен из вышеприведенного (без полного баланса) путем пропуска всех пар, содержащих 7. Росс табулировал сбалансированные схемы для нечетных $t \leqslant 17$.

Планы парных турниров также рассматривались — Крейчиком и позднее Шайдом [124]. Джилберт [53] изучал турниры смешанных пар. Парные турниры интересны для работы с парными сравнениями, когда мы хотим сравнить одну пару объектов с другой парой. Например, при создании цветовой шкалы наблюдателей просят сравнить *различия* (является ли различие между A_1 и A_2 бо̀льшим, чем различие между A_3 и A_4?). Смешанно-парная ситуация возникает, когда два объекта составляют одну «команду» при сравнениях пар различных объектов (предпочитается ли блюдо A_1 с напитком A_2 блюду A_3 с напитком A_4?). Однако аналогия неполная, так как игрок может быть лишь в одной команде, тогда как с объектами можно проводить также сравнения типа (A_1, A_2) против (A_1, A_3).

5.2. НЕПОЛНЫЕ ПЛАНЫ

Практические трудности в использовании экспериментов парных сравнений возникают при большом числе объектов, для которых требуется большое число сравнений. Мы можем сочувствовать тем мастерам, которым МакКормик и др. ([100], [101]) представили колоду в 1225 карточек, каждая из которых содержала некоторую пару из 50 работающих под их началом лиц, и для каждой пары требовали выразить предпочтение. Эти авторы эмпирически исследовали влияние пропуска различных долей сравнений, используя для этой цели неполные планы циклического типа, как показано, например, на рис. 5.1.

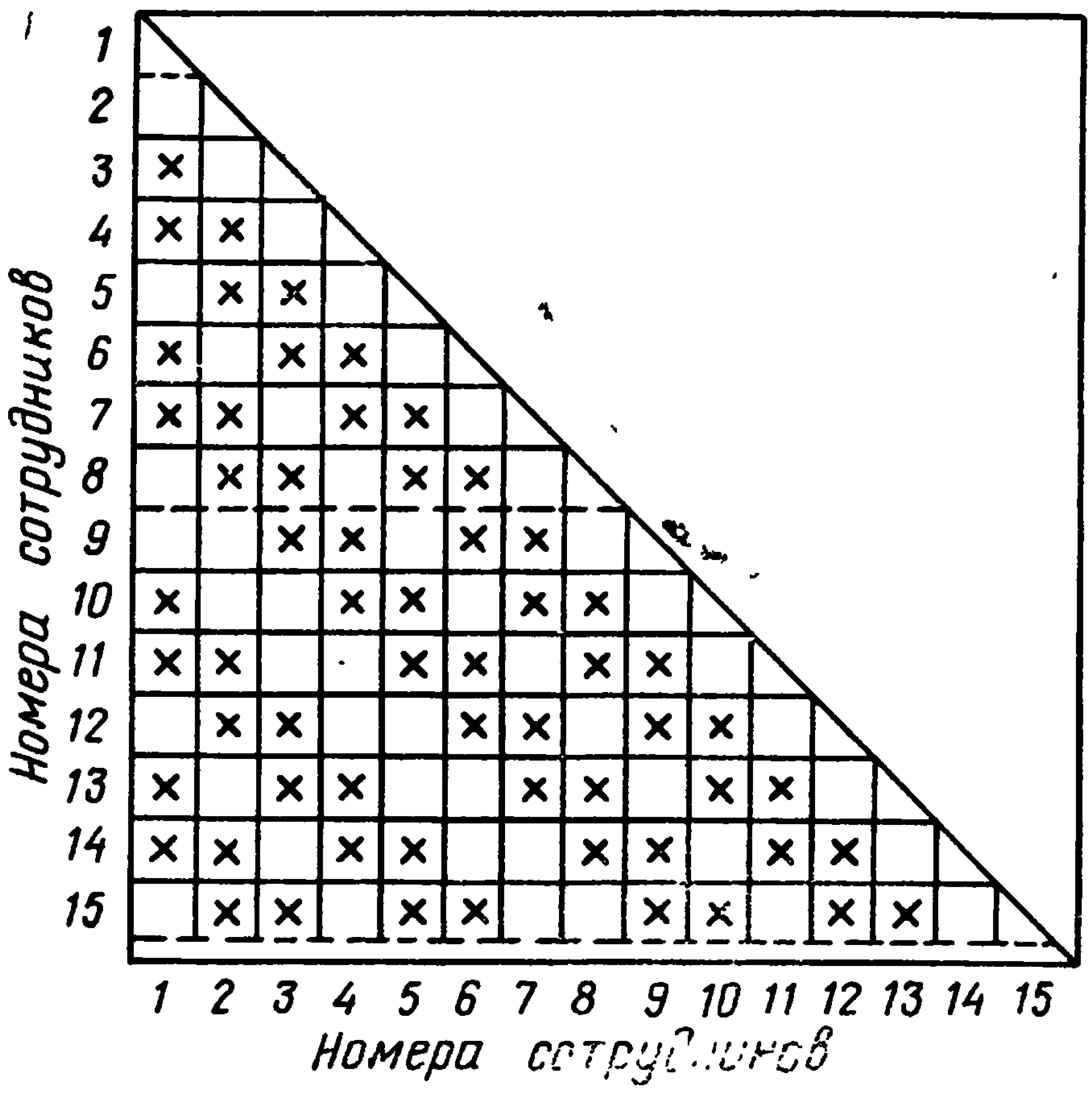


Рис. 5.1. Пример циклической схемы; × — сравниваемые пары

Этот пример ясен и соответствует пропуску 3/7 общего числа сравнений. Здесь, при аттестации персонала, читателя может заинтересовать пригодность парных сравнений. Иногда, если у мастера хорошая память, индивидуальные суждения не вполне независимы, но при столь большом числе сотрудников это не так уж важно при правильной рандомизации порядка сравнения пар. Почему не произвести непосредственное ранжирование? При большом числе объектов это непрактично и в любом случае не дает объяснения способности мастеров выносить суждения. Однако можно разбить сравнения на группы, например, по пять, а не по два. Снова выделим $\binom{50}{5}$ возможных блоков по пять, что приводит к BIB-планам. Было бы ошибочным заменить ранжирование внутри блока на десять (непротиворечивых транзитивных) парных сравнений, получаемых из ранжирования, как это иногда предлагают. Подходящий метод анализа дал Дарбин [34], который предложил использовать планы с квадратами Юдена (т. е. с неполными латин-

— 1 множеств из t и одно из $\frac{1}{2}t$ пар. Ведущее множество, которое можно обозначить $\{j, j+1\}$, всегда будет удовлетворять (а) и (б) так же, как и s-е множество $\{j, j+s\}$, при условии, что s и t взаимно просты ($s < \frac{1}{2}t$). Если s и t имеют наибольший общий множитель f, то s-е множество разбивается на f подмножеств. (I) или любое другое связанное множество можно объединять с другими множествами в план, удовлетворяющий (а) и (б). Для заданного числа пар, оцениваемых экспертами, мы будем интуитивно предпочитать «более связанный» план из тех, что можно построить с минимумом повторений.

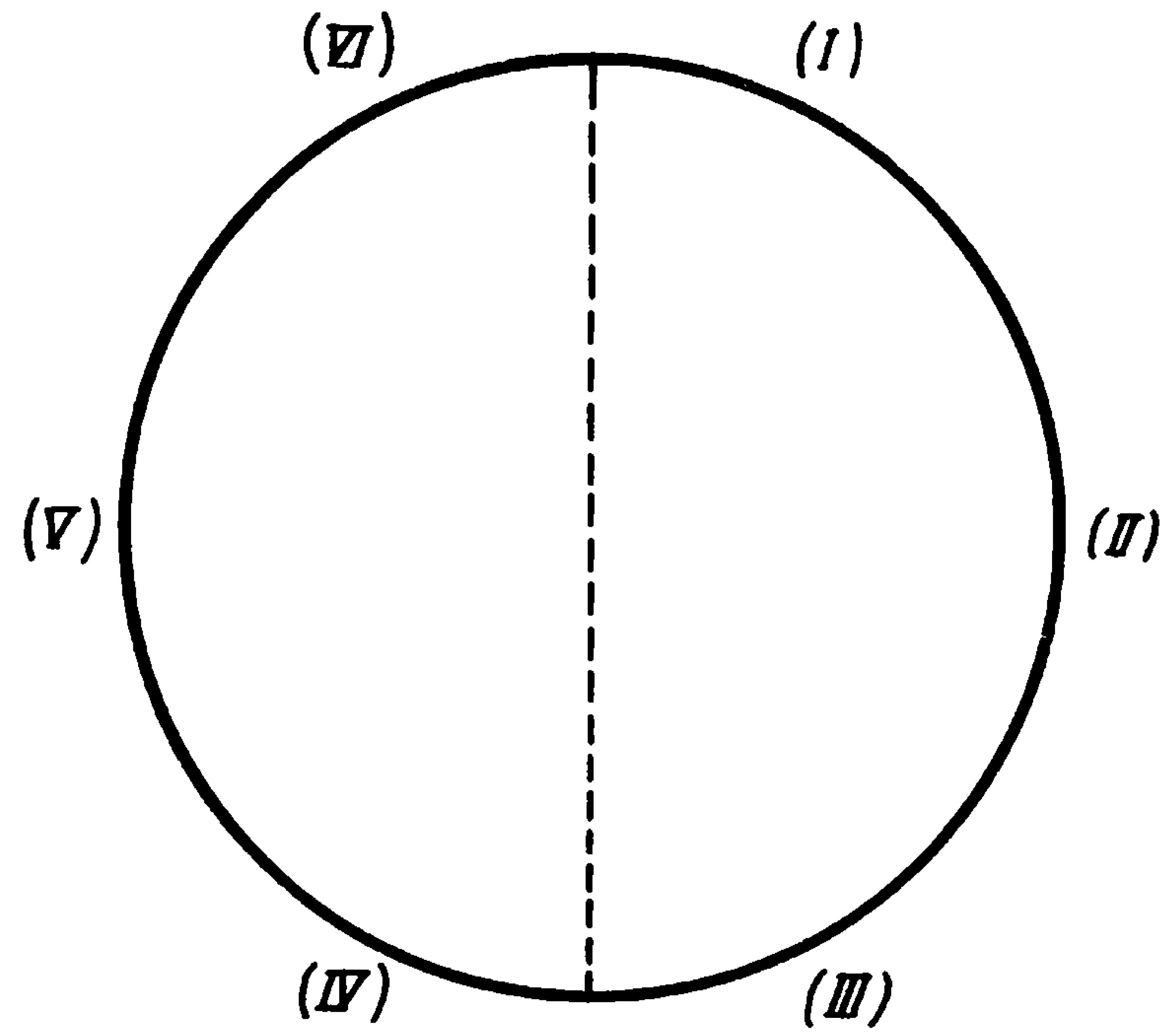


Рис. 5.2.

Вернемся к случаю $t = 7$. Множества (I)—(III) можно сделать сбалансированными и связанными. Ясно, что в отдельности они одинаково хороши, но если выбирать два из них, то будет ли какая-нибудь пара предпочтительнее других? Чтобы увидеть, что ответ отрицателен, расположим три множества, вместе с множествами (IV)—(VI), по кругу, как показано на рис. 5.2. Множество (IV) есть просто

15 26 37 41 52 63 74

и его можно назвать зеркальным отображением (III), которое эквивалентно (III), за исключением смены порядка в каждой паре. Аналогично (V) и (VI) есть зеркальные отображения (II) и (I) соответственно. Мы будем пренебрегать эффектом порядка, так что (i) и ($t - i$) эквивалентны (i = I, II, ..., $\frac{1}{2}(t-1)$; t нечетное). Рассмотрим теперь перенумерацию R объектов (1 2 3 4 5 6 7) в (1 3 5 7 2 4 6)[1]. Она превращает (I) в (II), (II) в (III) и (III) в (I), как легко проверить. Так, пары (I, II) заменяются на (II, III) и (II, III) заменяются на (III, I), чем демонстрируют, по сути дела, эквивалентность всех трех пар.

Наши соображения в пользу введения множеств (IV)—(VI) состоят в обобщении на другие, нечетные значения t. Для i = I, II,..., $\frac{1}{2}(t-1)$ преобразование R дает эффект замены множества (i) на ($2i$), где последнее множество должно быть делимо по модулю t и для эквивалентностей. Так, для $t = 7$ (I, II, III) заменяется на (II, IV, VI), что эквивалентно (II, III, I). Для $t = 9$ связанные множества есть (I, II, IV); R преобразует их в (II, IV, I), так что все пары связанных множеств эквивалентны. Действуя таким образом, можно выбрать

[1] Можно рассмотреть и другие пути перенумерации.

приемлемые комбинации подмножеств для любого t; подобные рассмотрения проводятся и для четного t, но мы не станем систематизировать их здесь[1]. Смотрите также [88] и [102].

*Частично сбалансированные неполноблочные планы**

Отношение циклических планов к частично сбалансированным неполноблочным планам с двумя ассоциативными классами представляет некоторый интерес. В нашем случае, когда каждый блок состоит из двух объектов, PBIB схемы определяются следующим образом:

а) существует t объектов и b блоков;

б) каждый из t объектов встречается r раз (так что $tr = 2b$);

в) каждая пара объектов встречается либо λ_1, либо λ_2 раз (и объекты называют соответственно или первично связанными, или вторично связанными);

г) существует соотношение ассоциативности между каждой парой объектов, удовлетворяющее следующим условиям:

1) каждый объект имеет n_1 первичных и n_2 вторичных связей;

2) для любых двух объектов, которые i-связаны, число объектов, j-связанных с первым и k-связанных со вторым, обозначается p^i_{jk}, и это число не зависит от пары объектов, с которой мы начали. Более того, $p^i_{jk} = p^i_{kj}$ $(i,\ j,\ k = 1,\ 2)$.

Клэтуорси [23] дал перечисление всех таких планов, для которых каждый объект повторяется не более 10 раз. Эти планы имеют высокую степень симметрии и потому весьма удобны для неполноблочных экспериментов парных сравнений. Можно различать четыре основных типа планов, один из которых циклический тип. Для $t = 5$ легко проверить, что каждое из множеств (I) и (II), выписанных выше, и любые комбинации из них, которые не полностью сбалансированы, образуют PBIB-план с двумя ассоциативными классами. Так, для (I) и (II) мы имеем $t = 5$, $b = 5$, $r = 2$, $\lambda_1 = 1$, $\lambda_2 = 0$, $n_1 = 2$, $n_2 = 2$; в (I) у объекта 1 соответственно 2 и 5 — первично связанные, 3 и 4 — вторично связанные объекты и т. д. Также

$$p^1_{jk} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad p^2_{jk} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Однако это лишь частный случай (отличный от $t = 3$), когда данное множество в такой высокой степени сбалансировано. Несколько другие циклические планы, полученные Клэтуорси, являются комбинациями множеств (II, V, VI) для $t = 13$ и множеств (III, V, VI, VII) для $t = 17$. Конечно, эквивалентные комбинации можно получить, меняя номера объектов.

Мы рассмотрим сейчас другие PBIB-планы с двумя ассоциативными классами:

---

[1] Такие исследования сейчас ведутся [29].

* См. [185], [186], [248]. — *Прим. пер.*

(1) *Планы с делением на группы*

Если t выразимо как произведение двух бо́льших, чем 1, чисел m, n', то объекты можно разделить на m групп размера n'. Пары встречаются λ_1 или λ_2 раз, в зависимости от того, принадлежат ли они одной и той же группе. Список всех таких планов для $r \leqslant 10$ и $\lambda_1 = 0$, $\lambda_2 = 1$ приведен в табл. 5.1 (сравните с [37]). При построении плана 2, например, мы можем взять объекты 1, 2, 3 в качестве одной группы, а 4, 5, 6 — как другую. Так как пары внутри групп не образуются, эксперимент состоит из блоков

14 15 16 24 25 26 34 35 36.

Планы с различными значениями λ можно построить из базисных[1] планов повторением приведенных случаев или комбинацией их с полными планами.

(2) *Треугольные планы*

Если $t = \frac{1}{2} n' (n' - 1)$ ассоциативная схема, то можно получить таблицу, размещая объекты по порядку и симметрично относительно вычеркнутой главной диагонали. Например, для $t=6$, $n'=4$ схема такова:

*	1	2	3
1	*	4	5
2	4	*	6
3	5	6	*

Таблица 5.1

Планы с делением на группы для $r \leqslant 10$, $\lambda_1 = 0$, $\lambda_2 = 1$
(из [23] с разрешения автора)

План №	Обработки (объекты) t	Блоки b	Число повторений r	Группы m	Число объектов в группе n'
1	4	4	2	2	2
2	6	9	3	2	3
3	6	12	4	3	2
4	8	16	4	2	4
5	8	24	6	4	2
6	9	27	6	3	3
7	10	25	5	2	5
8	10	40	8	5	2
9	12	36	6	2	6
10	12	48	8	3	4
11	12	54	9	4	3
12	12	60	10	6	2
13	14	49	7	2	7
14	15	75	10	3	5
15	16	64	8	2	8
16	18	81	9	2	9
17	20	100	10	2	10

[1] В этом контексте «базисные» можно интерпретировать как эквивалент условия $\lambda_1 + \lambda_2 = 1$.

Два объекта, находящихся в одной строке (или столбце), первично связаны, тогда как объекты, не лежащие в одной строке, связаны вторично. Для $\lambda_1 = 1$, $\lambda_2 = 0$ мы получаем план

$$\begin{matrix} 12 & 13 & 23 & 14 & 15 & 45 \\ 24 & 26 & 46 & 35 & 36 & 56, \end{matrix}$$

который, будучи «перевернут», эквивалентен плану 3 (упражнение 5.1). Другие базисные треугольные планы даны в табл. 5.2 и состоят из всех различных групп с делением.

Таблица 5.2

Треугольные планы для $r \leqslant 10$, $\lambda_1 + \lambda_2 = 1$
(из [23], с разрешения автора)

План №	Обработки (объекты) t	Блоки b	Повторения r	Размер таблицы n'	λ_1	λ_2
3	6	12	4	4	1	0
18	10	15	3	5	0	1
19	10	30	6	5	1	0
20	15	45	6	6	0	1
21	15	60	8	6	1	0
22	21	105	10	7	1	0
23	21	105	10	7	0	1

(3) *Квадратные планы*

Если t есть квадрат целого числа $s > 1$, ассоциативную схему можно получить, располагая объекты по порядку в квадрате $s \times s$. Так, для $t = 9$ мы имеем

$$\begin{matrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9. \end{matrix}$$

Здесь объекты, находящиеся в одной строке или в одном столбце, связаны первично, остальные—вторично. При $\lambda_1 = 1$, $\lambda_2 = 0$ мы получаем план 24 табл. 5.3, а именно

$$\begin{matrix} 12 & 13 & 14 & 17 & 23 & 25 & 28 & 36 & 39 \\ 45 & 46 & 47 & 56 & 58 & 69 & 78 & 79 & 89. \end{matrix}$$

Другие базисные квадратные планы даны в табл. 5.3.

Таблица 5.3

Квадратные планы для $r \leqslant 10$, $\lambda_1 + \lambda_2 = 1$
(из [23], с разрешения автора)

План №	Обработки (объекты) t	Блоки b	Повторения r	Размер квадрата s	λ_1	λ_2
1	4	4	2	2	1	0
24	9	18	4	3	1	0
25	16	48	6	4	1	0
26	16	72	9	4	0	1
27	25	100	8	5	1	0
28	36	180	10	6	1	0

Заканчивая наше описание базисных планов, мы можем добавить номера 29—31, соответствующие трем циклическим схемам для $t = 5$, 13, 17, которые упоминались выше. Клэтуорси дал некоторые другие планы (включающие 16 и 27 объектов), которые не относятся к четырем главным классам.

5.3. ПЛАНЫ ПАРНЫХ СРАВНЕНИЙ СО СВЯЗЯМИ (ЦЕПНЫЕ)

Если есть несколько экспертов и одна из целей эксперимента парных сравнений состоит в измерении согласия между ними, то естественно требовать, чтобы эксперимент был сбалансирован по экспертам в некотором смысле так же хорошо, как и по объектам. А это значит, что каждая пара экспертов должна иметь некоторые общие сравнения. Рассмотрим, например, случай 4 объектов, когда трудно делать сравнения. Мы можем взять план 1 из табл. 5.1 и заметить, что для PBIB-плана нужны 4 сравнения с двумя ассоциативными классами. Для получения баланса нам надо положить $n = 3$ (или кратным трем) экспертам и можно разместить сравнения следующим образом:

$$\begin{array}{lcccc} \text{эксперт } a & 13 & 14 & 23 & 24, \\ \text{эксперт } b & 12 & 14 & 23 & 34, \\ \text{эксперт } c & 12 & 13 & 24 & 34, \end{array} \qquad (5.3.1)$$

что немедленно получается из трех групп размером 2×2:

12, 34 13, 24 14, 23*

В результате получается эксперимент, в котором каждый эксперт имеет дело с высокосбалансированным множеством и каждая пара экспертов имеет по два общих сравнения. Однако требования к парам, сравниваемым одним экспертом, смягчаются в общем случае ради достижения полного баланса.

Бозе [12] рассмотрел этот предмет систематически, и приведенный выше план — в действительности первый из его планов, воспроизведенных в табл. 5.4. Пусть r' — число пар объектов, сравниваемых каждым экспертом. Для обеспечения симметрии между объектами и экспертами Бозе ввел *планы со связями*, требующие, чтобы:

а) среди r' пар, сравниваемых каждым из n экспертов, каждый объект появлялся одинаково часто, скажем, α раз;

б) каждая пара сравнивалась k экспертами, $k > 1$;

в) для любых двух данных экспертов было в точности λ пар, которые сравниваются ими обоими.

Так, для эксперимента (5.3.1) мы имеем:

$$t = 4,\ n = 3,\ r' = 4,\ \alpha = 2,\ k = 2,\ \lambda = 2,\ b = 6,$$

где $b = \frac{1}{2} t (t - 1)$ есть число возможных пар.

---

* Из которых берутся все $\binom{3}{2}$ сочетания, каждое из которых дает по 4 сравнения — строку в (5.3.1). — *Прим. пер.*

В общем случае ясно, что

$$r' = \frac{1}{2} t\alpha. \qquad (5.3.2)$$

Есть, видимо, некоторое неуловимое соответствие между планами парного сравнения со связями и сбалансированными неполноблочными планами. Можно рассматривать соответствия между экспертами и обработками, между парами объектов и блоками. Если пара сравнивается с экспертом, то блок, соответствующий паре, можно рассматривать как содержащий обработку, соответствующую эксперту. Рис. 5.3 иллюстрирует соответствие для (5.3.1).

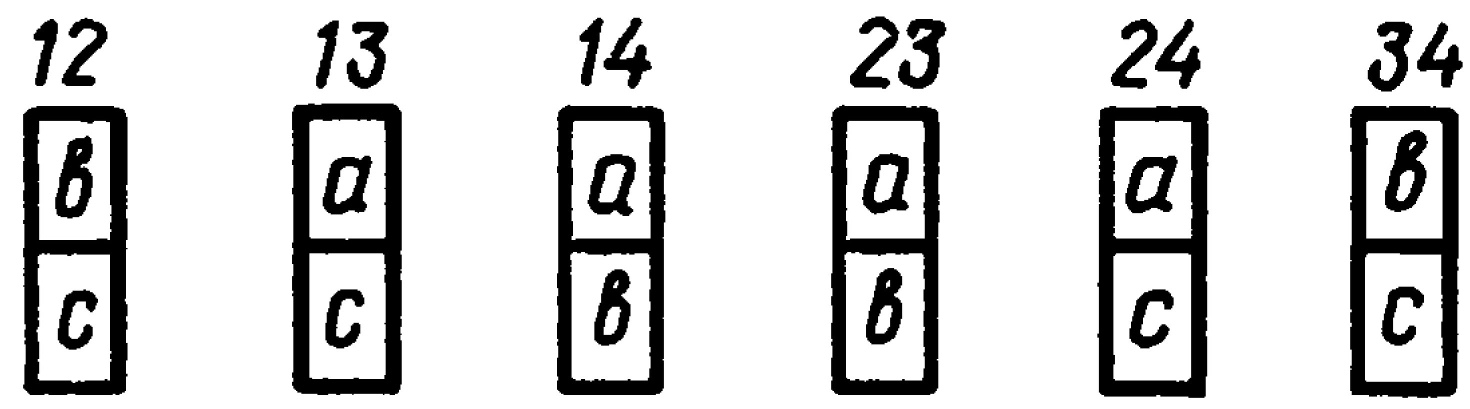


Рис. 5.3.

Отсюда, если план парных сравнений рассмотренного типа существует, должен существовать и соответствующий BIB-план с n обработками и b блоками, так что каждый блок содержит k обработок; каждая обработка встречается в r' блоках и две данные обработки встречаются в λ блоках. Из этого следует, что

$$bk = nr', \quad \lambda (n - 1) = r' (k - 1). \qquad (5.3.3)$$

Подставляя b и r', мы можем написать первое из этих уравнений как

$$k (t - 1) = n\alpha.$$

Надо помнить, что существование BIB-плана с параметрами n, $b = \frac{1}{2} t (t - 1)$, r', k, λ не обеспечивает существование соответствующего плана парных сравнений со связями, что обусловлено дополнительным ограничением (а). По поводу построения планов 1—3 табл. 5.4 (и большего плана для $t = 9$) мы отсылаем читателя к работе Бозе. Мы приведем теперь его весьма общий метод получения планов парных сравнений со связями. Сначала предположим, что число объектов четно, скажем, $t = 2i$. Тогда мы можем разделить $i (2i - 1)$ пары на $2i - 1$ множества по i пар в каждом, так что каждый объект встречается в точности один раз среди пар из некоторого множества.

Например, если $t = 6$, мы можем взять в качестве объектов 0, 1, 2, 3, 4, ∞. Тогда 5 наборов[1] таковы:

I	14	23	0∞
II	20	34	1∞
III	31	40	2∞
IV	42	01	3∞
V	53	12	4∞

[1] В отличие от приведенных множеств в скобках (I), (II) ... эти множества, вообще говоря, не циклические.

В общем случае $(2i - 1)$ набора могут быть получены разложением по mod $(2i - 1)$ начального набора

$$1{,}2i - 2 \quad 2{,}2i - 3, \ldots, \; i - 1, \; i \quad 0\,\infty,$$

объект ∞ остается неизменным.

