

561
425

Г. И. ИВЧЕНКО
Ю. И. МЕДВЕДЕВ

МАТЕМАТИЧЕСКАЯ
СТА-
ТИС-
ТИКА

«ВЫСШАЯ ШКОЛА»

Г. И. ИВЧЕНКО
Ю. И. МЕДВЕДЕВ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Допущено Министерством высшего
и среднего специального
образования СССР
в качестве учебного
пособия для студентов
высших технических учебных
заведений

БИБЛИОТЕКА
Бух. ТИП и ЛП
№ 2/4626/3



МОСКВА
«ВЫСШАЯ ШКОЛА» 1984

ББК 22.172
И 25
УДК 519.2

Рецензенты:

кафедра высшей математики Московского высшего
технического училища им. Н. Э. Баумана; проф. В. П. Чистяков
(Математический институт им. В. А. Стеклова АН СССР)

Ивченко Г. И., Медведев Ю. И.

И 25 Математическая статистика: Учеб. пособие для вузов. — М.: Высш. шк., 1984. — 248 с., ил.
В пер.: 85 к.

В пособии систематически, на современном научном уровне, изложены основные разделы статистической теории. Главное внимание уделено исследованию вопросов оптимальности соответствующих статистических процедур и использованию асимптотических методов теории вероятностей. Основные разделы сопровождаются задачами.

Предназначается для студентов вузов. Может быть полезно студентам, обучающимся на факультетах прикладной математики и кибернетики, аспирантам и научным работникам.

1702060000—320
001(01)—84 66—84

ББК 22.172
517.8

*Григорий Иванович Ивченко,
Юрий Иванович Медведев*

Математическая статистика

Зав. редакцией литературы по физике и математике Е. С. Гридасова. Редактор Ж. И. Яковлева. Мл. редакторы: С. А. Доровских, Н. П. Майкова. Оформление и макет художника В. Н. Хомякова. Художественный редактор В. И. Пономаренко. Технический редактор Л. А. Григорчук. Корректор Г. И. Костинова.

ИБ № 4682

Изд. № ФМ-783. Сдано в набор 24.08.83. Подп. в печать 25.05.84. Формат 60×90/16. Бум. тип. № 1. Гарнитура литературная. Печать высокая. Объем 15,5 усл. печ. л. 15,5 усл. кр.-отт. 16,33 уч.-изд. л. Тираж 28 000 экз. Зак. № 1064. Цена 85 коп. Издательство «Высшая школа», 101430, Москва, ГСП-4, Неглинная Ул., д. 29/14

Ордена Октябрьской Революции, ордена Трудового Красного Знамени Ленинградское производственно-техническое объединение «Печатный Двор» имени А. М. Горького Союзполиграфпрома при Государственном комитете СССР по делам издательства, полиграфии и книжной торговли. 197136, Ленинград, П-136, Чкаловский просп., 15.

© Издательство «Высшая школа», 1984

ПРЕДИСЛОВИЕ

В решениях XXVI съезда КПСС и последующих постановлениях партии и правительства перед высшей школой страны поставлены большие и ответственные задачи по совершенствованию системы подготовки высококвалифицированных специалистов для народного хозяйства. Важное место в этой системе занимают вопросы физико-математической подготовки специалистов инженерно-технических профилей. Повышенные требования, предъявляемые к математическому образованию современных инженеров, нашли отражение в новой программе по математике для высших технических учебных заведений. В ней уделено большое внимание вероятностно-статистическим методам ввиду возрастания их практической значимости. В соответствии с этими учебными планами различных инженерных специальностей предусматривается чтение обязательного курса математической статистики. Настоящее учебное пособие предназначено для студентов технических вузов и обеспечивает новую программу такого курса.

Авторы ставили перед собой задачу изложить в доступной для первоначального изучения форме элементы основных направлений современной статистической теории. При изложении материала акцент делается на исследование вопросов оптимальности соответствующих статистических процедур и их практической реализации. В книге широко используются асимптотические методы теории вероятностей, существенное внимание уделяется вопросам прикладной интерпретации решаемых задач и получаемых результатов.

Так как в вузах чтению курса математической статистики предшествует обязательный курс теории вероятностей, то авторы отказались от традиционного принципа, когда значительное место в начале книги отводится перечислению основных фактов и положений теории вероятностей, на которых базируется изложение собственно статистических вопросов. Предполагается, что необходимый минимум сведений из теории вероятностей (например, в объеме учебника В. П. Чистякова [22]) студентам уже известен;

некоторые дополнительные положения теории вероятностей приводятся в соответствующих местах текста. По мнению авторов, пособие позволит формировать курсы лекций различного объема и содержания, а также спецкурсы для различных специализаций и тем самым учитывать в той или иной мере профессиональную ориентацию и структурные особенности конкретного вуза. Книга может быть также полезной студентам университетов, обучающимся на факультетах прикладной математики и кибернетики, аспирантам и инженерам, применяющим в своих исследованиях теоретико-вероятностные и статистические методы (по крайней мере, как введение в соответствующие разделы статистической теории).

Изложение материала ведется на уровне, доступном студентам технических вузов; авторы стремились опираться только на знание студентами основ классического математического анализа и линейной алгебры, читаемых на первых семестрах в курсах высшей математики. Как правило, авторы стремились избегать громоздких математических выкладок и доказательств и в то же время сохранить и подчеркнуть теоретико-вероятностную и статистическую сущность рассматриваемых вопросов.

В конце книги приведен список литературы, в которой более полно и глубоко освещены различные аспекты статистической теории и ее приложений.

В тексте имеется большое число примеров, иллюстрирующих, а в ряде случаев и дополняющих излагаемую теорию. В конце каждой главы приведены задачи, которые могут служить материалом для практических занятий по курсу, а также заданий по учебно-исследовательским и курсовым работам.

В приложении приведены краткие статистические таблицы, необходимые для разбора примеров и решения задач. Знаки $\square$ и $\blacksquare$ в тексте означают соответственно начало и окончание доказательства.

Авторы выражают благодарность В. Я. Козлову, И. Н. Коваленко и В. П. Чистякову за полезные беседы по вопросам преподавания математической статистики и за поддержку идеи написания данного пособия. Весьма полезны были авторам также советы Б. А. Севастьянова, Д. М. Чибисова, А. Н. Ширяева и Г. Д. Карташова, сделанные ими при обсуждении различных разделов курса лекций по математической статистике, и замечания, связанные непосредственно с подготовкой рукописи к печати.

Авторы будут признательны всем, кто в той или иной форме поделится своими соображениями по улучшению ее содержания и стиля изложения материала. Замечания можно направлять по адресу: Москва, Ж-28, Б. Вузовский пер., 3/12, МИЭМ, кафедра теории вероятностей и математической статистики.

Авторы

1. Вероятностно-статистическая модель и задачи математической статистики. Математическая статистика — это прикладная математическая дисциплина, родственная теории вероятностей. Она базируется на понятиях и методах последней, но решает свои специфические задачи своими методами. Любая математическая теория развивается в рамках некоторой модели, описывающей определенный круг реальных явлений, изучением которых и занимается данная теория. Чтобы определить статистическую модель и объяснить специфику задач математической статистики, напомним некоторые факты из теории вероятностей.

Математические модели случайных явлений, изучаемых в теории вероятностей, основываются на понятии *вероятностного пространства* $(\Omega, \mathcal{A}, P)$, где $\Omega = \{\omega\}$ — непустое множество, называемое *пространством элементарных событий* (элементы ω интерпретируются как взаимно исключающие исходы изучаемого случайного явления); $\mathcal{A}$ — некоторая выделенная совокупность подмножеств множества Ω , называемых *событиями* (при этом требуется, чтобы $\mathcal{A}$ было σ -алгеброй, т. е. чтобы $\mathcal{A}$ содержало Ω и было замкнуто относительно операций взятия противоположного события и объединения событий в не более чем счетном числе); P — вероятность, заданная на событиях $A \in \mathcal{A}$ [22, с. 23]. При этом в каждой конкретной ситуации вероятность P считается полностью определенной и основной задачей теории вероятностей является разработка методов нахождения вероятностей различных сложных событий (исходя из известных вероятностей более простых событий) для данной вероятностной модели.

Однако на практике при изучении конкретного эксперимента вероятность P редко бывает известна полностью. Часто можно априори утверждать лишь, что P является элементом некоторого заданного класса (семейства) вероятностей $\mathcal{P}$. Этот класс $\mathcal{P}$ может включать в себя все вероятности, которые можно задать на $\mathcal{A}$ (ситуация полной неопределенности), в других же случаях представляет собой некоторое более узкое семейство вероятностей,

заданное в той или иной явной форме (ситуация, когда имеется определенная априорная информация). В любом случае $\mathcal{F}$ — это совокупность допустимых в данной ситуации (для описания данного эксперимента) вероятностей P . Если задан класс $\mathcal{F}$, то говорят, что имеется *вероятностно-статистическая модель* (или просто *статистическая модель*), понимая под этим набор $(\Omega, \mathcal{A}, \mathcal{F})$.

Приведем пример. Рассмотрим эксперимент, состоящий в проведении n независимых испытаний, в каждом из которых наблюдается либо 1 («успех»), либо 0 («неуспех») с вероятностями соответственно p и $q = 1 - p$ (схема Бернулли). Здесь исход эксперимента можно представить n -мерным вектором $\omega = (e_1, \dots, e_n)$ из 0 и 1, и если вероятность «успеха» p известна, то вероятностной моделью эксперимента является тройка $(\Omega, \mathcal{A}, P)$ с пространством элементарных исходов $\Omega = \{\omega: \omega = (e_1, \dots, e_n), e_i = 0, 1\}$ (содержащим 2^n точек ω), совокупностью $\mathcal{A}$ всех подмножеств Ω и вероятностью P , которая определяется в данном случае вероятностями появления отдельных элементарных исходов:

$$P(\omega) = p^{\sum e_i} q^{n - \sum e_i}.$$

Предположим теперь, что вероятность «успеха» заранее не известна. Обозначая ее через θ , можно только сказать, что $0 \in \Theta = [0, 1]$. Таким образом, здесь семейство допустимых вероятностей имеет вид $\mathcal{F} = \{P_\theta, \theta \in \Theta\}$, где P_θ задается вероятностями $P_\theta(\omega) = \theta^{\sum e_i} (1 - \theta)^{n - \sum e_i}$.

Итак, статистическая модель описывает такие ситуации, когда в вероятностной модели изучаемого эксперимента имеется та или иная неопределенность в задании вероятности P , и задача математической статистики состоит в том, чтобы уменьшить эту неопределенность, используя информацию, доставляемую наблюдаемым исходом эксперимента (статистическими данными). Математическая статистика — это наука о статистических выводах. В определенном смысле математическая статистика решает задачи, обратные задачам теории вероятностей: она уточняет (выявляет) структуру статистических моделей по результатам проводимых наблюдений. В настоящем пособии рассмотрены наиболее типичные статистические задачи и общие методы их решения.

Возникновение и развитие математической статистики, как и других математических дисциплин, определялось потребностями практики; в настоящее время ее методы широко используются в различных технических дисциплинах. Они играют важную роль в экономических исследованиях, сельском хозяйстве, биологии, медицине, физических науках, геологии, психологии, социологических исследованиях и других, считавшихся долго далекими от математики, науках.

Историю статистики как науки о статистических выводах обычно начинают с забавного эпизода, изложенного Ж. Берtrandом в предисловии к его курсу «Исчисление вероятностей»: «Однажды в Неаполе преподобный Галиани увидел человека из Базиликаты, который, встряхивая 3 игральные кости в чашке, держал пари, что выбросит 3 шестерки... Вы скажете, такая удача возможна.

Однако человеку из Базиликаты это удалось во второй раз, и пари повторилось. Он клал кости назад в чашку 3, 4, 5 раз и каждый раз выбрасывал 3 шестерки. «Черт возьми, — вскричал преподобный, — кости налиты свинцом! И так оно и было».

Здесь в шуточной форме дан типичный пример статистического вывода: если бы кости были симметричны, то наблюдаемое событие (5 раз подряд выпали 3 шестерки) имело бы ничтожно малую вероятность $(1/6^3)^5 = 4,71 \cdot 10^{-11}$ и, поскольку в эксперименте наблюдалось такое событие, вполне логично сделать вывод, что априорная модель (симметричность костей) ложна.

Первыми крупными работами, относящимися к математической статистике, были исследования Я. Бернулли и П. Лапласа. К. Гаусс разработал теорию ошибок наблюдений. Научное обоснование закономерностей случайного рассеивания связано с именами русских математиков П. Л. Чебышева, А. А. Маркова и А. М. Ляпунова, значительно углубивших результаты известных к тому времени классических исследований.

Ряд важнейших современных понятий и методов, оказавших большое влияние на развитие современной теории математической статистики, предложил Р. Фишер (метод максимума правдоподобия, дисперсионный анализ, понятия состоятельности, достаточности, эффективности и др.). Систематическое развитие теории проверки статистических гипотез началось с работ К. Пирсона по критерию хи-квадрат. Основной вклад в построение этой теории внесли Ю. Нейман и Э. Пирсон, которые ввели понятия ошибок первого и второго рода и показали общность и значение метода отношения правдоподобия для построения критериев. Общая теория статистических решающих функций, охватывающая как теорию проверки статистических гипотез, так и теорию оценивания, создана А. Вальдом.

Большую роль в развитии математической статистики сыграли работы Г. Крамера, Э. Лемана, С. Рао, М. Кендалла, А. Стьюарта, С. Уилкса, а также советских ученых А. Н. Колмогорова, Б. В. Гнеденко, Ю. В. Линника, Е. Е. Слуцкого, Н. В. Смирнова, В. И. Романовского, Л. Н. Большева, Ю. В. Прохорова и др.

Математическая статистика — это непрерывно и интенсивно развивающаяся наука. В последние годы появилось большое количество глубоких результатов, охватывающих, по существу, все основные направления современной теории математической статистики. Крупный вклад в развитие новых методов этой теории вносит советская математическая школа.

2. Терминология и обозначения. В большинстве случаев исходные статистические данные — результат наблюдения некоторой конечной совокупности случайных величин $X = (X_1, \dots, X_n)$, характеризующей исход изучаемого эксперимента. Обычно в этих случаях говорят, что эксперимент состоит в проведении n испытаний, в которых результат i -го испытания описывается случайной величиной X_i , $i = 1, \dots, n$. Совокупность наблюдаемых случайных величин $X = (X_1, \dots, X_n)$ называется *выборкой*, сами величины X_i , $i = 1, \dots, n$, — *элементами выборки*, а их число n — ее *объемом*. Реализации выборки X будем обозначать соответ-

ствующими строчными буквами $x = (x_1, \dots, x_n)$. Иногда для обозначения наблюдаемых случайных величин используются прописные буквы Y, Z, U, V и т. д., а для их наблюдавшихся значений — соответствующие строчные буквы y, z, u, v и т. д.

Пусть $\mathcal{X} = \{x\}$ — множество, на котором задано распределение случайного вектора X , т. е. множество всех возможных значений выборки X . В статистике множество $\mathcal{X}$ называется *выборочным пространством* (при этом имеется в виду, что на $\mathcal{X}$ задан класс $\mathcal{A}$ подмножеств, образующих σ -алгебру, однако в дальнейшем изложении нет необходимости выделять такой класс). Выборочное пространство может быть либо всем n -мерным евклидовым пространством R^n или его частью (если случайная величина X непрерывна), либо состоять из конечного или счетного числа точек в R^n (если случайная величина X дискретна). В данном случае под статистической моделью эксперимента понимается набор $(\mathcal{X}, \mathcal{F})$, где $\mathcal{F}$ — класс допустимых распределений случайной величины X , заданных на $\mathcal{X}$. Распределение вероятностей любой случайной величины однозначно определяется ее функцией распределения, поэтому обычно статистическая модель задается в терминах допустимых функций распределения выборки X . Таким образом, далее предполагается, что статистическая модель определяется выборочным пространством $\mathcal{X}$ и семейством функций распределения $\mathcal{F}$, которому принадлежит неизвестная функция распределения $F_X(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$, $-\infty < x_1, \dots, x_n < \infty$, выборки $X = (X_1, \dots, X_n)$.

Часто рассматриваются ситуации, когда компоненты $X_1, \dots, X_n$ независимы и все распределены так же, как и некоторая случайная величина ξ . Этот случай соответствует эксперименту, в котором проводятся повторные независимые наблюдения над случайной величиной ξ . Здесь $F_{X_i}(x_i) = F_\xi(x_i)$ для всех $i = 1, \dots, n$ и $F_X(x) = F_\xi(x_1) \dots F_\xi(x_n)$. Такую модель можно задать в терминах функции распределения F_ξ , и в этом случае говорят, что $X = (X_1, \dots, X_n)$ — выборка из распределения случайной величины ξ . Иногда множество возможных значений ξ с распределением F_ξ называют *генеральной совокупностью* (или просто *совокупностью*), имеющей функцию распределения $F_\xi(x)$, а величину X — выборкой из этой совокупности. В дальнейшем распределение ξ обозначается символом $\mathcal{L}(\xi)$. В этих обозначениях предыдущую фразу можно записывать кратко так: « $X = (X_1, \dots, X_n)$ есть выборка из $\mathcal{L}(\xi)$ ». В данном пособии в основном рассматриваются именно такие модели, хотя в ряде задач предположения независимости и одинаковой распределенности компонент X_i не выполняются.

Статистическая модель для повторных независимых наблюдений будет обозначаться кратко в виде $\mathcal{F} = \{F_\xi\}$, т. е. указанием лишь класса допустимых функций распределения исходной случайной величины ξ . Там, где это не приведет к недоразумениям, индекс u для краткости будет опускаться.

Если функции распределения из класса $\mathcal{F}$ заданы с точностью до значений некоторого параметра θ с множеством возможных

значений Θ , то такая модель обозначается $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$ и называется *параметрической*. Говорят, что в этом случае известен тип распределения наблюдаемой случайной величины, а неизвестен только параметр, от которого зависит распределение. Параметр θ может быть как скалярным, так и векторным, множество Θ называется *параметрическим*.

Например, пусть известно, что $\mathcal{L}(\xi)$ — нормальное распределение с известной дисперсией и неизвестным средним. Тогда статистическая модель имеет вид $\mathcal{F} = \{F(x; \theta), \theta \in \Theta = (-\infty, \infty)\}$, где функция распределения $F(x; \theta)$ имеет плотность

$$f(x; \theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\theta)^2/(2\sigma^2)}, \quad -\infty < x < \infty.$$

Если и дисперсия неизвестна, то статистическая модель имеет вид $\mathcal{F} = \{F(x; \theta), \theta = (\theta_1, \theta_2) \in \Theta\}$, где $\Theta = \{(\theta_1, \theta_2): -\infty < \theta_1 < \infty, 0 < \theta_2 < \infty\}$, и $F(x; \theta)$ имеет плотность

$$f(x; \theta) = \frac{1}{\sqrt{2\pi}\theta_2} e^{-(x-\theta_1)^2/(2\theta_2)}, \quad -\infty < x < \infty.$$

Модель $\mathcal{F} = \{F_x\}$ называется *абсолютно непрерывной* или *дискретной*, если таковыми соответственно являются все составляющие класс $\mathcal{F}$ функции распределения. В дальнейшем используется единое обозначение $f_x(x) = f(x)$ (для параметрических моделей $f(x; \theta)$) как для плотности распределения случайной величины ξ в случае абсолютно непрерывной модели, так и для вероятности $P(\xi = x)$ в случае дискретной модели.

Приведем один пример дискретной модели. Предположим, что $\mathcal{L}(\xi)$ — пуассоновское распределение с неизвестным параметром. Тогда статистическая модель имеет вид $\mathcal{F} = \{F(x; \theta), \theta \in \Theta = (0, \infty)\}$, где $F(x; \theta)$ определяется вероятностями

$$f(x; \theta) = P(\xi = x) = e^{-\theta} \frac{\theta^x}{x!}, \quad x = 0, 1, 2, \dots$$

В данном пособии рассматриваются только модели этих двух типов, которые наиболее распространены в приложениях.

В случае параметрической модели распределение вероятностей на выборочном пространстве $\mathcal{X}$, отвечающее параметру θ , обозначается символом P_θ . Аналогично, индекс θ при символах математического ожидания, дисперсии и т. д. означает, что соответствующие величины вычисляются для распределения P_θ . Например, $E_\theta T(X)$, $D_\theta T(X)$ — обозначения соответствующих моментов заданной функции $T(X)$ от выборки X в случае, когда функция распределения выборки есть $F_X(x; \theta)$. При этом для единообразия иногда используется сокращенная запись через интеграл Стильтеса:

$$E_\theta T(X) = \int T(x) dF_X(x; \theta).$$

Здесь интеграл понимается как n -мерный интеграл по всему выборочному пространству $\mathcal{X}$, который в случае абсолютно не-

прерывной модели вычисляется как обычный интеграл Римана $\int T(x_1, \dots, x_n) f(x_1; \theta) \dots f(x_n; \theta) dx_1 \dots dx_n$, а в случае дискретной модели — как соответствующая сумма $\sum T(x_1, \dots, x_n) f(x_1; \theta) \dots f(x_n; \theta)$.

Во многих задачах математической статистики рассматриваются последовательности случайных величин $\{\eta_n\}$, сходящиеся в том или ином смысле к некоторому пределу η (случайной величине или константе), когда $n \rightarrow \infty$. В настоящей книге используются два вида сходимости: сходимость по вероятности и сходимость по распределению, или слабая сходимость. Напомним, что последовательность $\{\eta_n\}$ называется сходящейся по вероятности к η , если

$$\lim_{n \rightarrow \infty} P(|\eta_n - \eta| > \varepsilon) \rightarrow 0, \quad \forall \varepsilon > 0,$$

что кратко записывается так: $\eta_n \xrightarrow{P} \eta$. Под сходимостью по распределению [или кратко: $\mathcal{L}(\eta_n) \rightarrow \mathcal{L}(\eta)$] понимается сходимость

Таблица В.1

Наименование модели	Обозначение модели	Функция $f(x; \theta)$	Параметрическое множество θ
Нормальная-1	$\mathcal{L}(\xi) \in \mathcal{N}(\theta, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}},$ $-\infty < x < \infty$	$\{\theta: -\infty < \theta < \infty\}$
Нормальная-2	$\mathcal{L}(\xi) \in \mathcal{N}(\mu, \theta^2)$	$\frac{1}{\sqrt{2\pi}\theta} e^{-\frac{(x-\mu)^2}{2\theta^2}},$ $-\infty < x < \infty$	$\{\theta: 0 < \theta < \infty\}$
Общая нормальная	$\mathcal{L}(\xi) \in \mathcal{N}(\theta_1, \theta_2^2)$	$\frac{1}{\sqrt{2\pi}\theta_2} e^{-\frac{(x-\theta_1)^2}{2\theta_2^2}},$ $-\infty < x < \infty$	$\{\theta = (\theta_1, \theta_2):$ $-\infty < \theta_1 < \infty$ $0 < \theta_2 < \infty\}$
Гамма	$\mathcal{L}(\xi) \in \Gamma(\theta, \lambda)$	$\frac{x^{\lambda-1} e^{-x/\theta}}{\Gamma(\lambda) \theta^\lambda}, x > 0$	$\{\theta: 0 < \theta < \infty\}$
Равномерная-1	$\mathcal{L}(\xi) \in R(0, \theta)$	$\frac{1}{\theta}, 0 \leq x \leq \theta$	$\{\theta: 0 < \theta < \infty\}$
Общая равномерная	$\mathcal{L}(\xi) \in R(\theta_1, \theta_2)$	$\frac{1}{\theta_2 - \theta_1}, \theta_1 \leq x \leq \theta_2$	$\{\theta = (\theta_1, \theta_2):$ $-\infty < \theta_1 < \theta_2 < \infty\}$
Коши	$\mathcal{L}(\xi) \in \mathcal{K}(\theta)$	$\frac{1}{\pi} \frac{1}{1 + (x-\theta)^2},$ $-\infty < x < \infty$	$\{\theta: -\infty < \theta < \infty\}$
Биномиальная	$\mathcal{L}(\xi) \in Bi(n, \theta)$	$C_n^x \theta^x (1-\theta)^{n-x},$ $x = 0, 1, \dots, n$	$\{\theta: 0 < \theta < 1\}$
Пуассоновская	$\mathcal{L}(\xi) \in \Pi(\theta)$	$e^{-\theta} \frac{\theta^x}{x!},$ $x = 0, 1, 2, \dots$	$\{\theta: 0 < \theta < \infty\}$
Отрицательная биномиальная	$\mathcal{L}(\xi) \in Bi(r, \theta)$	$C_{r+x-1}^x \theta^x (1-\theta)^r,$ $x = 0, 1, 2, \dots$	$\{\theta: 0 < \theta < 1\}$

соответствующих функций распределения в каждой точке непрерывности предельной функции. Известно, что из сходимости по вероятности следует слабая сходимость. Остальные обозначения и понятия вводятся по тексту.

3. Некоторые типичные статистические модели. В теории вероятностей наиболее часто встречающиеся законы распределения имеют общепринятое наименование и обозначение. Так, например, нормальный закон со средним μ и дисперсией σ^2 обозначается символом $\mathcal{N}(\mu, \sigma^2)$; пуассоновский закон со средним λ — символом $P(\lambda)$ и т. д. Если наблюдаемая случайная величина ξ имеет распределение некоторого стандартного типа, то соответствующая статистическая модель имеет такое же наименование. Например, говорят о нормальной модели, пуассоновской модели и т. д. В табл. В.1 приведены наиболее распространенные в приложениях статистические модели, которые встречаются в дальнейшем изложении (функция f равна нулю для не указанных в таблице значений x).

Например, запись $\mathcal{L}(\xi) \in \mathcal{N}(\theta_1, \theta_2)$ для общей нормальной модели означает, что модель задается классом допустимых функций распределения $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$, где плотность $f(x; \theta)$ и параметрическое множество Θ имеют указанный в таблице вид. Первые семь моделей в таблице являются абсолютно непрерывными, а три последние — дискретными.

Отметим также, что некоторые частные случаи приведенных моделей имеют специальные названия. Так, модель $Bi(1, \theta)$ называют *бернуллиевской моделью*, а модель $Bi(1, \theta)$ — *геометрической*.

В заключение несколько слов о физической природе перечисленных распределений. Замечательным фактом является то, что существует несколько распределений большой общности, встречающихся в самых разнообразных задачах теории вероятностей и математической статистики. Прежде всего это нормальное распределение, биномиальное распределение и распределение Пуассона.

Нормальная модель возникает в таких статистических экспериментах, когда на исход эксперимента оказывает влияние большое число независимо действующих случайных факторов, каждый из которых незначительно влияет на конечный результат. В таких ситуациях на основании центральной предельной теоремы теории вероятностей можно заключить, что наблюдаемая случайная величина ξ имеет приблизительно нормальное распределение $\mathcal{N}(\mu, \sigma^2)$, параметры которого (или их часть) могут быть неизвестны. Типичный пример такой ситуации — измерение некоторой физической величины: в теории ошибок измерений (разработанной К. Гауссом) считается, что результирующая ошибка измерения представляет собой сумму большого числа незначительных частных ошибок (связанных, например, с точностью настройки измерительного прибора, погрешностью округления при считывании данных, психофизическим состоянием оператора и т. д.) и является поэтому нормально распределенной случайной величиной.

Биномиальная модель $Bi(n, \theta)$ описывает распределение числа «успехов» в n независимых испытаниях с двумя исходами («успех» — «неуспех») и неизменной вероятностью «успеха» θ (схема Бернулли). Частный случай этой модели — бернуллиевская модель $Bi(1, \theta)$ — особенно часто встречается в приложениях теории вероятностей и математической статистики, так как эксперимент с двумя исходами является простейшим статистическим экспериментом. При этом случайная величина с распределением $Bi(n, \theta)$ может быть реализована как сумма

и независимых случайных величин с одинаковым распределением $Bi(1, \theta)$. Если рассматривать бесконечную последовательность испытаний Бернулли, то число испытаний, предшествующих первому «неуспеху», подчиняется геометрическому распределению $Bi(1, \theta)$, а число «успехов» предшествующих r -му «неуспеху», — отрицательному биномиальному распределению $Bi(r, \theta)$. Случайная величина с распределением $Bi(r, \theta)$ также может быть реализована как сумма r независимых случайных величин с одинаковым распределением $Bi(1, \theta)$. Указанные модели часто используются, например, при разработке математических методов контроля качества промышленной продукции.

Пуассоновская модель $\Pi(\theta)$ обычно описывает *схему редких событий*: при некоторых предположениях о характере процесса появления случайных событий число событий, происшедших за фиксированный промежуток времени или в фиксированной области пространства, часто подчиняется пуассоновскому распределению. Примерами могут служить число частиц радиоактивного распада, зарегистрированных счетчиком в течение некоторого времени t , число вызовов, поступивших на телефонную станцию за время t , число изюминок в кексе и т. д.

Модель гамма-распределения часто используется в задачах теории надежности и теории массового обслуживания. Особую роль играет при этом частный случай $\Gamma(\theta, 1)$, который называют также показательным или *экспоненциальным распределением*. Этому распределению часто подчиняются случайные длины интервалов между последовательными моментами наступления событий в пуассоновских потоках, времена «жизни» различных технических устройств и т. д. При целом $\lambda \geq 1$ распределение $\Gamma(\theta, \lambda)$ называют также *распределением Эрланга порядка λ* . Это распределение суммы λ независимых случайных величин с одинаковым распределением $\Gamma(\theta, 1)$. В задачах математической статистики гамма-распределение играет большую роль благодаря тесной связи с нормальным распределением, в частности там, где рассматриваются квадратичные формы от нормальных случайных величин. Наиболее ярким примером является распределение $\Gamma(2, n/2)$ суммы квадратов n независимых случайных величин с одинаковым распределением $\mathcal{N}(0, 1)$, которое называется *распределением хи-квадрат с n степенями свободы*.

Равномерное распределение $R(a, b)$ описывает процесс «выбора точки наудачу» в интервале (a, b) . Так, если (a, b) — интервал между последовательными отъездами автобусов от остановки, то время ожидания пассажира, прибывшего на остановку, если неизвестно расписание, есть случайная величина с распределением $R(a, b)$.

Распределение $R(0, 1)$ играет особую роль в методах моделирования с помощью ЭВМ случайных величин с заранее заданными распределениями. Такие методы широко используют для приближенных вычислений интегралов, решения дифференциальных и интегральных уравнений и т. д.

Распределение Коши интересно своими связями с другими распределениями, возникает в некоторых физических задачах, связанных с блужданием частиц, имеет ряд оригинальных аналитических свойств. Отметим, в частности, что по закону Коши распределены отношение двух независимых нормальных случайных величин и функция $\operatorname{tg} \eta$, где $\mathcal{L}(\eta) = R(-\pi/2, \pi/2)$; для распределения Коши не существует моментов, в том числе и математического ожидания,



Основные понятия и элементы выборочной теории

В этой главе рассматривается ситуация, соответствующая модели повторных независимых наблюдений над некоторой скалярной случайной величиной. Здесь вводятся основные понятия выборочной теории, которая изучает стохастические свойства случайной выборки; приводятся фундаментальные теоремы математической статистики; исследуются в точной и асимптотической (т. е. при большом объеме выборки) постановках свойства некоторых характеристик случайной выборки; рассматриваются распределения, играющие важную роль в статистике.

§ 1.1. Вариационный ряд выборки и эмпирическая функция распределения

1. Порядковые статистики и вариационный ряд выборки. Пусть $X = (X_1, \dots, X_n)$ — выборка объема n из распределения $\mathcal{L}(\xi)$ и $x = (x_1, \dots, x_n)$ — наблюдавшееся значение X . Каждой реализации x выборки X можно поставить в соответствие упорядоченную последовательность

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}, \quad (1.1)$$

где $x_{(1)} = \min(x_1, \dots, x_n)$, $x_{(2)}$ — второе по величине значение среди $x_1, \dots, x_n$ и т. д., $x_{(n)} = \max(x_1, \dots, x_n)$.

Обозначим через $X_{(k)}$ случайную величину, которая для каждой реализации x выборки X принимает значение $x_{(k)}$, $k = 1, \dots, n$. Так по выборке X определяют новую последовательность случайных величин $X_{(1)}, \dots, X_{(n)}$, называемых *порядковыми статистиками* выборки; при этом $X_{(k)}$ — k -я *порядковая статистика*, а $X_{(1)}$ и $X_{(n)}$ — *экстремальные значения* выборки. Из определения порядковых статистик следует, что они удовлетворяют неравенствам

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}. \quad (1.2)$$

Последовательность (1.2) называют *вариационным рядом* выборки. Симметричные относительно концов элементы последовательности (1.2) $X_{(m)}$ и $X_{(n-m+1)}$ иногда называют соответственно m -м *наименьшим* и m -м *наибольшим* значениями выборки ($m = 1, 2, \dots$); при $m = 1$ получаем экстремальные значения выборки. Итак, вариационный ряд — это расположенные в порядке возрастания их величин элементы выборки. Отметим, что для заданной ре-

лизации $x = (x_1, \dots, x_n)$ выборки $X = (X_1, \dots, X_n)$ реализацией последовательности (1.2) является последовательность (1.1).

2. Эмпирическая функция распределения. Определим для каждого действительного x случайную величину $\mu_n(x)$, равную числу элементов выборки $X = (X_1, \dots, X_n)$, значения которых не превосходят x , т. е.

$$\mu_n(x) = |\{j : X_j \leq x\}|, \quad (1.3)$$

где через $|A|$ обозначено число элементов конечного множества A , и положим $F_n(x) = \mu_n(x)/n$. Функция $F_n(x)$ называется *эмпирической функцией распределения* (соответствующей выборке X). Функцию распределения $F(x)$ наблюдаемой случайной величины ξ в этом случае называют иногда *теоретической функцией распределения*. По своему определению эмпирическая функция распределения — случайная функция: для каждого $x \in R$ значение $F_n(x)$ — случайная величина, реализациями которой являются числа $0, 1/n, 2/n, \dots, (n-1)/n, n/n = 1$, и при этом

$$P(F_n(x) = k/n) = P(\mu_n(x) = k).$$

Но из определения $\mu_n(x)$ следует, что $\mathcal{L}(\mu_n(x)) = Bi(n, p)$, где $p = P(\xi \leq x) = F(x)$. Поэтому

$$P(F_n(x) = k/n) = C_n^k F^k(x) (1 - F(x))^{n-k}, \quad k = 0, 1, \dots, n. \quad (1.4)$$

Итак, эмпирическая функция распределения (как и вариационный ряд) — некоторая сводная характеристика выборки. Для каждой реализации x выборки X функция $F_n(x)$ однозначно определена и обладает всеми свойствами функции распределения: изменяется от 0 до 1, не убывает и непрерывна справа. При этом она кусочно-постоянна и возрастает только в точках последовательности (1.1). Если все компоненты вектора x различны (в последовательности (1.1) все неравенства строгие), то функция $F_n(x)$ задается, очевидно, соотношениями

$$F_n(x) = \begin{cases} 0 & \text{при } x < x_{(1)}, \\ k/n & \text{при } x_{(k)} \leq x < x_{(k+1)}, \\ 1 & \text{при } x \geq x_{(n)}, \end{cases} \quad k = 1, \dots, n-1,$$

т. е. в этом случае величина всех скачков равна $1/n$ и типичный график функции $F_n(x)$ имеет вид, изображенный на рис. 1.1. В общем случае эмпирическую функцию распределения можно записать в виде

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n e(x - X_{(k)}), \quad (1.5)$$

где $e(x)$ — функция единичного скачка (функция Хевисайда): $e(x) = \begin{cases} 0 & \text{при } x < 0, \\ 1 & \text{при } x \geq 0. \end{cases}$ В представлении (1.5) хорошо видна зависимость $F_n(x)$ от выборки X .

Эмпирическая функция распределения играет фундаментальную роль в математической статистике. Важнейшее ее свойство состоит в том, что при увеличении числа испытаний над случайной величиной ξ происходит сближение этой функции с теоретической. Смысл утверждения раскрывает следующая теорема.

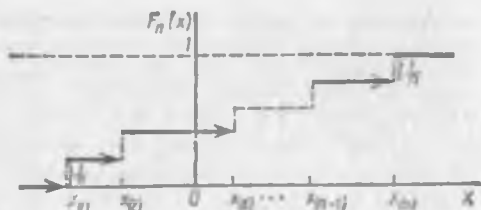


Рис. 1.1

Теорема 1.1. Пусть $F_n(x)$ — эмпирическая функция распределения, построенная по выборке $X = (X_1, \dots, X_n)$ из распределения $\mathcal{L}(\xi)$, и $F(x)$ — соответствующая теоретическая функция распределения. Тогда для любого x ($-\infty < x < \infty$) и любого $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|F_n(x) - F(x)| < \varepsilon) = 1. \quad (1.6)$$

□ Из (1.4) следует, что $F_n(x)$ — относительная частота события $\{\xi \leq x\}$ («успеха») в n испытаниях Бернулли с вероятностью «успеха» $F(x)$. Но в соответствии с законом больших чисел (теоремой Бернулли) относительная частота произвольного события в n независимых испытаниях сходится по вероятности при $n \rightarrow \infty$ к вероятности этого события. Следовательно, по теореме Бернулли, $F_n(x) \xrightarrow{P} F(x)$, т. е. имеет место равенство (1.6). ■

Таким образом, если объем выборки большой, то значение эмпирической функции распределения в каждой точке x может служить приближенным значением (оценкой) теоретической функции распределения в этой точке. Функцию $F_n(x)$ часто называют в этом случае *статистическим аналогом* для $F(x)$.

3. Предельные теоремы для эмпирической функции распределения. Справедлив следующий гораздо более сильный результат, принадлежащий В. И. Гливенко (1933 г.)

Теорема 1.2 (Гливенко). В условиях теоремы 1.1

$$P\left(\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |F_n(x) - F(x)| = 0\right) = 1. \quad (1.7)$$

Другими словами, соотношение (1.7) означает, что отклонение

$$D_n = D_n(X) = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|$$

эмпирической функции распределения от теоретической на всей оси с вероятностью 1 будет сколь угодно мало при достаточно большом объеме выборки.

Приведем еще один результат, принадлежащий А. Н. Колмогорову (1933 г.), который позволяет для больших n оценивать вероятности заданных отклонений случайной величины D_n от 0.

Теорема 1.3 (Колмогорова). Если функция $F(x)$ непрерывна, то при любом фиксированном $t > 0$

$$\lim_{n \rightarrow \infty} P(\sqrt{n} D_n \leq t) = K(t) = \sum_{l=-\infty}^{\infty} (-1)^l e^{-2l^2/t^2}. \quad (1.8)$$

При этом предельную функцию распределения $K(t)$ можно с хорошим приближением использовать для практических расчетов уже при $n \geq 20$.

Теорему Колмогорова обычно применяют для того, чтобы определить границы, в которых с заданной вероятностью находится теоретическая функция распределения $F(x)$, если она неизвестна. Пусть для заданного $\gamma \in (0, 1)$ число t_γ определяется уравнением $K(t_\gamma) = \gamma$. Тогда из (1.8) имеем $P(\sqrt{n} D_n \leq t_\gamma) = P(F_n(x) - t_\gamma/\sqrt{n} \leq F(x) \leq F_n(x) + t_\gamma/\sqrt{n} \text{ для всех } x) \xrightarrow{n \rightarrow \infty} K(t_\gamma) = \gamma$. Таким образом, при больших n с вероятностью, близкой к γ , значения функции $F(x)$ для всех x удовлетворяют неравенствам $F_n(x) - t_\gamma/\sqrt{n} \leq F(x) \leq F_n(x) + t_\gamma/\sqrt{n}$. Так как $0 \leq F(x) \leq 1$, то эти неравенства можно уточнить: $\max(0, F_n(x) - t_\gamma/\sqrt{n}) \leq F(x) \leq \min(F_n(x) + t_\gamma/\sqrt{n}, 1)$. Область, определяемая этими нижней и верхней границами, называется *асимптотической γ -доверительной зоной* для теоретической функции распределения. Для определения числовых значений t_γ при различных γ можно воспользоваться табулированными значениями функции $K(t)$ [2].

Сформулируем еще один важный результат выборочной теории, принадлежащий Н. В. Смирнову (1944 г.), который раскрывает другие свойства эмпирических функций распределения.

Теорема 1.4 (Смирнова). Пусть $F_{n_1}(x)$ и $F_{n_2}(x)$ — две эмпирические функции распределения, построенные на основе двух независимых выборок объемов n_1 и n_2 из одного и того же распределения $\mathcal{L}(\xi)$, и

$$D_{n_1, n_2} = \sup_{-\infty < x < \infty} |F_{n_1}(x) - F_{n_2}(x)|.$$

Тогда если теоретическая функция распределения $F(x)$ непрерывна, то для любого фиксированного $l > 0$

$$\lim_{n_1, n_2 \rightarrow \infty} P(\sqrt{n_1 n_2 / (n_1 + n_2)} D_{n_1, n_2} \leq l) = K(l),$$

где функция $K(l)$ определена равенством (1.8).

Эту теорему обычно используют для проверки предположения (гипотезы) о том, что две выборки получены из одного и того же распределения.

4. Гистограмма и полигон частот. Из изложенного выше следует, что эмпирическая функция распределения — это удобный способ представления статистических данных (выборки X), который позволяет делать выводы о распределении наблюдаемой случайной величины ξ , когда оно неизвестно.

Существуют и другие способы наглядного представления статистических данных; одним из них является построение *гистограммы*. В этом случае область значений наблюдаемой случайной величины ξ разбивают на равные интервалы, для заданной реализации $x = (x_1, \dots, x_n)$ выборки X из $\mathcal{L}(\xi)$ подсчитывают число координат x_i , попавших в соответствующие интервалы (говорят, что данные *группируются*), и на каждом интервале, как на основании, строят прямоугольник с высотой $v/(nh)$, где h — длина интервала, v — число выборочных точек в данном интервале. Получаемую при этом фигуру и называют *гистограммой* (рис. 1.2). Таким образом, площадь каждого прямоугольника равна v/n , т. е. относительной частоте попадания выборочных значений в соответствующий интервал, поэтому по теореме Бернулли она будет сходиться по вероятности при $n \rightarrow \infty$ к вероятности попадания значения случайной величины ξ в соответствующий интервал. Если длина интервалов h достаточно мала, а плотность $f(x)$ непрерывна, то последняя вероятность приблизительно равна $f(z)h$, где z — середина соответствующего интервала. Таким образом, при боль-

шом значении объема выборки n и достаточно малом h высоты построенных прямоугольников можно рассматривать в качестве приближенных значений для плотности f в средних точках соответствующих интервалов. Отсюда следует, что верхнюю границу гистограммы можно рассматривать как статистический аналог плотности распределения наблюдаемой случайной величины. Этот способ представления статистических данных рекомендуется применять только для непрерывных случайных величин; кроме того, он обладает очевидными недостатками: неопределенностью в способе построения интервалов, потерей информации при группировке данных (здесь используются не сами выборочные значения x_i , а лишь частоты попадания в интервалы). Поэтому гистограмму рекомендуется применять на предварительном этапе анализа статистических данных.

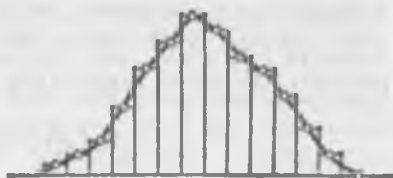


Рис. 1.2

В методе гистограмм неизвестная плотность распределения приближается кусочно-постоянным графиком. Если плотность $f(x)$ — достаточно гладкая, то, как известно из анализа, такие функции значительно лучше можно приблизить кусочно-линейными графиками. Отсюда следует, что для оценки гладких плотностей можно предложить более точную методику, которая основана на построении *полигона частот*. Полигон частот — это ломаная, которую строят так: если построена гистограмма, то ординаты, соответствующие средним точкам интервалов, последовательно соединяют отрезками прямых. Построенный таким образом кусочно-линейный график также является статистическим аналогом теоретической плотности (рис. 1.2).

Отметим, что один из современных подходов к решению задачи оценивания неизвестной плотности распределения наблюдаемой случайной величины по соответствующей выборке основан на использовании *ядерных оценок*. В этом случае в качестве статистического аналога теоретической плотности распределения рассматривают случайную функцию

$$f_n(x) = \frac{1}{na_n} \sum_{i=1}^n k\left(\frac{x - X_i}{a_n}\right) \quad (1.9)$$

при соответствующем выборе ядра $k(x)$ и последовательности чисел $a_n > 0$. Один из результатов, принадлежащий Э. Надарая (1970 г.), состоит в том, что если $k(x)$ — функция с ограниченным изменением, $a_n \rightarrow 0$ при $n \rightarrow \infty$ и при этом $\sum_{n=1}^{\infty} \exp\{-\gamma na_n^2\} < \infty$, $\forall \gamma > 0$, а $f(x)$ равномерно непрерывна на всей оси, то

$$P\left(\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |f_n(x) - f(x)| = 0\right) = 1.$$

БИБЛИОТЕКА

Бух. ТИП и ЛП

№ 5/4626/2

Имеются также результаты, позволяющие при большом n приближенно оценивать вероятности заданных отклонений оценок вида (1.9) от теоретической плотности $f(x)$ и тем самым рассчитывать границы, в которых с заданной вероятностью находится график $f(x)$.

§ 1.2. Выборочные характеристики

1. Выборочные моменты. Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{L}(\xi)$ и $F(x)$ и $F_n(x)$ — соответственно теоретическая и эмпирическая функции распределения. Точно так же, как функции $F(x)$ ставят в соответствие $F_n(x)$, любой теоретической характеристике $g = \int g(x) dF(x)$ можно поставить в соответствие ее статистический аналог $G = G(X)$, вычисляемый по формуле

$$G = \int g(x) dF_n(x) = \frac{1}{n} \sum_{i=1}^n g(X_i).$$

Случайную величину G называют *эмпирической* или *выборочной характеристикой*, соответствующей теоретической характеристике g . Таким образом, выборочная характеристика — это среднее арифметическое значений функции $g(x)$ для элементов выборки X .

Если $g(x) = x^k$, то G — *выборочный момент k -го порядка*, который будем обозначать A_k :

$$A_k = A_k(X) = \frac{1}{n} \sum_{i=1}^n X_i^k. \quad (1.10)$$

При $k=1$ величину A_k называют *выборочным средним* и обозначают символом $\bar{X}$:

$$\bar{X} = A_1 = \frac{1}{n} \sum_{i=1}^n X_i.$$

Значения случайных величин A_k и $\bar{X}$ для заданной реализации x выборки X будем обозначать соответствующими строчными буквами a_k и $\bar{x} = a_1$.

Аналогично, *выборочным центральным моментом k -го порядка* называют случайную величину

$$M_k = M_k(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k. \quad (1.11)$$

При $k=2$ величину M_k называют *выборочной дисперсией* и обозначают символом $S^2 = S^2(X)$:

$$S^2 = M_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Значения случайных величин M_k и S^2 для заданной реализации x выборки X будем обозначать соответствующими строчными буквами m_k и $s^2 = m_2$.

Между выборочными и выборочными центральными моментами сохраняются те же соотношения, что и между теоретическими $\alpha_k = E\xi^k$ и центральными моментами $\mu_k = E(\xi - \alpha_1)^k$. В общем случае при $k \geq 2$ из соотношения (1.11) имеем

$$M_k = \sum_{r=0}^k (-1)^r C_k^r \bar{X}^r A_{k-r} = \sum_{r=0}^{k-2} (-1)^r C_k^r \bar{X}^r A_{k-r} + (-1)^{k-1} (k-1) \bar{X}^k.$$

В частности,

$$S^2 = A_2 - \bar{X}^2, \quad M_3 = A_3 - 3\bar{X}A_2 + 2\bar{X}^3,$$

$$M_4 = A_4 - 4\bar{X}A_3 + 6\bar{X}^2A_2 - 3\bar{X}^4. \quad (1.12)$$

Аналогично вводят выборочные абсолютные моменты, выборочные семинварианты и т. д.

2. Моменты выборочного среднего и выборочной дисперсии. Выборочная характеристика G является случайной величиной, поэтому можно говорить о ее распределении (которое иногда называют *выборочным распределением*) и изучать различные характеристики этого распределения.

Рассмотрим некоторые характеристики распределений среднего $\bar{X}$ и дисперсии S^2 выборки. Так как X_i независимы и распределены так же, как наблюдаемая случайная величина ξ , то

$$E\bar{X} = \frac{1}{n} \sum_{i=1}^n EX_i = E\xi = \alpha_1, \quad D\bar{X} = \frac{1}{n^2} \sum_{i=1}^n DX_i = \frac{1}{n} D\xi = \frac{\mu_2}{n}. \quad (1.13)$$

Вычислим моменты выборочной дисперсии S^2 . Заметим, что при любом c

$$X_i - c - \frac{1}{n} \sum_{j=1}^n (X_j - c) = X_i - \bar{X},$$

т. е. выборочные центральные моменты M_k [см. обозначение (1.11)] не зависят от начала отсчета на шкале значений величины ξ и без ограничения общности можно с самого начала считать, что $\alpha_1 = E\xi = 0$ (в противном случае переходят к случайной величине $\xi' = \xi - \alpha_1$). Учитывая это замечание, из формул (1.12) и (1.13) получаем

$$ES^2 = EA_2 - E\bar{X}^2 = \mu_2 - \mu_2/n = (n-1)\mu_2/n. \quad (1.14)$$

Далее, используя формулы (1.12), представим S^2 в виде

$$S^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n^2} \left(\sum_{i=1}^n X_i \right)^2 = \frac{n-1}{n^2} \sum_{i=1}^n X_i^2 - \frac{2}{n^2} \sum_{i < j} X_i X_j.$$

Тогда, принимая во внимание, что $E\xi = 0$, имеем

$$\begin{aligned} E(S^2)^2 &= E\left\{\frac{(n-1)^2}{n^4} \sum_{i=1}^n X_i^2 + \frac{2(n-1)^2+4}{n^4} \sum_{i<j} X_i^2 X_j^2\right\} = \\ &= \frac{(n-1)^2}{n^3} \mu_4' + \frac{(n-1)^2+2}{n^3} (n-1) \mu_4''. \end{aligned}$$

Отсюда и из соотношения (1.14) окончательно получаем

$$DS^2 = E(S^2)^2 - (ES^2)^2 = \frac{(n-1)^2}{n^3} \left(\mu_4 - \frac{n-3}{n-1} \mu_4' \right). \quad (1.15)$$

Аналогично можно находить моменты и более высоких порядков, хотя с увеличением порядка вид формул и их вывод усложняются.

§ 1.3. Асимптотическое поведение выборочных моментов

1. Сходимость по вероятности выборочных моментов и функций от них. Рассмотрим поведение выборочных моментов A_n , определенных равенством (1.10), при неограниченном возрастании объема выборки n . Чтобы подчеркнуть зависимость моментов A_n от объема выборки, будем их в дальнейшем отмечать индексом n . Первые два момента случайной величины A_{nk} соответственно равны (везде предполагаем, что все соответствующие моменты наблюдаемой случайной величины ξ существуют)

$$EA_{nk} = \frac{1}{n} \sum_{i=1}^n EX_i^k = E\xi^k = \alpha_k, \quad (1.16)$$

$$DA_{nk} = \frac{1}{n^2} \sum_{i=1}^n DX_i^k = \frac{1}{n} D\xi^k = \frac{1}{n} (E\xi^{2k} - (E\xi^k)^2) = \frac{\alpha_{2k} - \alpha_k^2}{n}.$$

На основании неравенства Чебышева отсюда следует, что $A_{nk} \xrightarrow{P} \alpha_k$ при $n \rightarrow \infty$. Таким образом, выборочный момент A_{nk} можно рассматривать в качестве приближенного значения (оценки) соответствующего теоретического момента α_k , когда число наблюдений n достаточно велико. Аналогичное утверждение справедливо и для выборочных центральных моментов и вообще для любых выборочных характеристик, которые имеют вид непрерывных функций от конечного числа величин A_{nk} (центральный момент M_{nk} , как показано в § 1.2, выражается через $A_{n1}, \dots, A_{nk}$ в виде многочлена степени k). Этот вывод является следствием следующей общей теоремы о сходимости функций от случайных величин.

Теорема 1.5. Пусть случайные величины $\eta_1(n), \dots, \eta_r(n)$ сходятся по вероятности при $n \rightarrow \infty$ к некоторым постоянным $c_1, \dots, c_r$ соответственно. Тогда для любой непрерывной функции $\varphi(x_1, \dots, x_r)$ случайная величина

$$\zeta(n) = \varphi(\eta_1(n), \dots, \eta_r(n)) \xrightarrow{P} \varphi(c_1, \dots, c_r).$$

□ Функция φ непрерывна, поэтому для $\forall \varepsilon > 0$ найдется $\delta = \delta(\varepsilon)$ такое, что $|\varphi(x_1, \dots, x_r) - \varphi(c_1, \dots, c_r)| < \varepsilon$ при $|x_i - c_i| < \delta, i = 1, \dots, r$. Введем события $B_i = \{|\eta_i(n) - c_i| < \delta\}, i = 1, \dots, r$. Тогда событие $B = B_1 \dots B_r$ влечет событие $C = \{|\zeta(n) - \varphi(c_1, \dots, c_r)| < \varepsilon\}$. Отсюда

$$P(C) \geq P(B) = 1 - P(\bar{B}) = 1 - P(\bar{B}_1 \cup \dots \cup \bar{B}_r) \geq 1 - \sum_{i=1}^r P(\bar{B}_i). \quad (1.17)$$

Далее, из сходимости по вероятности случайной величины $\eta_i(n)$ имеем, что для данного δ и любого $\gamma > 0$ найдется $n_i = n_i(\gamma)$ такое, что $P(\bar{B}_i) = P(|\eta_i(n) - c_i| \geq \delta) < \gamma/r$ при $n \geq n_i$. Пусть $n_0 = \max(n_1, \dots, n_r)$; тогда при $n \geq n_0$ выполняются одновременно все неравенства $P(\bar{B}_i) < \gamma/r, i = 1, \dots, r$. Следовательно, из (1.17) получим

$$P(|\zeta(n) - \varphi(c_1, \dots, c_r)| < \varepsilon) \geq 1 - \gamma, \quad n \geq n_0. \quad \blacksquare$$

З а м е ч а н и е. Очевидно, что для справедливости теоремы 1.5 достаточно потребовать непрерывности функции φ только в некоторой окрестности точки $(c_1, \dots, c_r)$.

Отметим еще одно конкретное статистическое применение этой теоремы. При исследовании свойств графика плотности распределения непрерывной случайной величины часто рассматривают такие характеристики, как коэффициенты асимметрии γ_1 и эксцесса γ_2 , определяемые формулами

$$\gamma_1 = \mu_3/\mu_1^{3/2}, \quad \gamma_2 = \mu_4/\mu_1^2 - 3. \quad (1.18)$$

Если график плотности распределения симметричен, то $\gamma_1 = 0$ и по значению γ_1 судят о степени отклонения от симметрии. Аналогично, для нормального распределения $\gamma_2 = 0$, поэтому о кривых плотности распределения с $\gamma_2 = 0$ говорят, что они имеют нормальный эксцесс. Если $\gamma_2 > 0$ ($\gamma_2 < 0$), то считают, что эксцесс кривой положителен (отрицателен).

Если задана выборка $X = (X_1, \dots, X_n)$ из абсолютно непрерывного распределения $\mathcal{L}(\xi)$, то выборочные коэффициенты асимметрии Γ_{n1} и эксцесса Γ_{n2} определяются следующими формулами:

$$\Gamma_{n1} = M_{n3}/S_n^3, \quad \Gamma_{n2} = M_{n4}/S_n^4 - 3. \quad (1.19)$$

Из соотношений (1.19) и (1.12) следует, что $\Gamma_{ni}, i = 1, 2$, — непрерывные (при $\mu_2 > 0$) функции от выборочных моментов A_{nk} , поэтому в силу теоремы 1.5 выборочные характеристики (1.19) сходятся по вероятности при $n \rightarrow \infty$ к соответствующим теоретическим характеристикам (1.18).

2. Асимптотическая нормальность выборочных моментов. Введем некоторые дополнительные обозначения. Если распределение случайной величины η_n сходится при $n \rightarrow \infty$ к распределению случайной величины η и при этом $\mathcal{L}(\eta) = \omega_1^*(\mu, \sigma^2)$, то будем писать $\mathcal{L}(\eta_n) \rightarrow \omega_1^*(\mu, \sigma^2)$. Далее иногда будем говорить, что случайная величина η_n асимптотически нормальна с параметрами μ_n

и σ_n^2 или, короче, асимптотически нормальна $\mathcal{L}(\eta_n) \sim \mathcal{N}(\mu_n, \sigma_n^2)$, и записывать это следующим образом: $\mathcal{L}(\eta_n) \sim \mathcal{N}(\mu_n, \sigma_n^2)$. Это означает, что $\mathcal{L}\left(\frac{\eta_n - \mu_n}{\sigma_n}\right) \rightarrow \mathcal{N}(0, 1)$.

Продолжим исследование распределений выборочных характеристик для больших выборок (при $n \rightarrow \infty$). Каждый выборочный момент A_{nh} представляет собой сумму n независимых и одинаково распределенных случайных величин, поэтому к нему можно применить центральную предельную теорему теории вероятностей [22, с. 152]. Имеет место следующее общее утверждение.

Теорема 1.6. Выборочный момент A_{nh} асимптотически нормален $\mathcal{N}(\alpha_h, (\alpha_{2h} - \alpha_h^2)/n)$.

□ Так как [см. формулы (1.16)] $EX_i^h = \alpha_h$, $DX_i^h = \alpha_{2h} - \alpha_h^2$, то по центральной предельной теореме $\mathcal{L}(\eta_n) \rightarrow \mathcal{N}(0, 1)$, где

$$\eta_n = \frac{1}{\sqrt{n(\alpha_{2h} - \alpha_h^2)}} \left(\sum_{i=1}^n X_i^h - n\alpha_h \right) = \sqrt{\frac{n}{\alpha_{2h} - \alpha_h^2}} (A_{nh} - \alpha_h).$$

Следовательно, случайная величина A_{nh} асимптотически нормальна с параметрами α_h и $(\alpha_{2h} - \alpha_h^2)/n$.

Теорема 1.6 позволяет оценивать для больших выборок вероятности заданных отклонений значений выборочных моментов от теоретических. Действительно, из этой теоремы имеем, что при любом фиксированном $t > 0$ и $n \rightarrow \infty$

$$P\left(\sqrt{\frac{n}{\alpha_{2h} - \alpha_h^2}} |A_{nh} - \alpha_h| < t\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-t}^t e^{-x^2/2} dx = 2\Phi(t) - 1 \quad (1.20)$$

(здесь и далее используется обозначение $\Phi(x)$ для функции распределения закона $\mathcal{N}(0, 1)$). В частности, из теоремы 1.6 следует, что выборочное среднее $\bar{X} = A_{n1}$ асимптотически нормально $\mathcal{N}(\alpha_1, \mu_2/n)$. Отметим в этой связи, что если $\mathcal{L}(\xi) = \mathcal{N}(\alpha_1, \mu_2)$, то случайная величина $\bar{X}$, как сумма независимых нормальных случайных величин, также нормальна с параметрами α_1 и μ_2/n , т. е. в этом случае $\mathcal{L}(\bar{X}) = \mathcal{N}(\alpha_1, \mu_2/n)$ при любом n .

Аналогично, используя многомерную форму центральной предельной теоремы [22, с. 154], можно доказать, что совместное распределение любого конечного числа выборочных моментов A_{nh} асимптотически нормально. При этом первые два момента случайных величин A_{nh} определены в (1.16), а смешанные вторые моменты вычисляются аналогично и имеют вид

$$\text{cov}(A_{nh}, A_{ns}) = (\alpha_{h+s} - \alpha_h \alpha_s)/n.$$

Для центральных выборочных моментов также можно доказать утверждение об их асимптотической нормальности. Так например, для выборочной дисперсии S_n^2 при $n \rightarrow \infty$ имеет место соотношение

$$\mathcal{L}(S_n^2) \sim \mathcal{N}(\mu_2, (\mu_4 - \mu_2^2)/n),$$

в правой части которого стоят асимптотические значения среднего и дисперсии величины S_n^2 [см. формулы (1.14) и (1.15)].

§ 1.4. Порядковые статистики

1. Распределение порядковых статистик. При анализе свойств распределения $\mathcal{L}(\xi)$ по соответствующей выборке используются не только выборочные моменты, но и другие характеристики выборки, например порядковые статистики $X_{(k)}$, введенные в § 1.1. Найдем распределение случайной величины $X_{(k)}$ и исследуем его асимптотическое поведение для больших выборок. Предположим, что распределение $\mathcal{L}(\xi)$ абсолютно непрерывное, и пусть, как обычно, $F(x)$ и $f(x) = F'(x)$ — соответственно функция и плотность распределения наблюдаемой случайной величины ξ , а $X = (X_1, \dots, X_n)$ — выборка из $\mathcal{L}(\xi)$.

События $\{X_{(k)} \leq x\}$ и $\{\mu_n(x) \geq k\}$ [см. формулу (1.3)], очевидно, эквивалентны, поэтому

$$F_{X_{(k)}}(x) = P(\mu_n(x) \geq k) = \sum_{m=k}^n C_n^m F^m(x) (1 - F(x))^{n-m}.$$

Дифференцируя это равенство по x , выделяя первое слагаемое при $m=k$ и проводя несложные преобразования, получаем, что плотность $g_k(x) = F'_{X_{(k)}}(x)$ имеет вид

$$g_k(x) = n C_{n-1}^{k-1} F^{k-1}(x) (1 - F(x))^{n-k} f(x). \quad (1.21)$$

2. Выборочные квантили и их асимптотическая нормальность. Введем понятия *p-квантили* распределения $\mathcal{L}(\xi)$ и *выборочной p-квантили*. По определению, для $p \in (0, 1)$ *p-квантилью* ζ_p распределения $\mathcal{L}(\xi)$ называется корень уравнения

$$F(\zeta) = p. \quad (1.22)$$

Если функция $F(x)$ строго монотонна, то это уравнение имеет единственный корень; в противном случае (рис. 1.3) при некоторых p уравнению (1.22) удовлетворяют многие значения ζ (заполняющие целый интервал), тогда в качестве ζ_p берут минимальное из значений ζ , удовлетворяющих уравнению (1.22).

Выборочной *p-квантилью* $Z_{n,p}$ будем называть порядковую статистику $X_{([np] + 1)}$, где $[a]$ — целая часть числа a (для упрощения дальнейшего изложения мы несколько отступаем от строгого определения этого термина:

$$Z_{n,p} = \begin{cases} X_{([np] + 1)} & \text{при } np \text{ дробном,} \\ X_{np} & \text{при } np \text{ целом.} \end{cases}$$

Ясно, что $Z_{n,p}$ — это элемент выборки, левее которого находится доля $[np]/n \leq p$ наблюдений, и при этом $Z_{n,p}$ — порядковая статистика с максимальным номером, обладающая этим свойством. Следовательно, $Z_{n,p}$ можно рассматривать как статистический аналог ζ_p .

Следующая теорема об асимптотическом поведении выборочных квантилей дает дополнительное основание рассматривать $Z_{n,p}$

С другой стороны (см. замечание после теоремы 1.6), соответствующее выборочное среднее $\bar{X}$ при любом n нормально $\mathcal{N}(\mu, \sigma^2/n)$. Отсюда имеем, что если бы μ было неизвестно, то в качестве оценки для него можно было бы использовать (при больших значениях n) как $Z_{n, 1/2}$, так и $\bar{X}$. Однако последняя оценка более точная, так как ее дисперсия в $n/2 = 1,57 \dots$ раза меньше дисперсии $Z_{n, 1/2}$. Систематически вопросы сравнения точности различных оценок для теоретических характеристик исследованы в гл. 2.

Теорема 1.7 описывает асимптотическое поведение для больших выборок, как говорят, *средних членов* вариационного ряда, т. е. порядковых статистик $X_{(k)}$, номера которых удовлетворяют условию $k/n \rightarrow p$ при $n \rightarrow \infty$, где $0 < p < 1$. Таким образом, теорема 1.7 утверждает, что для больших выборок из достаточно гладких распределений средние члены вариационного ряда асимптотически нормальны; более того, средними членами вариационного ряда можно оценивать теоретические квантили произвольных уровней p , $0 < p < 1$. В частности, из этой теоремы следует, что $Z_{n, p} \xrightarrow{d} \xi_p$.

3. Предельные распределения крайних членов вариационного ряда. Исследуем асимптотическое поведение для больших выборок крайних членов вариационного ряда, т. е. порядковых статистик $X_{(n-m+1)}$ и $X_{(m)}$ при фиксированном $m \geq 1$. Заметим, что достаточно изучить m -е наибольшее значение выборки $X_{(n-m+1)}$, так как $X_{(m)} = -X'_{(n-m+1)}$, где $X'_{(1)}, \dots, X'_{(m)}$ — вариационный ряд выборки из $\mathcal{L}(\xi')$ при $\xi' = -\xi$.

Теорема 1.8. Для произвольного фиксированного $m \geq 1$ при $n \rightarrow \infty$

$$\mathcal{L}(n(1 - F(X_{(n-m+1)}))) \rightarrow \Gamma(1, m).$$

□ Пусть

$$\eta_n = n(1 - F(X_{(n-m+1)})) \quad (1.25)$$

и $F^{-1}(t)$, $0 \leq t \leq 1$, — функция, обратная $F(x)$. Тогда $P(\eta_n \leq x) = P(X_{(n-m+1)} \geq F^{-1}(1 - x/n)) = 1 - F_{X_{(n-m+1)}}(F^{-1}(1 - x/n))$ и из формулы (1.21) следует, что плотность распределения этой случайной величины

$$f_n(x) = C_{n-1}^{-1} (x/n)^{n-1} (1 - x/n)^{n-m}, \quad 0 \leq x \leq n.$$

В условиях теоремы при любом фиксированном x

$$f_n(x) \rightarrow x^{m-1} e^{-x} / (m-1)!,$$

т. е. в пределе получаем плотность гамма-распределения $\Gamma(1, m)$. При этом сходимость равномерна по x в каждом конечном интервале; следовательно, имеет место и интегральная сходимость. ■

Обратим внимание на следующие моменты. Во-первых, предельные распределения крайних членов вариационного ряда не являются нормальными (в отличие от средних членов и выборочных моментов). Во-вторых, теорема 1.8 определяет вид предельного распределения не самой порядковой статистики $X_{(n-m+1)}$, а некоторой случайной величины η_n , связанной с $X_{(n-m+1)}$ соотношением (1.25). Наконец, если функция $F(x)$ задана и уравнение (1.25) можно явно разрешить, т. е. получить явное представление

$$F_{(n-m+1)}(F^{-1}(1 - x/n)) = \eta_n/n, \quad (1.26)$$

то можно получить предельное распределение и самой величины $X_{(n-m+1)}$ через предельное распределение η_n . Однако в явном виде обратную функцию $F^{-1}(t)$ удастся найти только в редких случаях, поэтому отыскание предельного

распределения для $X_{(n-m+1)}$, в общем случае сложная задача и ее решение выходит за рамки настоящего пособия. Отметим лишь, что полное решение этой задачи получено Б. В. Гнеденко (1943 г.) и Н. В. Смирновым (1949 г.); приведем некоторые из основных результатов.

Класс предельных распределений для нормированных случайных величин $X_{(n-m+1)}$ (т. е. для $(X_{(n-m+1)} - a_n)/b_n$, где a_n и $b_n > 0$ — некоторые последовательности констант) исчерпывается распределениями следующих трех типов (далее $\pi_r(t) = e^{-t} (t^r/r!)$ и указаны функции распределения соответствующих законов):

$$\Lambda_1^{(m)}(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \sum_{r=0}^{m-1} \pi_r(x^{-\alpha}) & \text{при } x > 0, \end{cases} \quad \alpha > 0;$$

$$\Lambda_2^{(m)}(x) = \begin{cases} \sum_{r=0}^{m-1} \pi_r(|x|^\alpha) & \text{при } x \leq 0, \\ 1 & \text{при } x > 0, \end{cases} \quad \alpha > 0;$$

$$\Lambda_3^{(m)}(x) = \sum_{r=0}^{m-1} \pi_r(e^{-x}), \quad -\infty < x < \infty.$$

Установлены необходимые и достаточные условия, которые следует наложить на функцию $F(x)$, чтобы распределение случайной величины $X_{(n-m+1)}$ сходилось при $n \rightarrow \infty$ к одному из указанных предельных распределений.

Аналогично, три возможных типа предельных распределений для m -го минимального члена $X_{(m)}$ имеют соответственно вид

$$\bar{\Lambda}_1^{(m)}(x) = \begin{cases} 1 - \sum_{r=0}^{m-1} \pi_r(|x|^{-\alpha}) & \text{при } x \leq 0, \\ 1 & \text{при } x > 0, \end{cases} \quad \alpha > 0;$$

$$\bar{\Lambda}_2^{(m)}(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ 1 - \sum_{r=0}^{m-1} \pi_r(x^\alpha) & \text{при } x > 0, \end{cases} \quad \alpha > 0;$$

$$\bar{\Lambda}_3^{(m)}(x) = 1 - \sum_{r=0}^{m-1} \pi_r(e^x), \quad -\infty < x < \infty.$$

В заключение рассмотрим важный случай, когда удастся найти явный вид функции $F^{-1}(t)$, а тем самым и получить предельное распределение случайной величины $X_{(n-m+1)}$.

Пусть $\mathcal{L}(\xi) = \Gamma(1, 1)$. Это распределение задается плотностью $f(x) = e^{-x}$, $x > 0$, или функцией распределения $F(x) = 1 - e^{-x}$, $x > 0$, и называется экспоненциальным распределением. Здесь, очевидно, $F^{-1}(t) = -\ln(1-t)$, $0 < t < 1$, и из соотношения (1.26) имеем $X_{(n-m+1)} = \ln n - \ln \eta_n$. Отсюда по теореме 1.8

$$P(X_{(n-m+1)} - \ln n \leq x) = P(\eta_n \geq e^{-x}) \rightarrow 1 - \int_0^{e^{-x}} \frac{t^{m-1}}{(m-1)!} e^{-t} dt = \sum_{r=0}^{m-1} \pi_r(e^{-x}).$$

Таким образом, в данном случае $\mathcal{L}(X_{(n-m+1)} - \ln n) \rightarrow \Lambda_3^{(m)}$.

§ 1.5. Распределения некоторых функций от нормальных случайных величин

В этом параграфе будут введены и рассмотрены распределения, играющие важную роль в решении различных задач математической статистики. Здесь изучаются распределения функций $\psi(X)$ специального вида от выборки $X = (X_1, \dots, X_n)$ из заданного распределения $\mathcal{L}(\xi)$. Распределение случайной величины $\psi = \psi(X)$ удается вычислить в явном виде в редких случаях. Исключением является случай, когда $\mathcal{L}(\xi)$ — нормальное распределение; тогда соответствующее распределение можно найти для широкого класса функций $\psi(X)$.

С другой стороны, нормальное распределение часто используют в прикладных исследованиях в качестве математической модели изучаемых явлений. Поэтому нормальная модель выделяется среди других статистических моделей и ее исследованию уделяется в математической статистике особое внимание. Примеры решения различных статистических задач для нормальных моделей будут неоднократно встречаться в последующем изложении. Здесь же приводятся некоторые результаты для этой модели, которые будут использоваться в дальнейшем.

1. Распределение хи-квадрат. Среди семейства гамма-распределений (см. табл. В.1) выделим распределение $\Gamma(2, n/2)$, которое называют *распределением хи-квадрат с n степенями свободы* и обозначают $\chi^2(n)$. Соответствующую случайную величину условимся обозначать символом χ^2 . Таким образом, $\mathcal{L}(\chi^2) = \chi^2(n) = \Gamma(2, n/2)$; соответствующую плотность обозначим через $k_n(x)$. Она имеет вид

$$k_n(x) = \frac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)} e^{-x/2}, \quad x > 0. \quad (1.27)$$

Соответствующая этой плотности характеристическая функция

$$\varphi(t; n) = E e^{it\chi^2} = \int_0^\infty e^{itx} \frac{x^{n/2-1} e^{-x/2}}{2^{n/2}\Gamma(n/2)} dx = \frac{1}{(1-2it)^{n/2}}. \quad (1.28)$$

Отсюда получаем первые два момента:

$$E\chi^2 = (1/i)\varphi'(0; n) = n, \quad D\chi^2 = \varphi''(0; n)/i^2 - (E\chi^2)^2 = 2n. \quad (1.29)$$

Важным свойством распределения хи-квадрат является его *воспроизводимость* по параметру n , которое означает, что сумма независимых случайных величин, распределенных по закону хи-квадрат, распределена также по закону хи-квадрат с числом степеней свободы, равным сумме степеней свободы слагаемых. Действительно, если χ_1^2 и χ_2^2 независимы и $\mathcal{L}(\chi_k^2) = \chi^2(n_k)$, $k = 1, 2$, то в силу соотношения (1.28)

$$E e^{it(\chi_1^2 + \chi_2^2)} = \varphi(t; n_1) \varphi(t; n_2) = \frac{1}{(1-2it)^{(n_1+n_2)/2}} = \varphi(t; n_1 + n_2),$$

т. е. $\mathcal{L}(\chi_1^2 + \chi_2^2) = \chi^2(n_1 + n_2)$.

Замечание. Имеет место более общее утверждение: распределение $\Gamma(\theta, \lambda)$ воспроизводимо по параметру λ .

Таким образом, если $\mathcal{L}(\chi^2) = \chi^2(n)$, то случайную величину χ^2 можно представить в виде $\chi^2 = \chi_1^2 + \dots + \chi_n^2$, где случайные величины χ_k^2 , $k = 1, \dots, n$, независимы и $\mathcal{L}(\chi_k^2) = \chi^2(1)$. Но $k_1(x)$ [см. формулу (1.27)] совпадает с плотностью распределения случайной величины ξ^2 , если $\mathcal{L}(\xi) = \mathcal{N}(0, 1)$. В самом деле, $F_{\xi^2}(x) = 0$ при $x < 0$, при $x > 0$

$$F_{\xi^2}(x) = P(\xi^2 \leq x) = P(-\sqrt{x} \leq \xi \leq \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}),$$

$$f_{\xi^2}(x) = F'_{\xi^2}(x) = \Phi'(\sqrt{x})/(2\sqrt{x}) + \Phi'(-\sqrt{x})/(2\sqrt{x}) = k_1(x).$$

Таким образом, сумма квадратов n независимых нормальных $\mathcal{N}(0, 1)$ случайных величин имеет распределение $\chi^2(n)$. Применительно к статистическим задачам этот результат можно сформулировать следующим образом.

Лемма 1.1. Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — выборка из $\mathcal{L}(\xi) = \mathcal{N}(0, 1)$. Тогда

$$\mathcal{L}(X_1^2 + \dots + X_n^2) = \chi^2(n).$$

В дальнейшем понадобится одно важное обобщение распределения хи-квадрат. Рассмотрим независимые случайные величины $\eta_1, \dots, \eta_n$, где $\mathcal{L}(\eta_i) = \mathcal{N}(\mu_i, 1)$, $i = 1, \dots, n$. По лемме 1.1 сумма

$\sum_{i=1}^n (\eta_i - \mu_i)^2$ имеет распределение $\chi^2(n)$. Однако бывает необходимо

знать распределение суммы квадратов самих случайных величин η_i . Распределение случайной величины $U^2 = \eta_1^2 + \dots + \eta_n^2$ называют *нецентральным распределением хи-квадрат с n степенями свободы и параметром нецентральности $\lambda^2 = \mu_1^2 + \dots + \mu_n^2$* и обозначают символом $\chi^2(n; \lambda^2)$: $\mathcal{L}(U^2) = \chi^2(n; \lambda^2)$. Плотность $k_n(x; \lambda^2)$ этого распределения имеет вид

$$k_n(x; \lambda^2) = \frac{x^{n/2-1}}{2^{n/2}} e^{-(x+\lambda^2)/2} \sum_{j=0}^{\infty} \frac{(x\lambda^2)^j}{j! 2^{2j} \Gamma(n/2+j)}, \quad x > 0, \quad (1.30)$$

а первые два момента соответственно равны:

$$EU^2 = n + \lambda^2, \quad DU^2 = 2(n + 2\lambda^2).$$

Если параметр $\lambda^2 = 0$, то $\chi^2(n; \lambda^2) = \chi^2(n)$, т. е. нецентральное распределение хи-квадрат переходит в обычное (центральное) распределение хи-квадрат.

2. Квадратичные и линейные формы от нормальных случайных величин. Пусть $\mathbf{X} = (X_1, \dots, X_n)$ — выборка из $\mathcal{L}(\xi) = \mathcal{N}(0, 1)$.

Рассмотрим квадратичную форму $Q = \sum_{i,j=1}^n a_{ij} X_i X_j = \mathbf{X}' \mathbf{A} \mathbf{X}$, где

матрица $A = \|a_{ij}\|_1^n$ и $A' = A^*$, и m линейных форм $t_k = \sum_{i=1}^n b_{ki} X_i$,

$k = 1, \dots, m$, или в матричных обозначениях $t = BX$. Здесь B — прямоугольная матрица порядка $m \times n$, а $t = (t_1, \dots, t_m)$. Далее будем обозначать через O матрицу с нулевыми элементами и через E_n — единичную матрицу порядка n .

Следующая лемма дает условие независимости случайных функций Q и t .

Лемма 1.2. Если $BA = O$, то функции Q и t независимы.