Возьмем сейчас BIB-план с n_1 обработками, $b_1 = 2i - 1$ блоками, r_1 повторениями, блок размера k_1, в котором каждая пара обработок встречается вместе в любом блоке λ_1 раз; поставим в соответствие каждому блоку один набор и каждой обработке одного эксперта. Мы можем тогда получить план парных сравнений со связями, назначая каждому эксперту набор пар, соответствующих всем блокам, в которых обработка соответствует эксперту. Мы получим этим способом план парных сравнений с параметрами:

$$t = 2i, \; n = n_1, \; b = i\,(2i - 1), \; r' = ir_1, \; k = k_1,$$
$$\lambda = i\lambda_1, \; \alpha = r_1. \tag{5.3.4}$$

В случае $t = 6$ мы можем начать с BIB-плана, имеющего параметры:

$$n_1 = 5, \; b_1 = 5, \; r_1 = 1, \; k_1 = 4, \; \lambda_1 = 3.$$

Таблица 5.4

Планы парных сравнений со связями
(из [12], с разрешения автора и издателя)

№	Параметры	План	
		эксперт	пары, назначаемые эксперту
1	2	3	4
1	$t=4$, $n=3$, $b=6$, $r'=4$, $k=2$, $\lambda=2$, $\alpha=2$	*a*	14 13 24 23
		b	13 24 12 34
		c	14 12 23 34
2	$t=5$, $n=6$, $b=10$, $r'=5$, $k=3$, $\lambda=2$, $\alpha=2$	*a*	35 24 13 14 25
		b	23 34 14 15 25
		c	23 35 12 45 14
		d	35 12 34 24 15
		e	12 34 45 13 25
		f	23 45 24 13 15
3	$t=6$, $n=10$, $b=15$, $r'=6$, $k=4$, $\lambda=2$, $\alpha=2$	*a*	12 13 23 45 46 56
		b	23 24 34 15 16 56
		c	13 14 34 25 26 56
		d	14 24 12 35 36 56
		e	14 16 46 23 25 35
		f	13 15 35 24 26 46
		g	12 15 25 34 36 46
		h	14 15 45 23 26 36
		j	13 16 36 24 25 45
		k	12 16 26 34 35 45

№	Параметры	Множество пар	План	
			эксперт	множество пар, назначаемых эксперту
1	2	3	4	5
4	$t=6$, $n=5$, $b=15$, $r'=12$, $k=4$, $\lambda=9$, $\alpha=4$	I 14 23 56 II 25 34 16 III 31 45 26 IV 42 51 36 V 53 12 46	*a* *b* *c* *d* *e*	II III IV V I III IV V I II IV V I II III V I II III IV
5	$t=7$, $n=3$, $b=21$, $r'=14$, $k=2$, $\lambda=7$, $\alpha=4$	I 12 23 34 45 56 67 71 II 13 24 35 46 57 61 72 III 14 25 36 47 51 62 73	*a* *b* *c*	II III III I I II
6	$t=8$, $n=7$, $b=28$, $r'=12$, $k=3$, $\lambda=4$, $\alpha=3$	I 16 25 34 78 II 27 36 45 18 III 31 47 56 28 IV 42 51 67 38 V 53 62 71 48 VI 64 73 12 58 VII 75 14 23 68	*a* *b* *c* *d* *e* *f* *g*	I V VII II VI I III VII II IV I III V II IV VI III V VII IV VI
7	$t=8$, $n=7$, $b=28$, $r'=16$, $k=4$, $\lambda=8$, $\alpha=4$	как для № 6	*a* *b* *c* *d* *e* *f* *g*	III V VI VII IV VI VII I V VII I II VI I II III VII II III IV I III IV V II IV V VI

Если обработки *a*, *b*, *c*, *d*, *e*, то блоки

a	*b*	*c*	*d*
b	*c*	*d*	*e*
c	*d*	*e*	*a*
d	*e*	*a*	*b*
e	*a*	*b*	*c*

Так как встретились соответственно 1, 3, 4, 5, то мы назначаем множества I, III, IV, V эксперту *a* и получаем следующий план парных сравнений со связями:

Эксперт	Множества
a	I III IV V
b	I II IV V
c	I II III V
d	I II III IV
e	II III IV V

Параметры плана по (5.3.4):

$$t = 6,\ n = 5,\ b = 15,\ r' = 12,\ k = 4,\ \lambda = 9,\ \alpha = 4.$$

Это план 4 из табл. 5.4 с некоторым изменением номеров объектов.

Если число объектов нечетно, скажем, $t = 2i + 1$, то $i\,(2i + 1)$ пар могут быть разделены на i циклических множеств по $2i + 1$ пар каждое, как написано в 5.2. Для $t = 7$ эти множества таковы:

(I)	12	23	34	45	56	67	71
(II)	13	24	35	46	57	61	72
(III)	14	25	36	47	51	62	73

Возьмем теперь BIB-план с параметрами n_2, $b_2 = i$, r_2, k_2, λ_2. Посту-

Таблица 5.5

Планы парных сравнений со связями для $t \leqslant 20$, полученные с помощью (5.3.4) или (5.3.5) из BIB-планов по [24]

t	n	b	r'	k	λ	α	План
1	2	3	4	5	6	7	8
4	3	6	4	2	2	2	3×2*
6	5	15	12	4	9	4	5×4*
8	7	28	12	3	4	3	11.7
8	7	28	16	4	8	4	11.8
8	7	28	24	6	20	6	7×6*
9	4	36	27	3	18	6	4×3*
10	9	45	40	8	35	8	9×8*
11	5	55	44	4	33	8	5×4*
12	11	66	30	5	12	5	11.19
12	11	66	36	6	18	6	11.20
12	11	66	60	10	54	10	11.10*
13	4	78	39	2	13	6	11.1
13	6	78	65	5	52	10	6×5*
14	13	91	28	4	7	4	11.22
14	13	91	63	9	42	9	11.23
15	7	105	45	3	15	6	11.7
15	7	105	60	4	30	8	11.8
15	7	105	90	6	75	12	7×6*
16	6	120	40	2	8	5	11.3
16	6	120	80	4	48	10	11.6
16	10	120	48	4	16	6	11.16
16	10	120	72	6	40	9	11.18
16	15	120	56	7	24	7	11.25
16	15	120	64	8	32	8	11.26
17	8	136	119	7	102	14	8×7*
19	9	171	152	8	133	16	9×8*
20	19	190	90	9	40	9	11.31
20	19	190	100	10	50	10	11.32

* BIB-планы состоят из всех возможных комбинаций n групп по k.

пая как прежде, мы получим план парных сравнений со связями с параметрами:

$$t = 2i + 1,\ n = n_2,\ b = i\,(2i + 1),\ r' = (2i + 1)\, r_2,\ k = k_2,$$
$$\lambda = (2i + 1)\,\lambda_2,\quad \alpha = 2r_2. \tag{5.3.5}$$

В случае $t = 7$ BIB-план состоит из трех сравнений трех объектов, даваемых планом 5 табл. 5.4. Подводя итог, мы видим, что каждый BIB-план может быть превращен в план парных сравнений со связями: если число блоков в первом четное, мы можем использовать (5.3.5); если это число нечетное, мы можем выбирать между (5.3.4) и (5.3.5). Из списка BIB-планов (например, [46, табл. XVIII] или [24, табл. 9.5]) можно получить много планов парных сравнений со связями. Планы 1 и 4—7 табл. 5.4 — лишь некоторые из тех, что могут быть получены этим способом (упражнение 5.3). Каталог для $t \leqslant 20$ дается в табл. 5.5.

Некоторые из этих планов рассматривались Кендэлом [89], который также обсуждал, насколько хорошо сбалансированы размещения.

УПРАЖНЕНИЯ

5.1. Покажите, что планы 1 и 3 табл. 5.1 можно рассматривать так же, как квадратный и треугольный соответственно.

5.2. Пусть

$$n_{ij\gamma} = \begin{cases} 1, & \text{если } A_i \text{ и } A_j \text{ сравниваются экспертом } \gamma, \\ 0 & \text{в противном случае (включая } n_{ij\gamma} = 0) \end{cases}$$
$$(i,\ j = 1,\ 2,\ \ldots,\ t;\ i \neq j;\ \gamma = 1,\ 2,\ \ldots,\ n).$$

Тогда $N_\gamma = (n_{ij\gamma})$ можно называть *матрицей инциденций* для эксперта γ.

Покажите, что необходимым и достаточным условием для того, чтобы матрица N_γ была матрицей для плана парных сравнений со связями, служат:

а) $N_\gamma E = \alpha E \quad (\gamma = 1,\ 2,\ \ldots,\ n)$;

б) $\sum\limits_{\gamma=0}^{n} N_\gamma = kE$;

в)* $\operatorname{sp}(N_\gamma N'_{\gamma'}) = 2\lambda \quad (\gamma, \gamma' = 1, 2, \ldots, n;\ \gamma \neq \gamma')$,

где E — квадратная матрица порядка t, каждый элемент которой равен 1, $N_0 = kI$, I — единичная матрица порядка t [145].

5.3. Покажите, как план 1 может быть получен из общего подхода § 5.3. Так же можно получить схемы 7 и 8 из планов 11.7 и 11.8 в [24].

---

* *sp* ( ) — обозначение следа матрицы. — *Прим. пер.*

6 ГЛАВА

ВЫБОР НАИЛУЧШЕГО ОБЪЕКТА

6.1. КРУГОВЫЕ ТУРНИРЫ

Мы всегда говорим об аналогии между сбалансированным экспериментом парных сравнений, включающим t объектов, и круговым турниром t игроков. В этой главе мы поочередно используем язык этих двух ситуаций. Победитель турнира — игрок с наибольшей суммой (очков) $a_{(t)}$. В случае дележа первого места первый приз делится поровну среди игроков с высокими суммами очков или проводится повторная игра (или игры) для выяснения победителя. Этот подход освящен веками. Метод § 3.4 можно использовать для проверки того, действительно ли $a_{(t)}$ $(= x_1)$ значимо лучше, чем *среднее*. Только когда $a_{(t)}$ существенно больше, чем следующий наибольший результат $a_{(t-1)}$, победителя турнира можно считать также и лучшим игроком. Это отражается в процедуре множественных сравнений из § 3.6—3.8 как тенденция объекта с наибольшим числом очков образовывать свою собственную группу. Метод, который больше приспособлен для выявления лучшего игрока, дается в 6.2.

Рассмотрим таблицу предпочтений* § 2.1. Для наших целей удобно положить диагональный элемент равным $\frac{1}{2}$, что дает пять результатов: $3\frac{1}{2}$, $2\frac{1}{2}$, $\frac{1}{2}$, $3\frac{1}{2}$, $2\frac{1}{2}$. Мы можем иначе подсчитать очки, давая каждому игроку половину его собственных очков плюс очки игроков, проигравших ему:

$$A_1: \frac{1}{2}\left(3\frac{1}{2}\right) + 0 + \frac{1}{2} + 3\frac{1}{2} + 2\frac{1}{2} = 8\frac{1}{4};$$

$$A_2: 3\frac{1}{2} + \frac{1}{2}\left(2\frac{1}{2}\right) + \frac{1}{2} + 0 + 0 = 5\frac{1}{4};$$

$$A_3: 0 + 0 + \frac{1}{2}\cdot\frac{1}{2} + 0 + 0 = \frac{1}{4};$$

* Спектральные свойства (обобщенных) турнирных матриц см. [288], где изучаются свойства ранжирования, см. также [245], [253]. — *Прим. пер.*

$$A_4: 0+2\frac{1}{2}+\frac{1}{2}+\frac{1}{2}\left(3\frac{1}{2}\right)+\qquad 2\frac{1}{2}=7\frac{1}{4};$$

$$A_5: 0+2\frac{1}{2}+\frac{1}{2}+\quad 0\quad +\frac{1}{2}\left(2\frac{1}{2}\right)=4\frac{1}{4}.$$

Мы видим, что при таком подсчете A_1 — очевидный победитель. Новое приписывание очков имеет эффект большего поощрения игрока, победившего сильного противника (набравшего много очков), чем победившего слабого противника, что кажется обоснованным, но зато проигрыш слабому противнику наказывается меньше, чем проигрыш сильному. Однако эта процедура, которую предложил Уей [144] и в дальнейшем изучил Кендэл [89], имеет некоторую привлекательность как возможный путь борьбы с дележем мест, особенно когда переигрывание неудобно (см. упражнение 6.2). С очевидными модификациями этот подход можно применить в случае, когда отдельные игры могут оканчиваться вничью. Перераспределение очков можно продолжать, непрерывно назначая новые очки, как прежде, пока не наступит окончательное упорядочение; так что после некоторого этапа дальнейшее перераспределение не будет нарушать предыдущее ранжирование. Если начальная матрица предпочтений обозначается A, то очки после r-го перераспределения получаются как строчные суммы матрицы A^{r+1}. Окончательное упорядочение* определяется собственным вектором, отвечающим наибольшему собственному числу матрицы A**.

6.2. ПОДХОДЯЩИЙ РАЗМЕР ЭКСПЕРИМЕНТА

В этой книге часто предполагалось и будет предполагаться в этом параграфе, что линейная модель (§ 1.3) применима. Теоретическое ранжирование объектов можно выполнить в соответствии с их «ценностями». Мы поставим в соответствие объекту $A_{(i)}$ ценность $V_{(i)}$, так что

$$V_{(1)} \leqslant V_{(2)} \leqslant \ldots \leqslant V_{(t)},$$

и обозначим $a_{[i]}$ очки объекта $A_{(i)}$. Следует заметить, что $a_{[i]}$ — теоретическая величина (не следует путать ее с $a_{(i)}$), так как в практической ситуации мы не знаем, какой объект надо обозначить $A_{(i)}$. Для удобства записи мы будем в этом и следующем разделах опускать скобки индексов для π и обозначать π_{ij} вероятность предпочтения $\Pr\{A_{(i)} \to A_{(j)}\}$.

Постановка задачи

Во многих экспериментах по сравнению t объектов (или обработок) основной интерес состоит в выделении лучшего объекта. Для сбалансированного эксперимента естественно называть лучшим объект

* См. «задачу о лидере» в [154], а также [209]. Этот подход развивается в [159], обсуждение задач, связанных с этим подходом, результаты численных экспериментов и литературу см. в обзорах [256] и [216]. — *Прим. пер.*

** Сравните, например, [76, § 8.6] (или [167, гл. XII, § 11] — *Прим. пер.*).

с наибольшим числом очков, но это может привести к неверным результатам из-за случайных колебаний. Однако если $A_{(t)}$ строго больше, чем $A_{(t-1)}$, и если число повторений n достаточно велико, то $A_{(t)}$ будет иметь наибольшую сумму с вероятностью P, близкой к 1 настолько, насколько это желательно. Мы сейчас рассмотрим один из методов определения такого n.

Возможная процедура такова:

а) находим n, соответствующее заданным t, P и набору вероятностей предпочтения $C(\pi_{ij})$;

б) производим эксперимент и объявляем лучшим объект с наибольшей суммой очков; если есть m равных наибольших сумм, то случайным образом один из этих объектов объявляем лучшим.

Шаг б) прямолинеен, правило выбора довольно наивно, что упрощает дальнейшее обсуждение. Однако с а) необходимо еще поработать. Так, для t и P фиксированных, n зависит от $C(\pi_{ij})$, и мы уточним π_{ij} в духе общего подхода, развитого Бехгофером [6]. Мы предположим, что $A_{(t)}$ имеет вероятность предпочтения с любым другим объектом $\pi \left(>\frac{1}{2}\right)$ и что оставшиеся $t-1$ объектов равноценны. Это предположение, частный случай линейной модели, попросту означает, что главный объект известен и кажется разумным принять это за основу при определении числа повторений эксперимента. Так как есть полная аналогия с упрощением Бехгофера, то оказывается, что и в данном случае выбор π_{ij} не обязательно соответствует наименее желательному варианту. Значения n даются в табл. 4 приложения, дальнейшие обсуждения приводятся ниже.

Выбранную модель можно сформулировать так:

$$\begin{aligned} &\pi_{tj}=\pi>\frac{1}{2}, \quad j=1,2,\ldots,t-1; \\ &\pi_{ij}=\frac{1}{2}, \quad i,j=1,2,\ldots,t-1;\ i\neq j. \end{aligned} \tag{6.2.1}$$

Точная теория распределения

Из (2.4.3) и (2.4.4) совместное распределение очков для модели (6.2.1) можно записать так:

$$f(\mathrm{a}_{[t]})=2^{-n\binom{t-1}{2}}g(\mathrm{a};n)\,\pi^{a_{[t]}}(1-\pi)^{n(t-1)-a_{[t]}}. \tag{6.2.2}$$

Предположим, что m очков одинаковы и соответствующие объекты занимают первое место. Число перестановок очков $a_1, a_2, \ldots, a_t$ равно $G(\mathrm{a};n)/g(\mathrm{a};n)$, часть перестановок m/t должна иметь наибольшую сумму очков на последнем месте, что соответствует $A_{(t)}$. При рандомизации на шаге б) соответствующий вклад в вероятность правильного выбора P будет равен

$$\frac{1}{m}2^{-n\binom{t-1}{2}}\frac{m}{t}G(\mathrm{a};n)\,\pi^{a_{[t]}}(1-\pi)^{n(t-1)-a_{[t]}} \tag{6.2.3}$$

и независим от m. Поэтому P получается суммированием (6.2.3) по всем $a_{[t]}$, которые максимальны, и по всем другим возможным значениям очков, и его можно выразить как

$$P = 2^{-n\binom{t-1}{2}} \sum_{a_{[t]}=c}^{n(t-1)} \pi^{a_{[t]}} (1-\pi)^{n(t-1)-a_{[t]}} \times$$

$$\times \sum_{\sum\limits_{i=1}^{t-1} a_{[i]} = n\binom{t}{2} - a_{[t]}} \frac{1}{t} G(a_1, a_2, \ldots, a_t; n), \qquad (6.2.4)$$

где c — наименьшее целое число, большее или равное $\frac{1}{2} n(t-1)$. Например, если $t = 3$ и $n = 1$, (6.2.4) превращается в

$$P = 2^{-1} [\pi(1-\pi) \frac{1}{3} G(1, 1, 1; 1) + \pi^2 \frac{1}{3} G(0, 1, 2; 1)],$$

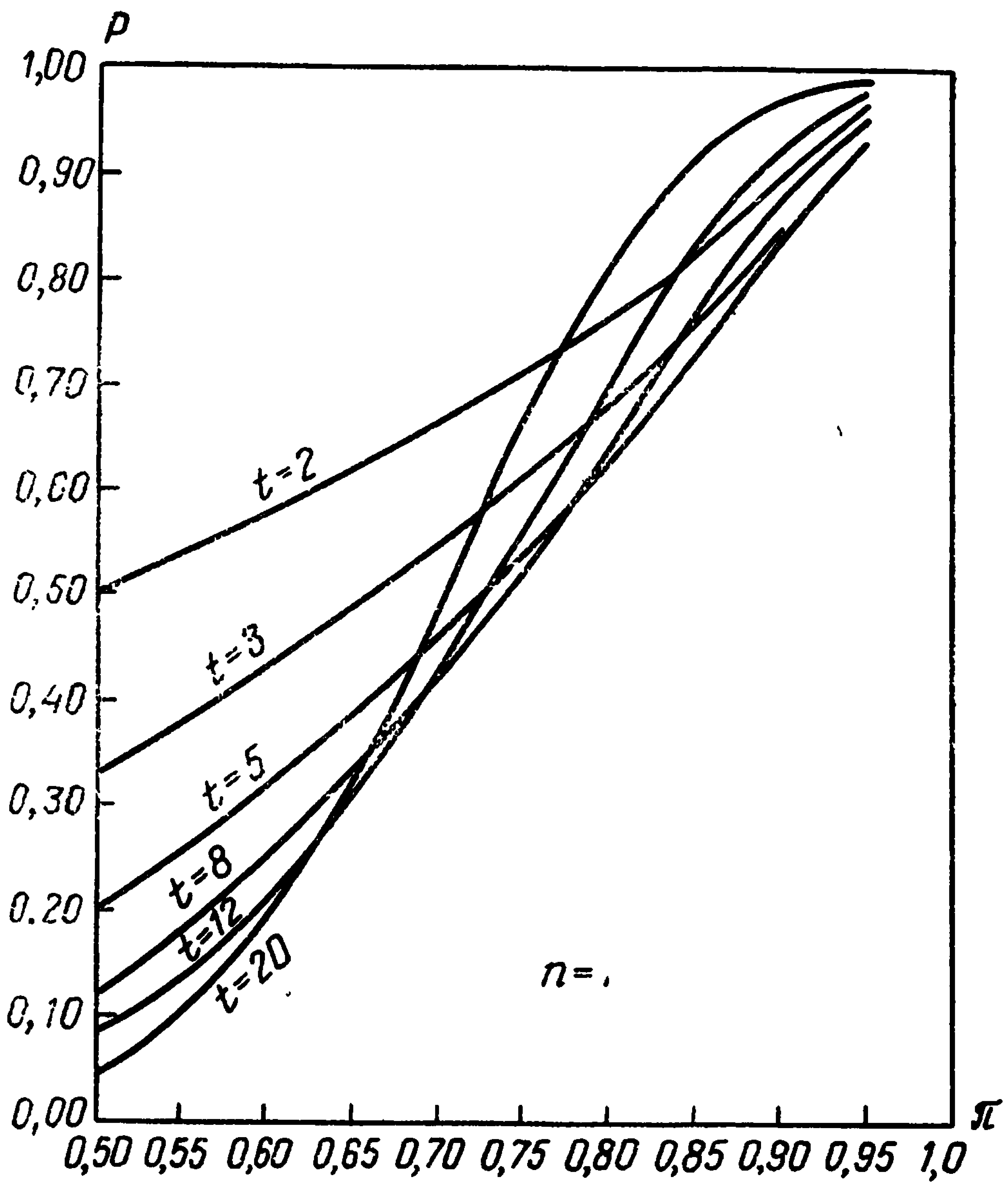
где $G(0, 1, 2; 1)$ — частота размещения [210], равная 6, и $G(1, 1, 1; 1) = 2$. Впрочем, для этого простого случая очевидно, что

$$P = \pi^2 + \frac{1}{3} \pi (1-\pi).$$

Главный член P всегда $\pi^{n(t-1)}$.

Случай $t = 2$ был предметом особого внимания Мостеллера [109] в контексте «Мировых серий», где две команды, финалисты чемпионата США по бейсболу, играли на большинство в серии из 7 игр. Ясно, что возможность окончания серий до того, как все игры будут сыграны, не изменяет вероятности P того, что лучшая команда выиграет. С другой стороны, нет достаточно удовлетворительного способа оценки вероятности того, что заданная команда выигрывает в определенном матче, хотя доля $a_1/(a_1 + a_2)$ побед более или менее годна для этого. Здесь мы имеем обобщение выбора из обратного биномиального распределения с серией, оканчивающейся в случае, когда либо a_1, либо a_2 достигают $c = 4$. Подходящая несмещенная оценка [54] есть $(c-1)/(c + a_2 - 1)$ или $a_1/(a_1 + c - 1)$ в соответствии с тем, выиграла ли данная команда или ее противник. Различные практические сложности, не вполне понятные для неамериканцев, обсуждаются у Мостеллера. См. также упражнение 6.1.

Комментарии к табл. 4 приложения. Для эксперимента, включающего t объектов с вероятностями предпочтений, удовлетворяющими (6.2.1), таблица дает наименьшее число повторений n, которое гарантирует, что высшая сумма очков в эксперименте размера (t, n) будет относиться к наилучшему объекту хотя бы с предварительно назначенной вероятностью P'. При создании таблицы точная теория использовалась для комбинаций t, n, не превышающих (2;268), (3;21), (4;8), (5;4), (6;1), (7;1) и (8;1). В другом месте будет описано асимптотическое приближение, полученное из асимптотической теории § 2.4 (см. [139]). Сравнение точных и асимптотических значений указывает на хорошее согласие для больших объемов эксперимента.



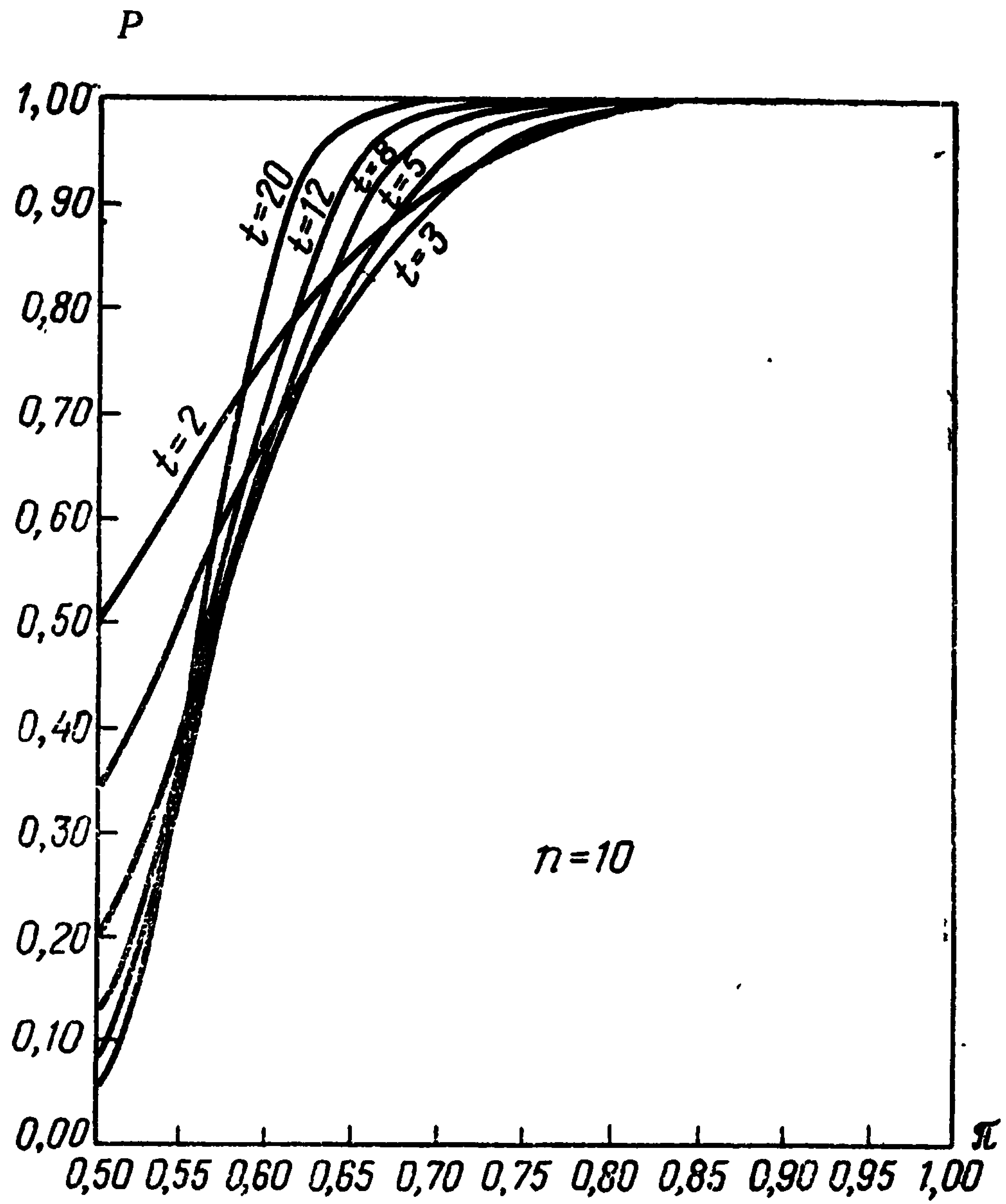


Рис. 6.1А и рис. 6.1Б. Асимптотическая вероятность правильного выбора наилучшего объекта $A_{(t)}$ при условии эквивалентности других t—1 объектов; n — число повторений, $\pi = \Pr\ (A_{(t)} \rightarrow A_{(i)})$, $i=1,2, \ldots, t-1$

Если модель (6.2.1) окажется менее подходящей, то n, полученное из таблицы, будет обеспечивать правильный выбор с вероятностью $P \geqslant P'$, пока

$$\Pr\{A_{(t)} \to A_{(t-1)}\} \equiv \pi_{t,\,t-1} \geqslant \pi. \qquad (6.2.5)$$

Это не необходимо, что довольно очевидно из рис. 6.1А и 6.1Б, которые показывают, как P возрастает в функции от π для $n = 1$ и $n = 10$, когда верна модель (6.2.1). Например, возьмем $t = 20$ и предположим, что первые пять объектов, рассматриваемые сами по себе, удовлетворяют модели (6.2.1) с $t = 5$. Пополнение набора объектов добавлением к пяти объектам пятнадцати не имеет значения, поскольку вероятность их предпочтения по отношению к первым пяти объектам равна нулю. Условия (6.2.5) очевидным образом удовлетворяются в этом расширенном случае. Теперь для $n = 1$ рис. 6.1А показывает, что P больше для $t = 20$, чем для $t = 5$, если $0{,}69 < \pi < 0{,}95$. Следовательно, для этого интервала π в случае расширенного опыта то же значение P, что и при $t = 5$, будет менее благоприятно, чем при $t = 20$. Нет, конечно, гарантии, что это самый неблагоприятный случай. На самом деле соображения такого рода указывают, что возможны трудности в общем случае при определении наименее благоприятной конфигурации и исключении ее в дальнейшем из практической ситуации. Так, таблицей 4 приложения можно спокойно пользоваться как справочником по определению подходящего n, только если верна модель (6.2.1). Однако верность этой модели имеет большое значение сама по себе, так как она соответствует ситуации единичного «выскочки».