□ Матрица A действительна и симметрична, поэтому можно найти такую ортогональную матрицу U (т. е. $U \cdot U' = E_n$), что $U'AU = D$. Здесь D — диагональная матрица с элементами $\lambda_i \geq 0$, $i = 1, \dots, n$, являющимися характеристическими числами матрицы A , т. е. корнями характеристического уравнения $\det(A - \lambda E_n) = 0$. Столбцами u_k матрицы $U = \|u_1 \dots u_n\|$ являются собственные векторы матрицы A , т. е. $Au_k = \lambda_k u_k$, $k = 1, \dots, n$.

Пусть r — ранг матрицы A и $\lambda_1, \dots, \lambda_r$ — отличные от 0 характеристические числа. Эквивалентной формой записи

$$A = UDU' \quad (1.31)$$

является спектральное представление матрицы A :

$$A = \sum_{k=1}^r \lambda_k u_k u_k'. \quad (1.32)$$

По условию леммы,

$$O = BA = \sum_{k=1}^r \lambda_k (Bu_k) u_k'.$$

Умножим это равенство на вектор u_s справа. В силу ортогональности векторов u_j ($u_k' u_s = 0$ при $k \neq s$) получим

$$Bu_s = 0, \quad s = 1, \dots, r. \quad (1.33)$$

Рассмотрим теперь случайный вектор $(t_1, \dots, t_m, u_1'X, \dots, u_r'X)$. Этот вектор распределен по нормальному закону, поскольку представляет собой линейное преобразование нормального вектора X . В то же время из представления (1.32) имеем

$$Q = \sum_{k=1}^r \lambda_k (X' u_k) (u_k' X) = \sum_{k=1}^r \lambda_k (u_k' X)^2;$$

следовательно, утверждение будет доказано, если показать, что $\text{cov}(t_i, u_s' X) = 0$, $i = 1, \dots, m$; $s = 1, \dots, r$ (отсюда следует независимость t_i и $u_s' X$). Обозначая через b_i' строки матрицы B ($i =$

* При матричных преобразованиях векторы понимают как вектор-столбцы; знак ' означает транспонирование.

$= 1, \dots, m)$, в силу соотношений (1.33) имеем

$$\begin{aligned} \text{cov}(t_i, u_i'X) &= \text{cov}(b_i'X, u_i'X) = E(b_i'Xu_i'X) - E(b_i'X)E(u_i'X) = \\ &= E(b_i'XX'u_i) = b_i'E(XX')u_i = b_i'E_n u_i = 0. \end{aligned}$$

Здесь и в дальнейшем под математическим ожиданием матрицы $|\xi_{ij}|$, составленной из случайных величин, понимаем равенство $E|\xi_{ij}| = |E\xi_{ij}|$. ■

Рассмотрим две квадратичные формы: $Q_1 = X'AX$ и $Q_2 = X'BX$.

Лемма 1.3. Если $AB = BA = 0$, то Q_1 и Q_2 независимы.

□ Пусть для матрицы A справедливо представление (1.32), а спектральное представление матрицы B имеет вид

$$B = \sum_{i=1}^s v_i v_i' v_i', \quad s = \text{rang } B.$$

По условию, $0 = AB = \sum_{k=1}^r \lambda_k v_k u_k (u_k' v_i) v_i'$. Умножив это равенство

слева на u_i' , а справа на v_j , получим $u_i' v_j = 0$, $i = 1, \dots, r$; $j = 1, \dots, s$, т. е. векторы u_i ортогональны всем векторам v_j . Отсюда, как и выше, имеем, что случайные величины $u_i'X$ и $v_j'X$ не коррелированы, а так как они совместно нормально распределены, то и независимы. Но $Q_1 = \sum_{k=1}^r \lambda_k (u_k'X)^2$, $Q_2 = \sum_{i=1}^s v_i (v_i'X)^2$,

откуда следует их независимость. ■

3. Распределения квадратичных форм от нормальных случайных величин. Обозначим через $\text{tr } A$ след квадратной матрицы A , т. е. сумму ее диагональных элементов. Имеет место следующее утверждение.

Лемма 1.4. Пусть $Q = X'AX$ и $\text{rang } A = r \leq n$. Если матрица A идемпотентна ($A^2 = A$), то $\mathcal{L}(Q) = \chi^2(r)$ и при этом $r = \text{tr } A$.

□ Пусть для A справедливо представление (1.32); тогда из условий симметричности и идемпотентности A следует, что

$$\lambda_1 = \dots = \lambda_r = 1 \text{ и поэтому } Q = \sum_{k=1}^r (u_k'X)^2. \text{ Из ортонормированности}$$

векторов u_k следует, что случайные величины $u_k'X$, $k = 1, \dots, r$, независимы и нормальны $N(0, 1)$; следовательно, $\mathcal{L}(Q) = \chi^2(r)$. Наконец, используя легко проверяемое равенство $\text{tr}(AB) = \text{tr}(BA)$, из формулы (1.31) находим

$$\text{tr } A = \text{tr}(U'UD) = \text{tr } D = \lambda_1 + \dots + \lambda_r = r. \quad \blacksquare$$

Отметим как следствие предыдущих результатов утверждение, представляющее самостоятельный интерес.

Теорема 1.9. Пусть n -мерный случайный вектор Y имеет невырожденное нормальное распределение $N(\mu, \Sigma)$. Тогда квадратичная форма $Q = (Y - \mu)' \Sigma^{-1} (Y - \mu)$ распределена по закону $\chi^2(n)$.

□ Пусть U — ортогональная матрица, приводящая Σ к диагональному виду: $U'\Sigma U = D$. По условию все диагональные элементы λ_i матрицы D положительны, поэтому определена матрица

$D^{-1/2}$ — диагональная матрица с диагональными элементами $\lambda_i^{-1/2}$. Рассмотрим вектор $Z = D^{-1/2}U'(Y - \mu)$. Используя известный факт, что если $\mathcal{L}(Y) = \mathcal{N}(\mu, \Sigma)$ и $V = LY$, где L — заданная матрица линейного преобразования, то $\mathcal{L}(V) = \mathcal{N}(L\mu, L\Sigma L')$, имеем $\mathcal{L}(Z) = \mathcal{N}(0, E_n)$. Далее, $Y - \mu = UD^{1/2}Z$; следовательно,

$$Q = Z'D^{1/2}U'\Sigma^{-1}UD^{1/2}Z = Z'E_nZ = Z'Z.$$

Применяя лемму 1.4 (или лемму 1.1), получаем $\mathcal{L}(Q) = \chi^2(n)$. ■

Докажем теперь следующее важное утверждение выборочной теории, принадлежащее Р. Фишеру (1925 г.).

Теорема 1.10. Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{N}(\mu, \sigma^2)$. Тогда выборочные среднее $\bar{X} = (1/n) \sum_{i=1}^n X_i$ и дисперсия $S^2 = S^2(X) = (1/n) \sum_{i=1}^n (X_i - \bar{X})^2$ независимы; при этом

$$\mathcal{L}(\sqrt{n}(\bar{X} - \mu)/\sigma) = \mathcal{N}(0, 1), \text{ а } \mathcal{L}(nS^2/\sigma^2) = \chi^2(n-1).$$

□ Перейдем к новым случайным величинам $Y_i = (X_i - \mu)/\sigma$, $i = 1, \dots, n$, которые образуют выборку Y из $\mathcal{N}(0, 1)$. Тогда $\bar{Y} = (\bar{X} - \mu)/\sigma$ и $S^2(Y) = (1/\sigma^2) S^2(X)$. Поэтому достаточно доказать, что $\bar{Y}$ и $S^2(Y)$ независимы и при этом $\mathcal{L}(\sqrt{n}\bar{Y}) = \mathcal{N}(0, 1)$, $\mathcal{L}(nS^2(Y)) = \chi^2(n-1)$. Рассмотрим n -мерный вектор-столбец $b = (1/n, \dots, 1/n)'$ и матрицу $B = [b \dots b]$, размер которой $n \times n$. Заметим, что $\bar{Y} = b'Y$, а $nS^2(Y) = (Y - BY)'(Y - BY)$. Отсюда $nS^2(Y) = Y'AY$, где матрица $A = E_n - B$ идемпотентна. Теперь $b'A = b' - b'B = b' - b' = 0$ и, следовательно, по лемме 1.2, $\bar{Y}$ и $S^2(Y)$ независимы.

Закон распределения $\bar{Y}$ очевиден, а так как $\text{tr } A = \text{tr } E_n - \text{tr } B = n - 1$, то на основании леммы 1.4 $\mathcal{L}(nS^2(Y)) = \chi^2(n-1)$. ■

4. Распределение Стьюдента. По определению *распределением Стьюдента с n степенями свободы* $S(n)$ называется распределение случайной величины

$$t = \xi / \sqrt{\chi^2/n}, \quad (1.34)$$

где случайные величины ξ и χ^2 независимы и при этом $\mathcal{L}(\xi) = \mathcal{N}(0, 1)$, $\mathcal{L}(\chi^2) = \chi^2(n)$. Иногда это распределение называют *t-распределением*. Плотность $s_n(x)$ распределения $S(n)$ можно найти с помощью стандартного метода вычисления плотности распределения частного двух независимых случайных величин. Именно: пусть ξ и η — независимые случайные величины с плотностями f_ξ и f_η соответственно. Тогда

$$\begin{aligned} F_{\xi/\eta}(z) &= \iint_{\{x, y: x/y \leq z\}} f_\xi(x) f_\eta(y) dx dy = \\ &= \int_{-\infty}^0 (1 - F_\xi(z\eta)) f_\eta(y) dy + \int_0^\infty F_\xi(z\eta) f_\eta(y) dy. \end{aligned} \quad (1.35)$$

Дифференцируя обе части равенства, получаем

$$f_{\xi/\eta}(z) = \int_{-\infty}^{\infty} f_{\xi}(zy) f_{\eta}(y) |y| dy. \quad (1.36)$$

Если $\eta = \sqrt{\chi^2/n}$, то, очевидно, $F_{\eta}(y) = 0$ при $y < 0$; $F_{\eta}(y) = P(\sqrt{\chi^2/n} \leq y) = P(\chi^2 \leq ny^2) = F_{\chi^2}(ny^2)$ при $y \geq 0$ и

$$f_{\eta}(y) = \frac{d}{dy} F_{\chi^2}(ny^2) = 2ny f_{\chi^2}(ny^2). \quad (1.37)$$

Полагая в (1.36) $\mathcal{L}(\xi) = t^*(0, 1)$, $\eta = \sqrt{\chi^2/n}$ и учитывая соотношения (1.37) и (1.27), нетрудно показать, что плотность распределения Стюдента имеет вид

$$s_n(x) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \frac{1}{(1+x^2/n)^{(n+1)/2}}, \quad -\infty < x < \infty. \quad (1.38)$$

Типичную статистическую ситуацию, где имеет место распределение Стюдента, выражает следующая теорема.

Теорема 1.11. Пусть $X = (X_1, \dots, X_n)$ — выборка из $t^*(\mu, \sigma^2)$ и $t = \sqrt{n-1} \frac{\bar{X} - \mu}{S}$, (1.39)

где $\bar{X}$ и S^2 — как обычно, выборочные среднее и дисперсия. Тогда при любом $\sigma^2 > 0$

$$\mathcal{L}(t) = S(n-1).$$

Эта теорема — очевидное следствие теоремы 1.10 и определения распределения Стюдента.

Тот факт, что определенная соотношением (1.39) дробь t и ее распределение не зависят от параметра σ^2 , используют при получении различных статистических выводов о среднем нормального распределения, когда дисперсия также неизвестна, т. е. является «мешающим» параметром. Подробнее об этом будет идти речь в соответствующих разделах (см. примеры 2.30, 4.13).

В некоторых рассмотренных далее задачах бывает необходимо исключить влияние не только дисперсии σ^2 распределения $t^*(\mu, \sigma^2)$ (как это делают в соответствии с теоремой 1.11), но также и среднего μ . В этом случае можно делать статистические выводы, не зависящие от параметров μ и σ^2 , т. е. являющиеся инвариантными относительно параметров модели. В таких задачах большую роль играет следующее утверждение.

Теорема 1.12. Пусть $X = (X_1, \dots, X_n)$ и $Y = (Y_1, \dots, Y_m)$ — две независимые выборки из одного и того же распределения $t^*(\mu, \sigma^2)$; $\bar{X}$, $S^2(X)$ и $\bar{Y}$, $S^2(Y)$ — соответствующие выборочные средние и дисперсии и пусть

$$t = \sqrt{\frac{mn(m+n-2)}{m+n}} \frac{\bar{X} - \bar{Y}}{\sqrt{nS^2(X) + mS^2(Y)}}. \quad (1.40)$$

Тогда при любых μ и $\sigma^2 > 0$

$$\mathcal{L}(t) = S(m+n-2).$$

□ По теореме 1.10,

$$\mathcal{L}\left(\frac{X-\mu}{\sigma}\right) = \mathcal{L}\left(0, \frac{1}{n}\right), \quad \mathcal{L}\left(\frac{Y-\mu}{\sigma}\right) = \mathcal{L}\left(0, \frac{1}{m}\right),$$

$$\mathcal{L}\left(\frac{n}{\sigma^2} S^2(X)\right) = \chi^2(n-1), \quad \mathcal{L}\left(\frac{m}{\sigma^2} S^2(Y)\right) = \chi^2(m-1).$$

В силу независимости выборок отсюда имеем:

$$\mathcal{L}\left(\frac{X-Y}{\sigma}\right) = \mathcal{L}\left(0, \frac{1}{n} + \frac{1}{m}\right) \text{ или } \mathcal{L}\left(\sqrt{\frac{mn}{m+n}} \frac{X-Y}{\sigma}\right) = \mathcal{L}(0, 1),$$

$$\mathcal{L}\left(\frac{1}{\sigma^2} [nS^2(X) + mS^2(Y)]\right) = \chi^2(m+n-2),$$

при этом случайные величины $X - Y$ и $nS^2(X) + mS^2(Y)$ независимы. Следовательно, в данном случае отношение t имеет вид (1.40). ■

Замечание. Теорема 1.12 допускает обобщение на случай, когда независимые выборки X и Y принадлежат нормальным распределениям с разными математическими ожиданиями μ_1 и μ_2 и одинаковыми дисперсиями. В этом случае вместо студентова отношения (1.40) следует ввести величину

$$t' = \sqrt{\frac{mn(m+n-2)}{m+n}} \frac{(X-\mu_1) - (Y-\mu_2)}{\sqrt{nS^2(X) + mS^2(Y)}}. \quad (1.41)$$

Очевидно, что $\mathcal{L}(X-\mu_1) = N(0, \sigma^2/n)$, $\mathcal{L}(Y-\mu_2) = N(0, \sigma^2/m)$ и для централизованных выборок $(X_1-\mu_1, \dots, X_n-\mu_1)$ и $(Y_1-\mu_2, \dots, Y_m-\mu_2)$ выполнены условия теоремы 1.12. Отсюда получаем, что $\mathcal{L}(t') = S(n+m-2)$ при любых μ_1, μ_2 и $\sigma^2 > 0$.

5. Распределение Снедекора. Пусть случайные величины χ_1^2 и χ_2^2 независимы и $\mathcal{L}(\chi_i^2) = \chi^2(n_i)$, $i=1, 2$. Рассмотрим отношение

$$F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2} = \frac{n_2 \chi_1^2}{n_1 \chi_2^2}. \quad (1.42)$$

Распределение случайной величины F называют *распределением Снедекора* с n_1 и n_2 степенями свободы и обозначают $S(n_1, n_2)$. Иногда это распределение называют *F-распределением* или *распределением дисперсионного отношения Фишера*.

Плотность $f_{n_1, n_2}(x)$ распределения $S(n_1, n_2)$ можно найти обычным методом. Используя формулы (1.36) и (1.27), после несложных преобразований получаем

$$f_{n_1, n_2}(x) = \int_{-\infty}^{\infty} f_{\chi_1^2/n_1}(xy) f_{\chi_2^2/n_2}(y) |y| dy = \int_0^{\infty} n_1 f_{\chi_1^2}(n_1 xy) n_2 f_{\chi_2^2}(n_2 y) y dy =$$

$$= \left(\frac{n_1}{n_2}\right)^{n_1/2} \frac{\Gamma((n_1+n_2)/2)}{\Gamma(n_1/2) \Gamma(n_2/2)} \frac{x^{n_1/2-1}}{(1+n_1 x/n_2)^{(n_1+n_2)/2}}, \quad x > 0. \quad (1.43)$$

Роль F -распределения в выборочной теории в определенной степени раскрывает следующее важное утверждение.

Теорема 1.13. Пусть $X = (X_1, \dots, X_n)$ и $Y = (Y_1, \dots, Y_m)$ — независимые выборки из распределений $\mathcal{L}(\mu_1, \sigma_1^2)$ и $\mathcal{L}(\mu_2, \sigma_2^2)$, $S^2(X)$ и $S^2(Y)$ — соответствующие выборочные дисперсии. Тогда

отношение

$$F = \frac{n(m-1)\sigma_1^2}{m(n-1)\sigma_1^2} \frac{S^2(X)}{S^2(Y)} \quad (1.44)$$

при любых μ_1, μ_2, σ_1^2 и σ_2^2 распределено по закону Снедекора с $n-1$ и $m-1$ степенями свободы.

□ По теореме 1.10, $\mathcal{L}((n/\sigma_1^2)S^2(X)) = \chi^2(n-1)$, $\mathcal{L}((m/\sigma_2^2)S^2(Y)) = \chi^2(m-1)$ и $S^2(X)$ и $S^2(Y)$ независимы по условию. Поэтому формула (1.42) в данном случае имеет вид (1.44). ■

С применением этой теоремы мы встретимся в § 2.6 и гл. 5.

Задачи

1. Пусть $(-0.8; 2.9; 4.4; -5.6; 1.1; -3.2)$ — наблюдавшиеся значения случайной величины. Построить соответствующую эмпирическую функцию распределения $F_n(x)$ и проверить, что $F_n(-5) = 1/6$, $F_n(0) = 1/2$, $F_n(4) = 5/6$.

2. В эксперименте наблюдалась целочисленная случайная величина ξ . Соответствующие выборочные значения оказались равными $(3, 0, 4, 3, 6, 0, 3, 1)$. Построить эмпирическую функцию распределения $F_n(x)$ и проверить, что $F_n(1) = 3/8$, $F_n(3) = 3/4$, $F_n(5) = 7/8$.

3. Вычислить для данных задач 1 и 2 наблюдавшиеся значения выборочных средних и дисперсий.

4. Проводились опыты с бросанием одновременно 12 игральных костей. Наблюдаемую случайную величину ξ брали равной числу костей, на которых выпадало 4, 5 или 6 очков. Пусть h_i — число опытов, в которых наблюдалось значение $\xi = i$ ($i = 0, 1, \dots, 12$). Данные для $n = 4096$ опытов приведены в следующей таблице [9, с.38]:

i	0	1	2	3	4	5	6	7	8	9	10	11	12	Всего
h_i	0	7	60	198	430	731	948	847	536	257	71	11	0	$n = 4096$

Взяв в качестве интервалов $\Delta_i = (i-1/2, i+1/2)$, $i = 0, 1, \dots, 12$, построить гистограмму и полигон частот и вычислить выборочные среднее и дисперсию, а также выборочные коэффициенты асимметрии и эксцесса.

5. Принимая в задаче 4 $\mathcal{L}(\xi) = Bi(12, 1/2)$, оценить вероятность $P(|A_{n1} - \alpha_1| \leq \delta)$, где δ — вычисленное по указанным данным отклонение выборочного среднего от теоретического. Указание. Воспользоваться теоремой 1.6 и соотношением (1.20).

Ответ. $2\Phi(3.7) - 1 = 0.99978$.

6. Вывести формулу для ковариации выборочных моментов: $\text{cov}(A_{nk}, A_{ns}) = (\alpha_{k+s} - \alpha_k \alpha_s)/n$.

7. Доказать, что $\text{cov}(\bar{X}, S^2) = (n-1)\mu_3/n^2$ для выборки объема n .

8. Доказать, что совместная функция распределения двух порядковых статистик $X_{(r)}$ и $X_{(s)}$ ($1 \leq r < s \leq n$) при $x < y$ имеет вид

$$F_{rs}(x, y) = \sum_{i=r}^n \sum_{j=\max(0, s-i)}^{n-i} \frac{n!}{i! j! (n-i-j)!} F^i(x) [F(y) - F(x)]^j [1 - F(y)]^{n-i-j},$$

и при $x \geq y$

$$F_{rs}(x, y) = F_s(y)$$

(здесь $F_s(y)$ — функция распределения для $X_{(s)}$). Показать также, что если распределение $F(x)$ абсолютно непрерывно, то совместная плотность распреде-

ления случайных величин $X_{(r)}$ и $X_{(s)}$ равна

$$g_{rs}(x, y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} F^{r-1}(x) f(x) [F(y) - F(x)]^{s-r-1} \times \\ \times f(y) [1 - F(y)]^{n-s}, \quad x \leq y.$$

9. Доказать, что совместная плотность всех порядковых статистик $X_{(1)}, \dots, X_{(n)}$ выборки объема n из абсолютно непрерывного распределения с плотностью $f(x)$ равна

$$g(x_1, \dots, x_n) = \begin{cases} n! f(x_1) \dots f(x_n) & \text{при } x_1 < x_2 < \dots < x_n, \\ 0 & \text{в остальных случаях.} \end{cases}$$

10. Доказать, что при выполнении условий теоремы 1.7 совместное распределение двух выборочных квантилей Z_{np_1} и Z_{np_2} ($p_1 < p_2$) асимптотически нормально со средними (ξ_{p_1}, ξ_{p_2}) и вторыми моментами $\|\sigma_{ij}/n\|$, где $\sigma_{ij} = = p_i(1-p_j)/[f(\xi_{p_i})f(\xi_{p_j})]$, $i \leq j$.

11. Доказать, что для абсолютно непрерывного распределения $F(x)$ крайние порядковые статистики $X_{(r)}$ и $X_{(n-s+1)}$ при $n \rightarrow \infty$ и фиксированных $r, s \geq 1$ асимптотически независимы. Указание. Перейти к случайным величинам $\eta_r = nF(X_{(r)})$ и $\eta_n = n[1 - F(X_{(n-s+1)})]$ и воспользоваться результатом задачи 8.

12. Пусть $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ — вариационный ряд выборки из экспоненциального распределения с плотностью $f(x) = e^{-x}$, $x > 0$. Доказать, что случайные величины $Y_r = (n-r+1)(X_{(r)} - X_{(r-1)})$, $r = 1, \dots, n$, $X_0 = 0$, независимы и одинаково распределены с плотностью $f(x)$. Получить отсюда формулу

$$EX_{(k)} = \sum_{j=n-k+1}^n 1/j.$$

Указание. Записать совместную плотность распределения величин $X_{(r)}$, $r = 1, \dots, n$, $n! \exp \left\{ - \sum_{r=1}^n x_r \right\}$, $0 \leq x_1 \leq x_2 \leq \dots \leq x_n < \infty$, в виде

$$n! \exp \left\{ - \sum_{r=1}^n (n-r+1)(x_r - x_{r-1}) \right\} \text{ и показать, что}$$

$$P(Y_r \geq t_r, r = 1, \dots, n) = \prod_{r=1}^n \int_{t_r}^{\infty} e^{-y_r} dy_r.$$

13. Пусть $X = (X_1, \dots, X_n)$ — выборка из равномерного на отрезке $[a, b]$ распределения. Показать, что совместная плотность распределения экстремальных значений выборки $X_{(1)}$ и $X_{(n)}$ имеет вид

$$f(x, y) = \frac{n(n-1)}{(b-a)^n} (y-x)^{n-2}, \quad a \leq x \leq y \leq b.$$

Получить отсюда, что

$$EX_{(1)} = (na + b)/(n+1), \quad EX_{(n)} = (nb + a)/(n+1),$$

$$DX_{(1)} = DX_{(n)} = n(b-a)^2/[(n+1)^2(n+2)],$$

$$\text{cov}(X_{(1)}, X_{(n)}) = (b-a)^2/[(n+1)^2(n+2)].$$

14. Доказать, что если $\mathcal{L}(\xi) = S(n)$, то моменты $E\xi^k$ существуют при $k \leq n$ и равны

$$E\xi^{2k} = \frac{1 \cdot 3 \dots (2k-1) n^k}{(n-2)(n-4) \dots (n-2k)}, \quad 2k \leq n, \quad E\xi^{2k+1} = 0, \quad 2k+1 \leq n.$$

15. Доказать, что если $\mathcal{L}(\chi^2) = \chi^2(n)$, то при $n \rightarrow \infty$

$$\mathcal{L}\left(\frac{\chi^2 - n}{\sqrt{2n}}\right) \rightarrow \varphi \in C^\infty(0, 1).$$

У к а з а н и е. Воспользоваться свойством воспроизводимости и применить центральную предельную теорему.

16. Доказать, что если $\mathcal{L}(\xi) = S(n)$ и $P(\xi \leq x) = S_n(x)$, то

$$\lim_{n \rightarrow \infty} S_n(x) = \Phi(x), \quad \lim_{n \rightarrow \infty} s_n(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

17. Говорят, что вектор $\mathbf{v} = (v_1, \dots, v_N)$ с целочисленными неотрицательными компонентами, удовлетворяющими условию $v_1 + \dots + v_N = n$, имеет *полиномиальное* или *мультиномиальное распределение* с параметрами n и $\mathbf{p} = (p_1, \dots, p_N)$ (обозначается $\mathcal{L}(\mathbf{v}) = M(n, \mathbf{p})$), если для $\mathbf{h} = (h_1, \dots, h_N)$

$$P(\mathbf{v} = \mathbf{h}) = \frac{n!}{h_1! \dots h_N!} p_1^{h_1} \dots p_N^{h_N}, \quad h_1 + \dots + h_N = n.$$

а) Показать, что производящая функция для $(v_1, \dots, v_k)$, $k \leq N$, имеет вид

$$E(x_1^{v_1} \dots x_k^{v_k}) = \left[1 + \sum_{i=1}^k p_i (x_i - 1) \right]^n;$$

б) найти производящую функцию линейной комбинации $\sum_{i=1}^N c_i v_i$

в) вывести следующие формулы для моментов: $E(v_i)_k = (n)_k p_i^k$, $E(v_i)_k (v_j)_l$, $i \neq j$, $(n)_{k+l} p_i^k p_j^l$ и вообще, $E((v_1)_{k_1} \dots (v_N)_{k_N}) = n(n-1) \dots (n-k_1-k_2-\dots-k_N) p_1^{k_1} \dots p_N^{k_N}$, где $(z)_a = z(z-1) \dots (z-a+1)$, $(z)_0 = 1$. Получить отсюда, что

$$\text{cov}(v_i, v_j) = \begin{cases} np_i(1-p_i) & \text{при } i=j, \\ -np_i p_j & \text{при } i \neq j; \end{cases}$$

г) пусть $\eta = c_1 v_1 + \dots + c_N v_N$, $\epsilon = \sum_{i=1}^N c_i p_i$, $\sigma^2 = \sum_{i=1}^N c_i^2 p_i - \epsilon^2$. Показать,

что $E\eta = \epsilon n$, $D\eta = \sigma^2 n$;

д) найти совместную производящую функцию двух линейных комбинаций

$$\eta_1 = \sum_{i=1}^N c_i v_i \text{ и } \eta_2 = \sum_{i=1}^N d_i v_i \text{ и вычислить с ее помощью } \text{cov}(\eta_1, \eta_2).$$

Оценивание неизвестных параметров распределений

В настоящей главе в рамках произвольной параметрической статистической модели излагаются некоторые вопросы теории оценок неизвестных параметров модели и функций от них. Рассматриваются два традиционных подхода к решению этих задач: точечное и интервальное оценивание (более общо — оценивание с помощью доверительных множеств). При изложении теории точечного оценивания в основном рассматриваются несмещенные оценки и за меру точности различных оценок принимается величина их дисперсии. Основное внимание уделяется методам построения оптимальных оценок, при этом рассматриваются случаи как скалярного, так и векторного параметра. Излагаются различные подходы для построения доверительных интервалов и доверительных множеств, в том числе принцип отношения правдоподобия — один из наиболее универсальных современных принципов решения многих статистических задач.

§ 2.1. Статистические оценки и общие требования к ним.

Несмещенные оценки с минимальной дисперсией

1. Понятие статистической оценки. Это понятие уже встречалось в гл. 1. Так, говорилось, что значение эмпирической функции распределения в каждой точке можно рассматривать в качестве оценки для значения в этой точке теоретической функции распределения, а различные выборочные характеристики (моменты, квантили и т. д.) — как оценки соответствующих теоретических характеристик. При этом использование термина «оценка» обосновывалось тем, что для выборок большого объема значительная разница между значениями реализаций выборочных характеристик и значениями соответствующих теоретических характеристик маловероятна, и поэтому разумно (по крайней мере для больших выборок) принять выборочную характеристику за приближенное значение соответствующей теоретической характеристики, когда последняя неизвестна. Таким образом, в этот термин вкладывался определенный асимптотический смысл. В то же время в случае применения статистической теории на практике

часто приходится строить приближенные значения для различных неизвестных теоретических характеристик изучаемой модели при любых объемах выборки, в том числе и ограниченных, и при этом обосновывать соответствующие рекомендации с точки зрения каких-либо критериев оптимальности. Общие методы решения подобных задач развиты в теории оценивания неизвестных параметров распределений, которой и посвящена настоящая глава.

Дальнейший анализ проводится в рамках произвольной параметрической статистической модели $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$, соответствующей схеме повторных независимых наблюдений над некоторой случайной величиной ξ . При этом на модель $\mathcal{F}$ по мере необходимости накладываются условия регулярности, обеспечивающие справедливость соответствующих утверждений и выводов.

Пусть имеется выборка $X = (X_1, \dots, X_n)$ из распределения $\mathcal{X}(\xi) \in \mathcal{F}$. Таким образом, априорная информация о наблюдаемой случайной величине ξ состоит в том, что ее функция распределения $F(x; \theta)$ является элементом заданного параметрического семейства функций распределения $\mathcal{F}$ (т. е. имеет известную функциональную форму, но зависит от неизвестного параметра θ , который может быть любой точкой заданного параметрического множества Θ). В общем виде задача оценивания формулируется так: используя статистическую информацию, доставляемую выборкой X , сделать статистические выводы об истинном значении θ^0 неизвестного параметра θ , т. е. оценить точку θ^0 .

Введем следующее общее понятие: *статистикой* называется всякая случайная величина, являющаяся функцией лишь от выборки X .

При *точечном оценивании* ищут статистику $T = T(X)$, значение которой при заданной реализации $x = (x_1, \dots, x_n)$ выборки X принимают за приближенное значение параметра θ^0 . В этом случае говорят, что *статистика $T(X)$ оценивает θ* или что статистика $T(X)$ есть *оценка θ* . Обычно (но не всегда!) область значений оценки T совпадает с Θ .

Интуитивно ясно, что для оценивания θ можно использовать различные оценки, и чтобы выбрать лучшую из них, надо иметь критерий сравнения качества оценок. В свою очередь, и критерии могут быть разными в зависимости от целей, для которых строятся оценки, но любой критерий определяется выбором *меры близости* оценки к истинному значению оцениваемого параметра. Кроме того, обычно класс рассматриваемых оценок ограничивают некоторыми требованиями. Таким образом, если определен класс рассматриваемых в данной ситуации (для данной модели) оценок и выбрана мера близости, то, по определению, оценка, минимизирующая меру близости, является *оптимальной* в этом классе. Ниже рассмотрен один из вариантов этих общих положений, чаще всего применяемый в приложениях.

2. *Несмещенные оценки.* Любая оценка $T = T(X)$ является случайной величиной, поэтому общим требованием к построению оценок является требование концентрации (в том или ином смысле)

распределения T около истинного значения оцениваемого параметра. Чем выше степень этой концентрации, тем лучше соответствующая оценка.

Предположим, что параметр θ — скалярный, и введем понятие несмещенной оценки. По определению, статистика $T(X)$ называется *несмещенной оценкой* для параметра θ , если выполняется условие

$$E_{\theta}T(X) = \theta, \quad \forall \theta \in \Theta. \quad (2.1)$$

Для оценок, не удовлетворяющих условию (2.1), можно ввести величину

$$b(\theta) = E_{\theta}T(X) - \theta,$$

называемую *смещением* оценки $T(X)$. Таким образом, можно также считать, что несмещенные оценки — это такие оценки, для которых смещение $b(\theta) = 0$, $\forall \theta \in \Theta$. Величину

$$E_{\theta}(T - \theta)^2 = D_{\theta}T + b^2(\theta) \quad (2.2)$$

называют *средним квадратом ошибки* или *среднеквадратической ошибкой* оценки T . Для несмещенных оценок среднеквадратическая ошибка совпадает с дисперсией оценки.

Иногда требуется оценить не сам параметр θ , а некоторую функцию от него $\tau(\theta)$ (такие функции называют *параметрическими*). В этом случае статистика $T = T(X)$ является несмещенной оценкой для $\tau(\theta)$, если выполняется соотношение

$$E_{\theta}T = \tau(\theta), \quad \forall \theta \in \Theta. \quad (2.3)$$

Часто ограничиваются рассмотрением класса несмещенных оценок. Это исходное требование интуитивно привлекательно: оно означает, что по крайней мере «в среднем» используемая оценка приводит к желаемому результату. К тому же, как это будет показано далее, для класса несмещенных оценок можно построить достаточно простую и практически полезную теорию, в которой критерием измерения концентрации (точности) оценки является ее дисперсия.

Однако не следует и преувеличивать значение понятия несмещенности: в некоторых случаях требование несмещенности может оказаться слишком «жестким» и привести к нежелательным результатам. Так, может оказаться, что несмещенные оценки (в данной модели и для данной параметрической функции) вообще не существуют, а в других случаях хотя и имеются, но практически бесполезны. Проиллюстрируем эти положения примерами.

Пример 2.1 (*пуассоновская модель, оценивание параметрической функции*). Пусть $\mathcal{L}(\xi) \in \Pi(\theta)$ и $n = 1$, т. е. производится одно наблюдение X над пуассоновской случайной величиной. Требуется оценить параметрическую функцию $\tau(\theta) = 1/\theta$.

Если $T(X)$ — несмещенная оценка $\tau(\theta)$, то условие (2.3) принимает вид

$$\sum_{x=0}^{\infty} T(x) e^{-\theta} \frac{\theta^x}{x!} = \frac{1}{\theta}, \quad \forall \theta \in (0, \infty), \text{ или}$$

$$\sum_{x=0}^{\infty} T(x) \frac{\theta^{x+1}}{x!} = e^{\theta} = \sum_{r=0}^{\infty} \frac{\theta^r}{r!}, \quad \forall \theta \in (0, \infty).$$

Очевидно, что функция $T(x)$, удовлетворяющая последнему условию и не зависящая от θ , как это требуется в соответствии с определением понятия оценки, не существует, т. е. в данном случае несмещенных оценок для $\tau(\theta)$ вообще нет.

Пример 2.2 (отрицательная биномиальная модель, оценивание параметра). Пусть $\mathcal{L}(\xi) \in Bi(1, \theta)$ и $n=1$. Требуется оценить параметр θ .

В данном случае условие несмещенности (2.1) эквивалентно условию

$$\sum_{x=0}^{\infty} T(x) \theta^x = \frac{\theta}{1-\theta} = \sum_{r=1}^{\infty} \theta^r, \quad \forall \theta \in (0, 1).$$

Приравнивая коэффициенты при одинаковых степенях θ , получаем, что единственной несмещенной оценкой в данном случае является

статистика $T(X) = \begin{cases} 0 & \text{при } X=0, \\ 1 & \text{при } X \geq 1. \end{cases}$ Но значения этой статистики

вообще не принадлежат параметрическому множеству $\Theta = (0, 1)$ данной модели, поэтому такая оценка практически бесполезна.

Эти примеры показывают, что не всегда надо ограничиваться рассмотрением только несмещенных оценок. Иногда, как показывает следующий пример, оценка с малым смещением и малой среднеквадратической ошибкой предпочтительнее несмещенной оценки с большой дисперсией.

Пример 2.3 (общая нормальная модель, оценивание дисперсии). Пусть $X = (X_1, \dots, X_n)$, $n \geq 2$, — выборка из $N(\tau(\eta), \theta_1^2)$. Оценим функцию $\tau(\eta) = \tau(\theta_1, \theta_2) = \theta_1^2$.

Из (1.14) имеем, что если $S^2 = S^2(X)$ — выборочная дисперсия, то $E_0(S^2) = (n-1)\theta_1^2/n$. Отсюда следует, что несмещенной оценкой для θ_1^2 является статистика

$$S'^2 = \frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (2.4)$$

При этом по теореме 1.10

$$\mathcal{L}_0\left(\frac{n-1}{\theta_1^2} S'^2\right) = \chi^2(n-1).$$

Отсюда и из (1.29) имеем

$$D_2 \left(\frac{n-1}{\theta_2^2} S'^2 \right) = 2(n-1), \quad \text{или} \quad D_2(S'^2) = \frac{2\theta_2^4}{n-1}. \quad (2.5)$$

Ниже (см. § 2.2, п. 5) будет показано, что по критерию минимума дисперсии оценка S'^2 является оптимальной среди всех несмещенных оценок функции $\tau(\theta) = \theta_2^4$ в модели $\mathcal{F}(\theta_1, \theta_2^2)$, т. е. что ее дисперсия (2.5) меньше дисперсии любой другой несмещенной оценки величины θ_2^4 .

Рассмотрим теперь класс оценок вида $T_\lambda = \lambda S'^2$. Так как $E_2 T_\lambda = \lambda E_2(S'^2) = \lambda \theta_2^4$, то в этом классе имеется лишь единственная несмещенная оценка $\tau(\theta)$, соответствующая значению $\lambda = 1$, т. е. оценка S'^2 .

Вычислим среднеквадратическую ошибку произвольной оценки T_λ :

$$E_2(T_\lambda - \theta_2^4)^2 = E_2[\lambda(S'^2 - \theta_2^4) + (\lambda - 1)\theta_2^4]^2 = \lambda^2 D_2(S'^2) + (\lambda - 1)^2 \theta_2^4.$$

Отсюда и из формулы (2.5) получим

$$E_2(T_\lambda - \theta_2^4)^2 = \left[\frac{2\lambda^2}{n-1} + (\lambda - 1)^2 \right] \theta_2^4 = \varphi(\lambda) \theta_2^4.$$

Функция $\varphi(\lambda)$ достигает минимума при $\lambda^* = (n-1)/(n+1)$ и $\varphi(\lambda^*) = 2/(n+1)$, откуда, учитывая соотношения (2.5), имеем

$$E_2(T_{\lambda^*} - \theta_2^4)^2 = \frac{2}{n+1} \theta_2^4 < \frac{2}{n-1} \theta_2^4 = E_2(S'^2 - \theta_2^4)^2.$$

Таким образом, на основании критерия минимума среднеквадратической ошибки при $n \geq 2$ смещенная оценка $T_{\lambda^*} = \frac{1}{n+1} \sum_{i=1}^n (X_i - \bar{X})^2$,

смещение которой $E_2 T_{\lambda^*} - \theta_2^4 = -\frac{2}{n+1} \theta_2^4$ мало при достаточно боль-

шом объеме выборки n , лучше оценивает дисперсию θ_2^2 модели $\mathcal{F}(\theta_1, \theta_2^2)$, чем несмещенная оценка S'^2 , определяемая равенством (2.4).

Этот пример показывает, что не может быть единственного критерия, по которому следует сравнивать все оценки, как не существует единственной оценки данного параметра, подходящей для всех случаев.

3. Оптимальные оценки. Требование несмещенности в силу сделанных выше замечаний нельзя рассматривать как универсальное, тем не менее во многих встречающихся на практике случаях оно уместно и обоснованно и далее в основном рассматриваются именно несмещенные оценки.

Итак, пусть требуется оценить заданную параметрическую функцию $\tau = \tau(\theta)$ в модели $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$ по статистической информации, доставляемой соответствующей выборкой $\mathbf{X} = (X_1, \dots, X_n)$. Предположим, что в данной задаче существуют несмещенные оценки, т. е. статистики $T = T(\mathbf{X})$, удовлетворяющие условию (2.3). Обозначим класс всех несмещенных оценок

в данной задаче через $\mathcal{F}_\tau$. Таким образом, $T \in \mathcal{F}_\tau$ тогда и только тогда, когда выполнено условие (2.3). Дополнительно предположим, что дисперсии всех оценок из класса $\mathcal{F}_\tau$ конечны: $D_\theta T = E_\theta (T - \tau(\theta))^2 < \infty$, $\forall T \in \mathcal{F}_\tau$ и $\forall \theta \in \Theta$. В этом случае точность оценок можно измерять величиной их дисперсии, и мы получаем простой критерий сравнения различных оценок из класса $\mathcal{F}_\tau$. Пусть T^* и T — оценки из класса $\mathcal{F}_\tau$. Если

$$D_\theta T^* \leq D_\theta T, \quad \forall \theta \in \Theta, \quad (2.6)$$

то по критерию минимума дисперсии оценка T^* равномерно (по параметру θ) не хуже оценки T ; если же в (2.6) имеет место строгое неравенство хотя бы при одном θ , то следует отдать предпочтение T^* , как более точной оценке. Если условие (2.6) выполняется для любой оценки $T \in \mathcal{F}_\tau$, то T^* называют *несмещенной оценкой с равномерно минимальной дисперсией*. Такую оценку T^* в дальнейшем для краткости будем называть *оптимальной* и иногда обозначать τ^* , чтобы подчеркнуть, что она относится к функции $\tau(\theta)$.

Итак, оптимальной является оценка $\tau^* \in \mathcal{F}_\tau$, для которой выполняется условие

$$D_\theta \tau^* = \inf_{\mathcal{F}_\tau} D_\theta T, \quad \forall \theta \in \Theta. \quad (2.7)$$

Требование равномерной минимальности дисперсии сильное и не всегда имеет место. Может оказаться, что из двух оценок $T_1, T_2 \in \mathcal{F}_\tau$ дисперсия $D_\theta T_1$ минимальна (в классе $\mathcal{F}_\tau$) для одних значений параметра θ , а дисперсия $D_\theta T_2$ — для других значений θ . В таких случаях с помощью одного критерия минимума дисперсии эти оценки сравнить нельзя. Однако это требование выделяет оптимальную оценку в классе $\mathcal{F}_\tau$ однозначно, если такая оценка существует, о чем свидетельствует следующая теорема.

Теорема 2.1. Пусть $T_i = T_i(X)$, $i = 1, 2$, — две оптимальные оценки для $\tau = \tau(\theta)$. Тогда $T_1 = T_2$.*

□ Рассмотрим новую оценку $T_3 = (T_1 + T_2)/2$. Ясно, что $T_3 \in \mathcal{F}_\tau$ и

$$D_\theta T_3 = (D_\theta T_1 + D_\theta T_2 + 2 \text{cov}_\theta(T_1, T_2))/4. \quad (2.8)$$

Для любых случайных величин η_1, η_2 имеет место неравенство Коши — Буняковского $|\text{cov}(\eta_1, \eta_2)| \leq \sqrt{D\eta_1 D\eta_2}$, причем знак равенства имеет место только если η_1 и η_2 линейно связаны. Отсюда и из равенства (2.8), положив $D_\theta T_1 = D_\theta T_2 = v = v(\theta)$, получим $D_\theta T_3 \leq (v + |\text{cov}_\theta(T_1, T_2)|)/2 \leq v$. (2.9)

Поскольку T_i ($i = 1, 2$) — оптимальные оценки, $v = D_\theta T_i \leq D_\theta T_3$, откуда $D_\theta T_3 = v$, т. е. T_3 также оптимальная оценка. Но так как

* Здесь и далее равенство статистик понимается в том смысле, что

$$P_\theta(X \in \{x: T_1(x) \neq T_2(x)\}) = 0, \quad \forall \theta \in \Theta.$$

в неравенствах (2.9) имеют место знаки равенства, то $\text{cov}_\theta(T_1, T_2) \geq 0$ и, более того, $\text{cov}_\theta(T_1, T_2) = D_\theta T_1 = v$. Следовательно, T_1 и T_2 линейно связаны, т. е. $T_1 = kT_2 + a$. Из условия несмещенности оценок имеем $\tau = k\tau + a$, т. е. $a = \tau(1 - k)$, и, следовательно, $T_1 - \tau = k(T_2 - \tau)$. Здесь коэффициент $k = k(\theta)$ — функция от параметра θ определяется цепочкой равенств

$$v = \text{cov}_\theta(T_1, T_2) = E_\theta(T_1 - \tau)(T_2 - \tau) = kE_\theta(T_2 - \tau)^2 = kD_\theta T_2 = kv.$$

Отсюда имеем $k \equiv 1$ и, следовательно, $T_1 = T_2$. ■

Приведем пример существования оптимальной оценки в конкретной модели.

Пример 2.4 (бернуллиевская модель, оценивание параметра). Пусть $X = (X_1, \dots, X_n)$ — выборка из $\mathcal{L}(\xi) \in \text{Bl}(1, \theta)$. Требуется оценить параметр θ .

Здесь $E_\theta X_i = \theta$, поэтому выборочное среднее $\bar{X}$ является несмещенной оценкой θ . Более того, из результатов § 1.3 следует, что $\bar{X}$ сходится по вероятности при $n \rightarrow \infty$ к оцениваемому параметру: $\bar{X} \xrightarrow{P_\theta} \theta, \forall \theta \in (0, 1)$. Однако $\bar{X}$ не единственная несмещенная оценка θ . Например, всякая статистика $T = \frac{1}{n} \sum_{i=1}^n b_i X_i$ при

$b_1 + \dots + b_n = n$ также является несмещенной оценкой θ . При этом так как

$$D_\theta T = \frac{1}{n^2} \sum_{i=1}^n b_i^2 \theta (1 - \theta) \leq \frac{b^2}{n} \theta (1 - \theta)$$

при $\max_i |b_i| \leq b < \infty$, то, согласно неравенству Чебышева, $T \xrightarrow{P_\theta} \theta$ при $n \rightarrow \infty$ и, таким образом, эти оценки так же «хороши», как и $\bar{X}$.

Итак, в данной задаче класс $\mathcal{T}_\tau$ содержит много оценок, и поэтому возникает вопрос о выборе среди них наилучшей. Покажем, что в данном случае оптимальная оценка T^* существует и при этом $T^* = \bar{X}$. Имеем $D_\theta \bar{X} = \theta(1 - \theta)/n$, поэтому в соответствии с определением (2.7) достаточно показать, что для любой несмещенной оценки $T = T(X)$ параметра θ

$$D_\theta T \geq \theta(1 - \theta)/n, \quad \forall \theta \in (0, 1). \quad (2.10)$$

В данном случае распределение наблюдаемой случайной величины ξ таково: $f(x; \theta) = \theta^x (1 - \theta)^{1-x}$, $x = 0, 1$; следовательно, распределение случайного вектора $X = (X_1, \dots, X_n)$ задается вероятностями

$$L(x; \theta) = \prod_{i=1}^n f(x_i; \theta) = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}, \quad x = (x_1, \dots, x_n). \quad (2.11)$$

Так как

$$1 = \sum_x L(x; \theta) \text{ и } \theta = E_\theta T(X) = \sum_x T(x) L(x; \theta),$$

то, дифференцируя эти тождества по θ , получаем

$$0 = \sum_x \frac{\partial L(x; \theta)}{\partial \theta} = \sum_x \frac{\partial \ln L(x; \theta)}{\partial \theta} L(x; \theta) = E_0 \left(\frac{\partial \ln L(X; \theta)}{\partial \theta} \right),$$

$$1 = \sum_x T(x) \frac{\partial \ln L(x; \theta)}{\partial \theta} L(x; \theta) = E_0 \left(T(X) \frac{\partial \ln L(X; \theta)}{\partial \theta} \right).$$

Отсюда можно записать

$$1 = E_0 \left[(T(X) - \theta) \frac{\partial \ln L(X; \theta)}{\partial \theta} \right]$$

и, согласно неравенству Коши — Буняковского,

$$1 \leq E_0 (T(X) - \theta)^2 E_0 \left(\frac{\partial \ln L(X; \theta)}{\partial \theta} \right)^2.$$

Но $E_0 (T(X) - \theta)^2 = D_0 T$, поэтому из последнего неравенства следует, что

$$D_0 T \geq 1/E_0 \left(\frac{\partial \ln L(X; \theta)}{\partial \theta} \right)^2, \quad \forall \theta \in (0, 1). \quad (2.12)$$

В рассматриваемом случае [см. формулу (2.11)]

$$\frac{\partial \ln L(x; \theta)}{\partial \theta} = \frac{1}{\theta} \sum_{i=1}^n x_i - \frac{1}{1-\theta} \left(n - \sum_{i=1}^n x_i \right) = \frac{1}{\theta(1-\theta)} \sum_{i=1}^n (x_i - \theta),$$

поэтому

$$\begin{aligned} E_0 \left(\frac{\partial \ln L(X; \theta)}{\partial \theta} \right)^2 &= \frac{1}{\theta^2(1-\theta)^2} E_0 \left[\sum_{i=1}^n (X_i - \theta) \right]^2 = \\ &= \frac{n}{\theta^2(1-\theta)^2} D_0 X_1 = \frac{n}{\theta(1-\theta)}. \end{aligned} \quad (2.13)$$

Отсюда и из неравенства (2.12) получаем соотношение (2.10).

Учитывая важность для приложений бернуллиевской модели $Bi(1, \theta)$, сформулируем доказанный результат в виде теоремы.

Теорема 2.2. *Относительная частота произвольного события в n независимых испытаниях является оптимальной оценкой для вероятности этого события.*

Как следствие этой теоремы отметим, что значение эмпирической функции распределения в каждой точке x является оптимальной оценкой для значения в этой точке теоретической функции распределения (это значительное усиление результата теоремы 1.1).

Докажем важное свойство оптимальных оценок.

Теорема 2.3. *Пусть T_1^* и T_2^* — оптимальные оценки функций $\tau_1 = \tau_1(\theta)$ и $\tau_2 = \tau_2(\theta)$ соответственно. Тогда статистика $T^* = a_1 T_1^* + a_2 T_2^*$ является оптимальной оценкой функции $\tau = a_1 \tau_1 + a_2 \tau_2$ для любых постоянных a_1, a_2 .*

□ Установим сначала следующее свойство оптимальных оценок, представляющее самостоятельный интерес: для любой ста-

истики $\psi = \psi(X)$ с $E_\theta \psi = 0$, $\forall \theta \in \Theta$, выполняются равенства $\text{cov}_\theta(T_i^*, \psi) = 0$, $\forall \theta \in \Theta$. Для доказательства рассмотрим статистику $\tilde{T}_i = T_i^* + \lambda \psi$. При любом λ это несмещенная оценка для τ_i , поэтому в силу оптимальности оценки T_i^*

$$D_0 \tilde{T}_i = D_0 T_i^* + \lambda^2 D_0 \psi + 2\lambda \text{cov}_\theta(T_i^*, \psi) \geq D_0 T_i^*, \quad \forall \theta \in \Theta. \quad (2.14)$$

Отсюда следует, что $\text{cov}_\theta(T_i^*, \psi) = 0$, $\forall \theta$, так как в противном случае при

$$-\frac{|\text{cov}_\theta(T_i^*, \psi)|}{D_0 \psi} - \frac{\text{cov}_\theta(T_i^*, \psi)}{D_0 \psi} < \lambda < \frac{|\text{cov}_\theta(T_i^*, \psi)|}{D_0 \psi} - \frac{\text{cov}_\theta(T_i^*, \psi)}{D_0 \psi}$$

имеем $D_0 \tilde{T}_i < D_0 T_i^*$, что противоречит неравенству (2.14).

Перейдем непосредственно к доказательству теоремы. Пусть T — произвольная несмещенная оценка τ . Тогда $\psi = T^* - T$ имеет нулевое математическое ожидание и, по предыдущему,

$$0 = a_1 \text{cov}_\theta(T_i^*, \psi) + a_2 \text{cov}_\theta(T_i^*, \psi) = \text{cov}_\theta(T^*, \psi) = D_0 T^* - \text{cov}_\theta(T^*, T),$$

т. е. $D_0 T^* = \text{cov}_\theta(T^*, T) \leq \sqrt{D_0 T^* D_0 T}$ или $D_0 T^* \leq D_0 T$, $\forall \theta$. ■

§ 2.2. Критерии оптимальности оценок, основанные на неравенстве Рао — Крамера и его обобщениях

В этом и следующем параграфах будут рассмотрены общие критерии существования оптимальных оценок и способы их нахождения.

1. Понятия функции правдоподобия, вклада выборки, функции информации. Пусть, как обычно, $f(x; \theta)$ — плотность распределения наблюдаемой случайной величины ξ (или вероятность в дискретном случае), $X = (X_1, \dots, X_n)$ — выборка из $\mathcal{X}(\xi) \in \mathcal{F}$ и $x = (x_1, \dots, x_n)$ — реализация X . Функция $L(x; \theta) = f(x_1; \theta) \dots f(x_n; \theta)$ является, очевидно, плотностью распределения случайного вектора X . Функция $L(x; \theta)$, рассматриваемая при фиксированном x как функция параметра $\theta \in \Theta$, называется *функцией правдоподобия*.

В дальнейшем предполагается, что функция $L(x; \theta) > 0$ при всех $x \in \mathcal{X}$ и $\theta \in \Theta$ и дифференцируема по параметру θ . Пусть параметр θ — скалярный. Случайная величина

$$U(X; \theta) = \frac{\partial \ln L(X; \theta)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \ln f(X_i; \theta)}{\partial \theta} \quad (2.15)$$

называется *вкладом* (или *функцией вклада*) *выборки* X (i -е слагаемое в правой части (2.15) называется *вкладом* i -го наблюдения, $i = 1, \dots, n$). В дальнейшем предполагается, что

$$0 < E_0 U^2(X; \theta) < \infty, \quad \forall \theta \in \Theta.$$

Для случая векторного параметра $\theta = (\theta_1, \dots, \theta_r)$ под вкладом выборки понимается случайный вектор $(U_1(X; \theta), \dots, U_r(X; \theta))$, где $U_i(X; \theta) = \frac{\partial \ln L(X; \theta)}{\partial \theta_i}$, $i = 1, \dots, r$.

В последующем неоднократно придется дифференцировать по θ интегралы от функций на выборочном пространстве $\mathcal{X}$, а также предполагать, что при этом можно менять порядок интегрирования и дифференцирования. Модели, для которых все перечисленные условия выполняются, обычно называют кратко *регулярными*. Точные аналитические условия, обеспечивающие регулярность модели, известны из математического анализа и вид их определяется в каждом конкретном случае. Отметим, в частности, общее необходимое условие, состоящее в том, что выборочное пространство $\mathcal{X}$ не должно зависеть от неизвестного параметра θ .

Рассмотрим некоторые свойства вклада $U(X; \theta)$ для регулярных моделей. Всегда имеет место тождество (по θ)

$$\int L(x; \theta) dx = 1 \quad (dx = dx_1 \dots dx_n)$$

(здесь и далее для дискретных моделей интегрирование заменяется суммированием). Если модель регулярна, то, дифференцируя это тождество по θ , получаем

$$0 = \int \frac{\partial L(x; \theta)}{\partial \theta} dx = \int \frac{\partial \ln L(x; \theta)}{\partial \theta} L(x; \theta) dx = E_{\theta} U(X; \theta).$$

Таким образом, для регулярной модели

$$E_{\theta} U(X; \theta) = 0, \quad \forall \theta \in \Theta. \quad (2.16)$$

Определим теперь *функцию информации Фишера*, или просто *информацию Фишера* о параметре θ , содержащуюся в выборке X :

$$i_1(\theta) = D_{\theta} U(X; \theta) = E_{\theta} U^2(X; \theta), \quad (2.17)$$

играющую важную роль в статистике. Величину

$$i(\theta) = i_1(\theta) = E_{\theta} \left(\frac{\partial \ln f(X_1; \theta)}{\partial \theta} \right)^2 \quad (2.18)$$

называют также *количеством (фишеровской) информации, содержащейся в одном наблюдении*. Из соотношений (2.15) — (2.18) непосредственно следует, что $i_n(\theta) = ni(\theta)$, т. е. количество информации, содержащейся в выборке, возрастает пропорционально объему выборки.

Если функция $f(x; \theta)$ дважды дифференцируема по θ , то, продифференцировав при $n=1$ выражение (2.16) еще раз, получим эквивалентное представление для $i(\theta)$:

$$0 = \int \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} f(x; \theta) dx + \int \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 f(x; \theta) dx, \text{ т. е.} \\ i(\theta) = -E_{\theta} \left(\frac{\partial^2 \ln f(X_1; \theta)}{\partial \theta^2} \right). \quad (2.19)$$

Пример вычисления информации Фишера для бернуллиевской модели $Bi(1, \theta)$ имеется в § 2.1 [формула (2.13)]; для этой модели $i(\theta) = 1/[\theta(1-\theta)]$.

Рассмотрим еще один пример вычисления функции $i(\theta)$ для нормальной модели $f^*(\theta, \sigma^2)$. Здесь вклад одного наблюдения

$$U(X_1; \theta) = \frac{\partial \ln f(X_1; \theta)}{\partial \theta} = \frac{X_1 - \theta}{\sigma^2}, \quad \frac{\partial^2 \ln f(X_1; \theta)}{\partial \theta^2} = -\frac{1}{\sigma^2}.$$

Отсюда по формуле (2.19) получаем $i(\theta) = 1/\sigma^2$. Вид функции $i(\theta)$ для некоторых моделей приведен в табл. 2.1.

Таблица 2.1

Модель	$\mathcal{N}(\theta, \sigma^2)$	$\mathcal{N}(\mu, \theta^2)$	$\Gamma(\theta, \lambda)$	$\mathcal{K}(\theta)$	$Bl(k, \theta)$	$\Pi(\theta)$	$\bar{B}l(r, \theta)$
$i(\theta)$	$1/\sigma^2$	$2/\theta^3$	λ/θ^2	$1/2$	$k/\theta(1-\theta)$	$1/\theta$	$r/\theta(1-\theta)^2$

Приведем пример нерегулярной модели. Пусть $\mathcal{L}(\xi) \in R(0, \theta)$.

Здесь из тождества $\int_0^\theta \frac{1}{\theta} dx = 1$ не следует, что $\int_0^\theta \frac{\partial}{\partial \theta} \left(\frac{1}{\theta} \right) dx = 0$,

так как при дифференцировании интеграла по верхнему пределу появляется еще одно слагаемое. В данном случае причина нерегулярности заключается в том, что выборочное пространство зависит от неизвестного параметра θ .

2. Неравенство Рао — Крамера и эффективные оценки. Рассмотрим задачу оценивания заданной параметрической функции $\tau(\theta)$ в модели $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$. Предположим, что модель $\mathcal{F}$ регулярна, функция $\tau(\theta)$ дифференцируема и пусть, как и выше, $\mathcal{F}_\tau$ — класс всех несмещенных оценок $\tau(\theta)$. Тогда имеет место следующее утверждение.

Теорема 2.4 (неравенство Рао — Крамера). Для любой оценки $T = T(X) \in \mathcal{F}_\tau$ справедливо неравенство

$$D_\theta T \geq \frac{[\tau'(\theta)]^2}{ni(\theta)}. \quad (2.20)$$

Равенство здесь имеет место тогда и только тогда, когда T — линейная функция вклада выборки, т. е.

$$T(X) - \tau(\theta) = a(\theta) U(X; \theta), \quad (2.21)$$

где $a(\theta)$ — некоторая функция от θ .

□ По условию,

$$E_\theta T(X) = \int T(x) L(x; \theta) dx = \tau(\theta), \quad \forall \theta \in \Theta. \quad (2.22)$$

Модель $\mathcal{F}$ регулярна, поэтому, дифференцируя это тождество по θ и учитывая (2.16), получаем

$$\begin{aligned} \tau'(\theta) &= \int T(x) \frac{\partial \ln L(x; \theta)}{\partial \theta} L(x; \theta) dx = E_\theta(T(X) U(X; \theta)) = \\ &= \text{cov}_\theta(T(X), U(X; \theta)). \end{aligned} \quad (2.23)$$

Используя неравенство Коши — Буняковского и определение (2.17), из этого соотношения получаем оценку $[\tau'(\theta)]^2 \leq i_n(\theta) D_\theta T(X)$,

причем неравенство здесь обращается в равенство тогда и только тогда, когда (при каждом θ) $T(X)$ и $U(X; \theta)$ линейно связаны. ■

Неравенство (2.20) называется *неравенством Рао — Крамера*. Оно определяет нижнюю границу дисперсий всех несмещенных оценок заданной параметрической функции $\tau(\theta)$ для регулярных моделей.

Замечание. Если $T = T(X)$ — оценка со смещением $b(\theta)$ и функция $b(\theta)$ дифференцируема, то, используя вместо соотношения (2.22) тождество

$$\int T(x) L(x; \theta) dx = \tau(\theta) + b(\theta),$$

с помощью аналогичных рассуждений можно установить неравенство (также называемое неравенством Рао — Крамера)

$$D_{\theta} T \geq \frac{[\tau'(\theta) + b'(\theta)]^2}{n I(\theta)},$$

обобщающее (2.20).

Если существует оценка $T^* \in \mathcal{T}_{\tau}$, для которой нижняя граница Рао — Крамера достигается, то ее называют *эффективной*. Эффективная оценка является оптимальной, и, согласно теореме 2.1, она единственна. Из теоремы 2.4 следует, что *критерием эффективности оценки является представление (2.21)*. Будем называть этот критерий оптимальности оценки *критерием Рао — Крамера*.

Если для оценки T^* выполнено соотношение (2.21), то из формулы (2.23) следует

$$\tau'(\theta) = \frac{1}{a(\theta)} D_{\theta} T^*,$$

т. е. для дисперсии эффективной оценки справедлива формула

$$D_{\theta} T^* = a(\theta) \tau'(\theta). \quad (2.24)$$

Отметим следующее обстоятельство. Вклад выборки $U(X; \theta)$ однозначно определяется моделью, поэтому представление (2.21) (когда оно имеет место) единственно и, следовательно, эффективная оценка может существовать только для одной определенной параметрической функции $\tau(\theta)$ и не существует ни для какой другой функции параметра θ , отличной от $a\tau(\theta) + b$, где a и b — константы.

3. Экспоненциальная модель. Введем важный класс параметрических моделей, называемых *экспоненциальными*, для которых всегда существует эффективная оценка некоторой параметрической функции. По определению, модель $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$ — *экспоненциальная*, если соответствующая функция $f(x; \theta)$ имеет вид

$$f(x; \theta) = \exp \{A(\theta) B(x) + C(\theta) + D(x)\}. \quad (2.25)$$

Многие часто встречающиеся в приложениях модели удовлетворяют этому условию. В частности, это модели $N(\theta, \sigma^2)$, $N(\mu, \theta^2)$, $\Gamma(\theta, \lambda)$, $Bi(k, \theta)$, $\Pi(\theta)$, $Bi(r, \theta)$, приведенные в табл. 2.1. Модель

Коши $\mathcal{K}(\theta)$, очевидно, не экспоненциальная, а равномерная-1 модель $R(0, \theta)$ хотя и удовлетворяет условию (2.25), но не является регулярной.

Вклад выборки для экспоненциальной модели равен

$$U(X; \theta) = A'(\theta) \sum_{i=1}^n B(X_i) + nC'(\theta) = nA'(\theta) \left[\frac{1}{n} \sum_{i=1}^n B(X_i) + \frac{C'(\theta)}{A'(\theta)} \right].$$

Это равенство можно записать в виде (2.21), если положить

$$T^* = T^*(X) = \frac{1}{n} \sum_{i=1}^n B(X_i), \quad \tau(\theta) = -\frac{C'(\theta)}{A'(\theta)}, \quad a(\theta) = \frac{1}{nA'(\theta)}. \quad (2.26)$$

Таким образом, для регулярной экспоненциальной модели существует эффективная (а значит, и оптимальная) оценка T^* для параметрической функции $\tau(\theta)$, где T^* и $\tau(\theta)$ определены в (2.26). При этом согласно формулам (2.24) и (2.26)

$$D_0 T^* = \frac{\tau'(\theta)}{nA'(\theta)}. \quad (2.27)$$

Верно и обратное утверждение: если эффективная оценка для некоторой функции $\tau(\theta)$ существует, то модель является экспоненциальной. Действительно, интегрируя по θ равенство (2.21), получаем

$$\ln L(X; \theta) = A_1(\theta) T(X) + C_1(\theta) + D_1(X).$$

Отсюда следует, что функция $f(x; \theta)$ должна иметь вид (2.25). Таким образом, эффективные оценки существуют для некоторых функций $\tau(\theta)$ только в случае экспоненциальных моделей.

Для перечисленных шести стандартных регулярных экспоненциальных моделей вид функции $\tau(\theta)$, оптимальной оценки τ^* для нее и соответствующей дисперсии (2.27) приведены в табл. 2.2.

Таблица 2.2

Модель	$\tau(\theta)$	τ^*	$D_0 \tau^*$
$N(\theta, \sigma^2)$	θ	$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	σ^2/n
$N(\mu, \theta^2)$	θ^2	$\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$	$2\theta^4/n$
$\Gamma(\theta, \lambda)$	θ	$\bar{X}/\lambda$	$\theta^2/\lambda n$
$Bi(k, \theta)$	θ	$\bar{X}/k$	$\theta(1-\theta)/kn$
$\Pi(\theta)$	θ	$\bar{X}$	θ/n
$Bi(r, \theta)$	$\theta/(1-\theta)$	$\bar{X}/r$	$\theta/[rn(1-\theta)^2]$

Отметим также, что для (регулярных) экспоненциальных моделей функцию информации $i(\theta)$ можно вычислять по формуле $i(\theta) = \tau'(\theta) A'(\theta)$.

Действительно, из соотношений (2.20) и (2.27) имеем равенство $[\tau'(\theta)]^2 / (n i(\theta)) = \tau'(\theta) / (n A'(\theta))$, откуда и следует предыдущее утверждение.

4. Критерий Бхаттачария оптимальности оценки. Неравенство Рао — Крамера (2.20) дает неулучшаемую границу дисперсии несмещенных оценок (для регулярных моделей) в тех случаях, когда существует эффективная оценка. Однако если этой оценки не существует [представление (2.21) не имеет места ни для какой статистики $T(X)$], то можно найти лучшую (т. е. большую) нижнюю границу для дисперсии несмещенных оценок. Основное условие достижения нижней границы дисперсии в неравенстве (2.20) состоит в существовании оценки T , для которой $T - \tau(\theta)$ — линейная (для каждого θ) функция от $\frac{\partial \ln L}{\partial \theta} = \frac{1}{L} \frac{\partial L}{\partial \theta}$. (Здесь и далее иногда будем для краткости писать просто L вместо $L(X; \theta)$.) Предположим, что такой оценки не существует, но зато имеет место оценка T^* , для которой разность $T^* - \tau(\theta)$ является линейной функцией от

$$\frac{1}{L} \frac{\partial L}{\partial \theta}, \frac{1}{L} \frac{\partial^2 L}{\partial \theta^2}, \dots, \frac{1}{L} \frac{\partial^r L}{\partial \theta^r}$$

при некотором $r \geq 2$. Тогда T^* — оптимальная оценка для $\tau(\theta)$. Этот критерий оптимальности, обобщающий критерий Рао — Крамера (2.21), называется критерием Бхаттачария и также основан на принципе достижения нижней границы дисперсии в соответствующем неравенстве. Приведем точную формулировку соответствующего утверждения. Пусть $L^{(k)} = \frac{\partial^k L}{\partial \theta^k}$, $\tau^{(k)} = \frac{d^k \tau(\theta)}{d\theta^k}$; тогда имеет место следующая теорема.

Теорема 2.5 (неравенство Бхаттачария). Пусть $T = T(X)$ — несмещенная оценка для $\tau = \tau(\theta)$. Тогда при всех $\theta \in \Theta$

$$D_\theta T \geq \sum_{i,j=1}^r c_{ij} a_i a_j, \quad (2.28)$$

где $c_{ij} = c_{ij}(\theta) = E_\theta \left(\frac{L^{(i)}}{L} \frac{L^{(j)}}{L} \right)$, $i, j = 1, \dots, r$, а коэффициенты $a_i = a_i(\theta)$ определяются системой уравнений

$$\sum_{j=1}^r c_{ij} a_j = \tau^{(i)}, \quad i = 1, \dots, r.$$

Если матрица $C = \|c_{ij}\|$ не вырождена и $C^{-1} = \|c^{ij}\|$, то неравенство (2.28) эквивалентно неравенству

$$D_\theta T \geq \sum_{i,j=1}^r c^{ij} \tau^{(i)} \tau^{(j)}. \quad (2.29)$$

Знак равенства в неравенствах (2.28) и (2.29) имеет место тогда и только тогда, когда

$$T - \tau = \sum_{i=1}^r a_i L^{(i)} / L. \quad (2.30)$$

Практически критерий Бхаттачария (2.30) используют так: учитывая старшие производные функции правдоподобия, подбирают такую их линейную комбинацию, чтобы получить представление вида (2.30); при этом последовательно полагают $r=2, 3, \dots$. Если при некотором значении r это удастся сделать, то получающаяся при этом пара $\tau(\theta)$ и $T^*(X)$ и есть соответственно параметрическая функция и оптимальная оценка для нее (случай $r=1$ в этой методике соответствует критерию Рао — Крамера).

Пример 2.5 (нормальная-1 модель, оценивание параметрической функции). Рассмотрим модель $\mathcal{N}(\theta, \sigma^2)$. Здесь

$$L(x; \theta) = \frac{1}{(\sqrt{2\pi} \sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right\},$$

и можно проверить, что

$$\frac{1}{L} \left[\frac{2\theta\sigma^2}{n} \frac{\partial L}{\partial \theta} + \frac{\sigma^4}{n^2} \frac{\partial^2 L}{\partial \theta^2} \right] = X^2 - \frac{\sigma^2}{n} - \theta^2.$$

По критерию Бхаттачария отсюда следует, что статистика $T^* = X^2 - \sigma^2/n$ — оптимальная (т. е. несмещенная с минимальной дисперсией) оценка для $\tau(\theta) = \theta^2$.

5. Критерии оптимальности в случае векторного параметра. Полученные выше критерии Рао — Крамера и Бхаттачария оптимальности оценок относятся к случаю скалярного параметра θ .

Обобщим их на случай векторного параметра. Будем предполагать, что $\theta = (\theta_1, \dots, \theta_r)$, т. е. θ — вектор некоторой размерности r . Предположим, что ищется оценка для заданной числовой функции $\tau(\theta)$, дифференцируемой по всем переменным. Исходную модель по-прежнему предполагаем регулярной. Тогда имеют место следующий многомерный аналог неравенства Рао — Крамера и соответствующий критерий оптимальности.

Теорема 2.6. Если $T = T(X)$ — произвольная несмещенная оценка функции $\tau = \tau(\theta)$, то при всех $\theta \in \Theta$

$$D_1 T \geq \sum_{i,j=1}^r g_{ij} c_i c_j, \quad (2.31)$$

где

$$g_{ij} = g_{ij}(\theta) = E_1 \left(\frac{\partial \ln L}{\partial \theta_i} \frac{\partial \ln L}{\partial \theta_j} \right), \quad i, j = 1, \dots, r,$$

а коэффициенты $c_i = c_i(\theta)$ определяются системой уравнений

$$\sum_{i=1}^r g_{ij} c_j = \tau'_i = \frac{\partial \tau}{\partial \theta_i}, \quad i = 1, \dots, r,$$

Если матрица $I_n = I_n(\theta) = \|g_{ij}\|_n$ не вырождена и $I_n^{-1} = \|g^{ij}\|_n$, то неравенство (2.31) эквивалентно неравенству

$$D_1 T \geq \sum_{i,j=1}^r g^{ij} \tau'_i \tau'_j. \quad (2.32)$$

Равенство в (2.31) и (2.32) имеет место тогда и только тогда, когда

$$T - \tau = \sum_{i=1}^r c_i \frac{\partial \ln L}{\partial \theta_i}. \quad (2.33)$$

Если функция правдоподобия $L(x; \theta)$ дважды дифференцируема по θ , то, продифференцировав тождество $\int L(x; \theta) dx = 1$ сначала по θ_i , а потом по θ_j , получим

$$E g_{ij} = -E \left(\frac{\partial^2 \ln L(X; \theta)}{\partial \theta_i \partial \theta_j} \right) = -n E \left(\frac{\partial^2 \ln f(X_1; \theta)}{\partial \theta_i \partial \theta_j} \right). \quad (2.34)$$

Матрица $I_n(\theta)$ — аналог функции информации $i_n(\theta)$ для случая векторного параметра и называется *информационной матрицей выборки*, а $I(\theta) = I_1(\theta)$ — *информационной матрицей одного наблюдения*. Соотношение (2.34) показывает, что для схемы повторных независимых наблюдений $I_n(\theta) = nI(\theta)$, т. е. здесь имеется аналогия со случаем скалярного параметра.

По аналогии со случаем одномерного параметра назовем несмещенную оценку $T^* = T^*(X)$ для $\tau(\theta)$ *эффективной*, если ее дисперсия совпадает при всех θ с правой частью неравенства (2.31) [или (2.32)]. Из теоремы 2.6 следует, что критерием эффективности (а значит, и оптимальности) оценки T является представление (2.33). Приведем пример использования этого критерия.

Пример 2.6 (общая нормальная модель, оценивание среднего). Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{N}^*(\theta_1, \theta_2)$. Требуется оценить параметрическую функцию $\tau(\theta) = \tau(\theta_1, \theta_2) = \theta_1$. Речь идет об оценивании теоретического среднего, поэтому в качестве несмещенной оценки можно взять, например, выборочное среднее $\bar{X}$. Покажем, что в данном случае $\bar{X}$ — оптимальная оценка θ_1 . Действительно,

$$L(x; \theta) = \frac{1}{(V^{2\pi} \theta_2)^n} \exp \left\{ -\frac{1}{2\theta_2^2} \sum_{i=1}^n (x_i - \theta_1)^2 \right\}$$

— функция правдоподобия для модели $\mathcal{N}^*(\theta_1, \theta_2)$ и непосредственно можно проверить, что

$$\frac{\theta_2}{n} \frac{\partial \ln L(X; \theta)}{\partial \theta_1} = \bar{X} - \theta_1.$$

По критерию (2.33) отсюда следует, что статистика $T^*(X) = \bar{X}$ является оптимальной оценкой θ_1 .

Было показано (см. табл. 2.2), что выборочное среднее $\bar{X}$ — оптимальная оценка теоретического среднего μ для модели $\mathcal{N}(\mu, \sigma^2)$. Таким образом, статистика $\bar{X}$ наилучшим образом оценивает неизвестное среднее нормальной модели независимо от того, известна дисперсия или нет.

Как и в одномерном случае, эффективные оценки не всегда существуют. Рассмотрим, например, задачу оценивания неизвестной дисперсии нормальной модели. Если среднее модели известно, то, как показано в п. 3 (см. табл. 2.2), выборочная дисперсия

$\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ является оптимальной и одновременно эффективной оценкой теоретической дисперсии. Для общей нормальной модели $\mathcal{N}(\theta_1, \theta_2^2)$ это уже не имеет места.

Во-первых, как уже было показано в § 2.1 (пример 2.3), выборочная дисперсия S^2 не является несмещенной оценкой функции $\tau(\theta) = \theta_2^2$, а таковой при $n \geq 2$ является статистика S'^2 , определенная в (2.4).

Во-вторых, оценка S'^2 не эффективная, что можно установить, сравнив ее дисперсию, определенную в (2.5), с нижней границей в неравенстве (2.32). Действительно, по формуле (2.34) в данном случае

$$g_{11} = n E_1(1/\theta_2^2) = n/\theta_2^2,$$

$$g_{12} = g_{21} = (2n/\theta_2^2) E_1(X_1 - \theta_1) = 0,$$

$$g_{22} = (3n/\theta_2^2) E_1(X_1 - \theta_1)^2 - n/\theta_2^2 = 2n/\theta_2^2,$$

т. е. для модели $\mathcal{N}(\theta_1, \theta_2^2)$ информационная матрица $I(\theta) = \begin{bmatrix} 1/\theta_2^2 & 0 \\ 0 & 2/\theta_2^3 \end{bmatrix}$. Отсюда $\det I_n(\theta) = g_{11}g_{22} \neq 0$, $g^{22} = g_{22}^{-1} = \theta_2^3/(2n)$ и правая часть в неравенстве (2.32) равна $g^{22}(\tau_1')^2 = 2\theta_2^3/n$, что меньше $2\theta_2^3/(n-1) = D_1(S'^2)$. Это еще не означает, что существует более точная, чем S'^2 , оценка для θ_2^2 . Покажем, что в данной задаче вообще не существует эффективной оценки (нижняя граница в неравенствах (2.31) — (2.32) не достигается ни для какой несмещенной оценки θ_2^2), но оптимальная оценка T^* существует, и это S'^2 .

Воспользуемся аналогом критерия Бхаттачария для векторного параметра, который имеет следующий вид: если при некотором целом $s \geq 2$ и коэффициентах $c_i = c_i(\theta)$ имеет место представление

$$T - \tau = \frac{1}{L} \left[\sum_i c_i \frac{\partial L}{\partial \theta_i} + \sum_{i,j} c_{ij} \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} + \dots + \sum_{i_1, \dots, i_s} c_{i_1 \dots i_s} \frac{\partial^s L}{\partial \theta_{i_1} \dots \partial \theta_{i_s}} \right], \quad (2.35)$$

то статистика $T = T(X)$ есть оптимальная оценка функции $\tau = \tau(\theta)$.