Может быть интересна перефразировка приведенного выше примера на языке турниров. Если игрок имеет вероятность выигрыша у любого соперника π $(0{,}69 < \pi < 0{,}95)$, причем все соперники одинаково сильны, он имеет больше шансов выиграть в простом круговом турнире против 19 соперников, чем в малом турнире против четырех из них. (См. также [84].)

6.3. ВЫБОР ПОДМНОЖЕСТВА, СОДЕРЖАЩЕГО НАИЛУЧШИЙ ОБЪЕКТ

Рассмотрим множество объектов $A = \{A_1, A_2, \ldots, A_t\}$ и пусть S будет подмножеством A, состоящим из объектов с высокими местами. В этом разделе наша цель состоит в выделении множества S, настолько большого, чтобы обеспечить с вероятностью, не меньшей наперед заданной вероятности P^*, что лучший объект $A_{(t)}$ включен в S. Следуя Гупте и Собелю [68], мы используем решающее правило $\mathcal{R}$: оставлять в S только те объекты A_i, для которых $a_i \geqslant a_{(t)} - \nu$, где $a_{(t)}$ есть высшая сумма и ν — неотрицательное целое число, функция от t, n и P^*.

Для $P^* = 0{,}75$; $0{,}90$; $0{,}95$; $0{,}99$ и широкого диапазона значений t и n значения ν, определенные как из точной, так и из асимптотической теории, даны в табл. 5 приложения. Следует заметить, что размер

Пример

Мы покажем, как работает решающее правило $\mathcal{R}$ на данных Е. Йенсена (Faellesforeningen for Danmarks Brugsforeninger, Copenhagen). 15 человек дегустировали все возможные пары из 4 различных продуктов. Была получена следующая таблица предпочтений:

	A_1	A_2	A_3	A_4	a_i
A_1	—	3	2	2	7
A_2	12	—	11	3	26
A_3	13	4	—	5	22
A_4	13	12	10	—	35

Мы имеем $a_{(t)} = 35$. Для обеспечения того, чтобы по крайней мере с наперед заданной вероятностью $P^* = 0{,}75$ наилучший образец был в выбранном подмножестве, мы обращаемся к табл. 5 приложения при $t = 4$, $n = 15$, $P^* = 0{,}75$ и находим $\nu = 7$. Поэтому оставляем в подмножестве лишь A_4. Для $P^* = 0{,}90$ мы имеем $\nu = 9$, так что подмножество состоит из A_4 и A_2 и т. д.

6.4. ОЛИМПИЙСКИЙ ТУРНИР (С ВЫБЫВАНИЕМ) И ДРУГИЕ ТУРНИРЫ

Аналогия между сбалансированными экспериментами парных сравнений и круговыми турнирами наводит на мысль, что другие типы турниров тоже могут заслуживать более внимательного изучения*. Особенностью таких турниров является то, что от уравновешивания отказываются либо для ускорения, либо для увеличения числа встреч между наиболее удачными игроками, или с обеими целями. Отсутствие баланса затрудняет изучение свойств этих турниров, но они интуитивно привлекательны как метод проведения эксперимента в том случае, когда наша главная цель — выбор наилучшего объекта. Хорошо известен олимпийский турнир (или игра на кубок), с которого мы начнем.

Олимпийский турнир

Будем предполагать, что число игроков t было уменьшено в предварительных состязаниях или другим способом так, что оно стало равно некоторой степени 2, скажем, $t = 2^p$. Победитель может быть объяв-

* См. в первую очередь книгу [287], где дается обзор различных систем турниров, работы [294], [295], [293] и [286], где обсуждаются свойства различных турниров. — *Прим. пер.*

лен в этом случае после p туров и $(t-1)$ игр, а в круговом турнире потребуется $(t-1)$ туров и $\frac{1}{2}t(t-1)$ игр, без переигровок.

Если некий игрок A_1 побеждает любого соперника с вероятностью π, то он выигрывает турнир с выбыванием с вероятностью π^p. В более сложном случае вероятность выигрыша игрока A_1, очевидно, зависит от первоначального хода турнира, который, как мы предположим с самого начала, будет совершенно случайным. Пусть $P_i^{(r)}$ ($i = 1, \ldots, t$; $r = 1, 2, \ldots, p$) означает вероятность того, что A_i доходит до r-го тура турнира и выигрывает свою встречу в этом туре, тогда $P_i^{(p)}$ — вероятность его победы в турнире. Пусть $M_{ij}^{(r)}$ — вероятность того, что A_i встречается с A_j ($j = 1, 2, \ldots, t$; $j \neq i$) в r-том туре и M_{ij} — вероятность их встречи в ходе турнира.

Ясно, что

$$M_{ij} = \sum_{r=1}^{p} M_{ij}^{(r)}. \qquad (6.4.1)$$

Предположим теперь, что

$$\pi_{1j} = \pi_1 \quad (j = 2, 3, \ldots, t),$$
$$\pi_{2j} = \pi_2 \quad (j = 1, 3, \ldots, t).$$

Тогда вероятность выигрыша A_1 в турнире

$$P_1^{(p)} = M_{12}\pi_{12}\pi_1^{p-1} + (1 - M_{12})\,\pi_1^p. \qquad (6.4.2)$$

Если известно, как проходил турнир*, то M_{12} и, следовательно, $P_1^{(p)}$ можно вычислить. Вместо этого мы можем получить выражения для $\overline{M}_{12}$ и $\overline{P}_1^{(p)}$ значений M_{12} и $P_{12}^{(p)}$ *до проведения* жеребьевки. Эти априорные вероятности до сих пор связывались с (6.4.2). Мы имеем

$$\overline{M}_{12}^{(1)} = 1/(t-1).$$

Если A_1 и A_2 встречаются во втором туре, в котором будут участвовать $\frac{1}{2}t$ игроков, они оба должны выиграть в первом туре, так что

$$\overline{M}_{12}^{(2)} = (1 - \overline{M}_{12}^{(1)})\,\pi_1\pi_2 \Big/ \left(\frac{1}{2}t - 1\right) = $$
$$= 2\pi_1\pi_2/(t-1).$$

Продолжая рассуждения, находим

$$\overline{M}_{12}^{(r)} = (2\pi_1\pi_2)^{r-1}/(t-1)$$

и, следовательно, из (6.4.1)

$$\overline{M}_{12} = \sum_{r=1}^{p} (2\pi_1\pi_2)^{r-1}/(t-1).$$

Интересно заметить, что $\overline{M}_{ij}^{(r)}$ зависит от π_1 и π_2 лишь как функция их произведения $\pi_1\pi_2$. Мы видим также, что вероятность того, что A_1 и A_2 встретятся в финале, равна $\overline{M}_{12}^{(p)}$. Если на время мы вообразим,

* Известна жеребьевка.—*Прим. пер.*

что A_1 и A_2 были фаворитами, которых поместили в противоположные половины жеребьевки, эта вероятность будет

$$(\pi_1\pi_2)^{p-1} \sim 2\bar{M}_{12}^{(p)}.$$

Для того чтобы рассмотреть общий случай с вероятностями предпочтения $C(\pi_{ij})$, возьмем $t = 8$ и предположим (без потери общности), что $A_1, A_2, \ldots, A_8$ выстроились по жребию в этом порядке. Так, мы имеем, например:

$$P_1^{(1)} = \pi_{12},\ P_2^{(1)} = \pi_{21},\ P_3^{(1)} = \pi_{34} \ldots$$

$$P_1^{(2)} = P_1^{(1)}(\pi_{13}P_3^{(1)} + \pi_{14}P_4^{(1)}) = \pi_{12}(\pi_{13}\pi_{34} + \pi_{14}\pi_{43}), \quad (6.4.3)$$

$$P_1^{(3)} = P_1^{(2)}(\pi_{15}P_5^{(2)} + \pi_{16}P_6^{(2)} + \pi_{17}P_7^{(2)} + \pi_{18}P_8^{(2)}). \quad (6.4.4)$$

Соответствующие вероятности до жеребьевки можно получить усреднением по всем возможным исходам жеребьевок. Пусть $P_1^{(2)}(i, j, k)$ означает совместную вероятность того, что A_1 победит A_2 в первом туре и выиграет у победителя встречи A_j и A_k во втором; тогда $P_1^{(2)}$ (2, 3, 4) — это выражение (6.4.3). Также мы можем (для $t = 8$) написать $P_1^{(3)}$ из (6.4.4) более подробно в виде $P_1^{(3)}$ (2,3, 4). Отсюда следует, что

$$\bar{P}_1^{(1)} = \frac{1}{7}\sum_{i=2}^{8} \pi_{ij}, \quad (6.4.5)$$

$$\bar{P}_1^{(2)} = \frac{1}{105}\sum_{i}\sum_{j<k}\sum_{1\neq i\neq j\neq k} P_1^{(2)}(i, j, k), \quad (6.4.6)$$

$$\bar{P}_1^{(3)} = \frac{1}{105}\sum_{i}\sum_{j<k}\sum_{1\neq i\neq j\neq k} P_1^{(3)}(i, j, k).$$

Для $t > 8$ число членов в правой части может стать огромным. Так, $\bar{P}_1^{(r)}$ есть среднее из

$$(t-1)\binom{t-2}{2}\binom{t-4}{4}\cdots\binom{t-2^{r-1}}{2^{r-1}} \text{ членов.}$$

*Сравнение эффективности турниров**

Для сравнения олимпийского турнира с круговым мы можем использовать как критерий вероятность того, что лучший игрок (т. е. игрок с наибольшей ожидаемой суммой очков) выиграет турнир. Конечно, такое сравнение несправедливо по отношению к олимпийскому турниру, который требует значительно меньше игр (однако не всегда, см. упражнение 6.3). Поэтому при сравнении можно рассмотреть олимпийский турнир с повторениями (в несколько кругов).

Например, если $t = 4$, два повторения требуют 6 игр, то же число, что и для кругового турнира с 4 игроками. Возможные исходы описы-

* Сравнение эффективности турниров см. в [307] и [286]. — *Прим. пер.*

ваются размещениями $[420^2]$, $[41^20]$, $[3210]$, $[3^20^2]$, $[2^21^2]$, $[2^30]$, первые три из которых отвечают единоличной победе одного из трех игроков. Трудность состоит в том, что для обоих типов турниров возможен дележ первого места. Простейший выход из этого положения, причем совсем неплохой, если вероятность дележа мала, — это бросание жребия для определения победителя. В практике проведения турниров устраивается повторная игра (переигрывание); если это возможно, то число игр становится случайной величиной, не меньшей 6. Возникает также вопрос: как повторять игры в олимпийском турнире? Можно использовать две различные случайные жеребьевки (назовем их K_2), но, вероятно, лучше просто поместить двух финалистов первого турнира в разные половины жеребьевки второго турнира, а их соперников из первого круга поменять местами (K_3). Гленн [55] исследовал этот и смежные вопросы, в основном в серии численных примеров, для различных наборов вероятностей предпочтения, а также при строгой стохастической транзитивности (§ 1.3). Дележ первого места устранялся кратчайшим возможным переигрыванием: дележ между двумя игроками требует в точности одной дополнительной игры, но тройной дележ требует уже *по крайней мере* трех дополнительных игр и может привести к бесконечному числу повторений (с вероятностью нуль). Гленн рассматривает также другой тип турнира с выбыванием (K_1), введенный Морис [104], в котором предполагается лучший выход — скорее тремя играми, чем одной, определять победителя каждой встречи, что требует всего от шести до десяти игр. В конце концов, хотя и нельзя исчерпать все разумные возможности, он работает с турниром с вторичным исключением (D). В этом турнире игрок не исключается до проигрыша двух игр. Первый круг такой же, как для обычного олимпийского турнира, но во втором круге в дополнение к встрече двух победителей проводится встреча двух проигравших. В конце этого круга один игрок выбывает из игры, а остаются один двукратный победитель и два игрока, у которых по одному выигрышу (B) и по одному проигрышу (П). Мы обозначим трех уцелевших *ВВ*, *ВП*, *ПВ* соответственно. Последние два встречаются в пятой игре и победитель встречи *ВПВ* и *ПВВ* встречается с *ВВ* в шестой игре. Если *ВВ* выиграет эту игру, он победитель турнира, если он проиграет, два других игрока встречаются в седьмой, решающей игре.

Для заданных значений π_{ij} можно вычислить не только вероятность, с которой лучший игрок выигрывает турнир, но и *ожидаемое* число игр, которые потребуются. В работе Гленна K_1 — лучший в первом смысле, но худший во втором, однако D — лучший в обоих смыслах по сравнению с K_2, K_3 и круговым турниром, который обозначен R. В частном случае

$$\pi_{12} = \pi_{13} = \pi_{14} = \pi \quad \text{и} \quad \pi_{23} = \pi_{24} = \pi_{34} = \frac{1}{2}.$$

Эти выводы верны для всех значений π (сравните упражнение 6.4). Для принятия решения о выборе K_1 или D необходимо определить какую-нибудь подходящую в данной ситуации функцию цены, в которой «цена» от выбора худшего игрока должна сравниваться с ценой проведения игры. Конечно, турниры с сильно отличающимися размерами также могут сравниваться с помощью функции цены, однако это целесообразно, только если есть веские основания для выбора какой-либо функции цены.

Другие проблемы турниров

Много лет тому назад Льюис Кэррол* (см., например, издание 1947 г.) выразил неудовлетворение использованием в теннисе турниров с выбыванием из-за очевидной ненадежности при определении игрока, занявшего второе место. Кэррол полагал, что соревнующиеся должны быть точно расположены в порядке их силы и что игрок будет *всегда* побеждать более слабого соперника. Случайным элементом в турнире будет лишь жеребьевка. При этих предположениях ясно, что лучший игрок должен победить наверняка. Средство против возможной несправедливости по отношению к игроку на втором месте Кэррол дает в виде подробного предписания, существенной чертой которого является то, что игрок исключается из турнира лишь тогда, когда вполне ясно, что два других игрока превосходят его. Это может быть прямой победой над ним этих игроков (как в турнире с вторичным исключением) или непрямой, когда игрок терпит первое поражение от соперника, который однажды проиграл. Скорее, хотя это и будет менее спортивным, стоит организовать турнир с выбыванием между игроками, проигравшими чемпиону. Для того чтобы показать, сколько матчей требуется для этой процедуры, рассмотрим обычный турнир с выбыванием с t игроками, где, вообще говоря,

$$t = 2^p + q \quad (1 \leqslant q \leqslant 2^p).$$

Понятно, что p предварительных матчей понадобятся для уменьшения числа игроков до 2^p, так что турнир закончится при

$$q + (2^p - 1) = t - 1 \text{ матчах.}$$

Следовательно, чемпион может побеждать столько же, сколько $p + 1$ его соперников в p дальнейших матчах, где определяется игрок, занявший второе место, участвуя всего в

$$t - 1 + p = t - 1 + [\log_2 (t - 1)] \text{ матчах,}$$

где квадратные скобки означают целую часть числа, как и далее, до конца этого раздела. Этот результат дан Штейнгаузом [130], который также рассматривал задачу обеспечения полного ранжирования за минимальное число матчей (единичных парных сравнений). Он предполагает, что множество игроков уже было ранжировано и что мы хотим найти место для нового игрока A. Сначала выделим медианного игрока M, т. е. такого, что в ранжировке одинаковое число игроков имеют ранг меньше и больше, чем у него; если же число игроков четное, то

---

* Математик и известный детский писатель Charles Lutwidge Dodgson; Lewis Carroll — его псевдоним (русское написание Кэррол, Кэролл, Кэрролл). — *Прим. пер.*

двух игроков в середине ранжировки мы называем медианными. Затем мы устроим матч A с M. Если $A \to M$, мы проведем матч с медианным среди игроков, превосходящих M; если $A \leftarrow M$, проведем матч A с медианным игроком нижней половины и т. д., пока не будет определено место A в ранжировке. Так, для ранжирования 3 игроков надо 3 матча (и два матча *могут* это обеспечить, но не так хорошо), четыре игрока можно проранжировать за 5 матчей, пять — за 8 матчей и т. д. Размещение $(k+1)$-го игрока в ранжировке требует $S(k) = 1 + [\log_2 k]$ матчей. Следовательно, общая формула для числа матчей при ранжировании t игроков

$$M(t) = S(1) + S(2) + \dots + S(t-1) =$$
$$= (t-1) + [\log_2 2] + [\log_2 3] + \dots + [\log_2 (t-1)].$$

Мы запишем t в виде

$$t = 2^{r-1} + s \qquad (0 \leqslant s < 2^{r-1}),$$

так что $r = 1 + [\log_2 t] = S(t)$.

Тогда

$$M(t) = t - 1 + 2(1) + 2^2(2) + 2^3(3) + \dots + 2^{r-2}(r - 2) + s(r-1) = t - 1 + 2\{1 + (r-3)2^{r-2}\} + (t - 2^{r-1})(r-1) = 1 + rt - 2^r.$$

Это не самый короткий из возможных путей, что было показано Фордом и Джонсоном [49]. Их оригинальная улучшенная процедура, оптимальность которой до сих пор открытый вопрос, состоит в следующем.

Предположим, $t = 2i$ или $2i + 1$. Тогда

а) разделим на пары $2i$ игроков и пусть пары играют в первом круге, оставляя одного игрока, если t нечетное;

б) распространение данного метода на i игроков дает полное ранжирование этих победителей первого круга;

в) третий шаг лучше изобразить на схеме. По этому пункту мы имеем иерархическую конструкцию, изображенную на рис. 6.2 для $t = 19$.

Победители первого круга $J, I, \dots, B$ ранжированы в этом порядке, J — лучший игрок, A в первом круге проиграл B, другие проигравшие в первом круге расположены прямо под своими победителями. Нечетный игрок, не принимавший участия в первом круге, рассматривается как проигравший и занимает позицию на левом краю схемы.

Выражение «главная цепь» первоначально относилось к цепи $JIH \dots CBA$, процедура введения игроков относится к занумерованным точкам в главной цепи, в указанном порядке. Процедура основывается на том, что размещение точек в цепи методом Штейнгауза наиболее эффективно, если их число имеет вид $2^K - 1$.

Теперь мы начнем с введения точки 1 в цепь ABC. После того как это будет сделано, главная цепь под точкой 2 будет состоять из AB и, возможно, точки 1; это введение можно выполнить за два сравнения.

Теперь мы обратимся к цепи длиной $2^3 - 1 = 7$ и заметим, что точка 3 расположена так высоко, как только возможно, главная цепь под 3 состоит из $ABCDE$, 1 и 2 и т. д.

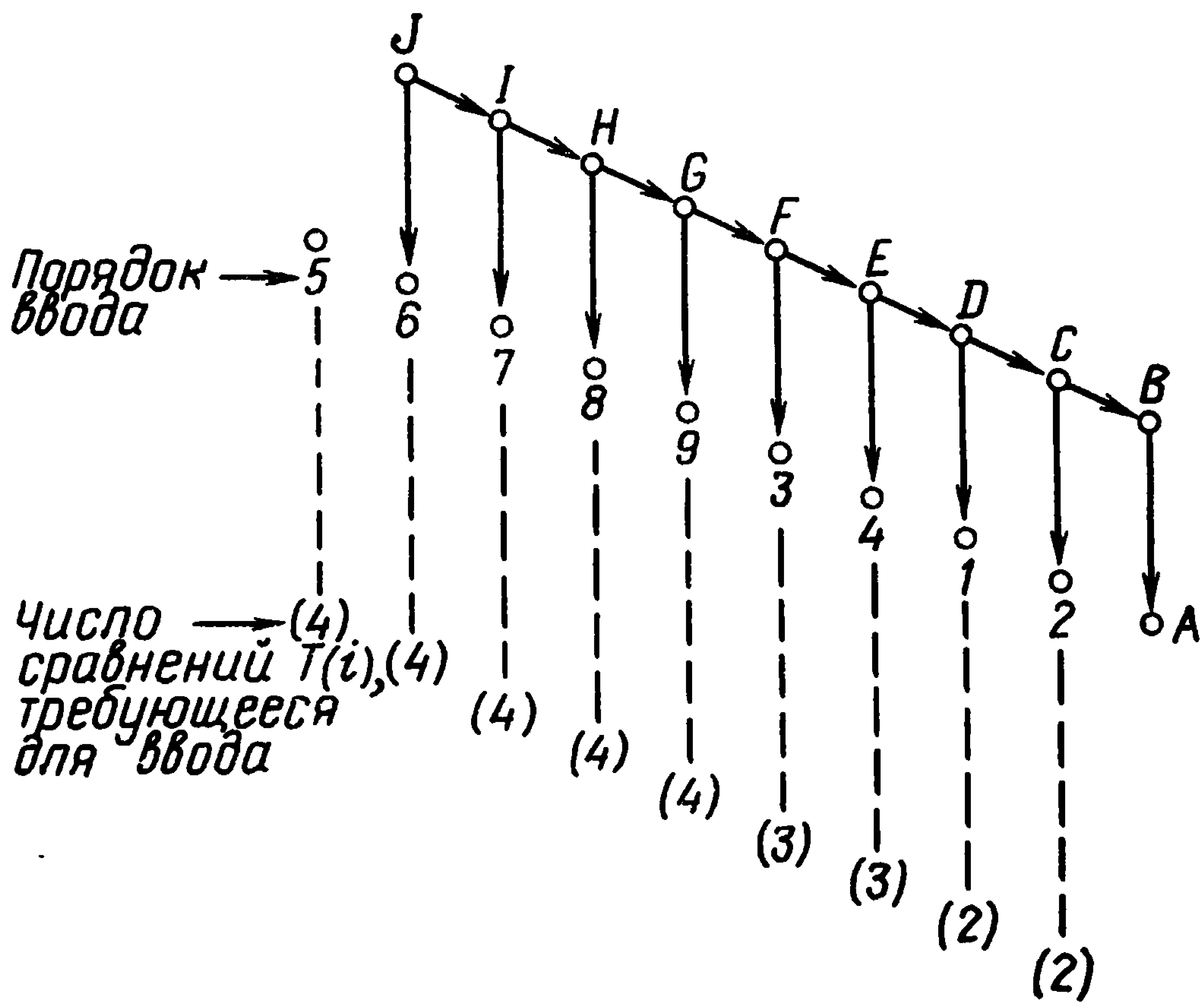


Рис. 6.2. Из работы [49], с разрешения авторов и редактора «*American Mathematical Monthly*»

Это образует методику ранжирования, требующую $U(t)$ сравнений, где $U(t)$ дается следующими рекуррентными соотношениями:

$$U(2k) = k + U(k) + \sum_{j=2}^{k} T(j);$$

$$U(2k+1) = k + U(k) + \sum_{j=2}^{k+1} T(j),$$

при $U(1) = 0$, $U(2) = 1$; $T(j)$ — число сравнений, требующееся для введения игроков в цепь длиной j; и

$$\begin{aligned} T(j) &= 2 \text{ для } 1 < j \leqslant 3, \\ &= 3 \text{ для } 3 < j \leqslant 5, \\ &= 4 \text{ для } 5 < j \leqslant 11, \\ &\dots\dots\dots\dots\dots \\ &= h \text{ для } t_{(h-1)} < j \leqslant t_h, \end{aligned}$$

где $t_h = \frac{1}{3}\{2^{h+1} + (-1)^h\}$.

Таблица 6.1

(из [49], с разрешения автора и издателя)

t	1	2	3	4	5	6	7	8	9	10	11	12	13
$M(t)$	0	1	3	5	8	11	14	17	21	25	29	33	37
$U(t)$	0	1	3	5	7	10	13	16	19	22	26	30	34
$L(t)$	0	1	3	5	7	10	13	16	19	22	26	29	33

В табл. 6.1 приведены значения $M(t)$, $U(t)$ для $t \leqslant 13$. Последняя строка дает $L(t)$ — нижнюю границу, которую Форд и Джонсон получили из теории информации. Каждое соединение в пару не может дать сведений более чем о разделении оставшихся возможностей на два взаимно дополняющих множества; в результате сравнения выделяется то или иное из этих множеств. Очевидные исходные $t!$ возможностей делятся пополам; это лучшее, что можно сделать на каждом шаге, поэтому мы приходим к формуле

$$L(t) = \text{наименьшее целое} \geqslant \log_2 t!$$

УПРАЖНЕНИЯ

6.1. Два теннисиста A и B играют матч, состоящий из n сетов. A выигрывает сет у B с постоянной вероятностью π. Если они заканчивают игру за четное число сетов ($n = 2r$) со счетом $r : r$, то победитель определяется по жребию. Покажите, что вероятность победы в матче одна и та же для $n = 2r - 1$ и для $n = 2r$.

6.2. Пусть в круговом турнире четырех игроков размещение очков $[2^3 0]$. Покажите, что тройной дележ нельзя устранить методом § 6.1 для пересчитанных очков.

Далее, предположим, что

$$\pi_{12} = \pi_{23} = a > \frac{1}{2}, \quad \pi_{13} = b > a, \quad \pi_{13} = \pi_{14},$$

$$\pi_{23} = \pi_{24}, \ \pi_{34} = \frac{1}{2}$$

и размещение очков есть $a_1 = 2$, $a_2 = 2$, $a_3 = 1$, $a_4 = 1$. Докажите, что пересчет очков будет в пользу сильнейшего игрока, если и только если

$$b < \frac{a^2}{a^2 + (1-a)^2},$$

и что равенства очков не нарушаются, если верна модель Брэдли—Терри.

6.3. Пусть в круговом турнире четырех игроков один из них имеет вероятность $\pi > \frac{1}{2}$ победы над другими, каждый из которых имеет одинаковую силу. Покажите, что, если дележ первого места устранять бросанием жребия, лучший игрок побеждает в турнире с вероятностью π^2, той же, что и в олимпийском турнире.

6.4. Если в упражнении 6.3 минимальное правило выбывания используется так, как описано в § 6.4, покажите, что лучший игрок выигрывает круговой турнир с вероятностью

$$P_R = \pi^3 (2 - \pi) + \pi^3 (1 - \pi) (2 + \pi + 2\pi^2)/4 (1 - \pi + \pi^3).$$

Покажите также, что

$$P_{K_1} = \pi^3 (2 - \pi) + \pi^3 (1 - \pi) (- 2 + 8\pi - 4\pi^2),$$

$$P_{K_2} = \pi^3 (2 - \pi) + 2\pi^4 (1 - \pi) (3 - 2\pi + \pi^2)/3 (1 - \pi + \pi^2),$$

$$P_{K_3} = \pi^3 (2 - \pi) + \pi^4 (1 - \pi) (4 - 3\pi + 2\pi^2)/2 (1 - \pi + \pi^2)$$

и $P_D = \pi^3 (2 - \pi) + 2\pi^4 (1 - \pi)$

и, следовательно, что для $1/2 < \pi < 1$

$$P_{K_1} > P_D > P_{K_3} > P_{K_2} > P_R. \qquad \text{(см. [55])}$$

7 ГЛАВА СМЕЖНЫЕ ВОПРОСЫ

7.1. ОТБОР ЭКСПЕРТОВ

До сих пор в наших исследованиях мы не обращали внимания на практически важную задачу выбора группы экспертов. На этот выбор, конечно, влияют такие факторы, как стоимость эксперимента и наличие подходящих экспертов. Еще более важный фактор — цель эксперимента парных сравнений, для которого отбираются эксперты.