Применим этот критерий к рассматриваемому случаю. Имеем:

$$\begin{aligned}\frac{1}{L} \frac{\partial L}{\partial \theta_2} &= \frac{\partial \ln L}{\partial \theta_2} = \frac{1}{\theta_2^2} \sum_{i=1}^n (X_i - \theta_1)^2 - \frac{n}{\theta_2^3} = \\ &= \frac{1}{\theta_2^3} \sum_{i=1}^n (X_i - \bar{X})^2 + \frac{n}{\theta_2^3} (\bar{X} - \theta_1)^2 - \frac{n}{\theta_2^3}, \\ \frac{1}{L} \frac{\partial^2 L}{\partial \theta_1^2} &= \frac{\partial^2 \ln L}{\partial \theta_1^2} + \left(\frac{\partial \ln L}{\partial \theta_1} \right)^2 = \frac{n^2}{\theta_1^4} (\bar{X} - \theta_1)^2 - \frac{n}{\theta_1^4}.\end{aligned}$$

Отсюда

$$\frac{1}{L} \left(\frac{\theta_2^2}{n-1} \frac{\partial L}{\partial \theta_2} - \frac{\theta_1^4}{n(n-1)} \frac{\partial^2 L}{\partial \theta_1^2} \right) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 - \theta_1^2 = S'^2 - \theta_1^2.$$

На основании критерия (2.35) это означает, что статистика $T^* = S'^2$ является оптимальной оценкой для θ_2^2 в модели $\mathcal{N}(\theta_1, \theta_2^2)$.

Итак, в задаче оценивания параметрической функции $\tau(\theta) = \theta_2^2$ для модели $\mathcal{N}(\theta_1, \theta_2^2)$ имеет место более точная, чем (2.32), оценка для дисперсии несмещенных оценок $\tau(\theta)$ — это оценка $D_0 T \geq 2\theta_1^4/(n-1)$; при этом равенство достигается для статистики $T^* = S'^2$.

§ 2.3. Принцип достаточности и оптимальные оценки

Рассмотренные в предыдущем параграфе критерии оптимальности оценок имеют сравнительно ограниченную применимость. Во-первых, они требуют жестких условий регулярности исходной модели и, во-вторых, в лучшем случае позволяют находить оптимальные оценки для отдельных параметрических функций $\tau(\theta)$. Так, например, было установлено, что оптимальной оценкой для дисперсии θ_2^2 модели $\mathcal{N}(\theta_1, \theta_2^2)$ является статистика S'^2 , но решить вопрос об оптимальном оценивании среднеквадратического отклонения θ_2 с помощью изложенных выше методов нельзя. Более эффективным способом построения оптимальных оценок является использование достаточных статистик, рассмотрению которых и посвящен настоящий параграф.

1. Достаточные статистики. По определению, статистика $T = T(X)$ (вообще говоря, векторная) называется *достаточной* для модели $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$ (или достаточной для параметра θ , когда ясно, о какой модели идет речь), если условная плотность (или вероятность в дискретном случае) $L(x|t; \theta)$ случайного вектора $X = (X_1, \dots, X_n)$ при условии $T(X) = t$ не зависит от параметра θ . Это свойство статистики T означает, что она содержит всю информацию о параметре θ , имеющуюся в выборке, и поэтому все заключения об этом параметре, которые можно сделать при наблюдении x , зависят только от $t = T(x)$. Можно также

сказать, что все статистические выводы о модели, обладающей достаточной статистикой, в итоге формулируются в терминах этой достаточной статистики. Достаточная статистика, следовательно, дает оптимальный в определенном смысле способ представления статистических данных, что особенно важно при обработке больших массивов статистической информации. При этом обычно стремятся найти достаточную статистику минимальной размерности, представляющую данные в наиболее сжатом виде, в этом смысле говорят о *минимальной достаточной статистике* (очевидно, сама выборка X всегда является достаточной статистикой, но эта статистика тривиальная, так как не сокращает данных).

Приведем *критерий факторизации*, позволяющий в каждом конкретном случае определить, существует ли достаточная статистика, и одновременно установить ее вид.

Теорема 2.7 (критерий факторизации). *Для того чтобы статистика $T(X)$ была достаточной для θ , необходимо и достаточно, чтобы функция правдоподобия $L(x; \theta)$ имела вид*

$$L(x; \theta) = g(T(x); \theta) h(x), \quad (2.36)$$

где множитель g может зависеть от θ , а от x зависит лишь через $T(x)$, множитель h же от параметра θ не зависит.

□ Приведем доказательство для дискретной модели. Если статистика T достаточна, то при любом t из области значений $T(x)$ функция $|L(x|t; \theta)$ не зависит от θ и ее можно записать в виде $h(x, t)$. Пусть $T(x) = t$; тогда событие $\{X = x\} \subseteq \{T(X) = t\}$, поэтому

$$\begin{aligned} L(x; \theta) &= P_{\theta}(X = x) = P_{\theta}(X = x, T(X) = t) = \\ &= P_{\theta}(T(X) = t) P_{\theta}(X = x | T(X) = t) = g(t; \theta) L(x|t; \theta) = \\ &= g(T(x); \theta) h(x, t), \end{aligned}$$

т. е. выполняется представление (2.36).

Обратно: пусть имеет место разложение (2.36). Тогда при любом x таком, что $T(x) = t$,

$$\begin{aligned} L(x|t; \theta) &= P_{\theta}(X = x | T(X) = t) = \frac{P_{\theta}(X = x, T(X) = t)}{P_{\theta}(T(X) = t)} = \\ &= L(x; \theta) \bigg| \sum_{x': T(x') = t} L(x'; \theta) = g(t; \theta) h(x) \bigg| \sum_{x': T(x') = t} g(t; \theta) h(x') = \\ &= h(x) \bigg| \sum_{x': T(x') = t} h(x'), \end{aligned}$$

т. е. не зависит от θ . Если же x таково, что $T(x) \neq t$, то, очевидно, $L(x|t; \theta) = 0$. Таким образом, в любом случае (при любом x) условная вероятность $L(x|t; \theta)$ не зависит от θ . ■

Отметим, что всякая эффективная оценка является одновременно достаточной статистикой, так как [см. (2.21)] из представления

$$T(X) - \tau(\theta) = a(\theta) \frac{\partial \ln L(X; \theta)}{\partial \theta}$$

следует форма (2.36) для $L(x; \theta)$. Итак, эффективная оценка существует только тогда, когда имеется достаточная статистика. Но последняя может существовать и при отсутствии эффективной оценки, т. е. условие достаточности является менее ограничительным, чем условие существования эффективной оценки.

Заметим также, что если T достаточна, то таковой же является и любая взаимно однозначная функция от T . Действительно, если $H = \varphi(T)$ и φ — взаимно однозначная функция, то существует обратная функция φ^{-1} : $T = \varphi^{-1}(H)$ и из представления (2.36) имеем

$$L(x; \theta) = g(\varphi^{-1}(H); \theta) h(x) = g_1(H; \theta) h(x),$$

т. е. H — достаточная статистика.

Примерами достаточных статистик для ряда моделей со скалярным параметром являются (с учетом сделанного замечания) эффективные оценки τ^* , приведенные в табл. 2.2. Рассмотрим еще несколько примеров нахождения достаточных статистик для моделей с векторным параметром и для нерегулярных моделей.

Пример 2.7 (общая нормальная модель, достаточная статистика для нее). Функцию правдоподобия для общей нормальной модели $\mathcal{N}(\bar{x}, \theta_1, \theta_2^2)$ можно записать в виде

$$L(x; \theta) = \frac{1}{(\sqrt{2\pi} \theta_2)^n} \exp \left\{ -\frac{1}{2\theta_2^2} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n(\bar{x} - \theta_1)^2}{2\theta_2^2} \right\},$$

откуда в силу критерия факторизации имеем, что статистика

$$T = (T_1, T_2), \quad T_1 = \bar{X}, \quad T_2 = \sum_{i=1}^n (X_i - \bar{X})^2,$$

является достаточной для $\theta = (\theta_1, \theta_2)$.

Пример 2.8 (равномерная модель, достаточная статистика для нее). Пусть модель $\mathcal{F} = R(0, \theta)$. Тогда функция правдоподобия

$$L(x; \theta) = \begin{cases} 1/\theta^n & \text{при } x_{(n)} = \max_{1 \leq i \leq n} x_i \leq \theta, \\ 0 & \text{в противном случае,} \end{cases}$$

или $L(x; \theta) = e(\theta - x_{(n)}) \cdot \theta^{-n}$, где $e(x)$ — функция Хевисайда [см. (1.5)].

На основании критерия факторизации отсюда имеем, что $T(X) = X_{(n)}$, т. е. максимальное значение выборки является достаточной статистикой для θ .

Аналогично, для общей равномерной модели $R(\theta_1, \theta_2)$ функцию правдоподобия можно записать в виде

$$L(x; \theta) = e(\theta_2 - x_{(n)}) e(x_{(1)} - \theta_1) / (\theta_2 - \theta_1)^n;$$

следовательно, статистика $T = (T_1, T_2)$, где $T_1 = X_{(1)}$, $T_2 = X_{(n)}$, достаточная для $\theta = (\theta_1, \theta_2)$.

Пример 2.9 (равномерная модель (продолжение)). Рассмотрим модель $R(a(\theta), b(\theta))$, где $a(\theta) < b(\theta)$, $\forall \theta$, — заданные непрерывные функции скалярного параметра θ . Как и в примере 2.8, можно записать

$$L(x; \theta) = e^{(x_{(1)} - a(\theta))} e^{(b(\theta) - x_{(n)})} (b(\theta) - a(\theta))^{-n}. \quad (2.37)$$

Следовательно, двумерная статистика $T = (X_{(1)}, X_{(n)})$ является достаточной для θ . Выясним теперь, нельзя ли найти в данном случае одномерную, т. е. более экономную, достаточную статистику. Числитель в (2.37) равен 1 тогда и только тогда, когда

$$\{x_{(1)} \geq a(\theta), b(\theta) \geq x_{(n)}\}. \quad (2.38)$$

Пусть $a(\theta) \uparrow$, $b(\theta) \downarrow$ с возрастанием θ . Тогда условие (2.38) эквивалентно условию

$$\{\theta \leq a^{-1}(x_{(1)}), \theta \leq b^{-1}(x_{(n)})\} \Leftrightarrow \{\theta \leq T_1(x) = \min(a^{-1}(x_{(1)}), b^{-1}(x_{(n)}))\}.$$

Таким образом, соотношение (2.37) можно записать в виде

$$L(x; \theta) = e^{(T_1(x) - \theta)} (b(\theta) - a(\theta))^{-n},$$

откуда следует, что $T_1 = T_1(X) = \min(a^{-1}(X_{(1)}), b^{-1}(X_{(n)}))$ — одномерная достаточная статистика для θ .

Аналогично, если $a(\theta) \downarrow$, $b(\theta) \uparrow$ с возрастанием θ , то одномерная достаточная статистика имеет вид $T_2 = T_2(X) = \max(a^{-1}(X_{(1)}), b^{-1}(X_{(n)}))$. Эти два случая исчерпывают ситуации, когда в модели $R(a(\theta), b(\theta))$ существует одномерная достаточная статистика. Обратим внимание на то, что как T_1 , так и T_2 являются функциями исходной двумерной достаточной статистики $T = (X_{(1)}, X_{(n)})$. Это обстоятельство имеет общий характер, т. е. минимальная достаточная статистика есть функция любых других достаточных статистик. В частности, для равномерной модели с нулевым средним $R(-\theta, \theta)$ одномерная достаточная статистика T_2 принимает вид

$$T_2 = \max(-X_{(1)}, X_{(n)}) = \max(|X_{(1)}|, |X_{(n)}|),$$

а для модели $R(\theta, \theta + 1)$ одномерной достаточной статистики не существует.

Пример 2.10 (модель Коши, достаточная статистика для нее). Для модели Коши $\mathcal{K}(\theta)$ функция правдоподобия имеет вид

$$L(x; \theta) = \frac{1}{\pi^n} \prod_{i=1}^n \frac{1}{1 + (x_i - \theta)^2}.$$

и здесь нельзя найти статистику T , дающую факторизацию (2.36), размерность которой меньше n .

2. Достаточные статистики и оптимальные оценки. Роль достаточных статистик в теории оценивания раскрывает следующее утверждение.

Теорема 2.8 (Рао — Блекуэлл — Колмогорова). *Оптимальная оценка, если она существует, является функцией от достаточной статистики*.*

□ Пусть $T = T(X)$ — достаточная статистика и $T_1 = T_1(X)$ — произвольная несмещенная оценка заданной параметрической функции $\tau(\theta)$. Рассмотрим функцию

$$H(t) = E_{\theta}(T_1 | T = t) = \int T_1(x) L(x | t; \theta) dx.$$

Эта функция не зависит от θ , так как условная плотность $L(x | t; \theta)$ от параметра не зависит. Далее, статистика $H(T(X))$ является также несмещенной оценкой $\tau(\theta)$. Действительно, если $g(t; \theta)$ — плотность распределения статистики $T(X)$, то

$$\begin{aligned} E_{\theta}H(T(X)) &= \int H(t)g(t; \theta)dt = \int T_1(x) \left[\int L(x | t; \theta)g(t; \theta)dt \right] dx = \\ &= \int T_1(x) L(x; \theta) dx = E_{\theta}T_1(X) = \tau(\theta). \end{aligned}$$

Наконец, справедливо неравенство

$$D_{\theta}H(T(X)) \leq D_{\theta}T_1(X), \quad \forall \theta,$$

причем равенство имеет место в том и только в том случае, когда $T_1 = H(T)$. Для доказательства этого заметим, что

$$\begin{aligned} D_{\theta}T_1 &= E_{\theta}[T_1 - H(T) + H(T) - \tau(\theta)]^2 = \\ &= E_{\theta}[T_1 - H(T)]^2 + D_{\theta}H(T) \geq D_{\theta}H(T), \end{aligned}$$

поскольку математическое ожидание произведения

$$\begin{aligned} E_{\theta}[T_1 - H(T)][H(T) - \tau(\theta)] &= E_{\theta}[(T_1 - H(T))H(T)] = \\ &= \int [E_{\theta}(T_1 | T = t) - H(t)]H(t)g(t; \theta)dt = 0. \end{aligned}$$

В предыдущем неравенстве знак равенства возможен только в случае $T_1 = H(T)$. Таким образом, для любой несмещенной оценки $\tau(\theta)$, не являющейся функцией от достаточной статистики, можно указать несмещенную оценку, которая зависит от достаточной статистики и имеет дисперсию меньшую, чем исходная оценка. Следовательно, *оптимальную оценку надо искать среди функций от достаточной статистики.* ■

При отыскании явного вида оптимальных оценок важную роль играет свойство полноты достаточной статистики. По определению, достаточная статистика $T = T(X)$ называется *полной*, если для всякой функции $\varphi(T(X))$ из того, что

$$E_{\theta}\varphi(T) = 0, \quad \forall \theta,$$

следует $\varphi(t) \equiv 0$ на всем множестве значений статистики T^{**} .

* Частный случай этого утверждения об эффективных оценках был отмечен в п. 1.

** Это утверждение понимается в том смысле, что

$$P_{\theta}(X \in \{x: \varphi(T(x)) \neq 0\}) = 0, \quad \forall \theta.$$

Теорема 2.9. Если существует полная достаточная статистика, то всякая функция от нее является оптимальной оценкой своего математического ожидания.

□ Пусть $T = T(X)$ — полная достаточная статистика и $H(T)$ — произвольная функция от T . Обозначим

$$E_{\theta} H(T) = \tau(\theta). \quad (2.39)$$

Из условия полноты следует, что $H(T)$ — единственная функция от T , удовлетворяющая соотношению (2.39), так как если бы была еще какая-нибудь функция $H_1(T)$, удовлетворяющая (2.39), то мы бы имели, что

$$E_{\theta}[H(T) - H_1(T)] = 0, \quad \forall \theta,$$

откуда $H_1(t) \equiv H(t)$.

По теореме 2.8 оптимальную оценку $\tau(\theta)$ надо искать в классе функций, зависящих от T . Но $H(T)$ — единственная функция от T , несмещенно оценивающая $\tau(\theta)$; следовательно, она и является искомой оптимальной оценкой. ■

Итак, пусть в рассматриваемой модели $\mathcal{F}$ существует полная достаточная статистика T и требуется оценить заданную параметрическую функцию $\tau(\theta)$. Тогда:

1) если существует какая-то несмещенная оценка $\tau(\theta)$, то существует и несмещенная оценка, являющаяся функцией от T ; можно также сказать, что если нет несмещенных оценок вида $H(T)$ (т. е. уравнение (2.39) не имеет решения), то класс $\mathcal{F}_{\tau}$ несмещенных оценок $\tau(\theta)$ пуст;

2) оптимальная оценка (когда она существует) всегда является функцией от T , и она однозначно определяется соотношением (2.39);

3) оптимальную оценку τ^* можно искать по формуле

$$\tau^* = H(T) = E_{\theta}(T_1 | T),$$

исходя из любой несмещенной оценки T_1 функции $\tau(\theta)$.

В конкретных задачах при отыскании оптимальных оценок последний критерий используют редко, так как вычисление условного математического ожидания обычно сопряжено с аналитическими трудностями. Чаше решают непосредственно уравнение (2.39), которое в развернутой форме для абсолютно непрерывной модели имеет вид

$$\int H(t) g(t; \theta) dt = \tau(\theta), \quad \forall \theta \in \Theta, \quad (2.40)$$

где $g(t; \theta)$ — плотность распределения достаточной статистики T . Для дискретной модели интегрирование в левой части уравнения (2.40) заменяется суммированием. Это уравнение в дальнейшем будем называть *уравнением несмещенности*. Его можно решать различными способами, например разлагая правую и левую части в ряды по степеням θ (в случае аналитических функций $\tau(\theta)$ и $g(t; \theta)$) и приравнивая соответствующие коэффициенты.

3. Примеры применения достаточных статистик. Рассмотрим применение изложенной теории к построению оптимальных оценок для различных параметрических функций $\tau(\theta)$.

Пример 2.11 (бернуллиевская модель, оценивание параметрических функций). Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{X}(\frac{n}{2}) \in Bi(1, \theta)$. Как отмечалось выше, статистика $T = \sum_{i=1}^n X_i$

(число «успехов» в n испытаниях Бернулли), будучи эффективной оценкой для $n\theta$, является достаточной. Покажем, что она обладает свойством полноты. Для этого прежде всего заметим, что $\mathcal{L}_\theta(T) = Bi(n, \theta)$ и, следовательно, распределение T имеет вид $g(t; \theta) = C_n^t \theta^t (1-\theta)^{n-t}$, $t = 0, 1, \dots, n$. Пусть $\varphi(t)$ — произвольная функция, заданная на множестве $\{0, 1, \dots, n\}$. Тогда условие $E_\theta \varphi(T) = 0$, $\forall \theta \in (0, 1)$, записывается в виде

$$\sum_{t=0}^n \varphi(t) C_n^t \theta^t (1-\theta)^{n-t} = 0, \quad \forall \theta \in (0, 1), \text{ или}$$

$$\sum_{t=0}^n \varphi(t) C_n^t x^t = 0, \quad \forall x > 0 \quad \left(x = \frac{\theta}{1-\theta}\right).$$

Отсюда следует, что все коэффициенты данного многочлена равны нулю, т. е. $\varphi(t) = 0$, $t = 0, 1, \dots, n$. Следовательно, T — полная достаточная статистика. Тем самым всякая функция от T является оптимальной оценкой своего среднего. Так как производящая функция случайной величины T равна

$$E_{\theta} z^T = \varphi(z; \theta) = [1 + (z-1)\theta]^n,$$

то, полагая $(a)_k = a(a-1)\dots(a-k+1)$, $k \geq 1$, имеем

$$E_{\theta}(T)_k = \frac{\partial^k \varphi(z; \theta)}{\partial z^k} \Big|_{z=1} = (n)_k \theta^k.$$

Отсюда и из предыдущего следует, что для любого целого k , $1 \leq k \leq n$, оптимальной оценкой степени θ^k является статистика $(T)_k / (n)_k$ (выше (см. теорему 2.2) соответствующий результат имел место только для $k=1$). Одновременно также получен вывод: более высокие степени θ^k , $k > n$, по выборке объема n не поддаются оценке в классе несмещенных оценок. Наконец, воспользовавшись теоремой 2.3, сразу получаем, что если $\tau(\theta)$ — много-

член степени $k \leq n$: $\tau(\theta) = \sum_{j=0}^k a_j \theta^j$, то оптимальной оценкой для него является статистика

$$\tau^* = \sum_{j=0}^k a_j (T)_j / (n)_j.$$

Например, теоретическую дисперсию $\tau(\theta) = \theta(1-\theta)$ распределения $Bi(1; \theta)$ оптимально оценивает статистика

$$\tau^* = T(n-T) / [n(n-1)].$$

Пример 2.12 (модель степенного ряда, оценивание параметрических функций). Рассмотрим семейство дискретных распределений типа степенного ряда, для которых вероятности

$$f(x; \theta) = a(x) \theta^x / f(\theta), \quad x = l, l+1, \dots, \quad f(\theta) = \sum_{x=l}^{\infty} a(x) \theta^x, \quad (2.41)$$

при этом последний ряд имеет ненулевой радиус сходимости R , в этом случае полагаем $\Theta = (0, R)$.

Распределения типа (2.41) охватывают многие хорошо известные дискретные распределения с бесконечным множеством значений; в частности, это следующие распределения: Пуассона $\Pi(\theta)$ ($f(\theta) = e^\theta$, $R = \infty$), отрицательное биномиальное $Bi(r, \theta)$ ($f(\theta) = (1 - \theta)^{-r}$, $R = 1$), логарифмическое ($f(x; \theta) = \theta^x / (x \ln \frac{1}{1-\theta})$, $x = 1, 2, \dots$, $R = 1$) и т. д., а также соответствующие усеченные слева распределения.

Примечание. Усеченным называют распределение, у которого некоторые значения запрещены, например усеченным в нуле распределением Пуассона является распределение $f(x; \theta) = c e^{-\theta} \frac{\theta^x}{x!}$, $x = 1, 2, \dots$ (константу c находят из условия нормировки $\sum_{x=1}^{\infty} f(x; \theta) = 1$, и в данном случае она равна $(1 - e^{-\theta})^{-1}$).

Отметим, что рассматриваемые распределения являются распределениями экспоненциального типа (см. п. 3 § 2.2), поэтому из соответствующих результатов для экспоненциальных моделей следует, что если имеется выборка $X = (X_1, \dots, X_n)$ из распределения (2.41), то статистика X является эффективной (значит, и оптимальной) оценкой функции $\tau(\theta) = \theta f'(\theta) / f(\theta)$. Также известно, что это единственная (параметрическая функция, для которой существует эффективная оценка. Выясним, используя теорию достаточных статистик, для каких еще параметрических функций в рассматриваемой модели существуют оптимальные оценки и как определить вид этих оценок.

Рассмотрим статистику $T = T(X) = X_1 + \dots + X_n$. В данном случае функция правдоподобия

$$L(x; \theta) = \theta^{T(x)} \prod_{i=1}^n a(x_i) / f^n(\theta),$$

откуда на основании критерия факторизации следует, что T — достаточная статистика. Ее распределение $g(t; \theta)$ имеет вид

$$g(t; \theta) = P_n(T(X) = t) = \sum_{x: T(x) = t} L(x; \theta) = \theta^t b_n(t) / f^n(\theta),$$

$$t = nl, nl+1, \dots, \quad (2.42)$$

$$\text{где } b_n(t) = \sum_{x_1 + \dots + x_n = t} a(x_1) \dots a(x_n) = \text{coef}_t f^n(z).$$

Таким образом, T также имеет распределение типа степенного ряда. Пусть, далее, $\varphi(t)$ — произвольная функция, заданная на множестве $\{nl, nl+1, \dots\}$, и такая, что $E_\theta \varphi(T) = 0$, $\forall \theta \in \Theta$, т. е.

$$\sum_{t=nl}^{\infty} \varphi(t) b_n(t) \theta^t = 0, \quad \forall \theta \in \Theta.$$

Отсюда следует, что $\varphi(t) = 0$ для всех t , для которых $b_n(t) \neq 0$, т. е. $\varphi(t) = 0$ на множестве всех возможных значений T . Таким образом, T — полная достаточная статистика.

Пусть теперь требуется оценить параметрическую функцию $\tau(\theta)$, представимую в виде сходящегося на Θ степенного ряда:

$\tau(\theta) = \sum_{j=r}^{\infty} a_j \theta^j$. Из формулы (2.42) следует, что в данном случае уравнение несмещенности (2.40) можно записать в виде

$$\begin{aligned} \sum_{t=nl}^{\infty} H(t) b_n(t) \theta^t &= \tau^n(\theta) \tau(\theta) = \sum_{t=nl}^{\infty} b_n(t) \theta^t \sum_{j=r}^{\infty} a_j \theta^j = \\ &= \sum_{k=nl+r}^{\infty} \theta^k \sum_{j=r}^{k-nl} a_j b_n(k-j). \end{aligned}$$

Поскольку это тождество по θ , приравнявая соответствующие коэффициенты, находим

$$H(t) b_n(t) = \begin{cases} \sum_{j=r}^{t-nl} a_j b_n(t-j) & \text{при } t \geq nl+r, \\ 0 & \text{при } t < nl+r. \end{cases}$$

Отсюда получаем, что оптимальная оценка τ^* для функции $\tau(\theta)$ имеет вид

$$\tau^* = H(T) = \begin{cases} b_n^{-1}(T) \sum_{j=r}^{T-nl} a_j b_n(T-j) & \text{при } T \geq nl+r, \\ 0 & \text{при } T < nl+r. \end{cases} \quad (2.43)$$

В частности, если $\tau(\theta) = \theta^r$ при некотором $r \geq 1$, то

$$\tau^* = \begin{cases} b_n(T-r)/b_n(T) & \text{при } T \geq nl+r, \\ 0 & \text{при } T < nl+r. \end{cases} \quad (2.44)$$

Таким образом, для важного класса дискретных распределений типа (2.41) можно строить оптимальные оценки для произвольных параметрических функций, представимых в виде степенного ряда от θ .

Пример 2.13 (пуассоновская модель, оценивание параметрических функций). Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{L}(\xi) \in \Pi(\theta)$. Требуется оценить вероятности появления отдель-

ных значений ξ , т. е. функции $\pi_k(\theta) = e^{-\theta k/k!}$, $k=0, 1, \dots$. Здесь

$$\pi_k(\theta) = \frac{1}{k!} \sum_{j=k}^{\infty} (-1)^{j-k} \frac{\theta^j}{(j-k)!}, \quad \text{т. е.}$$

$$a_j = \frac{(-1)^{j-k}}{k! (j-k)!}, \quad j=k, k+1, \dots$$

Далее, $f^n(\theta) = e^{-n\theta}$, поэтому $b_n(t) = n^t/t!$, $t=0, 1, 2, \dots$. Подставив эти значения в формулу (2.43), находим, что оптимальная оценка π_k^* имеет вид

$$\pi_k^* = \frac{T!}{n^T} \sum_{j=k}^T (-1)^{j-k} \frac{n^{T-j}}{k! (j-k)! (T-j)!} = C_T^k \frac{1}{n^k} \left(1 - \frac{1}{n}\right)^{T-k}.$$

Отсюда следует, что при заданном значении T отличными от нуля являются значения оценок π_k^* только для $k=0, 1, \dots, T$.

По теореме 2.3, оптимальной оценкой некоторой суммы вероятностей $\pi_k(\theta)$ является сумма соответствующих оценок π_k^* . В частности, статистика

$$\tau^* = \begin{cases} \sum_{k=r}^T C_T^k \frac{1}{n^k} \left(1 - \frac{1}{n}\right)^{T-k} & \text{при } T \geq r, \\ 0 & \text{при } T < r \end{cases}$$

— оптимальная оценка для $\tau(\theta) = P_\theta(\xi \geq r) = \sum_{k=r}^{\infty} \pi_k(\theta)$.

Пример 2.14 (отрицательная биномиальная модель, оценивание параметра). По выборке $X = (X_1, \dots, X_n)$ требуется оценить параметр θ модели $Bi(r, \theta)$. Здесь

$$f^n(\theta) = (1 - \theta)^{-nr} = \sum_{t=0}^{\infty} C_{nr+t-1}^t 1 - \theta^t,$$

т. е. $b_n(t) = C_{nr+t-1}^t$, $t=0, 1, 2, \dots$, поэтому по формуле (2.44) имеем, что оптимальной оценкой $\tau(\theta) = \theta$ является статистика

$$\tau^* = \begin{cases} T/(T + nr - 1) & \text{при } T \geq 1, \\ 0 & \text{при } T = 0. \end{cases}$$

Пример 2.15 (равномерная-1 модель, оценивание параметра). По выборке $X = (X_1, \dots, X_n)$ требуется оценить параметр θ равномерного распределения $R(0, \theta)$. В примере 2.8 было показано, что $T = X_{(n)}$ — достаточная статистика для θ . Убедимся в ее полноте. По формуле (1.21) плотность распределения величины T равна

$$g(t; \theta) = nt^{n-1}/\theta^n, \quad 0 \leq t \leq \theta.$$

Пусть теперь для функции $\varphi(t)$, $t \geq 0$, выполняется условие

$$E_{\theta} \varphi(T) = \frac{n}{\theta^n} \int_0^{\theta} \varphi(t) t^{n-1} dt = 0, \quad \forall \theta > 0.$$

Дифференцируя по θ тождество $\int_0^{\theta} \varphi(t) t^{n-1} dt = 0$, получаем $\varphi(\theta) \theta^{n-1} = 0$,

т. е. $\varphi(\theta) = 0$, $\forall \theta > 0$. Тем самым доказано, что статистика T — полная. Далее имеем

$$E_{\theta} T = \frac{n}{\theta^n} \int_0^{\theta} t^n dt = \frac{n}{n+1} \theta.$$

По теореме 2.9 отсюда следует, что $\frac{n+1}{n} T = \frac{n+1}{n} X_{(n)}$ — оптимальная оценка θ . Нетрудно показать, что

$$D_{\theta} \left(\frac{n+1}{n} X_{(n)} \right) = \frac{\theta^2}{n(n+2)}.$$

Пример 2.16 (достаточная статистика, не являющаяся полной). Пусть произведено одно наблюдение X над дискретной случайной величиной с распределением

$$f(x; \theta) = \begin{cases} \theta & \text{при } x = -1, \\ \theta^x (1 - \theta)^2 & \text{при } x = 0, 1, 2, \dots, \end{cases} \quad \theta \in (0, 1).$$

Тогда X — достаточная, но не полная статистика. Действительно, пусть $\varphi(x)$ — заданная на множестве $\{-1, 0, 1, 2, \dots\}$ функция, удовлетворяющая условию $E_{\theta} \varphi(X) = 0$, $\forall \theta \in (0, 1)$. Это условие можно записать в виде

$$\varphi(-1) \frac{\theta}{(1-\theta)^2} + \sum_{x=0}^{\infty} \varphi(x) \theta^x = 0$$

или (см. пример 2.14)

$$\varphi(0) + \sum_{x=1}^{\infty} [\varphi(x) + x\varphi(-1)] \theta^x = 0.$$

Этому условию удовлетворяет любая функция, для которой $\varphi(0) = 0$, $\varphi(x) = -x\varphi(-1)$, $x = 1, 2, \dots$, т. е. значение $\varphi(-1)$ можно задать произвольно. Следовательно, в данном случае критерий полноты не выполняется, поэтому нельзя ожидать, что уравнение несмещенности будет иметь однозначное решение. Рассмотрим, например, задачу оценивания параметрической функции $\tau(\theta) = (1 - \theta)^2$. В этом случае непосредственно можно проверить, что любая статистика $T = H(X)$, где функция H имеет вид

$$H(x) = \begin{cases} 1 & \text{при } x = 0, \\ ax & \text{при } x = -1, 1, 2, \dots, \end{cases} \quad a \text{ произвольно, является несмещенной оценкой } \tau(\theta).$$

§ 2.4. Оценки максимального правдоподобия

1. Определение и примеры оценок максимального правдоподобия. Одним из наиболее универсальных методов оценивания параметров распределений является метод максимального правдоподобия. Оценку параметра θ , получаемую с помощью этого метода, будем обозначать $\hat{\theta} = \hat{\theta}(X)$, а оценку параметрической функции $\tau(\theta)$ — записывать в виде $\hat{\tau} = \hat{\tau}(X)$.

Пусть, как обычно, задана выборка $X = (X_1, \dots, X_n)$ из распределения $\mathcal{L}(\xi) \in \mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$ и $L(x; \theta)$ — функция правдоподобия для реализации $x = (x_1, \dots, x_n)$ выборки X . По определению, оценкой максимального правдоподобия (о. м. п.) $\hat{\theta}$ параметра θ (точнее, значением о. м. п. при заданной реализации x выборки X) называется такая точка параметрического множества Θ , в которой функция правдоподобия $L(x; \theta)$ при заданном x достигает максимума. Таким образом,

$$L(x; \hat{\theta}) \geq L(x; \theta), \forall \theta, \text{ или } L(x; \hat{\theta}) = \sup_{\theta \in \Theta} L(x; \theta).$$

Если для каждого x из выборочного пространства $\mathcal{X}^*$ максимум $L(x; \theta)$ достигается во внутренней точке Θ и $L(x; \theta)$ дифференцируема по θ , то о. м. п. $\hat{\theta}$ удовлетворяет уравнению $\frac{\partial L(x; \hat{\theta})}{\partial \theta} = 0$ или $\frac{\partial \ln L(x; \hat{\theta})}{\partial \theta} = 0$. Если θ — векторный параметр: $\theta = (\theta_1, \dots, \theta_r)$, то это уравнение заменяется системой уравнений $\frac{\partial \ln L(x; \hat{\theta})}{\partial \theta_i} = 0$, $i = 1, \dots, r$. Последние уравнения называются *уравнениями правдоподобия*.

Отметим следующие свойства оценок максимального правдоподобия.

1) Если существует эффективная оценка $T(X)$ для скалярного параметра θ , то $\hat{\theta} = T(X)$. Это очевидное следствие критерия эффективности Рао — Крамера (см. теорему 2.4):

$$\frac{\partial \ln L(x; \hat{\theta})}{\partial \theta} = \frac{1}{a(\theta)} [T(x) - \theta].$$

2) Если имеется достаточная статистика $T = T(X)$, а оценка максимального правдоподобия $\hat{\theta}$ существует и единственна, то она является функцией от T . Действительно, из представления (2.36) следует, что в данном случае максимизация $L(x; \theta)$ сводится к максимизации $g(T(x); \theta)$ по θ . Следовательно, $\hat{\theta}$ зависит от статистических данных через $T(x)$.

Примерами оценок максимального правдоподобия $\hat{\theta}$ для ряда моделей со скалярным параметром являются приведенные в табл. 2.2 соответствующие эффективные оценки. Так, для моделей $N(\theta, \sigma^2)$, $\Gamma(\theta, 1)$, $Bi(1, \theta)$, $\Pi(\theta)$ оценка $\hat{\theta} = X$. Рассмотрим еще несколько примеров нахождения оценок максимального правдоподобия.

Пример 2.17 (общая нормальная модель, оценка максимального правдоподобия ее параметров). Рассмотрим общую нормальную модель $\mathcal{N}(\theta_1, \theta_2)$. Как следует из примера 2.7, максимизация функции правдоподобия $L(x; \theta)$, $\theta = (\theta_1, \theta_2)$, эквивалентна здесь минимизации по θ функции

$$\psi(x; \theta) = \frac{(x - \theta_1)^2}{2\theta_2} + \left[\frac{1}{2} \left(\frac{s^2}{\theta_2} - 1 \right) - \ln \frac{s}{\theta_2} \right],$$

где $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ — реализация выборочной дисперсии S^2

(см. § 1.2). Как легко проверить, $\ln a \leq a - 1$, $\forall a > 0$. Отсюда, в частности, имеем неравенство $\ln x \leq (x^2 - 1)/2$, $\forall x > 0$; знак равенства достигается только при $x = 1$. Учитывая это неравенство, получаем, что $\psi(x; \theta) \geq 0$; знак равенства имеет место лишь в точке $\theta = (x, s)$. Таким образом, в данном случае оценка максимального правдоподобия существует, единственна и при этом $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2) = (X, S)$. Отметим, что полученная о. м. п. $\hat{\theta}$ — функция достаточной статистики $T = (T_1, T_2)$, рассмотренной в примере 2.7. Можно проверить, что полученное решение совпадает с решением уравнений правдоподобия.

Пример 2.18 (многомерная нормальная модель, оценка максимального правдоподобия ее параметров). Предположим, что наблюдается многомерная (скажем, размерности k) случайная величина ξ , распределенная по невырожденному нормальному закону $\mathcal{N}(\mu, \Sigma)$, где вектор средних значений $\mu = (\mu_1, \dots, \mu_k)$ и матрица вторых моментов $\Sigma = |\sigma_{ij}|_k$ ($\det \Sigma = |\Sigma| \neq 0$) — неизвестные параметры. Общее число неизвестных параметров (с учетом симметричности матрицы Σ) равно $k + k(k+1)/2$. Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{L}(\xi)$, т. е. X_i — независимые k -мерные случайные величины с плотностью

$$f(x; \theta) = \frac{1}{V(2\pi)^k |\Sigma|} \exp \left\{ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right\}, \quad \theta = (\mu, \Sigma).$$

Найдем оценки параметров θ такой модели по методу максимального правдоподобия. В данном случае функция правдоподобия

$$\begin{aligned} L(x; \theta) &= \prod_{i=1}^n f(x_i; \theta) = \\ &= (2\pi)^{-kn/2} |\Sigma|^{-n/2} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n (x_i - \mu)' \Sigma^{-1} (x_i - \mu) \right\}. \end{aligned} \quad (2.45)$$

Пусть $\bar{x} = (x_1 + \dots + x_n)/n$ (сложение векторов производится по координатно); тогда

$$\begin{aligned} \sum_{i=1}^n (x_i - \mu)' \Sigma^{-1} (x_i - \mu) &= \\ &= \sum_{i=1}^n (x_i - \bar{x})' \Sigma^{-1} (x_i - \bar{x}) + n (\bar{x} - \mu)' \Sigma^{-1} (\bar{x} - \mu). \end{aligned} \quad (2.46)$$

Введем выборочную матрицу вторых моментов, положив, по определению,

$$\hat{\Sigma}(\mathbf{x}) = \frac{1}{n} \sum_{l=1}^n (\mathbf{x}_l - \bar{\mathbf{x}})(\mathbf{x}_l - \bar{\mathbf{x}})' = |\hat{\sigma}_{ij}|, \quad (2.45)$$

где (i, j) -й выборочный второй момент $\hat{\sigma}_{ij} = \hat{\sigma}_{ij}(\mathbf{x})$ вычисляется по формуле

$$\hat{\sigma}_{ij} = \frac{1}{n} \sum_{l=1}^n (x_{li} - \bar{x}_i)(x_{lj} - \bar{x}_j), \quad i, j = 1, \dots, k$$

(здесь $\bar{\mathbf{x}}_l = (x_{l1}, \dots, x_{lk})$, $l = 1, \dots, n$, $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_k)$, $\bar{x}_i = \frac{1}{n} (x_{1i} + \dots + x_{ni})$). Воспользовавшись равенством $\mathbf{y}'\mathbf{B}\mathbf{y} = \text{tr}(\mathbf{B}\mathbf{Y})$, $\mathbf{Y} = \mathbf{y}\mathbf{y}'$, и линейностью оператора tr (взятия следа от матрицы), получим

$$\sum_{l=1}^n (\mathbf{x}_l - \bar{\mathbf{x}})' \Sigma^{-1} (\mathbf{x}_l - \bar{\mathbf{x}}) = n \text{tr}(\Sigma^{-1} \hat{\Sigma}(\mathbf{x})). \quad (2.47)$$

Учитывая равенства (2.46) – (2.47), формулу (2.45) можно переписать в виде

$$L(\mathbf{x}; \mathbf{0}) = (2\pi)^{-kn/2} \exp \left\{ -\frac{n}{2} (\bar{\mathbf{x}} - \boldsymbol{\mu})' \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) - \frac{n}{2} \text{tr}(\Sigma^{-1} \hat{\Sigma}(\mathbf{x})) - \frac{n}{2} \ln |\Sigma| \right\}.$$

Отсюда следует, что максимизация по $\mathbf{0}$ функции $L(\mathbf{x}; \mathbf{0})$ эквивалентна минимизации по $\boldsymbol{\mu}$ и Σ функции

$$\psi(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = (\bar{\mathbf{x}} - \boldsymbol{\mu})' \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) + [\text{tr}(\Sigma^{-1} \hat{\Sigma}(\mathbf{x})) - k - \ln |\Sigma^{-1} \hat{\Sigma}(\mathbf{x})|]$$

(постоянные k и $\ln |\hat{\Sigma}(\mathbf{x})|$ введены здесь для упрощения дальнейших преобразований). Обозначим через $\lambda_1, \dots, \lambda_k$ корни характеристического уравнения $|\Sigma^{-1} \hat{\Sigma}(\mathbf{x}) - \lambda \mathbf{E}_k| = 0$ или уравнения $|\hat{\Sigma}(\mathbf{x}) - \lambda \Sigma| = 0$. Тогда выражение в квадратных скобках равно $\lambda_1 + \dots + \lambda_k - k - \ln(\lambda_1 \dots \lambda_k)$ и можно записать

$$\psi(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = (\bar{\mathbf{x}} - \boldsymbol{\mu})' \Sigma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) + \sum_{i=1}^k (\lambda_i - 1 - \ln \lambda_i).$$

Поскольку Σ^{-1} – положительно определенная матрица и $a - 1 - \ln a \geq 0$, $a \geq 0$, из последнего соотношения получаем $\psi(\mathbf{x}; \boldsymbol{\mu}, \Sigma) \geq 0$, причем равенство имеет место только при $\boldsymbol{\mu} = \bar{\mathbf{x}}$, $\lambda_1 = \dots = \lambda_k = 1$, т. е. при $\boldsymbol{\mu} = \bar{\mathbf{x}}$, $\Sigma = \hat{\Sigma}(\mathbf{x})$. Таким образом, оценками максимального правдоподобия параметров $\boldsymbol{\mu}$ и $\Sigma = |\hat{\sigma}_{ij}|$ являются в данном случае соответственно статистики $\bar{\mathbf{X}}$ и $\hat{\Sigma}(\mathbf{X}) = |\hat{\sigma}_{ij}(\mathbf{X})|$.

Пример 2.19 (равномерная-1 модель, оценка максимального правдоподобия ее параметра). Если $\mathbf{X} = (X_1, \dots, X_n)$ – выборка

из равномерного распределения $R(0, \theta)$, то из формулы для $L(x; \theta)$, полученной в примере 2.9, следует, что $L(x; \theta)$ монотонно убывает по θ для $\theta \geq x_{(n)}$. При $\theta = x_{(n)}$ функция правдоподобия достигает максимума. Таким образом, о. м. п. $\hat{\theta} = X_{(n)}$, т. е. совпадает с достаточной статистикой. В этой точке функция правдоподобия разрывна, поэтому производная $L(x; \theta)$ в этой точке не существует. Таким образом, в данном случае о. м. п. не является решением уравнения правдоподобия. Это характерно для ситуаций, когда выборочное пространство $\mathcal{X}$ зависит от неизвестного параметра.

Пример 2.20 (равномерная модель (продолжение)). Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $R(\theta, \theta + 1)$. Тогда из формулы (2.37) имеем

$$L(x; \theta) = e^{(\theta + 1 - x_{(n)})} e^{(x_{(1)} - \theta)}.$$

Отсюда следует, что любое $\theta \in [x_{(n)} - 1, x_{(1)}]$ максимизирует функцию правдоподобия; таким образом, о. м. п. не единственна. В качестве решения можно выбрать, например, $\hat{\theta} = (X_{(1)} + X_{(n)} - 1)/2$, в этом случае $\hat{\theta}$ — функция достаточной статистики $T = (X_{(1)}, X_{(n)})$ (см. пример 2.9).

2. Принцип инвариантности для о. м. п. Полезным свойством оценок максимального правдоподобия является их инвариантность относительно преобразований параметра. Это означает, что если $q = q(\eta)$ — произвольная функция от η , взаимно однозначно отображающая Θ в Q , где $Q = \{q = (q_1, \dots, q_r)\}$ — некоторое множество в R^r , то о. м. п. $\hat{q} = q(\hat{\eta})$. Действительно, поскольку q — взаимно однозначное отображение, существует обратная функция $\eta = \eta(q)$. Тогда

$$\sup_{\eta \in \Theta} L(x; \eta) = \sup_{q \in Q} L(x; q),$$

и если максимум по η функции $L(x; \eta)$ достигается в точке $\hat{\eta}$, то максимум по q функции $L(x; q)$ имеет место в точке $\hat{q}$, удовлетворяющей уравнению $\eta(\hat{q}) = \hat{\eta}$, т. е. в точке $\hat{q} = q(\hat{\eta})$. Согласно этому, в каждой конкретной задаче можно выбирать в качестве исходной наиболее удобную параметризацию, а о. м. п. получать затем с помощью соответствующих преобразований. Проиллюстрируем это на нескольких примерах.

Пример 2.21 (общая нормальная модель, оценка максимального правдоподобия параметрической функции). Пусть в условиях примера 2.17 требуется оценить параметрическую функцию $\tau(\eta) = \Phi\left(\frac{x_0 - \eta_1}{\eta_2}\right)$, представляющую собой вероятность $P_6(\xi \leq x_0)$.

В этом случае можно положить, например, $q(\eta) = (\tau(\eta), \eta_2)$, и, согласно принципу инвариантности, о. м. п. $\hat{q} = (\tau(\hat{\eta}), \hat{\eta}_2)$, откуда $\hat{\tau} = \tau(\hat{\eta}) = \Phi\left(\frac{x_0 - \hat{X}}{\hat{S}}\right)$.

Пример 2.22 (двумерная нормальная модель, оценка максимального правдоподобия ее параметров). Пусть $X = (X_1, \dots, X_n)$ — вы-

борка из двумерного нормального распределения $\bullet f^*(0, 0)$, $\left(\begin{smallmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{smallmatrix} \right)$ с неизвестными $\sigma^2 > 0$ и $\rho \in (-1, 1)$, т. е. X_i — двумерные случайные величины с плотностью распределения

$$f(x, y; \theta) = \frac{1}{2\pi} \exp \left\{ -\frac{x^2 + y^2}{2\sigma^2(1-\rho^2)} + \frac{\rho xy}{\sigma^2(1-\rho^2)} - \frac{1}{2} \ln [\sigma^4(1-\rho^2)] \right\},$$

$$\theta = (\sigma^2, \rho). \quad (2.48)$$

Найдем о. м. п. $\hat{\sigma}^2$ и $\hat{\rho}$. Здесь плотность сложным образом зависит от параметров, поэтому удобно перейти к новым параметрам $q = (q_1, q_2)$, положив $q_1 = q_1(\eta) = -1/[2\sigma^2(1-\rho^2)]$, $q_2 = q_2(\eta) = -\rho/[\sigma^2(1-\rho^2)]$. Тогда формулу (2.48) можно переписать в виде

$$f(x, y; q) = \frac{1}{2\pi} \exp \{q_1(x^2 + y^2) + q_2xy + \tau(q)\},$$

где $\tau(q) = (1/2) \ln(4q_1^2 - q_2^2)$. Отсюда сразу же получаем, что уравнения правдоподобия для нахождения о. м. п. $\hat{q}_1$ и $\hat{q}_2$ имеют вид

$$\frac{1}{n} \sum_{i=1}^n (x_i^2 + y_i^2) = -\frac{\partial \tau(q)}{\partial q_1} = -\frac{4q_1}{4q_1^2 - q_2^2},$$

$$\frac{1}{n} \sum_{i=1}^n x_i y_i = -\frac{\partial \tau(q)}{\partial q_2} = \frac{q_2}{4q_1^2 - q_2^2}. \quad (2.49)$$

Но $\sigma^2 = -\frac{2q_1}{4q_1^2 - q_2^2}$, $\rho = -q_2/(2q_1)$, поэтому из соотношения (2.49) находим, что значения о. м. п. $\hat{\sigma}^2$ и $\hat{\rho}$ для заданной реализации выборки X вычисляются по формулам

$$\hat{\sigma}^2 = \frac{1}{2n} \sum_{i=1}^n (x_i^2 + y_i^2), \quad \hat{\rho} = 2 \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n (x_i^2 + y_i^2)},$$

где (x_i, y_i) — наблюдавшееся значение X_i , $i = 1, \dots, n$.

3. Метод накопления для приближенного вычисления о. м. п.
Оценки максимального правдоподобия в явном виде не всегда удается получить; в таких случаях прибегают к приближенным методам решения уравнений правдоподобия (когда такие уравнения можно составить). Это, например, *метод накопления*, предложенный Фишером. Он заключается в следующем. Пусть θ — скалярный параметр и функция правдоподобия $L(\theta) = L(x; \theta)$ дважды дифференцируема по θ . Разложим функцию вклада $U(\theta) = U(x; \theta)$ [см. формулу (2.15)] в ряд Тейлора в окрестности точки θ_0 , выбранной в качестве начального приближения для $\hat{\theta} = \hat{\theta}(x)$, и вычислим это разложение при $\theta = \hat{\theta}$. В этом случае так как $U(\hat{\theta}) = 0$, то $0 = U(\theta_0) + (\hat{\theta} - \theta_0) U'(\theta^*)$, где θ^* — некоторая промежуточная точка между $\hat{\theta}$ и θ_0 . Отсюда

$$\hat{\theta} = \theta_0 - U(\theta_0)/U'(\theta^*). \quad (2.50)$$

Если заменить теперь в этом равенстве θ^* на θ_0 и $U'(\theta_0)$ на $-n\dot{l}(\theta_0) = E_0 U'(X; \theta_0)$ (см. п. 1 § 2.2), то получим первое приближение $\theta_1 = \theta_0 + U(\theta_0)/n\dot{l}(\theta_0)$. Теперь этот процесс можно повторить, взяв в качестве нового начального приближения θ_1 и т. д. Таким образом, $(k+1)$ -е приближение по методу накопления вычисляют рекуррентно по формуле

$$\theta_{k+1} = \theta_k + U(\theta_k)/n\dot{l}(\theta_k), \quad k = 0, 1, 2, \dots \quad (2.51)$$

Процесс вычислений продолжают до достижения желаемой точности: $|\theta_{k+1} - \theta_k| < \varepsilon$. Важным является выбор начальной точки θ_0 . Обычно в качестве θ_0 берут значение какой-нибудь легко вычисляемой *состоятельной* оценки (определение см. ниже в п. 4); тогда все три точки θ_0 , θ^* и $\bar{\theta}$ при больших объемах выборки n будут находиться вблизи истинного значения параметра. Метод накопления, как правило, быстро приводит к цели, хотя в некоторых случаях этот процесс может и не сходиться.

Пример 2.23 (*модель Коши, оценка параметра по методу накопления*). Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения Коши $\mathcal{K}(\theta)$. Здесь функция вклада

$$U(\theta) = 2 \sum_{i=1}^n \frac{x_i - \theta}{1 + (x_i - \theta)^2}$$

и получить точное решение уравнения правдоподобия $U(\theta) = 0$ невозможно. Функция информации для модели Коши равна $1/2$ (см. табл. 2.1), поэтому последовательность (2.51) принимает вид

$$\theta_{k+1} = \theta_k + (2/n) U(\theta_k), \quad k = 0, 1, \dots$$

В качестве начального приближения θ_0 можно взять значение выборочной медианы $X_{([n/2]+1)}$, которая, согласно теореме 1.7, сходится по вероятности при $n \rightarrow \infty$ к истинному значению параметра θ , являющегося в данном случае теоретической медианой.

Пример 2.24 (*модель степенного ряда, метод накопления для нее*). Рассмотрим распределение типа степенного ряда, введенное в примере 2.12. Здесь функция вклада

$$U(\theta) = \frac{\partial}{\partial \theta} \ln \left(\theta^{n\bar{x}} f^{-n}(\theta) \prod_{i=1}^n a(x_i) \right) = \frac{n\bar{x}}{\theta} - n \frac{f'(\theta)}{f(\theta)} = n \frac{\bar{x} - \mu(\theta)}{\theta},$$

где $\mu(\theta)$ — теоретическое среднее распределения. Действительно, $\mu(\theta) = \sum_x x a(x) \theta^x / f(\theta) = \theta f'(\theta) / f(\theta)$. Таким образом, уравнение правдоподобия $U(\theta) = 0$ имеет вид

$$\mu(\theta) = \bar{x}. \quad (2.52)$$

Вычислим функцию информации $i(\theta)$. Имеем

$$i(\theta) = E_0 U^2(X_1; \theta) = \frac{1}{\theta^2} E_0 (X_1 - \mu(\theta))^2 = \frac{1}{\theta^2} D_0 X_1 = \sigma^2(\theta) / \theta^2,$$

где $\sigma^2(\theta)$ — теоретическая дисперсия; кроме того,

$$i(\theta) = -E_0 \left(\frac{\partial}{\partial \theta} U(X_i; \theta) \right) = E_0 \left(\frac{X_i - \mu(\theta)}{\theta^2} + \frac{\mu'(\theta)}{\theta} \right) = \mu'(\theta)/\theta.$$

Если уравнение (2.52) нельзя решить точно, то для приближенного определения значения о. м. п. $\hat{\theta}$ можно применить метод накопления.

В частности, для усеченного в нуле распределения Пуассона $f(\theta) = e^{-\theta} - 1$ и уравнение (2.52) принимает вид $\theta = x(1 - e^{-\theta})$, а функция информации равна

$$i(\theta) = [1 - (1 + \theta)e^{-\theta}][\theta(1 - e^{-\theta})^2].$$

4. Асимптотические свойства о. м. п. Метод максимального правдоподобия не всегда приводит к несмещенным оценкам. Так, из примера 2.17 следует, что оценкой максимального правдоподобия дисперсии θ_2^2 модели $N^*(\theta_1, \theta_2^2)$ является выборочная дисперсия S^2 , которая (см. п. 1 § 2.1) является смещенной оценкой. Однако ее смещение $E_0 S^2 - \theta_2^2 = -\theta_2^2/n$, $\theta = (\theta_1, \theta_2)$, убывает при увеличении объема выборки и в пределе, при $n \rightarrow \infty$, оценка становится несмещенной. Оценки (вообще говоря, смещенные), обладающие указанным свойством, называются *асимптотически несмещенными*. Таким образом, в данном случае о. м. п. является асимптотически несмещенной, т. е. свойства этой оценки, связанные с несмещенностью, с увеличением объема выборки «улучшаются». Это положение имеет достаточно общий характер. Именно: важнейшие свойства оценок максимального правдоподобия имеют асимптотический характер, т. е. справедливы для больших выборок. Широкое применение оценок максимального правдоподобия связано именно с их «хорошими» асимптотическими свойствами. Изложение асимптотической теории оценок максимального правдоподобия и составляет дальнейшее содержание этого параграфа. Чтобы подчеркнуть зависимость рассматриваемых ниже статистик от объема выборки, будем отмечать их индексом n .

Когда говорят об асимптотических свойствах оценок (или свойствах для больших выборок), то прежде всего имеют в виду их состоятельность, или, что эквивалентно, сходимость по вероятности рассматриваемых оценок к соответствующим оцениваемым теоретическим характеристикам. Так, в гл. 1 было показано, что большинство выборочных характеристик (выборочные моменты, квантили, значение эмпирической функции распределения в каждой точке и т. д.) сходятся по вероятности при $n \rightarrow \infty$ к значениям соответствующих теоретических характеристик n , следовательно, являются состоятельными оценками последних. Дадим строгое определение понятия состоятельности.

Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{L}(\xi) \in \mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$, где параметрическое множество Θ — в общем случае некоторый невырожденный открытый интервал r -мерного евклидова пространства R^r . По определению, оценка $T_n = T_n(X)$ для заданной параметрической функции $\tau(\theta)$ назы-

вается *состоятельной*, если при $n \rightarrow \infty$

$$T_n \xrightarrow{P_\theta} \tau(\theta), \quad \forall \theta \in \Theta,$$

т. е., каково бы ни было истинное значение параметра θ , оценка T_n сходится по вероятности (относительно распределения P_θ) к истинному значению оцениваемой функции.

Свойство состоятельности обязательно для любого правила оценивания, однако оно, по существу, является асимптотическим и не связано со свойствами оценки при фиксированном объеме выборки (в отличие от свойств несмещенности и оптимальности).

Рассмотрим простой и важный критерий состоятельности.

Теорема 2.10. Пусть

$$E_\theta T_n = \tau(\theta) + \varepsilon_n, \quad D_\theta T_n = \delta_n$$

и $\varepsilon_n = \varepsilon_n(\theta) \rightarrow 0$, $\delta_n = \delta_n(\theta) \rightarrow 0$ при $n \rightarrow \infty$ для всех $\theta \in \Theta$, тогда T_n — состоятельная оценка функции $\tau = \tau(\theta)$.

□ Поскольку $|T_n - E_\theta T_n| \geq |T_n - \tau| - |E_\theta T_n - \tau|$, событие $\{|T_n - \tau| \geq \varepsilon\}$ влечет событие $\{|T_n - E_\theta T_n| \geq \varepsilon - |\varepsilon_n|\}$. Отсюда на основании неравенства Чебышева имеем

$$P_\theta(|T_n - \tau| \geq \varepsilon) \leq P_\theta(|T_n - E_\theta T_n| \geq \varepsilon - |\varepsilon_n|) \leq \frac{\delta_n}{(\varepsilon - |\varepsilon_n|)^2} \rightarrow 0$$

при $n \rightarrow \infty$ и всех $\theta \in \Theta$. ■

Приведем некоторые утверждения о сходимости функций от случайных величин, которые будут полезны в дальнейшем.

1°. Имеет место следующее непосредственное обобщение теоремы 1.5: если $\eta_n \xrightarrow{P} \eta$ и функция φ непрерывна, то $\varphi(\eta_n) \xrightarrow{P} \varphi(\eta)$, а также $\mathcal{L}(\varphi(\eta_n)) \rightarrow \mathcal{L}(\varphi(\eta))$.

2°. Пусть $\{\eta_n, \zeta_n\}$, $n = 1, 2, \dots$, — последовательность пар случайных величин. Тогда:

$$a) \eta_n \xrightarrow{P} 0, \zeta_n \xrightarrow{P} \zeta \Rightarrow \eta_n \zeta_n \xrightarrow{P} \zeta;$$

$$б) \eta_n \xrightarrow{P} 0, \mathcal{L}(\zeta_n) \rightarrow \mathcal{L}(\zeta) \Rightarrow \mathcal{L}(\eta_n) \rightarrow \mathcal{L}(\zeta);$$

$$в) \mathcal{L}(\eta_n) \rightarrow \mathcal{L}(\eta), \zeta_n \xrightarrow{P} 0 \Rightarrow \eta_n \zeta_n \xrightarrow{P} 0;$$

$$г) \mathcal{L}(\eta_n) \rightarrow \mathcal{L}(\eta), \zeta_n \xrightarrow{P} c = \text{const} \Rightarrow \mathcal{L}(\eta_n + \zeta_n) \rightarrow \mathcal{L}(\eta + c),$$

$$\mathcal{L}(\eta_n \zeta_n) \rightarrow \mathcal{L}(c\eta), \mathcal{L}(\eta_n / \zeta_n) \rightarrow \mathcal{L}(\eta / c) \text{ при } c \neq 0;$$

д) $\eta_n \xrightarrow{P} 0$, $\mathcal{L}(\zeta_n) \rightarrow \mathcal{L}(\zeta)$, функция φ непрерывна $\Rightarrow \varphi(\eta_n) - \varphi(\zeta_n) \xrightarrow{P} 0$. Проверка этих простых утверждений предоставляется читателю.

3°. Пусть $T_n = T_n(X)$, $X = (X_1, \dots, X_n)$, — оценка скалярного параметра θ в модели $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$ такая, что $\mathcal{L}_\theta \times \times (1/n(T_n - \theta)) \rightarrow \nu^*(0, \sigma^2(\theta))$ при $n \rightarrow \infty$ и всех $\theta \in \Theta$. Пусть, далее, функция φ дифференцируема и $\varphi' \neq 0$. Тогда

$$\mathcal{L}_\theta(1/n(\varphi(T_n) - \varphi(\theta))) \rightarrow \nu^*(0, [\varphi'(\theta)]^2 \sigma^2(\theta)). \quad (2.53)$$

Кроме того, если φ' непрерывна, то

$$\mathcal{L}_\theta(1/n(\varphi(T_n) - \varphi(\theta))/\varphi'(T_n)) \rightarrow \nu^*(0, \sigma^2(\theta)), \quad (2.54)$$

а если и $\sigma(\theta)$ непрерывна, то

$$\mathcal{L}_0 \left(\sqrt{n} \frac{\varphi(T_n) - \varphi(\theta)}{\varphi'(T_n) \sigma(T_n)} \right) \rightarrow \mathcal{N}^*(0, 1). \quad (2.55)$$

Доказательство основано на разложении Тейлора

$$\varphi(T_n) - \varphi(\theta) = (T_n - \theta) (\varphi'(\theta) + \xi_n),$$

где для $\forall \varepsilon > 0$ $|\xi_n| < \varepsilon$ при $|T_n - \theta| < \delta = \delta(\varepsilon)$. Отсюда $P_n(|\xi_n| < \varepsilon) \geq P_n(|T_n - \theta| < \delta)$. Но здесь правая часть по условию стремится к 1 при $n \rightarrow \infty$. Следовательно, $\xi_n \xrightarrow{P_n} 0$. Отсюда на основании свойства 2° в)

$$\sqrt{n}(\varphi(T_n) - \varphi(\theta)) = \sqrt{n}(T_n - \theta) \varphi'(\theta) = \sqrt{n}(T_n - \theta) \xi_n \xrightarrow{P_n} 0.$$

Это в силу свойства 2° б) означает, что случайная величина $\sqrt{n}(\varphi(T_n) - \varphi(\theta))$ имеет такое же предельное распределение, как и величина $\sqrt{n}(T_n - \theta) \varphi'(\theta)$, т. е. нормальное распределение $\mathcal{N}^*(0, [\varphi'(\theta)]^2 \sigma^2(\theta))$, что доказывает (2.53). Далее, если φ' непрерывна, то на основании свойства 1° $\varphi'(T_n) \xrightarrow{P_n} \varphi'(\theta)$. Отсюда и из предыдущего в силу свойства 2° г) следует (2.54). Рассуждая аналогично, можно убедиться в справедливости формулы (2.55).

Приведем без доказательства обобщение утверждения 3° на случай векторного параметра $\theta = (\theta_1, \dots, \theta_r)$.

4°. Пусть $T_n = (T_{1n}, \dots, T_{rn})$ — оценка параметра θ , удовлетворяющая условию $\mathcal{L}_2(\sqrt{n}(T_n - \theta)) \rightarrow \mathcal{N}^*(0, \Sigma(\theta))$ при $n \rightarrow \infty$ и всех $\theta \in \Theta$. Тогда для любой дифференцируемой функции φ от r переменных

$$\mathcal{L}_2(\sqrt{n}(\varphi(T_n) - \varphi(\theta))) \rightarrow \mathcal{N}^*(0, v^2(\theta)) \quad (2.56)$$

при условии, что $v(\theta) \neq 0$, где $v^2(\theta) = \mathbf{b}'(\theta) \Sigma(\theta) \mathbf{b}(\theta)$, $\mathbf{b}(\theta) = \left(\frac{\partial \varphi(\theta)}{\partial \theta_1}, \dots, \frac{\partial \varphi(\theta)}{\partial \theta_r} \right)$. Если, кроме того, функция φ непрерывно дифференцируема и все элементы матрицы вторых моментов $\Sigma(\theta)$ непрерывны по θ , то

$$\mathcal{L}_2(\sqrt{n}(\varphi(T_n) - \varphi(\theta))/v(T_n)) \rightarrow \mathcal{N}^*(0, 1). \quad (2.57)$$

Сформулируем теперь основные асимптотические свойства оценок максимального правдоподобия. Предположим, что модель $\mathcal{F}$ является регулярной в смысле § 2.2, а функция правдоподобия $L_n(x; \theta)$ при всех $n \geq 1$ и $x \in \mathcal{X}$ имеет лишь один локальный максимум по θ , лежащий внутри Θ , и притом достижимый (т. е. о. м. п. θ_n существует и совпадает с локальным максимумом). Тогда:

1) θ_n является состоятельной оценкой параметра θ (без доказательства);

2) если дополнительно функция $f(x; \theta)$ трижды дифференцируема по θ и при этом существует не зависящая от θ функция $M(x)$

такая, что для всех $\theta \in \Theta$

$$\left| \frac{\partial^2 \ln f(x; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \leq M(x), \quad E_\theta M(X_1) < \infty \quad (i, j, k = 1, \dots, r),$$

то при $n \rightarrow \infty$

$$\mathcal{L}_\theta(\sqrt{n}(\hat{\theta}_n - \theta)) \rightarrow \mathcal{N}^*(0, I^{-1}(\theta)), \quad (2.58)$$

где $I(\theta)$ — информационная матрица, определенная в (2.32) и являющаяся (по условию) невырожденной. Более того, если $\tau(\theta)$ — непрерывно дифференцируемая функция от θ и $\hat{\tau}_n = \tau(\hat{\theta}_n)$ — ее оценка максимального правдоподобия, то

$$\mathcal{L}_\theta(\sqrt{n}(\hat{\tau}_n - \tau(\theta))) \rightarrow \mathcal{N}^*(0, \sigma_\tau^2(\theta)), \quad (2.59)$$

где $\sigma_\tau^2(\theta) = b'(\theta) I^{-1}(\theta) b(\theta)$, $b(\theta) = \left(\frac{\partial \tau(\theta)}{\partial \theta_1}, \dots, \frac{\partial \tau(\theta)}{\partial \theta_r} \right)$.

Таким образом, для широкого класса моделей оценки максимального правдоподобия являются состоятельными и асимптотически нормальными.

Для случая скалярного параметра соотношения (2.58) и (2.59) принимают соответственно вид

$$\mathcal{L}_\theta(\sqrt{n}(\hat{\theta}_n - \theta)) \rightarrow \mathcal{N}^*(0, 1/i(\theta)) \quad \text{и} \quad (2.60)$$

$$\mathcal{L}_\theta(\sqrt{n}(\hat{\tau}_n - \tau(\theta))) \rightarrow \mathcal{N}^*(0, [\tau'(\theta)]^2/i(\theta)), \quad (2.61)$$

где $i(\theta)$ — функция информации, определенная в (2.18).

Доказательство свойства асимптотической нормальности о. м. п. (если ограничиться случаем скалярного параметра) основывается на разложении функции вклада $U_n(\theta) = U_n(X; \theta)$ в ряд Тейлора относительно истинного значения параметра θ и рассмотрении этого разложения в точке $\hat{\theta}_n$. Имеем

$$0 = U_n(\hat{\theta}_n) = U_n(\theta) + (\hat{\theta}_n - \theta) U'_n(\theta) + \frac{1}{2} (\hat{\theta}_n - \theta)^2 U''_n(\theta^*),$$

где θ^* — некоторая промежуточная точка между θ и $\hat{\theta}_n$. Это равенство можно записать в виде

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{U_n(\theta)}{\sqrt{n} i(\theta)} \left[-\frac{U'_n(\theta)}{ni(\theta)} + \varepsilon_n \right]^{-1}, \quad (2.62)$$

где

$$|\varepsilon_n| = |\hat{\theta}_n - \theta| |U''_n(\theta^*)| / 2ni(\theta) \leq \frac{|\hat{\theta}_n - \theta|}{2i(\theta)} \frac{1}{n} \sum_{j=1}^n M(X_j).$$

Поскольку $\hat{\theta}_n \xrightarrow{P_\theta} \theta$, из условия, наложенного на функцию $M(x)$, на основании свойства 2° в) следует, что $\varepsilon_n \xrightarrow{P_\theta} 0$. Применим к величине

$$\frac{1}{n} U'_n(\theta) = \frac{1}{n} \sum_{j=1}^n \frac{\partial \ln f(X_j; \theta)}{\partial \theta}$$

закон больших чисел, согласно которому [см. (2.19)]

$$-\frac{1}{ni(\theta)} U_n'(\theta) \xrightarrow{P_\theta} -\frac{1}{i(\theta)} E_\theta \left(\frac{\partial^2 \ln f(X_1; \theta)}{\partial \theta^2} \right) = 1.$$

К случайной величине

$$\frac{1}{\sqrt{ni(\theta)}} U_n(\theta) = \frac{1}{\sqrt{ni(\theta)}} \sum_{j=1}^n \frac{\partial \ln f(X_j; \theta)}{\partial \theta}$$

применима центральная предельная теорема, по которой в силу соотношений (2.16) и (2.18) при $n \rightarrow \infty$

$$\mathcal{L}_\theta \left(\frac{1}{\sqrt{ni(\theta)}} U_n(\theta) \right) \rightarrow \mathcal{N}^* \left(0, \frac{1}{i(\theta)} \right).$$

Отсюда и из соотношения (2.62) в силу свойства 2^о г) следует, что такое же предельное распределение имеет и случайная величина $\sqrt{n}(\hat{\theta}_n - \theta)$, т. е. справедливо соотношение (2.60).

Утверждение (2.61) является прямым следствием (2.60) и утверждения 3^о.

Назовем *асимптотической дисперсией* статистики T_n , удовлетворяющей при $n \rightarrow \infty$ условию

$$\mathcal{L}_\theta(\sqrt{n}(T_n - \tau(\theta))) \rightarrow \mathcal{N}^*(0, \sigma^2(\theta)),$$

величину $\sigma^2(\theta)/n$. Тогда из соотношений (2.60) — (2.61) следует, что асимптотическая дисперсия о. м. п. $\hat{\theta}_n$ (о. м. п. $\hat{\tau}_n$) совпадает с нижней границей в неравенстве Рао — Крамера (2.20) для дисперсий всех несмещенных оценок рассматриваемой параметрической функции. Введем следующее понятие, являющееся в некотором смысле асимптотическим аналогом понятия эффективности (см. п. 2 § 2.2): если оценка T_n (для параметра θ) является асимптотически нормальной $\mathcal{N}^*(\theta, 1/ni(\theta))$, то эта оценка — *асимптотически эффективная*.

Для любой оценки T_n , удовлетворяющей условию

$$\mathcal{L}_\theta(T_n) \sim \mathcal{N}^*(\theta, \sigma_T^2(\theta)/n), \quad (2.63)$$

ее *асимптотическая эффективность* $\text{eff}(T_n; \theta)$ определяется как отношение нижней границы Рао — Крамера к асимптотической дисперсии оценки T_n :

$$\text{eff}(T_n; \theta) = [i(\theta) \sigma_T^2(\theta)]^{-1}. \quad (2.64)$$

Из соотношений (2.60) и (2.64) следует третье свойство оценок максимального правдоподобия.

3) Оценка $\hat{\theta}_n$ является асимптотически эффективной оценкой параметра θ , т. е. ее асимптотическая эффективность $\text{eff}(\hat{\theta}_n; \theta) \equiv 1$.

Приведем пример, показывающий, что оценки максимального правдоподобия не всегда асимптотически нормальны.

Пример 2.25 (равномерная модель, распределение о. м. п. ее параметра). Пусть $X = (X_1, \dots, X_n)$ — выборка из равномерного распределения $R(0, \theta)$. Тогда, как показано в примере 2.19, $\hat{\theta}_n = X_{(n)}$. Но, как было показано в § 1.4, экстремальные значения выборки не являются асимптотически нормальными. В частности, в рассматриваемом случае по теореме 1.8 при $n \rightarrow \infty$

$$\mathcal{L}_{\theta} \left(\frac{n}{\theta} (\theta - \hat{\theta}_n) \right) \rightarrow \Gamma(1, 1),$$

т. е. в пределе имеем экспоненциальное распределение. Причина, по которой предельное распределение не является нормальным, заключается в нерегулярности модели.

Замечание 1. Из соотношения (2.61), вообще говоря, не следует, что $n D \hat{\theta}_n \rightarrow [\tau'(\theta)]' / i(\theta)$ при $n \rightarrow \infty$ (из сходимости распределений еще не следует сходимость моментов). Более того, можно привести примеры, когда у статистики $\hat{\theta}_n$ вообще не существуют конечные моменты ни при каком n , но утверждение (2.61) имеет место. Рассмотрим, например, задачу оценивания параметрической функции $\tau(\theta) = 1/\theta$ в пуассоновской модели $\Pi(\theta)$. Здесь теоретическое среднее $\mu(\theta) = 0$, поэтому из результата примера 2.24 следует, что о. м. п. $\hat{\theta}_n$ совпадает с выборочным средним $\bar{X}$. На основании принципа инвариантности о. м. п. $\hat{\tau}_n = 1/\bar{X}$ и для нее справедливо утверждение об асимптотической нормальности. Соотношение (2.61) в данном случае принимает вид

$$\mathcal{L}_{\theta} \left(\sqrt{n} \left(\hat{\tau}_n - \frac{1}{\theta} \right) \right) \rightarrow \phi^*(0, \theta^2), \quad \forall \theta > 0$$

(см. табл. 2.1). В то же время $\mathcal{L}_{\theta}(n\bar{X}) = \Pi(n\theta)$, откуда $P_0(\bar{X} = 0) = e^{-n\theta} > 0$. Следовательно, с положительной вероятностью $\bar{X}$ принимает нулевое значение и поэтому статистика $\hat{\tau}_n$ ни при каком n не имеет конечных моментов.

Замечание 2. В соотношениях (2.58) — (2.61) вторые моменты предельных распределений зависят от неизвестного параметра θ . Часто эту неопределенность бывает необходимо исключить, т. е. заменить неизвестные характеристики оценками, которые можно вычислить по выборочным данным, и при этом так, чтобы нормальная форма предельного распределения не изменилась. Такое преобразование можно сделать на основании утверждений 3° и 4° [см. соотношения (2.55) и (2.57)], заменив θ на его о. м. п. $\hat{\theta}_n$. Сохранение формы предельного распределения гарантируется при этом условием непрерывности по θ вторых моментов. Так, например, если $i(\theta)$ и $\tau'(\theta)$ — непрерывные функции, то справедливо вытекающее из (2.61) соотношение

$$\mathcal{L}_{\theta} \left(\sqrt{n i(\hat{\theta}_n)} \frac{\hat{\tau}_n - \tau(\hat{\theta}_n)}{\tau'(\hat{\theta}_n)} \right) \rightarrow \phi^*(0, 1). \quad (2.65)$$

Замечание 3. Определение асимптотически эффективной оценки предполагает, что невозможно построить оценку с меньшей асимптотической дисперсией. Как правило, это имеет место для регулярных моделей, удовлетворяющих указанным выше условиям асимптотической нормальности оценок максимального правдоподобия, хотя и существуют исключения, как это видно из примера 2.26. Для нерегулярных моделей (когда неприменимо неравенство Рао — Крамера) асимптотическая дисперсия о. м. п. может иметь порядок, меньший n^{-1} . Так, из результатов примеров 2.15 и 2.19 следует, что дисперсия о. м. п. $\hat{\theta}$ в модели $R(0, \theta)$ имеет при больших n порядок n^{-2} .

Пример 2.26 (нормальная-1 модель, сверхэффективная оценка ее параметра). Пусть $X = (X_1, \dots, X_n)$ — выборка из $\phi^*(\theta, 1)$. Тогда о. м. п. $\hat{\theta}_n = \bar{X}$ и

$\mathcal{L}_\theta(\hat{\theta}_n) = \varphi^*(\theta, 1/n)$. Рассмотрим оценку

$$T_n = \begin{cases} X & \text{при } |X| \geq a_n, \\ bX & \text{при } |X| < a_n, \end{cases}$$

где константа $a_n \rightarrow 0$, но $\sqrt{n}a_n \rightarrow \infty$ при $n \rightarrow \infty$. Найдём предельный закон распределения этой оценки. Имеем

$$P_\theta(\sqrt{n}(T_n - \theta) \leq x) = p_n P_\theta(\sqrt{n}(X - \theta) \leq x) + (1 - p_n) P_\theta(\sqrt{n}(bX - \theta) \leq x),$$

где

$$p_n = P_\theta(|X| \geq a_n) = \Phi(\sqrt{n}(\theta - a_n)) + \Phi(-\sqrt{n}(\theta + a_n)) \xrightarrow{n \rightarrow \infty} \begin{cases} 1 & \text{при } |\theta| > 0, \\ 0 & \text{при } \theta = 0. \end{cases}$$

Таким образом, при $n \rightarrow \infty$ для $|\theta| > 0$ и $\theta = 0$ соответственно имеем

$$\mathcal{L}_\theta(\sqrt{n}(T_n - \theta)) \rightarrow \varphi^*(0, 1), \quad \mathcal{L}_\theta(\sqrt{n}T_n) \rightarrow \varphi^*(0, b^2).$$

Оценка T_n удовлетворяет условию (2.63) с $\sigma_{\frac{1}{2}}^2(\theta) = \begin{cases} 1 & \text{при } |\theta| > 0, \\ b^2 & \text{при } \theta = 0. \end{cases}$ Следовательно, её асимптотическая эффективность (см. табл. 2.1)

$$\text{eff}(T_n; \theta) = \begin{cases} 1 & \text{при } |\theta| > 0, \\ 1/b^2 & \text{при } \theta = 0. \end{cases}$$

При $b < 1$ $\text{eff}(T_n; \theta) \geq 1$ со строгим неравенством в точке $\theta = 0$. Причина этого заключается в данном случае в нарушении непрерывности функции $\sigma_{\frac{1}{2}}^2(\theta)$ в нуле.

Точки θ , в которых $\text{eff}(T_n; \theta) > 1$, называют *точками сверхэффективности*. В рассмотренном примере это точка $\theta = 0$. Как показал Ле Кам (1953 г.), множество точек сверхэффективности не более чем счетно, поэтому явление сверхэффективности не играет существенной роли в теории оценивания. Известны дополнительные условия регулярности, которые исключают это явление. Так, Рао доказал, что для этого достаточно потребовать равномерности по параметру θ в предельном соотношении (2.63).

В заключение рассмотрим асимптотическую эффективность оценивания среднего θ нормальной модели $\varphi^*(\theta, \sigma^2)$ с помощью выборочной медианы, т. е. когда $T_n = X_{([n/2] + 1)}$. В силу соотношения (1.24) выборочная медиана удовлетворяет условию (2.63) с $\sigma_{\frac{1}{2}}^2(\theta) = \pi\sigma^2/2$, а так как в данном случае $i(\theta) = 1/\sigma^2$ (см. табл. 2.2), то $\text{eff}(T_n; \theta) = 2/\pi = 0,637 \dots$. Это означает, что при больших n выборочное среднее $\bar{X} (= \hat{\theta}_n)$ для выборки объема $n' = 2n/\pi$ оценивает истинное значение θ с такой же точностью, что и выборочная медиана $X_{([n/2] + 1)}$ выборки объема n , независимо от значений θ и σ^2 .

§ 2.5. Метод моментов и другие методы, основанные на группированных данных

1. Метод моментов. Кроме рассмотренных в предыдущих параграфах общих подходов для оценивания неизвестных параметров распределений на практике используют и другие методы получения оценок. Исторически первым общим методом точечного оценивания неизвестных параметров является *метод моментов*, предложенный К. Пирсоном. Суть этого метода состоит в следующем.

Пусть, как обычно, $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{L}(\xi) \in \mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$, где $\theta = (\theta_1, \dots, \theta_r)$ и $\Theta \subseteq R^r$. Предположим, что у наблюдаемой случайной величины ξ существуют первые r моментов $\alpha_k = E\xi^k$, $k = 1, \dots, r$. Они являются функциями от неизвестных параметров θ : $\alpha_k = \alpha_k(\theta)$. Рассмотрим соответствующие выборочные моменты $A_{nk}(X)$ (см. п. 1 § 1.2); пусть $a_k = A_{nk}(x)$ — значения этих величин для наблюдавшейся реализации x выборки X . Тогда метод моментов состоит в приравнивании значений a_k и теоретических моментов:

$$\alpha_k(\theta) = a_k, \quad k = 1, \dots, r. \quad (2.66)$$

Решая эти уравнения относительно $\theta_1, \dots, \theta_r$, получаем значения оценок параметров по методу моментов.

Обоснуем этот метод. Известно (см. § 2.3), что выборочные моменты $A_{nk}(X)$ являются несмещенными и состоятельными оценками теоретических моментов $\alpha_k(\theta)$. Предположим, что соответствие между $\theta_1, \dots, \theta_r$ и $\alpha_1, \dots, \alpha_r$ взаимно однозначное и взаимно непрерывное, т. е. существуют непрерывные функции $\varphi_1, \dots, \varphi_r$ такие, что $\theta_i = \varphi_i(\alpha_1, \dots, \alpha_r)$, $i = 1, \dots, r$. Тогда решения уравнений (2.66) можно записать в виде $\theta_i(x) = \varphi_i(a_1, \dots, a_r)$, $i = 1, \dots, r$, а соответствующие оценки $\hat{\theta}_i(X) = \varphi_i(A_{n1}(X), \dots, A_{nr}(X))$ в силу теоремы 1.5 при каждом θ будут сходиться по вероятности (по распределению P_θ) при $n \rightarrow \infty$ к θ_i , т. е. статистики $\hat{\theta}_i(X)$ являются состоятельными оценками θ_i , $i = 1, \dots, r$.

Таким образом, метод моментов при определенных условиях приводит к состоятельным оценкам; при этом уравнения (2.66) во многих случаях просты и их решение (в отличие, например, от метода максимального правдоподобия) не связано с большими вычислительными трудностями.

Пример 2.27 (модель гамма, оценивание параметров методом моментов). Рассмотрим модель гамма $\Gamma(\theta_1, \theta_2)$, когда оба параметра неизвестны (здесь $\Theta = \{\theta = (\theta_1, \theta_2): \theta_i > 0, i = 1, 2\}$). Имеем (см. табл. В.1)

$$\begin{aligned} \alpha_k &= \int_0^\infty x^k + k - 1 e^{-x/\theta_1} \Gamma(\theta_2) dx / \Gamma(\theta_2) \theta_1^{\theta_2} = \theta_1^k \frac{\Gamma(\theta_2 + k)}{\Gamma(\theta_2)} = \\ &= \theta_1^k \theta_2 (\theta_2 + 1) \dots (\theta_2 + k - 1). \end{aligned}$$

В частности, $\alpha_1 = \theta_1 \theta_2$, $\alpha_2 = \theta_1^2 \theta_2 (\theta_2 + 1)$, откуда $\theta_1 = (\alpha_2 - \alpha_1^2) / \alpha_1$, $\theta_2 = \alpha_1^2 / (\alpha_2 - \alpha_1^2)$. Окончательно имеем, что оценками параметров по методу моментов являются в данном случае статистики [см. формулы (1.10) и (1.12)]

$$\hat{\theta}_1(X) = \frac{A_{n2} - A_{n1}^2}{A_{n1}} = \frac{S^2}{\bar{X}}, \quad \hat{\theta}_2(X) = \frac{A_{n1}^2}{A_{n2} - A_{n1}^2} = \frac{\bar{X}^2}{S^2}.$$

Отметим, что метод моментов неприменим, когда теоретические моменты нужного порядка не существуют (например, для распределения Коши). Кроме того, оценки метода моментов, вообще

говоря, неэффективны. Поэтому их часто используют только в качестве первых приближений, ссылаясь на которых можно находить последующие приближения с большей эффективностью.

2. Метод минимума хи-квадрат. Систему различных методов построения оценок можно получить, конструируя тем или иным способом меру $\mathcal{D}$ отклонения эмпирической функции распределения $F_n(x)$ от $F(x; \theta)$. В любом случае такая мера является функцией от выборки $X = (X_1, \dots, X_n)$ и параметра θ : $\mathcal{D} = \mathcal{D}(X; \theta)$. Если такая мера имеется, то в качестве оценки параметра θ следует взять значение, минимизирующее меру $\mathcal{D}$.

Рассмотрим в качестве примера одну из наиболее известных и употребительных мер — меру хи-квадрат, предложенную К. Пирсоном. Это мера определяется следующим образом. Разобьем множество $\mathcal{E}$ возможных значений наблюдаемой случайной величины ξ на N непересекающихся подмножеств $\mathcal{E}_1, \dots, \mathcal{E}_N$, ($\mathcal{E} = \mathcal{E}_1 \cup \dots \cup \mathcal{E}_N$, $\mathcal{E}_i \cap \mathcal{E}_j = \emptyset$, $i \neq j$), и пусть v_j — число элементов выборки $X = (X_1, \dots, X_n)$, попавших в подмножество $\mathcal{E}_j$: $v_j = |\{i: X_i \in \mathcal{E}_j\}|$, $j = 1, \dots, N$ ($v_1 + \dots + v_N = n$). Обозначим через $p_j(\theta)$ вероятность попадания в подмножество $\mathcal{E}_j$: $p_j(\theta) = P_\theta(\xi \in \mathcal{E}_j) = \int_{\mathcal{E}_j} dF(x; \theta)$, $j = 1, \dots, N$ ($p_1(\theta) + \dots + p_N(\theta) = 1$). Относительная частота v_j/n попадания в подмножество $\mathcal{E}_j$ является состоятельной оценкой вероятности $p_j(\theta)$, поэтому мерой отклонения выборочных данных от теоретических значений может быть мера

$$\mathcal{D} = \sum_{j=1}^N c_j (v_j/n - p_j(\theta))^2.$$

Если здесь положить веса $c_j = n/p_j(\theta)$, то получим меру хи-квадрат:

$$\chi^2 = \sum_{j=1}^N \frac{n}{p_j(\theta)} \left(\frac{v_j}{n} - p_j(\theta) \right)^2 = \sum_{j=1}^N \frac{(v_j - np_j(\theta))^2}{np_j(\theta)} = \sum_{j=1}^N \frac{v_j^2}{np_j(\theta)} - n. \quad (2.67)$$

В дальнейшем иногда будем также использовать для этой меры название мера экс-квадрат и обозначать ее через X^2 , обозначая символом χ^2 величину, имеющую распределение хи-квадрат.

Оценку θ , полученную из условия обращения меры χ^2 в минимум, называют оценкой по методу минимума χ^2 . При достаточно общих условиях эта оценка оказывается состоятельной, асимптотически нормальной и асимптотически эффективной (как и оценка максимального правдоподобия).

Для нахождения указанной оценки надо решить систему уравнений

$$\sum_{j=1}^N \frac{v_j}{p_j^2(\theta)} \frac{\partial p_j(\theta)}{\partial \theta_k} = 0, \quad k = 1, \dots, r \quad (\theta = (\theta_1, \dots, \theta_r)).$$

Однако это трудно сделать даже в простейших случаях, поэтому вместо нее обычно рассматривают систему

$$\sum_{j=1}^N \frac{v_j}{p_j(\theta)} \frac{\partial p_j(\theta)}{\partial \theta_k} = 0, \quad k=1, \dots, r, \quad (2.68)$$

решение которой находить, как правило, гораздо проще. Оценки параметров θ , получающиеся как решение системы (2.68), называют *оценками по видоизмененному методу минимума χ^2* . При весьма общих условиях последний метод позволяет получить оценки, обладающие в случае больших выборок теми же асимптотическими свойствами, что и оценки, найденные с помощью метода минимума χ^2 .

3. Мультиномиальные оценки максимального правдоподобия. Описанные в п. 2 способы оценивания параметров предполагают использование не самих исходных наблюдений $X = (X_1, \dots, X_n)$, а частот $v = (v_1, \dots, v_N)$ попадания элементов выборки в соответствующие подмножества $\mathcal{E}_1, \dots, \mathcal{E}_N$. Такой «частотный» способ представления статистических данных называют *методом группировки наблюдений*, а подмножества $\mathcal{E}_1, \dots, \mathcal{E}_N$ — *множествами* (или *интервалами*, если $\mathcal{E}_j$ — интервалы на действительной прямой) *группировки*. В каждом конкретном случае метод группировки полностью определяется разбиением $\mathcal{E} = \mathcal{E}_1 \cup \dots \cup \mathcal{E}_N$. Если такое разбиение построено, то тем самым однозначно определены вероятности $p = p(\theta) = (p_1(\theta), \dots, p_N(\theta))$ попадания результата каждого наблюдения в соответствующие подмножества $\mathcal{E}_1, \dots, \mathcal{E}_N$ и каждая реализация $x = (x_1, \dots, x_n)$ выборки $X = (X_1, \dots, X_n)$ однозначно определяет соответствующую реализацию $h = (h_1, \dots, h_N)$ вектора частот $v = (v_1, \dots, v_N)$. Обозначим через L_{nN} множество всех реализаций вектора v , т. е.

$$L_{nN} = \{h = (h_1, \dots, h_N): h_j = 0, 1, \dots, n, j = 1, \dots, N, h_1 + \dots + h_N = n\}, \quad (2.69)$$

и найдем вероятности $P_1(v=h)$, $h \in L_{nN}$. Число различных способов зафиксировать h_1 элементов выборки, попавших в подмножество $\mathcal{E}_1, \dots, h_N$ элементов выборки, попавших в подмножество $\mathcal{E}_N$, очевидно, равно

$$C_n^{h_1} C_{n-h_1}^{h_2} \dots C_{n-h_1-\dots-h_{N-1}}^{h_N} = \frac{n!}{h_1! \dots h_N!}.$$

Вероятность того, что h_j элементов выборки попадут в подмножество $\mathcal{E}_j$, $j = 1, \dots, N$, в силу независимости испытаний равна

$$\prod_{j=1}^N p_j^{h_j}(\theta) \text{ и не зависит от конкретных номеров испытаний.}$$

Следовательно,

$$P_1(v=h) = L^{(1)}(h; \theta) = \frac{n!}{h_1! \dots h_N!} \prod_{j=1}^N p_j^{h_j}(\theta), \quad h \in L_{nN}. \quad (2.70)$$

Полученное распределение случайного вектора $\mathbf{v} = (v_1, \dots, v_N)$ называется *полиномиальным* или *мультиномиальным распределением* с параметрами $n = v_1 + \dots + v_N$ и $p_j = p_j(0)$, $j = 1, \dots, N$, и обозначается символом $M(n, \mathbf{p} = (p_1, \dots, p_N))$. Таким образом, с любым методом группировки наблюдений связано соответствующее полиномиальное распределение вектора частот попадания элементов выборки в подмножества группировки.

Если число подмножеств группировки $N=2$, то распределение (2.70) совпадает с биномиальным распределением $Bi(n, p_1)$ числа осуществлений события $\{\xi \in \mathcal{C}_1\}$ в n независимых испытаниях.

Определенную в (2.70) функцию $L^{(1)}(\mathbf{h}; 0)$ можно считать *частичной функцией правдоподобия* в методе группировки наблюдений в отличие от обычной функции правдоподобия $L(\mathbf{x}; 0) = \prod_{i=1}^n f(x_i; 0)$.

Оценку параметра 0 , получаемую из условия максимизации (при фиксированном $\mathbf{h}$) функции $L^{(1)}(\mathbf{h}; 0)$, называют *мультиномиальной оценкой максимального правдоподобия*. Уравнения для отыскания этой оценки $\partial \ln L^{(1)}(\mathbf{h}; 0) / \partial \theta_k = 0$, $k = 1, \dots, r$, совпадают с уравнениями (2.68). Таким образом, оценки по видоизмененному методу минимума χ^2 совпадают с мультиномиальными оценками максимального правдоподобия.

§ 2.6. Интервальное оценивание

В предыдущих параграфах были рассмотрены точечные оценки для параметра 0 исследуемой модели $\mathcal{F} = \{F(\mathbf{x}; 0), 0 \in \Theta\}$ и функций этого параметра. Любая точечная оценка представляет собой функцию $T = T(\mathbf{X})$ выборки $\mathbf{X} = (X_1, \dots, X_n)$, т. е. является случайной величиной, и при каждой реализации $\mathbf{x}$ выборки $\mathbf{X}$ эта функция определяет единственное значение $t = T(\mathbf{x})$ оценки, принимаемое за приближенное значение оцениваемой характеристики. При этом надо принимать во внимание, что в каждом конкретном случае значение оценки может отличаться от значения параметра; следовательно, полезно знать и возможную погрешность, возникающую при использовании предлагаемой оценки. Например, указать такой интервал (или область в случае векторного параметра), внутри которого с высокой вероятностью γ находится точное значение оцениваемого параметра. В этом случае говорят об *интервальном* или *доверительном оценивании*, а соответствующий интервал называют *доверительным*. Величину γ выбирают заранее; она отражает «степень готовности мириться с возможностью ошибки». При заданном γ длина доверительного интервала характеризует точность локализации значения параметра, поэтому желательно выбирать кратчайший интервал. Если этот интервал можно построить, то он обеспечивает наиболее точную (при заданном γ)

локализацию параметра. Подобный подход к задаче оценивания неизвестных параметров распределений и будет рассмотрен в данном параграфе.

1. Понятие доверительного интервала. Рассмотрим доверительное оценивание скалярного параметра. При интервальном оценивании ищут две такие статистики $T_1 = T_1(X)$ и $T_2 = T_2(X)$, что $T_1 < T_2$, для которых при заданном $\gamma \in (0, 1)$ выполняется условие

$$P_\theta(T_1(X) < \theta < T_2(X)) \geq \gamma, \quad \forall \theta \in \Theta. \quad (2.71)$$

В этом случае интервал $(T_1(X), T_2(X))$ называют γ -доверительным интервалом (для θ); число γ — доверительным уровнем или доверительной вероятностью, $T_1(X)$ и $T_2(X)$ — нижней и верхней доверительными границами соответственно. Таким образом, γ -доверительный интервал — это случайный интервал в параметрическом множестве Θ : $(T_1, T_2) \subset \Theta$, зависящий от выборки X (но не от θ), который содержит (накрывает) истинное значение неизвестного параметра θ с вероятностью, не меньшей γ .

Если выполнено условие (2.71), то это практически означает следующее. Пусть проводится большое число не зависящих друг от друга экспериментов, в каждом из которых по n наблюдений над случайной величиной ξ с $\mathcal{L}(\xi) \in \mathcal{F}$. Оценивается параметр θ и используется следующее статистическое правило: если результаты эксперимента описываются выборкой X , то неизвестное значение параметра θ лежит внутри интервала $(T_1(X), T_2(X))$. Тогда это правило иногда может оказаться ошибочным, но число таких случаев при γ , близком к 1, мало и составляет примерно $(1 - \gamma)$ 100% общего числа случаев применения этого правила. На практике обычно используют значения доверительного уровня γ из небольшого набора заранее выбранных, достаточно близких к 1 значений, например $\gamma = 0,9; 0,95; 0,99$ и т. д.

Иногда рассматривают односторонние доверительные интервалы, соответственно верхний (вида $\theta < T_2(X)$) и нижний (вида $T_1(X) < \theta$), определяемые условиями, аналогичными (2.71), в которых опускают соответствующую вторую границу.

Аналогично определяется доверительный интервал для отдельной компоненты (например, θ_1) в случае многомерного параметра θ :

$$P_\theta(T_1(X) < \theta_1 < T_2(X)) \geq \gamma, \quad \forall \theta \in \Theta,$$

а также доверительный интервал для скалярной параметрической функции $\tau(\theta)$ (θ может быть как скаляром, так и вектором):

$$P_\theta(T_1(X) < \tau(\theta) < T_2(X)) \geq \gamma, \quad \forall \theta \in \Theta.$$

2. Построение доверительного интервала с помощью центральной статистики. Общий прием, с помощью которого в ряде случаев можно построить доверительный интервал, состоит в следующем. Пусть модель $\mathcal{F}$ абсолютно непрерывна и существует случайная величина $G(X; \theta)$, зависящая от θ и такая, что:

1) распределение $G(X; \theta)$ не зависит от θ и 2) при каждом $x \in \mathcal{X}$ функция $G(x; \theta)$ непрерывна и строго монотонна по θ — такую случайную величину называют иногда *центральной статистикой* (для θ). Аналогично определяется центральная статистика для отдельной компоненты (например, θ_1) в случае многомерного параметра $\theta = G(X; \theta_1)$, а также для скалярной параметрической функции $\tau = \tau(\theta)$ (θ может быть как скаляром, так и вектором) — $G(X; \tau)$. Далее будем рассматривать для краткости лишь случай скалярного параметра θ .

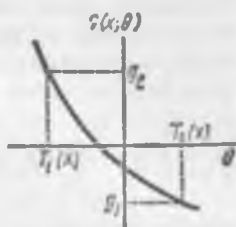


Рис. 2.1

Пусть для модели $\mathcal{F}$ построена центральная статистика $G(X; \theta)$ и $f_\theta(g)$ — ее плотность распределения. Функция $f_\theta(g)$ от параметра θ не зависит [условие 1)], поэтому для любого $\gamma \in (0, 1)$ можно выбрать величины $g_1 < g_2$ (многими способами) так, чтобы

$$P_\theta(g_1 < G(X; \theta) < g_2) = \int_{g_1}^{g_2} f_\theta(g) dg = \gamma. \quad (2.72)$$

Определим теперь при каждом $x \in \mathcal{X}$ числа $T_i(x)$, $i = 1, 2$, где $T_1(x) < T_2(x)$, как решения относительно θ уравнений

$$G(x; \theta) = g_1, \quad g_2 \quad (2.73)$$

[однозначность определения этих чисел обеспечивается условием 2), наложенным на функцию $G(x; \theta)$]. Тогда неравенства $g_1 < G(x; \theta) < g_2$ эквивалентны неравенствам $T_1(x) < \theta < T_2(x)$ (рис. 2.1) и, следовательно, (2.72) можно переписать в виде $P_\theta(T_1(X) < \theta < T_2(X)) = \gamma$, $\forall \theta$. Таким образом, построенный интервал $(T_1(X), T_2(X))$ является γ -доверительным интервалом для θ .

В каждой конкретной задаче при построении центральной статистики для оцениваемой характеристики обычно приходится учитывать специфику рассматриваемой модели (см. следующие ниже примеры), однако можно выделить класс моделей, для которых центральная статистика всегда существует и имеет простой вид. Именно: если функция распределения $F(x; \theta)$ непрерывна и монотонна по параметру θ , то можно положить

$$G(X; \theta) = - \sum_{i=1}^n \ln F(X_i; \theta). \quad (2.74)$$

Действительно, непрерывность и монотонность по θ здесь очевидны, а так как $\mathcal{L}_\theta(F(X; \theta)) = R(0, 1)$ при любом θ , то распределение $G(X; \theta)$ не зависит от θ . Далее, из $\mathcal{L}(\eta) = R(0, 1)$ следует, что $\mathcal{L}(-\ln \eta) = \Gamma(1, 1)$. Таким образом, слагаемые в (2.74) независимы и каждое из них имеет распределение $\Gamma(1, 1)$. По воспроизводящему свойству гамма-распределения окончательно получаем, что плотность распределения статистики $G(X; \theta)$ есть

плотность распределения $\Gamma(1, n)$, т. е. (см. табл. В.1) $f_0(g) = g^{n-1} e^{-g} / \Gamma(n)$, $g > 0$. Отсюда и из (2.72) получаем следующий метод построения доверительного интервала для θ : при заданном γ выбираем числа $g_1 < g_2$ так, чтобы

$$\frac{1}{\Gamma(n)} \int_{g_1}^{g_2} g^{n-1} e^{-g} dg = \gamma.$$

Решая уравнения

$$-\sum_{i=1}^n \ln F(x_i; \theta) = g_1, g_2, \quad (2.75)$$

находим корни $T_1(x) < T_2(x)$. Тогда $(T_1(X), T_2(X))$ — искомый γ -доверительный интервал для θ . Очевидно, наибольшая трудность в применении этой методики к конкретным задачам возникает при нахождении решений уравнений (2.75).

Рассмотрим несколько важных примеров использования центральных статистик при построении доверительных интервалов для параметров нормальных моделей.

Пример 2.28 (*нормальная-1 модель, доверительный интервал для среднего*). Пусть по выборке $X = (X_1, \dots, X_n)$ требуется построить доверительный интервал для неизвестного среднего θ в модели $\mathcal{N}(\theta, \sigma^2)$. Известно (см. теорему 1.10), что $\mathcal{L}_\theta(\sqrt{n}(X - \theta)/\sigma) = \mathcal{N}(0, 1)$. Следовательно, в данном случае центральная статистика

$$G(X; \theta) = \sqrt{n} \frac{X - \theta}{\sigma}.$$

Решения уравнений (2.73) имеют вид $T_1(x) = x - \frac{\sigma}{\sqrt{n}} g_2$, $T_2(x) = x - \frac{\sigma}{\sqrt{n}} g_1$, поэтому γ -доверительным для θ является любой интервал

$$\Delta_\gamma(X) = \left(X - \frac{\sigma}{\sqrt{n}} g_2, X - \frac{\sigma}{\sqrt{n}} g_1 \right), \quad (2.76)$$

где $g_1 < g_2$ — любые числа, удовлетворяющие условию

$$\Phi(g_2) - \Phi(g_1) = \gamma. \quad (2.77)$$

Отметим, что, хотя интервал $\Delta_\gamma(X)$ случаен, его длина постоянна и равна $l_\gamma(g_1, g_2) = \sigma(g_2 - g_1)/\sqrt{n}$, поэтому, чтобы среди всех интервалов вида (2.76) выбрать кратчайший, надо минимизировать функцию $l_\gamma(g_1, g_2)$ при условии (2.77). Применяя метод Лагранжа нахождения условного экстремума, получаем следующую систему уравнений:

$$\lambda \varphi(g_1) = \sigma/\sqrt{n}, \quad \lambda \varphi(g_2) = \sigma/\sqrt{n}, \quad \Phi(g_2) - \Phi(g_1) = \gamma,$$

где $\varphi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Отсюда находим, что $\varphi(g_1) = \varphi(g_2)$.

Так как функция $\varphi(x)$ — чётная, то $g_1 = -g_2$. Учитывая это,

а также последнее уравнение и соотношение $\Phi(-x) = 1 - \Phi(x)$, получаем равенство $\Phi(g_2) = (1 + \gamma)/2$, из которого находим, что $g_2 = c_\gamma = \Phi^{-1}((1 + \gamma)/2) = (1 + \gamma)/2$ -квантиль стандартного нормального распределения $\varphi^*(0, 1)$.

Итак, оптимальным [среди интервалов $\Delta_\gamma(X)$] γ -доверительным интервалом для параметра θ в модели $\varphi^*(\theta, \sigma^2)$ является интервал

$$\Delta_\gamma^*(X) = \left(X - \frac{\sigma}{\sqrt{n}} c_\gamma, X + \frac{\sigma}{\sqrt{n}} c_\gamma \right), \quad c_\gamma = \Phi^{-1} \left(\frac{1 + \gamma}{2} \right), \quad (2.78)$$

т. е. симметричный относительно случайной точки X интервал длины $2\sigma c_\gamma / \sqrt{n}$. В частности, для доверительного уровня $\gamma = 0,99$ величина $c_\gamma = 2,5758$; $c_\gamma = 3,0902$ для $\gamma = 0,998$. Симметрия оптимального интервала отражает тот факт, что распределение центральной статистики в данном случае оказалось симметричным относительно оцениваемого параметра.

Пример 2.29 (нормальная-2 модель, доверительный интервал для дисперсии). Построим доверительный интервал для неизвестной дисперсии θ^2 в модели $\varphi^*(\mu, \theta^2)$. В этом случае также легко найти центральную статистику, зависящую от $\tau = \tau(\theta) = \theta^2$:

$$G(X; \tau) = \frac{1}{\tau} \sum_{i=1}^n (X_i - \mu)^2.$$

Действительно, так как $\mathcal{L}_\theta \left(\frac{X_i - \mu}{\theta} \right) = \varphi^*(0, 1)$, то $\mathcal{L}_\theta \left(\frac{(X_i - \mu)^2}{\theta^2} \right) = \chi^2(1)$ (см. п. 1 § 1.5). Следовательно, $\mathcal{L}_\theta(G(X; \tau)) = \chi^2(n)$. Здесь

$$T_1(x) = \frac{1}{g_1} \sum_{i=1}^n (x_i - \mu)^2, \quad T_2(x) = \frac{1}{g_2} \sum_{i=1}^n (x_i - \mu)^2$$

— решения (относительно τ) уравнений $G(x; \tau) = g_1, g_2$. Следовательно, γ -доверительным для $\tau = \theta^2$ является в данном случае любой интервал

$$\Delta_\gamma(X) = \left(\frac{1}{g_1} \sum_{i=1}^n (X_i - \mu)^2, \frac{1}{g_2} \sum_{i=1}^n (X_i - \mu)^2 \right), \quad (2.79)$$

где $g_1 < g_2$ находят из условия $\int_{g_1}^{g_2} k_n(x) dx = \gamma$ [плотность $k_n(x)$ распределения $\chi^2(n)$ определяется формулой (1.28)]. Как правило, g_1 и g_2 следует выбирать так, чтобы выполнялись равенства

$$\int_0^{g_1} k_n(x) dx = \frac{1 - \gamma}{2}, \quad \int_{g_2}^{\infty} k_n(x) dx = \frac{1 - \gamma}{2}, \quad (2.80)$$

т. е. $g_1 = \chi_{(1 - \gamma)/2, n}^2$, $g_2 = \chi_{\gamma/2, n}^2$, где $\chi_{p, n}^2$ — p -квантиль распределения $\chi^2(n)$. В этом случае соответствующий доверительный

интервал называют иногда *центральный*. Соотношения (2.79)–(2.80) и определяют рекомендуемое для практического использования правило доверительного оценивания неизвестной дисперсии в модели $\mathcal{N}(\mu, \theta^2)$.

Задача отыскания наикратчайшего интервала среди интервалов вида (2.79) сводится к минимизации отношения g_2/g_1 при условии

$\int_{g_1}^{g_2} k_n(x) dx = \gamma$. Метод неопределенных множителей Лагранжа приводит в данном случае к уравнениям $g_1 k_n(g_1) = g_2 k_n(g_2)$, $\int_{g_1}^{g_2} k_n(x) dx = \gamma$ или, если положить $g_1 = \chi_{\alpha_1, n}^2$, $g_2 = \chi_{1-\alpha_2, n}^2$, где $\alpha_1 + \alpha_2 = 1 - \gamma$, к уравнению

$$(\chi_{1-\alpha_2, n}^2 / \chi_{\alpha_1, n}^2)^{n/2} = \exp \{(\chi_{1-\alpha_2, n}^2 - \chi_{\alpha_1, n}^2)/2\}. \quad (2.80')$$

Значения α_1 и α_2 , удовлетворяющие (2.80'), определяют оптимальный γ -доверительный интервал вида (2.79):

$$\Delta_\gamma^*(X) = \left(\frac{1}{\chi_{\alpha_1, n}^2} \sum_{i=1}^n (X_i - \mu)^2, \frac{1}{\chi_{1-\alpha_2, n}^2} \sum_{i=1}^n (X_i - \mu)^2 \right). \quad (2.79')$$

Таким образом, центральный интервал в данном случае не является наикратчайшим. В табл. 2.3 [10, с. 275] приведены значения $(\chi_{\alpha_1, n}^2; \chi_{1-\alpha_2, n}^2)$ и $(\chi_{(1-\gamma)/2, n}^2; \chi_{(1+\gamma)/2, n}^2)$ для $\gamma = 0,95$ и некоторых значений n .

Таблица 2.3

n	$(\chi_{\alpha_1, n}^2; \chi_{1-\alpha_2, n}^2)$	$(\chi_{(1-\gamma)/2, n}^2; \chi_{(1+\gamma)/2, n}^2)$	$(\chi_{\alpha_1, n}^2 - \chi_{(1-\gamma)/2, n}^2; \chi_{1-\alpha_2, n}^2 - \chi_{(1+\gamma)/2, n}^2)$
2	(0,08; 9,53)	(0,05; 7,38)	(0,03; 2,15)
5	(0,99; 14,37)	(0,83; 12,83)	(0,16; 1,54)
10	(3,52; 21,73)	(3,25; 20,48)	(0,27; 1,25)
20	(9,96; 35,23)	(9,59; 34,17)	(0,37; 1,06)

Заметим, что центральной статистикой для оценивания среднего квадратического отклонения θ является, очевидно,

$$G(X; 0) = \frac{1}{n} \left[\sum_{i=1}^n (X_i - \mu)^2 \right]^{1/2}.$$

Пример 2.30 (общая нормальная модель, доверительные интервалы для среднего и дисперсии). Рассмотрим доверительное оценивание параметров в общей нормальной модели $\mathcal{N}(\theta_1, \theta_2^2)$. Из теоремы 1.10 следует, что $G(X; \theta_2^2) = nS^2(X)/\theta_2^2$ — центральная статистика для оценивания дисперсии θ_2^2 . Здесь $S^2(X)$ — выборочная дисперсия. Доверительный интервал для θ_2^2 находим по схеме

предыдущего примера; окончательно получаем: центральным γ -доверительным интервалом для θ_1^2 является интервал

$$(nS^2(X)/\chi_{1-\gamma/2}^2, n-1, nS^2(X)/\chi_{\gamma/2}^2, n-1). \quad (2.81)$$

В частности, для выборки объема $n=10$ и доверительного уровня $\gamma=0,9$ имеем $\chi_{0,95;9}^2=3,3251$, $\chi_{0,05;9}^2=16,9190$, поэтому центральный 0,9-доверительный интервал для θ_1^2 имеет вид $(0,5911 S^2(X), 3,007 S^2(X))$. Нанкратчайший в данном случае интервал

$$(nS^2(X)/\chi_{1-\alpha_2}^2, n-1, nS^2(X)/\chi_{\alpha_1}^2, n-1), \quad \alpha_1 + \alpha_2 = 1 - \gamma, \quad (2.81')$$

где $\chi_{\alpha_1, n-1}^2$ и $\chi_{1-\alpha_2, n-1}^2$ определяются соотношением (2.80'), в котором n заменено на $n-1$.

В силу теоремы 1.11 центральной статистикой для оценивания среднего θ_1 является

$$G(X; \theta_1) = \sqrt{\frac{n-1}{n}} \frac{\bar{X} - \theta_1}{S(X)},$$

причем распределение этой статистики (распределение Стьюдента $S(n-1)$) симметрично относительно своей средней точки. Расчет доверительного интервала проводится так же, как в примере 2.28. Окончательно получаем: γ -доверительным для θ_1 является интервал

$$\left(\bar{X} - \frac{S(X)}{\sqrt{n-1}} t_{\gamma, n-1}, \bar{X} + \frac{S(X)}{\sqrt{n-1}} t_{\gamma, n-1} \right), \quad (2.82)$$

где $t_{\gamma, n-1} = (1+\gamma)/2$ -квантиль распределения $S(n-1)$. Построенный интервал имеет минимальную длину среди всех γ -доверительных интервалов вида $(\bar{X} - a_1 S(X), \bar{X} + a_2 S(X))$. В частности, $t_{0,95;9} = 2,262$, поэтому для выборки объема $n=10$ и доверительного уровня $\gamma=0,95$ интервал (2.82) имеет вид $(\bar{X} - 0,754 S(X), \bar{X} + 0,754 S(X))$.

Методику, основанную на использовании центральных статистик, можно применять и в других случаях. Приведем два практически важных соответствующих примера, относящихся к проблеме сравнения двух нормальных выборок.

Пример 2.31 (доверительный интервал для разности средних двух нормальных моделей). Пусть $X=(X_1, \dots, X_n)$ и $Y=(Y_1, \dots, Y_m)$ — две независимые выборки, причем первая — из распределения $\mathcal{N}(\theta_1^{(1)}, \theta_1^{(2)})$, а вторая — из распределения $\mathcal{N}(\theta_1^{(1)}, \theta_1^{(2)})$. Таким образом, соответствующие модели отличаются только своими средними и требуется оценить их разность $\theta_1^{(1)} - \theta_1^{(2)}$. Эту задачу называют *задачей сравнения двух средних*. Согласно замечанию к теореме 1.12, отношение Стьюдента

$$t = \sqrt{\frac{mn(m+n-2)}{m+n}} \frac{\bar{X} - \bar{Y} - (\theta_1^{(1)} - \theta_1^{(2)})}{[nS^2(X) + mS^2(Y)]^{1/2}}$$

имеет распределение $S(m+n-2)$ при любых значениях параметров $\theta_1^{(1)}, \theta_1^{(2)}, \theta_1^{(2)}$, поэтому величину t можно использовать в качестве центральной статистики для оценивания разности $\theta_1^{(1)} - \theta_1^{(2)}$. Соответствующий γ -доверительный интервал строится так же, как

и интервал (2.82) в примере 2.30, и имеет вид

$$(X - Y \pm t_{\gamma, m+n-2} \left[\frac{m+n}{mn(m+n-2)} (nS^2(X) + mS^2(Y)) \right]^{1/2}).$$

Если дисперсии моделей известны и равны соответственно σ_1^2 и σ_2^2 , то центральную статистику для оценивания разности средних $\theta^{(1)} - \theta^{(2)}$ можно легко построить, рассуждая следующим образом. Выборочные средние $\bar{X}$ и $\bar{Y}$ независимы и нормальны $N(\theta^{(1)}, \sigma_1^2/n)$ и $N(\theta^{(2)}, \sigma_2^2/m)$ соответственно, поэтому их разность $\bar{X} - \bar{Y}$ также нормальна $N(\theta^{(1)} - \theta^{(2)}, \sigma_1^2/n + \sigma_2^2/m)$. Следовательно, случайная величина $[\bar{X} - \bar{Y} - (\theta^{(1)} - \theta^{(2)})] / \sqrt{\sigma_1^2/n + \sigma_2^2/m}$ имеет распределение $N(0, 1)$ при любых значениях параметров $\theta^{(1)}$ и $\theta^{(2)}$ и поэтому является центральной статистикой для данной задачи. Расчет γ -доверительного интервала проводится здесь аналогично примеру 2.28, и искомый интервал для $\theta^{(1)} - \theta^{(2)}$ имеет вид

$$(\bar{X} - \bar{Y} \pm c_\gamma \sqrt{\sigma_1^2/n + \sigma_2^2/m}), \quad c_\gamma = \Phi^{-1}((1 + \gamma)/2).$$

Пример 2.32 (доверительный интервал для отношения дисперсий двух нормальных моделей). Рассмотрим теперь случай, когда $X = (X_1, \dots, X_n)$ — выборка из распределения $N(\theta_1^{(1)}, \theta_1^{(2)})$, а $Y = (Y_1, \dots, Y_m)$ — из распределения $N(\theta_2^{(1)}, \theta_2^{(2)})$ и выборки независимы. Требуется оценить отношение $\theta_1^{(2)}/\theta_2^{(2)}$ неизвестных дисперсий. Центральную статистику в этой задаче можно построить на основании теоремы 1.13, согласно которой случайная величина

$$F = \frac{n(m-1)S^2(X)}{m(n-1)S^2(Y)} \bigg/ \frac{\theta_1^{(2)}}{\theta_2^{(2)}}$$

имеет распределение Снедекора $S(n-1, m-1)$ при любых значениях параметров $\theta_1^{(1)}, \theta_1^{(2)}, \theta_2^{(1)}, \theta_2^{(2)}$. Следовательно, F является искомой центральной статистикой для оценивания отношения $\theta_1^{(2)}/\theta_2^{(2)}$. Соответствующий γ -доверительный интервал строится так же, как и интервал (2.81) в примере 2.30, и имеет вид

$$\left(\frac{n(m-1)S^2(X)}{m(n-1)S^2(Y)} \bigg/ F_{\frac{1+\gamma}{2}, n-1, m-1}, \frac{n(m-1)S^2(X)}{m(n-1)S^2(Y)} \bigg/ F_{\frac{1-\gamma}{2}, n-1, m-1} \right),$$

где $F_{p, n-1, m-1}$ — p -квантиль распределения $S(n-1, m-1)$. В частности, $F_{0.95, 8, 8} = F_{0.05, 8, 8} = 3.44$, $F_{0.95, 12, 12} = F_{0.05, 12, 12} = 2.69$, поэтому 0,9-доверительный интервал для выборок объемов $n = m = 9$ есть $(0,58S^2(X)/S^2(Y), 3,44S^2(X)/S^2(Y))$, а для выборок объемов $n = m = 13$ — $(0,37S^2(X)/S^2(Y), 2,69S^2(X)/S^2(Y))$.

Пример 2.33 (доверительный интервал для параметра равномерной-1 модели). Построим по выборке $X = (X_1, \dots, X_n)$ из равномерного распределения $R(0, \theta)$ доверительный интервал для параметра θ . Здесь $F(x; \theta) = x/\theta$, $0 \leq x \leq \theta$, т. е. функция распределения непрерывна и монотонна по θ при $\theta > 0$; следовательно, можно использовать, например, центральную статистику (2.74),

которая в данном случае принимает вид

$$G(X; \theta) = - \sum_{i=1}^n \ln X_i + n \ln \theta.$$

Решениями уравнений (2.75) являются числа

$$T_j(x) = e^{g_j/n} \left(\prod_{i=1}^n x_i \right)^{1/n}, \quad j=1, 2,$$

поэтому γ -доверительным для θ в данном случае будет любой интервал $(T_1(X), T_2(X))$, в котором $g_1 < g_2$ удовлетворяют усло-

вию $\frac{1}{\Gamma(n)} \int_{g_1}^{g_2} g^{n-1} e^{-g} dg = \gamma$. Кратчайшим среди всех таких интервалов при $n=1$ является интервал $(X_1, X_1/(1-\gamma))$, а при $n > 1$ — интервал

$$\left(e^{g_1^*/n} \left(\prod_{i=1}^n X_i \right)^{1/n}, e^{g_2^*/n} \left(\prod_{i=1}^n X_i \right)^{1/n} \right),$$

где $g_1^* < g_2^*$ однозначно определяются из соотношений

$$\frac{1}{g_1^*} e^{\frac{n+1}{(n-1)} g_1^*} = \frac{1}{g_2^*} e^{\frac{n+1}{(n-1)} g_2^*}, \quad \frac{1}{\Gamma(n)} \int_{g_1^*}^{g_2^*} g^{n-1} e^{-g} dg = \gamma.$$

В данной задаче можно указать и другую центральную статистику. Именно, из результата примера 2.15 следует, что

$$\mathcal{L}_\theta((X_{(n)}/\theta)^n) = R(0, 1), \quad \forall \theta > 0.$$

Таким образом, величину $(X_{(n)}/\theta)^n$ также можно использовать в качестве центральной статистики для оценивания θ . Согласно общей методике [см. соотношения (2.72) — (2.73)], γ -доверительным для θ является любой интервал вида $(X_{(n)}/\sqrt[n]{g}, X_{(n)}/\sqrt[n]{g-\gamma})$, $\gamma < g \leq 1$. Среди всех таких интервалов наименьшую длину имеет, очевидно, интервал, соответствующий значению $g=1$, т. е. интервал $(X_{(n)}, X_{(n)}/\sqrt[n]{1-\gamma})$.

При $n=1$ оба метода приводят к одному результату, при $n > 1$ однозначно ответить на вопрос о том, какой из этих двух методов приводит к более точной локализации параметра, трудно.

3. Построение доверительного интервала с использованием распределения точечной оценки параметра. Если уже имеется некоторая точечная оценка $T=T(X)$ для параметра θ и известна ее функция распределения $F_T(t; \theta)$, то доверительный интервал для θ можно построить, основываясь на этой функции. Предположим, что распределение оценки T непрерывно и функция $F_T(t; \theta)$ непрерывна и монотонна по θ . Пусть задан доверитель-

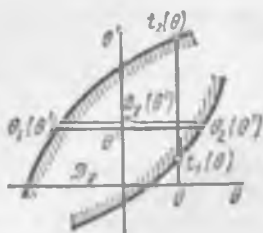


Рис. 2.2

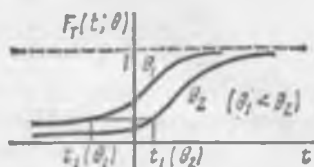


Рис. 2.3

ный уровень γ . Определим при каждом $\theta \in \Theta$ числа $t_i = t_i(\theta)$, $i = 1, 2$, где $t_1 < t_2$ и

$$P_\theta(t_1 < T(X) < t_2) = F_T(t_2; \theta) - F_T(t_1; \theta) = \gamma. \quad (2.83)$$

Чтобы данная процедура была однозначной, часто рекомендуется выбирать эти числа так, чтобы выполнялись условия

$$F_T(t_1; \theta) = (1 - \gamma)/2, \quad 1 - F_T(t_2; \theta) = (1 - \gamma)/2, \quad (2.84)$$

т. е. речь идет о построении *центрального* доверительного интервала. Обозначим через $\mathcal{D}_\gamma$ подмножество $\Theta \times \Theta$: $\mathcal{D}_\gamma = \{(\theta, \theta') : t_1(\theta) < \theta' < t_2(\theta)\}$ (рис. 2.2). Тогда $P_\theta((\theta, T(X)) \in \mathcal{D}_\gamma) = \gamma$ при $\forall \theta \in \Theta$. Определим теперь при фиксированном θ' сечение $\mathcal{D}_\gamma(\theta')$ множества $\mathcal{D}_\gamma$: $\mathcal{D}_\gamma(\theta') = \{\theta : (\theta, \theta') \in \mathcal{D}_\gamma\}$ — и рассмотрим случайное множество $\mathcal{D}_\gamma(T(X)) \subset \Theta$. Событие $\theta \in \mathcal{D}_\gamma(T(X))$ происходит тогда и только тогда, когда $T(X) \in (t_1(\theta), t_2(\theta))$ и, следовательно, при каждом θ имеет вероятность γ . Таким образом, построено случайное множество $\mathcal{D}_\gamma(T(X))$, которое накрывает истинное значение параметра θ с вероятностью γ . Если это множество является интервалом, то тем самым построен γ -доверительный интервал для θ . Это имеет место, если кривые $\theta' = t_i(\theta)$, $i = 1, 2$, являются монотонными одного типа (т. е. одновременно либо возрастают, либо убывают), что обеспечивается условием непрерывности и монотонности функции $F_T(t; \theta)$ по θ (рис. 2.3). Таким образом, при сделанных предположениях множество $\mathcal{D}_\gamma(\theta')$ при каждом θ' представляет собой интервал; следовательно, определены его концы $\theta_1(\theta') < \theta_2(\theta')$, а тем самым и соответствующий γ -доверительный интервал $(T_1(X), T_2(X))$ для θ , где $T_i(X) = \theta_i(T(X))$, $i = 1, 2$.

Построение диаграммы придает наглядный смысл этой методике. Диаграмму строят *по вертикали* (для любой абсциссы θ находят из условия (2.84) соответствующие ординаты t_1 и t_2); «читают» же ее *по горизонтали*, т. е. для наблюдавшейся ординаты $t = T(x)$ (x — реализация выборки X) «считывают» две величины θ_1 и θ_2 и утверждают, что $\theta \in (\theta_1, \theta_2)$. Таким образом, если «читать» диаграмму по вертикали, то границы области $\mathcal{D}_\gamma$ описываются парой переменных точек (θ, t_1) и (θ, t_2) таких, что выполняется условие (2.84). Если же «читать» диаграмму по горизонтали, то границы $\mathcal{D}_\gamma$ можно эквивалентно описать парой переменных точек

(t, θ_1) и (t, θ_2) , взяв за независимую переменную наблюдаемое значение t оценки $T(X)$. Тогда соотношения (2.84) можно переписать в терминах этих величин так: если функции $t_i(\theta)$ возрастают (см. рис. 2.2), что соответствует убыванию по θ функции $F_T(t; \theta)$, то $1 - F_T(t; \theta_1) = (1 - \gamma)/2$, $F_T(t; \theta_2) = (1 - \gamma)/2$, если же функции $t_i(\theta)$ убывают, что соответствует возрастанию по θ функции $F_T(t; \theta)$, то $F_T(t; \theta_1) = (1 - \gamma)/2$, $1 - F_T(t; \theta_2) = (1 - \gamma)/2$.

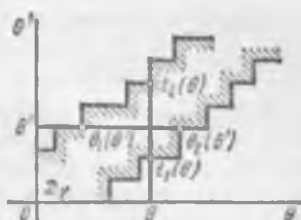


Рис. 2.4

Итак, алгоритм построения центрального γ -доверительного интервала для θ в случае, когда функция распределения $F_T(t; \theta)$ оценки $T = T(X)$ непрерывна и монотонна по θ , состоит в следующем. Пусть $t = T(x)$ — наблюдавшееся значение оценки. Решая относительно θ уравнения

$$F_T(t; \theta) = (1 - \gamma)/2, \quad (1 + \gamma)/2, \quad (2.85)$$

найдем два числа $\theta_1 < \theta_2$. Далее утверждаем, что $\theta \in (\theta_1, \theta_2)$. Приведенная теория гарантирует, что при γ , близком к 1, вероятность ошибки в случае применения этого правила равна $1 - \gamma$.

Аналогичные рассуждения можно провести и для случая дискретной модели. Единственное отличие состоит в том, что из-за ступенчатости функции распределения $F_T(t; \theta)$ в соотношении (2.83) можно, вообще говоря, обеспечить лишь выполнение неравенства

$$P_0(t_1 < T(X) < t_2) = F_T(t_2 - 0; \theta) - F_T(t_1; \theta) \geq \gamma, \quad (2.83')$$

поэтому вместо условий (2.84) следует ввести условия

$$F_T(t_1; \theta) \leq (1 - \gamma)/2, \quad 1 - F_T(t_2 - 0; \theta) \leq (1 - \gamma)/2, \quad (2.84')$$

где t_1 — наибольшее, а t_2 — наименьшее значения T , удовлетворяющие этим неравенствам.

Кривые $\theta' = t_i(\theta)$, $i = 1, 2$, в данном случае являются ступенчатыми (рис. 2.4) и при «считывании» величин $\theta_i(\theta')$, $i = 1, 2$, следует брать крайнюю правую точку отрезка пересечения горизонтальной прямой с левой кривой и крайнюю левую точку — с правой кривой (когда линия пересечения попадает на горизонтальные отрезки «ступенек» кривых). Алгоритм построения центрального γ -доверительного интервала для θ в дискретном случае в основном тот же, что и для непрерывного, только вместо уравнений (2.85) надо решать относительно θ уравнения

$$F_T(t; \theta) = (1 - \gamma)/2, \quad 1 - F_T(t - 0; \theta) = (1 - \gamma)/2, \quad (2.86)$$

где t — наблюдавшееся значение оценки T .

Замечание. Часто вместо центрального доверительного интервала в дискретном случае рекомендуется строить множество $\mathcal{M}_\gamma$ по принципу включения в него значений статистики T с наибольшими (при каждом θ) вероятностями их появления. В случае применения этого метода вместо условий (2.83') — (2.84') границы $t_i = t_i(\theta)$, $i = 1, 2$, определяют из условия

$$P_0(t_1 < T < t_2) = \sum_{t_1 < t < t_2} P_0(T = t) \geq \gamma.$$

При этом требуют, чтобы в сумму были включены значения t с наибольшими вероятностями их появления и чтобы интервал (t_1, t_2) был наименьшим, удов-

летворяющим данному условию. В ряде случаев, когда распределение статистики T асимметрично, такой способ позволяет получить более короткие доверительные интервалы.

Отметим, что, выбирая различные оценки $T(X)$, будем получать различные доверительные интервалы. Но конечная цель — это получить как можно более короткие интервалы (при фиксированном доверительном уровне γ). Предположим, что используются несмещенные и приблизительно нормальные оценки; тогда введенные выше интервалы тем короче, чем меньше дисперсия оценки. Таким образом, эффективные и асимптотически эффективные оценки приводят к кратчайшим или асимптотически кратчайшим интервалам. Выше было показано (см. п. 4 § 2.4), что при весьма общих условиях оценки максимального правдоподобия $\hat{\theta}_n$ являются асимптотически эффективными и асимптотически нормальными и при этом [см. соотношение (2.65)] при $n \rightarrow \infty$

$$P_{\theta}(|\theta_n - \theta| \sqrt{ni(\hat{\theta}_n)} \leq c_{\gamma}) \rightarrow \Phi(c_{\gamma}) - \Phi(-c_{\gamma}) = 2\Phi(c_{\gamma}) - 1 = \gamma, \quad (2.87)$$

если $c_{\gamma} = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$. Отсюда следует, что $(\hat{\theta}_n - c_{\gamma}/\sqrt{ni(\hat{\theta}_n)}, \hat{\theta}_n + c_{\gamma}/\sqrt{ni(\hat{\theta}_n)})$ — асимптотический кратчайший γ -доверительный интервал для параметра θ .

Пронлюстрируем эти общие принципы примерами.

Пример 2.34 (доверительный интервал для параметра бернуллиевской модели). Пусть требуется построить доверительный интервал для параметра θ модели $Bi(1, \theta)$. Если имеется соответствующая выборка $X = (X_1, \dots, X_n)$, то оптимальной оценкой для θ является статистика $T = X$. Случайная величина T принимает значения $0, 1/n, 2/n, \dots, n/n$, при этом

$$F_T\left(\frac{k}{n}; \theta\right) = \sum_{r=0}^k C_n^r \theta^r (1-\theta)^{n-r}.$$

Здесь $\frac{d}{d\theta} F_T\left(\frac{k}{n}; \theta\right) = -nC_{n-1}^k \theta^k (1-\theta)^{n-k-1} < 0$, $k < n$, так что функция $F_T(k/n; \theta)$ монотонно убывает по θ при $k < n$. Следовательно, применима описанная выше методика, согласно которой при $T = k/n$ (θ_1, θ_2) — центральный γ -доверительный интервал для θ , где θ_1 и θ_2 определяются в соответствии с уравнениями (2.86), которые в данном случае принимают вид

$$1 - F_T\left(\frac{k-1}{n}; \theta_1\right) = \sum_{r=k}^n C_n^r \theta_1^r (1-\theta_1)^{n-r} = \frac{1-\gamma}{2},$$

$$F_T\left(\frac{k}{n}; \theta_2\right) = \sum_{r=0}^k C_n^r \theta_2^r (1-\theta_2)^{n-r} = \frac{1-\gamma}{2}.$$

Доверительные интервалы (θ_1, θ_2) рассчитаны для широкого диапазона значений n и $\gamma = 0,9; 0,95; 0,99$ [2].

Если число наблюдений n велико, то для быстрого нахождения приближенного доверительного интервала для θ можно воспользоваться асимптотической теорией. В рассматриваемом случае о. м. п. $\hat{\theta}_n = \bar{X}$ и функция информации $i(\theta) = 1/[\theta(1-\theta)]$ (см. табл. 2.1), поэтому из соотношения (2.87) имеем, что

$$\left(\bar{X} - \frac{c_\gamma}{\sqrt{n}} \sqrt{\bar{X}(1-\bar{X})}, \bar{X} + \frac{c_\gamma}{\sqrt{n}} \sqrt{\bar{X}(1-\bar{X})} \right), \quad c_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right).$$

— асимптотический γ -доверительный интервал для θ .

Пример 2.35 (доверительный интервал для параметра пуассоновской модели). Рассмотрим теперь оценивание параметра θ модели $\Pi(\theta)$. Оптимальной оценкой для θ здесь также является статистика $T = \bar{X}$, принимающая значения k/n , $k = 0, 1, 2, \dots$

Так как $\mathcal{L}_\theta\left(\sum_{i=1}^n X_i\right) = \Pi(n\theta)$, то

$$F_T\left(\frac{k}{n}; \theta\right) = \sum_{r=0}^k e^{-n\theta} \frac{(n\theta)^r}{r!}.$$

Здесь $\frac{d}{d\theta} F_T\left(\frac{k}{n}; \theta\right) = -ne^{-n\theta} \frac{(n\theta)^k}{k!} < 0$, т. е. функция $F_T(k/n; \theta)$ монотонно убывает по θ . Согласно изложенной выше методике, если $T = k/n$, то центральным γ -доверительным интервалом для θ является интервал (θ_1, θ_2) , где θ_1 и θ_2 определяются соответственно уравнениями

$$1 - F_T\left(\frac{k-1}{n}; \theta_1\right) = \sum_{r=k}^{\infty} e^{-n\theta_1} \frac{(n\theta_1)^r}{r!} = \frac{1-\gamma}{2},$$

$$F_T\left(\frac{k}{n}; \theta_2\right) = \sum_{r=0}^k e^{-n\theta_2} \frac{(n\theta_2)^r}{r!} = \frac{1-\gamma}{2}.$$

Как и в предыдущем примере, для больших значений объема выборки можно просто рассчитать приближенный доверительный интервал. В рассматриваемом случае о. м. п. $\hat{\theta}_n = \bar{X}$ и $i(\theta) = 1/[\theta(1-\theta)]$, поэтому из соотношения (2.87) имеем, что $(\bar{X} \pm c_\gamma \sqrt{\bar{X}/n})$ — асимптотический γ -доверительный интервал для θ .

Пример 2.36 (доверительный интервал для параметра модели степенного ряда). Построим асимптотический γ -доверительный интервал для параметра θ распределения типа степенного ряда (см. пример 2.12). Из результатов примера 2.24 и соотношения (2.87) следует, что искомый интервал имеет вид

$$\left(\hat{\theta}_n \pm c_\gamma \sqrt{\frac{\hat{\theta}_n}{\mu'(\hat{\theta}_n)}} = \left(\hat{\theta}_n \pm c_\gamma \frac{\hat{\theta}_n}{\sigma(\hat{\theta}_n) \sqrt{n}} \right),$$

где $\hat{\theta}_n$ — решение уравнения $\mu(\theta) = \bar{X}$, а $\mu(\theta)$ и $\sigma^2(\theta)$ — теоретические среднее и дисперсия соответственно. Построенный в преды-

дущем примере приближенный интервал — частный случай этого решения.

4. **Доверительные области для многомерного параметра.** До сих пор рассматривалось доверительное оценивание скалярного параметра и речь шла о построении такого случайного интервала, который с заданной вероятностью накрывает истинное значение неизвестного параметра. Если оценивается векторный параметр $\theta = (\theta_1, \dots, \theta_r)$, то вместо доверительных интервалов определяют *доверительные области* (или *зоны*) $\mathfrak{J} = \mathfrak{J}(\mathbf{X}) \subset \Theta$ с помощью условия, аналогичного (2.71):

$$P_\theta(\theta \in \mathfrak{J}(\mathbf{X})) \geq \gamma, \quad \forall \theta \in \Theta. \quad (2.88)$$

Таким образом, γ -доверительная область — это такое случайное подмножество в параметрическом множестве Θ , зависящее от выборки $\mathbf{X}$ (но не от θ), которое накрывает истинное значение неизвестного параметра θ с вероятностью, не меньшей γ .

Общий метод построения доверительной области состоит в следующем. Сопоставим каждому значению $\theta \in \Theta$ такое подмножество $H(\theta)$ выборочного пространства $\mathcal{X}$, $H(\theta) \subset \mathcal{X}$, чтобы выполнялось условие

$$P_\theta(\mathbf{X} \in H(\theta)) \geq \gamma. \quad (2.89)$$

Таким образом, получаем в выборочном пространстве семейство подмножеств $\{H(\theta), \theta \in \Theta\}$. Определим теперь для каждого $x \in \mathcal{X}$ подмножество $\mathfrak{J}(x) \subset \Theta$ по следующему правилу: $\mathfrak{J}(x) = \{\theta: x \in H(\theta)\}$. Таким образом, в параметрическом множестве Θ получаем семейство подмножеств $\{\mathfrak{J}(x), x \in \mathcal{X}\}$. Рассмотрим случайное подмножество $\mathfrak{J}(\mathbf{X})$. События $\{\mathbf{X} \in H(\theta)\}$ и $\{\theta \in \mathfrak{J}(\mathbf{X})\}$ — эквивалентные, так как по построению каждое из них влечет за собой другое, поэтому их вероятности при каждом θ совпадают. Учитывая (2.89), получаем, что для $\mathfrak{J}(\mathbf{X})$ выполняются неравенства (2.88); следовательно, $\mathfrak{J}(\mathbf{X})$ — искомая γ -доверительная область для параметра θ .

Отметим, что в этой методике доверительная область строится неоднозначно (так как при заданном доверительном уровне γ множества $H(\theta)$ можно выбрать различными способами) и задача состоит в построении доверительной области минимальных размеров, обеспечивающей наиболее точную (при заданном γ) локализацию параметра. В конкретных задачах множества $H(\theta)$, удовлетворяющие условию (2.89), строят обычно с помощью некоторой статистики $T(\mathbf{X})$ (вообще говоря, векторной), распределение которой известно. Проиллюстрируем изложенный метод на следующем примере.

Пример 2.37 (доверительная область для параметра общей нормальной модели). В примере 2.30 были построены доверительные интервалы отдельно для неизвестных среднего θ_1 и дисперсии $\tau = \theta_2^2$ нормальной модели $\mathcal{N}(\theta_1, \theta_2^2)$. Было бы неверным считать, что двумерная точка $(\theta_1, \tau = \theta_2^2)$ с вероятностью γ лежит в прямо-

угольнике [см. формулы (2.81) — (2.82)]

$$X - \frac{S}{\sqrt{n-1}} t_{\gamma, n-1} < \theta_1 < X + \frac{S}{\sqrt{n-1}} t_{\gamma, n-1},$$

$$nS^2/\chi_{1+\gamma_1/2, n-1}^2 < \theta_2 < nS^2/\chi_{1-\gamma_1/2, n-1}^2, \quad S^2 = S^2(X),$$

так как случайные величины $(X - \theta_1)/S$ и S^2 , на основании которых строились эти интервалы, независимы; γ -доверительную область для пары (θ_1, τ) можно построить, например, используя двумерную статистику (X, S^2) , компоненты которой в силу теоремы 1.10 являются независимыми случайными величинами. Соответствующее построение можно выполнить, принимая во внимание результаты примеров 2.17 и 2.19. Сопоставим каждой паре $\theta = (\theta_1, \tau)$ подмножество $H(\theta)$ вида

$$H(\theta) = \left\{ x: \sqrt{\frac{n}{\tau}} |X - \theta_1| < t = c_{\gamma, 1} \right\}$$

$$t' = \chi_{(1-\gamma_1)/2, n-1}^2 < \frac{ns^2}{\tau} < t'' = \chi_{(1+\gamma_1)/2, n-1}^2 \},$$

где $s^2 = S^2(x)$, $\gamma_1 \gamma_2 = \gamma$, $c_{\gamma, 1} = \Phi^{-1} \left(\frac{1+\gamma_1}{2} \right)$, $\chi_{p, n-1}^2$ — p -квантиль распределения $\chi^2(n-1)$. Тогда в силу независимости X и S^2

$$P_\theta(X \in H(\theta)) = P_\theta \left(\sqrt{n/\tau} |X - \theta_1| < c_{\gamma, 1} \right) P_\theta(t' < ns^2/\tau < t'') = \gamma_1 \gamma_2 = \gamma$$

(см. теорему 1.10), т. е. построенные подмножества удовлетворяют условию (2.89). Разрешая неравенства, определяющие подмножества $H(\theta)$, относительно параметров θ_1 и τ , получаем искомые доверительные множества $\mathfrak{D}(x)$, определяемые условиями $\tau > (x - \theta_1)^2 n/t$, $ns^2/t' < \tau < ns^2/t''$. Область $\mathfrak{D}(x)$ представляет собой в данном случае часть плоскости (θ_1, τ) , ограниченную параболой $\tau = (x - \theta_1)^2 n/t$ и двумя прямыми $\tau = ns^2/t'$ и $\tau = ns^2/t''$ (рис. 2.5).

Другие примеры построения доверительных областей приведены в гл. 5.



Рис. 2.5

Задачи

1. Пусть требуется оценить функцию $\tau(\theta) = 1/\theta$ по выборке $X = (X_1, \dots, X_n)$ из распределения $Bi(1, \theta)$. Показать, что в данном случае несмещенных оценок не существует.

2. Пусть по одному наблюдению над случайной величиной ξ с распределением $f(x; \theta) = e^{-\theta} \frac{\theta^x}{x!} / (1 - e^{-\theta})$, $x = 1, 2, \dots$ (урезанное в нуле пуассоновское распределение), требуется оценить функцию $\tau(\theta) = 1 - e^{-\theta}$. Найти статистику, удовлетворяющую условию несмещенности $E_\theta T(X_1) = \tau(\theta)$, и показать, что она практически бесполезна.

3. Пусть производится одно наблюдение над случайной величиной ξ с распределением $Bi(n, \theta)$. Оценить функцию $\tau_{ab}(\theta) = \theta^a (1-\theta)^b$ ($a, b \geq \theta$ — целые числа) и показать, что при $a+b \leq n$ несмещенной оценкой $\tau_{ab}(\theta)$ является статистика $T(X_1) = (X_1)_a (n - X_1)_b / (n)_{a+b}$ (здесь $(z)_a = z(z-1)\dots(z-a+1)$, $(z)_0 = 1$) и что в других случаях несмещенных оценок не существует.

4. По результатам n испытаний оценивается неизвестная вероятность успеха θ в схеме Бернулли $Bi(1, \theta)$. Обозначая через r число успехов в этих испытаниях и рассматривая класс оценок вида $T(X) = (r + \alpha)/(n + \beta)$, вычислить среднеквадратическую ошибку оценки T и показать, что при $\alpha = \sqrt{n}/2$, $\beta = \sqrt{n}$ она не зависит от θ и равна $1/[4(n+1)^2]$. Сравнить эту ошибку с ошибкой обычной оценки r/n .

5. Вычислить функции информации $i(\theta)$, приведенные в табл. 2.1.

6. Проверить данные, приведенные в табл. 2.2.

7. Имеется выборка $X = (X_1, \dots, X_n)$ из распределения $\Gamma(a, \theta)$. Воспользовавшись критерием для экспоненциального семейства, показать, что статистика $T(X) = \frac{1}{n} \sum_{i=1}^n \ln X_i - \ln a$ является эффективной оценкой для $\tau(\theta) =$

$= \Gamma'(\theta)/\Gamma(\theta)$ и при этом $D_0 T = \tau'(\theta)/n$.

8. Доказать, что в модели $\varphi''(\mu, \theta^2)$ не существует эффективной оценки для $\tau(\theta) = \theta$; статистика $T(X) = \frac{1}{n} \sqrt{\frac{\pi}{2}} \sum_{i=1}^n |X_i - \mu|$ является несмещенной оценкой θ и $D_0 T = (\pi - 2)\theta^2/(2n)$.

9. Проверить с помощью непосредственных вычислений, что указанная в примере 2.5 статистика T^* является несмещенной оценкой θ^2 в модели $\varphi''(\theta, \sigma^2)$.

10. С помощью непосредственных вычислений убедиться, что в случае логистического распределения, плотность которого имеет вид $f(x; \theta) = e^{-x+\theta} / (1 + e^{-x+\theta})^2$, $-\infty < x < \infty$:

а) статистика $T(X) = X$ — несмещенная оценка θ и $D_0 T = \pi^2/(3n)$;

б) функция информации $i(\theta) = 1/3$ и, следовательно, X не является эффективной оценкой θ .

11. Доказать, что для экспоненциального семейства функция информации $i(\theta) = (C'(\theta)A'(\theta) - C''(\theta)A'(\theta))/A'(\theta)$. Указание. Сравнить представления (2.20) и (2.27) для дисперсии эффективной оценки.

12. Вычислить величину $E_0 B(\xi)$ для экспоненциального семейства. Указание. По определению,

$$i(\theta) = -E_0 \frac{\partial^2 \ln f(\xi; \theta)}{\partial \theta^2} = -A''(\theta) E_0 B(\xi) - C''(\theta);$$

далее использовать результаты задачи 11.

13. Продолжительность горения электрических ламп имеет распределение $\Gamma(\theta, 1)$. Чтобы оценить θ , берется выборка из n ламп и наблюдаются «время жизни» первых r перегоревших ламп $X_{(1)} < X_{(2)} < \dots < X_{(r)}$. Построить оптимальную оценку вида $T(X) = \sum_{k=1}^r \lambda_k X_{(k)}$. Указание. Воспользоваться

результатом задачи 12 гл. I и перейти к независимым величинам $Y_r = [(n-r+1)/\theta](X_{(r)} - X_{(r-1)})$, $r = 1, \dots, n$;

$$T(X) = \theta \sum_{i=1}^r \frac{\Lambda_i}{n-i+1} Y_i, \quad \Lambda_i = \sum_{k=i}^r \lambda_k, \quad i = 1, \dots, r.$$

Отсюда $E_0 T = \theta \sum_{i=1}^r \Lambda_i/(n-i+1)$, $D_0 T = \theta^2 \sum_{i=1}^r \Lambda_i^2/(n-i+1)^2$ и оптимальный выбор Λ_i таков: $\Lambda_i^* = (n-i+1)/r$, $i = 1, \dots, r$.

Ответ. $T^* = (X_{(1)} + \dots + X_{(r)})/r + (n-r)X_{(r)}/r$.

14. Пусть наблюдаемая случайная величина ξ имеет область изменения $[a(\theta), b]$, где $a(\theta)$ — заданная монотонная функция θ . Показать, что минимальное значение выборки $X_{(1)}$ является достаточной статистикой для θ тогда и только тогда, когда плотность $f(x; \theta)$ имеет вид $f(x; \theta) = g(x)/h(\theta)$, $a(\theta) <$

$\leq x \leq b$. Этот же результат справедлив и для статистики $X_{(n)}$ в случае области $[a, b(\theta)]$, где $b(\theta)$ — заданная монотонная функция θ . Указание. Воспользоваться критерием факторизации.

15. Доказать, что в случае модели $N(\theta, \sigma^2)$ среднее арифметическое выборок является полной достаточной статистикой. Указание. В данном

случае $E_0 \varphi(\bar{X}) \equiv 0 \Leftrightarrow \int_{-\infty}^{\infty} \varphi(t) e^{-nt^2/(2\sigma^2)} e^{nt\theta/\sigma^2} dt \equiv 0$. Этот интеграл есть

двустороннее преобразование Лапласа от $\varphi(t) e^{-nt^2/(2\sigma^2)}$, поэтому $\varphi(t) \equiv 0$.

16. Показать, что в случае модели $N(\theta, \gamma^2 \theta^2)$ достаточной статистикой является пара $T = (X, S^2)$, но эта статистика не полная. Указание. Рассмотреть функцию $\varphi(t) = (n + \gamma^2) S^2 / [(n-1) \gamma^2] - X^2$ и показать, что $E_0 \varphi(T) \equiv 0$.

17. Доказать полноту статистики $T = (X_{(1)}, X_{(n)})$ в случае модели $R(\theta_1, \theta_2)$. Получить отсюда, что $T_1 = (X_{(1)} + X_{(n)})/2$ и $T_2 = [(n+1)/(n-1)] \times \times (X_{(n)} - X_{(1)})$ — оптимальные оценки средней точки $(\theta_1 + \theta_2)/2$ и ширины $\theta_2 - \theta_1$ интервала, на котором сосредоточено распределение. Вычислить дисперсии этих оценок. Указание. Воспользоваться результатами задачи 12 гл. 1.

Ответ. $D_0 T_1 = (\theta_2 - \theta_1)^2 / [2(n+1)(n+2)]$, $D_0 T_2 = 2(\theta_2 - \theta_1)^2 / [(n-1)(n+2)]$.

18. В условиях задачи 17 построить оптимальные оценки для θ_1 и θ_2 .

19. Показать, что в случае *бухпараметрического экспоненциального распределения*, задаваемого плотностью

$$f(x; \theta) = e^{-(x-\theta_1)/\theta_2}, \quad x \geq \theta_1, \quad \theta = (\theta_1, \theta_2),$$

достаточной статистикой является пара $(X_{(1)}, \bar{X})$.

20. Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $\Gamma(\theta, \lambda)$:

а) показать, что $T(X) = \sum_{i=1}^n X_i$ — полная достаточная статистика для θ ;

б) пусть φ — заданная функция, для которой $\tau(\theta) = E_0 \varphi(\frac{T}{n})$ существует. Тогда оптимальная оценка для $\tau(\theta)$ имеет вид

$$\tau^* = \frac{\Gamma(\lambda n)}{\Gamma(\lambda) \Gamma(\lambda(n-1))} \int_0^1 \varphi(tT(X)) t^{\lambda-1} (1-t)^{n-1-\lambda-1} dt.$$

Получить отсюда при $\varphi(x) = x/\lambda$ результат, приведенный в табл. 2.2. Указание. Использовать тот факт, что $\mathcal{L}(T(X)) = \Gamma(\theta, \lambda n)$ (см. замечание в п. 1 § 1.5).

21. Проверить, что в условиях задачи 20 оптимальной оценкой для функции надежности $\tau(\theta, t) = P_\theta(\xi \geq t)$ является статистика

$$\tau^* = \begin{cases} 0 & \text{при } T(X) \leq t, \\ 1 - B(t/T(X); \lambda, \lambda(n-1)) & \text{при } T(X) > t, \end{cases}$$

где $B(z; a, b) = \int_0^z x^{a-1} (1-x)^{b-1} dx$, $0 \leq z \leq 1$ — неполная бета-функция. В частности, для экспоненциального распределения $\Gamma(\theta, 1)$

$$\tau^* = \begin{cases} 0 & \text{при } T(X) \leq t, \\ (1 - t/T(X))^{n-1} & \text{при } T(X) > t. \end{cases}$$

Указание. Взять $\varphi(x) = I_{(t, \infty)}(x) = \begin{cases} 0 & \text{при } x < t, \\ 1 & \text{при } x \geq t. \end{cases}$

22. Показать, что для *распределения Вейбулла*, задаваемого плотностью $f(x; \theta) = \frac{\lambda x^{\lambda-1}}{\theta^\lambda} e^{-(x/\theta)^\lambda}$, $x \geq 0$, полной достаточной статистикой для θ явля-

ется $T(X) = \sum_{i=1}^n X_i^{\lambda}$, а оптимальная оценка функции $\tau(\theta) = E_0 \varphi(\xi)$, где φ —

заданная функция, имеет вид $\tau^* = (n-1) \int_0^1 \varphi([tT(X)]^{1/\lambda}) (1-t)^{n-2} dt$.

23. Показать, что для функции $f(\theta)$, входящей в определение распределения типа степенного ряда [см. формулу (2.41)], $f^* = b_{n+1}(T); b_n(T)$, $T = X_1 + \dots + X_n$ — оптимальная оценка по выборке $X = (X_1, \dots, X_n)$ [$b_n(T)$ определены в (2.42)]. Указание. Применить формулу (2.43).

24. Показать, что $\tau^* = (1 + 1/n)^T$ — оптимальная оценка функции $\tau(\theta) = e^\theta$ в случае пуассоновской модели $\Pi(\theta)$. Вычислить $D_0 \tau^*$ и показать, что $D_0 \tau^* \sim \theta e^{\theta}/n$, $n \rightarrow \infty$. Указание. Воспользоваться результатом задачи 23 [здесь $b_n(t) = n^t/t!$, $D_0 \tau^* = e^{\theta} (e^{\theta/n} - 1)$].

25. Показать, что $\tau^* = (1 - 1/n)^T$ — оптимальная оценка для пуассоновской вероятности $\tau(\theta) = P_\theta(\xi=0) = e^{-\theta}$ и при этом $D_0 \tau^* = e^{-\theta} (e^{\theta/n} - 1) \sim \theta e^{-\theta}/n$, $n \rightarrow \infty$. Указание. Воспользоваться результатом примера 2.13.

26. Показать, что оптимальной оценкой для $\tau(\theta) = \theta$ усеченного в нуле пуассоновского распределения является статистика $\tau^* = TS(n, T-1)/S(n, T)$,

$T \geq n+1$, где $S(n, k) = \sum_{r=0}^n (-1)^{n-r} C_n^r r^k$. Указание. Положить в формуле (2.44) $f(\theta) = e^\theta - 1$.

27. Показать, что в случае выборки из распределения $Bi(r, \theta)$ статистика $\tau^* = \prod_{j=1}^n (1 - T/(T + rn - j))$ — оптимальная оценка для вероятности $\tau(\theta) = P_\theta(\xi=0) = (1-\theta)^r$. Указание. Воспользоваться результатами примера 2.14.

28. Составить уравнения правдоподобия для модели $\mathcal{N}^*(\theta_1, \theta_2^2)$ и найти их решения.

29. Доказать, что полученная в примере 2.18 оценка максимального правдоподобия $\hat{\Sigma}(X)$ ковариационной матрицы Σ многомерного нормального распределения $\mathcal{N}^*(\mu, \Sigma)$ удовлетворяет соотношению $E \hat{\Sigma}(X) = \frac{n-1}{n} \Sigma$ и,

следовательно, $\frac{n}{n-1} \hat{\Sigma}(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$ — несмещенная оценка Σ . Указание. Рассмотреть случайную величину $X_{ii} - X_{ij}$, дисперсия которой равна $\sigma_{ii} + 2\sigma_{ij} + \sigma_{jj}$, и воспользоваться известным результатом о несмещенном оценивании дисперсии скалярной случайной величины.

30. Найти выражение для абсолютного максимума функции правдоподобия в случае распределения $\mathcal{N}^*(\mu, \Sigma)$.

Ответ. $L(x; \hat{\mu}, \hat{\Sigma}) = (2\pi e)^{-kn/2} |\hat{\Sigma}|^{-n/2}$.

31. Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $R(\theta, 2\theta)$. Показать, что:

а) в данном случае не существует одномерной достаточной статистики;

б) любое значение $\theta \in [X_{(n)}/2, X_{(1)}]$ можно взять в качестве оценки максимального правдоподобия.

Рассмотреть оценки вида $\tilde{\theta}_{(\alpha, \beta)} = \alpha X_{(n)} + \beta X_{(1)}$, $\alpha, \beta \geq 0$, и найти среди них оптимальную (несмещенную с минимальной дисперсией). Указание. Воспользоваться результатами задачи 13 гл. 1.

Ответ. Оптимальная оценка определяется значениями

$$\alpha = \frac{(n+1)(n+2)(2n-1)}{5n+4}, \quad \beta = -\frac{2(n+1)(2n^2+4n^2-3n-3)}{(n+2)(5n+4)}.$$

32. Показать, что в случае выбора из распределения $\Gamma(\theta, \lambda)$ оценкой максимального правдоподобия параметрической функции $\tau(\theta) = 1/\theta$ является

$\hat{\tau}_n = \lambda/X$, где $X = \frac{1}{n} \sum_{i=1}^n X_i$. Проверить состоятельность этой оценки и найти

ее предельный при $n \rightarrow \infty$ закон распределения.

Ответ. $\mathcal{L}_\theta(\hat{\tau}_n) \sim \mathcal{N}(1/\theta, 1/(n\lambda\theta^2))$.

33. Показать, что $\hat{\theta} = X/(r+X)$ — оценка максимального правдоподобия параметра θ для модели $Bi(r, \theta)$. Вычислить ее асимптотическую дисперсию. У к а з а н и е. Воспользоваться результатами примеров 2.12 и 2.24.

34. Случайная величина ξ , характеризующая срок службы элементов электронной аппаратуры, имеет плотность $f(x, \theta) = (2x/\theta)e^{-x^2/\theta}$, $x \geq 0$. Построить по соответствующей выборке $X = (X_1, \dots, X_n)$ оценку максимального правдоподобия неизвестного параметра θ .