Изучать предпочтения в «полевых условиях» надо в большой *комиссии потребителей*, представляющих популяцию, которая нас интересует, а для выявления тонких различий в лабораторных условиях и установления стандартов качества привлекается небольшая *группа экспертов*.

Более подробно о комиссиях потребителей мы поговорим в § 7.2. Что касается экспертных групп, то первоначальный интерес представляет выявление различий между объектами, одно или более из которых могут быть «стандартными». Должна рассматриваться каждая попытка определить тип различия. В пищевой промышленности отбор экспертов обыкновенно основывается на простых тестах, таких, как парный, тройной и двойной-тройной. До обсуждения мы отметим, что они включают лишь два различных типа объектов. Чтобы охватить широкий диапазон, экспериментатор может организовать сбалансированный эксперимент парных сравнений с несколькими объектами. Следует выбирать экспертов с наименьшим числом циклов, если нет априорного упорядочения, или с наименьшим числом разделений (§ 2.1) плюс циклы, если есть такое упорядочение.

Ранжирование [143] и приписывание очков [44] также могут использоваться как методы отбора, но это представляется менее желательным, когда последующие эксперименты проводятся путем парных сравнений (см. также [10]).

Парный, тройной и двойной-тройной тесты

При парном тесте два различных объекта представляются эксперту. Один из них заведомо «лучший» по некоторому признаку, и верный ответ получается, если эксперт опознает этот объект. Другие два теста включают три объекта, два из которых близки. Назовем их A_1, A_1', и пусть A_2 означает непарный объект. В обоих случаях эксперт пытается указать A_2; в двойном-тройном тесте он дополнительно устанавливает идентичность[1] A_1 и A_1'. Можно видеть, что вероятность верного ответа вполне случайна и равна $\frac{1}{2}$ для парного и двойного-тройного тестов и $\frac{1}{3}$ для тройного теста.

Несколько свойств этих тестов были проверены в прекрасной работе [142], которая местами слишком сокращена со ссылкой на ограничение объема. Дальнейшие детали даны у Дэвида и Трайведи [30]. Пусть наблюдаемые ценности объектов A_1, A_1', A_2 равны соответственно y_1, y_1', y_2. Юэре предполагает фактически, что верна линейная модель и что наблюдаемые ценности есть непрерывные независимые случайные величины. Без потери общности мы можем считать A_2 лучшим объектом. Правильные ответы получаются, если

$y_1 < y_2$ для парного теста;

$|y_1 - y_1'| < |y_1 - y_2|$ и $|y_1 - y_1'| < |y_1' - y_2|$ для тройного теста;

$|y_1 - y_1'| < |y_1 - y_2|$ для двойного-тройного теста и A_1 опознан.

Обозначим соответствующие вероятности через P_P, P_Δ и P_D. Для y_1, y_1', распределенных нормально $N(\mu_1, \sigma^2)$, и y_2, распределенной нормально $N(\mu_2, \sigma^2)$, $(\mu_2 > \mu_1)$, мы имеем

$$P_P = \Phi(\delta/\sqrt{2}), \tag{7.1.1}$$

где Φ — стандартная нормальная функция распределения и $\delta = (\mu_2 - \mu_1)/\sigma$. Можно показать, что[2]

$$P_\Delta = 2\int_0^\infty \left[\Phi\left(\sqrt{\frac{2}{3}}\,\delta - \sqrt{3}\,x\right) + \Phi\left(-\sqrt{\frac{2}{3}}\,\delta - \sqrt{3}\,x\right)\right] d\Phi(x) \tag{7.1.2}$$

и (упражнение 7.1)

$$P_D = 1 - \Phi(\delta/\sqrt{2}) - \Phi(\delta/\sqrt{6}) + 2\,\Phi(\delta/\sqrt{2})\,\Phi(\delta/\sqrt{6}). \tag{7.1.3}$$

Юэре также получил соответствующие выражения для случая, когда y распределены равномерно с выписанными выше математическими ожиданиями и дисперсиями (упражнение 7.2). Результаты как функции от δ показаны на рис. 7.1 и неожиданно близки для двух распределений. Если на мгновение мы сочтем три размещения средством выявления различий между двумя похожими объектами (а не средством выбора экспертов), то мы сможем сравнить их действие, прове-

[1] Эксперта можно также спрашивать, предпочитает ли он выбранный объект или нет, но это сомнительное дело (см. [20]).

[2] Эквивалентные выражения в виде бесконечных рядов были ранее получены Брэдли в неопубликованной заметке.

ряя нуль-гипотезы $P_P = \frac{1}{2}$; $P_\Delta = \frac{1}{3}$; $P_D = \frac{1}{2}$ против соответствующих альтернатив $P_P > \frac{1}{2}$, $P_\Delta > \frac{1}{3}$, $P_D > \frac{1}{2}$. Предположим, что число повторений $n=20$. Таким образом, функции мощности указанных трех биномиальных критериев можно легко найти с помощью нормальной аппроксимации биномиального распределения. Преимущество за парным тестом. Так, например, на 5%-ном уровне значимости разность $\mu_2 - \mu_1 = 1{,}5\,\sigma$ выявляется с вероятностью 0,95 парным те-

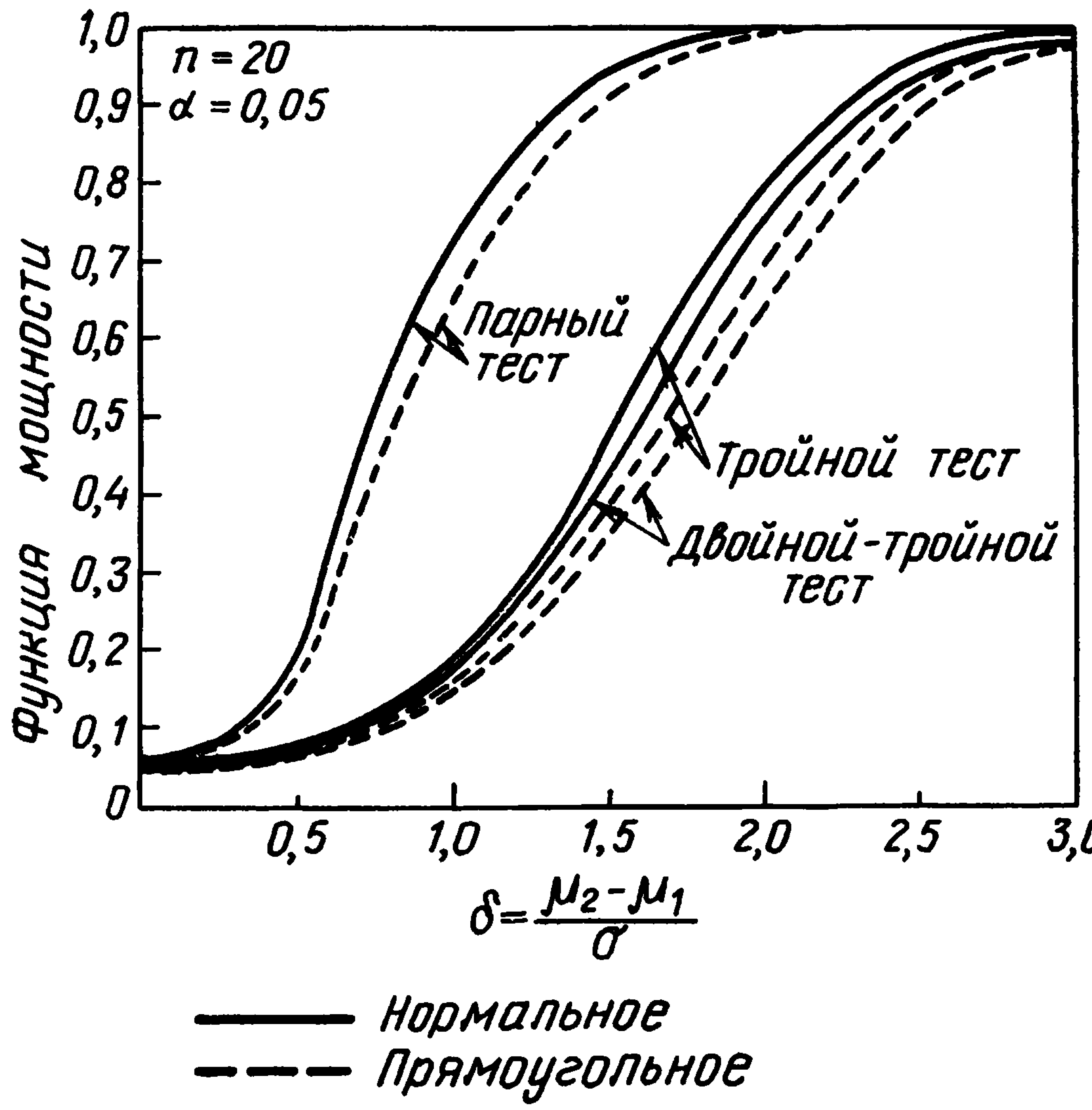


Рис. 7.1. Вероятность верного ответа. Из работы [142], с разрешения автора и редактора «*Reporte of Statistical Applications Research*» — журнала Союза японских ученых и инженеров (Union of Japanese Scientists and Engineers)

стом, соответствующие вероятности для других тестов меньше 0,5. Превосходство парного теста над двойным-тройным тестом очевидно из рис. 7.1, но тройной тест лишь немного лучше, чем двойной-тройной тест. Эти теоретические результаты согласуются с экспериментальными, которые получили Байер и Абрамс [20], Хопкинс и Гриджмен [83], так же, как и с исследованиями Юэре.

Когда мы выбираем экспертов, то мы интересуемся не столько тем, обладают ли они некоторой способностью, сколько тем, достаточно ли велика их способность.

Мы можем охарактеризовать способность различения эксперта его значением σ и выбирать лишь тех экспертов, для которых $\sigma \leqslant \sigma_0$, где σ_0 соответствует предписанной степени способности. Для разницы $\Delta\mu = \mu_2 - \mu_1$ между A_2 и A_1 вероятность правильного ответа может быть выражена как $p\ (\Delta\mu/\sigma)$ [1]. Так как это возрастающая функция от σ для фиксированного $\Delta\mu$, гипотезы

$$H_0 : \sigma \leqslant \sigma_0 \text{ и } H_1 : \sigma > \sigma_0 \tag{7.1.4}$$

эквивалентны

$$H_0 : p\ (\Delta\mu/\sigma) \geqslant p\ (\Delta\mu/\sigma_0) \text{ и} \tag{7.1.5}$$
$$H_1 : p\ (\Delta\mu/\sigma) < p\ (\Delta\mu/\sigma_0).$$

При известных очках x верных ответов экспертов в n испытаниях мы отвергаем H_0, применяя преобразование arcsin, если

$$\sin^{-1}\sqrt{x/n} \leqslant \sin^{-1}\sqrt{p\ (\Delta\mu/\sigma_0)} - u_\alpha/\sqrt{4n},$$

где u_α—верхняя α-процентная точка нормированного нормального распределения. Функция мощности этого критерия, приближенно равная

$$\Phi\left\{\sqrt{4n}\left[\sin^{-1}\sqrt{p\ (\Delta\mu/\sigma_0)} - \sin^{-1}\sqrt{p\ (\Delta\mu/\sigma)}\right] - u_\alpha\right\},$$

функция $\Delta\mu/\sigma_0$ и σ. Так как величина $\Delta\mu$ может варьировать, перед экспериментатором встает вопрос, каково ее оптимальное значение. Из изучения функций мощности Юэре заключает, что для $\alpha = 0{,}05$ $\Delta\mu$ должно быть примерно таким, чтобы вероятность верного ответа была 0,95 для эксперта, который вполне приемлем ($\sigma = \sigma_0$). Более того, он находит, что тройной тест наиболее приемлем в этих обстоятельствах.

Так как (7.1.5) означает просто нулевую и альтернативную гипотезы для биномиального параметра, можно ожидать, что последовательные методы для проверки биномиальной доли дадут уменьшение в среднем числа повторений для каждого эксперта [13]. В дополнение к σ_0 мы требуем теперь $\sigma_1 > \sigma_0$, чтобы эксперты с $\sigma \geqslant \sigma_1$ оказались нежелательными. Затем мы проверяем последовательно

$$H_0 : p\ (\Delta\mu/\sigma) = p\ (\Delta\mu/\sigma_0) \text{ против } H_1 : p\ (\Delta\mu/\sigma) = p\ (\Delta\mu/\sigma_1).$$

С другой стороны, можно работать с вероятностью правильного ответа p, фиксируя подходящие значения p_0 и p_1 ($< p_0$) и проверяя

$$H_0 : p = p_0 \text{ против } H_1 : p = p_1.$$

Это в то же время критерий

$$H_0 : p \geqslant p_0 \text{ против } H_1 : p \leqslant p_1.$$

Можно заметить, что если нет ясной методики определения приемлемых значений σ и p, то можно использовать методы принятия множественных решений для биномиальной совокупности [68] для отсеивания плохих экспертов.

---

[1] Эта вероятность также зависит от того, которое из трех размещений используется, но мы опускаем индексы P, Δ, D для удобства.

7.2. ТЕСТИРОВАНИЕ ПОТРЕБИТЕЛЕЙ *

Отбор комиссии потребителей связан с известными трудностями в нахождении отдельных представителей и всей группы. Будучи выбранными, члены группы будут сталкиваться с простыми, хорошо определенными альтернативами. Парные сравнения требуют принятия решения в пользу одной из двух альтернатив (и дозволяется объявлять об отсутствии предпочтений), что часто делается обычным путем, хотя иногда можно извлечь из респондента дополнительную информацию. Экспериментатор должен пытаться сбалансировать мешающие эффекты, такие, как порядок представления. Он должен принимать меры предосторожности по маскировке любых распознаваемых знаков на сравниваемых объектах, когда, например, требуется дегустация. Иногда лишь одна пара объектов может быть оценена одним экспертом из группы, так как работа по сравнению может быть велика, например при сравнении двух способов действий. Так как эксперименты с большим числом потребителей обыкновенно не доводятся до конца, неразумно проводить такие большие опыты. До сих пор мы интересовались в основном получением от группы экспертов общей картины их предпочтений и применение линейной модели было нежелательно; если же она используется, то можно работать методами, применимыми к неполным данным, как отмечено в гл. 4. Наконец, отметим хотя довольно очевидные, но важные вещи—мы отметим, что не стоит инструктировать членов комиссии потребителей, на чем они должны основывать свои предпочтения: хотелось бы, чтобы они были в естественном состоянии.

Модели для тестирования потребителей

Лица с плохой способностью различений не включаются в группы экспертов. Однако в тестировании потребителей главный интерес в том, чтобы узнать что-либо о доле тех лиц, которые не могут различить два продукта A и B. Такое знание полезно при решении того, продавать A, B или оба продукта. В тех ситуациях, когда можно выявить группу безразличных потребителей, сравнение A и B можно уточнить, отбросив эту группу (сравните [7]).

В первом приближении можно предположить [45], что π_0 — доля в совокупности тех, кто не может различить A и B; π_a и π_b — доли предпочитающих A и B ($\pi_a + \pi_0 + \pi_b = 1$). Пусть $P_i\,(j)$ — вероятность того, что потребитель типа i ($i = a, O, b$) предпочитает продукт j в любом случае, когда представляются объекты A и B. Мы можем положить j равным A, B или O, где O означает отсутствие предпочтения. Феррис предположил, что

$$P_a\,(A) = P_b(B) = 1,$$

$$P_a\,(B) = P_b(A) = P_a(O) = P_b(O) = 0,$$

* О психологических тестах см.: Лорд Ф. М. Отношение между тестовым баллом и исследуемой способностью [188, с. 54—90]. — *Прим. пер.*

$$P_0(O) = 1 - 2p,$$
$$P_0(A) = P_0(B) = p, \text{ где } 0 \leqslant p \leqslant \frac{1}{2}.$$

Это означает, что те, кто в действительности предпочитают A (или B), будут обыкновенно так и делать. Другие могут быть неспособны различить объекты или способны различить, но непоследовательны в своих предпочтениях; в одном случае они предпочитают A (или B) с вероятностью p, а в другом случае они не предпочитают с вероятностью $1 - 2p$.

Если n потребителей просят сравнить A и B дважды (с зашифрованными парами), то можно оценить все параметры (π_a, π_0, π_b, p) методом максимального правдоподобия.

7.3. ОБРАБОТКА РАВНЫХ РАНГОВ

Может привести в замешательство число способов обработки равных рангов. Мы можем предупредить их грубым вмешательством, требуя, чтобы эксперт мысленно бросал монету, если он не может принять решение, или позволяя ему для удобства суждения в действительности бросать монету. Эти два приема имеют то преимущество, что полученные с их помощью данные можно изучать методами предыдущих глав. Это не совсем верно, если для каждой пары очки просто делятся поровну. Тем не менее этот метод имеет преимущество независимости от более или менее произвольных способов преодоления дележа мест; в частности, физическая рандомизация ведет к различным выводам при одних и тех же экспериментальных данных. Другой возможностью является теоретическая рандомизация. Рассмотрим, например, коэффициент согласия и формулу (2.1.5) ([90], [78]):

$$u = \frac{2\sum\limits_{i\neq j}^{t}\binom{\alpha_{ij}}{2}}{\binom{n}{2}\binom{t}{2}} - 1. \qquad (7.3.1)$$

Если в результате n сравнений A_i и A_j появляется a «ничьих», деление поровну приводит к замене α_{ij} на $x + \frac{1}{2}a$ и α_{ji} на $n - x - \frac{1}{2}a$, где x — число ясных предпочтений объекта. Теоретическая рандомизация достигается заменой членов $\binom{\alpha_{ij}}{2}$ и $\binom{\alpha_{ji}}{2}$ в (7.3.1) на

$$\sum_{s=0}^{a}\binom{a}{s}2^{-a}\left\{\binom{x+s}{2}+\binom{n-x-s}{2}\right\}; \qquad (7.3.2)$$

это получается из-за случайного нарушения a ничьих, что может дать прирост, и взвешивания с учетом соответствующих вероятностей исходов. Выражение (7.3.2), как можно показать, больше

$$\left\{\binom{x+\frac{1}{2}a}{2}+\binom{n-x-\frac{1}{2}a}{2}\right\},$$

значения при равном делении ничьих по $\frac{1}{4}a$. Соответственно при теоретической рандомизации легко вычислить u.

При сравнении этих и других процедур обработки ничьих полезно делать различие между проверкой гипотез и оцениванием. Для $t = 2$, когда метод парных сравнений приводит к критерию знаков, показано, что игнорирование ничьих ведет в общем к более мощным критериям, чем их равное или случайное деление ([79], [118]). Этот интересный результат относится к анализу данных, содержащих ничьи; связанный с этим вопрос: можно ли разрешать ничьи? — должен быть решен в первую очередь. Кажется разумной стратегией инструктировать экспертов так, чтобы они выражали предпочтение лишь тогда, когда они вполне уверены. Экспериментальные результаты ухудшаются угадыванием. Однако на практике ситуация не вполне ясна, так как разрешение ничьих делает экспертов более небрежными [60].

Возвращаясь к оцениванию, мы ясно видим, что ничьими в экспериментальных данных не следует пренебрегать: при одном и том же числе ясных предпочтений два объекта A_i, A_j ближе друг к другу по ценности при бо́льшем количестве ничьих между ними. Четкие предпочтения в пользу A_i можно аргументировать тем, что A_i, вероятно, «выше ничьих». Это относится ко всем упомянутым выше «беспристрастным» методам, но может быть достигнуто и с помощью следующей модели [56]. Как обычно, пусть y_i, y_j — наблюдаемые ценности объектов A_i, A_j. Предположим, что эксперт объявляет объекты равными, если $|y_i - y_j| \leqslant \tau$ ($\tau \geqslant 0$); предпочитает A_i, если $y_i - y_j > \tau$, и предпочитает A_j, если $y_j - y_i > \tau$. В эту модель легко включить модель Тэрстоуна — Мостеллера и правило arcsin (4.2.2). Метод наименьших квадратов можно использовать не только для оценивания ценностей V_i ($i = 1, 2, \ldots, t$), но и для оценивания параметра порога (и, возможно, различного τ для различных экспертов; это утверждение можно проверить). Не разрушая равные ранги на самом деле, можно таким образом использовать исходные наблюдения*.

7.4. МЕТОДИКА ШЕФФЕ

Шеффе [122] предложил метод анализа экспериментов парных сравнений, включающих две особенности:

1) применение 7- или 9-точечной шкалы предпочтений;

2) введение специальных предположений о возможном эффекте порядка представления эксперту объектов в парах.

Вторая особенность требует пояснений. Хотя уже Фехнер всегда принимал во внимание, как важные, «пространственный» и «временной» эффекты, мы до сих пор не обращали на них внимания во всех способах анализа. Зная, что порядковые эффекты могут вызвать смещение результатов, мы имеем возможность сбалансировать их на стадии пла-

---

* Модель Гленна и Дэвида [56] относится к семейству «пороговых» моделей, таких, как модель Дэвидсона [258]. С ними конкурируют модели «вероятностного выбора». См. обзоры [244], [259] и [216], где обсуждаются такие модели. Свойства порогов см. в [153]. — *Прим. пер.*

а их разность — через 2 δ_{ij}:

$$\delta_{ij} = \frac{1}{2} (\mu_{ij} + \mu_{ji}). \qquad (7.4.2)$$

Так, $2\delta_{ij}$ есть разность «от порядка представления» в математическом ожидании предпочтений для A_i против A_j, в то время как τ_{ij} — среднее предпочтение для A_i против A_j, усредненное по двум упорядоченным парам. Мы замечаем, что

$$\tau_{ji} = -\tau_{ij}, \delta_{ji} = \delta_{ij}.$$

Мы можем также определить средний эффект порядка δ:

$$\frac{1}{2} t(t-1)\delta = \sum_{i<j} \delta_{ij}. \qquad (7.4.3)$$

Шеффе вводит теперь гипотезу о «разностности»*, которая есть просто гипотеза линейности, а именно, что существуют параметры $\alpha_1, \alpha_2, \ldots, \alpha_t$, для которых

$$\tau_{ij} = \alpha_i - \alpha_j \text{ для всех } i, j \ (i \neq j). \qquad (7.4.4)$$

Отклонение от разностности измеряется параметром γ_{ij}, так что в общем случае мы можем написать:

$$\tau_{ij} = \alpha_i - \alpha_j + \gamma_{ij}, \qquad (7.4.5)$$

где γ_{ij}, очевидно, удовлетворяет соотношениям $\gamma_{ji} = -\gamma_{ij}$; $\sum_j' \gamma_{ij} = 0$.

С другой стороны, α_i можно выразить через τ_{ij} так:

$$\alpha_i = \sum' \tau_{ij}/t. \qquad (7.4.6)$$

Оценки по методу наименьших квадратов получаются сразу. Начнем с

$$\hat{\mu}_{ij} = \sum_k x_{ijk}/r,$$

τ_{ij}, δ_{ij}, δ и α_i можно оценить по очереди из (7.4.1), (7.4.2), (7.4.3) и (7.4.6).

Очки s_{ijk} можно выразить так:

$$s_{ijk} = (\alpha_i - \alpha_j) + \gamma_{ij} + \delta + (\delta_{ij} - \delta) + e_{ijk},$$

где ошибки в правой части e_{ijk} есть случайные величины. Несмещенная оценка дисперсии σ^2 равна:

$$\hat{\sigma}^2 = S_e/[t(t-1)(r-1)].$$

Сумма квадратов ошибки S_e определяется в табл. 7.1, которая дает разложение общей суммы квадратов очков S_T. Можно построить приближенный критерий F-отношения.

* В оригинале «subtractivity»; можно ввести термин «сабтрактивность» по аналогии с «аддитивностью». — *Прим. пер.*

Таблица 7.1

Разбиение суммы квадратов очков

Источник	Суммы квадратов	Степени свободы	Ожидаемые средние квадраты
Главные эффекты	$S_\alpha = 2rt\sum_{i=1}^{t}\hat{\alpha}_i^2$	$\nu_\alpha = t-1$	$\sigma^2 + 2\,rt\sigma_\alpha^2$
Отклонения от разностности	$S_\gamma = S_\tau - S_\alpha$	$\nu_\gamma = \binom{t}{2} - (t-1)$	$\sigma^2 + r\,\sigma_\gamma^2$
Средние предпочтения	$S_\tau = r\sum_{i\neq j}^{t}\hat{\tau}_{ij}^2$	$\nu_\tau = \binom{t}{2}$	$\sigma^2 + r\sigma_\tau^2$
Порядковые эффекты	$S_\delta = S_\mu - S_\tau$	$\nu_\delta = \binom{t}{2}$	$\sigma^2 + r\sigma_\delta^2$
Средний	$S_\mu = r\sum_{i,j}\hat{\mu}_{ij}^2$	$\nu_\mu = 2\binom{t}{2}$	$\sigma^2 + r\,\sigma_\mu^2$
Ошибка	$S_e = S_T - S_\mu$	$\nu_e = t(t-1)(r-1)$	σ^2
Итого	$S_T = \sum_{i,j,k} s_{ijk}^2$	$\nu_T = rt(t-1)$	

Параметры нецентральности в последнем столбце табл. 7.1. определяются [141] следующим образом:

$$\sigma_\alpha^2 = \Sigma\alpha_i^2/\nu_\alpha,\quad \sigma_\gamma^2 = \sum_{i\neq j}\gamma_{ij}^2/\nu_\gamma,\quad \sigma_\tau^2 = \sum_{i\neq j}\tau_{ij}^2/\nu_\tau,$$

$$\sigma_\delta^2 = \sum_{i\neq j}\delta_{ij}^2/\nu_\delta \text{ и } \sigma_\mu^2 = \sum_{i\neq j}\mu_{ij}^2/\nu_\mu.$$

Шеффе снабдил анализ примерами и обсудил построение доверительных интервалов для разностей $\alpha_i - \alpha_j$.

Модель Тэрстоуна для дисперсионного анализа, предложенного Шеффе, была дана Юэре при следующих допущениях:

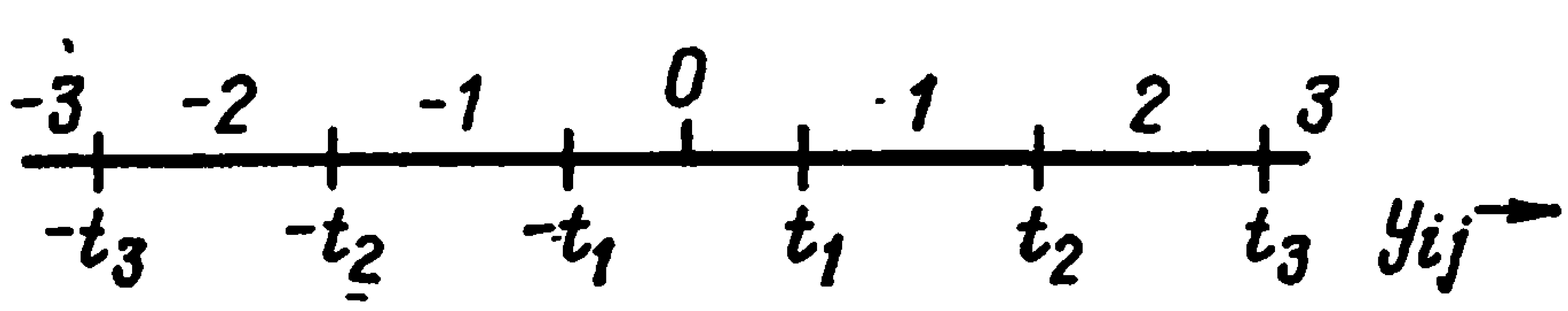
а) наблюдаемые ценности y_i ($i = 1, 2, \ldots, t$) объекта A_i нормальны $N(V_i, \sigma^2)$.

б) случайные величины y_{ij}, представляющие предпочтение A_i против A_j, когда они предъявляются в порядке (A_i, A_j), нормальны $N(V_{ij}, (1-\rho)\sigma^2)$. Величины V_{ij} можно разложить:

$$V_{ij} = (V_i - V_j) + \gamma'_{ij} + \delta'_{ij},$$

где γ'_{ij} есть отклонения от разности и δ'_{ij} — эффекты порядка, удовлетворяющие тем же условиям, которые были сформулированы выше для γ_{ij}, δ_{ij};

в) для $(2m+1)$-точечной системы шкала y_{ij} делится на $(2m+1)$ части и очки располагают для $m = 3$ следующим образом:



Вероятность $P_{ij}(s)$ того, что A_i наберет s очков против A_j в (A_i, A_j) поэтому равна:

$$P_{ij}(s) = \Phi\left(\frac{t_s - V_{ij}}{\sqrt{2(1-\rho)}\,\sigma}\right) - \Phi\left(\frac{t_{s-1} - V_{ij}}{\sqrt{2(1-\rho)}\,\sigma}\right),$$

где $\Phi(x)$ — функция нормированного нормального распределения. Если

$$\max|V_{ij}| \ll \sqrt{2(1-\rho)}\,\sigma,$$

то для малых различий между объектами можно найти приближенные соотношения между параметрами моделей Юэре и Шеффе. Юэре также рассмотрел мощность критерия Шеффе для главных эффектов и использовал ее для первого приближения к оптимальной системе очков.

7.5. МЕТОДИКА ГУТМАНА

Интересный метод превращения данных парных сравнений в количественные, который не использует специальной модели, был предложен Гутманом [69]. Как и раньше, пусть $a_{i\gamma}$ — очки эксперта γ для объекта A_i и пусть $a'_{i\gamma} = t - 1 - a_{i\gamma}$. Далее, пусть t_γ (u_γ) есть среднее очков v_i $(\Sigma v_i = 0)$ объектов, которые эксперт γ ранжирует выше (ниже), чем другие объекты, взвешенное на соответствующие относительные частоты

$$t_\gamma = \sum_i a_{i\gamma} v_i / F; \quad u_\gamma = \sum_i a'_{i\gamma} v_i / F, \tag{7.5.1}$$

где $F = \frac{1}{2} t(t-1)$. Также пусть

$$y_\gamma = \sum_i (v_i - t_\gamma)^2 a_{i\gamma} = \sum v_i^2 a_{i\gamma} - t_\gamma^2 F, \tag{7.5.2}$$

$$z_\gamma = \sum_i (v_i - u_\gamma)^2 a'_{i\gamma} = \sum v_i^2 a'_{i\gamma} - u_\gamma^2 F. \tag{7.5.3}$$

Сумму квадратов отклонений очков

$$W = n(t-1)\,\Sigma v_i^2 \tag{7.5.4}$$

можно разложить на 2 слагаемых:

$$R = \sum_\gamma (t_\gamma^2 + u_\gamma^2) F \tag{7.5.5}$$

и

$$S = \sum_\gamma (y_\gamma + z_\gamma), \tag{7.5.6}$$

суммы квадратов между экспертами и «внутри экспертов» соответственно.

Подход Гутмана сродни дискриминантному анализу и заключается в нахождении очков, минимизирующих вариацию «внутри экспертов», сравниваемых с вариацией группы экспертов в целом. Это значит, что мы делаем S как можно меньше или R как можно больше по сравнению с W. Если мы определим корреляционное отношение E через $E^2 = R/W$, то искомые очки v_i получают из уравнений, в которых $\partial E^2/\partial v_j$ приравнены к нулю, т. е. из уравнений стационарности

$$\frac{\partial R}{\partial v_j} = E^2 \frac{\partial W}{\partial v_j}. \qquad (7.5.7)$$

Производные от R можно вычислить из (7.5.5) и (7.5.1):

$$\frac{\partial R}{\partial v_j} = \frac{2}{F} \sum_i v_i \sum_\gamma (a_{i\gamma} a_{j\gamma} + a'_{ij} a'_{j\gamma}).$$

Из (7.5.4) — производные от W:

$$\frac{\partial W}{\partial v_j} = 2n(t-1) v_j.$$

Если мы положим

$$H_{ij} = \frac{1}{n(t-1)F} \sum_\gamma (a_{i\gamma} a_{j\gamma} + a'_{i\gamma} a'_{j\gamma}), \qquad (7.5.8)$$

то (7.5.7) можно переписать в виде

$$\sum_i v_i H_{ij} = E^2 v_j. \qquad (7.5.9)$$

Рассмотрим первое подтверждение того, что решение (7.5.9) удовлетворяет $\Sigma v_i = 0$. Суммируя обе части (7.5.9) по j и используя (7.5.8), мы получим

$$\sum_i v_i = E^2 \sum_j v_j,$$

так что $\Sigma v_i = 0$, если только не $E^2 = 1$, что практически можно исключить.

Всегда существует тривиальное решение (7.5.9), для которого E^2 формально равно 1. Это $v_j = 1$ для всех j, как легко проверить. Для получения нетривиального решения, обозначим $\mathbf{v}$ вектор-столбец v_i и пусть $\mathbf{H}$ $t \times t$ — симметричная матрица H_{ij}. Тогда (7.5.9) становится матричным уравнением

$$\mathbf{H}\mathbf{v} = E^2 \mathbf{v}, \qquad (7.5.10)$$

показывающим, что E^2 есть собственное число $\mathbf{H}$ с соответствующим собственным вектором $\mathbf{v}$. Поскольку мы хотим найти наибольшее корреляционное отношение, мы ищем наибольшее из нетривиальных собственных чисел. Численное решение (7.5.10) можно получить простым итеративным методом для получения собственных чисел и векторов[1]. Такие итерации сходятся, вообще говоря, к вектору, соответствующему наибольшему числу. Чтобы избежать сходимости к тривиальному решению (которое формально имеет наибольшее собственное число), исходные векторы должны удовлетворять условию $\Sigma v_i = 0$.

[1] См. примечание на с. 78.

Гутман полагал, что его подход ведет к очкам, которые для модели Тэрстоуна—Мостеллера лишь линейным преобразованием отличаются от тех, что получаются в 4.1. Однако этот метод применим в том случае, когда предположения модели не выполняются. Что касается недостатков, то следует заметить, что методика Гутмана—это скорее процедура оценивания, чем теория распределения.

Читатель может обратиться к работе Гутмана, в которой приведены дальнейшие обсуждения и обобщение подхода на ситуации, в которых сравниваются наборы, а не отдельные объекты.

7.6. ПРИЛОЖЕНИЯ

Следующий список классифицированной библиографии ограничивается работами, в которых представлены или анализируются экспериментальные результаты. Приведенные данные могли быть упрощены или специально получены для демонстрации процедуры, но случаев, когда они полностью придуманы, нет. Одни из работ уже упоминались, другие упоминаются здесь впервые. Мы не стремились при этом к полноте списка и упор в некоторой степени делался на недавние публикации. Названия групп по большей части не требуют пояснений. «Исследование предпочтений» включает и экспериментальные сравнения (например, сравнения запахов с помощью обоняния) и сравнение с помощью анкет. К работам под заголовком «Поведение в процессе выбора» относятся те, где основной интерес представляет самооценка, а не оценка сравниваемых объектов.

Тестирование потребителей: [4], [37], [45], [86].

Оптика: [25], [85], [107].

Аттестация персонала: [97], [100], [101].

Поведение в процессе выбора: [25], [26], [31], [38], [41], [73], [111], [125], [126], [135], [136].

Исследование предпочтений: [47], [56], [62], [63], [64], [65], [67], [69], [93], [102], [110], [129], [135], [145].

Спорт: [87], [108c], [109].

Дегустация: [7], [8], [14], [16], [17], [57], [58], [59], [81], [86], [122].

Будет кстати добавить здесь несколько работ, относящихся к технике выполнения экспериментов парных сравнений и выбора экспертов. В них освещается много практически важных вопросов, но не всегда приводятся оригинальные данные.

Технические приемы: [5], [7], [13], [20], [22], [41], [60], [61], [116], [142], [146].

УПРАЖНЕНИЯ

7.1. Покажите, что верный ответ в двойном-тройном тесте получается, когда либо

$$y_2 - y_1' < 0 \text{ и } y_1' + y_2 - 2\,y_1 < 0,$$

либо

$$y_2 - y_1' > 0 \text{ и } y_1' + y_2 - 2\,y_1 > 0.$$

Следовательно, верно (7.1.3).

7.2. Если y_1, y_1', y_2 — независимые равномерно распределенные случайные величины, первые две — на интервале $(\mu_1 - \sqrt{3}\,\sigma,\ \mu_1 + \sqrt{3}\,\sigma)$ и последняя — на $(\mu_2 - \sqrt{3}\,\sigma,\ \mu_2 + \sqrt{3}\,\sigma)$, то показать, что

$$
\begin{array}{ll}
P_P = \dfrac{1}{2} + \dfrac{\delta}{2\sqrt{3}} - \dfrac{\delta^2}{24} & 0 \leqslant \delta < 2\sqrt{3}, \\
1 = & 2\sqrt{3} \leqslant \delta; \\
P_\Delta = \dfrac{1}{3} + \dfrac{\delta^2}{12} - \dfrac{\delta^3}{48\sqrt{3}} & 0 \leqslant \delta < 2\sqrt{3}, \\
= -\dfrac{1}{3} + \dfrac{\delta}{\sqrt{3}} - \dfrac{\delta^2}{12} + \dfrac{\delta^3}{144\sqrt{3}} & 2\sqrt{3} \leqslant \delta < 4\sqrt{3}, \\
= 1 & 4\sqrt{3} \leqslant \delta; \\
P_D = \dfrac{1}{2} + \dfrac{\delta^2}{12} - \dfrac{7\delta^3}{288\sqrt{3}} & 0 \leqslant \delta < 2\sqrt{3}, \\
= \dfrac{1}{3} + \dfrac{\delta}{2\sqrt{3}} - \dfrac{\delta^2}{24} + \dfrac{\delta^3}{288\sqrt{3}} & 2\sqrt{3} \leqslant \delta < 4\sqrt{3}, \\
= 1 & 4\sqrt{3} \leqslant \delta;
\end{array}
$$

где $\delta = (\mu_2 - \mu_1)/\sigma$ [142].

Примечание к первому изданию

В гл. 2 мы рассмотрели (§ 2.2) лишь один из возможных подходов к перечислению. В моей недавней записке обращалось внимание на то, что использованный метод был предугадан Рапопортом [119] и Ландау [95] в работах по теории склевывания у кур. Некоторые другие вопросы, представляющие более теоретический интерес, таковы.

Сколько различных $2^{\frac{1}{2}t(t-1)}$ таблиц предпочтений различаются существенно (т. е. остаются различными, когда объекты неразличимы)? Это число для $t > 4$ больше, чем число размещений табл. 2.1, так как два или более существенно различных исхода могут приводить к различным размещениям ([95], [32]).

Как много существенно различных турниров можно сделать в случае, когда играется лишь часть возможных игр?

К задачам этого типа можно с пользой применить (но не во всех случаях ясно, как) методы теории графов. Читатель отсылается* к обширной библиографии по теории графов [105] и обзору [71].

Примечание ко второму изданию

Доктора Мун и Мозер обратили мое внимание на исторически интересную работу: E. Z e r m e l o. Die Berechnung der Turnier—Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. — *Math. Zeits*, 1929, *H*. 29, 436—460 («Вычисление результатов турнира как задача максимизации из теории вероятностей»). Эта работа касается оценивания силы игроков в круговом турнире на основе неполных данных и фактически использует для этой цели метод максимального правдоподобия и модель Брэдли—Терри.

* См. также [287] и [218]. — *Прим. пер.*

ПРИЛОЖЕНИЕ

Таблица 1

Критические значения для суммы квадратов очков $\sum_{i=1}^{t} a_i^2$

(основана на [138], с разрешения автора)

$t=3$		$t=3$		$t=4$		$t=5$	
$n=3$	45 —	$n=12$	488 518	$n=2$	52 56	$n=2$	104 110
4	72 80	13	569 603	3	106 114	3	216 228
5	101 113	14	660 692	4	176 188	$t=6$	
6	140 150	15	747 779	5	266 282	$n=1$	55 —
7	185 197	16	846 882	6	372 392	$t=7$	
8	234 248	17	945 989	7	498 522	$n=1$	85 89
9	285 305	18	1058 1098			$t=8$	
10	350 372	19	1179 1211			$n=1$	126 132
11	417 441	20	1296 1346				

Верхнее число в каждой паре относится к $\alpha=0{,}05$, нижнее—к $\alpha=0{,}01$. Тире указывает на то, что значимая точка невозможна.

Таблица 2

Критические значения для разности между очками двух заранее заданных объектов

(из [129], с разрешения авторов и издателя)

Объем эксперимента		$\alpha=0,01$		$\alpha=0,05$	
n	t	односторонний критерий m'_c	двусторонний критерий m_c	односторонний критерий m'_c	двусторонний критерий m_c
1	2	3	4	5	6
1	$\leqslant 4$	Нет значимых точек		Нет значимых точек	
1	5	4	невозможно	4	4
1	6	5	5	4	4
1	7	5	5	4	5
1	8	5	6	4	5
1	9	6	6	4	5
1	10	6	7	5	5
1	11	6	7	5	6
1	12	7	7	5	6
1	13	7	7	5	6
1	14	7	8	5	6
1	15	7	8	5	6
1	16	7	8	6	6
2	3	Нет значимых точек		4	4
2	4	5	6	4	5
2	5	6	6	5	5
3	3	6	6	4	5
3	4	6*	7	5	6
4	3	6**	7	5	6
4	4	7	8	6	6
Все большие значения n или t		m'_c = наименьшее целое $\geqslant 2,33\,\sigma+0,5$	m_c = наименьшее целое $\geqslant 2,56\sigma+0,5$	m'_c = наименьшее целое $\geqslant 1,64\,\sigma+0,5$ $\left[\sigma=\left(\frac{1}{2}nt\right)^{1/2}\right]$	m_c = наименьшее целое $\geqslant 1,96\sigma+0,5$

* $P_{436}=0,0103$.

** $P_{346}=0,01001$.

Таблица 3

Критические значения $R_{\beta(\alpha)}$ для множественного критерия сравнений размахов*

(из [129], с разрешения авторов и издателя)

t	n	$\alpha=0,01$		$\alpha=0,05$	
		$R_\beta(\alpha)$	β	$R_\beta(\alpha)$	β
1	2	3	4	5	6
3	1	Нет	Нет	Нет	Нет
3	2	Нет	Нет	Нет	Нет
3	3	6	0,01	Нет	Нет
3	4	7	0,01	6	0,05
3	5	8	0,01	7	0,04
3	6	9	0,01	8	0,03
3	7	10	0,01	8	0,05
3	8	10	0,01	9	0,03
3	9	11	0,01	9	0,05
3	10	12	0,01	10	0,03
4	1	Нет	Нет	Нет	Нет
4	2	6	0,01	Нет	Нет
4	3	8	0,005	7	0,03
4	4	9	0,01	8	0,03
4	5	10	0,01	9	0,03
4	6	11	0,01	9	0,06
4	7	12	0,01	10	0,05
4	8	13	0,01	11	0,04
5	1	Нет	Нет	Нет	Нет
5	2	7	0,0154	6	0,08
5	3	9	0,01	8	0,04
5	4	11	0,005	9	0,05
5	5	12	0,01	10	0,05
6	1	Нет	Нет	5	0,0586
7	1	Нет	Нет	6	0,0205
8	1	7	0,0068	6	0,0738

* Эти критические значения $R_\beta(\alpha)$ выбирались скорее так, чтобы сделать β возможно ближе к α, чем таким образом, чтобы $R_\beta(\alpha)$ обеспечивало $\beta \leqslant \alpha$, как определено в (3.7.1).

Таблица 4

Наименьшее число повторений, требующееся для обеспечения с вероятностью, не меньшей наперед заданной вероятности P', отбора лучшего объекта в случае, когда

$$\Pr\,(A_{(t)} \to A_{(i)}) \geqslant \pi \qquad (i=1,\,2,\;\ldots,\;t-1),$$

$$\Pr\,(A_{(i)} \to A_{(j)}) = \frac{1}{2} \qquad (i \neq j;\; i,\; j=1,2,\;\ldots,\;t-1).$$

(а) $P'=0{,}75$

t \ π	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95
2	45	11	5	3	1	1	1	1	1
3	69	17	8	4	3	2	1	1	1
4	71	18	8	5	3	2	2	1	1
5	68	17	8	4	3	2	2	1	1
6	65	16	7	4	3	2	2	1	1
7	61	15	7	4	3	2	2	1	1
8	58	15	7	4	3	2	1	1	1
9	54	14	6	4	2	2	1	1	1
10	52	13	6	4	2	2	1	1	1
12	47	12	5	3	2	2	1	1	1
14	43	11	5	3	2	2	1	1	1
16	39	10	5	3	2	1	1	1	1
18	37	9	4	3	2	1	1	1	1
20	34	9	4	3	2	1	1	1	1

(б) $P'=0{,}90$

t \ π	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95
2	163	41	17	9	7	5	3	1	1
3	165	41	18	10	6	4	3	2	1
4	150	37	16	9	6	4	3	2	1
5	135	33	15	8	5	4	3	2	1
6	122	30	13	7	5	3	2	2	1
7	112	28	12	7	4	3	2	2	1
8	103	26	11	6	4	3	2	2	1
9	95	24	11	6	4	3	2	1	1
10	89	22	10	6	4	3	2	1	1
12	79	20	9	5	3	2	2	1	1
14	71	18	8	5	3	2	2	1	1
16	64	16	7	4		2	2	1	1
18	59	15	7		3	2	1	1	1
20	55	14	6	4		2	1	1	1

(в) $P' = 0{,}95$

t \ π	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95
2	269	67	29	15	9	7	5	3	1
3	243	60	26	15	9	6	4	3	2
4	212	52	23	12	8	5	4	3	2
5	186	46	20	11	7	5	3	3	2
6	166	41	18	10	6	4	3	2	2
7	150	37	16	9	6	4	3	2	2
8	137	34	15	8	5	4	3	2	2
9	126	31	14	8	5	3	2	2	1
10	117	29	13	7	5	3	2	2	1
12	102	26	11	6	4	3	2	2	1
14	91	23	10	6	4	3	2	1	1
16	82	21	9	5	3	2	2	1	1
18	75	19	9	5	3	2	2	1	1
20	69	17	8	5	3	2	2	1	1

(г) $P' = 0{,}99$

t \ π	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95
2	537	133	57	31	19	13	9	5	3
3	433	106	45	24	15	10	7	5	3
4	358	88	38	20	12	8	6	4	3
5	306	75	33	18	11	7	5	4	3
6	267	66	29	16	10	6	4	3	2
7	238	59	26	14	9	6	4	3	2
8	214	53	23	13	8	5	4	3	2
9	195	48	21	12	7	5	3	2	2
10	180	45	20	11	7	4	3	2	2
12	155	39	17	9	6	4	3	2	2
14	137	34	15	8	5	4	3	2	1
16	123	31	14	8	5	3	2	2	1
18	112	28	12	7	4	3	2	2	1
20	102	26	11	6	4	3	2	2	1

Таблица 5

Значения ν для решающего правила $\mathscr{R}$ из § 6.3

(a) $P^* = 0{,}75$

n \ t	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	1	2	2	2	3	3	3	4	4	4	4	5	5	5	5	5	6	6
2	0	2	2	3	3	4	4	5	5	5	6	6	6	7	7	7	8	8	8
3	1	2	3	4	4	5	5	6	6	7	7	7	8	8	9	9	9	10	10
4	2	2	3	4	5	5	6	7	7	8	8	9	9	10	10	10	11	11	12
5	1	3	4	5	5	6	7	7	8	9	9	10	10	11	11	12	12	13	13
6	2	3	4	5	6	7	7	8	9	9	10	11	11	12	12	13	13	14	14
7	1	3	4	5	6	7	8	9	9	10	11	11	12	13	13	14	14	15	15
8	2	4	5	6	7	8	9	9	10	11	12	12	13	14	14	15	15	16	16
9	3	4	5	6	7	8	9	10	11	12	12	13	14	14	15	16	16	17	17
10	2	4	5	7	8	9	10	10	11	12	13	14	14	15	16	16	17	18	18
11	3	4	6	7	8	9	10	11	12	13	14	14	15	16	17	17	18	19	19
12	2	4	6	7	8	9	10	11	12	13	14	15	16	17	17	18	19	19	20
13	3	5	6	7	9	10	11	12	13	14	15	16	16	17	18	19	20	20	21
14	2	5	6	8	9	10	11	12	13	14	15	16	17	18	19	19	20	21	22
15	3	5	7	8	9	11	12	13	14	15	16	17	18	18	19	20	21	22	23
16	2	5	7	8	10	11	12	13	14	15	16	17	18	19	20	21	22	22	23
17	3	5	7	9	10	11	12	14	15	16	17	18	19	20	21	21	22	23	24
18	2	5	7	9	10	12	13	14	15	16	17	18	19	20	21	22	23	24	25
19	3	5	7	9	10	12	13	14	16	17	18	19	20	21	22	23	24	24	25
20	4	6	8	9	11	12	14	15	16	17	18	19	20	21	22	23	24	25	26
25	3	6	8	10	12	14	15	17	18	19	20	22	23	24	25	26	27	28	29
30	4	7	9	11	13	15	17	18	20	21	22	24	25	26	27	29	30	31	32
35	3	7	10	12	14	16	18	20	21	23	24	26	27	28	30	31	32	33	34
40	4	8	11	13	15	17	19	21	23	24	26	27	29	30	32	33	34	36	37
45	5	8	11	14	18	18	20	22	24	26	27	29	31	32	34	35	36	38	39
50	4	9	12	15	17	19	21	23	25	27	29	31	32	34	35	37	38	40	41
60	6	10	13	16	19	21	23	26	28	30	32	33	35	37	39	40	42	44	45
70	6	10	14	17	20	23	25	28	30	32	34	36	38	40	42	44	45	47	49
80	6	11	15	18	22	24	27	30	32	34	37	39	41	43	45	47	48	50	52
90	6	12	16	20	23	26	29	31	34	36	39	41	43	45	47	49	51	53	55
100	6	12	17	21	24	27	30	33	36	38	41	43	46	48	50	52	54	56	58

(б) $P^* = 0,90$

n \ t	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	2	2	3	3	4	4	4	5	5	5	6	6	6	6	7	7	7	7
2	2	3	3	4	5	5	6	6	7	7	8	8	8	9	9	9	10	10	10
3	3	3	4	5	6	6	7	8	8	9	9	10	10	11	11	12	12	12	13
4	2	4	5	6	7	7	8	9	9	10	11	11	12	12	13	13	14	14	15
5	3	4	5	6	7	8	9	10	11	11	12	12	13	14	14	15	15	16	16
6	4	5	6	7	8	9	10	11	12	12	13	14	14	15	16	16	17	17	18
7	3	5	6	8	9	10	11	12	12	13	14	15	16	16	17	18	18	19	20
8	4	5	7	8	9	10	11	12	13	14	15	16	17	17	18	19	20	20	21
9	3	6	7	9	10	11	12	13	14	15	16	17	18	18	19	20	21	21	22
10	4	6	8	9	10	12	13	14	15	16	17	18	19	19	20	21	22	23	23
11	5	6	8	10	11	12	13	15	16	17	18	18	19	20	21	22	23	24	24
12	4	7	8	10	11	13	14	15	16	17	18	19	20	21	22	23	24	25	26
13	5	7	9	10	12	13	15	16	17	18	19	20	21	22	33	24	25	26	27
14	4	7	9	11	12	14	15	16	18	19	20	21	22	23	24	25	26	27	28
15	5	8	9	11	13	14	16	17	18	19	21	22	23	24	25	26	27	28	29
16	6	8	10	12	13	15	16	18	19	20	21	22	24	25	26	27	28	29	29
17	5	8	10	12	14	15	17	18	19	21	22	23	24	25	26	27	28	29	30
18	6	8	10	12	14	16	17	19	20	21	23	24	25	26	27	28	29	30	31
19	5	8	11	13	14	16	18	19	21	22	23	24	26	27	28	29	30	31	32
20	6	9	11	13	15	17	18	20	21	22	24	25	26	27	29	30	31	32	33
25	7	10	12	15	17	18	20	22	24	25	27	28	29	31	32	33	34	36	37
30	8	11	13	16	18	20	22	24	26	28	29	30	32	34	35	36	38	39	40
35	7	11	15	17	20	22	24	26	28	30	31	33	35	36	38	39	41	42	44
40	8	12	16	18	21	23	26	28	30	32	34	35	37	39	40	42	44	45	47
45	9	13	16	19	22	25	27	29	32	34	36	37	39	41	43	45	46	48	49
50	10	14	17	21	23	26	29	31	33	36	38	39	42	43	45	47	49	50	52
60	10	15	19	23	26	29	31	34	37	39	41	43	46	48	50	52	53	55	57
70	12	16	21	24	28	31	34	37	39	42	44	46	49	51	54	56	58	60	62
80	12	17	22	26	30	33	36	39	42	45	48	50	53	55	57	60	62	64	66
90	12	18	23	28	31	35	38	42	45	48	50	53	56	58	61	63	65	68	70
100	12	19	25	29	33	37	41	44	47	50	53	56	59	61	64	67	69	71	74

(в) $P^* = 0{,}95$

n \ t	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	2	3	3	4	4	5	5	5	6	6	6	7	7	7	8	8	8	8
2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	11	11	11	12
3	3	4	5	6	7	7	8	9	9	10	10	11	12	12	13	13	14	14	14
4	4	5	6	7	8	9	9	10	11	11	12	13	13	14	15	15	16	16	17
5	3	5	6	8	9	10	10	11	12	13	14	14	15	16	16	17	17	18	19
6	4	6	7	8	9	10	11	12	13	14	15	16	16	17	18	18	19	20	20
7	5	6	8	9	10	11	12	13	14	15	16	17	18	16	19	20	21	21	22
8	4	7	8	10	11	12	13	14	15	16	17	18	19	20	21	21	22	23	24
9	5	7	9	10	12	13	14	15	16	17	18	19	20	21	22	23	23	24	25
10	6	7	9	11	12	14	15	16	17	18	19	20	21	22	23	24	25	25	26
11	5	8	10	11	13	14	16	17	18	19	20	21	22	23	24	25	26	27	28
12	6	8	10	12	13	15	16	17	19	20	21	22	23	24	25	26	27	28	29
13	5	8	11	12	14	15	17	18	19	21	22	23	24	25	26	27	28	29	30
14	6	9	11	13	14	16	18	19	20	21	23	24	25	26	27	28	29	30	31
15	7	9	11	13	15	17	18	20	21	22	23	25	26	27	28	29	30	31	32
16	6	9	12	14	15	17	19	20	22	23	24	26	27	28	29	30	31	32	33
17	7	10	12	14	16	18	19	21	22	24	25	26	28	29	30	31	32	33	34
18	6	10	12	14	16	18	20	21	23	24	26	27	28	30	31	32	33	34	35
19	7	10	13	15	17	19	20	22	24	25	26	28	29	30	32	33	34	35	36
20	8	10	13	15	17	19	21	23	24	26	27	29	30	31	32	34	35	36	37
25	9	12	15	17	19	21	23	25	27	29	30	32	33	35	36	38	39	40	42
30	8	13	16	19	21	23	26	28	30	31	33	35	37	38	40	41	43	44	46
35	9	14	17	20	23	25	28	30	32	34	36	38	40	41	43	45	46	48	49
40	10	15	18	22	24	27	30	32	34	36	38	40	42	44	46	48	49	51	53
45	11	16	20	23	26	29	31	34	36	39	41	43	45	47	49	51	52	54	56
50	12	17	21	24	27	30	33	36	38	41	43	45	47	49	51	53	55	57	59
60	12	18	23	26	30	33	36	39	42	44	47	49	52	54	56	58	60	62	64
70	14	20	24	29	32	36	39	42	45	48	51	53	56	58	61	63	65	67	70
80	14	21	26	31	35	38	42	45	48	51	54	57	60	62	65	67	70	72	74
90	16	22	28	32	37	41	44	48	51	54	58	60	63	66	69	71	74	76	79
100	16	23	29	34	39	43	47	51	54	57	61	64	67	70	73	75	78	81	83