Ответ. $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i^2$.

35. Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения Коши $\mathcal{K}(\theta)$. Показать, что статистика $T_n(X) = X$ не является состоятельной оценкой параметра θ . У к а з а н и е. Характеристическая функция, а следовательно, и распределение T_n не зависят от n и поэтому $P_\theta(|T_n(X) - \theta| \geq \epsilon)$ одна и та же при всех n ($E_{\theta} e^{itT_n} = e^{it\theta - |t|}$).

36. Доказать, что выборочная дисперсия $S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ является

состоятельной оценкой θ_2^2 в модели $\mathcal{N}(\theta_1, \theta_2^2)$. У к а з а н и е. Воспользоваться критерием состоятельности и результатами примера 2.7.

37. Имеется выборка $((X_1, Y_1), \dots, (X_n, Y_n))$ из двумерного нормального распределения $\mathcal{N}((0, 0), \begin{bmatrix} 1 & \theta \\ \theta & 1 \end{bmatrix})$, $\theta \in (-1, 1)$. Показать, что:

а) асимптотическая дисперсия оценки $\hat{\theta}_n$ равна $(1 - \theta^2)^2/n(1 + \theta^2)$;

б) $\text{eff}(T_n; \theta) = [(1 - \theta^2)/(1 + \theta^2)]^2$ — асимптотическая эффективность оценки $\hat{\theta}$

вида $T_n = \frac{1}{n} \sum_{i=1}^n X_i Y_i$ (выборочный коэффициент корреляции). У к а з а н и е.

С помощью характеристической функции установить, что $E_0(X_i^2 Y_i^2) = 1 + 2\theta^2$ и, следовательно, $D_0 T_n = (1 + \theta^2)/n$.

38. Доказать, что для распределения Лапласа $f(x; \theta) = (1/2)e^{-|x-\theta|}$, $x \in R$, о. м. п. $\hat{\theta}$ совпадает с выборочной медианой.

39. Оценить с помощью метода максимального правдоподобия параметр θ в модели $\mathcal{N}(\mu, \theta^2)$ и найти предельное распределение $\hat{\theta}_n$. Вычислить асимптотическую эффективность оценки в задаче 8.

Ответ. $\mathcal{L}_\theta(\hat{\theta}_n) \sim \mathcal{N}\left(\hat{\theta}_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2}, \frac{\theta}{\sqrt{2n}}\right)$; $\text{eff}(T_n; \theta) = 1/(2 - 1) = 0,88$.

40. Показать, что асимптотическая эффективность выборочной медианы распределения Коши равна $8/\pi^2 \approx 0,8$. У к а з а н и е. Воспользоваться теоремой 1.7.

41. Пусть $X = (X_1, \dots, X_n)$ — выборка из логнормального распределения, т. е. $X_i = e^{Y_i}$, где $\mathcal{L}(Y_i) = \mathcal{N}(\theta_1, \theta_2^2)$. Показать, что:

а) $E_0 X_1 = q_1 = \exp\{\theta_1 + \theta_2^2/2\}$, $D_0 X_1 = q_2 = q_1^2(e^{\theta_2^2} - 1)$;

б) оценки максимального правдоподобия для q_1 и q_2 имеют соответственно вид $\hat{q}_1 = \exp \{Y + (1/2) S^2(Y)\}$, $\hat{q}_2 = Y^2 (e^{S^2(Y)} - 1)$, где $Y = (Y_i = \ln X_i, i = 1, \dots, n)$, а Y и $S^2(Y)$ — соответствующие выборочные среднее и дисперсия. Указание. Воспользоваться свойством инвариантности о. м. п.

42. Показать, что если существует эффективная оценка T^* для функции $\tau(\theta)$, то оценка $\hat{\theta}$ однозначно определяется уравнением $T^*(X) = \tau(\hat{\theta})$. Указание. Применив равенство (2.24), убедиться, что $\frac{\partial^2 \ln L}{\partial \theta^2} < 0$.

43. Убедиться в том, что оценки в задачах 24 и 25 являются асимптотически эффективными.

44. Показать, что γ -доверительный интервал для параметра θ в модели $\mathcal{L}^*(\theta, \theta^2)$ имеет вид $(X/(1+c_\gamma/\sqrt{n}), X/(1-c_\gamma/\sqrt{n}))$, $c_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$.

Указание. Использовать тот факт, что $\mathcal{L}_\theta((X-\theta)/(\theta/\sqrt{n})) = \mathcal{L}^*(0, 1)$.

45. Проверить, что в случае больших выборок асимптотически крайними γ -доверительным интервалом является:

а) для параметра θ гамма-распределения $\Gamma(\theta, \lambda) = (X/\lambda)(1+c_\gamma/\sqrt{n\lambda})$;

б) для параметра θ отрицательного биномиального распределения $B(r, \theta) =$

$$(\hat{\theta}_1 \pm c_\gamma \sqrt{\hat{\theta}_n(1-\hat{\theta}_n)^2/(rn)}) = \frac{1}{1+X} (X \pm c_\gamma \sqrt{X/(n(r+X))});$$

в) для параметра θ нормального распределения $\mathcal{L}^*(\mu, \theta^2) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 (1 \pm c_\gamma/\sqrt{2n})$. Указание. Использовать результаты задач 32, 33 и 39.

46. Показать, что в случае экспоненциального распределения с плотностью $f(x; \theta) = e^{-(x+\theta)}$, $0 \leq x < \infty$, γ -доверительный интервал для θ имеет вид $(X_{(1)} + \frac{\ln(1-\gamma)}{n}, X_{(1)})$, где $X_{(1)} = \min_{1 \leq i \leq n} X_i$. Указание. Найти распределение статистики $X_{(1)}$ и учесть, что событие $\{X_{(1)} \geq \theta\}$ является достоверным.

47. Пусть в полиномиальном распределении $M(n, p = (p_1, \dots, p_N))$ вероятности $p_i = p_i(\theta)$, $i = 1, \dots, N$, где θ — неизвестный скалярный параметр. Записать уравнения для приближенной оценки максимального правдоподобия $\hat{\theta}$ методом накопления. Указание. В данном случае $\frac{\partial \ln L}{\partial \theta} =$

$$= \sum_{i=1}^N \frac{h_i}{p_i(\theta)} p_i'(\theta), \text{ а количество информации о } \theta \text{ равно}$$

$$I_\theta(\theta) = -E_\theta \left(\frac{\partial^2 \ln L}{\partial \theta^2} \right) = n \frac{(p_i'(\theta))^2}{p_i(\theta)}.$$

48. Пусть n , X и S^2 — соответственно объем, выборочное среднее и дисперсия выборки из нормального распределения с неизвестными параметрами. Показать, что с вероятностью γ результат следующего, $(n+1)$ -го испытания находится в интервале $(X \pm t_{(1+\gamma)/2, n-1} S \sqrt{(n+1)/(n-1)})$. Указание. Применить теорему 1.10, согласно которой $\mathcal{L}(\sqrt{(n-1)/(n+1)}(X - X_{n+1})/S) = S(n-1)$.

49. В результате пяти независимых взвешиваний одного и того же тела получены следующие результаты (в граммах): 4,12; 3,92; 4,55; 4,04; 4,35. Считая погрешности измерений нормальными $\mathcal{L}^*(0, \sigma^2)$ случайными величинами

нами, указать доверительные границы для результата предстоящего шестого взвешивания (доверительный уровень принять равным 0,95).

50. По результатам $n = 2$ независимых измерений диаметра круга построить оптимальную оценку его площади (погрешности измерений распределены по закону $N(0, \sigma^2)$ с неизвестным σ^2). Указание. Использовать тот факт, что выборочные среднее $\bar{X}$ и дисперсия S^2 образуют полную достаточную статистику.

Отв.т. $T(\bar{X}, S^2) = (\pi/4)(\bar{X}^2 - S^2/(n-1))$.

51. Доказать, что оптимальная оценка всегда является симметрической функцией наблюдений. Указание. Если $T = T(X)$ — несмещенная оценка τ ,

то рассмотреть симметрическую статистику $\bar{T} = \frac{1}{n!} \sum_{\pi} T(\pi X)$, где $\pi =$

$= \begin{pmatrix} 1 & \dots & n \\ i_1 & \dots & i_n \end{pmatrix}$ — перестановка из n элементов, $\pi X = (X_{i_1}, \dots, X_{i_n})$ и суммирование производится по всем $n!$ перестановкам. Показать, что $D\bar{T} \leq DT$.

Проверка статистических гипотез

Эта глава представляет собой введение в теорию проверки статистических гипотез. Здесь приводятся основные понятия этой теории (статистической гипотезы, статистического критерия, критической области, функции мощности и др.). Сформулированы типичные и наиболее распространенные в приложениях статистические гипотезы и на примерах решения задач проверки этих гипотез излагаются общие принципы построения и исследования критериев согласия. Рассматривается метод группировки наблюдений с последующим применением классического критерия согласия χ^2 . Излагаются некоторые результаты для схемы группировки с растущим числом интервалов.

§ 3.1. Понятие статистической гипотезы и статистического критерия. Критерии согласия

1. Статистические гипотезы. Задача разработки рациональных методов проверки статистических гипотез — одна из основных задач математической статистики. *Статистической гипотезой* (или просто *гипотезой*) называют любое утверждение о виде или свойствах распределения наблюдаемых в эксперименте случайных величин. Такие утверждения можно делать на основании теоретических соображений или статистических исследований других наблюдений. Пусть, например, эксперимент состоит в многократном измерении некоторой физической величины, точное значение a которой не известно и в процессе измерений не меняется. На результаты измерений влияют многие случайные факторы (точность настройки измерительного прибора, погрешность округления при считывании данных и т. д.), поэтому результат i -го измерения X_i можно записать в виде $X_i = a + \varepsilon_i$, где ε_i — случайная погрешность измерения. Обычно считают, что общая ошибка ε_i складывается из большого числа ошибок, каждая из которых невелика. На основании центральной предельной теоремы можно предположить, что случайные величины X_i имеют нормальное распределение. Такое предположение является статистической гипотезой о виде распределения наблюдаемых случайных величин.

Если для исследуемого явления (процесса, ситуации и т. д.) сформулирована та или иная гипотеза (обычно ее называют *основной* или *нулевой* гипотезой и обозначают символом H_0), то задача состоит в том, чтобы сформулировать такое правило, которое позволяло бы по результатам соответствующих наблюдений (по имеющимся статистическим данным) принять или отклонить эту гипотезу. Правило, согласно которому проверяемая гипотеза H_0 принимается или отвергается, называется *статистическим критерием* (или просто *критерием*) проверки гипотезы H_0 . Разработка таких правил и их обоснование с точки зрения требований оптимальности и составляют предмет теории проверки статистических гипотез.

Приведем несколько примеров математических формулировок наиболее распространенных в приложениях статистических гипотез.

Пример 3.1 (гипотеза о виде распределения). Пусть производится n независимых наблюдений над некоторой случайной величиной ξ с неизвестной функцией распределения $F_{\xi}(x)$. Гипотезой, подлежащей проверке, может быть утверждение типа $H_0: F_{\xi}(x) = F(x)$, где функция $F(x)$ полностью задана, либо типа $H_0: F_{\xi}(x) \in \mathcal{F}$, где $\mathcal{F}$ — заданное семейство функций распределения. При этом обычно семейство $\mathcal{F}$ задается в параметрическом виде: $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$. Например, наблюдается неотрицательная целочисленная случайная величина ξ и требуется проверить гипотезу $H_0: \mathcal{L}(\xi) \in \Pi(\theta)$, где $\Pi(\theta)$ — семейство всех пуассоновских распределений. Во всех этих случаях H_0 — это *гипотеза о виде распределения* наблюдаемой случайной величины. В других случаях гипотеза состоит в том, что фиксируются некоторые частные характеристики функции распределения $F_{\xi}(x)$. Например, она может иметь вид $H_0: F_{\xi}(\xi_i) = p_i, i = 1, \dots, k$, где $-\infty < \xi_1 < \dots < \xi_k < \infty$ и $0 < p_1 < \dots < p_k < 1$ — заданные числа. Здесь H_0 — это гипотеза о том, что распределение $\mathcal{L}(\xi)$ имеет заданные квантили ξ_i для заданных уровней p_i .

Пример 3.2 (гипотеза однородности). Произведено k серий независимых наблюдений, результаты которых таковы: $(x_{i1}, \dots, x_{in_i}), i = 1, \dots, k$. Имеются ли основания рассматривать эти данные как результаты наблюдений над одной и той же случайной величиной, или, другими словами, можно ли с достаточной надежностью считать, что закон распределения наблюдений от серии к серии не меняется? Если это так, то говорят, что статистические данные *однородны*. Пусть $F_i(x)$ — функция (вообще говоря, неизвестная) распределения наблюдений i -й серии, $i = 1, \dots, k$. Тогда задача состоит в проверке *гипотезы однородности* $H_0: F_1(x) \equiv \dots \equiv F_k(x)$.

Пример 3.3 (гипотеза независимости). В эксперименте наблюдается двумерная случайная величина $\xi = (\xi_1, \xi_2)$ с неизвестной функцией распределения $F_{\xi}(x, y)$ и есть основания предполагать, что компоненты ξ_1 и ξ_2 независимы. В этом случае надо проверить *гипотезу независимости* $H_0: F_{\xi}(x, y) = F_{\xi_1}(x)F_{\xi_2}(y)$, где $F_{\xi_1}(x)$ и $F_{\xi_2}(y)$ — некоторые одномерные функции распределения.

В общем случае можно рассматривать k -мерную случайную величину ξ и проверять гипотезу независимости ее компонент.

Пример 3.4 (гипотеза случайности). Результат эксперимента описывается n -мерной случайной величиной $X = (X_1, \dots, X_n)$ с известной функцией распределения $F_X(x)$, $x = (x_1, \dots, x_n)$. Можно ли рассматривать X как случайную выборку из распределения некоторой случайной величины ξ (т. е. являются ли компоненты X_i независимыми и одинаково распределенными)? В данном случае требуется проверить *гипотезу случайности* $H_0: F_X(x) = F_{\xi}(x_1) \dots F_{\xi}(x_n)$, где $F_{\xi}(x)$ — некоторая одномерная функция распределения.

Эти примеры не охватывают возникающих в приложениях задач проверки статистических гипотез. В частности, бывают довольно часто ситуации, когда проверяемая гипотеза состоит в том, что некоторый параметр семейства распределений соответствующей совокупности, например среднее значение, дисперсия и т. п., имеет наперед заданное значение или множество значений. Такие гипотезы называются *параметрическими*. Более подробно они будут рассмотрены ниже.

Рассмотрим общие методы проверки гипотез описанных выше типов. Во всех этих случаях формулируется только одна гипотеза H_0 и требуется проверить, согласуются ли имеющиеся статистические данные с этой гипотезой или же они ее опровергают. Соответствующие критерии называют *критериями согласия*. Введем понятие простой и сложной гипотезы. Если гипотеза H_0 однозначно фиксирует распределение наблюдений, то ее называют *простой*, в противном случае — *сложной*. В предыдущих примерах только гипотеза $H_0: F_{\xi}(x) = F(x)$ является простой, а остальные — сложные.

2. Критерии согласия и основные их характеристики. Рассмотрим общий метод построения критериев согласия. Пусть о распределении случайной величины $X = (X_1, \dots, X_n)$, описывающей результат изучаемого эксперимента, сформулирована некоторая гипотеза H_0 . Чтобы построить критерий проверки этой гипотезы, в большинстве случаев поступают следующим образом. Пытаются найти такую статистику $T = T(X)$, характеризующую отклонение эмпирических данных от соответствующих (гипотезе H_0) гипотетических значений (обычно такие статистики неотрицательны), распределение которой в случае справедливости H_0 можно было бы определить (точно или приближенно). В частности, если H_0 — сложная гипотеза, то распределение $T(X)$ должно быть одним и тем же для всех простых гипотез, составляющих H_0 .

Предположим, что такая статистика и ее распределение при гипотезе H_0 найдены. Пусть $\mathcal{T} = \{t: t = T(x), x \in \mathcal{X}\}$ — множество всех возможных значений статистики T ; определим для фиксированного заранее достаточно малого числа $\alpha > 0$ подмножество $\mathcal{T}_{1-\alpha} \subset \mathcal{T}$ так, чтобы вероятность осуществления события $\{T(X) \in \mathcal{T}_{1-\alpha}\}$ в случае справедливости гипотезы H_0 (в дальнейшем для обозначения этой вероятности используется символиче-

ская запись $P(T(X) \in \mathcal{T}_{1\alpha} | H_0)$ удовлетворяла условию

$$P(T(X) \in \mathcal{T}_{1\alpha} | H_0) = \alpha. \quad (3.1)$$

Тогда правило проверки гипотезы H_0 можно сформулировать следующим образом. Пусть x — наблюдавшаяся реализация случайной величины X и $t = T(x)$ — соответствующее значение статистики T . Если окажется, что $t \in \mathcal{T}_{1\alpha}$, то в предположении справедливости гипотезы H_0 произошло маловероятное событие и эта гипотеза должна быть отвергнута как противоречащая статистическим данным. В противном случае (т. е. если $t \notin \mathcal{T}_{1\alpha}$) нет оснований отказываться от принятой гипотезы и следует считать, что наблюдения не противоречат гипотезе (или согласуются с ней).

В описанной выше ситуации критерий имеет следующий вид: если $t = T(x)$ — наблюдавшееся значение статистики $T(X)$, то при $t \in \mathcal{T}_{1\alpha}$ гипотеза H_0 отвергается; в противном случае считается, что данные не противоречат гипотезе. Отметим, что факт $t \in \mathcal{T}_{1\alpha} = \mathcal{T} \setminus \mathcal{T}_{1\alpha}$ не является доказательством истинности гипотезы H_0 ; он лишь свидетельствует о том, что согласие опытных данных и теоретических предположений достаточно хорошее («на уровне α »).

В описанной методике статистику T называют *статистикой критерия*, а подмножество ее значений $\mathcal{T}_{1\alpha}$ — *критической областью* для гипотезы H_0 (этот термин отражает тот факт, что значения $t \in \mathcal{T}_{1\alpha}$ рассматривают как свидетельствующие против этой гипотезы). Число α в соотношении (3.1) называют *уровнем значимости* критерия, и его можно считать вероятностью ложного отвержения гипотезы H_0 , когда она верна (т. е. вероятностью ошибочного решения в ситуации, когда H_0 истинна). В конкретных задачах величину α выбирают обычно равной 0,1; 0,05; 0,01 и т. д.

Итак, согласно описанной методике, критерий определяется заданием соответствующей критической области в множестве значений статистики T . По своему смыслу критическая область должна включать все маловероятные при гипотезе H_0 значения статистики критерия. Обычно используют области вида $\{t \geq t_{1\alpha}\}$ (для неотрицательных статистик T) или вида $\{|t| \geq t_{1\alpha}\}$, хотя в конкретных задачах возможны и другие варианты выбора критической области. Вид критической области во многом определяется целью, для которой строится критерий. Ограничимся пока следующим общим замечанием: каждый критерий строится для того, чтобы определять, имеют ли место те или иные отклонения от основной гипотезы. Характер таких отклонений может быть разным, поэтому надо иметь критерии как универсального типа («улавливающие» любые отклонения от основной гипотезы), так и предназначенные для выявления отклонений только определенного типа. Так, часто большие и малые значения статистики $T(X)$ указывают на разный характер отклонения от нулевой гипотезы, поэтому может оказаться, что в одних случаях лучше использовать критерий, основанный на критической области $\{t \geq t_{1\alpha}\}$,

а в других — на критической области $\{t \leq t_\alpha\}$. Примеры таких ситуаций приводятся в дальнейшем.

Для проверки одной и той же гипотезы H_0 можно строить различные критерии согласия, основываясь на разных статистиках $T(X)$, и чтобы выбрать в конкретной ситуации тот или иной критерий, надо иметь принципы сравнения различных критериев. Идея построения таких принципов состоит в исследовании поведения критериев при тех или иных отклонениях от основной гипотезы. Введем понятия альтернативного распределения (альтернативной гипотезы) и мощности критерия.

Любое распределение $F_X = F$ наблюдаемой случайной величины X , которое может оказаться истинным (т. е. допустимо в данной ситуации), но отличающееся от гипотетического (т. е. распределения при основной гипотезе H_0), называют *альтернативным распределением* или *альтернативой*. Совокупность всех альтернативных распределений называют *альтернативной гипотезой* и обозначают символом H_1 .

Пусть, далее, для проверяемой гипотезы H_0 построен некоторый критерий с уровнем значимости α , основанный на статистике $T(X)$, и пусть $\mathcal{T}_{1\alpha}$ — соответствующая критическая область. Величину $W(F) = W(\mathcal{T}_{1\alpha}; F)$, представляющую собой вероятность попадания значения статистики критерия в критическую область, когда истинным распределением наблюдений является распределение F , называют *функцией мощности критерия*. (Условимся записывать в дальнейшем эту вероятность в виде $P(T(X) \in \mathcal{T}_{1\alpha} | F)$, так что $W(F) = P(T(X) \in \mathcal{T}_{1\alpha} | F)$.) Таким образом, функция мощности — это некоторый функционал на множестве всех допустимых распределений $\{F\}$. Для распределений F , составляющих нулевую гипотезу (условимся этот факт записывать как $F \in H_0$), значения функции мощности удовлетворяют по построению критерия условию (3.1); перепишем это условие в новых терминах в виде

$$W(F) \leq \alpha, \quad \forall F \in H_0. \quad (3.2)$$

Если $F \in H_1$, то значение $W(F)$ называют *мощностью критерия при альтернативе F* .

Функция мощности играет в теории проверки гипотез фундаментальную роль. Она полностью характеризует критерий, так как показывает, насколько хорошо критерий соответствует своему основному назначению — «улавливать» возможные отклонения от основной гипотезы. В терминах функции $W(F)$ можно сказать, что критерий тем лучше («тем мощнее»), чем больше его мощность при альтернативах. Действительно, если наблюдавшееся значение $t(x)$ попадает в критическую область, то нулевую гипотезу отклоняют, и если истинной действительно является некоторая альтернатива (нулевая гипотеза не верна), то тем самым принимают правильное решение. Следовательно, значение $W(F)$ при $F \in H_1$ характеризует вероятность принятия правильного решения в ситуации, когда нулевая гипотеза ложна.

Желательным свойством критерия является свойство *несмещенности*, которое означает, что одновременно с условием (3.2) должно выполняться условие

$$W(F) > \alpha, \quad \forall F \in H_1. \quad (3.3)$$

Другими словами, несмещенность критерия означает, что вероятность отвергнуть нулевую гипотезу, когда она истинна, не превышает заданного уровня значимости α , и в то же время если гипотеза H_0 ложна, то она отвергается с вероятностью, большей α .

Вычисление функции мощности критерия — трудная задача, так как для этого требуется знать распределение статистики критерия не только при нулевой гипотезе, но и при альтернативах. Поэтому функцию $W(F)$ удастся найти далеко не во всех случаях.

В заключение отметим, что важным показателем каждого критерия является трудоемкость практической реализации соответствующего алгоритма. На практике, когда требуется быстро получить ответ, предпочтение нередко отдается просто реализуемому критерию, даже если он не является оптимальным в теоретическом смысле. Продемонстрируем общую методику построения критериев согласия на примерах решения задач проверки описанных выше гипотез.

§ 3.2. Проверка гипотезы о виде распределения

Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{L}(\xi)$ с неизвестной функцией распределения $F_\xi(x)$, о которой выдвинута простая гипотеза $H_0: F_\xi(x) = F(x)$. Наиболее известными критериями проверки этой гипотезы являются критерий Колмогорова и критерий χ^2 .

1. Критерий согласия Колмогорова. Его применяют в тех случаях, когда функция $F(x)$ непрерывна. Статистикой критерия является величина

$$\mathcal{D}_n = \mathcal{D}_n(X) = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|, \quad (3.4)$$

представляющая собой максимальное отклонение эмпирической функции распределения $F_n(x)$ от гипотетической функции распределения $F(x)$. При каждом x величина $F_n(x)$ является оптимальной оценкой для $F(x)$ [см. замечание к теореме (2.2)] и с увеличением объема выборки n происходит сближение $F_n(x)$ с $F(x)$, поэтому, по крайней мере при больших n , в тех случаях, когда гипотеза H_0 истинна, значение $\mathcal{D}_n$ не должно существенно отклоняться от нуля. Отсюда следует, что критическую область критерия, основанного на статистике $\mathcal{D}_n$, следует задавать в виде $\mathcal{K}_{1\alpha} = \{t \geq t_\alpha\}$.

Особенностью статистики $\mathcal{D}_n$ является то, что ее распределение при гипотезе H_0 не зависит от вида функции $F(x)$. Действи-

тельно, полагая в формуле (3.4) $x = F^{-1}(u)$, $0 \leq u \leq 1$, где $F^{-1}(u)$ — функция, обратная к $F(x)$, получаем

$$\mathcal{D}_n = \sup_{0 \leq u \leq 1} |F_n(F^{-1}(u)) - u|.$$

Перейдем к новым случайным величинам, используя формулу $U_i = F(X_i)$, $i = 1, \dots, n$; пусть $U_{(1)}, \dots, U_{(n)}$ — их вариационный ряд. Функция $F(x)$ монотонна, поэтому $U_{(k)} = F(X_{(k)})$, $k = 1, \dots, n$, и неравенства $F^{-1}(u) \geq X_k$ эквивалентны неравенствам $u \geq U_{(k)}$. Но тогда из представления (1.5) имеем

$$F_n(F^{-1}(u)) = \frac{1}{n} \sum_{k=1}^n e(F^{-1}(u) - X_{(k)}) = \frac{1}{n} \sum_{k=1}^n e(u - U_{(k)}) = \Phi_n(u).$$

Независимо от вида функции $F(x)$ (если только $F(x)$ непрерывна) $\mathcal{L}(U_i) = R(0, 1)$ и $\Phi_n(u)$ — эмпирическая функция распределения для выборки из равномерного $R(0, 1)$ распределения. Таким образом, статистика $\mathcal{D}_n$ совпадает с $\sup_{0 \leq u \leq 1} |\Phi_n(u) - u|$ и, следова-

тельно, ее распределение не зависит от $F(x)$ (если справедлива гипотеза H_0). Этот факт имеет принципиальное значение, так как тем самым достаточно вычислить и протабулировать распределение $\mathcal{D}_n$ только один раз, а именно для выборки из равномерного $R(0, 1)$ распределения, и использовать ее для проверки гипотезы относительно произвольной непрерывной функции распределения $F(x)$.

Вторым замечательным фактом является то, что распределение статистики $\mathcal{D}_n$ при достаточно больших n (уже при $n \geq 20$) практически от n не зависит и имеет указанный вид (1.8). Отсюда следует, что при $n \geq 20$ критическую границу $t_\alpha = t_\alpha(n)$ можно полагать равной $\lambda_\alpha / \sqrt{n}$, где $K(\lambda_\alpha) = 1 - \alpha$ [функция $K(t)$ определена в (1.8)]. Действительно, в этом случае

$$P(\mathcal{D}_n \in \mathcal{T}_{1\alpha} | H_0) = P(\sqrt{n} \mathcal{D}_n \geq \lambda_\alpha | H_0) \approx 1 - K(\lambda_\alpha) = \alpha.$$

Так, $\lambda_{0.05} = 1,3581$ при $\alpha = 0,05$ и $\lambda_{0.01} = 1,6276$ при $\alpha = 0,01$.

Таким образом, при заданном уровне значимости α число λ_α определяют из соотношения $K(\lambda_\alpha) = 1 - \alpha$ и в этом случае правило проверки гипотезы H_0 имеет (при $n \geq 20$) следующий вид: если наблюдавшееся значение $t = \mathcal{D}_n(x)$ статистики (3.4) удовлетворяет неравенству $\sqrt{n} t \geq \lambda_\alpha$, то гипотезу H_0 отвергают; в противном случае делают вывод, что статистические данные не противоречат гипотезе. Следуя этому правилу, можно ошибочно отклонить гипотезу H_0 , когда она верна, с вероятностью, приблизительно равной α . Это правило называется критерием согласия Колмогорова.

Для небольших значений n точное распределение $\mathcal{D}_n$ также протабулировано и для расчета точных значений критической границы t_α [10, с. 611] можно воспользоваться соответствующими таблицами.

2. Критерий согласия хи-квадрат К. Пирсона. На практике вычисление статистики D_n — трудоемкая задача, поэтому часто применяют другой критерий, называемый *критерием χ^2* . Его можно использовать для любых распределений, в том числе и многомерных. Чтобы воспользоваться этим критерием, выборочные данные предварительно группируют, как это описано в п. 2 § 2.5, т. е. переходят к частотному представлению исходных данных. Пусть $\mathbf{v} = (v_1, \dots, v_N)$ — вектор частот попадания выборочных точек в соответствующие интервалы группировки $\mathcal{E}_1, \dots, \mathcal{E}_N$ ($v_1 + \dots + v_N = n$) и $\mathbf{p}^0 = (p_1^0, \dots, p_N^0)$, где $p_j^0 = P(\xi \in \mathcal{E}_j | H_0)$, $j = 1, \dots, N$. В этом случае $\mathcal{L}(\mathbf{v} | H_0) = M(n; \mathbf{p}^0)$ и гипотеза H_0 сводится к гипотезе о том, что вероятности полиномиального распределения построенного вектора частот $\mathbf{v}$ имеют заданные значения p_j^0 , $j = 1, \dots, N$. В качестве статистики, характеризующей отклонение выборочных данных (т. е. частот v_j) от соответствующих гипотетических значений (в данном случае от средних $E(v_j | H_0) = np_j^0$), принимают величину

$$X_n^2 = X_n^2(\mathbf{v}) = \sum_{j=1}^N (v_j - np_j^0)^2 / (np_j^0) = \sum_{j=1}^N v_j^2 / (np_j^0) - n, \quad (3.5)$$

а критическую область задают в виде $\mathcal{T}_{1-\alpha} = \{t \geq t_\alpha\}$. Точное распределение $\mathcal{L}(X_n^2 | H_0)$ неудобно для вычисления (при заданном уровне значимости) критической границы t_α , но для больших объемов выборок n статистика X_n^2 имеет при гипотезе H_0 простое предельное распределение, не зависящее от гипотезы (т. е. от чисел p_j^0). Справедливо следующее утверждение.

Теорема 3.1. Если $0 < p_j^0 < 1$, $j = 1, \dots, N$, то при $n \rightarrow \infty$

$$\mathcal{L}(X_n^2 | H_0) \rightarrow \chi^2(N-1).$$

□ Используя формулу

$$(a_1 + \dots + a_N)^n = \sum_{h_1 + \dots + h_N = n} \frac{n!}{h_1! \dots h_N!} a_1^{h_1} \dots a_N^{h_N}$$

(здесь суммирование производится по всем целым неотрицательным значениям $(h_1, \dots, h_N)$, удовлетворяющим условию $h_1 + \dots + h_N = n$), получаем, что характеристическая функция вектора $\mathbf{v} = (v_1, \dots, v_N)$ при гипотезе H_0 имеет вид

$$E e^{i\mathbf{t} \cdot \mathbf{v}} = (p_1^0 e^{it_1} + \dots + p_N^0 e^{it_N})^n, \quad \mathbf{t} = (t_1, \dots, t_N).$$

Введем нормированный вектор $\mathbf{v}^* = (v_1^*, \dots, v_N^*)$, где $v_j^* = (v_j - np_j^0) / \sqrt{n}$, $j = 1, \dots, N$. Имеем

$$\varphi_n(\mathbf{t}) = E e^{i\mathbf{t} \cdot \mathbf{v}^*} = e^{-i\mathbf{t} \cdot \mathbf{V}^* \mathbf{p}^0} \left[1 + \sum_{j=1}^N p_j^0 (e^{it_j / \sqrt{n}} - 1) \right]^n.$$

Логарифмируя это соотношение и применяя формулу $\ln(1+\varepsilon) = \varepsilon - \varepsilon^2/2 + O(\varepsilon^3)$, $\varepsilon \rightarrow 0$, получаем, что при $n \rightarrow \infty$ и $|\mathbf{t}| \leq c < \infty$

$$\begin{aligned} \ln \varphi_n(\mathbf{t}) &= -i\sqrt{n} \mathbf{t}' \mathbf{p}^0 + n \sum_{j=1}^N p_j^0 (e^{it_j/\sqrt{n}} - 1) - \\ &- \frac{n}{2} \left[\sum_{j=1}^N p_j^0 (e^{it_j/\sqrt{n}} - 1)^2 \right] + O\left(\frac{1}{\sqrt{n}}\right) = \\ &= -\frac{1}{2} \sum_{j=1}^N p_j^0 t_j^2 + \frac{1}{2} \left(\sum_{j=1}^N p_j^0 t_j \right)^2 + O\left(\frac{1}{\sqrt{n}}\right) = -\frac{1}{2} \mathbf{t}' \Sigma \mathbf{t} + O\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

где $\Sigma = \|\sigma_{jk}\|_1^N$ и $\sigma_{jk} = \begin{cases} p_j^0(1-p_j^0) & \text{при } j=k, \\ -p_j^0 p_k^0 & \text{при } j \neq k. \end{cases}$ Отсюда следует, что предел характеристической функции вектора $\mathbf{v}^*$ есть $\exp\{-(1/2)\mathbf{t}'\Sigma\mathbf{t}\}$ — характеристическая функция нормального закона $\mathcal{N}(\mathbf{0}, \Sigma)$. По теореме непрерывности для характеристических функций отсюда имеем $\mathcal{L}(\mathbf{v}^* | H_0) \rightarrow \mathcal{N}(\mathbf{0}, \Sigma)$ при $n \rightarrow \infty$. Матрица вторых моментов Σ предельного распределения вырождена. (Это является следствием того, что компоненты вектора $\mathbf{v}^*$

связаны линейной зависимостью $\sum_{j=1}^N v_j^* = 0$.) Но определитель

$(N-1)$ -го порядка матрицы $\Sigma(N-1) = \|\sigma_{jk}\|_1^{N-1}$ уже отличен от нуля. Таким образом, предельное распределение подвектора $\mathbf{v}^*(N-1) = (v_1^*, \dots, v_{N-1}^*)$ является уже невырожденным нормальным законом $\mathcal{N}(\mathbf{0}, \Sigma(N-1))$. Отсюда по теореме 1.9 следует, что при $n \rightarrow \infty$

$$\mathcal{L}(Q_n = \mathbf{v}^{*'}(N-1)\Sigma^{-1}(N-1)\mathbf{v}^*(N-1) | H_0) \rightarrow \chi^2(N-1). \quad (3.6)$$

С другой стороны, из формулы (3.5) имеем

$$\begin{aligned} X_n^2 &= \sum_{j=1}^N \frac{1}{p_j^0} (v_j^*)^2 = \sum_{j=1}^{N-1} \frac{1}{p_j^0} (v_j^*)^2 + \frac{1}{p_N^0} (v_1^* + \dots + v_{N-1}^*)^2 = \\ &= \mathbf{v}^{*'}(N-1) \mathbf{A} \mathbf{v}^*(N-1), \end{aligned}$$

где $\mathbf{A} = \|a_{jk}\|_1^{N-1}$ и $a_{jk} = \begin{cases} 1/p_j^0 + 1/p_N^0 & \text{при } j=k, \\ 1/p_N^0 & \text{при } j \neq k, \end{cases}$ Непосредственной проверкой убеждаемся, что $\mathbf{A}\Sigma(N-1) = \Sigma(N-1)\mathbf{A} = \mathbf{E}_{N-1}$, т. е. $\mathbf{A} = \Sigma^{-1}(N-1)$. Таким образом, X_n^2 совпадает с квадратичной формой Q_n в соотношении (3.6). ■

На практике предельное распределение $\chi^2(N-1)$ можно использовать с хорошим приближением уже при $n \geq 50$ и $v_j \geq 5$. При выполнении этих условий в соответствии с теоремой 3.1 критическую границу t_α выбирают равной $\chi_{1-\alpha, N-1}^2$, т. е. $(1-\alpha)$ —

квантили распределения $\chi^2(N-1)$. Действительно, в этом случае

$$P(X_n^2 \in \mathcal{F}_{1-\alpha} | H_0) = P(X_n^2 \geq \chi_{1-\alpha, N-1}^2 | H_0) \approx \int_{\chi_{1-\alpha, N-1}^2}^{\infty} k_{N-1}(x) dx = \alpha$$

(здесь $k_{N-1}(x)$ — плотность распределения $\chi^2(N-1)$).

Таким образом, критерий согласия χ^2 имеет следующий вид: пусть заданы уровень значимости α и объем выборки n и наблюдавшиеся значения $\mathbf{h} = (h_1, \dots, h_N)$ вектора частот $\mathbf{v} = (v_1, \dots, v_N)$ удовлетворяют условиям $n \geq 50$, $h_j \geq 5$, $j = 1, \dots, N$; тогда если наблюдавшееся значение $t = X_n^2(\mathbf{h})$ статистики (3.5) удовлетворяет неравенству $t \geq \chi_{1-\alpha, N-1}^2$, то гипотезу H_0 отвергают; в противном случае гипотеза H_0 не противоречит результатам испытаний.

Сделаем несколько общих замечаний. Критерий согласия χ^2 применяется в тех случаях, когда в каждом опыте наблюдается одно из N несовместных событий $A_1, \dots, A_N$ и заданы частоты появлений этих событий в n испытаниях (говорят также, что наблюдается дискретная случайная величина, принимающая N различных значений). Если же выборка имеет непрерывный закон распределения, то, применяя предварительно метод группировки данных, приходят к рассмотрению дискретной схемы, в которой в качестве событий A_j рассматриваются события $\{\xi \in \mathcal{E}_j\}$, где $\mathcal{E}_1, \dots, \mathcal{E}_N$ — интервалы группировки. Недостатком метода является то, что группировка данных по классам (интервалам) приводит к некоторой потере информации. Кроме того, остается еще вопрос о выборе числа интервалов N и длине самих интервалов $\mathcal{E}_j$. (Более подробно эти вопросы освещены в [10, гл. 30].) Однако критерий χ^2 имеет и некоторые достоинства: при его применении нет необходимости учитывать точные значения наблюдений (бывают случаи, когда исходные статистические данные носят не числовой характер; см. пример 3.7). Несомненным преимуществом этого критерия является его универсальность.

Приведем несколько примеров применения критерия χ^2 .

Пример 3.5. При $n = 4040$ бросаниях монеты Бюффон получил $h_1 = 2048$ выпадения «герба» и $h_2 = n - h_1 = 1992$ выпадений решетки. Проверим, используя критерий χ^2 , совместимы ли эти данные с гипотезой H_0 о том, что монета была симметричной, т. е. что вероятность выпадения «герба» $p = 1/2$. Здесь $N = 2$, $p_1 = p = 1/2$, $p_2 = 1 - p = q = 1/2$ и из (3.5) имеем $t = X_n^2(\mathbf{h}) = (h_1 - np)^2/(np) + (h_2 - nq)^2/(nq) = (h_1 - np)^2/(npq) = 0,776$. Пусть уровень значимости α был задан равным 0,05. По таблицам распределения χ^2 находим $\chi_{0,95; 1} = 3,841$. Сравниваем полученное значение t с табличной величиной $\chi_{0,95; 1}^2$. Так как $t < \chi_{0,95; 1}^2$, то делаем вывод, что данные не противоречат гипотезе.

Рассмотрим пример [13, с. 459], когда гипотетическое распределение является непрерывным.

Пример 3.6. Наблюдались показания 500 наугад выбранных часов, выставленных в витринах часовщиков. Пусть i — номер промежутка от i -го часа до $(i+1)$ -го, $i = 0, 1, \dots, 11$, а h_i —

число часов, показания которых принадлежали i -му промежутку. Результаты наблюдений оказались следующими:

i	0	1	2	3	4	5	6	7	8	9	10	11	Всего
h_i	41	34	54	39	49	45	41	33	37	41	47	39	$n=500$

Согласуются ли эти данные с гипотезой H_0 о том, что показания часов равномерно распределены на интервале $(0, 12)$?

Здесь $N=12$ и, согласно гипотезе H_0 , $p_1^0 = \dots = p_{12}^0 = 1/12$. Отсюда значение статистики

$$\chi_n^2(h) = \sum_{j=1}^N (h_j - np_j^0)^2 / (np_j^0) = 10,000.$$

По таблицам распределения χ^2 находим $\chi_{0.95; 11}^2 = 19,675$, поэтому следует признать, что согласие предположения с опытными данными хорошее.

Пример 3.7 [10, с. 563]. В экспериментах с селекцией гороха Мендель наблюдал частоты различных видов семян, получаемых при скрещивании растений с круглыми желтыми семенами и растений с морщинистыми зелеными семенами. Эти данные и значения теоретических вероятностей, определяемые в соответствии с теорией наследственности Менделя, приведены в следующей таблице:

Семена	Частота h_i	Вероятность p_i^0
Круглые и желтые	315	9/16
Морщинистые и желтые	101	3/16
Круглые и зеленые	104	3/16
Морщинистые и зеленые	32	1/16
Σ	$n = 556$	1

Следует проверить гипотезу H_0 о согласовании частотных данных с теоретическими вероятностями.

Здесь $\chi_n^2(h) = 0,47$. Из таблиц распределения χ^2 следует, что при любом уровне значимости $\alpha \leq 0,90$ критерий χ^2 не отвергает гипотезу, или, другими словами, между наблюдениями и гипотезой имеется очень хорошее согласие.

Для критерия χ^2 можно исследовать предельное при $n \rightarrow \infty$ поведение мощности при произвольной альтернативе. В рассматриваемой методике гипотезы характеризуются вектором $p = (p_1, \dots, p_N)$ вероятностей, с которыми появляются в каждом опыте события $A_1, \dots, A_N$, поэтому для функции мощности будем

использовать обозначение $W(p)$, а о соответствующей гипотезе будем говорить для краткости как о гипотезе p . Чтобы подчеркнуть зависимость функции мощности от объема выборки, будем писать $W_n(p)$.

Исследуя асимптотические свойства критериев (т. е. поведение функций мощности при $n \rightarrow \infty$), прежде всего рассматривают вопрос, является ли критерий состоятельным. По определению, критерий называют *состоятельным*, если при $n \rightarrow \infty$ $W_n(F) \rightarrow 1$, $\forall F \in H_1$. Состоятельность критерия означает, что с ростом числа наблюдений он позволяет с вероятностью, близкой к 1, «улавливать» любые отклонения от основной гипотезы. В частности, состоятельный критерий является асимптотически несмещенным [см. (3.3)]. В рассматриваемом случае справедливо следующее утверждение.

Теорема 3.2. Для любого сектора $p \neq p^0$ при $n \rightarrow \infty$ функция мощности $W_n(p)$ стремится к 1, т. е. критерий χ^2 является состоятельным.

□ Вычислим среднее и дисперсию статистики X_n^2 при гипотезе p . Для этого перепишем формулу (3.5) в виде

$$X_n^2 = \sum_{j=1}^N (v_j - np_j)^2 / (np_j) + 2 \sum_{j=1}^N (v_j - np_j) (p_j - p_j^0) / p_j^0 + \\ + n \sum_{j=1}^N (p_j - p_j^0)^2 / p_j^0.$$

Так как $E(v_j | p) = np_j$, $E[(v_j - np_j)^2 | p] = D(v_j | p) = np_j(1 - p_j)$, то

$$E(X_n^2 | p) = n \sum_{j=1}^N (p_j - p_j^0)^2 / p_j^0 + \sum_{j=1}^N p_j(1 - p_j) / p_j^0. \quad (3.7)$$

Отсюда, в частности, имеем $E(X_n^2 | p^0) = N - 1$. Этот точный результат согласуется с асимптотическим результатом теоремы 3.1, поскольку среднее предельного распределения $\chi^2(N - 1)$ равно $N - 1$ [см. (1.29)]. Приведем без доказательства формулу дисперсии:

$$D(X_n^2 | p) = 4 \frac{(n-1)(n-2)}{n} (R_{32} - R_{11}^2) + \\ + 2 \frac{n-1}{n} (3R_{22} - 2R_{21}R_{11} - R_{11}^2) + \frac{1}{n} (R_{12} - R_{11}^2), \quad (3.8)$$

где $R_{ks} = \sum_{j=1}^N p_j^k / p_j^{s0}$. Отметим частные случаи этой формулы. Если все $p_j = p_j^0$, т. е. дисперсия вычисляется при нулевой гипотезе, то

$$R_{ks} = \sum_{j=1}^N (p_j^0)^{k-s} \text{ и, в частности, } R_{11} = R_{22} = N, \quad R_{21} = R_{32} = 1, \quad R_{12} = \\ = \sum_{j=1}^N 1/p_j^0. \text{ В этом случае из формулы (3.8) имеем}$$

$$D(X_n^2 | p^0) = 2(N-1) + \frac{1}{n} \left(\sum_{j=1}^N 1/p_j^0 - N^2 - 2N + 2 \right).$$

Отсюда, в частности, следует, что $\lim_{n \rightarrow \infty} D(X_n^a | p^0) = 2(N-1)$, что также согласуется с теоремой 3.1.

Пусть теперь p — любой вектор вероятностей, удовлетворяющий условию $p \neq p^0$. Тогда $\sum_{i=1}^N (p_i - p_i^0)^2 / p_i^0 > 0$ и из формул (3.7) — (3.8) следует, что при $n \rightarrow \infty$ среднее и дисперсия статистики X_n^a при гипотезе p имеют порядок роста n . Отсюда на основании неравенства Чебышева имеем

$$1 - W_n(p) = P(X_n^a < \chi_{1-\alpha, N-1}^2 | p) = P(E(X_n^a | p) - X_n^a > E(X_n^a | p) - \chi_{1-\alpha, N-1}^2 | p) \leq P(|E(X_n^a | p) - X_n^a| > |E(X_n^a | p) - \chi_{1-\alpha, N-1}^2| | p) \leq D(X_n^a | p) / [E(X_n^a | p) - \chi_{1-\alpha, N-1}^2]^2 = O\left(\frac{1}{n}\right). \blacksquare$$

3. Критерий согласия хи-квадрат для сложной гипотезы. Метод группировки наблюдений с последующим применением критерия согласия χ^2 применим и в более сложной ситуации, когда требуется проверить гипотезу о принадлежности неизвестной функции распределения наблюдаемой в опыте случайной величины ξ заданному семейству функций распределения. В общем виде задача формулируется так. Пусть $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$ — заданное параметрическое семейство функций распределения (параметр θ может быть как скалярным, так и векторным) и $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{L}(\xi)$ с неизвестной функцией распределения. Требуется проверить гипотезу $H_0: \mathcal{L}(\xi) \in \mathcal{F}$. Таким образом, в данном случае речь идет о проверке сложной гипотезы.

Пусть исходные данные сгруппированы и $v = (v_1, \dots, v_N)$ — соответствующий вектор частот попадания наблюдений в интервалы группировки. Составим статистику, аналогичную (3.5). В данном случае вероятности попадания в интервалы группировки при гипотезе H_0 уже не будут заданы однозначно, а представляют собой некоторые функции от параметра θ :

$$p_j(\theta) = P(\xi \in \mathcal{E}_j | H_0) = \int_{\mathcal{E}_j} dF(x; \theta), \quad j = 1, \dots, N,$$

поэтому статистика X_n^a имеет вид

$$X_n^a = X_n^a(\theta) = \sum_{j=1}^N (v_j - np_j(\theta))^2 / [np_j(\theta)]. \quad (3.9)$$

Эта статистика зависит от неизвестного параметра; следовательно, непосредственно использовать ее для построения критерия пока нельзя — требуется предварительно исключить в (3.9) неопределенность, связанную с неизвестным параметром θ . Для этого поступают следующим образом: заменяют θ некоторой оценкой $\bar{\theta}_n = \bar{\theta}_n(X)$ и получают, таким образом, статистику

$$\bar{X}_n^a = X_n^a(\bar{\theta}_n) = \sum_{j=1}^N (v_j - np_j(\bar{\theta}_n))^2 / [np_j(\bar{\theta}_n)]. \quad (3.10)$$

Эта статистика уже представляет собой функцию только от выборочных данных; следовательно, ее значение можно однозначно вычислить для каждой заданной реализации выборки X .

Если бы распределение статистики $\bar{X}_n^2$ при гипотезе H_0 можно было найти (хотя бы приближенно) и при этом распределение не зависело бы от конкретных функций $F(x; \theta)$, составляющих гипотезу H_0 , то, основываясь на $\bar{X}_n^2$, можно было бы построить критерий согласия для гипотезы H_0 .

В данном случае величины $p_j(\bar{\theta}_n)$ уже не постоянные, а представляют собой функции от выборки (случайные величины). Поэтому теорема 3.1 к статистике $\bar{X}_n^2$ неприменима. Более того, следует ожидать, что распределение этой статистики (даже если оно существует) будет, вообще говоря, зависеть от способа построения оценки $\bar{\theta}_n$. Проблема нахождения предельного (при $n \rightarrow \infty$) распределения для $\bar{X}_n^2$ при этих усложненных условиях впервые была рассмотрена Р. Фишером (1924 г.), который показал, что существуют методы оценивания параметра θ , при которых предельное распределение имеет простой вид, а именно является распределением хи-квадрат с числом степеней свободы $N - 1 - r$, где r — размерность оцениваемого параметра θ . Одним из таких методов оценивания является описанный в п. 2 § 2.5 метод максимального правдоподобия, основанный на частотах $v_1, \dots, v_N$, т. е. когда в качестве $\bar{\theta}_n$ в формуле (3.10) используют мультиномиальную оценку максимального правдоподобия.

Теорема 3.3. Пусть функции $p_j(\theta)$, $j = 1, \dots, N$, $\theta = (\theta_1, \dots, \theta_r)$, $r < N - 1$, удовлетворяют следующим условиям:

- а) $\sum_{j=1}^N p_j(\theta) = 1, \forall \theta \in \Theta$;
- б) $p_j(\theta) \geq c > 0$, $j = 1, \dots, N$, и существуют непрерывные производные $\frac{\partial p_j(\theta)}{\partial \theta_k}$ и $\frac{\partial^2 p_j(\theta)}{\partial \theta_k \partial \theta_l}$, $k, l = 1, \dots, r$;
- в) матрица $\left[\frac{\partial p_j(\theta)}{\partial \theta_k} \right]$ размера $N \times r$ имеет ранг r для всех $\theta \in \Theta$.

Тогда если $\bar{\theta}_n = \hat{\theta}_n$ — мультиномиальная оценка максимального правдоподобия для параметра θ и $\bar{X}_n^2 = X_n^2(\hat{\theta}_n)$, то при $n \rightarrow \infty$

$$\mathcal{L}(\bar{X}_n^2 | H_0) \rightarrow \chi^2(N - r - 1). \quad (3.11)$$

Доказательство этой теоремы можно найти в [13, с. 462—470].

Приведем схему использования критерия согласия χ^2 . Пусть в опыте наблюдается одно из N несовместных событий $A_1, \dots, A_N$ и о вероятностях $p_1, \dots, p_N$ появления этих событий выдвинута гипотеза H_0 : $p_j = p_j(\theta)$, $j = 1, \dots, N$, где $\theta = (\theta_1, \dots, \theta_r) \in \Theta$ — некоторому невырожденному интервалу в R^r , и функции $p_j(\theta)$ удовлетворяют условиям теоремы 3.3 (если в опыте наблюдается случайная величина ξ непрерывного типа, то, как уже было отмечено, задачу сводят к такой дискретной схеме, предварительно группируя данные по N интервалам $\mathcal{E}_1, \dots, \mathcal{E}_N$ и рассматривая

в качестве A_j события $\{\xi \in \mathcal{E}_j\}$. Пусть произведено $n \geq 50$ опытов и наблюдавшиеся частоты $h_1, \dots, h_N$ событий удовлетворяют условиям $h_j \geq 5, j = 1, \dots, N$. Определим значение оценки $\hat{\theta}_n$, решая относительно θ уравнения

$$\sum_{j=1}^N \frac{h_j}{p_j(\theta)} \frac{\partial p_j(\theta)}{\partial \theta_k} = 0, \quad k = 1, \dots, r. \quad (3.12)$$

Вычислим $\hat{p}_j = p_j(\hat{\theta}_n), j = 1, \dots, N$, и найдем значение статистики $\hat{X}_n^2$ по формуле $\hat{X}_n^2 = \sum_{j=1}^N (h_j - n\hat{p}_j)^2 / (n\hat{p}_j)$. Пусть задан уровень значимости α . Определим по таблицам распределения $\chi^2(N - r - 1)$ значение $(1 - \alpha)$ -квантили $\chi_{1-\alpha, N-r-1}^2$ и сравним с ним найденное значение $\hat{X}_n^2$. Если $\hat{X}_n^2 \geq \chi_{1-\alpha, N-r-1}^2$, то гипотезу H_0 отвергают; в противном случае можно только сказать, что гипотеза H_0 не противоречит результатам испытаний. Изложенная теория гарантирует, что используя это правило, можно ошибочно отклонить гипотезу H_0 , когда она истинна, с вероятностью, приближенно равной α .

Замечание. Если данные предварительно группируются, то оценивать параметр θ можно и до группировки наблюдений, например максимизируя по θ функцию правдоподобия $L(X; \theta) = f(X_1; \theta) \dots f(X_n; \theta)$. В этом случае оценка максимального правдоподобия параметра θ использует сами наблюдения X , а не частоты интервалов $\mathcal{E}_1, \dots, \mathcal{E}_N$. Можно было бы ожидать, что такой метод оценивания должен приводить к более точным выводам; кроме того, обычную оценку максимального правдоподобия часто находить гораздо проще, чем решать систему уравнений (3.12). Однако, как показали Чернов и Леман (1954 г.), при таком методе оценивания предельное соотношение (3.11), вообще говоря, уже не имеет места, а поэтому асимптотические результаты не будут иметь такую простую форму [21].

Пример 3.8. (пуассоновская модель, критерий χ^2 для нее). Пусть производится n независимых наблюдений над неотрицательной целочисленной случайной величиной ξ . Требуется проверить гипотезу $H_0: \mathcal{X}(\xi) \in \Pi(\theta)$.

Положим $\mathcal{E}_j = \{j - 1\}, j = 1, \dots, N - 1, \mathcal{E}_N = \{N - 1, N, \dots\}$.

Тогда $p_j(\theta) = f(j - 1; \theta), j = 1, \dots, N - 1, p_N(\theta) = \sum_{k=N-1}^{\infty} f(k; \theta)$,

где $f(k; \theta) = e^{-\theta} \theta^k / k!, k \geq 0$. Найдем оценку $\hat{\theta}_n$. В данном случае неизвестный параметр только один, поэтому система (3.12) сводится к уравнению

$$\sum_{j=0}^{N-2} \left(\frac{j}{\theta} - 1 \right) h_{j+1} + h_N \sum_{k=N-1}^{\infty} \left(\frac{k}{\theta} - 1 \right) f(k; \theta) - \sum_{k=N-1}^{\infty} f(k; \theta) = 0.$$

Отсюда

$$\hat{\theta} = \frac{1}{n} \left[\sum_{j=0}^{N-2} j h_{j+1} + h_N \sum_{k=N-1}^{\infty} k f(k; \hat{\theta}) \right] / \left[\sum_{k=N-1}^{\infty} f(k; \hat{\theta}) \right].$$

Первый член в скобках равен сумме всех x_i таких, что $x_i \leq N-2$, а последний член приближенно равен сумме всех x_i , которые больше или равны $N-1$ ($x_1, \dots, x_n$ — наблюдавшиеся значения ξ). Таким образом, оценкой $\hat{\theta}_n$ для θ может служить среднее арифметическое выборочных значений: $\hat{\theta}_n = \bar{x}$. (Напомним, что обычная оценка максимального правдоподобия для параметра θ пуассоновского распределения точно равна $\bar{x}$.)

Теперь находим

$$\hat{p}_j = \begin{cases} e^{-\bar{x}} \bar{x}^{j-1} / (j-1)! & \text{при } j=1, \dots, N-1, \\ e^{-\bar{x}} \sum_{k=N-1}^{\infty} \frac{\bar{x}^k}{k!} & \text{при } j=N, \end{cases}$$

и гипотеза H_0 отвергается тогда и только тогда, когда

$$\sum_{j=1}^N (h_j - n\hat{p}_j)^2 / (n\hat{p}_j) \geq \chi^2_{1-\alpha, N-2}.$$

4. Критерий квантилей. Метод χ^2 можно применять для построения критериев согласия и в задачах другого типа. Рассмотрим задачу проверки описанной в примере 3.1 гипотезы H_0 о том, что неизвестное распределение $\mathcal{L}(\xi)$ имеет некоторое количество заданных квантилей.

Пусть выборка $\mathbf{X} = (X_1, \dots, X_n)$ извлечена из непрерывного на R^1 распределения $\mathcal{L}(\xi)$, о котором выдвинута гипотеза H_0 : $F_k(\xi_j) = p_j$, $j=1, \dots, N-1$, где $-\infty < \xi_1 < \dots < \xi_{N-1} < \infty$ и $0 < p_1 < \dots < p_{N-1} < 1$ — заданные числа. Гипотеза H_0 сложная и включает все непрерывные распределения с указанными квантилями. Построим критерий проверки этой гипотезы. Положим $\xi_j = (\xi_{j-1}, \xi_j]$, $j=1, \dots, N$, $\xi_0 = -\infty$, $\xi_N = \infty$, и пусть v_j — число выборочных точек X_i , принадлежащих интервалу ξ_j . Таким образом, имеем схему группировки данных с естественными (т. е. порождаемыми самой проверяемой гипотезой) интервалами. Тогда

$$\mathcal{L}(\mathbf{v} = (v_1, \dots, v_N) | H_0) = M(n; p^0 = (p_1^0, \dots, p_N^0)),$$

где $p_j^0 = p_j - p_{j-1}$, $j=1, \dots, N$, $p_0^0 = 0$, $p_N^0 = 1$, и гипотеза H_0 эквивалентна утверждению, что вероятности исходов в построенной полиномиальной схеме равны заданным числам $p_1^0, \dots, p_N^0$. Следовательно, для проверки этой гипотезы можно применить описанный в п. 2 критерий согласия χ^2 , который основан на статистике (3.5). Этот критерий в данном случае называют *критерием квантилей*. При $N=2$ и $p_1=0.5$ соответствующий критерий называют *критерием знаков*. В этом случае гипотеза H_0 — это гипотеза о том, что выборка $\mathbf{X}$ извлечена из распределения $\mathcal{L}(\xi)$ с медианой ξ_1 , а статистика (3.5) сводится к статистике $(4/n)(v_1 - n/2)^2$, где v_1 — число компонент вектора $\mathbf{X}$, значения которых лежат в интервале $(-\infty, \xi_1]$, или, что то же самое, число отрицательных разностей $X_i - \xi_1$, $i=1, \dots, n$.

На практике критерий знаков чаще всего применяют в следующей ситуации. Пусть $((X_1, Y_1), \dots, (X_n, Y_n))$ — выборка из двумерного распределения $\mathcal{L}(\xi = (\xi_1, \xi_2))$. Требуется проверить гипотезу H_0 о том, что компоненты ξ_1 и ξ_2 независимы и одинаково распределены, т. е. что $F_{\xi}(x, y) = F(x)F(y)$, где $F(x)$ — некоторая одномерная функция распределения. Для проверки этой гипотезы поступают следующим образом: составляют разности $Z_i = X_i - Y_i$, $i = 1, \dots, n$. Тогда если гипотеза H_0 верна, то $P(Z_i < 0) = P(Z_i > 0) = 1/2$ и исходная гипотеза сводится к утверждению, что выборка $(Z_1, \dots, Z_n)$ извлечена из распределения, медиана которого равна 0. В этом случае статистика v_1 есть число отрицательных величин среди $Z_1, \dots, Z_n$ и по теореме 3.1 при $n \rightarrow \infty$ $\mathcal{L}(4(v_1 - n/2)^2/n | H_0) \rightarrow \chi^2(1)$, что позволяет рассчитать соответствующую критическую границу в критерии согласия χ^2 .

§ 3.3. Симметрические критерии в схеме группировки с растущим числом интервалов. Критерий пустых ящиков

1. Критерий согласия χ^2 для непрерывных распределений. Вопросы его состоятельности. Вернемся к задаче проверки простой гипотезы $H_0: F_{\xi}(x) = F(x)$ о виде распределения и условимся считать, что функция распределения $F(x)$ непрерывна. В случае применения описанного в п. 2 § 3.2 критерия согласия χ^2 обычно рекомендуется интервалы группировки $\xi_1, \dots, \xi_N$ выбирать *равновероятными* при нулевой гипотезе, т. е. чтобы все $p_j^* = 1/N$. В этом случае

$$\xi_j = (\zeta_{j-1}, \zeta_j], \quad j = 1, \dots, N, \quad \zeta_0 = -\infty, \quad \zeta_N = \infty, \quad (3.13)$$

где $F(\zeta_j) = j/N$, $j = 1, \dots, N-1$, и критерий согласия χ^2 является, по существу, критерием квантилей (см. п. 4 § 3.2). Но этот критерий учитывает лишь вероятности попадания в интервалы группировки и не фиксирует локального поведения функции распределения, поэтому по критерию χ^2 все непрерывные распределения с указанными квантилями не различаются. Таким образом, при применении критерия χ^2 задачу проверки исходной простой гипотезы заменяют задачей проверки некоторой сложной гипотезы, в которую помимо $F(x)$ входит множество отличных от $F(x)$ функций распределения с теми же квантилями $\zeta_1, \dots, \zeta_{N-1}$, что и $F(x)$. Можно сказать, что в случае применения методики критерия χ^2 в задаче проверки гипотезы $H_0: F_{\xi}(x) = F(x)$ получают критерий согласия, не чувствительный к любым таким отклонениям от H_0 , при которых распределение $\mathcal{L}(\xi)$ сохраняет заданные квантили $\zeta_1, \dots, \zeta_{N-1}$, т. е. для любого распределения F_{ξ} , удовлетворяющего условиям $F_{\xi}(\zeta_j) = j/N$, $j = 1, \dots, N-1$, выполняется предельное при $n \rightarrow \infty$ соотношение

$$P\left(X_n^2 = \frac{N}{n} \sum_{j=1}^N \left(v_j - \frac{n}{N}\right)^2 \geq \chi^2_{1-\alpha, N-1} | F_{\xi}\right) \rightarrow \alpha.$$

В частности, критерий χ^2 оказывается несостоятельным против альтернатив F_{ξ} , имеющих данные квантили (в этой связи отметим, что рассмотренный в п. 1 § 3.2 критерий согласия Колмогорова является состоятельным против любых альтернатив $F_{\xi} \neq F$). Например, на рис. 3.1 изображены графики трех функций распределения, у которых точка $x=0$ является медианой. Если для проверки гипотезы $H_0: F_{\xi}(x) = F(x)$ применить методику χ^2 с двумя ($N=2$) равновероятными интервалами, то по соответствующему критерию все три

функции не будут отличимы. Этот недостаток критерия χ^2 можно попытаться «уменьшить», увеличивая вместе с ростом числа наблюдений число интервалов группировки N , сужая тем самым класс неотличимых от $F(x)$ (по критерию χ^2) функций распределения. Однако тогда критерий χ^2 становится трудоемким, так как подсчет и обработка большого (при больших N) числа частот $v_1, \dots, v_N$ — сложная вычислительная задача.

Следует также иметь в виду и то обстоятельство, что теорема 3.1, определяющая предельное распределение статистики X_N^2 при гипотезе H_0 , справедлива только при фиксированных значениях параметров N и $\rho_1^0, \dots, \rho_N^0$ и ее уже нельзя использовать для расцата критической границы критерия, если $N \rightarrow \infty$.

Выход из этой противоречивой ситуации можно найти, сконструировав более простые, чем X_N^2 , статистики критерия и доказав предельные теоремы для распределений этих статистик, которые учитывают одновременное неограниченное возрастание параметров n и N . Изложенные ниже результаты позволяют получить ответ на некоторые возникающие здесь вопросы.

2. Симметрические статистики в схеме группировки. Итак, пусть $\mathcal{E}_1, \dots, \mathcal{E}_N$ — определенная в (3.13) система равновероятных (при гипотезе H_0) интервалов и $v = (v_1, \dots, v_N)$ — соответствующий выборке $X = (X_1, \dots, X_n)$ вектор частот попадания в эти интервалы. Рассмотрим класс статистик вида

$$S_{nN}(g) = \sum_{j=1}^N g(v_j), \quad (3.14)$$

где $g(x)$ — заданная функция, определенная во всех целых неотрицательных точках $x=0, 1, 2, \dots$. В выражение для $S_{nN}(g)$ все частоты $v_1, \dots, v_N$ входят симметрично, поэтому будем называть такие статистики *симметрическими*. Если в (3.14) в качестве $g(x)$ взять функцию $g(x) = \begin{cases} 1 & \text{при } x=r, \\ 0 & \text{при } x \neq r \end{cases}$ для некоторого целого $r \in \{0, 1, \dots, n\}$, то $S_{nN}(g)$, очевидно, равно числу интервалов среди $\mathcal{E}_1, \dots, \mathcal{E}_N$, каждый из которых содержит ровно r выборочных точек (элементов выборки X); такую статистику обозначают символом $\mu_r = \mu_r(n, N)$. Таким образом, с любой схемой группировки связан набор симметрических статистик $\mu_0, \mu_1, \dots, \mu_n$. Эти статистики удовлетворяют двум очевидным линейным соотношениям:

$$\sum_{r=0}^n \mu_r(n, N) = N, \quad \sum_{r=0}^n r \mu_r(n, N) = n. \quad (3.15)$$

В терминах статистик μ_r формулу (3.14) можно переписать в виде

$$S_{nN}(g) = \sum_{r=0}^n g(r) \mu_r,$$

откуда следует, что любая симметрическая статистика является линейной комбинацией статистик $\mu_0, \mu_1, \dots, \mu_n$. Верно и обратное: любая линейная комбинация этих статистик — симметрическая статистика. Действительно,

$\sum_{r=0}^n c_r \mu_r = \sum_{j=1}^N g(v_j)$, где $g(r) = c_r$, $r=0, 1, \dots, n$. Таким образом, класс симметрических статистик совпадает с классом всех линейных комбинаций величин $\mu_0, \mu_1, \dots, \mu_n$.

Обметим, в частности, что статистика X_N^2 [см. (3.5)] в случае равновероятных интервалов — симметрическая статистика, для которой $g(x) = Nx^2/n - x$.

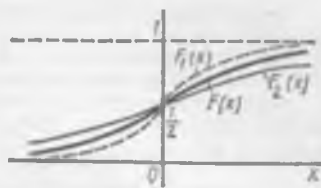


Рис. 3.1

3. Критерий пустых ящиков. Простейшей симметрической статистикой является величина μ_0 , которая определяет число интервалов, не содержащих ни одной выборочной точки. Критерий, основанный на этой статистике, называют *критерием пустых ящиков*.

Вычислим первые два момента этой статистики для произвольного распределения наблюдений F_ξ . Введем для этого случайные величины $\eta_1, \dots, \eta_N$, положив, по определению,

$$\eta_j = \begin{cases} 1, & \text{если интервал } \mathcal{E}_j \text{ пуст,} \\ 0 & \text{в остальных случаях.} \end{cases}$$

Очевидно, что $\mu_0 = \eta_1 + \dots + \eta_N$, откуда

$$E\mu_0 = \sum_{j=1}^N E\eta_j = \sum_{j=1}^N P(\eta_j = 1),$$

$$\begin{aligned} D\mu_0 &= \sum_{j=1}^N D\eta_j + 2 \sum_{i < j} \text{cov}(\eta_i, \eta_j) = \\ &= \sum_{j=1}^N P(\eta_j = 1)(1 - P(\eta_j = 1)) + 2 \sum_{i < j} [P(\eta_i = \eta_j = 1) - P(\eta_i = 1)P(\eta_j = 1)]. \end{aligned}$$

Испытания независимы, поэтому $P(\eta_j = 1) = (1 - p_j)^n$, $P(\eta_i = \eta_j = 1) = (1 - p_i - p_j)^n$. Следовательно,

$$E\mu_0 = \sum_{j=1}^N (1 - p_j)^n, \quad D\mu_0 = 2 \sum_{i < j} (1 - p_i - p_j)^n + E\mu_0 - (E\mu_0)^2. \quad (3.16)$$

С помощью метода неопределенных множителей Лагранжа легко проверить, что среднее $E\mu_0$ как функция параметров $p_1, \dots, p_N$ достигает минимума при $p_1 = \dots = p_N = 1/N$, т. е. при $F_\xi = F$. В этом случае формулы (3.16) принимают соответственно вид

$$\begin{aligned} E\mu_0 &= N(1 - 1/N)^n, \\ D\mu_0 &= N(N-1)(1 - 2/N)^n + N(1 - 1/N)^n - N^2(1 - 1/N)^{2n}. \end{aligned} \quad (3.17)$$

Таким образом, при любом отклонении от нулевой гипотезы, таком, что вероятности попадания в интервалы группировки не все равны $1/N$, статистика μ_0 имеет тенденцию увеличиваться, т. е. слишком большие значения μ_0 «свидетельствуют» против гипотезы H_0 . Отсюда следует, что критическую область критерия следует задавать в виде

$$\{\mu_0 \geq t_\alpha(n, N)\}. \quad (3.18)$$

Для нахождения критической границы $t_\alpha(n, N)$ надо знать распределение статистики μ_0 при гипотезе H_0 . Точное ее распределение имеет сложный вид и неудобно для практических расчетов. Однако если n и N одновременно велики, то случайная величина μ_0 имеет простое предельное распределение при любом абсолютно непрерывном распределении наблюдений F_ξ (а не только для $F_\xi = F$). Это предельное распределение дается в приводимой ниже теореме 3.4 (доказательство теоремы можно найти, например, в [8]). Без ограничения общности все распределения можно рассматривать заданными на отрезке $[0, 1]$, а в качестве распределения при нулевой гипотезе использовать равномерное распределение $R(0, 1)$. Действительно, если это не так, то всегда можно перейти от исходной случайной величины ξ к случайной величине $\eta = F(\xi)$, принимающей значения уже из $[0, 1]$. Тогда $\mathcal{L}(\eta | H_0) = R(0, 1)$ и для любого альтернативного распределения F_ξ

$$P(\eta \leq x | F_\xi) = P(F(\xi) \leq x | F_\xi) = P(\xi \leq F^{-1}(x) | F_\xi) = F_\xi(F^{-1}(x)), \quad x \in [0, 1].$$

Таким образом, если сформулирована некоторая нулевая гипотеза $H_0: F_x = F$, то с помощью указанного преобразования случайной величины ξ задачу можно свести к проверке нулевой гипотезы о равномерном распределении на отрезке $[0, 1]$. В этом случае система интервалов (3.13) принимает вид

$$X_j = \left(\frac{j-1}{N}, \frac{j}{N} \right], \quad j = 1, \dots, N.$$

Собозначим через $p_j = \int_{(j-1)/N}^{j/N} f_x(x) dx$, $j = 1, \dots, N$, вероятности попадания наблюдений в интервалы группировки для распределения с плотностью $f_x(x)$. В дальнейшем предполагается, что все допустимые плотности $f_x(x) > 0$ и непрерывны.

Теорема 3.4. Пусть $n, N \rightarrow \infty$ так, что $n/N \rightarrow \rho > 0$. Тогда

$$X(\mu_0(n, N) | f_x) \sim f^*(Nm(f_x), N\sigma^2(f_x)),$$

где

$$\begin{aligned} m(f_x) &= \int_0^1 e^{-\rho f_x(x)} dx, \\ \sigma^2(f_x) &= \int_0^1 e^{-\rho f_x(x)} (1 - e^{-\rho f_x(x)}) dx - \rho \left| \int_0^1 f_x(x) e^{-\rho f_x(x)} dx \right|^2. \end{aligned} \quad (3.19)$$

Для нулевой гипотезы $f_x(x) \equiv 1$, $x \in [0, 1]$, поэтому утверждение теоремы 3.4 принимает вид

$$X\left(\frac{\mu_0(n, N) - Ne^{-\rho}}{\sqrt{Ne^{-\rho}(1 - e^{-\rho}(1 + \rho))}} | H_0\right) \rightarrow f^*(0, 1). \quad (3.20)$$

Этот результат позволяет приближенно рассчитать критическую границу в (3.18) при больших n и N и заданном уровне значимости α . В самом деле, из (3.20) следует, что если в соотношении (3.18) выбрать

$$t_\alpha(n, N) = Ne^{-\rho} + \sqrt{Ne^{-\rho}(1 - e^{-\rho}(1 + \rho))} t_\alpha, \quad \Phi(-t_\alpha) = \alpha, \quad (3.21)$$

то при $n, N \rightarrow \infty$, $n/N \rightarrow \rho > 0$ критерий (3.18) будет иметь в пределе уровень значимости α . Соотношения (3.18) и (3.21) и определяют критерий пустых ящиков, который формулируется следующим образом: если наблюдающееся число пустых интервалов удовлетворяет неравенству $\mu_0 \geq t_\alpha(n, N)$, где $t_\alpha(n, N)$ задано в (3.21), то гипотезу H_0 отвергают; в противном случае считают, что данные не противоречат гипотезе. Следуя этому правилу можно ошибиться, отвергнув истинную гипотезу с вероятностью, приблизительно равной α , если параметры n и N достаточно велики и $n/N = \rho > 0$.

4. Асимптотическое поведение мощности критерия пустых ящиков. Теорема 3.4 позволяет не только найти критическую границу в критерии пустых ящиков, но и исследовать предельное поведение мощности $W_n(f)$ критерия при произвольной альтернативе, задаваемой плотностью $f(x) = f_x(x) \neq 1$. Из (3.18) и (3.21) по этой теореме имеем

$$\begin{aligned} W_n(f) &= P(\mu_0(n, N) \geq t_\alpha(n, N) | f) = \\ &= P\left(\frac{\mu_0(n, N) - Nm(f)}{\sqrt{N\sigma(f)}} \geq z_\alpha(N) | f\right) = \Phi(z_\alpha(N)), \end{aligned} \quad (3.22)$$

$$\text{где } z_\alpha(N) = \sqrt{N} \frac{m(f) - e^{-\rho}}{\sigma(f)} - \frac{t_\alpha \sigma_0}{\sigma(f)}, \quad \sigma_0^2 = e^{-\rho}(1 - e^{-\rho}(1 + \rho)).$$

Далее, для любой неотрицательной случайной величины η с конечным математическим ожиданием $E\eta \geq \exp\{E \ln \eta\}^*$; равенство имеет место только

* Это неравенство является частным случаем неравенства Пенсена: для любой случайной величины η и любой выпуклой вниз функции $g(x)$ имеет место неравенство $g(E\eta) \leq Eg(\eta)$ (предполагается, что указанные математические ожидания существуют).

в случае, когда $\eta = \text{const}$. Полагая $\eta = e^{-\rho/(1-\rho)}$, получаем при $Z(\frac{\rho}{1-\rho}) = R(0, 1)$

$$\int_0^1 e^{-\rho/(1-x)} dx \geq \exp \left\{ -\rho \int_0^1 f(x) dx \right\} = e^{-\rho};$$

равенство имеет место только в случае, когда $f(x) \equiv \text{const} = 1$ (так как $f(x)$ — это функция плотности на отрезке $[0, 1]$). Отсюда и из (3.19) следует, что для любой альтернативной плотности $f(x) \neq 1$ разность $m(f) - e^{-\rho} > 0$, поэтому $z_\alpha(N)$ в (3.22) при $N \rightarrow \infty$ неограниченно возрастает. В силу соотношения (3.22) это означает, что $\lim_{N \rightarrow \infty} \Psi_N(f) = 1$, т. е. критерий пустых ящиков является состоятельным.

Итак, если выполнены условия теоремы 3.4, то мощность критерия пустых ящиков при любой фиксированной альтернативе стремится к 1. Однако если при изменении n и N изменяется также альтернатива, «приближаясь» к основной гипотезе H_0 , то мощность критерия уже не обязательно будет сходиться к 1 — это зависит от скорости сближения альтернативы с нулевой гипотезой.

Теорема 3.4 дает возможность ответить и на вопрос о том, с какой скоростью может сближаться альтернатива с нулевой гипотезой, чтобы критерий пустых ящиков обладал еще способностью «реагировать» на такие отклонения от H_0 . В данном случае альтернативами являются абсолютно непрерывные распределения на $[0, 1]$, отличающиеся от равномерного распределения $R(0, 1)$; следовательно, «близкую» альтернативу F_n можно задать, например, плотностью

$$f_n(x) = 1 + a(n)b(x), \quad (3.23)$$

где $a(n) \rightarrow 0$ при $n \rightarrow \infty$, а $b(x)$ — непрерывная функция, удовлетворяющая условию $\int_0^1 b(x) dx = 0$. Здесь $a(n)$ и определяет скорость сближения (в смысле поточечной сходимости) альтернативной плотности $f_n(x)$ с плотностью при нулевой гипотезе.

Следующее утверждение является простым следствием теоремы 3.4.

Теорема 3.5. Если в соотношении (3.23) положить $a(n) = n^{-1/4}$, то при $n, N \rightarrow \infty, n/N \rightarrow \rho > 0$

$$\Psi_n(f_n) \rightarrow \Phi(\Delta(b^2, \rho) - I_\alpha), \quad (3.24)$$

$$\text{где } \Delta(b^2, \rho) = \frac{b^2}{2} \rho^{3/2} / \sqrt{e^\rho - 1 - \rho}, \quad b^2 = \int_0^1 b^2(x) dx.$$

Из этой теоремы следует, что критерий пустых ящиков позволяет отличать от гипотезы H_0 альтернативы, сближающиеся с ней со скоростью $n^{-1/4}$. Более близкие альтернативы вида (3.23) (т. е. при $a(n)n^{1/4} \rightarrow 0$) этот критерий «не отличает» (т. е. $\Psi_n(f_n) \rightarrow \alpha$), а для более удаленных альтернатив (т. е. при $a(n)n^{-1/4} \rightarrow \infty$) он сохраняет свойство состоятельности (т. е. $\Psi_n(f_n) \rightarrow 1$).