(г) $P^* = 0{,}99$

n \ t	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	2	3	4	4	5	6	6	7	7	7	8	8	9	9	9	10	10	10
2	2	4	5	6	7	8	8	9	9	10	11	11	12	12	13	13	13	14	14
3	3	5	6	8	9	9	10	11	12	12	13	14	14	15	15	16	16	17	18
4	4	6	7	9	10	11	12	13	13	14	15	16	16	17	18	18	19	20	20
5	5	7	8	10	11	12	13	14	15	16	17	18	18	19	20	21	21	22	23
6	6	7	9	11	12	13	14	15	16	17	18	19	20	21	22	23	23	24	25
7	5	8	10	12	13	14	16	17	18	19	20	21	22	23	23	24	25	26	27
8	6	9	11	12	14	15	17	18	19	20	21	22	23	24	25	26	27	28	29
9	7	9	11	13	15	16	18	19	20	21	22	24	25	26	27	28	29	29	30
10	8	10	12	14	16	17	19	20	21	22	24	25	26	27	28	29	30	31	32
11	7	10	13	15	16	18	19	21	22	24	25	26	27	28	29	30	32	33	34
12	8	11	13	15	17	19	20	22	23	25	26	27	28	30	31	32	33	34	35
13	9	11	14	16	18	19	21	23	24	26	27	28	30	31	32	33	34	35	36
14	8	12	14	16	18	20	22	24	25	27	28	29	31	32	33	34	36	37	38
15	9	12	15	17	19	21	23	24	26	28	29	30	32	33	34	36	37	38	39
16	10	12	15	18	20	22	23	25	27	28	30	31	33	34	35	37	38	39	40
17	9	13	16	18	20	22	24	26	28	29	31	32	34	35	37	38	39	40	42
18	10	13	16	19	21	23	25	27	28	30	32	33	35	36	38	39	40	42	43
19	9	14	17	19	21	24	26	27	29	31	33	34	36	37	39	40	41	43	44
20	10	14	17	20	22	24	26	28	30	32	33	35	37	38	40	41	43	44	45
25	11	16	19	22	25	27	29	31	34	36	37	39	41	43	44	46	48	49	51
30	12	17	21	24	27	30	32	34	37	39	41	43	45	47	49	50	52	54	55
35	13	19	22	26	29	32	35	37	40	42	44	46	49	51	52	54	56	58	60
40	14	20	24	28	31	34	37	40	42	45	47	50	52	54	56	58	60	62	64
45	15	21	25	29	33	36	39	42	45	48	50	53	55	57	60	62	64	66	68
50	16	22	27	31	35	38	41	45	47	50	53	55	58	60	63	65	67	69	72
60	18	24	29	34	38	42	45	49	52	55	58	61	64	66	69	71	74	76	78
70	20	26	32	37	41	45	49	53	56	59	63	66	69	71	74	77	80	82	85
80	20	28	34	39	44	48	52	56	60	64	67	70	73	76	79	82	85	88	90
90	22	30	36	42	47	51	56	60	64	67	71	74	78	81	84	87	90	93	96
100	24	31	38	44	49	54	59	63	67	71	75	78	82	85	89	92	95	98	101

БИБЛИОГРАФИЯ

1. Abelson R. M. and Bradley R. A. (1954). A2×2 factorial with paired comparisons. *Biometrics*, **10**, 487—502.
2. Alway G. G. (1962). Matrices and sequences. *Math. Gaz.*, **46**, 208—213.
3. Anderson T. W. (1958). Introduction to multivariate statistical analysis. New York: Wiley. Андерсон Т. Введение в многомерный статистический анализ. М., Физматгиз, 1963, 500 с.
4. Baker G. A., Amerine M. A., Roessler E. B. and Filipello F. (1960). The nonspecificity of differences in taste testing for preference. *Food Research*, **25**, 810—816.
5. Baker G. A., Mrak Vera and Amerine M. A. (1958). Errors of the second kind in an acid threshold test. *Food Research*, **23**, 150—154.
6. Bechhofer R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.*, **35**, 16—39.
7. Bliss C. I. (1960). Some statistical aspects of preference and related tests. *Applied Statistics*, **9**, 8—19.
8. Bliss C. I., Greenwood Mary L. and White Edna S. (1956). A rankit analysis of paired comparisons for measuring the effect of sprays on flavor. *Biometrics*, **12**, 381—403.
9. Block H. D. and Marschak J. (1960). Random orderings and stochastic theories of responses. Contributions to probability and statistics, ed. I. Olkin et al. Stanford: University Press, 97—132.
10. Bock R. D. (1956). The selection of judges for preference testing. *Psychometrika*, **21**, 349—366.
11. Bock R. D. (1958). Remarks on the test of significance for the method of paired comparisons. *Psychometrika*, **23**, 323—334.
12. Bose R. C. (1956). Paired comparison designs for testing concordance between judges. *Biometrika*, **43**, 113 — 121.
13. Bradley R. A. (1953). Some statistical methods in taste testing and quality evaluation. *Biometrics*, **9**, 22—38.
14. Bradley R. A. (1954). Incomplete block rank analysis: On the appropriateness of the model for a method of paired comparisons. *Biometrics*, **10**, 375—390.
15. Bradley R. A. (1954). The rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika*, **41**, 502—537.
16. Bradley R. A. (1955). Rank analysis of incomplete block designs. III. Some lagre-sample results on estimation and power for a method of paired comparisons. *Biometrika*, **42**, 450—470.
17. Bradley R. A. and Terry M. E. (1952). The rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, **39**, 324—345.
18. Brunk H. D. (1960). Mathematical models for ranking from paired comparisons. *J. Am. Statist. Ass.*, **55**, 503—520.
19. Burros R. H. and Gibson W. A. (1954). A solution for Case III of the law of comparative judgment. *Psychometrika*, **19**, 57—64.

46. F i s h e r Sir R. A. and Y a t e s F. (1957). Statistical Tables for Biological, Agricultural and Medical Research, 5 th ed. London: Oliver and Boyd.
47. F l e c k e n s t e i n Mary, F r e u n d R. A. and J a c k s o n J. E. (1958). A paired comparison test of typewriter carbon papers. *Tappi*, **41**, 128—130.
48. F o r d L. R. Jr. (1957). Solution of a ranking problem from binary comparisons. *Am. Math. Monthly*, **64**, 28—33.
49. F o r d L. R. Jr. and J o h n s o n S. M. (1959). A tournament problem. *Am. Math. Monthly*, **66**, 387—389.
50. F r e u n d J. E. (1956). Round Robin Mathematics. *Am. Math. Monthly*, **63**, 112—114.
51. G e r a r d H. B. and S h a p i r o H. N. (1958). Determining the degree of inconsistency in a set of paired comparisons. *Psychometrika*, **23**, 33—46.
52. G i b s o n W. A. (1953). A least-squares solution for Case IV of the law of comparative judgement. *Psychometrika*, **18**, 15—21.
53. G i l b e r t E. N. (1961). Design of mixed doubles tournament. *Am. Math. Monthly*, **68**, 124—131.
54. G i r s c h i c k M. A., M o s t e l l e r F. and S a v a g e L. J. (1946). Unbiased estimates of certain binomial sampling problems with applications. *Ann. Math. Statist.*, **17**, 13—23.
55. G l e n n W. A. (1960). A comparison of the effectiveness of tournaments. *Biometrika*, **47**, 253—262.
56. G l e n n W. A. and D a v i d H. A. (1960). Ties in paired — comparison experiments using a modified Thurstone—Mosteller method. *Biometrics*, **16**, 86—109.
57. G r e e n w o o d Mary L., P o t g i e t e r Martha and B l i s s C. I. (1951). The effect of certain prefreezing treatments on the quality of eight varieties of cultivated highbush blueberries. *Food Research*, **16**, 154—160.
58. G r i d g e m a n N. T. (1955). The Bradley—Terry probability model and preference testing. *Biometrics*, **11**, 335—343.
59. G r i d g e m a n N. T. (1958). Application of quantal-response theory to the cross-comparison of taste-stimuli intensities. *Biometrics*, **14**, 548—557.
60. G r i d g e m a n N. T. (1959). Pair comparison, with and without ties. *Biometrics*, 15, 382—388.
61. G r i d g e m a n N. T. (1960). Statistics and taste testing. *Applied Statistics*, **9**, 103—112.
62. G u i l f o r d J. P. (1954). Psychometric methods, 2nd ed. New York: McGraw-Hill.
63. G u l l i k s e n H. (1954). A least-squares solution for paired comparisons with incomplete data. *Psychometrika*, **19**, 117—139.
64. G u l l i k s e n H. (1956). Measurement of subjective values. *Psychometrika*, **21**, 229—244.
65. G u l l i k s e n H. (1958). Comparatal dispersion, a measure of accuracy of judgment. *Psychometrika*, **23**, 137—150.
66. G u l l i k s e n H. and T u c k e r L. R. (1961). A general procedure for obtaining paired comparisons from multiple rank orders. *Psychometrika*, **26**, 173—183.
67. G u l l i k s e n H. and T u k e y J. W. (1958). Reliability for the law of comparative judgment. *Psychometrika*, **23**, 95—110.
68. G u p t a S. S. and S o b e l M. (1960). Selecting a subset containing the best of several binomial populations. Contributions to probability and statistics, ed. I. Olkin et al. Stanford: University Press, 224—248.
69. G u t t m a n L. (1946). An approach for quantifying paired comparisons and rank order. *Ann. Math. Statist.*, **17**, 144—163.
70. H a l d A. (1952). Statistical Tables and Formulas. New York: Wiley.
71. H a r a r y F. (1960). Unsolved problems in the enumeraiton of graphs *Math. Inst. Hung. Acad. Sci.*, 5*a*, 63—95.
72. H a r p e r R. (1955). Fundamental problems in the subjective appraisal of foodstuffs. *Applied Statistics*, **4**, 145—161.
73. H a r r i s W. P. (1957). A revised law of comparative judgement. *Psychometrika*, **22**, 189—198.

74. H a r t e r H. L. (1961). Expected values of normal order statistics. *Biometrika*, **48**, 151—165.
75. H a r t l e y H. O. (1950). The use of range in analysis of variance. *Biometrika*, **37**, 271—280.
76. H a r t r e e D. R. (1952). Numerical Analysis. Oxford, Clarendon Press.
77. Harvard University Computation Laboratory (1955). Tables of the Cumulative Binomial Distribution. Cambridge, Mass.: Harvard University Press.
78. V a n d e r H e i d e n J. A. (1952). On a correction term in the method of paired comparisons. *Biometrika*, **39**, 211—212.
79. H e m e l r i j k J. (1952). A theorem on the sign test when ties are present. *Indag. Math.*, **14**, 322—326.
80. H o e f f d i n g W. (1956). On the distribution of the number of successes in independent trials. *Ann. Math. Statist.*, **27**, 713—721.
81. H o p k i n s J. W. (1954). Incomplete block rank analysis: some taste test results. *Biometrics*, **10**, 391—399.
82. H o p k i n s J. W. (1954). Some observations on sensitivity and repeatability of triad taste difference tests. *Biometrics*, **10**, 521—530.
83. H o p k i n s J. W. and G r i d g e m a n N. T. (1955). Comparative sensitivity of pair and triad flavor intensity difference tests. *Biometrics*, **11**, 63—68.
84. H u b e r P. J. (1963). A remark on a paper of Trawinski and David entitled: «Selection of the best treatment in a paired-comparison experiment». *Ann. Math. Statist*, **34**, 92—94.
85. J a c k s o n J. E. and F l e c k e n s t e i n Mary (1957). An evaluation of some statistical techniques used in the analysis of paired comparison data. *Biometrics*, **13**, 51—64.
86. J o n e s L. V. and B o c k R. D. (1957). Methodology of preference measurement. Final report, Quartermaster Food and Container Institute for the Armed Forces, 1—202.
87. K a i s e r H. F. (1959). Contributions to the method of paired comparisons. Unpublished memorandum, Univ. of Illinois.
88. K e m p t h o r n e O. (1953). A class of experimental designs using blocks of two plots. *Ann. Math. Statist.*, **24**, 76—84.
89. K e n d a l l M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics*, **11**, 43—62.
90. K e n d a l l M. G. (1962). Rank correlation methods, 3rd ed. London: Griffin. К е н д э л М. Ранговые корреляции. М., «Статистика», 1975, 214 с. (пер. с англ. изд. 1970 г.).
91. K e n d a l l M. G., S t u a r t A. (1963). The Advanced Theory of Statistics, vol. 1, 2nd ed., London: Griffin. К е н д а л л М., С т ь ю а р т А. Теория распределений. М., «Наука», 1966.
92. K e n d a l l M. G., S t u a r t A. (1967). The Advanced Theory of Statistics, vol. 2, 2nd ed., London: Griffin. К е н д а л л М., С т ь ю а р т А. Статистические выводы и связи. М., «Наука», 1973.
93. K e n d a l l M. G. and S m i t h B. Babington (1940). On the method of paired comparisons. *Biometrika*, **31**, 324—345.
94. Kraitchik M. (1953). Mathematical Recreations. New York: Dover Publications, 230—237.
95. L a n d a u H. G. (1951). On dominance relations and the structure of animal societies: I. Effect of inherent characteristics. *Bull. Math. Biophysics*, **13**, 1—19.
96. L a n d a u H. G. (1953). On dominance relations and the structure of animal societies: III. The condition for a score structure. *Bull. Math. Biophysics*, **15**, 143—148.
97. L a w s h e C. H., K e p h a r t N. C. and M c C o r m i c k E. J. (1949). The paired comparison technique for rating performance of industrial employees. *J. Appl. Psychol.*, **33**, 69—77.
98. L u c e R. D. (1959). Individual choice behavior. New York: Wiley.
99. L u c e R. D. (1961). A choice theory analysis of similarity judgments. *Psychometrika*, **26**, 151—163.

100. McCormick E. J. and Bachus J. A. (1952). Paired comparison ratings: 1. The effect on ratings of reductions in the number of pairs. *J. Appl. Psychol.*, **36**, 123—127.
101. McCormick E. J. and Roberts W. K. (1952). Paired comparison ratings: 2. The reliability of ratings based on partial pairings. *J. Appl. Psychol.*, **36**, 188—192.
102. McKeon J. J. (1960). Some cyclical incomplete paired comparisons designs. Tech. Rep. N **24**, Psychom. Lab., Univ. of North Carolina.
103. Mallows C. L. (1957). Non-null ranking models. I. *Biometrika*, **44**, 114—130.
104. Maurice Rita J. (1958). Selection of the population with the largest mean when comparisons can be made only in pairs. *Biometrika*, **45**, 581—586.
105. Moon J. W. and Moser L. (1962). Bibliography on graph theory. Univ. of Alberta (unpublished).
106. Moran P. A. P. (1947). On the method of paired comparisons. *Biometrika*, **34**, 363 — 365.
107. Morrissey J. H. (1955). New method for the assignment of psychometric scale values from incomplete paired comparisons. *J. Optic. Soc.Am.*,**45**, 373—378.
108. Mosteller F. (1951a, b, c). Remarks on the method of paired comparisons: I. The least-squares solution assuming equal standard deviations and equal correlations. II. The effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed. III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. *Psychometrika*, **16**, 3 — 9, 203 — 206, 207 — 218.
109. Mosteller F. (1952). The world series competition. *J. Am. Statist. Ass.*, **47**, 355 — 400.
110. Mosteller F. (1958). The mystery of the missing corpus. *Psychometrika*, **23**, 279 — 289.
111. Mosteller F. and Nogee P. (1951). An experimental measurement of utility. *J. Political Economy*, **59**, 371 — 404.
112. Noether G. E. (1960). Remarks about a paired comparison model. *Psychomertika*, **25**, 357—367.
113. Pearson E. S., and Hartley H. O. (1951). Charts of the power function for analysis of variance tests, derived from the noncentral *F*-distribution. *Biometrika*, **38**, 112 — 130; also N XIII of New Statistical Tables Separates.
114. Pearson E. S. and Hartley H. O. (1954). Biometrika Tables for Statisticians, vol. 1. Cambridge: University Press.
115. Pendergrass R. N. and Bradley R. A. (1960). Ranking in triple comparisons. Contributions to Probability and Statistics, ed. I. Olkin et al. Stanford: University Press, 331 — 351.
116. Peryam D. R. (1958). Sensory difference tests. Flavor research and food acceptance. Arthur D. Little, Inc., New York: Reinhold, 47 — 64.
117. Pfanzagl J. (1962). Über die stochastische Fundierung des psychophysischen Gesetzes. *Biom. Zeit.*, **4**, 1 — 14.
118. Putter J. (1955). The treatment of ties in some nonparametric tests. *Ann. Math. Statist.*, **26**, 368 — 386.
119. Rapoport A. (1949). Outline of a probabilistic approach to animal sociology: I. *Bull. Math. Biophysics*, **11**, 183 — 196.
120. Restle F. (1961). Psychology of judgment and choice: a theoretical essay. New York: Wiley.
121. Ross R. T. (1934). Optimum orders for the presentation of pairs in the method of paired comparisons. *J. Educ.Psychol.*, **25**, 375 — 382.
122. Scheffé H. (1952). An analysis of variance for paired comparisons. *J. Am. Statist.Ass.*, **47**, 381 — 400.
123. Scheffé H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika*, **40**, 87 — 104.

124. Scheid F. (1960). A tournament problem. *Am. Math.Monthly*, **67**, 39 — 41.
125. Shuford E. H., Jones L. V. and Bock R.D. (1960). A rational origin obtained by the method of contingent paired comparisons. *Psychometrika*, **25**, 343 — 356.
126. Slater P. (1960). The analysis of personal preferences. *Brit. J. Statist. Psychol.*, **13**, 119 — 135.
127. Slater P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika*, **48**, 303 — 312.
128. Starks T. H. (1958). Significance tests in experiments involving paired comparisons. Unpublished Ph. D. dissertation, Virginia Poly. Inst.
129. Starks T. H. and David H. A. (1961). Significance tests for paired comparison experiments. *Biometrika*, **48**, 95 — 108.
130. Steinhaus H. (1950). Mathematical Snapshots. Oxford: University Press. Штейнгауз Г. Математический калейдоскоп. М.—Л., Гостехиздат, 1949. (С более раннего издания).
131. Steinhaus H. and Trybula S. (1959). On a paradox in applied probabilities. *Bulletin de l'acadêmie polonaise des sciences.* Série des sci. math. astr. et phys., **7**, 67 — 79.
132. Thurstone L.L. (1927). A law of comparative judgment. *Psychol. Rev.*, **34**, 273 — 286.
133. Thurstone L. L. (1927). Psychophysical analysis. *Am. J: Psychol.*, **38**, 368 — 389. Терстон Л. Л. Психофизический анализ. — В сб.: Проблемы и методы психофизики. Под ред. А. Г. Асмолова, М. Б. Михалевской. М., Изд-во МГУ, 1974, с. 33 — 55.
134. Thurstone L.L. (1927). The method of paired comparisons for social values. *J. Abnorm. Soc.Psychol.*, **21**, 384 — 400.
135. Thurstone L. L. (1959). The measurement of values. Chicago: University Press.
136. Thurstone L. L. and Jones L. V. (1957). The rational origin for measuring subjective values. *J. Am. Statist. Ass.*, **52**, 458 — 471.
137. Torgerson W. S. (1958). Theory and methods of scaling. New York: Wiley.
138. Trawinski B.J. (1961). Frequencies of partitions for paired comparison experiments and related tables. Virginia Poly. Inst., Tech. Rep. 53.
139. Trawinski B. J. and David H. A. (1963). Selection of the best treatment in a paired — comparison experiment. *Ann. Math. Statist.*, **34**, 75 — — 91.
140. Trybula S. (1961). On the paradox of three random variables. *Zastosowania Mat.*, **5**, 321 — 332.
141. Ura S. (1957). On Scheffé's analysis of variance for paired comparisons. *Rep.Stat.Appl.Res., JUSE*, **4**, 132 — 146.
142. Ura S. (1960). Pair, triangle and duo-trio test. *Rep. Stat. Appl. Res., JUSE*, **7**, 107 — 119.
143. Ura S. (1960). Selection of judges by ranking method. *Rep. Stat. Appl. Res., JUSE*, **7**, 120 — 130.
144. Wei T. H. (1952). The algebraic foundations of ranking theory. Unpublished thesis, Cambridge University.
145. Wilkinson J. W. (1957). An analysis of paired-comparison designs with incomplete repetitions. *Biometrika*, **44**, 97 — 113.
146. Wolfe H. D. (1958). Consumer product testing. *Flavor research and food acceptance.* Arthur D. Little, Inc., New York: Reinhold, 135 — 149.

ПРИЛОЖЕНИЕ К РУССКОМУ ПЕРЕВОДУ

ПАРНЫЕ СРАВНЕНИЯ В ПРОШЛОМ, НАСТОЯЩЕМ И БУДУЩЕМ

Книга, которую вы только что прочитали, представляет собой одну из первых в мировой литературе попыток построения теории простейшей разновидности сравнений, когда объекты сравниваются попарно. Данное послесловие ставит три цели: 1) общее представление о сравнениях и их возможностях, 2) краткая характеристика книги Г. Дэвида, 3) краткий обзор более поздних результатов с библиографией ключевых работ. В таком порядке и построено изложение материала.

Упорядочение, оценивание и принятие решений. Одна из главных целей сравнения — упорядочение объектов или некоторых представлений о них, систематика явлений окружающего нас мира. Такие сравнения привели, например, к систематике животных и растений, классификации горных пород по твердости, шкалам силы землетрясений, силы ветра, высоты штормовых волн. При описании явлений природы упорядочение—часто цель работы.

Несколько иную роль играют сравнения в оценочных суждениях такого типа, как, например, «в этом году яблок уродилось больше, чем в прошлом». Здесь сравнение не претендует на роль элемента системы, а приводит к единичному суждению, по крайней мере внешне.

Характерно для рассмотренных случаев то, что сравнение не предполагало выбора. В практической деятельности сравнение чаще всего выступает именно как инструмент выбора. Тогда оно оказывается включенным в общую процедуру принятия решений как средство сопоставления или упорядочения альтернатив перед выбором. Теперь сравнение уже не беспристрастно, в нем заложена идея предпочтения. Значит, возник вопрос (гипотеза), для ответа на который решено привлечь некоторое сравнение или систему сравнений; определен критерий, позволяющий предпочесть ту или иную из альтернатив; указана операция, приводящая к значению заданного критерия. Конечно, определены множество альтернатив, внешние условия сравнений и т. п.

Собственно операция, устанавливающая отношения между альтернативами (объектами), определяется числом объектов, используемых одновременно: при установлении эквивалентности некоторого объекта самому себе применяется одноместная операция, так как фигурирует лишь один объект, а когда проводится двухместная операция, то как раз и получаются парные сравнения. Существуют и более сложные операции, приводящие к множественным сравнениям.

Когда мы сравниваем, скажем, размеры двух земельных участков, говоря, что один из них больше другого, то предполагается возможность прямых инструментальных измерений этих размеров. Но если мы скажем, что один из этих участков красивее другого, то метр окажется непригодным инструментом сравнения. В такой ситуации приходится прибегать к суждениям компетентных людей, называемых экспертами. Подобные суждения чрезвычайно важны в экономике (в самом широком смысле этого слова), а также в социологии и психологии (сошлемся на обзоры [336], [337], книги [333], [198] и [215]). Более узкие области применения парных сравнений — научно-техническое прогнозирование ([166], [168], [171], [207]) и особенно оценка качества продукции ([170], [201], [208]), где методы экспертных оценок часто единственно возможны. Такой интенсивно развиваемый раздел науки, как принятие решений ([338], [164]), тесно связан с методами парных сравнений — общими приложениями и некоторыми математическими задачами. Парные сравнения используются также в различных областях техники, гуманитарных наук, спорта (например, фигурное катание). Но и это не все. Частная жизнь каждого из нас в значительной степени определяется теми сравнениями, которые мы непрерывно осуществляем, и теми предпочтениями, которые мы при этом делаем.

Сравнить—очень часто означает измерить. Поэтому операция сравнения лежит в основе процедур измерения и приводит к построению измерительных шкал. Это помимо всего прочего имеет большое значение для научных исследований, для которых измерение — один из главных инструментов познания. В последние годы закладываются основы общей теории измерений, сравнения играют достойную роль.

Особая роль сравнений в построении измерительных шкал объясняет их отношение к построению шкал полезности и желательности и их место в методах исследования операций, системотехники, сетевого планирования, ПАТТЕРН, в задачах органолептических измерений, прогнозирования, педагогики и психологии.

Самые простые и естественные среди сравнений — парные сравнения, которым и посвящена данная книга. Перейдем теперь к ее краткой характеристике.

Парные сравнения и книга Г. Дэвида. Данная книга, как всякая классическая работа, суммирует наиболее простые, четкие и разработанные модели и результаты. Любопытно, что второе издание 1969 г., с которого сделан настоящий перевод, отличается от первого издания 1963 г. лишь добавлением в несколько строк.

Центральное место в книге занимают линейные модели Тэрстоуна — Мостеллера (ТМ), Брэдли — Терри (БТ) и Бока, проблемы планирования экспериментов в задачах парных сравнений и тесно связанные с экспериментальными планами вопросы построения турниров, особенно направленных на выявление лучшего объекта (игрока) из заданного конечного множества.

Источником моделей для парных сравнений служит в основном психология. Там накоплен уже большой опыт, сложились традиции, и вряд ли стоит говорить здесь об этом подробно. Что же касается интерпретации парных сравнений как эмпирической процедуры и вызванной ею идей планирования экспериментов парных сравнений, то это, пожалуй, самая нетривиальная часть книги. Связать воедино парные сравнения, планы и турниры — значит обеспечить высокую эффективность парных сравнений за счет разумного сокращения числа элементарных операций сравнения; расширить область приложения теории планирования эксперимента в масштабах, которые, несомненно, окажут влияние на судьбу самой теории; наконец, получить удобные и рациональные схемы проведения спортивных состязаний — турниров. Последнее достижение кажется слишком узким. Но на самом деле это совсем не так. Проблемы, изоморфные турнирам, постоянно возникают в значительно более серьезных областях, чем спорт, скажем, при сравнении алгоритмов, предназначаемых для решения некоторой задачи, да и вообще при выборе «наилучшего» объекта, будь то проект завода или стиральная машина.

Теория вероятностей, комбинаторный анализ, дискретная математика и математическая статистика образуют теоретический фундамент, на котором построено здание парных сравнений. Сюда нужно было бы добавить еще и теорию графов, но Дэвид ею практически не пользуется* (из-за чего иногда не вполне точен, см. сноску на с. 21). Вообще эклектичность «начал», из которых складывается современная теория парных сравнений, создает при изложении большие трудности. Только последовательно оставаясь на статистической точке зрения, автор сохраняет сбалансированность материала.