5. Общие симметрические критерии. Критерий пустых ящиков весьма прост для практического применения, однако статистика μ_0 включает далеко не всю информацию, содержащуюся в выборке $X = (X_1, \dots, X_n)$. Очевидно, более полно статистическая информация будет использована, если рассматривать статистику вида $\sum_{r=0}^n c_r \mu_r$, где $c_0, c_1, \dots, c_n$ — некоторые «веса», выбираемые

из следующего условия: соответствующий критерий должен иметь наибольшую асимптотическую мощность в классе всех симметрических критериев. Более подробно с теорией симметрических критериев можно ознакомиться в [8]. Отметим только, что в условиях теоремы 3.5 наибольшую асимптотическую мощность среди всех симметрических критериев имеет известный уже критерий

χ^2 , определяемый статистикой $X_n^2 = \frac{N}{n} \sum_{j=1}^N v_j^2 - n = \frac{N}{n} \sum_{r=0}^n r^2 \mu_r - n$. Критическая граница для этого критерия имеет вид $t_\alpha(n, N) = N + \sqrt{2N} t_\alpha$, $\Phi(-t_\alpha) = \alpha$, а предельная мощность равна $\Phi(b^2 \sqrt{\rho/2} - t_\alpha)$.

Можно сравнить, насколько уступает по мощности критерий пустых ящиков оптимальному критерию χ^2 . Из (3.24) следует, что

$$\Delta(b^2, \rho) = b^2 \sqrt{\rho/2} e(\rho),$$

где $e(\rho) = \rho / \sqrt{2(e^\rho - 1 - \rho)}$. Как функция от ρ величина $e(\rho)$ монотонно убывает от 1 при $\rho=0$ до 0 при $\rho=\infty$, т. е. при малых значениях параметра ρ простейший из симметрических критериев — критерий пустых ящиков — по мощности мало уступает оптимальному, но сравнительно трудоемкому для практической реализации критерию χ^2 . В заключение приведем краткую таблицу значений функции $e(\rho)$.

Таблица 3.1

ρ	0,05	0,10	0,15	0,20	0,30	0,40	0,50	0,60	0,70	0,80
$e(\rho)$	0,992	0,983	0,975	0,967	0,950	0,933	0,917	0,900	0,884	0,867

ρ	0,90	1,00	1,10	1,20	1,30	1,40	1,50	1,60	1,80	2,00
$e(\rho)$	0,851	0,834	0,818	0,802	0,786	0,769	0,753	0,738	0,706	0,675

§ 3.4. Гипотеза однородности

Одной из важных прикладных задач математической статистики является задача проверки однородности статистического материала. Пусть имеются две независимые выборки $X = (X_1, \dots, X_n)$ и $Y = (Y_1, \dots, Y_m)$, описывающие один и тот же процесс, явление и т. д., но полученные в разное время или, вообще говоря, в разных условиях; требуется установить, являются ли они выборками из одного и того же распределения или же закон распределения наблюдений от выборки к выборке меняется. Такая задача может возникнуть, например, при контроле качества некоторой продукции, когда по контрольным выборкам из различных партий надо установить, не менялось ли ее качество от смены к смене или в результате изменения технологического процесса и т. д. В общем случае можно рассматривать произвольное конечное число независимых выборок.

В общем виде задачу можно сформулировать следующим образом. Пусть $X = (X_1, \dots, X_n)$ — выборка из распределения $\mathcal{L}(\xi)$ с некоторой (неизвестной) функцией распределения $F_1(x)$, а $Y = (Y_1, \dots, Y_m)$ — выборка из распределения $\mathcal{L}(\eta)$ с неизвестной функцией распределения $F_2(x)$. Требуется проверить гипотезу

однородности $H_0: F_1(x) = F_2(x)$. Рассмотрим несколько способов построения критерия согласия для этой гипотезы.

1. Критерий однородности Смирнова. Одним из критериев проверки гипотезы однородности является критерий Смирнова, который применяют в случае непрерывных распределений. Этот критерий основан на статистике $\mathcal{D}_{nm} = \sup_{-\infty < x < \infty} |F_{1n}(x) - F_{2m}(x)|$, где $F_{1n}(x)$ и $F_{2m}(x)$ — эмпирические функции распределения, построенные по выборкам X и Y соответственно. Эмпирическая функция распределения является оптимальной оценкой для теоретической функции распределения, и с увеличением объема выборки они сближаются, поэтому в случаях, когда справедлива гипотеза H_0 , функции $F_{1n}(x)$ и $F_{2m}(x)$ оценивают одну и ту же неизвестную функцию распределения. Таким образом, в этих случаях (по крайней мере при больших n и m) статистика $\mathcal{D}_{nm}$ не должна существенно отклоняться от нуля. Отсюда следует, что если имеют место слишком большие значения этой статистики, то этот факт следует расценивать как свидетельство против нулевой гипотезы H_0 . Таким образом, в данном случае разумно выбирать критическую область в виде $\mathcal{T}_1 = \{t \geq t_\alpha(n, m)\}$. Критическую границу $t_\alpha(n, m)$ при заданном уровне значимости α находят на основании известного при гипотезе H_0 предельного при $n, m \rightarrow \infty$ распределения статистики $\mathcal{D}_{nm}$. По теореме Смирнова при больших n и m можно положить $t_\alpha(n, m) = \sqrt{(n+m)/(nm)} \lambda_\alpha$, где $\mathcal{K}(\lambda_\alpha) = 1 - \alpha$ [функция $\mathcal{K}(l)$ определена в (1.8)]. Действительно, в этом случае

$$P(\mathcal{D}_{nm} \in \mathcal{T}_1 | H_0) = P\left(\sqrt{\frac{nm}{n+m}} \mathcal{D}_{nm} \geq \lambda_\alpha | H_0\right) \approx 1 - \mathcal{K}(\lambda_\alpha) = \alpha.$$

Сформулируем критерий однородности Смирнова: если объемы выборок достаточно велики, то, вычислив по выборочным данным значение t статистики $\mathcal{D}_{nm}$, принимают решение отвергнуть гипотезу H_0 в том и только том случае, когда $t \geq \sqrt{(n+m)/(nm)} \lambda_\alpha$. Вероятность ошибочно отвергнуть при этом истинную гипотезу приблизительно равна α . Отметим также, что этот критерий состоятельный [альтернативой здесь является любая пара распределений (F_1, F_2) такая, что $F_1(x) \neq F_2(x)$].

Указанное правило проверки неизменности функции распределения не зависит от конкретного вида функции. Для приложений это имеет важное значение, так как истинное распределение наблюдаемой случайной величины, как правило, бывает неизвестно, а интерес представляет вопрос о том, не изменялось ли неизвестное распределение от выборки к выборке. Для применения критерия Смирнова необходимо выполнение только условия непрерывности, которое обычно вытекает из физической природы изучаемого явления и не требует специальной проверки.

2. Критерий однородности χ^2 . Часто применяемым критерием является критерий однородности χ^2 . Его используют для проверки однородности данных, имеющих дискретную структуру.

т. е. когда в опытах наблюдается некоторый переменный признак, принимающий конечное число, например s , различных значений. Но, как известно, к такой схеме можно свести любую другую модель, применяя предварительно метод группировки данных. Поэтому метод χ^2 применим на самом деле к анализу любых данных, т. е. является в этом смысле универсальным. Кроме того, с помощью этого метода можно анализировать одновременно любое конечное число выборок, в то время как с помощью критерия Смирнова можно сравнивать только две выборки.

Итак, предположим, что осуществлено k последовательных серий независимых наблюдений, состоящих из $n_1, \dots, n_k$ наблюдений соответственно. При этом в каждом опыте наблюдается некоторый переменный признак, принимающий одно из s различных значений (исходов). Пусть v_{ij} — число реализаций i -го исхода в j -й серии, так что

$$\sum_{i=1}^s v_{ij} = n_j, \quad j = 1, \dots, k.$$

Требуется проверить гипотезу H_0 о том, что все наблюдения производились над одной и той же случайной величиной. Другими словами, если p_{ij} — (неизвестная) вероятность появления i -го исхода в испытаниях j -й серии ($i = 1, \dots, s$; $j = 1, \dots, k$), то гипотеза H_0 означает утверждение: $(p_{1j}, \dots, p_{sj}) = (p_1, \dots, p_s)$, $j = 1, \dots, k$, где $p = (p_1, \dots, p_s)$ — некоторый (неизвестный) вектор вероятностей ($p_1 + \dots + p_s = 1$).

Так как $E(v_{ij} | H_0) = n_j p_i$, то, следуя принципу χ^2 , в качестве меры отклонения опытных данных от их гипотетических (при гипотезе H_0) значений в данном случае следовало бы выбрать статистику

$$\chi_s^2(p) = \sum_{i=1}^s \sum_{j=1}^k (v_{ij} - n_j p_i)^2 / (n_j p_i). \quad (3.25)$$

Однако здесь $p_1, \dots, p_s$ неизвестны, поэтому, чтобы воспользоваться данной статистикой, следует предварительно оценить неизвестные параметры. Для этого воспользуемся методом максимального правдоподобия. Здесь функция правдоподобия (при гипотезе H_0) равна $L(p) = c \prod_{i,j} p_i^{v_{ij}} = c \prod_i p_i^{v_i}$, $v_i = \sum_{j=1}^k v_{ij}$ (c от параметров p_i не зависит). Применяя метод неопределенных множителей Лагранжа, получаем, что оценки максимального правдоподобия $\hat{p}_i$ параметров p_i таковы: $\hat{p}_i = v_i/n$, $i = 1, \dots, s$, где $n =$

$= n_1 + \dots + n_k = \sum_{i,j} v_{ij}$ — общее число наблюдений.

Таким образом, получена следующая статистика критерия:

$$\chi_s^2(p) = n \sum_{i=1}^s \sum_{j=1}^k \frac{(v_{ij} - n_j v_i/n)^2}{n_j v_i} = n \left(\sum_{i=1}^s \sum_{j=1}^k \frac{v_{ij}}{n_j v_i} - 1 \right). \quad (3.26)$$

Критическую область задают в виде $\mathcal{F}_{1\alpha} = \{t \geq t_\alpha\}$, а для нахождения критической границы t_α применяют следующий предельный результат [13, с. 483], аналогичный теореме 3.2: при $n \rightarrow \infty$

$$\mathcal{L}(X_n^s(\hat{p}) | H_0) \rightarrow \chi^2((s-1)(k-1)).$$

В силу этого предельного соотношения при больших n можно полагать $t_\alpha = \chi_{1-\alpha, (s-1)(k-1)}^2$. Окончательно критерий однородности χ^2 имеет следующий вид: гипотезу однородности H_0 отвергают тогда и только тогда, когда вычисленное по фактическим данным значение t статистики (3.26) удовлетворяет неравенству $t \geq \chi_{1-\alpha, (s-1)(k-1)}^2$. Вероятность ошибочно отклонить при этом истинную гипотезу приблизительно равна α , если n достаточно велико. Эту же методику можно использовать и для проверки гипотезы о том, что k серий наблюдений произведены над одной и той же случайной величиной, имеющей распределение заданного типа, например распределение Пуассона, нормальное и т. п. В этом случае предварительно следует найти мультиномиальную оценку максимального правдоподобия $\hat{\theta}_n$ параметров распределения при гипотезе H_0 (т. е. рассматривая все данные как одну выборку с групповыми частотами $v_1, \dots, v_s$) и заменить в (3.25) $p_i = p_i(\theta)$ на $p_i(\hat{\theta}_n)$. Число степеней свободы в предельном распределении χ^2 заменяют при этом на $(s-1)k - r$, где r — число параметров, определяющих гипотетическое распределение (размерность параметрического вектора θ).

Выделим два важных частных случая общей ситуации. Случай $s=2$ соответствует испытаниям с двумя исходами A и $\bar{A}$, а гипотеза однородности представляет собой утверждение, что событие A имеет во всех испытаниях одну и ту же постоянную (хотя и неизвестную) вероятность реализации p . В этом случае оценкой для p является относительная частота появления события A во

всей совокупности данных: $\hat{p} = 1 - \hat{q} = \frac{1}{n} \sum_{j=1}^k v_j$, где v_j — число появлений события A в испытаниях j -й серии, а статистика (3.26) принимает вид

$$X_n^s(\hat{p}) = \frac{1}{p} \sum_{j=1}^k \frac{v_j^2}{n_j} + \frac{1}{q} \sum_{j=1}^k \frac{(n_j - v_j)^2}{n_j} - n = \frac{1}{p\hat{q}} \sum_{j=1}^k \frac{v_j}{n_j} - n \frac{\hat{p}}{q}.$$

Для случая двух выборок ($k=2$) статистика (3.26) принимает следующий вид:

$$X_n^2(\hat{p}) = n_1 n_2 \sum_{i=1}^2 \frac{1}{v_{i1} + v_{i2}} (v_{i1}/n_1 - v_{i2}/n_2)^2.$$

Если положить $\omega_i = v_{i1}/(v_{i1} + v_{i2})$, $\omega = n_1/(n_1 + n_2)$, то последнее выражение можно преобразовать к виду, более удобному для

практических вычислений:

$$X_n^2(\hat{p}) = \frac{1}{\omega(1-\omega)} \left(\sum_{i=1}^s \omega_i v_{i1} - \omega n_1 \right).$$

Описанный критерий однородности χ^2 является состоятельным, т. е. с вероятностью, стремящейся к 1 при $n \rightarrow \infty$, он «улавливает» любые отклонения от нулевой гипотезы, при которых вероятности появления исходов от серии к серии не сохраняют постоянного значения.

3. Другие критерии однородности для двух выборок из непрерывных распределений. Кроме двух описанных критериев для проверки гипотезы однородности разработаны и другие методы, основанные на различных принципах. Приведем краткий обзор некоторых из них, применяемых в случае двух выборок из непрерывных распределений.

а) Критерий знаков. Простой для применения критерий однородности, не требующий сложных вычислений, представляет собой описанный в п. 4 § 3.2 критерий знаков. Основным недостатком этого критерия является «неэкономное» использование информации, содержащейся в результатах наблюдений, поэтому его обычно рекомендуют применять только на стадии предварительного анализа. Кроме того, этот критерий можно использовать лишь для выборок одинакового объема.

б) Критерий пустых блоков. Целый класс критериев однородности можно построить, исходя из следующих соображений. Рассмотрим вариационный ряд $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ выборки $X = (X_1, \dots, X_n)$. Он порождает естественное разбиение оси Ox на интервалы $B_i = (X_{(i-1)}, X_{(i)})$, $i = 1, \dots, n+1$ ($X_0 = -\infty$, $X_{(n+1)} = \infty$), которые называются *выборочными блоками*. Пусть $s_r = s_r(n, m)$ — число этих блоков, каждый из которых содержит ровно r элементов второй выборки $Y = (Y_1, \dots, Y_m)$, $r = 0, 1, \dots, m$. Тогда в качестве статистики критерия можно взять произвольную линейную комбинацию $S_l(n, m) =$

$$= \sum_{r=0}^m c_r s_r, \text{ где } c_0, c_1, \dots, c_l \text{ — некоторые заранее выбранные «веса»}. \text{ В частности, при } l=0 \text{ получаем критерий пустых блоков, статистикой которого является } s_0 \text{ — число блоков, не содержащих ни одного элемента второй выборки.}$$

Основным при построении и расчете асимптотического варианта этого критерия является следующее утверждение о предельном распределении статистики s_0 [19, с. 453]: *если $n, m \rightarrow \infty$ так, что $m/n \rightarrow \rho > 0$, то*

$$\mathcal{L}(s_0(n, m) | H_0) \sim \mathcal{N}(n/(1+\rho), n\rho^2/(1+\rho)^3).$$

При выполнении этих условий критерий пустых блоков формулируется следующим образом: *гипотезу однородности H_0 отвергают тогда и только тогда, когда*

$$s_0(n, m) \geq \frac{n}{1+\rho} + \sqrt{n} \frac{\rho}{(1+\rho)^{3/2}} t_\alpha, \quad \Phi(-t_\alpha) = \alpha.$$

Доказано [19, с. 454], что этот критерий является состоятельным против альтернатив (F_1, F_2) , $F_1 \not\equiv F_2$, удовлетворяющих следующему требованию: функция $F_2(F_1^{-1}(u))$, $u \in [0, 1]$, имеет производную $g(u)$, отличную от 1 на множестве положительной левовой меры (если справедлива нулевая гипотеза H_0 , то $g(u) \equiv 1$, $u \in [0, 1]$).

в) Критерий серий. В ряде случаев особый интерес представляют такие отклонения от нулевой гипотезы $H_0: F_1(x) \equiv F_2(x)$, когда одно распределение сдвинуто относительно другого, т. е. если, например, $F_1(x) > F_2(x)$. В этом случае говорят, что случайная величина η «стохастически больше», чем ξ (при каждом x случайная величина η с большей вероятностью превосходит x , чем ξ). Простой критерий, хорошо «улавливающий» такие отклонения, можно построить

следующим образом. Объединим обе выборки X и Y в одну $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ объема $n+m$ и построим вариационный ряд объединенной выборки. После этого заменим все элементы выборки X буквой C , а все элементы выборки Y — буквой S . В результате получим некоторую последовательность из n букв S и m букв C . Число всех таких последовательностей равно C_{n+m}^n , и интуитивно ясно, что при гипотезе H_0 (элементы обеих выборок X и Y должны вести себя одинаково), все возможные последовательности из X и Y равновероятны. Описанные выше отклонения от нее приводят к тому, что с повышенной вероятностью будут наблюдаться последовательности, в которых элементы одного вида имеют тенденцию смещаться к какому-нибудь краю последовательности (в рассмотренном выше примере буквы S смещаются к правому краю последовательности). Одной из статистик, с помощью которой можно количественно характеризовать степень перемешивания в получаемой последовательности букв S и C , является число серий $W(n, m)$, составленных из этих букв. *Серией* называется участок последовательности, состоящий из подряд идущих одинаковых букв и ограниченный с обеих сторон (или с одной стороны, если речь идет о концах последовательности) буквами другого вида. Число серий будет мало, если одинаковые буквы группируются в одном месте, поэтому в данном случае следует задать критическую область в виде $\mathcal{S}_{1\alpha} = \{t \leq t_\alpha(n, m)\}$. Критерий, основанный на статистике $W(n, m)$ и задаваемый такой критической областью, называется *критерием серий* [предложен в 1940 г. Вальдом и Вольфовитцем].

Асимптотический вариант этого критерия можно рассчитать на основании следующего утверждения [19, с. 459]: если $n, m \rightarrow \infty$ так, что $m/n \rightarrow \rho > 0$, то

$$\mathcal{L}(W(n, m) | H_0) \sim t^2 \left(\frac{2n\rho}{1+\rho}, \frac{4n\rho^2}{(1+\rho)^3} \right).$$

При этих условиях критерий серий формулируется следующим образом: *гипотезу однородности H_0 отвергают тогда и только тогда, когда*

$$W(n, m) \leq \frac{2n\rho}{1+\rho} + \sqrt{n} \frac{2\rho}{(1+\rho)^{3/2}} t_\alpha, \quad \Phi(-t_\alpha) = \alpha.$$

Построенный критерий является состоятельным против тех же альтернатив, что и рассмотренный в п. б) критерий пустых блоков.

г) *Ранговые критерии*. Иногда исходная статистическая информация может быть задана не числовыми значениями наблюдений, а отношением порядка между ними (типа «больше-меньше»). Особенно часто это имеет место в психологических исследованиях. В таких случаях наблюдения *ранжируют*, т. е. упорядочивают по степени их предпочтения. Номер места, которое занимает наблюдение в таком упорядоченном ряду, называют *рангом* соответствующего наблюдения. Таким образом, статистическая информация с самого начала может быть задана рангами наблюдений, статистические методы, которые применяют в таких ситуациях, называют *ранговыми методами*, статистики, являющиеся функциями только рангов, — *ранговыми статистиками*, а критерии, основанные на таких статистиках, — *ранговыми критериями*.

Ранговые методы можно применять и в тех случаях, когда заданы числовые значения наблюдений, т. е. выборка $X = (X_1, \dots, X_n)$, так как при этом всегда можно упорядочить элементы выборки, построив соответствующий вариационный ряд. В этом случае рангом i -го наблюдения (случайной величины X_i) является номер места R_i , которое занимает X_i в вариационном ряду $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Теория ранговых критериев является важной и хорошо разработанной частью математической статистики, она достаточно полно изложена в [4]. В дальнейшем будут рассмотрены примеры ранговых критериев. Одним из таких критериев в задаче проверки гипотезы однородности является *критерий Вилкоксона*.

Рассмотрим, как и в предыдущем случае, объединенную выборку $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ и построим ее вариационный ряд. Пусть $R_1, \dots, R_n$ — ранги величин $X_1, \dots, X_n$ соответственно и $T = R_1 + \dots + R_n$. Таким образом, T — сумма номеров мест, которые занимают в общем вариационном ряду эле-

менты первой выборки. Критерий, основанный на ранговой статистике T , и есть критерий Вилкоксона. Он был впервые предложен Вилкоксоном (1945 г.) для выборок одинакового объема и распространен на случай выборок произвольных объемов Манном и Уитни (1947 г.).

Введем случайные величины $Z_{rs} = \begin{cases} 1 & \text{при } X_r < Y_s, \\ 0 & \text{в противном случае} \end{cases}$ и положим

$$U = U(n, m) = \sum_{r=1}^n \sum_{s=1}^m Z_{rs}, \quad (3.27)$$

так что U — общее число тех случаев, когда элементы выборки X предшествуют в общем вариационном ряду элементам выборки Y .

Можно показать, что статистики T и U связаны линейным соотношением $T + U = nm + n(n+1)/2$, поэтому критерий Вилкоксона эквивалентен критерию, основанному на статистике U .

Вычислим первые два момента статистики U при любой гипотезе, задаваемой парой непрерывных функций распределения (F_1, F_2) . Из (3.27) имеем

$$EU = nmEZ_{11} = nmP(X_1 < Y_1) = nm \int_{-\infty}^{\infty} F_1(x) dF_2(x) = nma. \quad (3.28)$$

В частности, для случая $F_1(x) \equiv F_2(x)$ (т. е. при нулевой гипотезе) $a = 1/2$. Аналогично можно получить, что

$$DU = nm[a + (n-1)b + (m-1)c - (n+m-1)a^2], \quad (3.29)$$

$$\text{где } b = \int_{-\infty}^{\infty} F_1^2(x) dF_2(x), \quad c = \int_{-\infty}^{\infty} (1 - F_2(x))^2 dF_1(x).$$

При нулевой гипотезе $b = c = 1/3$, поэтому $DU = nm(n+m+1)/12$. Известно, что при $n, m \rightarrow \infty$

$$\mathcal{L}(U(n, m) | H_0) \sim \mathcal{N}\left(\frac{nm}{2}, \frac{nm(n+m+1)}{12}\right).$$

Это предельное соотношение можно с хорошим приближением использовать уже при $n, m \geq 4, n+m \geq 20$.

Основываясь на полученных результатах, можно построить и рассчитать критерий согласия проверки гипотезы H_0 . Критическая область $\mathcal{S}_{1\alpha}$ имеет различный вид в зависимости от целей, для которых строится критерий. Если хотят получить критерий, состоятельный против альтернатив, для которых величина a в (3.28) удовлетворяет условию $a < 1/2$, то критическую область задают в виде $\mathcal{S}_{1\alpha} = \{t \leq t_{\alpha}(n, m)\}$, где при больших n и m можно положить

$$t_{\alpha}(n, m) = \frac{nm}{2} + \sqrt{\frac{nm(n+m+1)}{12}} t_{\alpha}, \quad \Phi(-t_{\alpha}) = \alpha.$$

Случай $a > 1/2$ сводится к рассмотренному перестановкой F_1 и F_2 . Наконец, критерий, состоятельный против альтернатив, определяемых условием $a \neq 1/2$, задают двусторонней критической областью

$$\mathcal{S}_{1\alpha} = \left\{ \left| t - \frac{nm}{2} \right| \geq \sqrt{\frac{nm(n+m+1)}{12}} t_{\alpha/2} \right\}, \quad \Phi(-t_{\alpha/2}) = \frac{\alpha}{2}.$$

Отметим, что в последнем случае критерий не отличает от нулевой гипотезы те альтернативы (F_1, F_2) , $F_1 \equiv F_2$, для которых $a = 1/2$.

Рассмотренный в п. 1 критерий Смирнова также, по существу, является ранговым критерием [4, с. 112].

§ 3.5. Гипотеза независимости

В данном параграфе будут рассмотрены несколько вариантов проверки гипотезы независимости, описанной в примере 3.3. В этом случае имеется выборка $((X_1, Y_1), \dots, (X_n, Y_n))$ из двумерного распределения $\mathcal{L}(\xi = (\xi_1, \xi_2))$ с известной функцией распределения $F_\xi(x, y)$, для которой требуется проверить гипотезу $H_0: F_\xi(x, y) = F_{\xi_1}(x)F_{\xi_2}(y)$, где F_{ξ_i} , $i = 1, 2$, — некоторые одномерные функции распределения.

1. Критерий независимости χ^2 . Простой критерий согласия для гипотезы H_0 можно построить, основываясь на методике χ^2 . Как известно, эту методику применяют для дискретных моделей с конечным числом исходов, поэтому условимся считать, что случайная величина ξ_1 принимает конечное число s некоторых значений, которые будем обозначать буквами $a_1, \dots, a_s$, а вторая компонента ξ_2 — k значений $b_1, \dots, b_k$. Если исходная модель имеет другую структуру, то предварительно группируют возможные значения случайных величин отдельно по первой и второй компонентам. В этом случае множество значений ξ_1 разбивается на s интервалов $\mathcal{E}_1^{(1)}, \dots, \mathcal{E}_s^{(1)}$, множество значений ξ_2 — на k интервалов $\mathcal{E}_1^{(2)}, \dots, \mathcal{E}_k^{(2)}$, а само множество значений $\xi = (\xi_1, \xi_2)$ — на $N = sk$ прямоугольников $\mathcal{E}_i^{(1)} \times \mathcal{E}_j^{(2)}$.

Обозначим через v_{ij} число наблюдений пары (a_i, b_j) (число элементов выборки, принадлежащих прямоугольнику $\mathcal{E}_i^{(1)} \times \mathcal{E}_j^{(2)}$,

если данные группируются), так что $\sum_{i=1}^s \sum_{j=1}^k v_{ij} = n$. Результаты

наблюдений удобно расположить в виде *таблицы сопряженности двух признаков* (табл. 3.2). В приложениях ξ_1 и ξ_2 обычно означают два признака, по которым производится классификация результатов наблюдений.

Таблица 3.2

ξ_1	ξ_2				Сумма
	b_1	b_2	...	b_k	
a_1	v_{11}	v_{12}	...	v_{1k}	$v_{1\cdot}$
a_2	v_{21}	v_{22}	...	v_{2k}	$v_{2\cdot}$
...	...	...	...	...	...
a_s	v_{s1}	v_{s2}	...	v_{sk}	$v_{s\cdot}$
Сумма	$v_{\cdot 1}$	$v_{\cdot 2}$	...	$v_{\cdot k}$	n

Пусть $p_{ij} = P(\xi_1 = a_i, \xi_2 = b_j)$, $i = 1, \dots, s$, $j = 1, \dots, k$. Тогда гипотеза независимости означает, что существуют $s + k$ постоян-

ных $p_{i.}$, $p_{.j}$ таких, что $\sum_{i=1}^s p_{i.} = \sum_{j=1}^k p_{.j} = 1$ и $p_{ij} = p_{i.} p_{.j}$, т. е.

$$\mathcal{L} = (v_{ij}, i = 1, \dots, s, j = 1, \dots, k | H_0) = M(n; \mathbf{p} = (p_{i.}, p_{.j}, i = 1, \dots, s, j = 1, \dots, k)).$$

Таким образом, гипотеза H_0 сводится к утверждению, что частоты v_{ij} (число их равно $N = sk$) распределены по полиномиальному закону с вероятностями исходов, имеющими указанную специфическую структуру (вектор вероятностей исходов $\mathbf{p}$ определяется значениями $r = s + k - 2$ неизвестных параметров $p_{1.}, \dots, p_{s-1.}, p_{.1}, \dots, p_{.k-1}$).

Для проверки этой гипотезы, следовательно, можно применить описанную в п. 3 § 3.2 методику χ^2 . Найдем оценки максимального правдоподобия для определяющих рассматриваемую схему неизвестных параметров. Если справедлива нулевая гипотеза, то функция правдоподобия имеет вид

$$L(\mathbf{p}) = c \prod_{i,j} (p_{i.} p_{.j})^{v_{ij}} = c \prod_i p_{i.}^{v_{i.}} \prod_j p_{.j}^{v_{.j}},$$

где множитель c от неизвестных параметров не зависит. Отсюда по методу неопределенных множителей Лагранжа получаем, что искомые оценки имеют вид $\hat{p}_{i.} = v_{i.}/n$, $i = 1, \dots, s$; $\hat{p}_{.j} = v_{.j}/n$, $j = 1, \dots, k$. Следовательно, статистика

$$\hat{\chi}_n^2 = n \sum_{i=1}^s \sum_{j=1}^k \frac{(v_{ij} - v_{i.} v_{.j}/n)^2}{v_{i.} v_{.j}} = n \left(\sum_{i,j} \frac{v_{ij}^2}{v_{i.} v_{.j}} - 1 \right) \quad (3.30)$$

и, по теореме 3.2, $\mathcal{L}(\hat{\chi}_n^2 | H_0) \rightarrow \chi^2((s-1)(k-1))$ при $n \rightarrow \infty$, поскольку число степеней свободы в предельном распределении χ^2 равно $N - 1 - r = sk - 1 - (s + k - 2) = (s-1)(k-1)$.

Итак, при достаточно больших n можно использовать следующее правило проверки гипотезы: гипотезу H_0 отвергают тогда и только тогда, когда вычисленное по фактическим данным значение t статистики (3.30) удовлетворяет неравенству $t \geq \chi_{1-\alpha, (s-1)(k-1)}^2$. Этот критерий имеет асимптотически (при $n \rightarrow \infty$) заданный уровень значимости α и называется критерием независимости χ^2 .

Пример 3.9 [10, с. 781]. В эксперименте каждый индивидуум классифицировался по двум признакам: цвету глаз и цвету волос; при этом по первому признаку ξ_1 индивидуум относился к одной из трех категорий a_1, a_2, a_3 , а по второму ξ_2 — к одной из четырех категорий $b_1, \dots, b_4$. Соответствующие данные для $n = 6800$ индивидуумов приведены в табл. 3.3.

Здесь значение статистики (3.30) равно $t = 1075,2$. По таблицам распределения χ^2 находим, что, например, $\chi_{0,999; 6}^2 = 22,5$. Таким образом, гипотезу о независимости этих двух признаков следует отклонить; вероятность ошибки при этом значительно меньше 0,001.

Таблица 3.3

Цвет глаз	Цвет волос				Сумма
	b_1	b_2	b_3	b_4	
a_1	1768	807	189	47	2811
a_2	946	1387	746	53	3132
a_3	115	438	288	16	857
Сумма	2829	2632	1223	116	6800

2. Критерий Спирмена. На практике для проверки гипотезы независимости часто используют ранговые критерии (см. п. 2г) предыдущего параграфа). Наиболее известным из них является *критерий Спирмена*, который можно построить следующим образом. Обозначим через R_i ранг X_i среди элементов $X_1, \dots, X_n$ (т. е. номер места, занимаемого величиной X_i в вариационном ряду $X_1, \dots, X_n$); аналогично, пусть S_i — ранг Y_i среди элементов $Y_1, \dots, Y_n$. Таким образом, исходная выборка порождает множество пар рангов $(R_1, S_1), \dots, (R_n, S_n)$. Переставив эти пары в порядке возрастания первой компоненты, обозначим полученное множество пар через $(1, T_1), \dots, (n, T_n)$. Рассмотрим теперь ранговую статистику

$$\rho = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\left[\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{i=1}^n (S_i - \bar{S})^2 \right]^{1/2}},$$

представляющую собой коэффициент корреляции двух множеств рангов $(R_1, \dots, R_n)$ и $(S_1, \dots, S_n)$ (здесь R и S — соответствующие арифметические средние). Но $(R_1, \dots, R_n)$ и $(S_1, \dots, S_n)$ — некоторые перестановки множества $(1, \dots, n)$, поэтому

$$\bar{R} = \bar{S} = \frac{1}{n} \sum_{i=1}^n i = \frac{n+1}{2},$$

$$\sum_{i=1}^n (R_i - \bar{R})^2 = \sum_{i=1}^n (S_i - \bar{S})^2 = \sum_{i=1}^n i^2 - n \left(\frac{n+1}{2} \right)^2 = \frac{n(n^2-1)}{12}.$$

Отсюда

$$\begin{aligned} \rho &= \frac{12}{n(n^2-1)} \sum_{i=1}^n \left(R_i - \frac{n+1}{2} \right) \left(S_i - \frac{n+1}{2} \right) = \\ &= \frac{12}{n(n^2-1)} \sum_{i=1}^n \left(i - \frac{n+1}{2} \right) \left(T_i - \frac{n+1}{2} \right). \end{aligned} \quad (3.31)$$

Таким образом, ρ — линейная функция рангов T_i . Часто используют также формулы

$$\rho = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^n (R_i - S_i)^2 = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^n (i - T_i)^2, \quad (3.32)$$

совпадение которых с (3.31) проверяют непосредственно. Величину ρ называют *статистикой Спирмена*, а критерий проверки гипотезы H_0 , основанный на этой статистике, — *критерием Спирмена*.

Исследуем некоторые свойства этой статистики при гипотезе H_0 . Множество рангов $(T_1, \dots, T_n)$ — это некоторая перестановка $(1, \dots, n)$, и интуитивно ясно, что при гипотезе H_0 все $n!$ таких перестановок равновероятны. Поэтому

$$ET_i = \sum_{j=1}^n j \frac{(n-1)!}{n!} = \frac{n+1}{2}$$

и из (3.32) имеем

$$E\tau = 1 - \frac{6}{n(n^2-1)} \left[2 \sum_{i=1}^n i^2 - 2 \sum_{i=1}^n i ET_i \right] = 0.$$

Аналогично можно найти

$$D\rho = 1/(n-1). \quad (3.33)$$

При полном соответствии рангов $(R_i = S_i, i=1, \dots, n)$ $\rho=1$, а при противоположных рангах $(T_i = n-i+1, i=1, \dots, n)$ $\rho=-1$, вообще, $-1 \leq \rho \leq 1$. Значения ρ , близкие к крайним, рассматривают как свидетельствующие против гипотезы H_0 , поэтому критическую область критерия Спирмена задают в виде $\mathcal{F}_{1\alpha} = \{|\rho| \geq t_\alpha(n)\}$. Для определения численного значения критической границы $t_\alpha(n)$ при заданных объеме выборки n и уровне значимости α используют таблицы табулированного распределения статистики ρ , рассчитанные для $n=2, \dots, 30$. При больших n можно воспользоваться приближенным распределением. Известно [10, с. 644], что $\mathcal{L}(\sqrt{n}\rho | H_0) \rightarrow \mathcal{L}^*(0, 1)$ при $n \rightarrow \infty$. Отсюда следует, что если выбрать $t_\alpha(n) = c_\alpha \sqrt{n}$, где $\Phi(-c_\alpha) = \alpha/2$, то при больших n

$$P(\rho \in \mathcal{F}_{1\alpha} | H_0) = P(\sqrt{n}|\rho| \geq c_\alpha | H_0) \approx 2\Phi(-c_\alpha) = \alpha,$$

т. е. уровень значимости критерия приблизительно равен α .

3. Критерий Кендалла. Другой известный ранговый критерий предложен Кендаллом и основан на статистике

$$\tau = \frac{1}{n^2} \sum_{i < j} \text{sign}(T_j - T_i),$$

где $\text{sign } a = 1$, если $a > 0$, и $\text{sign } a = -1$, если $a < 0$. Известно [10, с. 643], что $E(\tau | H_0) = 0$, $D(\tau | H_0) = 2(2n+5)/[9n(n-1)]$ и $\mathcal{L}(\tau | H_0) \sim \mathcal{L}^*(0, 4/(9n))$ при $n \rightarrow \infty$. Отсюда следует, что при больших n критическую область следует выбрать в виде $\mathcal{F}_{1\alpha} = \{|\tau| \geq 2c_\alpha/(3\sqrt{n})\}$, $\Phi(-c_\alpha) = \alpha/2$.

Статистики ρ и τ имеют разную форму, однако они сильно коррелированы; если гипотеза H_0 истинна, то [10, с. 683] $\text{corr}(\rho, \tau) = k_n = 2(n+1)/\sqrt{2n(2n+5)}$.

Функция k_n убывает от 1 при $n=2$ до минимального значения 0,98 при $n=5$ и затем возрастает до 1 при $n \rightarrow \infty$, т. е. критерии Спирмена и Кендалла асимптотически эквивалентны.

§ 3.6. Гипотеза случайности

В различных статистических задачах исходные данные $X = (X_1, \dots, X_n)$ часто рассматривают как случайную выборку из некоторого распределения $\mathcal{L}(\xi)$, т. е. считают компоненты X_i вектора данных X независимыми и одинаково распределенными случайными величинами. Как правило, это предположение оправдано и вытекает из самого характера задачи, но иногда оно нуждается в проверке. Математическую задачу можно сформулировать следующим образом: проверить гипотезу

$H_0: F_X(x) = F(x_1) \dots F(x_n)$, $x = (x_1, \dots, x_n)$, где $F(x)$ — некоторая функция распределения. Такую гипотезу называют *гипотезой случайности*. Критерий согласия для проверки этой гипотезы можно построить, исходя из следующих соображений (далее предполагается, что вектор X имеет непрерывное распределение).

Если гипотеза случайности действительно имеет место, то компоненты вектора X «равноправны» и поэтому данные не должны быть ни в каком смысле упорядочены. Другими словами, ситуацию, соответствующую гипотезе H_0 , можно охарактеризовать как «полный хаос», или «полный беспорядок». При отклонениях от H_0 исходные данные имеют тот или иной порядок, проявляются связи. Следовательно, критерий проверки H_0 можно построить на основании статистик, измеряющих степень «беспорядка» исходных данных. Одной из таких статистик является число инверсий в выборке. Эта статистика определяется следующим образом. Построим вариационный ряд $X_{(1)} < \dots < X_{(n)}$ выборки $X = (X_1, \dots, X_n)$. Говорят, что компоненты X_i и X_j образуют *инверсию*, если $i < j$, но X_i стоит правее X_j в вариационном ряду, т. е. наблюдению с меньшим номером соответствует большее значение. Пусть η_i — число инверсий, образованных компонентой X_i (в вариационном ряду левее X_i стоит η_i элементов выборки с большими номерами), $i = 1, \dots, n-1$. Тогда $T_n = T_n(X) = \eta_1 + \dots + \eta_{n-1}$ — общее число инверсий для выборки X . Статистика T_n является естественной мерой «беспорядка» среди наблюдений, и ее можно использовать для проверки гипотезы H_0 . Крайние случаи, когда вариационный ряд имеет вид $X_1 < X_2 < \dots < X_n$ или $X_n < X_{n-1} < \dots < X_1$, естественно рассматривать как свидетельства «полного отсутствия беспорядка», т. е. противоречащие гипотезе H_0 . В первом случае статистика T_n принимает минимальное значение, равное 0, а во втором случае она максимальна и равна $(n-1) + (n-2) + \dots + 1 = n(n-1)/2$. Таким образом, слишком малые значения T_n и слишком большие (близкие к $(n-1)/2$) естественно рассматривать как критические для гипотезы H_0 . Чтобы определить числовые характеристики этого критерия, найдем распределение статистики T_n при гипотезе H_0 .

Из соображений симметрии ясно, что при гипотезе H_0 любое из $n!$ относительных расположений элементов выборки в соответствующем вариационном ряду имеет одинаковую вероятность $1/n!$. Введенная случайная величина η_i определяется расположением компоненты X_i по отношению к $X_{i+1}, \dots, X_n$ в вариационном ряду и не зависит от относительного расположения последних между собой, т. е. η_i при любом $i = 1, \dots, n-2$ не зависит от $\eta_{i+1}, \dots, \eta_{n-1}$. Таким образом, $\eta_1, \dots, \eta_{n-1}$ взаимно независимы.

Далее, η_i может с одной и той же вероятностью $1/(n-i+1)$ принимать значения $0, 1, \dots, n-i$, поэтому ее производящая функция имеет вид

$$\varphi_i(z) = \sum_{r=0}^{n-i} P(\eta_i=r) z^r = \frac{1}{n-i+1} (1+z+\dots+z^{n-i}),$$

а производящая функция статистики T_n — вид

$$\Phi_n(z) = \sum_r P(T_n=r) z^r = \prod_{i=1}^{n-1} \varphi_i(z) = \frac{1}{n!} \prod_{r=1}^{n-1} (1+z+\dots+z^r).$$

Отсюда имеем:

$$E\eta_i = \varphi_i'(1) = \frac{n-i}{2}, \quad D\eta_i = \varphi_i''(1) + E\eta_i - (E\eta_i)^2 = \frac{(n-i)(n-i+2)}{12},$$

$$ET_n = \sum_{i=1}^{n-1} E\eta_i = \frac{n(n-1)}{4}, \quad DT_n = \sum_{i=1}^{n-1} D\eta_i = \frac{2n^3+3n^2-5n}{72}.$$

Итак, среднее значение статистики T_n при нулевой гипотезе совпадает с серединой промежутка $[0, n(n-1)/2]$, и в критическую область $\mathcal{F}_{1\alpha}$ следует включать все целые точки этого промежутка, достаточно удаленные от середины, т. е. можно положить $\mathcal{F}_{1\alpha} = \{t \mid t - n(n-1)/4 \geq t_\alpha(n)\}$ (в данном случае t (возможное значение статистики T_n) пробегает все целые точки $0, 1, \dots, n(n-1)/2$). Границу $t_\alpha(n)$ при заданном уровне значимости α выбирают из условия $P(T_n \in \mathcal{F}_{1\alpha} | H_0) \leq \alpha$ или, что эквивалентно, из условия

$$P(T_n \in \mathcal{F}_{1\alpha} | H_0) = P\left(\frac{n(n-1)}{4} - t_\alpha(n) \leq T_n \leq \frac{n(n-1)}{4} + t_\alpha(n) \mid H_0\right) \geq 1 - \alpha$$

($t_\alpha(n)$ — это минимальное число, удовлетворяющее данному соотношению).

Раскладывая функцию $\Phi_n(z)$ в ряд по степеням z и вычисляя коэффициент при z^r , можно вычислить вероятности $P(T_n=r | H_0)$ при заданном n и любом r и использовать их для нахождения критической границы $t_\alpha(n)$. Распределение статистики T_n протабулировано для значений $n=2, 3, \dots, 12$ [2].

Для больших объемов выборки n применяют простой асимптотический вариант этого критерия. Используя производящую функцию $\Phi_n(z)$, можно показать, что характеристическая функция нормированной статистики $T_n^* = (T_n - n(n-1)/4) / (6/n^{3/2})$ сходится при $n \rightarrow \infty$ и любом конечном t к $e^{-t^2/2}$ — характеристической функции нормального распределения. Это означает, что $\mathcal{L}(T_n^* | H_0) \rightarrow \rightarrow \mathcal{L}^*(0, 1)$ при $n \rightarrow \infty$.

Последний результат дает возможность построить следующее правило проверки гипотезы H_0 , когда n велико: для заданного уровня значимости α определяют число t_α из условия $\Phi(-t_\alpha) = \alpha/2$; по фактически наблюдавшимся данным $x = (x_1, \dots, x_n)$ вычисляют значение $t = T_n(x)$ числа инверсий в выборке; если $|t - n(n-1)/4| / (6/n^{3/2}) > t_\alpha$, то гипотезу H_0 отвергают как противоречащую исходным данным; в противном случае признают, что гипотеза независимости и одинаковости распределенности наблюдений согласуется с опытными данными.

Вероятность ошибочно отвергнуть при этом истинную гипотезу H_0 равна

$$P\left(\left|T_n - \frac{n(n-1)}{4}\right| \left| \frac{6}{n^{3/2}} > t_{\alpha} \mid H_0 \right) \xrightarrow{n \rightarrow \infty} 2\Phi(-t_{\alpha}) = \alpha.$$

Это правило можно использовать уже при $n > 10$.

Задачи

1. При $n=4000$ независимых испытаний события A_1 , A_2 и A_3 , составляющие полную группу, осуществились 1905, 1015 и 1080 раз соответственно. Проверить, согласуются ли эти данные на уровне 0,05 с гипотезой $H_0: p_1=1/2$, $p_2=p_3=1/4$, где $p_i=P(A_i)$.

2. Доказать формулу (3.8) для дисперсии статистики χ^2_n .

3. В опытах наблюдалась неотрицательная непрерывная случайная величина ξ . Ее значения (упорядоченные по величине и округленные с точностью до 0,01) для $n=50$ опытов оказались равными: 0,01; 0,01; 0,04; 0,17; 0,18; 0,22; 0,22; 0,25; 0,25; 0,29; 0,42; 0,46; 0,47; 0,56; 0,59; 0,67; 0,68; 0,70; 0,72; 0,76; 0,78; 0,83; 0,85; 0,87; 0,93; 1,00; 1,01; 1,01; 1,02; 1,03; 1,05; 1,32; 1,34; 1,37; 1,47; 1,50; 1,52; 1,54; 1,59; 1,71; 1,90; 2,10; 2,35; 2,46; 2,46; 2,50; 3,73; 4,07; 6,03. Проверить гипотезу $H_0: F_{\xi}(x) = 1 - e^{-x}$, $x \geq 0$, применяя метод группировки с четырьмя равновероятными интервалами (уровень значимости принять равным 0,1). У к а з а н и е. Здесь $\zeta_1=0,288$; $\zeta_2=0,693$; $\zeta_3=1,386$ ($F(\zeta_j)=j/4$, $j=1, 2, 3$). $\chi^2_{0,95; 3}=6,25$; $\chi^2_{1, 3}=3,9$.

4. Реализацией выборки $(X_1, \dots, X_n)$ является целочисленный вектор (47, 46, 49, 53, 50). Можно ли с уровнем значимости 0,1 считать распределение наблюдавшейся случайной величины пуассоновским?

5. Поступающие в институт абитуриенты разбиты на два потока по 300 человек в каждом. Итоги экзамена по одному и тому же предмету на каждом потоке оказались следующими:

1-й поток: баллы «2», «3», «4» и «5» получили соответственно 33, 43, 80 и 144 человека;

2-поток: баллы «2», «3», «4» и «5» получили соответственно 39, 35, 72 и 154 человека.

Можно ли считать оба потока однородными (на уровне значимости 0,05)? У к а з а н и е. Критический уровень $\chi^2_{0,95; 3}=7,82$.

6. Из 300 абитуриентов, поступивших в институт, 97 человек имели балл «5» в школе и 48 получили «5» на вступительных экзаменах по тому же предмету, причем только 18 человек имели «5» и в школе и на экзамене. С уровнем значимости 0,1 проверить гипотезу о независимости оценок «5» в школе и на экзаменах. У к а з а н и е. Таблица сопряженности двух признаков (см. табл. 3.2) в данном случае имеет следующий вид:

ξ_1	ξ_2		Сумма
	«5»	«не 5»	
«5»	18	30	48
«не 5»	79	173	252
Сумма	97	203	300

Использовать значение $\chi^2_{0,95; 1}=2,71$.

7. Таблица «случайных чисел» содержит реализации 10 000 независимых и одинаково распределенных случайных величин, принимающих значения 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Корректно ли предположение о равновероятности этих значений, если в упомянутой таблице числа, не превосходящие 4, встречаются 4806 раз? При каком уровне значимости гипотеза равновероятности отвергается?

8. Можно ли с уровнем значимости 0,001 считать, что последовательность чисел 1,05; 1,12; 1,37; 1,50; 1,51; 1,73; 1,85; 1,98 является реализацией слу-

чайного вектора, все 8 компонент которого — независимые одинаково распределенные случайные величины?

9. Среди 2020 семей, имеющих двух детей, 527 семей, в которых два мальчика, и 476 — две девочки (в остальных 1017 семьях дети разного пола). Можно ли с уровнем значимости 0,05 считать, что количество мальчиков в семье с двумя детьми — биномиальная случайная величина?

ξ_1	ξ_2			Сумма
	b_1	b_2	b_3	
a_1	3000	2832	3008	8840
a_2	3047	3051	2997	9095
a_3	2974	3038	3018	9030
Сумма	9030	8921	9023	26974

10. Для заданной таблицы сопряженности двух признаков проверить гипотезу независимости (уровень значимости взять равным 0,05).

11. Используя формулы (3.17), убедиться, что при $n, N \rightarrow \infty, n/N \rightarrow p > 0$ $E\mu_0/N \rightarrow e^{-p}$, $D\mu_0/N \rightarrow e^{-p}(1 - e^{-p}(1 + p))$.

12. Доказать теорему 3.5 об асимптотическом поведении мощности критерия пустых ящиков.

13. Доказать, что число инверсий в повторной выборке объема n распределено асимптотически нормально $\mathcal{N}(n(n-1)/4, n^3/36)$.

4

Глава

Параметрические гипотезы

Параметрические гипотезы, рассматриваемые в настоящей главе,— это гипотезы об истинном значении неизвестного параметра, определяющего заданное параметрическое семейство распределений. В данной главе изложены и продемонстрированы на конкретных примерах основные принципы построения оптимальных или асимптотически оптимальных критериев проверки таких гипотез, в основе которых лежит предложенный Ю. Нейманом и Э. Пирсоном метод отношения правдоподобия. Отдельный параграф посвящен последовательному анализу, основные элементы которого рассматриваются на примере различения двух простых гипотез.

§ 4.1. Общие положения

1. Понятие параметрической гипотезы. Важный класс статистических гипотез составляют гипотезы о параметрических моделях. В этом случае класс $\mathcal{F}$ допустимых распределений наблюдаемой случайной величины ξ имеет вид $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$, т. е. является классом специального функционального вида. Функции этого класса находят в соответствии со значениями вещественного параметра $\theta = (\theta_1, \dots, \theta_r)$ из некоторого параметрического множества $\Theta \subseteq R^r$, поэтому гипотезы, по существу, относятся к неизвестным параметрам распределения и называются *параметрическими*. Примерами параметрических гипотез являются утверждения следующего типа:

1) $H_0: \theta = \theta_0$, где $\theta_0 \in \Theta$ — некоторое фиксированное значение параметра;

2) $H_0: \theta_1 = \dots = \theta_r$;

3) $H_0: g(\theta) = g_0$, где $g(\theta)$ — некоторая (в общем случае векторная) функция θ , g_0 — фиксированное значение.

В общем случае параметрическая гипотеза задается указанием некоторого подмножества $\Theta_0 \subset \Theta$, элементом которого является, по предположению, неизвестная параметрическая точка θ . Записывается это так: $H_0: \theta \in \Theta_0$. Альтернативная гипотеза имеет вид $H_1: \theta \in \Theta_1 = \Theta \setminus \Theta_0$; точки $\theta \in \Theta_1$ называют *альтернативами*. Если множество Θ_0 (Θ_1) состоит из одной точки, то гипотезу H_0 (альтернативу H_1) называют *простой*; в противном случае гипотезу (или альтернативу) называют *сложной*. Так, например, гипотеза 1) —

простая, гипотеза 2) — сложная, а гипотеза 3) может быть как простой, так и сложной.

Приведем пример конкретной параметрической гипотезы. Пусть класс $\mathcal{F} = \mathcal{N}(\theta_1, \theta_2^2)$. Тогда утверждение $H_0: \theta_1 = \theta_{10}, \theta_2 = \theta_{20}$, где θ_{10}, θ_{20} — заданные числа, есть простая гипотеза о нормальном распределении наблюдений. Гипотеза, выраженная равенством $\theta_1 = \theta_{10}$ и оставляющая значение дисперсии θ_2^2 неопределенным, — сложная.

2. Критерии проверки гипотез. Пусть имеется выборка $X = (X_1, \dots, X_n)$ из распределения $\mathcal{L}(\xi) \in \mathcal{F}$, о котором сформулирована некоторая гипотеза $H_0: \theta \in \Theta_0$ (θ может быть как скаляром, так и вектором). Требуется выяснить, верна или не верна гипотеза H_0 , т. е. надо построить такое правило (критерий), которое позволяло бы для каждой реализации x выборки X принять одно из двух решений: принять гипотезу H_0 или отклонить ее (принять H_1). Тем самым каждому критерию соответствует разбиение выборочного пространства $\mathcal{X}$ на два взаимно дополнительных множества $\mathcal{X}_0$ и $\mathcal{X}_1$ ($\mathcal{X}_0 \mathcal{X}_1 = \emptyset, \mathcal{X}_0 \cup \mathcal{X}_1 = \mathcal{X}$), где $\mathcal{X}_0$ состоит из точек x , для которых гипотеза H_0 принимается, а $\mathcal{X}_1$ — из точек, для которых H_0 отвергается. Множество $\mathcal{X}_0$ называют *областью принятия гипотезы H_0* , а $\mathcal{X}_1$ — областью ее отклонения или *критической областью*. Таким образом, выбор правила проверки гипотезы H_0 эквивалентен заданию критической области $\mathcal{X}_1$. Если выбрана критическая область $\mathcal{X}_1$, то критерий можно сформулировать следующим образом: *пусть x — наблюдавшаяся реализация выборки X ; тогда при $x \in \mathcal{X}_1$ гипотезу H_0 отвергают (принимают альтернативную гипотезу H_1), если же $x \in \mathcal{X}_0 = \mathcal{X}_1$, то гипотезу H_0 принимают.* Критерий, определяемый критической областью $\mathcal{X}_1$, часто для краткости называют критерием $\mathcal{X}_1$.

В некоторых случаях удобно рассматривать критерии более сложной структуры — так называемые *рандомизированные* критерии, когда при наблюдении x гипотезу H_0 отвергают с некоторой вероятностью $\varphi(x)$ и принимают с дополнительной вероятностью $1 - \varphi(x)$. Рандомизированный критерий, таким образом, полностью характеризуется *критической функцией* $\varphi(x)$ ($0 \leq \varphi(x) \leq 1, \forall x \in \mathcal{X}$). Если функция $\varphi(x)$ принимает только два значения 0 и 1, то приходим, очевидно, к случаю *нерандомизированного* критерия с критической областью $\mathcal{X}_1 = \{x: \varphi(x) = 1\}$. Далее будут рассмотрены в основном нерандомизированные критерии, которые, как правило, и используют на практике.

3. Общий принцип выбора критической области критерия. В процессе проверки гипотезы H_0 можно прийти к правильному решению или совершить *ошибку первого рода* — отклонить H_0 , когда она верна, или *ошибку второго рода* — принять H_0 , когда она ложна. Иными словами, ошибка первого рода имеет место, если точка x попадает в критическую область $\mathcal{X}_1$, в то время как верна нулевая гипотеза H_0 , а ошибка второго рода — когда $x \in \mathcal{X}_0$, но гипотеза H_0 не верна (верна альтернатива H_1).

Вероятности этих ошибок можно выразить через функцию мощности $W(\theta)$ критерия $\mathcal{X}_1$:

$$W(\theta) = W(\mathcal{X}_1; \theta) = P_\theta(X \in \mathcal{X}_1), \quad \theta \in \Theta.$$

Именно: вероятности ошибок первого и второго рода равны соответственно $W(\theta)$, $\theta \in \Theta_0$, и $1 - W(\theta)$, $\theta \in \Theta_1$. Иногда удобно эти вероятности записывать в символическом виде: $P(H_1 | H_0)$ (вероятность ошибки первого рода) и $P(H_0 | H_1)$ (вероятность ошибки второго рода). В случае рандомизированного критерия, задаваемого критической функцией $\varphi(x)$, функция мощности определяется соотношением $W(\theta) = W(\varphi; \theta) = E_\theta \varphi(X)$.

Желательно провести проверку гипотезы так, чтобы свести к минимуму вероятности обоих типов ошибок. Однако при данном числе испытаний n в общем случае невозможно ни при каком выборе критической области одновременно обе эти вероятности сделать как угодно малыми. В то же время, выбирая критическую область, можно добиться произвольной малости вероятности какой-либо одной из ошибок первого и второго рода. Так, положив $\mathcal{A}_1 = \mathcal{A}$, будем иметь $W(\theta) \equiv 1$ и поэтому вероятность ошибки второго рода равна 0; если $\mathcal{A}_1 = \emptyset$, то $W(\theta) \equiv 0$ и нулю равна вероятность ошибки первого рода. Ясно, что ни с одним из этих решений согласиться нельзя. Рациональный принцип выбора критической области можно сформулировать следующим образом: при заданном числе испытаний n устанавливается граница для вероятности ошибки первого рода и при этом выбирается та критическая область $\mathcal{A}_1$, для которой вероятность ошибки второго рода минимальна. Иными словами, выбирается число α между 0 и 1 и налагается условие

$$W(\theta) \leq \alpha \quad \text{для всех } \theta \in \Theta_0; \quad (4.1)$$

при этом условии желательно сделать минимальной (за счет выбора критической области $\mathcal{A}_1$) величину $1 - W(\theta)$ для всех $\theta \in \Theta_1$, или, что то же самое, максимальной мощность

$$W(\theta) \quad \text{для всех } \theta \in \Theta_1. \quad (4.2)$$

Величину α в соотношении (4.1) называют *уровнем значимости*, а тот факт, что критерий $\mathcal{A}_1$ имеет уровень значимости α , часто подчеркивают обозначением $\mathcal{A}_{1\alpha}$.

В конкретных задачах выбор уровня значимости до некоторой степени произволен и связан с практической стороной вопроса. Так, часто ошибочное принятие или отбрасывание гипотезы H_0 связано с материальными затратами. Если принятие гипотезы H_0 в то время, когда она не верна (ошибка второго рода), приводит к большим затратам, тогда как отклонение истинной гипотезы H_0 (ошибка первого рода) приводит к небольшим потерям, то ясно, что желательно сделать как можно меньшей вероятность ошибки второго рода, допуская сравнительно большие значения α . Обычно для α выбирают одно из следующих стандартных значений: 0,005; 0,01; 0,05; для этих значений рассчитывают соответствующие таблицы, используемые при проведении различных испытаний.

4. Равномерно наиболее мощные критерии. Пусть $\mathcal{A}_{1\alpha}$ и $\mathcal{A}_{1\alpha}^*$ — два критерия одного и того же уровня значимости α для гипотезы H_0 . Если

$$W(\mathcal{A}_{1\alpha}^*; \theta) \leq W(\mathcal{A}_{1\alpha}; \theta), \quad \theta \in \Theta_0, \quad \text{и} \quad (4.3)$$

$$W(\mathcal{X}_{1\alpha}^*; \theta) \geq W(\mathcal{X}_{1\alpha}; \theta), \theta \in \Theta_1, \quad (4.4)$$

причем строгое неравенство в (4.4) имеет место хотя бы при одном значении θ , то говорят, что критерий $\mathcal{X}_{1\alpha}^*$ *равномерно мощнее* критерия $\mathcal{X}_{1\alpha}$. В этом случае, очевидно, следует отдать предпочтение критерию $\mathcal{X}_{1\alpha}^*$, как приводящему к меньшим ошибкам. Если соотношения (4.3) и (4.4) выполняются для любого критерия $\mathcal{X}_{1\alpha}$, то $\mathcal{X}_{1\alpha}^*$ называют *равномерно наиболее мощным* (р. н. м.) *критерием* для проверки гипотезы H_0 . В случае когда множество Θ_1 состоит из одной точки (гипотеза H_1 простая), вместо термина р. н. м. критерий используют термин *наиболее мощный критерий*.

Равномерно наиболее мощный критерий не всегда существует, так как, как правило, критерий, максимизирующий мощность при определенной альтернативе $\theta \in \Theta_1$, зависит от этой альтернативы и экстремальная задача (4.2) при ограничении (4.1) имеет решение только в некоторых специальных случаях. Примеры таких ситуаций встретятся в дальнейшем.

Часто ограничиваются рассмотрением подкласса *несмещенных* критериев, для которых одновременно с (4.1) выполняется условие $W(\theta) \geq \alpha$ для всех $\theta \in \Theta_1$. (4.5)

В ряде задач, для которых р. н. м. критерии не существуют, могут иметь место р. н. м. несмещенные критерии.

В заключение отметим, что обычно критическая область задается с помощью некоторой статистики $T(X)$ и имеет, как правило, следующий вид: $\mathcal{X}_1 = \{x : T(x) \geq c\}$, или $\mathcal{X}_1 = \{x : T(x) < c\}$, или $\mathcal{X}_1 = \{x : |T(x)| \geq c\}$. Функцию наблюдений $T(X)$ называют в этом случае *статистикой критерия*.

§ 4.2. Выбор из двух простых гипотез.

Критерий Неймана — Пирсона

1. Постановка задачи. В этом параграфе рассмотрен важный частный случай описанной в § 4.1 общей ситуации, а именно когда основная и альтернативная гипотезы являются простыми. В этом случае параметрическое множество Θ состоит из двух точек: $\Theta = \{\theta_0, \theta_1\}$ — и проверяемая (основная) гипотеза означает утверждение $H_0: \theta = \theta_0$, а альтернатива — утверждение $H_1: \theta = \theta_1$. Другими словами, допустимыми распределениями наблюдаемой случайной величины ξ являются только два распределения (две функции распределения): $F_0(x) = F(x; \theta_0)$ и $F_1(x) = F(x; \theta_1)$; требуется по выборке $X = (X_1, \dots, X_n)$ из распределения $\mathcal{X}(\xi)$ определить, какое из этих двух распределений истинно.

Предположим, что практические соображения приняты в расчет и уровень значимости α выбран. Тогда, согласно изложенному в § 4.1 общему принципу, задача построения наилучшего критерия $\mathcal{X}_{1\alpha}^*$ сводится к решению экстремальной задачи максимизации по $\mathcal{X}_{1\alpha}$ мощности $W(\mathcal{X}_{1\alpha}; \theta_1)$ при ограничении $W(\mathcal{X}_{1\alpha}; \theta_0) = \alpha$. Решению этой задачи, впервые найденному Ю. Нейманом и Э. Пирсоном (1933 г.), и посвящен настоящий параграф.

Напомним, что в рассматриваемом случае наилучший критерий $\mathcal{X}_{1\alpha}^*$ называется наиболее мощным критерием.

2. Критерий Неймана — Пирсона в случае абсолютно непрерывных распределений. Предположим, что распределения F_0 и F_1 абсолютно непрерывны и соответствующие плотности $f_0(x)$ и $f_1(x)$ удовлетворяют условию $f_j(x) > 0$, $j = 0, 1$. Рассмотрим статистику отношения правдоподобия

$$l(X) = \frac{L(X; \theta_1)}{L(X; \theta_0)} = \prod_{i=1}^n \frac{f_1(X_i)}{f_0(X_i)} \quad (4.6)$$

и определим функцию $\psi(c) = P_{\theta_0}(l(X) \geq c)$. С ростом c эта функция может только убывать; кроме того, $\psi(0) = 1$. Далее имеем

$$\begin{aligned} P_{\theta_1}(l(X) \geq c) &= \int_{x: l(x) \geq c} L(x; \theta_1) dx \geq c \int_{x: l(x) \geq c} L(x; \theta_0) dx = \\ &= c P_{\theta_0}(l(X) \geq c) = c\psi(c), \end{aligned} \quad (4.7)$$

поэтому $\psi(c) \leq 1/c$, откуда следует, что $\psi(c) \rightarrow 0$ при $c \rightarrow \infty$. Будем далее предполагать, что существует такое значение $c = c_\alpha$, для которого $\psi(c) = \alpha$ (в частности, это всегда имеет место, если функция $\psi(c)$ непрерывна). Тогда имеет место следующее утверждение.

Теорема 4.1 (Неймана — Пирсона). При сделанных предположениях существует наиболее мощный критерий проверки гипотезы H_0 . Этот критерий задается критической областью

$$\mathcal{X}_{1\alpha}^* = \{x: l(x) \geq c\}, \quad (4.8)$$

где критическая граница c определяется из условия $\psi(c) = \alpha$.

□ Рассмотрим любой другой критерий $\mathcal{X}_{1\alpha}$ уровня значимости α . Тогда

$$\begin{aligned} W(\mathcal{X}_{1\alpha}; \theta_1) &= \int_{\mathcal{X}_{1\alpha}} L(x; \theta_1) dx = \\ &= \int_{\mathcal{X}_{1\alpha} \cap \mathcal{X}_{1\alpha}^*} L(x; \theta_1) dx + \int_{\mathcal{X}_{1\alpha} \setminus \mathcal{X}_{1\alpha}^*} L(x; \theta_1) dx. \end{aligned}$$

Аналогично имеем

$$W(\mathcal{X}_{1\alpha}^*; \theta_1) = \int_{\mathcal{X}_{1\alpha} \cap \mathcal{X}_{1\alpha}^*} L(x; \theta_1) dx + \int_{\bar{\mathcal{X}}_{1\alpha} \cap \mathcal{X}_{1\alpha}^*} L(x; \theta_1) dx.$$

Отсюда

$$\begin{aligned} W(\mathcal{X}_{1\alpha}; \theta_1) &= W(\mathcal{X}_{1\alpha}^*; \theta_1) + \\ &+ \int_{\mathcal{X}_{1\alpha} \cap \bar{\mathcal{X}}_{1\alpha}^*} l(x) L(x; \theta_0) dx - \int_{\bar{\mathcal{X}}_{1\alpha} \cap \mathcal{X}_{1\alpha}^*} l(x) L(x; \theta_0) dx. \end{aligned}$$

Но согласно определению множества $\mathcal{X}_{1\alpha}^*$, вне этого множества (первый интеграл) $l(x) < c$, а в точках этого множества (второй интеграл) $-l(x) \leq -c$; следовательно, из последнего соотношения

получаем неравенство

$$W(\mathcal{X}_{1\alpha}; \theta_1) < W(\mathcal{X}_{1\alpha}^*; \theta_1) + c \left(\int_{\mathcal{X}_{1\alpha} \bar{\mathcal{X}}_{1\alpha}^*} L(x; \theta_0) dx - \int_{\bar{\mathcal{X}}_{1\alpha}^* \mathcal{X}_{1\alpha}^*} L(x; \theta_0) dx \right). \quad (4.9)$$

По условию,

$$\int_{\mathcal{X}_{1\alpha}} L(x; \theta_0) dx = \int_{\mathcal{X}_{1\alpha}^*} L(x; \theta_0) dx = \alpha,$$

откуда имеем, что каждый из интегралов в правой части (4.9) отличается от α на одну и ту же величину $\int_{\mathcal{X}_{1\alpha} \bar{\mathcal{X}}_{1\alpha}^*} L(x; \theta_0) dx$.

Таким образом, оба этих интеграла равны и получено требуемое неравенство $W(\mathcal{X}_{1\alpha}; \theta_1) < W(\mathcal{X}_{1\alpha}^*; \theta_1)$. ■

Построенный критерий проверки гипотезы H_0 называют *критерием Неймана — Пирсона*.

Покажем, что этот критерий является несмещенным [см. условие (4.5)]. Если в (4.8) $c > 1$, то из (4.7) получаем

$$W(\mathcal{X}_{1\alpha}^*; \theta_1) \geq c\alpha > \alpha.$$

Если $c < 1$, то

$$\begin{aligned} W(\mathcal{X}_{1\alpha}^*; \theta_1) &= 1 - \int_{\bar{\mathcal{X}}_{1\alpha}^*} L(x; \theta_1) dx > 1 - \int_{\mathcal{X}_{1\alpha}^*} L(x; \theta_0) dx = \\ &= \int_{\mathcal{X}_{1\alpha}^*} L(x; \theta_0) dx = \alpha, \end{aligned}$$

поскольку $L(x; \theta_1) < cL(x; \theta_0) \leq L(x; \theta_0)$ при $x \in \bar{\mathcal{X}}_{1\alpha}^*$. Следовательно, в любом случае $W(\mathcal{X}_{1\alpha}^*; \theta_1) > \alpha$.

3. Критерий Неймана — Пирсона в случае дискретных распределений. Соответствующие рассуждения можно провести и для дискретных распределений, для которых вероятности $f_j(z_k) > 0$ ($j=0, 1$) для всех z_k -возможных значений наблюдаемой случайной величины ξ ($\sum_k f_j(z_k) = 1$, $j=0, 1$). В этом

случае также «упорядочивают» выборочные точки $x = (x_1, \dots, x_n)$ соответственно величине отношения

$$l(x) = \frac{L(x; \theta_1)}{L(x; \theta_0)} = \frac{\prod_{i=1}^n l_1(x_i)}{\prod_{i=1}^n l_0(x_i)}$$

и включают в критическое множество $\mathcal{X}_1$ максимальное число точек, согласующееся с требованием $\sum_{x \in \mathcal{X}_1} L(x; \theta_0) \leq \alpha$. Однако в отличие от непрерыв-

ного случая здесь в силу дискретности распределения выборки, вообще говоря, уже нельзя получить точное значение α за счет выбора границы c в неравенстве $l(x) \geq c$, определяющем множество $\mathcal{X}_1$: может быть так, что, включив в $\mathcal{X}_1$ очередную точку, мы еще не достигнем уровня α , а включив следующую — превзойдем его. Подробнее это можно объяснить следующим образом. Обозначим через $\{\dots < l_k < l_{k+1} < \dots\}$ возможные значения статистики $l(X)$.

Тогда в общем случае можно определить такое $k = k(\alpha)$, что

$$\sum_{x: l(x) \geq l_{k+1}} L(x; \theta_0) < \alpha \leq \sum_{x: l(x) \geq l_k} L(x; \theta_0). \quad (4.10)$$

Если в правой части соотношения (4.10) имеет место равенство, то, полагая $\mathcal{X}_{\alpha}^+ = \{x: l(x) \geq l_{k(\alpha)}\}$ и повторяя рассуждения, проведенные в п. 2, можно показать, что это критическое множество задает наиболее мощный критерий уровня значимости α для проверки гипотезы H_0 против альтернативы H_1 .

Рассмотрим теперь более подробно случай, когда в (4.10) имеют место строгие неравенства (как это чаще всего и бывает). Пусть

$$\alpha_j = \sum_{x: l(x) \geq l_{k+1}} L(x; \theta_j) \text{ и} \quad (4.11)$$

$$p_j = P_{\theta_j}(l(X) = l_k) = \sum_{x: l(x) = l_k} L(x; \theta_j), \quad j = 0, 1.$$

Тогда соотношение (4.10) запишется в данном случае в виде $0 < \alpha - \alpha_0 < p_0$. Чтобы построить критерий с уровнем значимости α , надо использовать прием рандомизации, т. е. поступить следующим образом. Когда выполняется равенство $l(x) = l_{k(\alpha)}$, — произвести дополнительный случайный эксперимент с исходами R и $\bar{R}$, вероятности которых равны соответственно $(\alpha - \alpha_0)/p_0$ и $1 - (\alpha - \alpha_0)/p_0$, и отвергнуть гипотезу H_0 , если наблюдается R , и принять H_0 в противном случае. Если $l(x) > l_{k(\alpha)}$, то гипотезу H_0 следует отвергнуть, а при $l(x) < l_{k(\alpha)}$ — принять. Другими словами, строят рандомизированный критерий, основанный на критической функции

$$\varphi^*(x) = \begin{cases} 1 & \text{при } l(x) > l_{k(\alpha)}, \\ (\alpha - \alpha_0)/p_0 & \text{при } l(x) = l_{k(\alpha)}, \\ 0 & \text{при } l(x) < l_{k(\alpha)}. \end{cases}$$

Вероятность ошибки первого рода такого критерия равна [см. (4.11)]

$$\begin{aligned} P(H_1 | H_0) &= E_{\theta_0} \varphi^*(X) = P_{\theta_0}(l(X) > l_{k(\alpha)}) + \frac{\alpha - \alpha_0}{p_0} P_{\theta_0}(l(X) = l_{k(\alpha)}) = \\ &= \alpha_0 + \frac{\alpha - \alpha_0}{p_0} p_0 = \alpha, \end{aligned}$$

что и требовалось показать. Мощность вычисляют по аналогичной формуле $W(\varphi^*; \theta_1) = E_{\theta_1} \varphi^*(X) = \alpha_1 + (\alpha - \alpha_0) p_1/p_0$.

Покажем, что это наиболее мощный критерий. Рассмотрим произвольный критерий с уровнем значимости α , задаваемый некоторой критической функцией $\varphi(x): E_{\theta_0} \varphi(X) \leq \alpha$, и обозначим $\mathcal{X}^\pm = \{x: \varphi^*(x) - \varphi(x) \geq 0\}$ [знак «+» («-») соответствует неравенству $>$ ($<$)]. Если $x \in \mathcal{X}^+$, то $\varphi^*(x) > 0$ и поэтому $L(x; \theta_1) \geq l_k L(x; \theta_0)$; если $x \in \mathcal{X}^-$, то $\varphi^*(x) < 1$ и, следовательно, $L(x; \theta_1) \leq l_k L(x; \theta_0)$. Таким образом,

$$\sum_x (\varphi^*(x) - \varphi(x)) (L(x; \theta_1) - l_k L(x; \theta_0)) = \sum_{x \in \mathcal{X}^+} + \sum_{x \in \mathcal{X}^-} \geq 0.$$

Отсюда для разности мощностей $\Delta = E_{\theta_1} \varphi^*(X) - E_{\theta_1} \varphi(X)$ получаем

$$\begin{aligned} \Delta &= \sum_x (\varphi^*(x) - \varphi(x)) L(x; \theta_1) \geq l_k \sum_x (\varphi^*(x) - \varphi(x)) L(x; \theta_0) = \\ &= l_k (E_{\theta_0} \varphi^*(X) - E_{\theta_0} \varphi(X)) \geq l_k (\alpha - \alpha) = 0, \end{aligned}$$

что и требовалось доказать.

Итак, в дискретном случае всегда можно построить наиболее мощный критерий, который, вообще говоря, является рандомизированным. Отметим, что рандомизация необходима лишь на пограничном множестве $\{x: l(x) = l_k(\alpha)\}$ и только в том случае, когда желательно иметь вероятность ошибки первого рода равной точно α . На практике предпочитают в таких случаях несколько изменить уровень значимости так, чтобы отпала необходимость в рандомизации, т. е. заменить α на α_0 и использовать нерандомизированный критерий $\mathcal{X}_{1\alpha_0}^* = \{l(x) \geq l_k(\alpha_0)\}$, который, по предыдущему, является наиболее мощным критерием уровня значимости α_0 ($< \alpha$). Можно также использовать нерандомизированный критерий $\mathcal{X}_{1\alpha'}^* = \{l(x) \geq l_k(\alpha_0)\}$ с большим уровнем значимости $\alpha' = \alpha_0 + p_0$. Эти общие положения проиллюстрированы ниже в примере 4.2.

4. Примеры применения критерия Неймана — Пирсона.

Пример 4.1 (нормальная-1 модель, проверка простых гипотез). Пусть θ нормально распределенной случайной величины ξ с известной дисперсией σ^2 и неизвестным средним θ имеются две гипотезы: $H_0: \theta = \theta_0$ и $H_1: \theta = \theta_1$ (для определенности будем считать, что $\theta_1 > \theta_0$). Если $X = (X_1, \dots, X_n)$ — выборка из $\mathcal{L}(\xi)$ и x — наблюдавшаяся реализация X , то

$$l(x) = \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [(x_i - \theta_1)^2 - (x_i - \theta_0)^2] \right\} = \\ = \exp \left\{ \frac{n}{\sigma^2} (\theta_1 - \theta_0) \bar{x} - \frac{n}{2\sigma^2} (\theta_1^2 - \theta_0^2) \right\}$$

и неравенство $l(x) \geq c$ эквивалентно неравенству

$$\bar{x} \geq \sigma^2 \ln c / [n(\theta_1 - \theta_0)] + (\theta_1 + \theta_0)/2,$$

которое можно переписать в виде

$$\frac{\sqrt{n}}{\sigma} (\bar{x} - \theta_0) \geq \frac{\sigma}{\sqrt{n}(\theta_1 - \theta_0)} \ln c + \frac{\sqrt{n}}{2\sigma} (\theta_1 - \theta_0) = t(c).$$

Так как $\mathcal{L}_{\theta_0}(X) = \mathcal{N}(\theta_0, \sigma^2/n)$, то отсюда имеем

$$\psi(c) = P_{\theta_0}(l(X) \geq c) = P_{\theta_0} \left(\frac{\sqrt{n}}{\sigma} (\bar{X} - \theta_0) \geq t(c) \right) = \Phi(-t(c)).$$

При $c > 0$ функция $t(c)$ непрерывно зависит от c , поэтому $\psi(c)$ — непрерывная функция и для любого $\alpha \in (0, 1)$ однозначно определена величина c_α , где $t(c_\alpha) = t_\alpha$, а $\Phi(-t_\alpha) = \alpha$.