Собранные в последней главе вариации моделей парных сравнений дают известные представления о расширении их возможностей при учете некоторых дополнительных обстоятельств — качества экспертов, равенства рангов, порядка представления пар и т. п. В последние годы в этих (и других) направлениях наметился значительный прогресс, о котором мы скажем дальше. А прежде остановимся на особенностях перевода на русский язык ряда терминов. Начнем с одного уже давнего недоразумения, которое пора рассеять. Речь идет об английских терминах *tied* и *tied ranks*. Первый из них мы переводим в зависимости от контекста как *равенство*, *совпадение*, *ничья*, а второй — *равные ранги* или *совпадающие ранги*, вместо получившего распространение термина *связанные ранги*, *связи*, что не оправдано ни лингвистически, ни логически.

Для выражения *taste testing* нам представляется наиболее подходящим русский термин *дегустация*.

* За последние 10 лет теория графов очень широко применялась в парных сравнениях.

Удобное, короткое и емкое английское слово *score* приходится передавать в зависимости от контекста как *число очков*, *очки*, *счет*, *результат*, *значение*, *данные*, *отметка*. Получивший некоторое распространение перевод *метка* представляется малоудачным, поскольку метка ассоциируется со шкалой прибора или зарубкой на дереве и не выражает всех нужных оттенков. Другой возможный вариант перевода—*оценка* — ведет к синонимии с распространенным статистическим термином и потому нецелесообразен.

В литературе на русском языке прижился термин *пробит* (*probit* = *probability unit*), введенный Финни. Поэтому было решено для его рангового аналога (*rankit*) поступить аналогично, воспользовавшись транслитерацией *ранкит*.

Уже ясно, что в книге много терминов спортивного происхождения. Для ряда из них приведем принятые переводы *«Round Robin» tournament* — круговой турнир (или *турнир по круговой системе*): *knock-out tournament* — олимпийский турнир (или *турнир с выбыванием после проигрыша*); *double elimination tournament* — *турнир с выбыванием после двух проигрышей*; *round* — *тур*, *период*, *тайм*, *раунд*.

Три разных английских слова *merit*, *value* и *worth* мы воспринимали как синонимы и переводили одним словом — *ценность*, с редкими вариантами — *цена*, *стоимость*. Конечно, в соответствующих контекстах слово *value* сохраняло свой другой смысл — *значение*, *величина*.

Наконец укажем еще перевод терминов *rating methods* и *personal rating*. Первый мы переводили так: *методы упорядочения* (*определения рейтинга*), *методы аттестации*. Слово *рейтинг* завоевало права гражданства, например, в шахматах, где принято говорить о рейтинге, или индивидуальном коэффициенте шахматиста, определяемом, кстати сказать, с помощью системы парных сравнений (вообще говоря, несбалансированной). Второй термин получил значения *индивидуальный рейтинг* или *аттестация*.

Эти не претендующие на полноту терминологические заметки должны, как мы надеемся, помочь упорядочению русской терминологии в области парных сравнений.

Парные сравнения—новый этап. Можно констатировать, что классический период развития парных сравнений уже принадлежит прошлому. Можно надеяться, что мы стоим на пороге нового этапа. Во всяком случае, поток литературы стремителен. Наши трудности усугубляются рассеянием информации по самым различным и часто неожиданным изданиям на разных языках и в разных странах. Начнем с обзоров и библиографий последнего времени, чтобы затем попытаться пополнить их наиболее важными из оказавшихся в нашем распоряжении работ, подвергнутых грубой классификации.

Если говорить об обзорах и библиографиях, то, несомненно, наибольшее значение имеют работа Брэдли [244] и работы [280], [281] вместе с сопровождающей дискуссией и соответствующей обширной библиографией [259], а также обзор [336], на русском языке—недавно изданный (и, видимо, пока единственный) обзор [216]. Далее мы просто перечислим в алфавитном порядке иностранные и отечественные монографии, обзоры и библиографии, имеющие прямое или косвенное отношение к нашему предмету и в основном относящиеся к последнему времени. Пересечения с работами [216], [244] и [259] минимальны. Итак, [147] — [219]. Эта библиография избавляет нас от необходимости перечисления многих работ, ссылки на которые можно найти в указанных источниках.

Теперь остановимся кратко на нескольких ведущих направлениях развития парных сравнений.

Модели и предпосылки. Суждения экспертов есть всего только суждения экспертов и, что бы мы ни говорили, имеют субъективную окраску. Это обстоятельство долгие годы служило пугалом, мешавшим привлечению столь «ненаучной» информации к научному рассмотрению тех или иных сложных проблем. Потребовалось много времени, чтобы осознать, что, во-первых, у нас часто просто нет альтернативы, а во-вторых, что в самых респектабельных научных концепциях, по существу, присутствует неизбежная доля той самой субъективности, к которой мы столь пренебрежительно относимся. Поэтому произошло включение в научный обиход рассмотрения, анализа и использования субъективных суждений, что, на наш взгляд, служит вехой в истории науки.

Всякое субъективное суждение связано с некоторой долей неопределенности, неуверенности, сомнительности. Это вполне естественно, ибо мы обращаемся к суждениям экспертов только тогда, когда не в силах справиться с проблемой классическими методами. Но эта неопределенность приводит к необходимости ее описания, что, в свою очередь, ведет к аппарату теории вероятностей. Это послужило одной из причин создания концепции субъективных вероятностей, порождаемых механизмом выбора, напоминающим пари, в котором ставка делается на альтернативу, предпочитаемую экспертом в некоторой процедуре сравнения альтернатив. Теперь на смену призраку субъективности пришли действительные проблемы построения адекватных моделей выбора (предпочтения), процедур обработки данных и механизмов принятия решений.

Линейные модели Брэдли—Терри—Льюса (БТЛ) и Тэрстоуна — Мостеллера (ТМ), рассматриваемые Дэвидом, обобщались многими авторами в различных направлениях.

Одним из первых обобщений линейной модели было распространение ТМ-модели на случай «ничьих» в работе [56]. Позднее в работе [301] была предложена модель, основанная на гипотезе о порогах чувствительности [153], как и модель Гленна — Дэвида.

Рассмотрим обобщение модели Рао — Купера. Пусть $\eta_{ij} = \ln\theta_{ij}$ — порог для объектов A_i, A_j, $1 \leqslant i < j \leqslant t$, отражающий способность экспертов различать объекты, $\theta_{ij}=\theta_{ji}$, $i \neq j$. Обозначим вероятности $P(A_i \to A_j) = P(Y_i - Y_j > \eta_{ij}) = \pi_{ij}$, $P(A_j \to A_i) = P(Y_i - Y_j) < -\eta_{ij}) = \pi_{ji}$, $P(A_i \sim A_j) = P(|Y_i - Y_j| < \eta_{ij}) = q_{ij}$ (см. гл. 1 книги Дэвида). Предлагается модель $\pi_{ij} = \pi_i/(\pi_i + \theta_{ij}\pi_j$, $q_{ij} = (\theta_{ij}^2 - 1)\,\pi_i\pi_j/(\pi_i + \theta_{ij}\pi_j)\,(\theta_{ij}\pi_i + \pi_j)$, причем $\pi_{ij} + \pi_{ij} + q_{ij} = 1$.

Если $\theta_{ij} = \theta$, то получается модель Рао — Купера. Возможен другой подход, основанный на гипотезе о том, что $(A_i \to A_j)$, $(A_i \sim$

$\sim A_j$) — случайные события и неразличение происходит или не происходит с некоторой вероятностью. Рассмотрим одну довольно общую модель такого типа. Частный случай такой модели был предложен Томпсоном и Сингхом [315] и развит Бивером и Рао [230], Сингхом и Гуптой [309]. Томпсон и Сингх [315] рассматривали два основных психофизических механизма различения объектов — «аддитивный» и «заменяющий» (*substitution*), предложенные Стивенсом (см. [313]). При использовании предельных теорем эти два механизма приводят к моделям Тэрстоуна и Брэдли—Терри соответственно.

Упомянем о некоторых работах, где изучалась линейная модель. Это работы Дэниелса [253], где изучалась система присвоения очков в круговом турнире, Муна и Пальмэна [288], обобщивших подход Дэниелса. Бюльман и Хубер [247] занимались задачей «корректного ранжирования» в парных сравнениях в духе теории статистических решений. Работу Лемана (см. [244] и [259]) о мощности ранговых критериев можно применить к парным сравнениям, что дает известную БТЛ-модель.

Упомянем еще работы Аткинсона [225], который, следуя книге Кокса [251], рассмотрел логарифмическую разность в логистической модели (т. е. линейной модели с логистической плотностью распределения, частный случай которой—БТЛ-модель) $\lambda_{ij} = \ln [\pi_{ij}/(1 - \pi_{ij})] = \rho_i - \rho_j$, $\rho_i = \ln \pi_i$, $i \neq j$, $i, j = 1, \ldots, t$.

Другой моделью Аткинсона, альтернативной по отношению к логистической, является

$$\lambda_{ij} = (\rho_i - \rho_j)\,[1 + \exp \{k\,(\rho_i - \rho_j)^2\}],\ i \neq j,$$

причем предлагается критерий проверки гипотезы $k = 0$.

Дэвид лишь упоминает о таком важном моменте, как порядок предъявления пар экспертам, который может оказывать существенное влияние на результат сравнения, вызывая усталость или меняя порог чувствительности. В упоминаемой у Дэвида работе Харриса [73] ТМ-модель модифицируется с учетом порядка представления (во времени или пространстве) объектов путем введения смещения в математическое ожидание разности откликов между двумя стимулами $Y_i - Y_j$, которое равно $\mu_i - \mu_j$ (при дисперсии $2\sigma^2 (1 - \delta\rho)^2$).

Обобщение БТЛ-модели с учетом влияния предъявления предложено Бивером и Джокхэйлом [228]. Вводятся вероятности предпочтения $A_i \to A_j$ с учетом порядка предъявления:

$$P_{ij}(A_i \to A_j) = \frac{\pi_i + \delta_{ij}}{\pi_i + \pi_j};\quad P_{ij}(A_j \to A_i) = \frac{\pi_j - \lambda_{ij}}{\pi_i + \pi_j};$$

$$P_{ji}(A_i \to A_j) = \frac{\pi_i - \delta_{ij}}{\pi_i + \pi_j};\quad P_{ji}(A_j \to A_i) = \frac{\pi_j + \delta_{ij}}{\pi_i + \pi_j},$$

т. е. аддитивный параметр δ_{ij}, $|\delta_{ij}| \leqslant \min(\pi_i, \pi_j)$, прибавляется, когда находится вероятность превосходства первого из предъявляемых объектов, и вычитается для второго.

Модель с мультипликативным увеличением или уменьшением (и параметром) предложена Бивером в обсуждении доклада Брэдли [244].

Упомянем еще работу Шаафсма [306] и исследование Коусгаарда [285], в которых порядок предъявления учитывался наряду с учетом различной способности экспертов к сравнению.

Альтернативой к парным сравнениям служат тройные сравнения, анализ которых иногда представляет значительный интерес, поскольку это шаг от парных сравнений к прямому ранжированию.

Бивер и Рао ([229], [230]) применили психофизический подход Томпсона и Сингха [315] для тройных сравнений и получили модель (1.3.4) — из Дэвида — с обобщением на случай «ничьих». Более подробной является публикация Бивера и Рао [229]; Рей [300] применил модель (1. 4. 2).

При проведении экспериментов парных сравнений явно или неявно предполагается, что имеется некий признак (критерий типа полезности, ценности и т. п.), по которому производится сравнение. Однако часто имеется несколько критериев, по которым требуется производить сравнение (изучение качества продукции, многокритериальные задачи принятия решений [201]). Если в этом случае производятся парные сравнения, то говорят о многокритериальных парных сравнениях. Модель многокритериальных парных сравнений, обобщающая БТЛ-модель, была введена в работе Дэвидсона и Брэдли [258].

Многокритериальная модель применялась для данных дегустационного эксперимента ([258], [224], [216]).

Дальнейшее изучение упомянутой модели выполнено Дэвидсоном и Брэдли [258], Импри, Джонсоном и Кохом [282], Финбергом и Ларнцем [266]. Другая модель предложена Сеном и Дэвидом [308].

Модели парных сравнений являются в некотором смысле частными по отношению к так называемым моделям «выбора и ценности» (*choice and worth*) или «вероятностного выбора». Первой из них является модель Льюса [98], описанная в книге Дэвида (см. также [151]). Эти модели интенсивно развивались в последние годы. Одли (см. [244] и [259]) рассмотрел процесс выбора и отбора как динамический процесс во времени, Блок и Маршак [9] развили математическую теорию полезности, связанную с выбором и ранжированием, где основой для парных сравнений служит ранжирование, другой подход принадлежит Бранку [18]. Оба подхода в частных случаях приводят к линейной модели из книги Дэвида. Развитие моделей этого класса дано в работах Тверского ([320], [321]), где имеются ссылки на литературу. Обзор моделей «выбора и ценности» сделан в работах [213] и отчасти [303].

Обработка данных: *ранжирование*. У процедуры парного сравнения две основные цели — упорядочение объектов, называемое еще ранжированием, и поиск наилучшего из объектов — лидера, чемпиона и т. п. Первая задача фактически есть задача построения шкалы порядка (или иногда даже метрической шкалы). О второй мы скажем ниже.

При обработке данных в задаче ранжирования исследовались многие пути, среди которых упомянем здесь следующие:

ранжирование максимального правдоподобия ([303], [314], [279], [280], [281], [160] — [162]) с обобщениями на случай равных рангов [311];

медиану Кемени — Снелла ([175], [182], [189], [190], [279] — [281]). Другой подход к этой задаче развит Миркиным Б. Г. ([191], [192]);

правило большинства с ограничениями транзитивности ([183], [191]).

Задача о правиле большинства возникла, видимо, независимо от работы Кемени — Снелла [175], опубликованной в 1962 г., в работе [237] и была исследована в серии работ [238] и др. (см. [280], [281], [241] — [243]). В книге Миркина Б. Г. [191, гл. 2] задача Кемени — Снелла обобщается на случай, когда экстремум ищется на множестве бинарных отношений.

Поиск «последовательного» упорядочения Слейтера [126], [127]. Подход Слейтера породил многочисленные работы: Томпсона, Римейджа, Сингха, упоминавшиеся выше и являющиеся обобщением, а также работы в теории ориентированных графов, точнее говоря, турниров — см. в первую очередь [210]. Обзоры работ даны Дэвидом [256] и Хьюбертом [280], причем они хорошо дополняют друг друга.

Поиск перестановки с минимальным потенциалом. В работе Буркова и Гроппена [160] формулируется понятие потенциала перестановки R для матрицы $\|Q_{ij}\|$, $i, j = 1, \ldots, t$, равного

$$\Gamma(R) = \sum_{1 \leqslant i < j \leqslant t} \theta_{ij} C(R_i, R_j),$$

и указывается на связь задачи о поиске перестановки с минимальным потенциалом и задачи о разрезе на ориентированном графе. Вычислительные аспекты данной задачи обсуждаются как в упомянутой работе, где предложен алгоритм, так и в работах [161], [162].

Перечисленные выше подходы (ранжирование максимального правдоподобия, медиана Кемени — Снелла, правило большинства, поиск перестановки с минимальным потенциалом) и некоторые другие являются частным случаем задачи о квадратичном значении Купманса и Бекмана [284], которая развивалась в экономике независимо от парных сравнений (см. Хьюберт [280], [281]).

Ранжирование объектов и экспертов в соответствии с их точками зрения методами многомерного неметрического шкалирования — Каменский ([172], [173]).

С точки зрения математической статистики в основе обработки данных часто лежит та или иная модель дисперсионного анализа [214].

Турниры и планы. Уже упоминалось, что совокупность парных сравнений для некоторого множества объектов можно рассматривать как турнир; способ организации турнира — как план проведения эксперимента парных сравнений, а теорию графов — как удобный инструмент анализа турниров и их результатов [286].

Одна из основных целей турнира — выбор лучшего игрока. Этому вопросу посвящена гл. 6 в книге Дэвида. В дальнейшем исследования продолжены в работах Нарайаны и Шидека ([294], [295]). Во второй части работы [295] изучается задача проверки гипотезы эквивалентности (одинаковой силы) n игроков турнира (т. е. n объектов) против альтернативы о наличии одного сильнейшего игрока. Предполагается m независимых повторов («кругов»). В работе изучается также задача отбора сильнейшего игрока, причем так же, как и для проверки гипотез, используются как обычный, так и байесовский подход. Турниры как полные ориентированные графы [211] есть объект интенсивных исследований в теории графов и комбинаторике, где получены многочисленные результаты ([287]—[290], [218], [219], [262], [263], [220], [221]), имеющие интересную интерпретацию в рамках парных сравнений.

Задаче планирования экспериментов парных сравнений посвящена обширная литература, лишь малая часть которой упомянута в гл. 5 книги Дэвида. Укажем в этой связи ряд дополнительных источников ([184] — [186], [254], [255], [257], [258], [265], [283]).

Любопытно отметить, что если планирование эксперимента полезно в проведении парных сравнений, то и парные сравнения весьма полезны при проведении конкретных работ с применением планирования эксперимента. Они наряду с ранжированием широко используются на этапе предпланирования при выборе откликов и факторов [148].

В последние годы сочетание идей теории турниров и теории планов породило ряд новых направлений, например анализ турниров с гандикапом (форой, которая дается для уравнивания шансов) [288].

Парные сравнения, множественные сравнения и таблицы сопряженности. В предыдущих разделах рассматривались модели БТЛ (1952 г., 1959 г.) для парных сравнений, Льюса — для вероятностного выбора (множественных сравнений), Пендерграсса и Брэдли [115] — для множественных сравнений, Сена — Дэвида [308] и Дэвидсона — Брэдли [258] — для многокритериальных парных сравнений и др. Рассматривая данные парных сравнений как категоризованные данные и представляя их в виде различным образом составленных таблиц сопряженности ([91], [92]), можно облегчить анализ результатов парных сравнений ([282], [266]).

Парные сравнения сейчас в некоторой степени представляют собой пример того, как, действуя параллельно методами математической статистики (в том числе и свободными от распределения, непараметрическими методами [244]) и так называемыми методами «анализа данных» [280], можно продвинуться достаточно далеко в направлении создания математического аппарата для обеспечения практических потребностей. При таком развитии в двух направлениях выявляются многочисленные и иногда довольно глубокие связи между совершенно различными на первый взгляд экономическими, техническими и психологическими задачами [192], [281].

Понимание таких возможностей приводит к общим концепциям, что весьма важно само по себе.

З а к л ю ч е н и е. Таким образом, нам представляется, что парные сравнения, как и сопутствующие им задачи, ожидает большое будущее. Они находят и будут находить применение во все более широком спектре областей человеческой деятельности—везде, где требуются суждения компетентных людей и где эти суждения нуждаются в научных формах фиксирования: квантификации, классификации, шкалировании.

Книга Дэвида, которую вы прочитали, — важный шаг на пути прогресса в этих направлениях.

БИБЛИОГРАФИЯ

147. Абрамов Л. М., Шляго Н. Н. Об одном методе нахождения медианы множества упорядочения (деп. рук. в ВИНИТИ), 1976.
148. Адлер Ю. П. Введение в планирование эксперимента. М., «Металлургия», 1969.
149. Апинис А. Аксиоматическое определение правила α-большинства. — В сб.: Современные направления теории игр. Под ред. Э. Й. Вилкаса, А. А. Корбута. Вильнюс, «Моксклас», 1976, с. 12—17.
150. Апинис А. О. О транзитивности правил большинства и более общих правил. — В кн.: Математические методы в социальных науках. Вып. 4. Вильнюс, 1974, с. 9—36.
151. Аткинсон Р., Бауэр Г., Кротерс Э. Введение в математическую теорию обучения. М., «Мир», 1969.
152. Байнеке Л. В., Харари Ф. Максимальное число сильно связанных подтурниров. — В кн.: Теория графов. Покрытия. Укладки. Турниры. М., «Мир», 1974, с. 163—168.
153. Бардин К. В. Проблема порогов чувствительности и психофизические методы. М., «Наука», 1976.
154. Берж К. Теория графов и ее применения. М., ИЛ, 1962.
155. Бешелев С. Д., Гурвич Ф. Г. Математико-статистические методы экспертных оценок. М., «Статистика», 1974.
156. Битинас Б. П. Многомерный анализ в педагогике и педагогической психологии. Вильнюс, 1971.
157. Большев Л. Н. Рецензия на книгу: Molenaar W. Approximation to the Poison, binomial and hipergeometric distribution functions, *Math. Centrum Tracts.*, v. 31, *Amsterdam*, 1970. Теория вероятностей и ее применения. 1971, т. 16, вып. 1, с. 196—198.
158. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М., ВЦ АН СССР, 1968.
159. Брук Б. Н., Бурков В. Н. Методы экспертных оценок в задачах упорядочения объектов. — «Изв. АН СССР. Техн. кибернетика», 1972, № 3 с. 29—39.
160. Бурков В. Н., Гроппен В. О. Разрезы в сильно связанных графах и потенциалы перестановок. — «Автоматика и телемеханика», 1972, № 2, с. 111—119.
161. Бурков В. Н., Гроппен В. О. Решение задачи о минимальном разрезе в бисвязном орграфе алгоритмами типа «ветвей и границ». — «Автоматика и телемеханика», 1974, № 9, с. 104—110.
162. Бурков В. Н., Гроппен В. О. Максимальная циркуляция и минимальный разрез на планарных графах. — «Кибернетика», 1975, № 6, с. 85—89.
163. Вилкас Э. Й. Теория полезности. — В кн.: Теория вероятностей. Математическая статистика. Теоретическая кибернетика. М., ВИНИТИ, 1977, с. 123—151. (Итоги науки и техники. Т. 14).
164. Вопросы анализа и процедуры принятия решений. Сб. переводов под ред. Шахнова И. Ф. М., «Мир», 1976.
165. Глотов В. А., Павельев В. В. Экспертные методы определения весовых коэффициентов (обзор). — «Автоматика и телемеханика», 1976, № 12, с. 95—107.

166. Голованов Л. В., Ефимушкин С. Н., Козырев В. П., Литвиненко С. Ф., Никитина Л. В. Экспертные оценки в практике управления наукой в техническом вузе. М., МВТУ, 1973.
167. Демидович Б. П., Марон И. А. Основы вычислительной математики. М., Физматгиз, 1960.
168. Добров Г. М., Ершов Ю. В., Левин Е. И., Смирнов Л. Г. Экспертные оценки в научно-техническом прогнозировании. Киев, «Наукова думка», 1974.
169. Зиновьев В. А., Козырев В. П. Сильно регулярные графы и комбинаторные конфигурации. — В кн.: Вопросы кибернетики. Вып. 16. М., «Сов. радио», 1975, с. 30—53.
170. Измерение качества продукции. Вопросы квалиметрии. Сб. под ред. Гличева А. В. М., «Стандарты», 1976.
171. Информационное обеспечение в задачах управления научными подразделениями НИИ. — В кн.: Сб. трудов Ин-та проблем упр. 1976, вып. 9.
172. Каменский В. С. Методы и модели неметрического многомерного шкалирования (обзор). — «Автоматика и телемеханика», 1977, № 8, с. 118—156.
173. Каменский В. С. Парное неметрическое многомерное развертывание. — В кн.: Проблемы педагогической квалиметрии. Вып. 3. М., 1977. (в печати).
174. Кемени Дж., Снелл Дж., Томпсон Дж. Введение в конечную математику. Изд. 2-е (Пер. с англ.). М., «Мир», 1965.
175. Кемени Дж., Снелл Дж. Кибернетическое моделирование. М., «Сов. радио», 1972.
176. Козырев В. П. Теория графов. — В кн.: Теория вероятностей. Математическая статистика. Теоретическая кибернетика. М., ВИНИТИ, 1972, с. 25—74. (Итоги науки и техники. Т. 10).
177. Кузьмин В. Б., Овчинников С. В. Априорные модели предпочтения. — В кн.: Сб. трудов Ин-та проблем упр. 1976, вып. 9, с. 51—54.
178. Кузьмин В. Б., Овчинников С. В. Об измерениях в порядковых шкалах. — «Автоматика и телемеханика», 1974, № 11, с. 106—112.
179. Кузьмин В. Б., Овчинников С. В. Геометрия пространств предпочтений. — «Автоматика и телемеханика», 1975, № 12, с. 140—145; 1976, № 1, с. 174—178.
180. Кузьмин В. Б. Модельный геометрический подход к проблеме группового выбора. — В кн.: Сб. трудов Ин-та проблем упр. 1976, вып. 9, с. 40—50.
181. Леман Е. Проверка статистических гипотез. М., «Наука», 1964.
182. Литвак Б. Г. Об упорядочении объектов по предпочтениям. — В кн.: Математические методы управления производством. Вып. 5. М., МГУ, 1973, с. 47—56.
183. Льюс Р. Д., Райфа Х. Игры и решения (Пер. с англ.). М., ИЛ, 1961.
184. Маркова Е. В. Руководство по применению латинских планов при планировании эксперимента с качественными факторами. Челябинск, УралНИИ—Стройпроект, 1971.
185. Маркова Е. В. Неполноблочные планы. Препринт № 15. Лаборатория статистических методов МГУ. М., Изд-во МГУ, 1970.
186. Маркова Е. В., Лисенков А. Н. Планирование экспериментов в условиях неоднородностей. М., «Наука», 1973.
187. Математические методы в современной буржуазной социологии. Сб. пер. М., «Прогресс», 1966.
188. Математические методы в социальных науках. Сб. пер. с англ. М., «Прогресс», 1973.
189. Метев Б. С. Приближенный метод для предварительной обработки мнений экспертов. — «Журн. вычислит. математики и мат. физики», 1976, т. 16, № 5, с. 1344—1348.
190. Метев Б. С., Попчев И. П. Метод решения задачи группового выбора с использованием метрических пространств отношений. — «Автоматика и телемеханика», 1977, № 2, с. 81—87.

191. Миркин Б. Г. Проблема группового выбора. М., «Наука», 1974.
192. Миркин Б. Г. Анализ качественных признаков (математические модели и методы). М., «Статистика», 1976.
193. Наппельбаум Э. Л., Поспелов Д. А. Проблема коллективов решений и экспертные оценки. — В кн.: Вопросы кибернетики. Вып. № 8. Теория принятия решения. М., «Сов. радио», 1975, с. 86—102.
194. Орлов А. И. Проблемы устойчивости и обоснованности решений в теории экспертных оценок. — В сб.: Статистические методы анализа экспертных оценок. М., «Наука», 1977, с. 7—30. (Учен. зап. по статистике. Т. 29).
195. Орлов А. И. Проблема учета неопределенностей реальных явлений в математических моделях. — В сб.: Проблемы педагогической квалиметрии. Вып. 2. М., МГПИ им. Ленина, 1975, с. 180—187.
195а. Орлов А. И. Устойчивость в социально-экономических моделях. М., «Наука» (в печати).
196. Оуэн Д. Б. Сборник статистических таблиц. М., ВЦ АН СССР, 1973.
197. Прогноз в речевой деятельности. М., «Наука», 1974.
198. Процесс социального исследования (вопросы методологии, методики и организации марксистско-ленинских социальных исследований). (Пер. с нем.). М., «Прогресс», 1975.
199. Психологические измерения (Пер. с англ.). М., «Мир», 1967.
200. Пфанцагль И. при участии Бауманна В., Хубера Г. Теория измерений. М., «Мир», 1976.
201. Райхман Э. П., Азгальдов Г. Г. Экспертные методы в оценке качества товаров. М., «Экономика», 1974.
202. Рао С. Р. Линейные статистические методы и их применения. М., «Наука», 1968.
203. Скворцов В. В. Методы экспертных оценок и их приложение в задачах теории фильтрации. Казань, Татар. кн. изд-во, 1976.
204. Статистическое измерение качественных характеристик (Пер. с англ.). М., «Статистика», 1972.
205. Статистические методы анализа экспертных оценок. М., «Наука», 1977. (Учен. зап. по статистике. Т. 29).
206. Стивенс С. С. Математика, измерение и психофизика. — В кн.: Экспериментальная психология. Т. 1. М., ИЛ, 1960.
207. Тарасенко Ф. П. О принципиальных трудностях балльных оценок научно-технического прогресса. — «Вестн. АН СССР», 1976, № 6, с. 69—75.
208. Тильгнер Д. Е. Органолептический анализ пищевых продуктов (Пер. с польск.). М., Пищепромиздат, 1962.
209. Ушаков И. А. Задача о выборе предпочтительного объекта. — «Изв. АН СССР. Техн. кибернетика», 1971, № 4, с. 3—7.
210. Фалкерсон Д. Р. Нарушения в круговых турнирах. — В кн.: Теория графов. Покрытия, укладки, турниры. М., «Мир», 1974, с. 175—189.
211. Харари Ф. Теория графов. М., «Мир», 1973.
212. Черный Л. Б. Метод пространственных разбиений в анализе качественных признаков. Канд. дис. Новосибирск, 1973. (Ин-т экономики и организации пром. производства).
213. Шейнин Р. Л. Вероятностный выбор альтернатив (обзор). — «Автоматика и телемеханика», 1979 (в печати).
214. Шеффе Г. Дисперсионный анализ. М., «Наука», 1963.
215. Шляпентох В. Э. Проблемы достоверности статистической информации в социологических исследованиях. М., «Статистика», 1973.
216. Шмерлинг Д. С., Дубровский С. А., Аржанова Т. Д., Френкель А. А. Экспертные оценки. Методы и применения (обзор). — В сб.: Статистические методы анализа экспертных оценок. М., «Наука», 1977, с. 290—382. (Учен. зап. по статистике. Т. 29).
217. Шрейдер Ю. А. Равенство, сходство, порядок. М., «Наука», 1971.
218. Эрдеш П., Спенсер Дж. Вероятностные методы в комбинаторике. М., «Мир», 1976.
219. Эрдеш П., Мун Дж. В. О множествах согласованных дуг в турнире. — В кн.: Теория графов. М., «Мир», 1974, с. 160—162.

245. B r a d l e y R. A., El-H e l b a w y A. T. Treatment contrasts in paired comparisons: Basic procedures with application to factorials. *Biometrika*, 1976, v. **63**, N 2, p. 255 — 262.
246. B r a u e r A., G e n t r y I. C. On the characteristic roots of tournaments. *Bull. Amer. Math. Soc.*, 1968, v. **74**, p. 1133 — 1135.
247. B ü h l m a n n H., H u b e r P. J. Pairwise comparison and ranking in tournaments. *Ann. Math. Statist.*, 1963, v. **34**, p. 501 — 510.
248. B u r k e C. J., Z i n n e s J. L. A Paired comparison of pair comparisons. *J. Math. Psychol.*, v. **2**, 1965, N 1, p. 53 — 76.
249. C l a t w o r t h y W. H. Table of two-associate-class partially balanced designs. Washington, *Gov. print. off.*, 1973. (US. Nat. bur. stand. Appl. Math. ser. 63).
250. C o o m b s C. H. A theory of data. N. Y., Wiley, 1964.
251. C o x D. R. The analysis of binary data. London, Methuen, 1970.
252. C o x D. R. Planning of experiments. N. Y., Wiley, 1958.
253. D a n i e l s H. E. Round-robin tournament scores. *Biometrika*, 1969, v. **56**, N 2, p. 295 — 299.
254. D a v i d H. A. Enumeration of cyclic paired-comparison designs. *Amer. Math. Monthly*, 1965, v. **72**, p. 241 — 248.
255. D a v i d H. A. Enumeration of cyclic graphs and cyclic designs. *J. Comb. Theory*, 1972, v. **13**, p. 303 — 308.
256. D a v i d H. A. Ranking of the players in a Round Robin tournament. *Rev. Int. Statist. Inst.*, 1971, v. **39**, p. 137 — 147.
257. D a v i d H. A. Resolvable cyclic designs. *Sankhyā*, A, 1967, v. **29**, p. 191 — 198.
258. D a v i d s o n R. R., Bradley R. A. Multivariate paired comparisons: some large-sample results on estimation and tests for equality of preference. — In: Nonparametric techniques in statistical inference. Ed. M. L. Puri. Cambridge Univ. press, 1970, p. 111 — 125.
259. D a v i d s o n R. R., F a r q u h a r P. H. A bibliography on the method of paired comparisons. *Biometrics*, 1976, v. **32**, p. 241 — 252, bibl. 352.
260. D e C a n i J. Maximum likelihood paired comparison ranking by linear programming. *Biometrika*, 1969, v. **56**, N 3, p. 537—545.
261. D e C a n i J. A branch and bound algorithm for maximum likelihood paired comparison ranking. *Biometrika*, 1972, v. **59**, N 1, p. 131—135.
262. E r d ö s P., H a j n a l A., M i l n e r E. C. Simple one-point extensions of tournaments. *Mathematika*, 1972, v. **19**, 57—62.
263. E r d ö s P., F r i e d E., H a j n a l A., M i l n e r E. C. Some remarks on simple tournaments. *Algebra universalis*, 1972, v. **2**, p. 238—245.
264. F e c h n e r G. I. Elements of psychophysics. V. 1. N. Y., Holt, Rinehart and Winston, 1965.
265. F e d e r e r W. T., B a l a a m L. N. Bibliography on experiment and treatment design pre-1968. Edinburgh, Oliver and Boyd, 1972.
266. F i e n b e r g S. E., L a r n t z K. Loglinear representation for paired and multiple comparisons models. *Biometrika*, 1976, v. **63**, N 2, p. 245—254.
267. F i s h b u r n P. C. Binary choice probabilities: on the varieties of stochastic transitivity. *J. Math. Psychol.*, 1973, v. **10**, N 4, p. 327—352.
268. F i s h b u r n P. C. Should social choice be based on binary comparisons? *J. Math. Sociology*, 1971, v. 1, p. 133—142.
269. F i s h b u r n P. S. Transitive binary social choice and intransitive conditions. *Econometrica*, 1973, v. **41**, N 4, p. 603—615.
270. F i s h b u r n P. C. On linear extension majority graphs of partial orders *J. Comb. Theory*, 1976, B, v. **21**, p. 65—70.
271. F l u e c k J. A., K o r s h J. F. A branch search algorithm for maximum likelihood paired comparison ranking. *Biometrika*, 1974, v. **61**, p. 621—626.
272. F l u e c k J. A., K o r s c h J. F. A generalized approach to maximum likelihood paired comparison ranking. *Ann. Statist*, 1975, v. **3**, N 4, p. 846—861.
273. G l e a s o n J. R., H a l p e r i n S. A paired compositions model for Round—Robin experiments. *Psychometrika*, 1975, v. **40**, N 4, p. 433—454..
274. H a b e r m a n S. J. The analysis of frequency data. Chicago, Univ. of Chicago Press, 1974.

275. Hanan M., Kurtzberg J. M. A review of the placement and quadratic assignment problems. *SIAM Rev*, 1972, v. **14**, N 2, p. 324—342.
276. Hartigan J. A. Probabilistic competition of a khockout tournament. *Ann. Math. Statist.*, 1966, v. 37, p. 495—503.
277. Holland P. W., Leinhardt S. Conditions for eliminating intransitivities in binary digraphs. *J. Math. Sociology*, 1976, v. **4**, p. 315—318.
278. Hoyer R. W., Mayer L. S. Notes and comments social preference orderings under majority rule. *Econometrica*, 1975, v. **43**, N 4, p. 803—806.
279. Hubert L., Schultz P. Maximum likelihood paired comparison ranking and quadratic assignment. *Biometrika*, 1975, v. **62**, N 3, p. 655—659.
280. Hubert L., Schultz P. Quadratic assignment as a general data analysis strategy. *Brit. J. Math. Statist. Psychol.*, 1976, v. **29**, N 2, p. 190—241.
281. Hubert L. Seriation using asymmetric proximity measures. *Brit. J. Math. Statist. Psychol.*, 1976, v. **29**, N 1, p. 32—52.
282. Imprey P. B., Johnson W. D., Koch G. G. An incomplete contigency table approach to paired—comparison experiment. *J. Amer. Statist. Assoc.*, 1976, v. **71**, N 355, p. 614—623.
283. John J. A., Wolock F. W., David H. A. Cyclic designs. Washington, Gov. print. off., 1972, 76 p. (*US. Nat. bur. stand. Appl. math. ser.* 62).
284. Koopmans T. C., Beckmann M. Assignment problems and the location of economic activities. *Econometrica*, 1957, v. **25**, N 1, p. 53—76.
285. Kousgaard N. Models for paired comparisons with ties. *Scand. J. Statist. Theory and Applications*, 1976, v. 3, N 1, p. 1—14.
286. Maurer W. On most effective tournament plans with fewer games than competitors. *Ann. Statist.*, 1975, v. **3**, N 3, p. 717—727.
287. Moon J. W. Topics on tournaments. N. Y., Holt, Rinehart and Winston, 1968.
288. Moon J. W., Pullman N. F. On generalized tournament matrices. *SIAM Rev.*, 1970, v. **12**, p. 384—399.
289. Moon J. W. A problem on arcs without bypasses in tournament. *J. Comb. Theory*, 1976, B, v. **21**, p. 71—75.
290. Moon J. W. A problem on ranking by committees. *Econometrica*, 1976, v. **44**, N 2, p. 241—246.
291. Muller V., Nesetriil F., Pelant J. Either tournaments or algebras. *Discrete Math.*, 1975, v. **11**, N 1, p. 37—66.
292. Narayana T. V., Bent D. H. Computation of the number of score sequence in round-robin tournaments. *Canad. Math. Bull.*, 1964, v. 7, p. 133—136.
293. Narayana T. V. Quelques resultats relatifs aux tournois «knockout» et leurs applications of paired comparisons. *C. r. Acad. Sci.*, 1968, A, v. **267**, p. 32—33.
294. Narayana T. V., Zidek J. Contributions to the theory of tournaments. I. The combinatorics of knock-out tournaments. *Cah. Bur. Univ. Rech. Oper.*, 1969, v. **13**, p. 1—18.
295. Narayana T. V., Zidek J. Contributions to the theory of tournaments. II. Statistical inference in random tournaments. *Rev. Roum. Math. Pures et Appl.*, 1969, v. **10**, p. 1563—1576.
296. Navarick D. J., Fantino E. Stochastic transitivity and unidimensional behavior theories. *Psycholog. Rev.*, 1974, v. **81**, N 5, p.426—441.
297. O'Neill R., Wetherill G. B. The present state of multiple comparison methods (with discussion). *J. Royal Statist. Soc.*, 1971, B, v. **33**, N 2, p.218—255.
298. Phillips J. P. N. On an algorithm of Smith and Payne for determining Slater's i and all nearest adjoining orders. *Brit. J.Math. Statist, Psychol.*, 1976, v. **29**, N 1, p. 126—127.
299. Puri M. L., Sen P. K. Rank tests for some linear hypotheses in paired comparison designs. *Sankhyā*, 1972, A, N 34, p. 257—264.
300. Rai S. C. Ranking in fractional triad comparisons. *Ind. Soc. Agric. Statist.*, 1971, v. **23**, N 7, p. 52—61.

301. Rao P. V., Kupper L. L. Ties in paired-comparison experiments: generalization of the Bradley—Terry model. *J. Amer. Statist. Assoc.*, 1967, v. **62**, N 317, p. 194—204.
302. Ranyard R. N. An algorithm for maximum likelihood ranking and Slater's i from paired comparisons. *Brit. J. Math. Statist. Psychol.*, 1976, v. **29**, N 2, p. 242—248.
303. Remage R., Thompson W. A. Maximum-likelihood paired comparison rankings. *Biometrika*, 1966, v. **53**, p.143—149.
304. Saposnik R. On transitivity of the preference relation under simple majority rule. *J. Econ. Theory*, 1975, v. **10**, N 1, p. 1—7.
305. Sattath S., Tversky A. United and conquer: a multiplicative inequality for choice probabilities. *Econometrica*, 1976, v. **44**, N 1, p. 79—89.
306. Schaafsma W. Paired comparison with order effects. *Ann. Statist.*, 1973, v. **1**, N 6, p.1027—1045.
307. Searls D. T. On the probability of winning with different tournament procedures. *J. Amer. Statist. Assoc.*, 1963, v. **58**, N 304, p. 1064—1081.
308. Sen P. K., David H. A. Paired comparisons for paired characteristics. *Ann. Math. Statist.*, 1968, v. **39**, N 1, p. 200—208.
309. Singh J., Gupta R. S. Derivations at paired comparison model. — In: Applied statistics. Ed. by R. P. Gupta. Amsterdam — Oxford, North—Holland publ. co., 1975, p. 295—300.
310. Singh J. Paired comparison ranking by linear programming. *Communs. Statist.*, 1973, v. **1**, p. 351—364.
311. Singh J., Thompson W. A. A treatment of ties in paired comparisons. *Ann. Math. Statist.*, 1968, v. **39**, p. 2002—2015.
312. Springall A. Responce surface fitting using a generalization of the Bradley—Terry paired comparison model. *Appl. Statist.*, 1973, v. **22**, p. 59—68.
313. Stevens S. S. Psychophysics: Introduction to its perceptual, neural and social aspects. N. Y., Wiley, 1975, v.1.
314. Thompson W. A., Remage R. Rankings from paired comparisons. *Ann. Math. Statist.*, 1964, v. **35**, p. 739—747.
315. Thompson W. A., Singh J. The use of limit theorems in paired comparison model building. *Psychometrika*, 1967, v. **32**, p. 255—264.
316. Travinski B. J. An exact probability distribution over sample spaces of paired comparisons. *Biometrics*, 1965, v. **21**, p. 986—1000.
317. Trawinski B. J. Asymptotic approximation to the expected size of a selected subset. *Biometrika*, 1969, v. **56**, p. 207—213.
318. Trotter W. T., Bogart K. P. Maximal dimensional partially ordered sets III: A characterisation of Hiraguchi's inequality for interval dimensions. *Discrete Math.*, 1977, v. **15**, p. 389—400.
319. Tversky A. Intransitivity of preferences. *Psychol. Rev.*, 1969, v. **76**, p. 31—38.
320. Tversky A. Elimination by aspects: A theory of choice. *Psychol. Rev.*, 1972, v. **79**, p. 281—299.
321. Tversky A. Choice by elimination. *J. Math. Psychol.*, 1972, v. **9**, p. 341—367.
322. Uppuluri V. R., Blot W. J. Asymptotic properties of the number of replications of a paired comparison. *J. Appl. Prob.*, 1974, v. **11**, p. 43—52.
323. Walker W. J. Algebraic and combinatorial results for ranking competitors in a sequence of races. *Discrete Math.*, 1976, v. **14**, p. 297—304.
324. Weisner V. Planung von Turnieren, Kombinatorische Analyse und Algorithmen. Diss. Dokt. Naturwiss. Eidgenöss. Techn. Hochschule, Zürich, 1974, S. 97.
325. Wierenga B. Paired comparison product testing when individual preferences are stochastic. *Appl. Statist*, 1974, v. **23**, N 3, p. 384—396.
326. Williams E. R. Resolvable paired-comparisons designs. *J. Roy. Statist. Soc.*, 1976, **B**, v. **38**, N 2, p. 171—174.
327. Williams E. R., Patterson H. D., John J. A. Resolvable designs with two replications. *J. Roy. Statist. Soc*, 1976, B, v. **38**, p. 296—301.

328. Z i m m e r m a n n H., R a h l t s V. W. Die Erweiterung des Bradley—Terry—Luce Models auf bindungen in Rahmen verallgemeinerter (Log), Linearmodelle. *Biometr. Z.*, 1976, Bd. 18, N 1, S. 23—32.

За время работы над книгой редактору и переводчикам стало известно о существовании нескольких важных работ по парным сравнениям. Кроме того, вышло несколько новых работ, среди которых особенно выделяется обзор [336].

329. М о р к е л ю н а с А. Некоторые понятия теории графов применительно к определению правила групповых решений. — В сб.: Математические методы в социальных науках. Вып. 8. Вильнюс, 1976, с. 39—49.
330. М о р к е л ю н а с А. Нетранзитивные «индивидуальные» бинарные отношения и аксиомы Эрроу. — В сб.: Математические методы в социальных науках. Вып. 8. Вильнюс, 1976, с. 65—71.
331. М о р к е л ю н а с А. О нейтральности и независимости альтернатив в групповых решениях. Liet. mat. rinkinus. Литовский математический сборник, 1977, т. 17, N 1, с. 143—149.
332. C h a k r a b u r t y S. K. A note on the relation between binary and multiple choice probabilities. *Econometrica*, 1969, v. **37**, p.726—727.
333. F a r a r o T. J. Mathematical sociology. N. Y., Wiley, 1973.
334. G e o r g e s c u-R o e g e n N. The relation between binary and multiple choices: some comments and further results. *Econometrica*, 1969, v. **37**, p. 728—730.
335. I n d o w T. On choice probability, *Behaviometrika*, 1975, v. **2**, p.13—31.
336. L u c e R. D. The choice axiom after twenty years. *J. Math. Psychol.*, 1977, v. **15**, N 3, p. 215—233.
337. M c F a d d e n D. Quantal choice analysis: a survey. *Ann. of Economic and Social Measurement*, 1976, v. **5**, p. 363—390.
338. J e m i e l j a n o w S. W., O z i e r n o j W. M., L a r i c z e w O. I. Problemy, metody podejmowania decyzji. Przegląd. *Prace Instytutu organizacji i kierowania*, 1976, N 32.
339. M o r k e l u n a s A. Monotonisity and extension of the simple majority rule. — В сб.: Математические методы в социальных науках. Вып. 8. Вильнюс, 1976, с. 51—63.
340. S k v o r e t z J., W i n d e l l P., F a r a r o T. J. Luce's axiom and occupational prestige: test of a measurement model. *J. Math. Sociology.*, 1974, v. **3**, p.147—162.
341. О с и п о в Г. В. А н д р е е в Э. П. Методы измерения в социологии. М., «Наука», 1978.

Кандидат технических наук *Ю. Адлер, Д. Шмерлинг*

АВТОРСКИЙ УКАЗАТЕЛЬ К ОСНОВНОМУ ТЕКСТУ

agreement between judges — согласованность экспертов; согласие между экспертами
angular transformation — угловое (тригонометрическое) преобразование
asymmetric proximity measure — антисимметрическая мера близости
asymptotic power function — асимптотическая функция мощности
asymptotic relative efficiency — асимптотическая относительная эффективность

balanced incomplete block (BIB) designs — сбалансированные неполноблочные планы
balanced paired-comparison experiment — сбалансированный эксперимент парных сравнений
beat — побеждать, побивать, подавить сопротивление
binary choice — двоичный (парный) выбор
binary comparisons — парные (двоичные) сравнения
binary data — двоичные данные (0 и 1)
branch and bound algorithm — алгоритм метода ветвей и границ
bypasse in tournament — обход в турнире

categorical (qualitative) data — категоризованные (качественные) данные (о частотах попадания в классы)
cell — клетка в таблице, ячейка
chi-square goodness of fit test — критерий согласия хи-квадрат
choice — выбор
— axiom (Luce's) — аксиома выбора (Льюса)
—, theory of — теория выбора
— by elimination — выбор с исключением
circular triad, cyclic triple — циклическая триада, тройка, нетранзитивная тройка
combination of experiments — объединение экспериментов
comparison of tournaments — сравнение турниров (сравнение эффективности турниров)
conquer — побеждать
consistence — совместимость во мнениях, последовательность суждений
—, coefficient of — коэффициент последовательности суждений
consistent arcs — совместимые (согласованные) дуги в турнире
consumer panel testing — тестирование потребителей
contingency table — таблица сопряженности
—, incomplete — неполная таблица сопряженности
cup-tie tournament — состязание на кубок (нокаут-турнир, турнир с выбыванием)
cycle, cyclic path — цикл (в графе)
cyclic designs — циклические планы

degree of a point — степень вершины
defeat — побеждать, наносить поражение
difference of scores — разность очков (меток)
directed walk — ориентированный маршрут в графе
distribution of partitions — распределение размещений по ячейкам
distribution theory in non-null case — теория распределения при ненулевой гипотезе
duo-trio test — двойной-тройной тест
embedding tournament — вложение турнира (в граф)
enumeration of graphs, designs — перечисление графов, планов
equitable system of awards — справедливая система выигрышей, наград
estimates — оценки
least square — — оценки по методу наименьших квадратов (МНК)
maximum likelihood — — оценки по методу максимального правдоподобия (ММП)
factorial treatment combinations — сочетание факторов обработки
frequency data — данные в виде частот (таблицы сопряженности)

gain — побеждать
generalyzed tournament martrix — обобщенная турнирная матрица (матрица вероятностей предпочтения)
goodness of fit test — критерий согласия
graph — граф
directed — (digraph) — ориентированный граф (орграф)
—, mutually connected — связный граф
oriented — — направленный граф (без симметричных пар ориентированных ребер)
—, strongly cyclic edge connected — граф, ориентированно-циклически-реберно связный
group divisible designs — планы с делением на группы

handicapping of tournament — уравнение условий при организации турнира

incomplete data — неполные данные
incomplete designs — неполные планы
inconsistency — несовместимость, непоследовательность во мнениях
indegree — полустепень захода

judges, selection of — отбор экспертов
knock-out plan (k. o. plan) — план турнира с выбыванием

law of comparative judgment — закон сравнительных суждений (Тэрстоуна)
likelihood ratio test — критерий отношения правдоподобия
linked designs — планы со связями, цепные планы
linked paired-comparison designs — планы парного сравнения со сцепленными блоками

majority rule — правило большинства (в групповом выборе)
maximum likelihood ranking — ранжирование по методу максимального правдоподобия
maximum-likelihood paired comparison ranking — ранжирование по методу максимального правдоподобия на основе парных сравнений
multi-binomial test — мульти-биномиальный критерий
merit (worth) — достоинство, «стоимость», ценность
multiple choice — множественный выбор (выбор из 3 и более объектов)
— models — модели множественного выбора
multiple comparison test — критерий множественных сравнений (в дисперсионном анализе)
multivariate paired comparisons — многомерные парные сравнения (парные сравнения по многим критериям)

net gain — чистый выигрыш
nonmetric multidimensional scaling — неметрическое многомерное шкалирование (метод представления данных)
nonreconstructable tournament — неразложимый турнир

one-parameter system of awards — однопараметрическая система наград (зависящая лишь от номера игрока, объекта)
order effect — влияние порядка предъявления объектов эксперту
order of presentation — порядок предъявления стимулов (объектов)
ordinal data — «порядковые» данные (частоты упорядочений)
outdegree — полустепень исхода (вершины графа)

pair test — парный тест
paired compositions — парные композиции, сравнения
pairwise comparisons — парные (попарные) сравнения
panel — группа специалистов
— of experts — группа экспертов
— of consumers — группа потребителей, комиссия потребителей
partial order — частичный порядок
— relation — отношение частичного порядка
partially balanced incomplete block (PBIB) designs — частично сбалансированные неполные планы
partitions — размещения по ячейкам (в комбинаторике)

path — цепь ребер (графа)
personnel rating — оценивание персонала
predilection — предпочтение, пристрастие, склонность к чему-либо
preference — предпочтение (в сравнении)
— analysis of — анализ предпочтений
— behaviour (behavior) — поведение в процессе предпочтения
— matrix — матрица вероятностей предпочтения
— polygon — многоугольник предпочтений, таблица предпочтений
— probabilities — вероятности предпочтения
— table — таблица предпочтений
— testing — проверка (тестирование) предпочтений

quadratic assignment problem — задача о квадратичном назначении

randomness, hypothesis of — гипотеза случайности
ranking — ранжирование, ранжировка
ratings — баллы, оценки (не в статистическом смысле), параметры
resolvable design — разрешимая схема
robustness — устойчивость статистической процедуры при отклонении условий от предполагаемых
round robin tournament — круговой турнир, турнир по круговой системе
row sum — строчная сумма, сумма по одной из строк матрицы

scores — очки (метки)
—, contrasts of — контрасты меток (в дисперсионном анализе)
—, range of — размах меток
—, as sufficient statistics — метки как достаточные статистики
scoring — присуждение, приписывание очков (меток)
seeding of players — рассеивание сильных игроков по разным группам
selection of best judges — отбор лучших экспертов
seriation — упорядочение (объектов)
sequence (directed) — маршрут в графе (ориентированный)
sequential methods — последовательные методы (в статистике)
similarity data — данные о расстояниях, «мерах» близости
simple majority rule — правило простого большинства
simple tournament — простой турнир
size — размер
—, appropriate, of experiment — необходимый объем эксперимента
spanning tree — остов графа
strong digraph — сильный орграф
strongly connected subtournament — сильно связный подтурнир
strongly connected digraph — сильно связный орграф
system of bets (fair), (fair betting system) — «справедливая» система ставок, пари (в турнире)

taste testing — дегустация
Thurstone model — модель типа Тэрстоуна
tie — ничья, связь в ранжировке
top score — наибольшая метка (строчная сумма)
tournaments (double) — двойные турниры
trail, chain — цепь (в графе)
transitive closure graph — транзитивное замыкание графа
transitive triple — транзитивная тройка
transitivity constraint — ограничение транзитивности
transitivity, stochastic — стохастическая транзитивность
—, strong stochastic — сильная стохастическая транзитивность
triangle test — тройной тест
triple comparisons — тройное сравнение

unequal repetitions — неравные повторения
unidimential scale — одномерная шкала
upset in tournament — нарушение транзитивности в турнире

weakly connected digraph — слабо связный граф (ориентированный)
win — выигрывать
winner — победитель

ОГЛАВЛЕНИЕ

Дэвид Г.

Д94 Метод парных сравнений. Пер. с англ. Н. Космарской и Д. Шмерлинга. Под ред. Ю. Адлера. С прил. к русскому переводу. М., «Статистика», 1978.

144 с. с ил. (Б-чка иностр. книг для экономистов и статистиков).

Парные сравнения приобретают все большее значение по мере распространения экспертных методов. Они лежат в основе многих методов упорядочения объектов. В книге рассматриваются вероятностные модели, комбинаторные аспекты теории, статистическая проверка, проблемы планирования эксперимента при парных сравнениях, проблема поиска наилучшего объекта среди заданного множества. В заключительной главе приведены приложения метода парных сравнений. Работа снабжена библиографией и специальными таблицами.

Читателями могут быть специалисты научно-исследовательских экономических институтов, преподаватели и студенты старших курсов вузов.

517.8

Д $\dfrac{\mathbf{10805^1\text{-}091}}{\mathbf{008(01)\text{-}78}}$ **38-78**

[1] Второй индекс 10803.

Г. Дэвид

МЕТОД ПАРНЫХ СРАВНЕНИЙ

Редактор *К. М. Чижевская* Мл. редактор *О. Б. Степанченко*
Корректоры *Я. Б. Островский, И. П. Елкина* Техн. редактор *К. К. Букалова*
Худ. редактор *Т. В. Стихно*

ИБ № 341

Сдано в набор 28/XI 1977 г. Подписано к печати 6/VI 1978 г.
Формат бумаги 60×90[1]/16. Бумага № 2. Объем 9 печ. л. Уч.-изд. л. 10,63
Усл. п. л. 9 Тираж 9500 экз. (Тематич. план 1978 г. № 38)
Заказ № 2179 Цена 65 коп.

Издательство «Статистика», Москва, ул. Кирова, 39.
Московская типография № 4 Союзполиграфпрома
при Государственном комитете СССР
по делам издательств, полиграфии и книжной торговли
Москва И-41, Б. Переяславская ул., д. 46