Таким образом, все условия теоремы 4.1 выполнены и, следовательно, наиболее мощный критерий для проверки гипотезы H_0 против альтернативы H_1 задается в данном случае критической областью

$$\mathcal{X}_{1\alpha}^* = \{x: \sqrt{n}(\bar{x} - \theta_0)/\sigma \geq t_\alpha\}, \quad \Phi(-t_\alpha) = \alpha. \quad (4.12)$$

В данном случае статистикой критерия является выборочное среднее $\bar{X}$, а критическая область не зависит от конкретного значения альтернативы $\theta_1 > \theta_0$.

Вычислим мощность критерия (4.12). Так как $\mathcal{L}_{\theta_1}(X) = \mathcal{N}(\theta_1, \sigma^2/n)$, то

$$\begin{aligned} W(\mathcal{X}_{1\alpha}^*; \theta_1) &= P_{\theta_1}\left(X \geq \theta_0 + \frac{\sigma}{\sqrt{n}} t_{\alpha}\right) = \\ &= P_{\theta_1}\left(\frac{\sqrt{n}}{\sigma}(X - \theta_1) \geq -\frac{\sqrt{n}}{\sigma}(\theta_1 - \theta_0) + t_{\alpha}\right) = \\ &= \Phi\left(\frac{\sqrt{n}}{\sigma}(\theta_1 - \theta_0) - t_{\alpha}\right). \end{aligned} \quad (4.13)$$

В частности, отсюда видно, что $W(\mathcal{X}_{1\alpha}^*; \theta_1) > \Phi(-t_{\alpha}) = \alpha$ (несмещенность).

Аналогично, при $\theta_1 < \theta_0$ наиболее мощный критерий имеет вид

$$\mathcal{X}_{1\alpha}^* = \{x : \sqrt{n}(\bar{x} - \theta_0)/\sigma \leq -t_{\alpha}\}, \quad \Phi(-t_{\alpha}) = \alpha, \quad (4.14)$$

а его мощность

$$W(\mathcal{X}_{1\alpha}^*; \theta_1) = \Phi(\sqrt{n}(\theta_0 - \theta_1)/\sigma - t_{\alpha}). \quad (4.15)$$

Из (4.13) следует, что вероятность ошибки второго рода критерия (4.12) равна

$$\beta = \beta(\alpha, n) = \Phi(t_{\alpha} - \sqrt{n}(\theta_1 - \theta_0)/\sigma). \quad (4.16)$$

Наглядной иллюстрацией приведенных рассуждений служит рис. 4.1.

Рассмотрим следующую задачу. Пусть заранее заданы вероятности ошибок первого и второго рода α и β . Определим, каким должно быть минимальное число $n^* = n^*(\alpha, \beta)$ испытаний, необходимых для того, чтобы ошибочные заключения могли быть сделаны с вероятностями, не превосходящими α и β .

Из (4.16) следует, что $\beta(\alpha, n)$ с ростом n убывает к нулю, поэтому искомое число n^* есть наименьшее из тех n , для которых $\beta(\alpha, n) \leq \beta$. Для определения n имеем два уравнения: $\Phi(-t_{\alpha}) = \alpha$ и $\Phi(t_{\alpha} - \sqrt{n}(\theta_1 - \theta_0)/\sigma) = \beta$. Обозначим p -квантиль распределения $\mathcal{N}(0, 1)$ через ζ_p , т. е. ζ_p — решение уравнения $\Phi(\zeta) = p$. Отсюда имеем $-t_{\alpha} = \zeta_{\alpha}$, $t_{\alpha} - \sqrt{n}(\theta_1 - \theta_0)/\sigma = \zeta_{\beta}$, т. е.

$$n = \sigma^2(\zeta_{\alpha} + \zeta_{\beta})^2/(\theta_1 - \theta_0)^2.$$

Но число n^* должно быть целым, поэтому надо положить

$$n^* = [\sigma^2(\zeta_{\alpha} + \zeta_{\beta})^2/(\theta_1 - \theta_0)^2] + 1, \quad (4.17)$$

где $[a]$ — целая часть числа a . Отсюда следует, что при фиксированных ошибках число наблюдений пропорционально дисперсии и обратно пропорционально квадрату разности между средними значениями. Если, например, $\alpha = \beta = 0,001$, то $\zeta_{\alpha} = \zeta_{\beta} = -3,09$ и из формулы (4.17) имеем $n^* = [38,2\sigma^2/(\theta_1 - \theta_0)^2] + 1$. При $\sigma = 1$, $\theta_1 - \theta_0 = 1$ значение $n^* = 39$. Таким образом, если требуется различить гипотезы с указанными параметрами, то только при числе

испытаний, не меньшем 39, можно быть уверенным в том, что, поступая согласно критерию (4.12), будет принято ошибочное решение с вероятностью, не большей 0,001.

Пример 4.2 (бернуллиевская модель, проверка простых гипотез). Пусть о неизвестной вероятности «успеха» θ в бернуллиевской модели $Bi(1, \theta)$ имеются две гипотезы: $H_0: \theta = \theta_0$ и $H_1: \theta = \theta_1 > \theta_0$. Здесь для любой точки $x = (x_1, \dots, x_n)$, $x_i \in \{0, 1\}$, $i = 1, \dots, n$,

$$L(x; \theta_j) = \theta_j^r (1 - \theta_j)^{n-r}, \quad j = 0, 1,$$

где $r = r(x) = \sum_{i=1}^n x_i$ — наблюдавшееся число «успехов», а

$$l(x) = \left[\frac{\theta_1 (1 - \theta_0)}{\theta_0 (1 - \theta_1)} \right]^r \left(\frac{1 - \theta_1}{1 - \theta_0} \right).$$

Функция $\phi(\theta) = \theta/(1 - \theta)$ возрастает на интервале $(0, 1)$, поэтому при $\theta_1 > \theta_0$ $\phi(\theta_1)/\phi(\theta_0) > 1$ и неравенство $l(x) \geq c$ эквивалентно неравенству $r(x) \geq (\ln c - \ln \rho_1)/\rho$, где $\rho = \ln \{[\theta_1 (1 - \theta_0)]/[\theta_0 (1 - \theta_1)]\}$, $\rho_1 = \ln \{(1 - \theta_1)/(1 - \theta_0)\}$. Следовательно, критическая область задается в данном случае условием

$$\mathcal{X}_1^* = \{x : r(x) \geq r_\alpha\}. \quad (4.18)$$

Для определения критической границы $r = r_\alpha$ заметим, что $\mathcal{L}_{\theta_0}(r(X)) = Bi(n, \theta_0)$, поэтому условие (4.10) принимает вид

$$\sum_{m=r+1}^n C_n^m \theta_0^m (1 - \theta_0)^{n-m} < \alpha \leq \sum_{m=r}^n C_n^m \theta_0^m (1 - \theta_0)^{n-m}. \quad (4.19)$$

Если в этом соотношении имеет место знак равенства, то тем самым построен критерий Неймана—Пирсона $\mathcal{X}_1^*$ с уровнем значимости α для гипотезы H_0 против альтернативы H_1 . Здесь также критическая область не зависит от альтернативы. Поскольку $\mathcal{L}_{\theta_1}(r(X)) = Bi(n, \theta_1)$, мощность этого критерия

$$W(\mathcal{X}_1^*; \theta_1) = \sum_{m=r_\alpha}^n C_n^m \theta_1^m (1 - \theta_1)^{n-m}.$$

Чаще всего при заранее заданном α в (4.19) имеет место случай строгого неравенства; следовательно, критерий, определяемый критической областью (4.18), будет иметь уровень значимости не точно равный α , а больший, именно

$$\alpha' = \sum_{m=r_\alpha}^n C_n^m \theta_0^m (1 - \theta_0)^{n-m}.$$

Если бы в данной ситуации требовалось построить критерий с уровнем значимости, точно равным α , то следовало бы прибегнуть к рандомизации и поступить следующим образом. Положив

$$P_\alpha = C_n^r \theta_0^r (1 - \theta_0)^{n-r} = P_{\theta_0}(r(X) = r_\alpha),$$

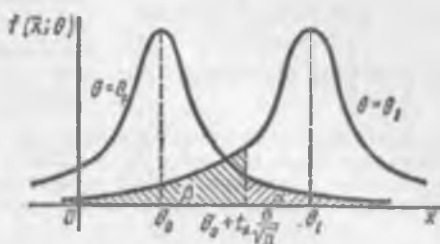


Рис. 4.1

задать критическую функцию

$$\varphi^*(x) = \begin{cases} 1 & \text{при } r(x) > r_\alpha, \\ (\rho_0 + \alpha - \alpha')/\rho_0 & \text{при } r(x) = r_\alpha, \\ 0 & \text{при } r(x) < r_\alpha, \end{cases}$$

т. е. условиться отвергать гипотезу H_0 , если $r(x) > r_\alpha$, и принимать ее, если $r(x) < r_\alpha$; если $r(x) = r_\alpha$, то отвергать H_0 с вероятностью $(\rho_0 + \alpha - \alpha')/\rho_0$ и принимать с дополнительной вероятностью $(\alpha' - \alpha)/\rho_0$. Тогда вероятность ошибки первого рода такого рандомизированного критерия равна

$$\begin{aligned} P(H_1 | H_0) &= E_{\theta_0} \varphi^*(X) = P_{\theta_0}(r(X) > r_\alpha) + \frac{\rho_0 + \alpha - \alpha'}{\rho_0} P_{\theta_0}(r(X) = r_\alpha) = \\ &= \alpha' - \rho_0 + \frac{\rho_0 + \alpha - \alpha'}{\rho_0} \rho_0 = \alpha. \end{aligned}$$

Мощность этого критерия вычисляется по формуле

$$W(\varphi^*; \theta_1) = E_{\theta_1} \varphi^*(X) =$$

$$= \sum_{m=r_\alpha+1}^n C_n^m \theta_1^m (1-\theta_1)^{n-m} + (\rho_0 + \alpha - \alpha') \left(\frac{\theta_1}{\theta_0}\right)^{r_\alpha} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{n-r_\alpha}.$$

Случай $\theta_1 < \theta_0$ рассматривается аналогично, при этом критическая область имеет вид $\{r(x) \leq r_\alpha\}$.

§ 4.3. Выбор из двух простых гипотез. Понятие о последовательном анализе

В примере 4.1 при рассмотрении задачи выбора одной из двух гипотез о нормальном распределении была установлена [см. соотношение (4.17)] связь между числом необходимых наблюдений и значениями вероятностей ошибок α и β . Это число можно рассчитать заранее (до проведения испытаний), и оно не зависит от исходов самих испытаний. Кроме правил проверки гипотез, основанных на выборках фиксированного объема, известны *последовательные* правила (критерии). В случае использования этих правил вопрос о числе необходимых наблюдений решают в процессе наблюдений; следовательно, это число (объем выборки) является случайной величиной. Последовательные правила впервые были предложены А. Вальдом (1947 г.), и их изучение составляет предмет важного раздела математической статистики — *последовательного анализа*. В этом параграфе кратко на простейшем примере различения двух простых гипотез о распределении наблюдаемой случайной величины ξ будут рассмотрены некоторые особенности последовательных критериев. С общей теорией последовательного анализа можно ознакомиться в [3] и [24], где дана современная трактовка соответствующих задач.

1. Определение критерия Вальда. Пусть, как обычно, x_i — наблюдавшаяся реализация случайной величины ξ в i -м испытании, $i = 1, 2, \dots, n$ и $L_{j/n} =$

$$= L(x_1, \dots, x_n; \theta_j) = \prod_{i=1}^n f_j(x_i) \text{ — функция правдоподобия для первых } n \text{ испы-}$$

таний при условии, что истинной является гипотеза $H_j: \theta = \theta_j$, $j = 0, 1$ (как и в § 4.2, $f_j(x)$ — плотность распределения ξ (или вероятность в дискретном случае) при гипотезе H_j). Согласно теории Неймана — Пирсона, наилучшая процедура проверки гипотезы H_0 против альтернативы H_1 состоит в принятии или отклонении гипотезы H_0 в зависимости от того, меньше или больше отношение правдоподобия $L_{1/n}/L_{0/n}$ некоторой выбранной константы c , при этом объем выборки n фиксируется заранее и не зависит от наблюдений. Однако если объем выборки сделать случайным и зависящим от исходов наблюдений, то можно добиться выигрыша в среднем числе наблюдений до принятия окончательного решения.

Последовательный критерий отношения вероятностей (критерий Вальда) состоит в следующем. Задаются две положительные константы $A_0 < 1 < A_1$ и наблюдения производятся до тех пор, пока не будет впервые нарушено какое-нибудь из неравенств

$$A_0 < L_{1n}/L_{0n} < A_1. \quad (4.20)$$

Если в момент прекращения испытаний (момент остановки) $L_{1n}/L_{0n} \leq A_0$, то принимается гипотеза H_0 ; если $L_{1n}/L_{0n} \geq A_1$, то принимается гипотеза H_1 . Эта процедура характеризуется обычно вероятностями ошибок первого и второго рода $\alpha = P(H_1 | H_0)$ и $\beta = P(H_0 | H_1)$ и средним числом $E_j(v) =$

$E(v | H_j)$ наблюдений v до момента остановки ($j=0, 1$). Если вероятности ошибок α и β заданы, то любой критерий с такими ошибками называют критерием силы (α, β) . В классе критериев данной силы (α, β) предпочтительным является тот, который требует меньшего числа наблюдений. Критерий, минимизирующий одновременно как $E_0(v)$, так и $E_1(v)$, называют оптимальным. Оптимальным свойством обладает критерий Вальда. В частности, этот критерий требует в среднем меньше наблюдений, чем критерий Неймана—Пирсона с такими же вероятностями ошибок. Эти и другие свойства критерия Вальда рассмотрены ниже.

2. О числе испытаний до момента остановки в критерии Вальда. Пусть функции $f_j(x) > 0$, $j=0, 1$, при всех возможных значениях случайной величины ξ и не тождественны (иначе гипотезы H_0 и H_1 неразличимы). Это означает, что определена и не вырождена случайная величина $Z = \ln(f_1(\xi)/f_0(\xi))$; будем предполагать, что существуют $E_0 Z \neq 0$ и $D_0 Z = \sigma^2(\theta) > 0$ ($\theta = \theta_0, \theta_1$). Обозначим также $Z_i = \ln(f_1(X_i)/f_0(X_i))$, $i=1, 2, \dots$, где $X_1, X_2, \dots$ — последовательные независимые наблюдения над ξ . Тогда $Z_1, Z_2, \dots$ — независимые наблюдения над Z , и если $z_1, z_2, \dots$ — наблюдавшиеся реализации этих величин, то в этих обозначениях процедура проверки заканчивается принятием гипотезы H_0 или H_1 при первом n , при котором нарушается какое-нибудь из неравенств

$$a_0 = \ln A_0 < z_1 + \dots + z_n < \ln A_1 = a_1, \quad a_0 < 0, a_1 > 0. \quad (4.21)$$

Критерию можно дать следующую наглядную геометрическую интерпретацию. Рассмотрим блуждающую на плоскости частицу, находящуюся в начальный момент времени в начале координат, ордината которой в каждый целочисленный момент времени $t=1, 2, \dots$ получает приращение z_t . Тогда ордината частицы в момент $t=n$ будет равна $z_1 + \dots + z_n$ и блуждание продолжится, пока траектория частицы впервые не выйдет за пределы полосы, ограниченной двумя горизонтальными прямыми на уровнях a_0 и a_1 (рис. 4.2). Выход траектории на верхнюю (нижнюю) границу приводит к принятию гипотезы H_1 (H_0). Ответ на вопрос о том, не может ли блуждание продолжаться бесконечно долго (в этом случае процесс проверки гипотезы никогда бы не закончился), дает следующее утверждение.

Теорема 4.2. Критерий Вальда с вероятностью 1 заканчивается за конечное число шагов, т. е. $\lim_{n \rightarrow \infty} P_0(v > n) = 0$ ($\theta = \theta_0, \theta_1$).

□ Зафиксируем некоторое целое r и введем случайные величины $\eta_1 = Z_1 + \dots + Z_r$, $\eta_2 = Z_{r+1} + \dots + Z_{2r}$, ... Тогда событие $\{v > rk\}$, эквивалентное событию $a_0 < Z_1 + \dots + Z_i < a_1$, $i \leq rk$, включается в событие $a_0 < \eta_1 + \dots + \eta_j < a_1$, $j \leq k$, которое, в свою очередь, включается в событие $|\eta_j| < b = a_1 - a_0$, $j \leq k$. Величины η_j независимы и одинаково распределены, поэтому

$$P_0(v > rk) \leq p^k(\theta), \quad (4.22)$$

где $p(\theta) = P_0(|\eta_1| < b) = P_0(\eta_1^2 < b^2)$. Но $E_0 \eta_1^2 = D_0 \eta_1 = r\sigma^2(\theta) > b^2$, если выбрать $r > \max(b^2/\sigma^2(\theta_0), b^2/\sigma^2(\theta_1))$. Отсюда следует, что, выбирая r , можно обеспечить неравенство $p(\theta) < 1$. Но тогда, переходя в неравенство (4.22)



Рис. 4.2

к пределу при $k \rightarrow \infty$, получаем

$$\lim_{n \rightarrow \infty} P_0(v > n) = \lim_{k \rightarrow \infty} P_0(v > rk) = 0. \blacksquare$$

Итак, в критерии Вальда число испытаний v до момента остановки с вероятностью 1 принимает конечное значение. Справедливо и более сильное утверждение.

Теорема 4.3. Все моменты случайной величины v конечны.

□ Пусть $t \geq 0$; тогда для производящей функции моментов

$$\varphi(t) = E_0 e^{tv} = \sum_{k=1}^{\infty} e^{tk} P_0(v=k)$$

имеет место следующая оценка:

$$\varphi(t) \leq \sum_{k \geq 0} e^{t(k+1)} r P_0(kr < v \leq (k+1)r) \leq e^{tr} \sum_{k \geq 0} e^{tkr} P_0(v > kr).$$

Но из соотношения (4.22) следует, что $P_0(v > kr) \leq p^k$, $p = \max(p(\theta_0), p(\theta_1))$, и, выбирая r , можно обеспечить неравенство $p < 1$. Из этих неравенств имеем, что для любого $t \geq 0$

$$\varphi(t) \leq e^{tr}/(1 - pe^{tr}),$$

при этом $pe^{tr} < 1$. Аналогично, если $t \leq 0$, то

$$\varphi(t) \leq \sum_{k \geq 0} e^{tkr} P_0(kr < v \leq (k+1)r) \leq 1/(1 - pe^{tr}).$$

Для любого $t < (1/r) \ln(1/p)$ ряд для $\varphi(t)$ сходится; следовательно, существуют все производные $\varphi^{(k)}(0)$, а тем самым и все моменты $E_0(v^k) = \varphi^{(k)}(0)$. ■

Из этой теоремы, в частности, следует, что для критерия Вальда величины $E_0(v)$ и $E_1(v)$ конечны. Вычисление этих величин при заданных ошибках (α, β) рассмотрено в п. 4.

3. О выборе границ в критерии Вальда. Выясним теперь, как связаны границы A_0 и A_1 в (4.20) [или, что то же самое, границы a_0 и a_1 в (4.21)] с вероятностями ошибок α и β .

Теорема 4.4. Постоянные A_0 и A_1 критерия Вальда силы (α, β) удовлетворяют неравенствам

$$A_0 \geq A'_0 = \beta/(1 - \alpha), \quad A_1 \leq A'_1 = (1 - \beta)/\alpha. \quad (4.23)$$

При этом если границы A_0 и A_1 заменить их оценками A'_0 и A'_1 , то сила полученного критерия будет равна (α', β') , где

$$\alpha' \leq \alpha/(1 - \beta), \quad \beta' \leq \beta/(1 - \alpha) \quad \text{и} \quad \alpha' + \beta' \leq \alpha + \beta. \quad (4.24)$$

□ Обозначим через $\mathcal{X}_{0n}(\mathcal{X}_{1n})$ множество тех результатов наблюдений $(x_1, \dots, x_n)$, для которых процедура заканчивается на $v=n$ шаге принятием $H_0(H_1)$, т. е., например,

$$\mathcal{X}_{0n} = \{(x_1, \dots, x_n): A_0 < L_{1k}/L_{0k} < A_1, k=1, \dots, n-1, L_{1n}/L_{0n} \leq A_0\}.$$

В силу теоремы 4.2

$$\sum_{n=1}^{\infty} P_0(v=n) = \sum_{n=1}^{\infty} P_0(\mathcal{X}_{0n}) + \sum_{n=1}^{\infty} P_0(\mathcal{X}_{1n}) = 1 \quad (\theta = \theta_0, \theta_1).$$

Далее имеем

$$\begin{aligned} \alpha = P(H_1 | H_0) &= \sum_{n=1}^{\infty} P_{\theta_0}(\mathcal{X}_{1n}) \leq \frac{1}{A_1} \sum_{n=1}^{\infty} P_{\theta_1}(\mathcal{X}_{1n}) = \\ &= \frac{1}{A_1} (1 - P(H_0 | H_1)) = \frac{1 - \beta}{A_1}, \end{aligned}$$

так как в точках множества $\mathcal{X}_{1n}$ выполняется неравенство $L_{0n} \leq L_{1n}/A_1$. Аналогично получаем

$$\begin{aligned} \beta &= P(H_0 | H_1) = \sum_{n=1}^{\infty} P_{\theta_1}(\mathcal{X}_{0n}) \leq A_0 \sum_{n=1}^{\infty} P_{\theta_0}(\mathcal{X}_{0n}) = \\ &= A_0 (1 - P(H_1 | H_0)) = A_0 (1 - \alpha), \end{aligned}$$

поскольку в точках множества $\mathcal{X}_{0n}$ выполняется неравенство $L_{1n} \leq A_0 L_{0n}$. Тем самым неравенства (4.23) доказаны.

Рассмотрим теперь критерий Вальда с границами A'_0 и A'_1 , определенными в (4.23), и пусть α' и β' — его ошибки. Тогда на основании первой части теоремы должны выполняться неравенства $\beta/(1-\alpha) \geq \beta'/(1-\alpha')$, $(1-\beta)/\alpha \leq (1-\beta')/\alpha'$. Из первого и второго неравенств соответственно имеем $\beta' \leq (1-\alpha')\beta/(1-\alpha) \leq \beta/(1-\alpha)$, $\alpha' \leq (1-\beta')\alpha/(1-\beta) \leq \alpha/(1-\beta)$. Складывая неравенства $\beta - \beta' \geq \beta' - \alpha\beta'$ и $-\alpha' + \alpha'\beta \geq -\alpha + \alpha\beta'$, получаем, что $\beta - \alpha' \geq \beta' - \alpha$ или $\alpha' + \beta' \leq \alpha + \beta$. ■

На практике теорему 4.4 используют следующим образом. Если требуется построить критерий Вальда силы (α, β) , то границы в (4.20) полагают равными соответственно A'_0 и A'_1 [см. (4.23)]. В этом случае последнее неравенство (4.24) гарантирует, что сумма действительных ошибок α' и β' такого критерия не превосходит суммы $\alpha + \beta$ заданных ошибок; далее, обычно α и β не превышают 0,1, поэтому из остальных неравенств (4.24) следует, что разность между действительными и заданными ошибками в этих случаях незначительна. Таким образом, можно утверждать, что такой специальный выбор границ в критерии Вальда приводит к более сильному критерию.

Отметим в этой связи интересную особенность последовательного критерия по сравнению с обычными критериями. В обычном критерии для определения критической области при выбранном уровне значимости и вычисления вероятности ошибки второго рода (или мощности) надо знать распределение статистики критерия и при нулевой гипотезе и при альтернативе, при расчете же критерия Вальда не возникает проблемы отыскания распределений. Действительно, границы A'_0 и A'_1 зависят только от заданных ошибок α и β и отношение L_{1n}/L_{0n} можно вычислить на основании данных задачи без отыскания каких-либо распределений. Необходимость иметь информацию о распределениях при использовании последовательного критерия возникает только при нахождении закона распределения числа наблюдений v до принятия решения.

4. О среднем числе наблюдений в критерии Вальда. Рассмотрим вопрос о вычислении среднего числа наблюдений v в критерии Вальда. Предварительно установим одно интересное равенство, которое называется тождеством Вальда.

Пусть $S_v = Z_1 + \dots + Z_v$ — ордината блуждающей частицы в момент прекращения блуждания (см. п. 2). Тогда имеет место *тождество Вальда*

$$E_{\theta}(S_v) = E_{\theta}(Z) E_{\theta}(v). \quad (4.25)$$

□ Введем случайные величины $Y_1, Y_2, \dots$, где

$$Y_n = \begin{cases} 1, & \text{если решение не принято до } n\text{-го шага,} \\ 0 & \text{в противном случае.} \end{cases}$$

Тогда Y_n , очевидно, есть функция только $Z_1, \dots, Z_{n-1}$ и, следовательно, не зависит от Z_n . Нетрудно показать, что $S_v = \sum_{n=1}^{\infty} Y_n Z_n$, откуда

$$E_{\theta}(S_v) = \sum_{n=1}^{\infty} E_{\theta}(Y_n Z_n) = E_{\theta}(Z) \sum_{n=1}^{\infty} E_{\theta}(Y_n) = E_{\theta}(Z) \sum_{n=1}^{\infty} P_{\theta}(Y_n = 1). \quad (4.26)$$

Событие $\{Y_n = 1\}$ эквивалентно событию $\{v \geq n\}$, а так как для любой целочисленной положительной случайной величины x $E x = P(x=1) + 2P(x=2) +$

$$+ 3P(x=3) + \dots = \sum_{k=1}^{\infty} P(x=k) + \sum_{k=2}^{\infty} P(x=k) + \sum_{k=3}^{\infty} P(x=k) + \dots = \\ = \sum_{n=1}^{\infty} P(x \geq n), \text{ то из (4.26) следует тождество (4.25). } \blacksquare$$

Пусть теперь заданы значения ошибок α и β . Как показано в п. 3, практически при малых α и β границы в (4.21) можно заменить соответственно на

$$a'_0 = \ln A'_0 = \ln \frac{\beta}{1-\alpha} \text{ и } a'_1 = \ln A'_1 = \ln \frac{1-\beta}{\alpha}. \quad (4.27)$$

При малых α и β длина интервала (a'_0, a'_1) велика, а величина $E_0(Z)$ от α и β не зависит и, по предположению (см. п. 2), конечна. Поэтому можно пренебречь эффектом превышения суммой S_v границы в момент остановки, т. е. положить $S_v \approx a'_j$, если принимается гипотеза H_j ($j=0, 1$). Другими словами, можно считать среднее значение суммы S_v равным приблизительно соответствующей границе a'_0 или a'_1 . Из этих рассуждений и тождества Вальда (4.25) имеем

$$E_{\theta_0}(v) E_{\theta_0}(Z) = E_{\theta_0}(S_v) \approx a'_0 P(H_0 | H_0) + a'_1 P(H_1 | H_0) = (1-\alpha) a'_0 + \alpha a'_1.$$

Аналогично получаем

$$E_{\theta_1}(v) E_{\theta_1}(Z) = E_{\theta_1}(S_v) \approx a'_0 P(H_0 | H_1) + a'_1 P(H_1 | H_1) = \beta a'_0 + (1-\beta) a'_1.$$

Таким образом, $E_j(v) = E_{\theta_j}(v)$ можно вычислять по следующим приближенным формулам:

$$E_0(v) \approx \frac{(1-\alpha) a'_0 + \alpha a'_1}{E_0(Z)}, \quad E_1(v) \approx \frac{\beta a'_0 + (1-\beta) a'_1}{E_1(Z)}, \quad (4.28)$$

где a'_0 и a'_1 определены в (4.27).

Для произвольного последовательного критерия такой же силы (α, β) можно показать, что среднее число наблюдений до момента остановки при гипотезе H_j оценивается снизу правой частью соответствующего равенства (4.28). Таким образом, среднее число наблюдений для критерия Вальда близко к наименьшему значению, т. е. этот критерий является (с принятой выше степенью приближения) оптимальным.

5. Пример «экономичности» последовательного критерия. Рассмотрим на примере, насколько «экономична» последовательная процедура. Пусть для различения двух простых гипотез о среднем нормальной случайной величины ξ , сформулированных в примере 4.1, применяется последовательный критерий отношения вероятностей. Тогда

$$z_i = \ln \frac{f_1(x_i)}{f_0(x_i)} = \frac{(x_i - \theta_0)^2 - (x_i - \theta_1)^2}{2\sigma^2} = \frac{\theta_1 - \theta_0}{\sigma^2} \left(x_i - \frac{\theta_0 + \theta_1}{2} \right) \quad (4.29)$$

и критерий Вальда силы (α, β) определяется в данном случае следующим образом. Наблюдения продолжают до тех пор, пока впервые не нарушится какое-нибудь из неравенств

$$\frac{\sigma^2}{\theta_1 - \theta_0} \ln \frac{\beta}{1-\alpha} < \sum_{i=1}^n x_i - \frac{n}{2} (\theta_0 + \theta_1) < \frac{\sigma^2}{\theta_1 - \theta_0} \ln \frac{1-\beta}{\alpha};$$

если нарушается левое (правое) неравенство, то принимается гипотеза H_0 (H_1). Вычислим среднее число наблюдений до принятия решения. Из (4.29) имеем

$$E_{\theta_j}(Z) = \frac{\theta_1 - \theta_0}{\sigma^2} \left(\theta_j - \frac{\theta_0 + \theta_1}{2} \right) = \begin{cases} -(\theta_1 - \theta_0)^2 / (2\sigma^2) & \text{при } j=0, \\ (\theta_1 - \theta_0)^2 / (2\sigma^2) & \text{при } j=1. \end{cases}$$

Отсюда и из (4.28) находим

$$E_0(v) = -\frac{2\sigma^2}{(\theta_1 - \theta_0)^2} \left[(1-\alpha) \ln \frac{\beta}{1-\alpha} + \alpha \ln \frac{1-\beta}{\alpha} \right],$$

$$E_1(v) = -\frac{2\sigma^2}{(\theta_1 - \theta_0)^2} \left[\beta \ln \frac{\beta}{1-\alpha} + (1-\beta) \ln \frac{1-\beta}{\alpha} \right].$$

Для критерия Неймана—Пирсона с теми же ошибками α и β необходимое число испытаний n^* определяется формулой (4.17). Отношение среднего объема выборки в последовательном критерии Вальда, когда истинной является гипотеза H_0 , к объему n^* равно

$$\frac{E_0(v)}{n^*} \approx -\frac{2}{(\zeta_\alpha + \zeta_\beta)^2} \left((1-\alpha) \ln \frac{\beta}{1-\alpha} + \alpha \ln \frac{1-\beta}{\alpha} \right) \quad (4.30)$$

и не зависит ни от σ^2 , ни от $\theta_1 - \theta_0$. (Напомним, что ζ_α определяется уравнением $\Phi(\zeta_\alpha) = \alpha$.) В том случае, когда справедлива гипотеза H_1 , это отношение равно

$$\frac{E_1(v)}{n^*} \approx -\frac{2}{(\zeta_\alpha + \zeta_\beta)^2} \left(\beta \ln \frac{\beta}{1-\alpha} + (1-\beta) \ln \frac{1-\beta}{\alpha} \right). \quad (4.31)$$

Если задаться ошибками $\alpha = \beta$, то $\zeta_\alpha = \zeta_\beta$, $\ln \frac{\beta}{1-\alpha} = -\ln \frac{1-\beta}{\alpha}$ и из соотношений (4.30)–(4.31) следует, что

$$\frac{E_0(v)}{n^*} \approx \frac{E_1(v)}{n^*} \approx \frac{1-2\alpha}{2\zeta_\alpha^2} \ln \frac{1-\alpha}{\alpha} = \psi(\alpha). \quad (4.32)$$

При $\alpha = 0.05$ имеем $\zeta_\alpha = 1.6449$ и, вычисляя по формуле (4.32), получаем $\psi(\alpha) = 0.4897$; это свидетельствует, что экономия наблюдений составляет приблизительно 50 %.

Если воспользоваться известной асимптотической формулой для нормальной функции распределения $\Phi(x)$ [20, с. 183]

$$1 - \Phi(x) = \frac{e^{-x^2/2}}{x\sqrt{2\pi}} (1 + o(1)), \quad x \rightarrow \infty, \quad (4.33)$$

то можно показать, что при $\alpha \rightarrow 0$

$$\psi(\alpha) = \frac{1}{4} \left(1 + \frac{\ln \ln(1/\alpha) + \ln 4\pi + o(1)}{2 \ln(1/\alpha)} \right). \quad (4.34)$$

Действительно, так как $\Phi(\zeta_\alpha) = \alpha$, то $\zeta_\alpha \rightarrow -\infty$ при $\alpha \rightarrow 0$. Но $\Phi(\zeta_\alpha) = 1 - \Phi(-\zeta_\alpha)$, поэтому на основании (4.33) можно записать асимптотическое равенство

$$-\frac{1+o(1)}{\zeta_\alpha \sqrt{2\pi}} e^{-\zeta_\alpha^2/2} = \alpha.$$

Непосредственной подстановкой нетрудно убедиться, что этому соотношению удовлетворяет величина ζ_α , определяемая равенством

$$\zeta_\alpha^2 = 2 \ln(1/\alpha) - \ln \ln(1/\alpha) - \ln 4\pi + o(1).$$

Отсюда и из (4.32) получаем представление (4.34), из которого следует, что при ошибках $\alpha = \beta \rightarrow 0$ последовательная процедура в среднем приблизительно в четыре раза экономнее оптимальной процедуры с фиксированным объемом выборки.

§ 4.4. Сложные гипотезы

Случай, когда основная и альтернативная гипотезы являются простыми, встречается в приложениях сравнительно редко. На практике чаще имеют место ситуации, когда

обе гипотезы (или по крайней мере одна из них) сложные. В этих случаях равномерно наиболее мощные (р. н. м.) критерии существуют только если допустимые распределения (статистическая модель) удовлетворяют определенным ограничениям. Рассмотрим некоторые наиболее типичные случаи.

1. Р. н. м. критерии против сложных альтернатив. Модели с монотонным отношением правдоподобия. Пусть проверяется простая гипотеза $H_0: \theta = \theta_0$ против сложной альтернативной гипотезы $H_1: \theta \in \Theta_1 = \Theta \setminus \{\theta_0\}$. Построим для каждой фиксированной альтернативы $\theta_1 \in \Theta_1$ критерий Неймана — Пирсона $\mathcal{X}_{1\alpha}^* = \mathcal{X}_{1\alpha}^*(\theta_0, \theta_1)$, как это описано в § 4.2. Если $\mathcal{X}_{1\alpha}^*(\theta_0, \theta_1)$ зависит от альтернативы θ_1 , то это означает, что не существует критической области, которая была бы наилучшей для всех $\theta_1 \in \Theta_1$. Другими словами, в данной задаче не существует р. н. м. критерия. Однако может оказаться, что критерий $\mathcal{X}_{1\alpha}^*(\theta_0, \theta_1)$ не зависит от конкретной альтернативы θ_1 : $\mathcal{X}_{1\alpha}^*(\theta_0, \theta_1) = \mathcal{X}_{1\alpha}^*(\theta_0)$. Тогда соответствующая критическая область максимизирует мощность при любой допустимой альтернативе и поэтому $\mathcal{X}_{1\alpha}^*(\theta_0)$ является р. н. м. критерием проверки гипотезы $H_0: \theta = \theta_0$ против альтернативы $H_1: \theta \in \Theta_1$. Такие случаи уже были рассмотрены в § 4.2. Так, в примере 4.1 построен критерий Неймана — Пирсона для проверки гипотезы $H_0: \theta = \theta_0$ относительно неизвестного среднего нормальной модели $\mathcal{N}(\theta, \sigma^2)$. Было получено, что при альтернативе $\theta = \theta_1 > \theta_0$ критическая область $\mathcal{X}_{1\alpha}^*$, определенная в (4.12), не зависит от θ_1 ; следовательно, этот критерий является одновременно р. н. м. критерием при сложной *правосторонней* альтернативе $H_1^+: \theta > \theta_0$. Аналогично, р. н. м. критерием при *левосторонней* альтернативе $H_1^-: \theta < \theta_0$ в рассматриваемой модели является критерий (4.14). Таким образом, здесь имеем примеры существования р. н. м. критериев при сложных *односторонних* альтернативах. Если же рассматривать полный класс альтернатив $H_1 = H_1^- \cup H_1^+: \theta \neq \theta_0$, то р. н. м. критерия не существует, так как всегда один из двух критериев (4.12) или (4.14) лучше любого другого критерия с тем же уровнем значимости. Аналогичные выводы можно сделать и на основании результатов примера 4.2.

Сформулируем общее достаточное условие существования р. н. м. критерия в случае односторонних альтернатив. Итак, предположим, что параметр θ — скалярный и альтернативная гипотеза H_1 — односторонняя, т. е. либо $H_1 = H_1^+: \theta > \theta_0$, либо $H_1 = H_1^-: \theta < \theta_0$. Пусть, далее, семейство допустимых распределений $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$ (здесь $\Theta = \{\theta \geq \theta_0\}$ или $\Theta = \{\theta \leq \theta_0\}$) обладает достаточной статистикой $T = T(X)$, $X = (X_1, \dots, X_n)$ (см. § 2.3). Тогда из критерия факторизации [см. (2.36)] следует, что статистика отношения правдоподобия в данном случае имеет вид $l(X) = g(T(X); \theta_1)/g(T(X); \theta_0)$

и, следовательно, критерий Неймана — Пирсона формулируется в терминах достаточной статистики [соответствующая критическая область выражается через $T(X)$].

Выделим важный класс моделей $\mathcal{F}$, для которых отношение $g(T; \theta_1)/g(T; \theta_0)$ является монотонной функцией T ; говорят, что такие модели имеют *монотонное отношение правдоподобия* (многие модели, встречающиеся в статистике, обладают этим свойством). Для таких моделей в задаче проверки простой гипотезы $H_0: \theta = \theta_0$ против односторонней альтернативы H_1 существует р. н. м. критерий, который совпадает с критерием Неймана—Пирсона проверки гипотезы H_0 против произвольной фиксированной альтернативы из H_1 .

Действительно, пусть, например, $H_1 = H_1^+$ и при фиксированном $\theta_1 > \theta_0$ отношение $g(T; \theta_1)/g(T; \theta_0)$ монотонно возрастает по T . Рассмотрим критерий Неймана—Пирсона проверки гипотезы H_0 против альтернативы $\theta = \theta_1$. В данном случае неравенство $l(x) \geq c$ эквивалентно неравенству $T(x) \geq c_\alpha^*$, причем граница c_α^* определяется заданным уровнем значимости α и распределением $F(x; \theta_0)$. Таким образом, этот критерий не зависит от выбора альтернативы θ_1 и поэтому одновременно является р. н. м. критерием при сложной альтернативе H_1^+ . Если отношение $g(T; \theta_1)/g(T; \theta_0)$ убывает по T , то неравенство для $T(x)$ заменяется на противоположное. Аналогично рассматривается случай левосторонней альтернативы H_1^- .

Пример 4.3 (*экспоненциальная модель, р. н. м. критерий при односторонней альтернативе*). Рассмотрим экспоненциальную модель, введенную в п. 3 § 2.2. Из представления (2.25) следует,

что статистика $T(X) = \sum_{i=1}^n B(X_i)$ является достаточной и

$$l(X) = \exp \{ (A(\theta_1) - A(\theta_0)) T(X) + n(c(\theta_1) - c(\theta_0)) \}.$$

Следовательно, если функция $A(\theta)$ строго монотонна, то l — монотонная функция T , поэтому для таких моделей всегда существует р. н. м. критерий проверки гипотезы $H_0: \theta = \theta_0$ против односторонних альтернатив. При этом если $A(\theta)$ — возрастающая функция, то при альтернативе $H_1^+: \theta > \theta_0$ критическая область имеет вид $\{T(x) \geq c_\alpha^*\}$, а при альтернативе $H_1^-: \theta < \theta_0$ — вид $\{T(x) \leq c_\alpha^*\}$. Если же функция $A(\theta)$ убывает, то неравенства, определяющие критические области, меняются на противоположные.

Пример 4.4 (*нормальная-2 модель, р. н. м. критерий при односторонней альтернативе*). Пусть в модели $\mathcal{N}(\mu, \theta^2)$ требуется проверить простую гипотезу $H_0: \theta^2 = \theta_0^2$ о неизвестной дисперсии.

Здесь имеет место экспоненциальная модель с $T(X) = \sum_{i=1}^n (X_i - \mu)^2$ и $A(\theta) = -1/(2\theta^2)$. Поэтому р. н. м. критерий против альтернативы $H_1^+: \theta^2 > \theta_0^2$ задается критической областью $\mathcal{X}_{1-\alpha}^* = \{T(x) \geq c_\alpha^*\}$, а против альтернативы $H_1^-: \theta^2 < \theta_0^2$ — областью $\mathcal{X}_{1-\alpha}^* = \{T(x) \leq c_\alpha^*\}$. При этом так как $\mathcal{L}_\theta(T(X)/\theta^2) = \chi^2(n)$, то $c_\alpha^+ = \theta_0^2 \chi_{1-\alpha, n}^2$, $c_\alpha^- = \theta_0^2 \chi_{\alpha, n}^2$, где $\chi_{p, n}^2$ — p -квантиль распределения $\chi^2(n)$.

Замечание 1. Можно доказать [14, с. 101], что для моделей с монотонным отношением правдоподобия р. н. м. критерий проверки простой гипотезы

тезы $H_0: \theta = \theta_0$ против правосторонней альтернативы $H_1: \theta > \theta_0$ является одновременно р. н. м. критерием проверки сложной гипотезы $H_0: \theta \leq \theta_0$ против $H_1: \theta > \theta_0$ того же уровня значимости.

Аналогичное утверждение справедливо и для двойственной проблемы проверки $H_2: \theta \geq \theta_0$ против левосторонней альтернативы $H_1: \theta < \theta_0$.

Пример 4.5 (бернуллиевская модель, р. н. м. критерий для односторонних гипотез). В процессе производства изделия обычно подвергают выборочному статистическому контролю. Предположим, что каждое изделие независимо от других может оказаться дефектным с некоторой одной и той же, но неизвестной вероятностью θ ($0 < \theta < 1$) и исправным с дополнительной вероятностью $1 - \theta$. Пусть для контроля взято n изделий и результат описывается вектором $X = (X_1, \dots, X_n)$, где $X_i = 1$, если i -е изделие дефектно, и $X_i = 0$ в противном случае. Часто бывает нужно проверить гипотезу $H_0: \theta \geq \theta_0$, где θ_0 — некоторая критическая доля брака (если гипотеза H_0 истинна, то процесс производства необходимо приостановить и усовершенствовать для уменьшения доли брака).

При сделанных предположениях число $r(X) = \sum_{i=1}^n X_i$ дефектных изделий в выборке является достаточной статистикой и $\mathcal{L}_\theta(r(X)) = Bi(n, \theta)$. Но биномиальная модель является частным случаем экспоненциальной модели с монотонным отношением правдоподобия, поэтому в силу замечания 1 существует р. н. м. критерий проверки гипотезы H_0 против альтернативы $H_1: \theta < \theta_0$, который задается критической областью вида $\{r(x) \leq r_\alpha\}$. Рассчитывают этот критерий (вычисляют границу r_α) так же, как в примере 4.2.

Пример 4.6 (бернуллиевская модель (продолжение)). В условиях предыдущего примера иногда применяют другой план контроля, который называется *обратным биномиальным выбором*. В этом случае испытания продолжают до получения заданного числа r «успехов» (здесь под «успехом» понимается появление дефектного изделия). Пусть Y_i — число испытаний между $(i-1)$ -м и i -м успехами. Тогда $P_\theta(Y_i = y) = \theta(1-\theta)^{y-1}$, $y = 0, 1, \dots$, т. е. $Y = (Y_1, \dots, Y_r)$ — выборка из распределения $Bi(1, 1-\theta)$ (см. табл. В.1), которое является распределением экспоненциального типа с $T(Y) =$

$= \sum_{i=1}^r Y_i$ и $A(\theta) = \ln(1-\theta)$. Так как $A(\theta)$ — убывающая функция θ ,

то существует р. н. м. критерий проверки гипотезы $H_0: \theta \geq \theta_0$, при альтернативе $H_1: \theta < \theta_0$, который задается критической областью вида $\{T(y) \geq c\}$. Этот вывод совпадает и с интуитивным представлением, так как при больших значениях θ естественно ожидать, что r -й успех произойдет достаточно быстро. Для определения критической границы c используют тот факт, что $\mathcal{L}_\theta(T(Y)) = Bi(r, 1-\theta)$.

Пример 4.7 (модель гамма, р. н. м. критерий для односторонних гипотез). Предположим, что время «жизни» некоторого устройства есть случайная величина ξ , распределенная по экспоненциальному закону с неизвестным параметром $\theta > 0$, т. е. $\mathcal{L}(\xi) =$

$= \Gamma(0, 1)$. Пусть $X_1, X_2, \dots, X_n$ — времена «жизни», полученные при испытании n контрольных устройств. Так как гамма-модель является экспоненциальной моделью и для нее $T(X) = \sum_{i=1}^n X_i$ и $A(\theta) = -1/\theta$, то р. н. м. критерий проверки гипотезы $H_0: \theta \geq \theta_0$ при альтернативе $H_1: \theta < \theta_0$ существует и задается критической областью вида $\{T(X) \leq c\}$. Это также ожидаемый результат, поскольку параметр θ имеет здесь смысл среднего величины $\xi: E_0 \xi = \theta$, и слишком малые значения T естественно интерпретировать в пользу альтернативы.

Заметим далее, что случайная величина $2X_i/\theta$ имеет плотность $e^{-x}/2, x > 0$, т. е. $\mathcal{L}_\theta(2X_i/\theta) = \chi^2(2)$ [см. формулу (1.27)]. Отсюда на основании воспроизводящего свойства распределения χ^2 (см. п. 1 § 1.5) имеем, что $\mathcal{L}_\theta(2T(X)/\theta) = \chi^2(2n)$ и, следовательно, границу критической области можно определить по таблицам χ^2 -распределения. Именно: так как $P_{\theta_0}\left(\frac{2}{\theta_0} T(X) \leq \chi^2_{\alpha, 2n}\right) = \alpha$, то критическая область

$$\mathcal{R}_{1\alpha}^* = \left\{x: T(x) \leq \frac{\theta_0}{2} \chi^2_{\alpha, 2n}\right\}$$

задает р. н. м. критерий уровня значимости α в данной задаче.

Пример 4.8 (экспоненциальная модель с векторным параметром, р. н. м. критерий для нее). Пусть семейство допустимых распределений задается плотностью $f(x; \theta) = \frac{1}{\theta_2} e^{-(x-\theta_1)/\theta_2}, x \geq \theta_1$, т. е. здесь $\Theta = \{\theta = (\theta_1, \theta_2): -\infty < \theta_1 < \infty, \theta_2 > 0\}$. Проверим простую гипотезу $H_0: \theta = \theta_0 = (\theta_{10}, \theta_{20})$ против альтернативы $H_1: \theta_1 < \theta_{10}, \theta_2 < \theta_{20}$. Следуя общему принципу, зафиксируем некоторую альтернативу $\theta_1 = (\theta_{11}, \theta_{21})$ и построим критерий Неймана — Пирсона проверки гипотезы H_0 против этой альтернативы. Здесь

$$l(x) = \frac{L(x; \theta_1)}{L(x; \theta_0)} = \begin{cases} \left(\frac{\theta_{20}}{\theta_{21}}\right)^n \exp\left\{n\left(\frac{1}{\theta_{20}} - \frac{1}{\theta_{21}}\right)x + n\left(\frac{\theta_{11}}{\theta_{21}} - \frac{\theta_{10}}{\theta_{20}}\right)\right\} & \text{при } x_{(1)} = \min_{1 \leq i \leq n} x_i \geq \theta_{10}, \\ \infty & \text{в противном случае,} \end{cases}$$

поэтому неравенство $l(x) \geq c$ эквивалентно неравенствам $\{x_{(1)} < \theta_{10}\} \cup \{x \leq c, x_{(1)} \geq \theta_{10}\}$. При гипотезе H_0 событие $\{x_{(1)} \geq \theta_{10}\}$ имеет вероятность 1, поэтому, определив c_α из условия $P_{\theta_0}(X \leq c_\alpha) = \alpha$, получим, что критерий Неймана — Пирсона имеет вид $\mathcal{R}_{1\alpha}^* = \{x: x_{(1)} < \theta_{10} \text{ либо } x \leq c_\alpha\}$ и не зависит от альтернативы θ_1 ; следовательно, он одновременно является р. н. м. критерием для всего класса альтернатив H_1 .

2. Проверка простой гипотезы против двусторонней альтернативы. р. н. м. несмещенные критерии. Как отмечалось выше, даже для экспоненциальных моделей с монотонным отношением правдоподобия р. н. м. критерия для двусторонней альтернативы $H_1 = H_1^- \cup H_1^+: \theta \neq \theta_0$ (θ — скалярный параметр), вообще говоря, не

существует. В этих случаях обычно поступают следующим образом: используют статистику $T(X)$, с помощью которой строятся р. н. м. критерии против односторонних альтернатив H_1^- и H_1^+ , и задают критическую область в виде $\mathcal{X}_{1\alpha} = \{T(x) \leq c_{\alpha_1}^-\} \cup \{T(x) \geq c_{\alpha_2}^+\}$, т. е. объединяют две соответствующие односторонние критические области с уровнями значимости α_1 и α_2 такими, что $\alpha_1 + \alpha_2 = \alpha$. Таким образом получают критерий с уровнем значимости α .

Рассмотрим применение этого правила на примере проверки гипотезы $H_0: \theta = \theta_0$ против альтернативы $H_1: \theta \neq \theta_0$ для нормальной модели $\mathcal{N}^*(\theta, \sigma^2)$. Как отмечено в начале п. 1, р. н. м. критерий уровня значимости α_1 при левосторонней альтернативе H_1^- задается критической областью

$$\mathcal{X}_{1\alpha_1}^- = \{x: (\sqrt{n}/\sigma)(x - \theta_0) \leq -t_{\alpha_1}\}, \quad \Phi(-t_{\alpha_1}) = \alpha_1,$$

а р. н. м. критерий уровня значимости α_2 при правосторонней альтернативе H_1^+ — критической областью

$$\mathcal{X}_{1\alpha_2}^+ = \{x: (\sqrt{n}/\sigma)(x - \theta_0) \geq t_{\alpha_2}\}, \quad \Phi(-t_{\alpha_2}) = \alpha_2.$$

Следовательно, в данной задаче можно использовать любой критерий вида

$$\mathcal{X}_{1\alpha} = \{x: (\sqrt{n}/\sigma)(x - \theta_0) \leq -t_{\alpha_1} \text{ либо } (\sqrt{n}/\sigma)(x - \theta_0) \geq t_{\alpha_2}\}, \quad \alpha_1 + \alpha_2 = \alpha. \quad (4.35)$$

Интуитивно предпочтителен выбор «симметричного» критерия, т. е. когда $\alpha_1 = \alpha_2 = \alpha/2$; в этом случае критерий имеет вид

$$\mathcal{X}_{1\alpha}^i = \{x: (\sqrt{n}/\sigma)|x - \theta_0| \geq t_{\alpha/2}\}, \quad \Phi(-t_{\alpha/2}) = \alpha/2. \quad (4.36)$$

Чтобы исследовать свойства этих критериев и выбрать среди них наилучший, рассмотрим функцию мощности $W(\mathcal{X}_{1\alpha}; \theta) = P_\theta(X \in \mathcal{X}_{1\alpha}^-) + P_\theta(X \in \mathcal{X}_{1\alpha}^+)$. Из формул (4.13) и (4.15) имеем

$$W(\mathcal{X}_{1\alpha}; \theta) = \Phi(\sqrt{n} \Delta/\sigma - t_{\alpha_1}) + \Phi(-\sqrt{n} \Delta/\sigma - t_{\alpha_2}) = \psi(\Delta), \\ \Delta = \theta - \theta_0.$$

Исследуем мощность как функцию Δ . Ее первые две производные соответственно равны

$$\psi'(\Delta) = \frac{1}{\sigma} \sqrt{\frac{n}{2\pi}} \left[\exp\left\{-\frac{1}{2}\left(\frac{\sqrt{n}}{\sigma}\Delta - t_{\alpha_1}\right)^2\right\} - \exp\left\{-\frac{1}{2}\left(\frac{\sqrt{n}}{\sigma}\Delta + t_{\alpha_2}\right)^2\right\} \right]$$

$$\text{и} \\ \psi''(\Delta) = \frac{n}{\sigma^2 \sqrt{2\pi}} \left[\left(t_{\alpha_1} - \frac{\sqrt{n}}{\sigma}\Delta\right) \exp\left\{-\frac{1}{2}\left(\frac{\sqrt{n}}{\sigma}\Delta - t_{\alpha_1}\right)^2\right\} + \right. \\ \left. + \left(\frac{\sqrt{n}}{\sigma}\Delta + t_{\alpha_2}\right) \exp\left\{-\frac{1}{2}\left(\frac{\sqrt{n}}{\sigma}\Delta + t_{\alpha_2}\right)^2\right\} \right].$$

Отсюда имеем, что $\psi'(\Delta) = 0$ только при $\Delta = \Delta_0 = \sigma(t_{\alpha_1} - t_{\alpha_2})/(\sqrt{n})$. В этом случае

$$\psi''(\Delta_0) = \frac{n}{\sigma^2 \sqrt{2\pi}} (t_{\alpha_1} + t_{\alpha_2}) \exp\left\{-\frac{1}{8}(t_{\alpha_1} + t_{\alpha_2})^2\right\}.$$

При малых α величины t_{α_1} и t_{α_2} положительны, поэтому $\psi^*(\Delta_0) > 0$, т. е. Δ_0 — точка минимума функции $\psi(\Delta)$. Поскольку $\Delta_0 = 0$ только при $\alpha_1 = \alpha_2 = \alpha/2$, а $\psi(0) = \Phi(-t_{\alpha_1}) + \Phi(-t_{\alpha_2}) = \alpha_2 + \alpha_1 = \alpha$, лишь для «симметричного» критерия (4.36) выполняется условие $W(\tilde{\mathcal{X}}_{1\alpha}; \theta) \geq \alpha$. В остальных случаях (т. е. при $\alpha_1 \neq \alpha_2$) критерий (4.35) при некоторых альтернативах имеет мощность, меньшую α , т. е. является смещенным критерием [см. условие (4.5)]. Таким образом, среди критериев вида (4.35) только «симметричный» критерий (4.36) — несмещенный и, следовательно, ему надо отдавать предпочтение.

Критерий (4.36) на самом деле обладает наибольшей мощностью среди всех несмещенных критериев уровня значимости α , т. е. является р. н. м. несмещенным критерием. Это утверждение есть следствие приведенной ниже теоремы об общем виде р. н. м. несмещенного критерия.

Пусть Θ — интервал действительной оси и θ_0 — внутренняя точка Θ . Пусть, далее, допустимые распределения абсолютно непрерывны и функция правдоподобия $L(x; \theta)$ имеет во всех внутренних точках интервала Θ производную $\frac{\partial L(x; \theta)}{\partial \theta} = L_1(x; \theta)$ такую, что $|L_1(x; \theta)| < M(x)$, где $M(x)$ — интегрируемая функция на $\mathcal{X}$ (тогда $\frac{\partial}{\partial \theta} \int_S L(x; \theta) dx = \int_S L_1(x; \theta) dx$ для любого подмножества $S \subseteq \mathcal{X}$ и любого θ). Определим, наконец, множество

$$\tilde{\mathcal{X}}_{1\alpha} = \{x: L(x; \theta_1) \geq cL(x; \theta_0) + c_1 L_1(x; \theta_0)\}, \quad (4.37)$$

где $\theta_1 \neq \theta_0$ и постоянные $c \geq 0$ и c_1 определяются (по предположению однозначности) из условий

$$P_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha}) = \int_{\tilde{\mathcal{X}}_{1\alpha}} L(x; \theta_0) dx = \alpha, \quad P'_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha}) = \int_{\tilde{\mathcal{X}}_{1\alpha}} L_1(x; \theta_0) dx = 0 \quad (4.38)$$

(для краткости здесь и далее пишем $P_\theta(S)$ вместо более строгой записи $P_\theta(X \in S)$, где $S \subset \mathcal{X}$).

Теорема 4.5. Если множество $\tilde{\mathcal{X}}_{1\alpha}$, определенное в (4.37), не зависит от θ_1 , то оно задает р. н. м. несмещенный критерий уровня значимости α для проверки простой гипотезы $H_0: \theta = \theta_0$ против двусторонней альтернативы $H_1: \theta \neq \theta_0$.

□ Рассмотрим любой несмещенный критерий $\mathcal{X}_{1\alpha}$ проверки гипотезы $\theta = \theta_0$. Так как $W(\mathcal{X}_{1\alpha}; \theta_0) = \alpha$ и производная функции мощности по θ существует, то

$$W'(\mathcal{X}_{1\alpha}; \theta_0) = 0. \quad (4.39)$$

Но

$$P_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \mathcal{X}_{1\alpha}) = \alpha - P_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \mathcal{X}_{1\alpha}) = P_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \mathcal{X}_{1\alpha}) \quad (4.40)$$

и в силу соотношений (4.39) и (4.38) имеем

$$P_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \mathcal{X}_{1\alpha}) + P'_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \mathcal{X}_{1\alpha}) = 0,$$

$$P_{\theta_0}(\mathcal{X}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \mathcal{X}_{1\alpha}) + P'_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \mathcal{X}_{1\alpha}) = 0,$$

т. е. выполняются равенства

$$P'_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \cdot \mathcal{X}_{1\alpha}) = -P'_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \cdot \mathcal{X}_{1\alpha}) = P'_{\theta_0}(\mathcal{X}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \cdot \mathcal{X}_{1\alpha}). \quad (4.41)$$

Из определения (4.37) множества $\tilde{\mathcal{X}}_{1\alpha}$ и равенств (4.40) и (4.41) получаем следующую цепочку соотношений: $P_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \cdot \mathcal{X}_{1\alpha}) \geq c P_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \cdot \mathcal{X}_{1\alpha}) + c_1 P_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha} \cdot \mathcal{X}_{1\alpha}) = c P_{\theta_0}(\mathcal{X}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \cdot \mathcal{X}_{1\alpha}) + c_1 P_{\theta_0}(\mathcal{X}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \cdot \mathcal{X}_{1\alpha}) \geq P_{\theta_0}(\mathcal{X}_{1\alpha} \setminus \tilde{\mathcal{X}}_{1\alpha} \cdot \mathcal{X}_{1\alpha})$, откуда следует, что $P_{\theta_0}(\tilde{\mathcal{X}}_{1\alpha}) \geq P_{\theta_0}(\mathcal{X}_{1\alpha})$. Тем самым показано, что критерий, опирающийся на критическое множество $\tilde{\mathcal{X}}_{1\alpha}$, является р. н. м. критерием проверки гипотезы $\theta = \theta_0$ против альтернативы θ_1 в классе всех несмещенных критериев $\tilde{\mathcal{X}}_{1\alpha}$. По условию, множество $\tilde{\mathcal{X}}_{1\alpha}$ не зависит от конкретной альтернативы θ_1 , поэтому этот критерий одновременно является р. н. м. несмещенным критерием против любой допустимой альтернативы $\theta \neq \theta_0$. ■

Замечание 2. Аналогичное утверждение справедливо также и для дискретных распределений с теми же оговорками, что и при построении критерия Неймана — Пирсона в § 4.2.

Применим доказанный результат к критерию (4.36). Здесь

$$L_1(x; \theta) = \frac{n}{\sigma^2} (x - \theta) L(x; \theta); \quad (4.42)$$

следовательно, неравенство в (4.37) эквивалентно неравенству

$$\frac{L(x; \theta_1)}{L(x; \theta_0)} \geq c + c_1 \frac{n}{\sigma^2} (x - \theta_0),$$

которое можно легко привести к неравенству

$$\exp \left\{ MT(x) - \frac{M^2}{2} \right\} \geq c + c'_1 T(x), \quad (4.43)$$

где $T(x) = \sqrt{n} (x - \theta_0)/\sigma$, $M = \sqrt{n} (\theta_1 - \theta_0)/\sigma$, $c'_1 = c_1 \sqrt{n}/\sigma$. Неравенство, противоположное (4.43), эквивалентно неравенствам

$$t^{(1)} \leq T(x) \leq t^{(2)}. \quad (4.44)$$

Статистика $T(X)$ при нулевой гипотезе распределена по закону $\mathcal{N}(0, 1)$, симметричному относительно нуля. Отсюда и из (4.42) следует, что для обеспечения второго из соотношений (4.38) необходимо, чтобы интервал (4.44) был симметричен относительно нуля. Это означает, что $t^{(2)} = -t^{(1)}$. Для обеспечения одновременного выполнения и первого из соотношений (4.38) надо положить $t^{(2)} = t_{\alpha/2}$, где $\Phi(-t_{\alpha/2}) = \alpha/2$. Таким образом, условия (4.38) выполняются тогда и только тогда, когда область (4.37) имеет вид

$$\tilde{\mathcal{X}}_{1\alpha} = \{x : |T(x)| \geq t_{\alpha/2}\}, \quad \Phi = (-t_{\alpha/2}, t_{\alpha/2}),$$

т. е. мы пришли к (4.36). Следовательно, по теореме 4.5 критерий (4.36) является р. н. м. несмещенным критерием проверки гипотезы $H_0 : \theta = \theta_0$ против двусторонней альтернативы $H_1 : \theta \neq \theta_0$. Функция мощности этого критерия

$$W(\tilde{\mathcal{X}}_{1\alpha}; \theta) = \Phi(\sqrt{n}(\theta - \theta_0)/\sigma - t_{\alpha/2}) + \Phi(\sqrt{n}(\theta_0 - \theta)/\sigma - t_{\alpha/2})$$

симметрична относительно точки $\theta = \theta_0$, в которой она имеет минимум, равный α . При $\theta \neq \theta_0$ мощность строго больше α и стремится к единице при $\theta \rightarrow \pm \infty$. В соответствии с теоремой 4.5 график этой функции лежит выше соответствующего графика любого другого несмещенного критерия в данной задаче.



Рис. 4.3

Сравним функцию $W(\tilde{x}_{1\alpha}^*; \theta)$ с функциями мощности соответствующих односторонних р. н. м. критериев $W(\tilde{x}_{1\alpha}^*; \theta)$ и $W(\tilde{x}_{1\alpha}^*; \theta)$ (рис. 4.3).

Функция $W(\tilde{x}_{1\alpha}^*; \theta)$ всегда заключена между этими двумя функциями (они изображены на рисунке пунктирными линиями с соответствующим знаком), за исключением точки $\theta = \theta_0$, в которой все три графика совпадают. Это означает, что критерий $\tilde{x}_{1\alpha}^*$ всегда менее мощный, чем один из односторонних критериев $\tilde{x}_{1\alpha}^*$ или $\tilde{x}_{1\alpha}^*$, но всегда более мощный, чем другой односторонний критерий.

Замечание 3. В приложениях теорему 4.5 чаще всего применяют для построения р. н. м. несмещенных критериев в случае экспоненциальных моделей с монотонным отношением правдоподобия [см. рассмотренную задачу для модели $\mathcal{N}(\theta, \sigma^2)$]. Изложение общей соответствующей теории можно найти, например, в [14, гл. 4].

3. Локальные наиболее мощные критерии. На практике особый интерес представляют малые отклонения от нулевой гипотезы $H_0: \theta = \theta_0$. В этом случае при исследовании свойств критерия можно ограничиться только анализом локального поведения функции мощности критерия $W(\theta)$ в окрестности точки θ_0 . При таком подходе часто удается построить локальный наиболее мощный (л. н. м.) критерий, даже тогда, когда р. н. м. критерия не существует.

Пусть θ — вещественный параметр и выполнены условия регулярности, приведенные перед формулировкой теоремы 4.5. Кроме того, будем предполагать, что для любой критической области $\tilde{x}_{1\alpha}^*$ функция мощности $W(\theta) = W(\tilde{x}_{1\alpha}^*; \theta)$ ($W(\theta_0) = \alpha$) допускает разложение в ряд Тейлора в окрестности точки θ_0 :

$$W(\theta) = \alpha + (\theta - \theta_0) W'(\theta_0) + \frac{(\theta - \theta_0)^2}{2} W''(\theta_0) + \dots \quad (4.45)$$

Пусть гипотеза $H_0: \theta = \theta_0$ проверяется против альтернативы $H_1: \theta > \theta_0$. Тогда для получения л. н. м. одностороннего критерия необходимо максимизировать величину $W'(\theta_0) = \int_{\tilde{x}_{1\alpha}^*} L_1(x; \theta_0) dx$ при ограничении $W(\theta_0) = \int_{\tilde{x}_{1\alpha}^*} L(x; \theta_0) dx = \alpha$.

Повторяя рассуждения, использованные при доказательстве теоремы 4.1, получаем, что оптимальная критическая область имеет вид

$$\tilde{x}_{1\alpha}^* = \{x: L_1(x; \theta_0)/L(x; \theta_0) \geq c_\alpha^+\}, \quad (4.46)$$

где константа c_α^+ определяется из условия $W(\theta_0) = \alpha$. Аналогично, критическая область

$$\tilde{x}_{1\alpha}^* = \{x: L_1(x; \theta_0)/L(x; \theta_0) \leq c_\alpha^-\} \quad (4.47)$$

задает л. н. м. критерий в случае левосторонней альтернативы $H_1: \theta < \theta_0$.

При двусторонней альтернативе $H_1: \theta \neq \theta_0$ необходимо ввести дополнительное условие несмещенности критерия: $W'(\theta_0) = 0$; тогда л. н. м. критерием

является критерий, максимизирующий $W'(\theta_0)$ — коэффициент при $(\theta - \theta_0)^2$ в разложении (4.45). Как и при доказательстве теоремы 4.5, можно показать, что в этом случае оптимальная критическая область имеет вид

$$\mathcal{X}_{1\alpha} = \{x: L_2(x; \theta_0) \geq c_1 L_1(x; \theta_0) + c L(x; \theta_0)\}. \quad (4.48)$$

где $L_2(x; \theta) = \frac{\partial}{\partial \theta} L_1(x; \theta)$ и константы c и c_1 определяются условиями $W(\theta_0) = \alpha$, $W'(\theta_0) = 0$.

Пример 4.9 (нормальная-1 модель, л. н. м. критерий для среднего). Рассмотрим задачу проверки гипотезы $H_0: \theta = \theta_0$ против альтернативы $H_1: \theta \neq \theta_0$ для модели $\mathcal{N}(\theta, \sigma^2)$. Здесь $L_2(x; \theta) = [(n^2/\sigma^4)(x - \theta)^2 - n/\sigma^2] L(x; \theta)$, поэтому

$$\frac{L_1(x; \theta_0)}{L(x; \theta_0)} = \frac{n}{\sigma^2} (x - \theta_0), \quad \frac{L_2(x; \theta_0)}{L(x; \theta_0)} = \frac{n^2}{\sigma^4} (x - \theta_0)^2 - \frac{n}{\sigma^2}.$$

Следовательно, область (4.48) задается условием $T^2(x) \geq k_1 T(x) + k_2$, $T(x) = \sqrt{n}(x - \theta_0)/\sigma$. Чтобы было выполнено условие несмещенности, эта область должна быть симметричной (поскольку распределение статистики $T(X)$ при гипотезе H_0 симметрично относительно нуля), что имеет место только при $k_1 = 0$. Таким образом, условие можно переписать в виде $|T(x)| \geq k$. Наконец, из условия $W(\theta_0) = \alpha$ находим $k = t_{\alpha/2}$, где $\Phi(-t_{\alpha/2}) = \alpha/2$. Окончательно получаем, что критическая область (4.48) имеет вид $\{x: |T(x)| \geq t_{\alpha/2}\}$, т. е. результат совпадает с полученным в п. 2. Это ожидаемый результат, так как л. н. м. несмещенный критерий должен совпадать с р. н. м. несмещенным критерием, когда последний существует.

Рассмотрим снова односторонние альтернативы и проанализируем более подробно условия в (4.46) — (4.47). На основании определения вклада выборки $U(X; \theta)$ [см. (2.15)] имеем

$$\frac{L_1(x; \theta_0)}{L(x; \theta_0)} = \frac{\partial}{\partial \theta} \ln L(x; \theta_0) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln l(x_i; \theta_0) = U(x; \theta_0).$$

Как показано в § 2.2, для регулярных моделей $E_{\theta_0} U(X; \theta_0) = 0$, $D_{\theta_0} U(X; \theta_0) = -n i(\theta_0)$, где $i(\theta)$ — функция информации Фишера. Если n велико, то, по центральной предельной теореме,

$$\mathcal{X}_{\theta_0}(U(X; \theta_0)/\sqrt{n i(\theta_0)}) \sim \mathcal{N}(0, 1). \quad (4.49)$$

Следовательно, при большом объеме выборки n при вычислении границ в (4.46) — (4.47) можно использовать нормальную аппроксимацию (4.49) и в (4.46) приближенно полагать $\bar{\alpha} \approx t_{\alpha} \sqrt{n i(\theta_0)}$ а в (4.47) считать $\bar{\alpha} \approx -t_{\alpha} \sqrt{n i(\theta_0)}$, где $\Phi(-t_{\alpha}) = \alpha$. Таким образом, получено общее решение задачи построения л. н. м. односторонних критериев для больших выборок.

Применяя описанный в п. 2 прием объединения двух односторонних критических областей можно построить и асимптотический двусторонний критерий

$$|U(x; \theta_0)| \geq t_{\alpha/2} \sqrt{n i(\theta_0)}, \quad \Phi(-t_{\alpha/2}) = \alpha/2 \quad (4.50)$$

проверки гипотезы $H_0: \theta = \theta_0$ против альтернативы $H_1: \theta \neq \theta_0$. Из свойства симметрии предельного распределения $U(X; \theta_0)/\sqrt{n}$ вытекает, что критерий (4.50) — несмещенный и не существует другого несмещенного критерия, который имел бы асимптотически большую мощность. Таким образом критерий (4.50) является асимптотически оптимальным критерием против локальных альтернатив. Построенный приближенный критерий (4.50) совпадает с полученным в примере 4.9 точным критерием для нормальной модели $\mathcal{N}(\theta, \sigma^2)$, основанным на статистике $T(x) = \sqrt{n}(x - \theta_0)/\sigma$, поскольку в этом случае $U(x; \theta_0) = \sqrt{n} T(x)/\sigma$, а $i(\theta) = 1/\sigma^2$ (см. табл. 2.1).

Пример 4.10 (модель Коши, л. н. м. критерий для параметра). Пусть требуется проверить гипотезу $H_0: \theta = \theta_0$ против альтернативы $H_1: \theta \neq \theta_0$ для модели

Коши $\mathcal{H}(\theta)$. Здесь функция вклада и функция информации соответственно равны (см. пример 2.23)

$$U(x; \theta) = 2 \sum_{i=1}^n \frac{x_i - \theta}{1 + (x_i - \theta)^2}, \quad i(\theta) = \frac{1}{2},$$

поэтому при больших n критическая область критерия задается условиями

$$\frac{2\sqrt{2}}{\sqrt{n}} \left| \sum_{i=1}^n \frac{x_i - \theta_0}{1 + (x_i - \theta_0)^2} \right| \geq t_{\alpha/2}, \quad \Phi(-t_{\alpha/2}) = \frac{\alpha}{2}.$$

4. Проверка гипотез и доверительное оценивание. Между задачей проверки простой гипотезы о параметре θ и построением доверительной области для этого параметра имеется тесная связь. Рассмотрим для каждого $\theta_0 \in \Theta$ какой-либо критерий $\mathcal{X}_{1\alpha} = \mathcal{X}_{1\alpha}(\theta_0)$ проверки гипотезы $H_0: \theta = \theta_0$. Пусть $\mathcal{X}_{0\alpha}(\theta_0) = \mathcal{X}_{1\alpha}(\theta_0)$ — область принятия H_0 . Тем самым в выборочном пространстве $\mathcal{X}$ задано семейство подмножеств $\{\mathcal{X}_{0\alpha}(\theta), \theta \in \Theta\}$. Определим при каждом $x \in \mathcal{X}$ подмножество $\mathfrak{S}(x) \in \Theta$, положив $\mathfrak{S}(x) = \{\theta: x \in \mathcal{X}_{0\alpha}(\theta)\}$. Таким образом, в параметрическом множестве Θ получаем семейство подмножеств $\{\mathfrak{S}(x), x \in \mathcal{X}\}$.

Рассмотрим случайное множество $\mathfrak{S}(X)$. События $\{\theta \in \mathfrak{S}(X)\}$ и $\{X \in \mathcal{X}_{0\alpha}(\theta)\}$ эквивалентны, так как по построению каждое из них влечет за собой другое, поэтому их вероятности при каждом θ совпадают. Но вероятность второго события равна по построению $1 - \alpha$; следовательно, $P_\theta(\theta \in \mathfrak{S}(X)) = P_\theta(X \in \mathcal{X}_{0\alpha}(\theta)) = 1 - \alpha$, т. е. $\mathfrak{S}(X)$ является $(1 - \alpha)$ -доверительной областью для θ [см. (2.88)]. Верно и обратное, т. е. если имеется семейство γ -доверительных множеств $\{\mathfrak{S}_\gamma(x), x \in \mathcal{X}\}$ для θ , то множество $\mathcal{X}_{0,1-\gamma} = \{x: \theta_0 \in \mathfrak{S}_\gamma(x)\}$ определяет область принятия гипотезы $H_0: \theta = \theta_0$ с уровнем значимости $\alpha = 1 - \gamma$.

Таким образом, задача построения доверительного множества для параметра θ и задача проверки простой гипотезы относительно θ взаимно обратные: если для некоторой модели известно решение одной из них, то по описанному алгоритму можно получить решение другой. При этом р. н. м. критерий (когда он существует) соответствует наикратчайшему доверительному множеству и наоборот.

Пример 4.11 (нормальная-1 модель, оптимальный доверительный интервал для среднего). Рассмотрим модель $\mu^*(\theta, \sigma^2)$. В п. 2 построен оптимальный (р. н. м. несмещенный) критерий проверки гипотезы $H_0: \theta = \theta_0$, который задается критической областью (4.36). Здесь $\mathcal{X}_{0\alpha}(\theta) = \{x: \sqrt{n} |x - \theta| / \sigma < t_{\alpha/2}\}$, $\Phi(-t_{\alpha/2}) = \alpha/2$, и множество $\mathfrak{S}(x)$ имеет вид

$$\mathfrak{S}(x) = \left\{ \theta: \frac{\sqrt{n}}{\sigma} |x - \theta| < t_{\alpha/2} \right\} = \left\{ \theta: x - \frac{\sigma}{\sqrt{n}} t_{\alpha/2} < \theta < x + \frac{\sigma}{\sqrt{n}} t_{\alpha/2} \right\}.$$

Следовательно, интервал $(X \pm (\sigma/\sqrt{n}) t_{\alpha/2})$ является наикратчайшим среди всех $(1 - \alpha)$ -доверительных интервалов для неизвестного

среднего закона $\varphi^*(\theta, \sigma^2)$ — это значительное усиление результата примера 2.28.

Далее, так как р. н. м. критерий против правосторонней альтернативы $H_1: \theta > \theta_0$ задается критической областью, определенной в (4.12), то $\mathcal{X}_{\alpha_2}(\theta) = \{x: \sqrt{n}(x - \theta)/\sigma < t_{\alpha}\}$, $\Phi(-t_{\alpha}) = \alpha$, и поэтому множество $\mathcal{S}(x) = \{\theta: \theta > x - \sigma t_{\alpha}/\sqrt{n}\}$. Отсюда имеем, что оптимальный односторонний (нижний) $(1 - \alpha)$ -доверительный интервал для θ имеет вид $(X - \sigma t_{\alpha}/\sqrt{n}, \infty)$.

Аналогично, используя критическую область (4.14), можно построить и верхний доверительный интервал.

Пример 4.12 (нормальная-2 модель, оптимальный доверительный интервал для дисперсии). В примере 2.29 построен доверительный интервал $\Delta_{\gamma}^*(X)$ для неизвестной дисперсии θ^2 в модели $\varphi^*(\mu, \theta^2)$. Используя этот результат и применяя описанную методику, получаем, что критерий уровня значимости $\alpha = 1 - \gamma$ для проверки гипотезы $H_0: \theta = \theta_0$ против альтернативы $H_1: \theta \neq \theta_0$ задается критической областью

$$\mathcal{X}_{1-\alpha} = \left\{ x: \frac{1}{\theta_0^2} \sum_{i=1}^n (x_i - \mu)^2 \leq \chi_{\alpha_1, n}^2 \right\} \cup \left\{ x: \frac{1}{\theta_0^2} \sum_{i=1}^n (x_i - \mu)^2 \geq \chi_{1-\alpha_2, n}^2 \right\}, \quad (4.51)$$

где $\alpha_1 + \alpha_2 = \alpha$ и границы определяются из уравнения (2.80'). Эта область представляет собой объединение двух р. н. м. односторонних критических областей, приведенных в примере 4.4.

Проверим, является ли этот критерий несмещенным. Пусть $F_n(t)$ — функция распределения закона $\chi^2(n)$, тогда, учитывая, что $\mathcal{X}_{\theta} \left(\frac{1}{\theta^2} \sum_{i=1}^n (X_i - \mu)^2 \right) = \chi^2(n)$ (см. пример 2.29), получаем

$$W(\theta) = W(\mathcal{X}_{1-\alpha}; \theta) = F_n(t\chi_{\alpha_1, n}^2) + 1 - F_n(t\chi_{1-\alpha_2, n}^2) = \psi(t), \quad t = \theta_0^2/\theta^2.$$

Отсюда

$$\psi'(t) = \chi_{\alpha_1, n}^2 k_n(t\chi_{\alpha_1, n}^2) - \chi_{1-\alpha_2, n}^2 k_n(t\chi_{1-\alpha_2, n}^2).$$

Из соотношения (2.80') следует, что

$$\psi'(t) = \frac{t^{n-1}}{2^{n/2} \Gamma(n/2)} \chi_{\alpha_1, n}^n e^{-\chi_{\alpha_1, n}^2 / 2} \left[e^{\frac{1-t}{2} \chi_{\alpha_1, n}^2} - e^{\frac{1-t}{2} \chi_{1-\alpha_2, n}^2} \right].$$

Так как $\chi_{\alpha_1, n}^2 < \chi_{1-\alpha_2, n}^2$, то

$$\psi'(t) \begin{cases} < 0 & \text{при } t < 1, \\ = 0 & \text{при } t = 1, \\ > 0 & \text{при } t > 1. \end{cases}$$

Это означает, что функция мощности $W(\theta)$ монотонно убывает при изменении θ от 0 до θ_0 и монотонно возрастает при изменении θ от θ_0 до ∞ , а $W(\theta_0) = \alpha$, т. е. критерий (4.51) не смещен.

Применяя теорему 4.5, можно убедиться, что построенный критерий является р. н. м. несмещенным критерием в задаче проверки гипотезы $H_0: \theta = \theta_0$ против двусторонней альтернативы $H_1: \theta \neq \theta_0$. Одновременно это означает, что интервал $\Delta_{\gamma}^*(X)$, определенный в (2.79'), — наикратчайший среди всех $(1 - \alpha)$ -доверительных интервалов для неизвестной дисперсии θ^2 в модели $\mathcal{N}^*(\mu, \theta^2)$.

Методику, устанавливающую связь между задачами доверительного оценивания и проверки гипотез, можно применять и в более общих ситуациях, когда требуется оценить не весь параметрический вектор θ (в случае многомерного параметра), а только его некоторый подвектор $\theta^{(1)}$ (так что $\theta = (\theta^{(1)}, \theta^{(2)})$) и о координатах, образующих подвектор $\theta^{(2)}$, говорят как о «мешающих» параметрах), а также когда соответствующие статистические выводы требуется сделать о заданной параметрической функции $\tau(\theta)$ (θ может быть как скаляром, так и вектором).

Пусть, например, имеется семейство γ -доверительных множеств $\mathcal{S}_{\gamma}(x)$ для $\theta^{(1)}$, т. е. семейство подмножеств параметрического множества $\Theta^{(1)} = \{\theta^{(1)}: (\theta^{(1)}, \theta^{(2)}) \in \Theta\}$, удовлетворяющих условию $P_{(\theta^{(1)}, \theta^{(2)})}(\theta^{(1)} \in \mathcal{S}_{\gamma}(X)) \geq \gamma, \forall (\theta^{(1)}, \theta^{(2)}) \in \Theta$. Определим для фиксированного $\theta_0^{(1)}$ подмножество $\mathcal{X}_{0, 1-\gamma} = \mathcal{X}_{0, 1-\gamma}(\theta_0^{(1)})$ в выборочном пространстве $\mathcal{X}$ с помощью правила $\mathcal{X}_{0, 1-\gamma} = \{x: \theta_0^{(1)} \in \mathcal{S}_{\gamma}(x)\}$. Тогда по построению

$$P_{(\theta_0^{(1)}, \theta^{(2)})}(X \in \mathcal{X}_{0, 1-\gamma}) = P_{(\theta_0^{(1)}, \theta^{(2)})}(\theta_0^{(1)} \in \mathcal{S}_{\gamma}(X)) \geq \gamma, \text{ или}$$

$$P_{(\theta_0^{(1)}, \theta^{(2)})}(X \in \mathcal{X}_{0, 1-\gamma}) \leq 1 - \gamma.$$

Следовательно, область $\mathcal{X}_{1, \alpha} = \mathcal{X}_{0, 1-\gamma}$, $\alpha = 1 - \gamma$, задает критерий уровня значимости α для проверки гипотезы $H_0: \theta^{(1)} = \theta_0^{(1)}$ (это гипотеза сложная, так как оставляет неопределенным значение $\theta^{(2)}$). Как и выше, можно провести обратные рассуждения: от задачи проверки гипотезы перейти к задаче доверительного оценивания.

Продемонстрируем эту методику на примерах проверки некоторых гипотез для общей нормальной модели.

Пример 4.13 (общая нормальная модель, проверка гипотез для среднего и дисперсии). В примере 2.30 построены доверительные интервалы для неизвестных среднего θ_1 и дисперсии θ_2^2 закона $\mathcal{N}^*(\theta_1, \theta_2^2)$. Используя эти результаты и применяя описанную методику, получаем, что критическое множество

$$\mathcal{X}_{1, \alpha} = \left\{ x: \sqrt{n-1} \frac{|\bar{x} - \theta_{10}|}{S(x)} \geq t_{1-\alpha/2, n-1} \right\} \quad (4.52)$$

задает критерий уровня значимости α для проверки гипотезы о среднем $H_0: \theta_1 = \theta_{10}$ (θ_2 произвольно) против альтернативы $H_1: \theta_1 \neq \theta_{10}$, а критерий проверки гипотезы о дисперсии $H_0: \theta_2 = \theta_{20}$ (θ_1 произвольно) против альтернативы $H_1: \theta_2 \neq \theta_{20}$ определяется

критической областью [ср. с (4.51)]

$$\mathcal{X}_{1\alpha} = \{x : nS^2(x) \leq \theta_{1\alpha}^2 \chi_{2n-1}^2\} \cup \{x : nS^2(x) \geq \theta_{1\alpha}^2 \chi_{1-\alpha, 2n-1}^2\} \quad (4.53)$$

Здесь приняты такие же обозначения, что и в (2.81').

Известно [14, с. 222—228], что критерии (4.52) и (4.53) являются оптимальными (р. н. м. несмещенными), поэтому и соответствующие доверительные интервалы являются наиболее точными.

Пример 4.14 (*гипотеза о равенстве средних двух нормальных моделей*). Пусть, как и в примере 2.31, X и Y — независимые выборки из нормальных совокупностей с одинаковыми дисперсиями. Используя результат этого примера, определяющий доверительный интервал для параметра $\Delta = \theta_{1'} - \theta_{1''}$ — разности неизвестных средних, получаем критерий проверки гипотезы $H_0: \Delta = 0$ (гипотезы о равенстве средних), который определяется критической областью

$$\mathcal{X}_{1\alpha} = \{(x, y) : |\bar{x} - \bar{y}| \geq t_{1-\alpha/2, n+m-2} \sqrt{\frac{(n+m)(nS^2(x) + mS^2(y))}{nm(n+m-2)}}\}.$$

Эта область задает р. н. м. несмещенный критерий [14, с. 234], тем самым и соответствующий доверительный интервал для Δ является наиболее точным.

§ 4.5. Критерий отношения правдоподобия

1. Метод отношения правдоподобия проверки общих гипотез. Одним из наиболее универсальных методов построения критериев проверки сложных гипотез является *метод отношения правдоподобия*, суть которого состоит в следующем. Для проверки гипотезы $H_0: \theta \in \Theta_0 \subset \Theta$ против альтернативы $H_1: \theta \in \Theta_1 = \Theta \setminus \Theta_0$ вводится *статистика отношения правдоподобия*

$$\Lambda_n^* = \Lambda_n^*(X; \Theta_0) = \sup_{\theta \in \Theta_0} L_n(X; \theta) / \sup_{\theta \in \Theta} L_n(X; \theta), \quad X = (X_1, \dots, X_n),$$

представляющая собой естественное обобщение на случай сложных гипотез статистики $l(X)$ критерия Неймана — Пирсона [см. (4.6)]; как и в критерии Неймана — Пирсона, здесь в критическую область включаются большие значения этой статистики. Предпочтительнее использовать эквивалентную статистику

$$\Lambda_n = \Lambda_n(X; \Theta_0) = \sup_{\theta \in \Theta} L_n(X; \theta) / \sup_{\theta \in \Theta_0} L_n(X; \theta),$$

которая связана с Λ_n^* соотношением $\Lambda_n = \max(1, \Lambda_n^*)$. В литературе, однако, чаще рассматривают обратную статистику

$$\lambda_n = \lambda_n(X; \Theta_0) = 1/\Lambda_n(X; \Theta_0) \quad (4.54)$$

и критическую область задают в виде

$$\mathcal{X}_1 = \{x : \lambda_n(x; \Theta_0) \leq c\}. \quad (4.55)$$

Критическую границу c в (4.55) следует выбрать так, чтобы критерий имел заданный уровень значимости α :

$$P_1(X \in \mathcal{X}_1) = \int_{\mathcal{X}_1} L_n(x; \theta) dx = P_1(\lambda_n(X; \theta_0) \leq c) \leq \alpha, \quad (4.56)$$

$$\forall \theta \in \theta_0.$$

Соотношения (4.54) — (4.56) определяют критерий отношения правдоподобия (к. о. п.) для проверки гипотезы $H_0: \theta \in \theta_0$.

Как было показано в § 4.2, метод отношения правдоподобия для простых гипотез приводит к оптимальному критерию. В общем случае для сложных гипотез это свойство оптимальности не выполняется. Тем не менее этот метод широко применяют, получая удовлетворительные решения во многих практических задачах. Кроме того, как будет показано ниже, при некоторых условиях к. о. п. обладает свойством асимптотической оптимальности для больших выборок.

Пример 4.15 (общая нормальная модель, к. о. п. для среднего)
Продemonстрируем изложенный метод на уже рассмотренном примере 4.13 — задаче проверки гипотезы $H_0: \theta_1 = \theta_{10}$ в модели $N(\theta_1, \theta_2^2)$. Здесь $\theta = \{\theta = (\theta_1, \theta_2): -\infty < \theta_1 < \infty, \theta_2 > 0\}$ — плоскость, $\theta_0 = \{\theta = (\theta_1, \theta_2): \theta_1 = \theta_{10}, \theta_2 > 0\}$ — полупрямая. Учитывая результаты примера 2.17, получаем

$$\sup_{\theta \in \theta_0} L_n(x; \theta) = L_n(x; \hat{\theta} = (\bar{x}, s)) = (2\pi es^2)^{-n/2}.$$

Далее имеем

$$\sup_{\theta \in \theta} L_n(x; \theta) = L_n(x; (\theta_{10}, s_0)) = (2\pi es_0^2)^{-n/2},$$

где $s_0^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \theta_{10})^2$ — оценка максимального правдоподобия для дисперсии θ_2^2 при гипотезе H_0 . Так как $s_0^2 = s^2 + (\bar{x} - \theta_{10})^2$, то отсюда находим $\lambda_n = (s_0^2/s^2)^{-n/2} = (1 + t^2/(n-1))^{-n/2}$, где $t = t(x) = \sqrt{n-1}(\bar{x} - \theta_{10})/s$. Таким образом, между значениями λ_n и t^2 существует взаимно однозначное соответствие. Отсюда, в частности следует, что неравенство $\lambda_n \leq c$ эквивалентно неравенству $|t| \geq c'$. Но, по теореме 1.11, $\mathcal{L}(t(X) | H_0) = S(n-1)$, т. е. статистика $t(X)$ при гипотезе H_0 имеет распределение, не зависящее от «мешающего» параметра θ_2 , а именно распределение Стьюдента с $n-1$ степенями свободы. Таким образом, можно рассчитать границу c' при заданном уровне значимости α и получить следующий критерий, эквивалентный к. о. п.:

$$\mathcal{X}_{1\alpha} = \{x: |t(x)| \geq t_{1-\alpha/2, n-1}\}$$

[здесь $t_{p, n-1}$ — p -квантиль распределения $S(n-1)$], т. е. приходим к критерию (4.52). Таким образом, в данном случае метод отношения правдоподобия приводит к оптимальному критерию.

В рассмотренном примере найдено точное распределение статистики λ_n (точнее, некоторой взаимно однозначной функции от

нее) при нулевой гипотезе, что и позволило рассчитать соответствующий критерий. Такое обстоятельство характерно для нормальной модели. Для других моделей точное распределение λ_n удается найти не всегда, поэтому для конечных выборок метод отношения правдоподобия не всегда позволяет получить решение задачи. В то же время для выборок большого объема при достаточно общих условиях удается получить простое асимптотическое распределение статистики $-2 \ln \lambda_n$, которое можно использовать для построения приближенной критической области. Таким образом, в большинстве случаев можно рассчитать простой асимптотический вариант критерия отношения правдоподобия, основанный на критической области вида $\{x: -2 \ln \lambda_n(x; \theta_0) \geq c\}$. Далее излагаются основные элементы асимптотической теории критерия отношения правдоподобия.

2. К. о. п. для больших выборок. Будем далее предполагать, что выполняются условия регулярности, обеспечивающие существование, единственность и асимптотическую нормальность оценки максимального правдоподобия $\hat{\theta}_n = (\hat{\theta}_{1n}, \dots, \hat{\theta}_{rn})$ параметра $\theta = (\theta_1, \dots, \theta_r)$ (см. п. 4 § 2.4). Рассмотрим случай простой нулевой гипотезы.

Теорема 4.6. Пусть требуется проверить простую гипотезу $H_0: \theta = \theta_0$, где $\theta_0 = (\theta_{10}, \dots, \theta_{r0})$ — фиксированная внутренняя точка множества Θ . Тогда для больших выборок ($n \rightarrow \infty$) при выполнении указанных условий регулярности к. о. п. задается асимптотически критической областью

$$\mathcal{X}_{1\alpha} = \{x: -2 \ln \lambda_n(x; \theta_0) \geq \chi_{1-\alpha, r}^2\}, \quad (4.57)$$

т. е. при $n \rightarrow \infty$

$$P_{\theta_0}(X \in \mathcal{X}_{1\alpha}) = P_{\theta_0}(-2 \ln \lambda_n(X; \theta_0) \geq \chi_{1-\alpha, r}^2) \rightarrow \alpha.$$

□ Покажем, что в условиях теоремы

$$\mathcal{L}_{\theta_0}(-2 \ln \lambda_n(X; \theta_0)) \rightarrow \chi^2(r), \quad (4.58)$$

отсюда следует (4.57). Если справедлива гипотеза H_0 , то в силу состоятельности о. м. п. при больших n точка $\hat{\theta}_n$ близка к θ_0 , поэтому для $\ln L_n(\theta_0) = \ln L_n(X; \theta_0)$ можно записать разложение Тейлора относительно точки $\hat{\theta}_n$:

$$\ln L_n(\theta_0) = \ln L_n(\hat{\theta}_n) + \frac{1}{2} \sum_{i,j=1}^r \frac{\partial^2 \ln L_n(\theta^*)}{\partial \theta_i \partial \theta_j} (\hat{\theta}_{in} - \theta_{i0}) (\hat{\theta}_{jn} - \theta_{j0}),$$

где $|\theta^* - \theta_0| < |\hat{\theta}_n - \theta_0|$. Отсюда следует, что

$$\begin{aligned} -2 \ln \lambda_n &= 2[\ln L_n(\hat{\theta}_n) - \ln L_n(\theta_0)] = \\ &= \sum_{i,j=1}^r -\frac{1}{n} \frac{\partial^2 \ln L_n(\theta^*)}{\partial \theta_i \partial \theta_j} \sqrt{n} (\hat{\theta}_{in} - \theta_{i0}) \sqrt{n} (\hat{\theta}_{jn} - \theta_{j0}). \end{aligned} \quad (4.59)$$

Так как θ^* — состоятельная оценка для θ_0 , а вторые производные функции правдоподобия, по предположению, непрерывны по θ ,

то, по теореме 1.5,

$$\frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta_i \partial \theta_j} \xrightarrow{P_{\theta_0}} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta_i \partial \theta_j}.$$

На основании закона больших чисел при $n \rightarrow \infty$ величина

$$\frac{1}{n} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta_i \partial \theta_j} = \frac{1}{n} \sum_{k=1}^n \frac{\partial^2 \ln f(X_k; \theta_0)}{\partial \theta_i \partial \theta_j}$$

сходится по вероятности (по распределению P_{θ_0}) к среднему $E_{\theta_0} \left(\frac{\partial^2 \ln f(X_1; \theta_0)}{\partial \theta_i \partial \theta_j} \right)$. Таким образом, матрица предельных значений коэффициентов квадратичной формы в (4.59) совпадает с информационной матрицей $I(\theta_0)$ [см. (2.34)]. Далее, из соотношения (2.58) следует, что случайный вектор $\eta_n = \sqrt{n}(\bar{\theta}_n - \theta_0)$ имеет в пределе такое же распределение, что и нормальный $\mathcal{N}(0, I^{-1}(\theta_0))$ случайный вектор η . Таким образом, правая часть (4.59) имеет в пределе такое же распределение, как и квадратичная форма $Q = \eta' I(\theta_0) \eta$. Но, по теореме 1.9, $\mathcal{L}(Q) = \chi^2(r)$, откуда и следует соотношение (4.58). ■

З а м е ч а н и е 1. Предельные распределения статистик $-2 \ln \lambda_n$ и $Q_n^{(1)} = \eta_n' I(\bar{\theta}_n) \eta_n$ при нулевой гипотезе совпадают, поэтому к. о. п. асимптотически эквивалентен критерию вида

$$\mathcal{Q}_{1\alpha}^{(1)} = \{Q_n^{(1)} \geq \chi_{1-\alpha}^2, r\}. \quad (4.60)$$

Если здесь матрицу $I(\bar{\theta}_n)$ заменить на $I(\theta_0)$, то получим еще один критерий, асимптотически эквивалентный к. о. п.

З а м е ч а н и е 2. Из доказательства асимптотической нормальности о. м. п. следует [см. соотношение (2.62)], что в случае одномерного параметра предельные распределения η_n и $U_n(\theta_0)/[\sqrt{n} i(\theta_0)]$ совпадают. В случае векторного параметра этот вывод остается в силе, если $U_n(\theta_0)/[\sqrt{n} i(\theta_0)]$ заменить на $[I^{-1}(\theta_0) U_n(\theta_0)]/\sqrt{n}$, где $U_n(\theta) = (U_{1n}(\theta), \dots, U_{rn}(\theta))$ и $U_{in}(\theta) = \frac{\partial \ln L_n(\theta)}{\partial \theta_i}$, $i = 1, \dots, r$. Отсюда и из предыдущего замечания следует, что статистики $Q_n^{(1)}$ и $Q_n^{(2)} = (1/n) U_n'(\theta_0) I^{-1}(\theta_0) U_n(\theta_0)$ имеют при гипотезе H_0 одно и то же предельное распределение, поэтому к. о. п. асимптотически эквивалентен также критерию вида

$$\mathcal{Q}_{1\alpha}^{(2)} = \{Q_n^{(2)} \geq \chi_{1-\alpha}^2, r\}. \quad (4.61)$$

Достоинство этого варианта критерия заключается в том, что при его использовании не надо вычислять о. м. п. $\bar{\theta}_n$ и его удобно использовать в тех случаях, когда явный вид $\bar{\theta}_n$ получить невозможно.

Итак, в задаче проверки простой гипотезы $H_0: \theta = \theta_0$ против сложной альтернативы $H_1: \theta \in \Theta \setminus \{\theta_0\}$ все три критерия (4.57), (4.60) и (4.61) асимптотически эквивалентны, т. е. уровень значимости каждого из них при $n \rightarrow \infty$ стремится к α .

Рассмотрим важный пример применения изложенных результатов к полиномиальному распределению $M(n, p = (p_1, \dots, p_N))$.

Пример 4.16 (метод отношения правдоподобия для полиномиального распределения). Пусть производятся независимые испытания,

в каждом из которых реализуется один из N исходов $A_1, \dots, A_N$, т. е. наблюдается случайная величина ξ , принимающая значения $1, \dots, N$ ($\xi = j$, если реализуется исход A_j). Обозначим через $\mathbf{p} = (p_1, \dots, p_N)$ вектор вероятностей этих исходов ($p_1 + \dots + p_N = 1$) и через $\mathbf{v} = (v_1, \dots, v_N)$ вектор частот реализаций соответствующих исходов в n испытаниях ($v_1 + \dots + v_N = n$). Как известно (см. п. 3 § 2.5), распределение вектора $\mathbf{v}$ и является полиномиальным распределением $M(n, \mathbf{p})$. Предположим теперь, что вероятности исходов $p_1, \dots, p_N$ неизвестны и требуется проверить гипотезу $H_0: \mathbf{p} = \mathbf{p}^0$, где $\mathbf{p}^0 = (p_1^0, \dots, p_N^0)$ — заданный вектор, удовлетворяющий условиям $0 < p_j^0 < 1$, $j = 1, \dots, N$, $p_1^0 + \dots + p_N^0 = 1$. Альтернативная гипотеза имеет вид $H_1: \mathbf{p} \neq \mathbf{p}^0$.

Здесь роль параметра $\theta = (\theta_1, \dots, \theta_r)$ играет вектор $\mathbf{p}$, но так как на значения параметров наложено ограничение $p_1 + \dots + p_N = 1$, то желательно это ограничение устранить, исключив, например, $p_N = 1 - p_1 - \dots - p_{N-1}$. Таким образом, далее полагаем $\theta = (p_1, \dots, p_{N-1})$, так что $r = \dim \theta = N - 1$.

Оценками максимального правдоподобия для параметров p_i являются относительные частоты реализаций соответствующих исходов, т. е. $\hat{p}_j = v_j/n$, $j = 1, \dots, N$, поэтому в данном случае статистика отношения правдоподобия имеет вид

$$\lambda_n(\mathbf{X}; \mathbf{p}^0) = \prod_{j=1}^N p_j^{v_j} / \prod_{j=1}^N \left(\frac{v_j}{n}\right)^{v_j} = \prod_{j=1}^N \left(\frac{np_j^0}{v_j}\right)^{v_j}.$$

Отсюда

$$-2 \ln \lambda_n(\mathbf{X}; \mathbf{p}^0) = 2 \sum_{j=1}^N v_j \ln \frac{v_j}{np_j^0}.$$

Вычислим информационную матрицу $\mathbf{I}(\theta_0)$. В рассматриваемом случае вероятности $f(x; \theta) = \mathbf{P}_\theta(\xi = x) = p_x$, $x = 1, 2, \dots, N$, удобно записать в виде

$$f(x; \theta) = \prod_{i=1}^N p_i^{\delta_{ix}} = \prod_{i=1}^{N-1} p_i^{\delta_{ix}} (1 - p_1 - \dots - p_{N-1})^{\delta_{Nx}},$$

где δ_{ij} — символ Кронекера ($\delta_{ij} = 1$ при $i = j$ и $\delta_{ij} = 0$ при $i \neq j$). Отсюда

$$\frac{\partial^2 \ln f(x; \theta)}{\partial p_i \partial p_l} = \begin{cases} \delta_{ix}/p_i + \delta_{Nx}/p_N & \text{при } i = l, \\ \delta_{Nx}/p_N^2 & \text{при } i \neq l. \end{cases}$$

Так как $E_\theta \delta_{ij} = \mathbf{P}_\theta(\xi = j) = p_j$, $j = 1, \dots, N$, то

$$-E_\theta \left(\frac{\partial^2 \ln f(X_1; \theta_0)}{\partial p_i \partial p_l} \right) = \begin{cases} 1/p_i + 1/p_N & \text{при } i = l, \\ 1/p_N & \text{при } i \neq l. \end{cases}$$

Используя обозначения, введенные при доказательстве теоремы 3.1, имеем, что $\mathbf{I}(\theta_0)$ совпадает с матрицей $\mathbf{A}$, а обратная матрица $\mathbf{I}^{-1}(\theta_0)$ — с матрицей $\Sigma(N-1)$. Заметим также, что вектор $\eta_n =$

$= \sqrt{n}(\hat{\theta}_n - \theta_0)$ в данном случае совпадает с вектором $\mathbf{v}^*(N-1)$, поэтому квадратичная форма $Q_n = \eta_n' I(\theta_0) \eta_n = \mathbf{v}^{*'}(N-1) \mathbf{A} \mathbf{v}^*(N-1)$ совпадает со статистикой X_n^2 , введенной в (3.5). Далее, для квадратичной формы Q_n^{**} в (4.60) имеем представление

$$Q_n^{**} = n \left[\sum_{j=1}^{N-1} \left(\frac{n}{v_j} + \frac{n}{v_N} \right) \left(\frac{v_j}{n} - p_j^0 \right)^2 + \sum_{\substack{j=1 \\ j \neq l}}^{N-1} \frac{n}{v_N} \left(\frac{v_l}{n} - p_l^0 \right) \left(\frac{v_j}{n} - p_j^0 \right) \right] = \sum_{j=1}^N \frac{(v_j - np_j^0)^2}{v_j}.$$

Наконец,

$$U_{jN}(\theta) = \sum_{i=1}^n \frac{\partial \ln f(X_{ji}, \theta)}{\partial p_j} = \sum_{i=1}^n \left(\frac{\delta_{ij} X_i}{p_j} - \frac{\delta_{Nj} X_i}{p_N} \right) = \frac{v_j}{p_j} - \frac{v_N}{p_N}$$

и квадратичная форма в (4.61) также совпадает со статистикой X_n^2 . Таким образом, в рассматриваемой задаче критерий можно строить, используя любую из статистик

$$-2 \ln \lambda_n(X; p^0) = 2 \sum_{j=1}^N v_j \ln \frac{v_j}{np_j^0},$$

$$Q_n^{**} = \sum_{j=1}^N \frac{(v_j - np_j^0)^2}{v_j}, \quad Q_n = Q_n^{**} = X_n^2 = \sum_{j=1}^N \frac{(v_j - np_j^0)^2}{np_j^0}.$$

Если справедлива гипотеза $H_0: \mathbf{p} = \mathbf{p}^0$, то в пределе при $n \rightarrow \infty$ все эти статистики имеют одно и то же распределение $\chi^2(N-1)$, поэтому при заданном уровне значимости α критическую границу выбирают равной $\chi_{1-\alpha, N-1}^2$. Последнюю статистику называют статистикой хи-квадрат Пирсона. В гл. 3 она была введена из других соображений. В § 3.2 было получено то же самое предельное распределение статистики X_n^2 при нулевой гипотезе и построен асимптотический вариант критерия согласия χ^2 . Здесь эти же результаты получены как простое следствие метода отношения правдоподобия применительно к полиномиальному распределению. Таким образом, для полиномиального распределения метод отношения правдоподобия и метод χ^2 асимптотически эквивалентны. С точки зрения простоты вычислений статистика X_n^2 предпочтительнее статистики $-2 \ln \lambda_n$, но по сравнению со статистикой Q_n^{**} у нее преимуществ нет.

3. Асимптотические свойства к. о. п. Рассмотрим асимптотические свойства критерия (4.57) [а также эквивалентных ему критериев (4.60) и (4.61)] при альтернативах $\theta \neq \theta_0$ и покажем, что критерий отношения правдоподобия *состоятелен*, т. е. что его мощность $W_n(\theta) = W_n(\mathcal{A}_{1\alpha}; \theta)$ удовлетворяет предельному соотношению $\lim_{n \rightarrow \infty} W_n(\theta) = 1, \forall \theta \neq \theta_0$. Для простоты рассмотрим случай скалярного параметра ($r = 1$).

Пусть $\theta_1 \neq \theta_0$ — произвольная фиксированная альтернатива, являющаяся внутренней точкой Θ . Запишем статистику $\lambda_n(X; \theta_0) = \lambda_n(\theta_0)$ в виде

$$\lambda_n(\theta_0) = \frac{L_n(\theta_1)}{L_n(\theta_0)} \frac{L_n(\theta_0)}{L_n(\theta_1)} = \lambda_n(\theta_1) v_n(\theta_0, \theta_1).$$

Отсюда

$$-2 \ln \lambda_n(\theta_0) = -2 \ln \lambda_n(\theta_1) + n v_n(\theta_0, \theta_1), \quad (4.62)$$

где

$$\begin{aligned} v_n(\theta_0, \theta_1) &= -\frac{2}{n} \ln v_n(\theta_0, \theta_1) = \frac{2}{n} [\ln L_n(\theta_1) - \ln L_n(\theta_0)] = \\ &= 2 \left[\frac{1}{n} \sum_{i=1}^n \ln f(X_i; \theta_1) - \frac{1}{n} \sum_{i=1}^n \ln f(X_i; \theta_0) \right]. \end{aligned}$$

Введем функцию $H(\theta) = E_{\theta_1}(\ln f(X_1; \theta))$; тогда на основании закона больших чисел, если истинной параметрической точкой является θ_1 , то при $n \rightarrow \infty$

$$\frac{1}{n} \sum_{i=1}^n \ln f(X_i; \theta_k) \xrightarrow{P_{\theta_1}} H(\theta_k), \quad k=0, 1.$$

Далее имеем

$$H^{(j)}(\theta) = E_{\theta_1} \left(\frac{\partial^j \ln f(X_1; \theta)}{\partial \theta^j} \right), \quad j=1, 2.$$

Отсюда [см. (2.16) и (2.19)] $H'(\theta_1) = 0$, $H''(\theta_1) = -i(\theta_1) < 0$. Таким образом, в точке θ_1 функция $H(\theta)$ имеет максимум, поэтому случайная величина $v_n(\theta_0, \theta_1)$ сходится по вероятности к положительному числу $2[H(\theta_1) - H(\theta_0)]$. По теореме 4.6, $\mathcal{L}_{\theta_1}(-2 \ln \lambda_n(\theta_1)) \rightarrow \chi_{(1)}^2$; следовательно, из (4.62) имеем, что при альтернативе θ_1 случайная величина $-2 \ln \lambda_n(\theta_0)$ по вероятности неограниченно возрастает при $n \rightarrow \infty$. Отсюда следует, что при $n \rightarrow \infty$

$$W_n(\theta_1) = P_{\theta_1}(-2 \ln \lambda_n(\theta_0) \geq \chi_{(1)}^2 - \alpha, 1) \rightarrow 1.$$

Таким образом, для к. о. п. последовательность функций мощности $W_n(\theta)$, $n=1, 2, \dots$, сходится к функции $\zeta(\theta) = \begin{cases} \alpha & \text{при } \theta = \theta_0, \\ 1 & \text{при } \theta \neq \theta_0. \end{cases}$ Это означает, в частности, что критерий отношения правдоподобия асимптотически не смещен.

Итак, если альтернатива θ_1 фиксирована, а объем выборки $n \rightarrow \infty$, то мощность к. о. п. стремится к единице. Однако если с изменением n изменяется и альтернатива ($\theta_1 = \theta_1^{(n)}$), «приближаясь» к нулевой гипотезе ($\theta_1^{(n)} \rightarrow \theta_0$ при $n \rightarrow \infty$), то мощность $W_n(\theta_1^{(n)})$, вообще говоря, уже не обязательно стремится к единице. Характер асимптотического поведения мощности при таких «близких» альтернативах зависит от скорости сближения альтернативы с проверяемой гипотезой, и для ее вычисления надо исследовать

предельное распределение статистики $-2 \ln \lambda_n(\mathbf{X}; \theta_0)$ не только при гипотезе H_0 (что было сделано в теореме 4.6), но и при альтернативах.

Теорема 4.7. Пусть близкая альтернатива $\theta_1^{(n)}$ имеет вид $\theta_1^{(n)} = \theta_0 + \beta/\sqrt{n}$, где $\beta = (\beta_1, \dots, \beta_r)$ — фиксированный ненулевой вектор. Тогда при выполнении условий регулярности теоремы 4.6 и $n \rightarrow \infty$

$$\mathcal{L}_{\theta_1^{(n)}}(-2 \ln \lambda_n(\mathbf{X}; \theta_0)) \rightarrow \chi^2(r; \lambda^2), \quad (4.63)$$

т. е. (см. п. 1 § 1.5) предельное распределение статистики $-2 \ln \lambda_n(\mathbf{X}; \theta_0)$ при рассматриваемой альтернативе является нецентральным распределением χ^2 с числом степеней свободы $r = \dim \Theta$ и параметром нецентральности λ^2 , вычисляемым по формуле

$$\lambda^2 = \beta' I(\theta_0) \beta. \quad (4.64)$$

Приведем идею доказательства этого утверждения, ограничившись случаем скалярного параметра (для многомерного параметра рассуждения обобщаются непосредственно). Имеем

$$-2 \ln \lambda_n(\theta_0) = 2 [\ln L_n(\theta_0^*) - \ln L_n(\theta_1^{(n)})] + 2 [\ln L_n(\theta_1^{(n)}) - \ln L_n(\theta_0)]. \quad (4.65)$$

Пусть справедлива альтернатива, т. е. $\theta_1^{(n)}$ — истинная параметрическая точка. Тогда из доказательства теоремы 4.6 имеем, что первое слагаемое в (4.65) имеет в пределе такое же распределение, что и случайная величина $\zeta^2 = i(\theta_0) \eta^2$, где $\mathcal{L}(\eta) = \mathcal{N}(0, I^{-1}(\theta_0))$; следовательно, $\mathcal{L}(\zeta) = \mathcal{N}(0, 1)$. Далее, по формуле Тейлора второе слагаемое в (4.65) равно

$$2 \sqrt{n}(\theta_1^{(n)} - \theta_0) \frac{1}{\sqrt{n}} U_n(\theta_1^{(n)}) - n(\theta_1^{(n)} - \theta_0)^2 \frac{1}{n} U_n'(\theta^*), \quad (4.66)$$

где $|\theta^* - \theta_0| < |\theta_1^{(n)} - \theta_0|$. Здесь при больших n на основании закона больших чисел случайную величину $(1/n) U_n'(\theta^*)$ можно заменить константой $-i(\theta_0)$, а случайная величина $(1/\sqrt{n}) U_n(\theta_1^{(n)})$ на основании центральной предельной теоремы асимптотически нормальна $\mathcal{N}(0, i(\theta_0))$. Поэтому все выражение (4.66) неограниченно возрастает, если $\sqrt{n}(\theta_1^{(n)} - \theta_0) \rightarrow \infty$; сходится к нулю, если $\sqrt{n}(\theta_1^{(n)} - \theta_0) \rightarrow 0$, и сходится к $2\beta \sqrt{i(\theta_0)} \zeta + \beta^2 i(\theta_0)$, если $\sqrt{n}(\theta_1^{(n)} - \theta_0) \rightarrow \beta$. Окончательно можно записать

$$\mathcal{L}_{\theta_1^{(n)}}(-2 \ln \lambda_n(\mathbf{X}; \theta_0)) \rightarrow \mathcal{L}((\zeta + \beta \sqrt{i(\theta_0)})^2) = \chi^2(1; \beta^2 i(\theta_0)),$$

т. е. получены утверждения (4.63) — (4.64) для $r = 1$.

Итак, для близких альтернатив вида $\theta_1^{(n)} = \theta_0 + \beta/\sqrt{n}$ мощность к. о. п. удовлетворяют при $n \rightarrow \infty$ соотношению

$$W_n(\theta_1^{(n)}) \rightarrow 1 - F_r(\chi^2_{1-\alpha, r}; \lambda^2), \quad (4.67)$$

где $F_r(t; \lambda^2)$ — функция распределения закона $\chi^2(r; \lambda^2)$. Для более близких альтернатив (случай $\lambda^2 \rightarrow 0$ или $\sqrt{n}(\theta_1^{(n)} - \theta_0) \rightarrow 0$) мощность стремится к уровню значимости α , т. е. такие альтернативы по к. о. п. асимптотически от нулевой гипотезы не отли-

чаются. Более далекие альтернативы (случай $\lambda^2 \rightarrow \infty$ или $\sqrt{n}(\theta_1^{(n)} - \theta_0) \rightarrow \infty$) критерий улавливает с вероятностью, стремящейся к 1 при $n \rightarrow \infty$, когда они верны, поскольку для таких альтернатив мощность стремится к единице. Эти же выводы справедливы и для критериев (4.60) и (4.61).

Пример 4.17 (полиномиальное распределение). Для локальных альтернатив $p_i^{(n)} = p_i^0 + \beta_j / \sqrt{n}$, $j = 1, \dots, N$, $\sum_{j=1}^N \beta_j = 0$, параметр нецентральности λ^2 равен (см. пример 4.16)

$$\begin{aligned} \lambda^2 &= \sum_{j=1}^{N-1} \beta_j^2 \left(\frac{1}{p_j^0} + \frac{1}{p_N^0} \right) + \frac{1}{p_N^0} \sum_{\substack{j=1 \\ j \neq i}}^{N-1} \beta_i \beta_j = \\ &= \sum_{j=1}^{N-1} \frac{\beta_j^2}{p_j^0} + \frac{1}{p_N^0} \left(\sum_{j=1}^{N-1} \beta_j \right)^2 = \sum_{j=1}^N \frac{\beta_j^2}{p_j^0}. \end{aligned}$$

Следовательно, мощность любого из полученных в примере 4.16 критериев при такой альтернативе равна в пределе (при $n \rightarrow \infty$)

$$1 - F_{N-1} \left(\chi^2_{1-\alpha, N-1}; \sum_{j=1}^N \beta_j^2 / p_j^0 \right).$$

4. Асимптотические свойства к. о. п. (сложная нулевая гипотеза). Ранее рассматривалась задача проверки простой гипотезы. Но оказывается, что все важные асимптотические свойства критерия отношения правдоподобия сохраняются и для случая сложной нулевой гипотезы. Приведем формулировки соответствующих утверждений (доказательства можно найти в [26] и [27]).

Пусть требуется проверить сложную гипотезу $H_0: \theta \in \Theta_0 \subset \Theta$, т. е. подмножество Θ_0 не сводится к одной точке. Альтернативная гипотеза также сложная и имеет вид $H_1: \theta \in \Theta \setminus \Theta_0$. Рассмотрим типичную задачу, соответствующую случаю, когда Θ_0 — евклидово подпространство размерности меньшей, чем размерность всего параметрического множества Θ (напомним, что Θ — либо все r -мерное евклидово пространство R^r при некотором $r \geq 1$, либо его подмножество той же размерности r). Пусть $s = \dim \Theta_0$, так что $s < r = \dim \Theta$; тогда значение параметра $\theta \in \Theta_0$ можно выразить в виде функции некоторого нового параметра $\delta = (\delta_1, \dots, \delta_s)$ с некоторым множеством возможных значений Δ , т. е. Θ_0 имеет параметрическое представление в терминах параметра δ :

$$\Theta_0 = \{ \theta = (\theta_1, \dots, \theta_r) : \theta = h(\delta), \quad \delta = (\delta_1, \dots, \delta_s) \in \Delta \}. \quad (4.68)$$

В общем случае будем предполагать, что функции $h(\delta) = (h_1(\delta), \dots, h_r(\delta))$ удовлетворяют следующим условиям регулярности: они трижды непрерывно дифференцируемы, и ранг матрицы их производных $\|\partial h_i(\delta) / \partial \delta_j\|$ равен s всюду в Δ .

Другой, эквивалентный, способ задания нулевой гипотезы состоит в следующем. Обозначим через $\theta^{(s)}$ совокупность тех координат вектора θ , которые являются взаимно однозначными функциями от δ . Предположим для простоты, что $\theta^{(s)} = (\theta_{r-s+1}, \dots, \theta_r)$. Тогда δ можно выразить через $\theta^{(s)}$ и записать $\delta = \delta(\theta^{(s)})$. Если теперь ввести функции $H_j(\theta) = \theta_j - h_j(\delta(\theta^{(s)}))$, $j = 1, \dots, r-s$, то гипотеза $H_0: \theta \in \Theta_0$ принимает вид

$$H_0: H_j(\theta) = 0, \quad j = 1, \dots, r-s. \quad (4.69)$$

Таким образом, гипотезу H_0 можно задавать также указанием некоторого числа ограничений на возможные значения параметра δ .

Выделим часто встречающийся в приложениях случай, когда Θ_n задается как цилиндрическое подмножество Θ , например

$$\Theta_n = \{\theta: \theta = (\theta_{10}, \dots, \theta_{r-s,0}, \theta_{r-s+1}, \dots, \theta_r)\}, \quad (4.70)$$

где θ_{i0} , $i=1, \dots, r-s$, — некоторые фиксированные значения координат $\theta_1, \dots, \theta_{r-s}$. Это частный вид гипотез (4.68) и (4.69) и сводится к ним, если положить $\delta = (\theta_{r-s+1}, \dots, \theta_r)$ и $h_j(\delta) \equiv \theta_{j0}$, $j=1, \dots, r-s$, $h_j(\delta) = \theta_j$, $j = r-s+1, \dots, r$. Если Θ_n имеет вид (4.70), то гипотезу $H_0: \theta \in \Theta_n$ называют *линейной*. Таким образом, линейная гипотеза — это гипотеза, фиксирующая значения части координат $\theta^{(1)}$ параметрического вектора θ [в данном случае $\theta^{(1)} = (\theta_1, \dots, \theta_{r-s})$], остальные координаты $\theta^{(2)}$ [в данном случае $\theta^{(2)} = (\theta_{r-s+1}, \dots, \theta_r)$] — «мешающие» параметры.

Пусть проверяемая гипотеза задана в параметрической форме (4.68). Если функцию правдоподобия наблюдений рассматривать при этой гипотезе, то она является функцией нового параметра $\delta: L_n(x; \theta) = L_n(x; h(\delta))$ и при выполнении всех условий регулярности оценка максимального правдоподобия $\hat{\delta}_n$ параметра δ будет обладать стандартными асимптотическими свойствами. Это позволяет провести асимптотическое исследование статистики отношения правдоподобия (4.54) (аналогично проведенному в п. 2 и 3), которая в данном случае имеет вид

$$\lambda_n(X; \Theta_n) = L_n(X; h(\hat{\delta}_n)) / L_n(X; \hat{\delta}_n),$$

и получить для больших выборок (при $n \rightarrow \infty$) следующие результаты:

$$1) \mathcal{L}(-2 \ln \lambda_n(X; \Theta_n) | H_0) \rightarrow \chi^2(r-s); \quad (4.71)$$

2) критерий отношения правдоподобия, основанный на критической области

$$\mathcal{R}_{1\alpha} = \{x: -2 \ln \lambda_n(x; \Theta_n) \geq \chi^2_{1-\alpha, r-s}\}, \quad (4.72)$$

состоятелен;

3) в случае линейной гипотезы (4.70) этот критерий имеет асимптотически невырождающуюся мощность (т. е. мощность, стремящуюся при $n \rightarrow \infty$ к некоторому пределу γ , $\alpha < \gamma < 1$) для локальных альтернатив вида $\theta_{i1}^{(n)} = \theta_{i0} + \beta_i / \sqrt{n}$, $i=1, \dots, r-s$, равную

$$1 - F_{r-s}(\chi^2_{1-\alpha, r-s}; \lambda^2), \quad (4.73)$$

где $\lambda^2 = \lambda^2(\theta^*) = \beta' (I^{-1}(r-s))^{-1} \beta$, $\theta^* = (\theta_{10}, \dots, \theta_{r-s,0}, \theta_{r-s+1}, \dots, \theta_r)$, $\beta = (\beta_1, \dots, \beta_{r-s})$, $I^{-1}(r-s)$ — главный минор порядка $r-s$ матрицы $I^{-1}(\theta^*)$ (выражение для параметра нецентральности в случае задания гипотезы в виде (4.68) или (4.69) можно найти в [26]).

Отметим, что все полученные ранее результаты для случая простой гипотезы H_0 можно получить из приведенных при $s=0$.

Пример 4.18 (общая нормальная модель, к. о. п. для дисперсии). Применим сформулированные результаты к задаче проверки сложной гипотезы $H_0: \theta_1 = \theta_{10}$ (θ_1 произвольно) для нормальной модели $\mathcal{N}(\theta_1, \theta_1^2)$. Используя результаты примера 4.15, найдем

$$\lambda_n(x) = (s^2/\theta_{10}^2)^{n/2} \exp \{- (n/2) (s^2/\theta_{10}^2 - 1)\} \text{ или } \\ -2 \ln \lambda_n(x) = n (s^2/\theta_{10}^2 - 1) - n \ln [1 + (s^2/\theta_{10}^2 - 1)].$$

Выборочная дисперсия S^2 является состоятельной оценкой теоретической дисперсии, поэтому в случае справедливости гипотезы H_0 отношение s^2/θ_{10}^2 при больших n близко к единице и

$$-2 \ln \lambda_n(x) \simeq (n/2) (s^2/\theta_{10}^2 - 1)^2. \quad (4.74)$$

В рассматриваемом случае параметры $r=2$, $s=1$; следовательно, к. о. п. (4.72), если учесть (4.74), можно записать в виде

$$\mathcal{R}_{1\alpha} = \{x: (n/2) (s^2/\theta_{10}^2 - 1)^2 \geq \chi^2_{1-\alpha, 1}\}. \quad (4.75)$$

В данном случае утверждение (4.71) легко получить непосредственно из (4.74). Напомним, что, по теореме 1.10, $\mathcal{L}(nS^2/\theta_{20}^2 | H_0) = \chi^2(n-1)$. Далее, если воспользоваться свойством воспроизводимости распределения χ^2 и записать представление $\chi^2 = \chi_1^2 + \dots + \chi_{n-1}^2$, где $\mathcal{L}(\chi^2) = \chi^2(n-1)$, слагаемые независимы и $\mathcal{L}(\chi_k^2) = \chi^2(1)$, $k=1, \dots, n-1$, то по центральной предельной теореме при $n \rightarrow \infty$ имеем

$$\mathcal{L}\left(\frac{\chi^2 - (n-1)}{\sqrt{2(n-1)}}\right) \rightarrow \phi^*(0, 1)$$

(см. п. 1 § 1.5). Таким образом, при $n \rightarrow \infty$

$$\mathcal{L}\left(\frac{1}{\sqrt{2(n-1)}} \left[\frac{nS^2}{\theta_{20}^2} - (n-1) \right] | H_0\right) \rightarrow \phi^*(0, 1).$$

Последний результат справедлив и для статистики $(1/\sqrt{2n})(nS^2/\theta_{20}^2 - n) = \sqrt{n/2}(S^2/\theta_{20}^2 - 1)$. Отсюда и из (4.74) следует, что при гипотезе H_0 статистика $-2 \ln \lambda_n(X)$ распределена при больших n приблизительно так же, как квадрат нормальной $\phi^*(0, 1)$ случайной величины, т. е. по закону $\chi^2(1)$, что и утверждается в (4.71).

Рассмотрим теперь локальную альтернативу $\theta_{21}^{(n)} = \theta_{20} + \beta/\sqrt{n}$, где $\beta \neq 0$, и вычислим предельную мощность (4.73) при этой альтернативе для критерия (4.75). Информационная матрица $I(\theta)$ для модели $\phi^*(\theta_1, \theta_2)$ вычислена в п. 5 § 2.2 и имеет вид $I(\theta) = \begin{bmatrix} 1/\theta_2^2 & 0 \\ 0 & 2/\theta_2^2 \end{bmatrix}$. Отсюда $I^{-1}(\theta) = \begin{bmatrix} \theta_2^2 & 0 \\ 0 & \theta_2^2/2 \end{bmatrix}$. Главный минор

этой матрицы, соответствующий фиксируемому гипотезой параметру, есть $\theta_2^2/2$, поэтому параметр нецентральности λ^2 в (4.73) в данном случае равен $2\beta^2/\theta_{20}^2$. Окончательно имеем: искомый предел мощности равен $1 - F_1(\chi_{1-\alpha, 1}^2; 2\beta^2/\theta_{20}^2)$.

5. Доверительные области максимального правдоподобия. Изложенные результаты дают возможность получить общее решение задачи доверительного оценивания параметров распределений для общих параметрических моделей. При этом используют методику, изложенную в п. 4 § 4.4; в результате получают *доверительные области максимального правдоподобия*. Далее будем предполагать, что выполняются стандартные условия регулярности, формулируемые в асимптотической теории критерия отношения правдоподобия.

Рассмотрим к. о. п. уровня значимости $\alpha = 1 - \gamma$ для простой гипотезы $\theta = (\theta_1, \dots, \theta_r)$. Согласно (4.57), область принятия этой гипотезы в случае больших выборок имеет вид $\mathcal{X}_{ny}(\theta) = \{x: -2 \ln \lambda_n(x; \theta) < \chi_{\gamma, r}^2\}$. Отсюда имеем, что

$$\mathcal{Y}_\gamma(X) = \{\theta: -2 \ln \lambda_n(X; \theta) < \chi_{\gamma, r}^2\} \quad (4.76)$$

— асимптотическая γ -доверительная область для параметра θ , так как, по теореме 4.6, при $n \rightarrow \infty$

$$P_\theta(\theta \in \mathcal{Y}_\gamma(X)) = P_\theta(-2 \ln \lambda_n(X; \theta) < \chi_{\gamma, r}^2) \rightarrow \gamma, \quad \forall \theta \in \Theta.$$

Определенное таким образом случайное подмножество параметрического множества Θ называется *доверительной областью максимального правдоподобия*. Таким образом, множество (4.76) строится на основании статистики отношения правдоподобия $\lambda_n(X; \theta) = L_n(X; \theta)/L_n(X; \hat{\theta}_n)$ и включает все значения параметра, при которых функция правдоподобия близка к своему максимальному значению $L_n(X; \hat{\theta}_n)$, или, как говорят, все достаточно *правдоподобные значения* θ^* . В частности, наиболее правдоподобное значение параметра —

* Если $L(X; \theta_1) > L(X; \theta_2)$, то говорят, что значение параметра θ_1 *более правдоподобно*, чем значение θ_2 .

оценка максимального правдоподобия $\hat{\theta}_n$ — всегда принадлежит $\mathcal{G}_\gamma(X)$ (рис. 4.4).

Приближенную доверительную область максимального правдоподобия можно построить и для произвольной совокупности координат параметрического вектора θ при наличии мешающих параметров. Пусть, например, требуется оценить координаты $\theta^{(1)} = (\theta_1, \dots, \theta_{r-s})$ (остальные координаты $\theta^{(2)} = (\theta_{r-s+1}, \dots, \theta_r)$ являются мешающими параметрами). В этом случае следует рассмотреть линейную гипотезу относительно $\theta^{(1)}$ (см. предыдущий пункт), вычислить статистику

$$\lambda_n(X; \theta^{(1)}) = \sup_{\theta^{(2)}} L_n(X; (\theta^{(1)}, \theta^{(2)})) / L_n(X; \hat{\theta}_n)$$

и на основании (4.72) положить

$$\mathcal{G}_\gamma(X) = \{\theta^{(1)} : -2 \ln \lambda_n(X; \theta^{(1)}) < \chi_{r-s}^2\}.$$

Это и есть искомая асимптотическая γ -доверительная область для $\theta^{(1)}$, так как из (4.71) — (4.72) следует, что при $n \rightarrow \infty$

$$P_\theta(\theta^{(1)} \in \mathcal{G}_\gamma(X)) = P_\theta(-2 \ln \lambda_n(X; \theta^{(1)}) < \chi_{r-s}^2) \rightarrow \gamma, \quad \forall \theta \in \Theta.$$

Сделаем несколько общих замечаний о критерии отношения правдоподобия. Применяя метод отношения правдоподобия и используя приведенные предельные теоремы, можно получать асимптотическое решение широкого класса задач. В то же время этот подход имеет границы применимости. Так, в случае его использования необходимо выполнение довольно жестких условий регулярности от исходной модели, что не всегда имеет место. Кроме того, с его помощью можно проверять только гипотезы вида $H_0: \theta \in \Theta_0$, где Θ_0 — евклидово подпространство пространства Θ , $\dim \Theta_0 < \dim \Theta$; многие же интересные задачи к этому типу не относятся.

Задачи

1. В последовательности независимых испытаний вероятности положительных исходов одинаковы и равны p . Построить критерий проверки гипотезы $H_0: p=0$ против альтернативы $H_1: p=0,01$ и определить наименьший объем выборки, для которого вероятности ошибок первого и второго рода не превосходят 0,01.

2. Выборочные значения непрерывной случайной величины ξ оказались равными $-0,460, -0,114, -0,325, 0,196, -0,174$. Проверить гипотезу $H_0: \mathcal{L}(\xi) = R(-0,5; 0,5)$ против альтернативы $H_1: \mathcal{L}(\xi) = \mathcal{L}^*(0; 0,09)$ (вероятность ошибки первого рода принять равной 0,1).

3. Пусть для распределения Коши $K(\theta)$ проверяется гипотеза $H_0: \theta=0$ против альтернативы $H_1: \theta=1$. Показать, что р. н. м. критерий уровня значимости $\alpha = 1/2 - (1/\pi) \arctg(1/2) \approx 0,352$ по одному наблюдению задается критической областью $\{x \geq 1/2\}$ и что его мощность равна $1/2 + (1/\pi) \arctg(1/2) \approx 0,648$. Если $\alpha = (1/\pi)(\arctg 3 - \arctg 1) \approx 0,148$, то критическая область имеет вид $\{1 \leq x \leq 3\}$, а мощность равна $(1/\pi) \arctg 2 \approx 0,352$.

4. Пусть $X = (X_1, \dots, X_n)$ — выборка из экспоненциального распределения $\Gamma(\theta, 1)$. Построить критерий Неймана — Пирсона проверки гипотезы $H_0: \theta = \theta_0$ против альтернативы $H_1: \theta = \theta_1$. Найти функцию мощности этого критерия.

5. Пусть для пуассоновской модели $\Pi(\theta)$ проверяется гипотеза $H_0: \theta = \theta_0$ против сложной альтернативы $H_1: \theta > \theta_0$. Показать, что р. н. м. критерий задается критической областью вида $\{x \geq c\}$. Вычислить функцию мощности критерия и показать его несмещенность. Установить, что при большом объ-

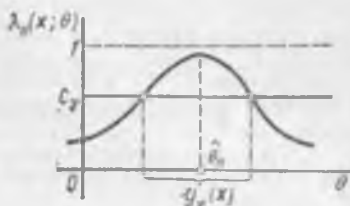


Рис. 4.4

еме выборки n критическая граница $c=c_\alpha$ с уровнем значимости α может быть выбрана равной $\theta_0 + t_\alpha \sqrt{\theta_0/n}$, $\Phi(-t_\alpha) = \alpha$. Рассмотреть альтернативу вида $H_1: \theta < \theta_0$.

6. Показать, что при большом объеме выборки критическую границу t_α в критерии примера 4.5 можно вычислять с помощью приближенной формулы $r_\alpha = n\theta_0 - t_\alpha \sqrt{n\theta_0(1-\theta_0)}$, $\Phi(-t_\alpha) = \alpha$.

7. Построить асимптотическую форму к. о. п. проверки гипотезы $H_0: \theta = \theta_0$ в модели $\Pi(\theta)$ и вычислить его предельную мощность при альтернативе $\theta_1^{(n)} = \theta_0 + \beta/\sqrt{n}$.

8. Установить равенство $Q_n^{**} = X_n^*$ в примере 4.16.

9. Пусть в модели $Bi(1, \theta)$ проверяется гипотеза $H_0: \theta = \theta_0$. Построить к. о. п. и вычислить его предельную мощность при альтернативе $\theta_1^{(n)} = \theta_0 + \beta/\sqrt{n}$ (n — объем выборки).

10. Показать, что предельная (при $n \rightarrow \infty$) мощность построенного в примере 4.15 к. о. п. при альтернативе $\theta_1^{(n)} = \theta_0 + \beta/\sqrt{n}$ равна $1 - F_1(\chi_{1-\alpha}^2; 1; \beta^2/\theta_0^2)$.

11. Имеются две независимые выборки из пуассоновских совокупностей $\Pi(\theta_1)$ и $\Pi(\theta_2)$, соответственно $X_1 = (X_{11}, \dots, X_{1n_1})$ и $X_2 = (X_{21}, \dots, X_{2n_2})$. Требуется проверить гипотезу $H_0: \theta_1 = \theta_2$ (гипотезу однородности). Построить к. о. п. и установить, что статистика этого критерия имеет вид $\lambda_{n_1, n_2} = \frac{X_1 X_2}{(X_1^{n_1} X_2^{n_2})}$, $n = n_1 + n_2$, $X = (n_1 X_1 + n_2 X_2)/n$. Найти предельное распределение $-2 \ln \lambda_{n_1, n_2}$ при $n_1, n_2 \rightarrow \infty$. Обобщить на случай k выборок.

12. Пусть $S_1^2, \dots, S_k^2$ — выборочные дисперсии, построенные по независимым выборкам объемов $n_1, \dots, n_k$ из совокупностей $\mathcal{P}^*(\theta_{11}, \theta_{11}^2), \dots, \mathcal{P}^*(\theta_{k1}, \theta_{k1}^2)$ соответственно. Показать, что к. о. п. для гипотезы $H_0: \theta_{11}^2 = \dots = \theta_{k1}^2$ основан на статистике

$$\lambda_{n_1 \dots n_k} = \prod_{i=1}^k (S_i^2/S_1^2)^{n_i/2}, \quad S_1^2 = \frac{n_1 S_1^2 + \dots + n_k S_k^2}{n_1 + \dots + n_k}.$$

Построить асимптотическую при $n_1, \dots, n_k \rightarrow \infty$ форму этого критерия. Убедиться, что в случае $k=2$ статистика λ_{n_1, n_2} взаимно однозначно связана с отношением Снедекора $F = [n_1(n_2-1)S_1^2]/[n_2(n_1-1)S_2^2]$, которое имеет при гипотезе H_0 распределение Снедекора $S(n_1-1, n_2-1)$.

13. Построить и рассчитать асимптотический вариант к. о. п. проверки гипотезы однородности в случае независимых выборок из распределений $Bi(1, \theta_j)$, $j=1, \dots, k$.

14. Пусть $n_i, \bar{X}_i$ и S_i^2 — соответственно объем, выборочные среднее и дисперсия для выборки из совокупности $\mathcal{P}^*(\theta_{i1}, \theta_{i1}^2)$, $i=1, 2$. Построить к. о. п. для гипотезы $H_0: \theta_{11} = \theta_{21}$. Показать, что статистика отношения правдоподобия эквивалентна в этом случае студентову отношению

$$t = \sqrt{n_1 n_2 (n_1 + n_2 - 2) / (n_1 + n_2)} (\bar{X}_1 - \bar{X}_2) / \sqrt{n_1 S_1^2 + n_2 S_2^2},$$

которое имеет при гипотезе H_0 распределение Стюдента $S(n_1 + n_2 - 2)$.

Линейная регрессия и метод наименьших квадратов

В этой главе рассматривается практически важный случай неодинаково распределенных наблюдений с одинаковыми дисперсиями, когда их средние — линейные функции неизвестных параметров (схема линейной регрессии). Излагаются классический метод наименьших квадратов для оценивания параметров модели и его оптимальные свойства, рассматриваются вопросы его практического применения. В предположении нормальности наблюдений приводится полное решение задач доверительного оценивания произвольных линейных функций параметров регрессии и проверки линейных гипотез для них. Рассматривается применение модели нормальной регрессии к задачам дисперсионного анализа. Излагаются элементы теории статистической регрессии и корреляции.

§ 5.1. Модель линейной регрессии

До сих пор рассматривались статистические выводы для моделей, соответствующих повторным независимым наблюдениям над некоторой случайной величиной ξ . В этих случаях исходные статистические данные представляют собой реализацию случайного вектора $X = (X_1, \dots, X_n)$, компоненты которого независимы и одинаково распределены, а именно $F_{X_i} = F_\xi$, $i = 1, \dots, n$. В приложениях математической статистики предположения о независимости и одинаковой распределенности компонент X_i не всегда выполняются. В этой главе будет рассмотрен важный случай таких ситуаций, часто встречающихся в приложениях, которые описываются в терминах *линейной регрессионной модели*. В этой модели предполагается, что математические ожидания наблюдений X_i являются *линейными* функциями $\varphi_i(\beta)$ от неизвестных параметров $\beta = (\beta_1, \dots, \beta_k)$ и делаются некоторые предположения о вторых моментах.

В общем случае речь идет об изучении таких случайных экспериментов, когда на их исход влияют некоторые неслучайные переменные (факторы) $z_1, \dots, z_k$, значения которых меняются от опыта к опыту. Например, $z_1, \dots, z_k$ могут быть входными характеристиками прибора, при каждой задаваемой произвольно комбинации значений которых наблюдается некоторый эффект («отклик») как исход эксперимента. В других случаях значения факторов для каждого опыта могут быть обусловлены не зависящими от исследователя причинами (как правило, это имеет место,

например, в экономических исследованиях). Но в любом случае статистические данные состоят из множества наблюдавшихся значений «откликов» $X_1, \dots, X_n$ и соответствующих значений факторов, т. е. имеют вид $(x_i; z_1^{(i)}, \dots, z_k^{(i)})$, $i = 1, \dots, n$; при этом всегда предполагается, что $k \leq n$.

Часто можно считать, что факторы $z_1, \dots, z_k$ оказывают влияние только на среднее значение «отклика», при этом математическое ожидание исхода i -го опыта имеет вид

$$EX_i = \sum_{j=1}^k z_j^{(i)} \beta_j = \mathbf{z}^{(i)'} \boldsymbol{\beta}, \quad \mathbf{z}^{(i)} = (z_1^{(i)}, \dots, z_k^{(i)}),$$

т. е. является линейной функцией неизвестных параметров $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)$. (Напомним, что при матричных преобразованиях векторы всегда понимаются как вектор-столбцы.) Другими словами, случайную величину X_i , описывающую результат i -го опыта, можно представить в виде

$$X_i = \mathbf{z}^{(i)'} \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n,$$

где $E\varepsilon_i = 0$ и распределение «ошибки» ε_i от параметров $\boldsymbol{\beta}$ не зависит. Введем матрицу плана $\mathbf{Z} = [\mathbf{z}^{(1)} \dots \mathbf{z}^{(n)}]$ размером $k \times n$, составленную из вектор-столбцов $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(n)}$, и вектор ошибок $\mathbf{e} = (\varepsilon_1, \dots, \varepsilon_n)$. Тогда в матричных обозначениях предыдущие равенства принимают вид

$$\mathbf{X} = \mathbf{Z}' \boldsymbol{\beta} + \mathbf{e}, \quad E(\mathbf{e}) = \mathbf{0}. \quad (5.1)$$

Далее обычно предполагают, что случайные величины $\varepsilon_1, \dots, \varepsilon_n$ (или, что то же самое, $X_1, \dots, X_n$) не коррелированы и имеют одинаковые дисперсии $DX_i = D\varepsilon_i = \sigma^2 > 0$, $i = 1, \dots, n$, где σ^2 обычно неизвестна. В этом случае матрица вторых моментов вектора наблюдений $\mathbf{X}$ имеет следующий вид:

$$\mathbf{D}(\mathbf{X}) = \mathbf{D}(\mathbf{e}) = E(\mathbf{e}\mathbf{e}') = \sigma^2 \mathbf{E}_n. \quad (5.2)$$

Если выполняются условия (5.1) — (5.2), то говорят, что имеет место модель *линейной регрессии*. Параметры $\beta_1, \dots, \beta_k$ называют *коэффициентами регрессии*, а σ^2 — *остаточной дисперсией*.

В описанной схеме переменные $z_1, \dots, z_k$ могут быть функционально зависимы. В частности, все z_i могут являться функциями одной переменной t , например $z_j = a_j(t)$, где $a_j(t)$ — полином степени $j > 1$, в этом случае говорят о *параболической регрессии*.

В более общем случае возможны корреляции между наблюдениями, т. е. вместо условия (5.2) вводится условие $\mathbf{D}(\mathbf{e}) = \sigma^2 \mathbf{G}$.

Если матрица $\mathbf{G}$ известна, то, положив $\mathbf{Y} = \mathbf{G}^{-1/2} \mathbf{X}$, получим, что преобразованные таким образом данные удовлетворяют модели (5.1) — (5.2) (при линейном преобразовании $\mathbf{Y} = \mathbf{LX}$ первые и вторые моменты случайного вектора преобразуются следующим образом: $E(\mathbf{Y}) = \mathbf{L}E(\mathbf{X})$, $\mathbf{D}(\mathbf{Y}) = \mathbf{L}\mathbf{D}(\mathbf{X})\mathbf{L}'$). Таким образом, подробного изучения заслуживает только стандартная модель (5.1) — (5.2).

Далее определяющую роль играет матрица

$$\mathbf{A} = \mathbf{Z}\mathbf{Z}'. \quad (5.3)$$

Эта матрица всегда неотрицательно определена и положительно определена тогда и только тогда, когда $\text{rang } Z = k$, т. е. когда строки матрицы Z линейно независимы. Действительно, квадратичная форма $t'At = t'ZZ't = (Z't)'(Z't) \geq 0$, $\forall t = (t_1, \dots, t_k)$, причем равенство нулю возможно только при $Z't = 0$. В свою очередь, это равенство эквивалентно равенству $t_1 z_1 + \dots + t_k z_k = 0$, где $z_1, \dots, z_k$ — столбцы матрицы Z' . Таким образом, равенство нулю $t'At$ при некотором $t \neq 0$ имеет место только в случае линейной зависимости векторов $z_1, \dots, z_k$, т. е. при $\text{rang } Z < k$.

Излагаемая далее теория строится в предположении, что матрица A не вырождена (или, что эквивалентно, $\text{rang } Z = k$). Полные результаты, относящиеся и к вырожденному случаю, можно найти в [16, гл. 4].

§ 5.2. Оценивание неизвестных параметров модели

В этом параграфе будет рассмотрена задача построения точечных оценок параметров $\beta_1, \dots, \beta_k$ и σ^2 модели линейной регрессии (5.1) — (5.2).

1. Метод наименьших квадратов. Общим методом оценивания неизвестных коэффициентов регрессии $\beta_1, \dots, \beta_k$ является разработанный К. Гауссом (1809 г.) и А. Марковым (1900 г.) *метод наименьших квадратов*, в соответствии с которым оценки этих параметров находят из условия обращения в минимум квадратичной формы

$$S(\beta) = S(X; \beta) = (X - Z'\beta)'(X - Z'\beta), \quad (5.4)$$

представляющей собой сумму квадратов разностей между наблюдениями и их математическими ожиданиями. Точку $b = (b_1, \dots, b_k)$, удовлетворяющую равенству $S(b) = \min_{\beta} S(\beta)$, называют, по определению, *оценкой наименьших квадратов* (о. н. к.) параметра $\beta = (\beta_1, \dots, \beta_k)$.

Пусть $Y = ZX$; тогда с помощью непосредственных вычислений можно убедиться, что система уравнений $\partial S(\beta) / \partial \beta_j = 0$, $j = 1, \dots, k$, в матричной форме записывается в виде

$$A\beta = Y, \quad (5.5)$$

где матрица A задана в (5.3). Это уравнение для экстремальных точек β называют *нормальным уравнением* метода наименьших квадратов. Справедливо следующее утверждение.

Теорема 5.1. Пусть β — любое решение нормального уравнения. Тогда

$$\min_{\beta} S(\beta) = S(\beta)$$

и, следовательно, этот минимум одинаков для всех β .

Если $|A| \neq 0$, то о. н. к. единственна и определяется равенством

$$b = \hat{\beta} = A^{-1}Y = A^{-1}ZX. \quad (5.6)$$

□ Пусть β^* — произвольное фиксированное значение β ; тогда из (5.4) имеем

$$S(\beta) = [X - Z'\beta^* + Z'(\beta^* - \beta)]' [X - Z'\beta^* + Z'(\beta^* - \beta)] = \\ = S(\beta^*) + 2(\beta^* - \beta)'(Y - A\beta^*) + (\beta^* - \beta)'A(\beta^* - \beta). \quad (5.7)$$

Если $\beta^* = \hat{\beta}$, то

$$S(\beta) = S(\hat{\beta}) + (\beta - \hat{\beta})'A(\beta - \hat{\beta}) \geq S(\hat{\beta}), \quad (5.8)$$

поскольку матрица A неотрицательно определена. Таким образом, минимум $S(\beta)$ равен $S(\hat{\beta})$, достигается при $\beta = \hat{\beta}$ и одинаков для всех решений $\hat{\beta}$ уравнения (5.5). Отсюда следует, что любое решение нормального уравнения является о. н. к. для β . Для невырожденной матрицы A уравнение (5.5) однозначно разрешимо и, следовательно, в этом случае о. н. к. единственна и имеет вид (5.6). ■

В ряде случаев интерес представляют не сами параметры $\beta_1, \dots, \beta_k$, а их некоторые линейные комбинации, т. е. новый параметрический вектор $t = (t_1, \dots, t_m)$, $m \leq k$, связанный с β соотношением $t = T\beta$, где T — заданная матрица размером $m \times k$. В этом случае о. н. к. $\hat{t}$ для t определяется равенством $\hat{t} = T\hat{\beta}$, где $\hat{\beta}$ — любое решение нормального уравнения (5.5). Если $|A| \neq 0$, то из (5.6) следует, что $\hat{t}$ определяется однозначно и имеет вид

$$\hat{t} = TA^{-1}Y = TA^{-1}ZX. \quad (5.9)$$

2. Оптимальность оценок наименьших квадратов. Исследуем свойства полученных оценок. В общем случае будем рассматривать задачу оценивания вектора $t = T\beta$ в классе линейных оценок, т. е. оценок вида $l = LX$, являющихся линейными функциями от наблюдений $X = (X_1, \dots, X_n)$.

Теорема 5.2. Пусть матрица A не вырождена. Тогда для произвольного вектора $t = T\beta$ о. н. к. $\hat{t}$, определенная равенством (5.9), является несмещенной оценкой с минимальной дисперсией в классе всех линейных несмещенных оценок t ; при этом матрица вторых моментов случайного вектора $\hat{t}$ имеет вид

$$D(\hat{t}) = \sigma^2 TA^{-1}T' = \sigma^2 D. \quad (5.10)$$

□ Из (5.9) и (5.1) имеем

$$E(\hat{t}) = E(TA^{-1}ZX) = TA^{-1}ZE(X) = TA^{-1}ZZ'\beta = T\beta = t,$$

т. е. $\hat{t}$ — линейная несмещенная оценка t . Пусть $l = LX$ — произвольная линейная несмещенная оценка t , т. е. $E(l) = LE(X) = LZ'\beta = T\beta$. Это равенство должно выполняться для всех β , по-

этому отсюда следует, что

$$LZ' = T. \quad (5.11)$$

Из (5.2) находим

$$D(l) = LD(X) L' = \sigma^2 LL'. \quad (5.12)$$

Наша цель — минимизировать дисперсии оценок $l_1, \dots, l_m$, т. е. диагональные элементы матрицы LL' . Для этого запишем тождество

$$LL' = (TA^{-1}Z)(TA^{-1}Z)' + (L - TA^{-1}Z)(L - TA^{-1}Z)',$$

которое непосредственно следует из равенства (5.11). Каждое слагаемое правой части этого тождества имеет вид HH' , откуда следует неотрицательность диагональных элементов. Но от L зависит только второе слагаемое, поэтому диагональные элементы $D(l)$ одновременно достигают минимума тогда и только тогда, когда $L = TA^{-1}Z$. Соответствующая оптимальная оценка имеет вид $l^* = TA^{-1}ZX = \hat{l}$, т. е. совпадает с о. н. к. (5.9). Наконец, формула (5.10) следует из соотношения (5.12), если подставить вместо L найденное оптимальное решение. ■

В качестве простого следствия теоремы 5.2 получаем, что $D(\beta) = \sigma^2 A^{-1}$ или

$$\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 a^{ij}, \quad i, j = 1, \dots, k, \quad (5.13)$$

$$\text{где } |a^{ij}| = A^{-1} = |a_{ij}|^{-1}.$$

Замечание. Если матрица T имеет вид $T = \Lambda A$, то формулы (5.9) и (5.10) принимают соответственно вид

$$t = \Lambda Y, \quad D(t) = \sigma^2 \Lambda \Lambda A \Lambda'$$

т. е. необходимость в вычислении обратной матрицы A^{-1} для нахождения о. н. к. и их вторых моментов отпадает.

Итак, теорема 5.2 позволяет решить задачу о построении оптимальных оценок для произвольных линейных функций от коэффициентов регрессии; это оценки наименьших квадратов.

3. Оценивание остаточной дисперсии. Из равенства (5.4) имеем

$$ES(\beta) = \sum_{i=1}^n DX_i = n\sigma^2.$$

Далее, учитывая (5.13), найдем

$$\begin{aligned} E(\hat{\beta} - \beta)' A (\hat{\beta} - \beta) &= \sum_{i,j=1}^k a_{ij} E(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j) = \\ &= \sigma^2 \sum_{i,j=1}^k a_{ij} a^{ij} = \sigma^2 \text{tr}(AA^{-1}) = \sigma^2 \text{tr}(E_k) = k\sigma^2. \end{aligned}$$

Отсюда и из тождества (5.8) следует, что $ES(\hat{\beta}) = (n - k)\sigma^2$, т. е. несмещенной оценкой для остаточной дисперсии σ^2 является ста-

$$\hat{\sigma}^2 = \frac{1}{n-k} S(\hat{\beta}) = \frac{1}{n-k} (X - Z'\hat{\beta})' (X - Z'\hat{\beta}). \quad (5.14)$$

Вектор $U = X - Z'\hat{\beta}$ называют *остаточным вектором*, а его компоненты — *остатками*. Таким образом, оценка $\hat{\sigma}^2$ равна сумме квадратов остатков, поделенной на $n - k$ (разность между числом наблюдений и числом параметров β_i).

Приведем другое выражение для (5.14), которое понадобится в дальнейшем. Используя разложение (5.7), запишем остаточный вектор в виде

$$U = (E_n - Z'A^{-1}Z)X = BX. \quad (5.15)$$

Непосредственно можно проверить, что матрица B , определяемая этим равенством, симметрична и идемпотентна ($B^2 = B$); следовательно, вместо (5.14) можно использовать представление

$$\hat{\sigma}^2 = \frac{1}{n-k} X'BX, \quad (5.16)$$

показывающее явную зависимость оценки $\hat{\sigma}^2$ от наблюдений. Наконец, из (5.15) имеем, что первые и вторые моменты остаточного вектора имеют вид

$$E(U) = 0, \quad D(U) = \sigma^2 B = D(X) - D(Z'\hat{\beta}). \quad (5.17)$$

4. **Обобщенные о. н. к.** Ранее рассматривался случай, когда на возможные значения параметров $\beta = (\beta_1, \dots, \beta_k)$ не накладывалось никаких ограничений, т. е. областью их возможных значений было все евклидово пространство R^k . Однако в ряде задач допустимые значения β бывают ограничены теми или иными условиями. Часто эти условия имеют вид линейных ограничений на параметры $\beta_1, \dots, \beta_k$, что в общем виде будем записывать так:

$$T\beta = t_0, \quad (5.18)$$

где T — некоторая заданная матрица размером $m \times k$ ($m \leq k$) и t_0 — заданный m -мерный вектор такой, что система (5.18) совместна. Другими словами, условие (5.18) означает, что допустимые значения коэффициентов регрессии $\beta_1, \dots, \beta_k$ удовлетворяют m заданным линейным ограничениям

$$t_j'\beta = t_{j0}, \quad j = 1, \dots, m, \quad (5.19)$$

где $t_0 = (t_{10}, \dots, t_{m0})$ и $t_1', \dots, t_m'$ — строки матрицы T . Естественно предполагать, что ограничения (5.19) линейно независимы (иначе можно перейти к меньшему числу уравнений, исключив линейно зависимые). Таким образом, в дальнейшем будем считать, что $\text{rang } T = m$ (случай $m = k$, однозначно фиксирующий вектор β , в последующих рассуждениях формально допускается).

Рассмотрим задачу оценивания параметров β в этой усложненной ситуации. Обозначим

$$S_T = \min_{\beta: T\beta = t_0} S(\beta) \quad (5.20)$$

и назовем обобщенной оценкой наименьших квадратов β , то значение β (удовлетворяющее условию (5.18)), при котором $S = S(\beta_T)$. Нахождение обобщенной о. н. к. — это задача нахождения условного экстремума функции $S(\beta)$, и ее можно решить методом неопределенных множителей Лагранжа. Приведем только окончательный результат:

$$\beta_T = \beta - A^{-1}T'D^{-1}(T\beta - t_0), \quad (5.21)$$

где β — обычная о. н. к. (без ограничений на параметры β), определенная равенством (5.6), а матрица $D = TA^{-1}T'$ размером $m \times m$ положительно определена.

Докажем тот факт, что β_T — точка условного минимума квадратичной формы $S(\beta)$. Из (5.21) непосредственно получаем, что $T\beta_T = t_0$, т. е. точка β , удовлетворяет условию (5.18). Воспользуемся далее разложением (5.7) для $S(\beta)$ и положим в нем $\beta^* = \beta_T$:

$$S(\beta) = S(\beta_T) + 2(\beta_T - \beta)'(Y - A\beta_T) + (\beta_T - \beta)'A(\beta_T - \beta).$$

Так как $A\beta = Y$, то из (5.21) следует, что $Y - A\beta_T = T'D^{-1} \times \times (T\beta - t_0)$. Учитывая также, что $T(\beta_T - \beta) = t_0 - T\beta$, получаем равенство

$$S(\beta) = S(\beta_T) + 2(t_0 - T\beta)'D^{-1}(T\beta - t_0) + (\beta_T - \beta)'A(\beta_T - \beta), \quad (5.22)$$

справедливое при всех β . Пусть теперь β удовлетворяет условию (5.18); тогда средний член обращается в нуль и

$$S(\beta) = S(\beta_T) + (\beta_T - \beta)'A(\beta_T - \beta) \geq S(\beta_T),$$

причем равенство достигается только при $\beta = \beta_T$.

Замечание. Полагая в (5.22) $\beta = \beta$ и учитывая, что в силу (5.21) $(\beta_T - \beta)'A(\beta_T - \beta) = (T\beta - t_0)'D^{-1}(T\beta - t_0)$, получаем следующее представление условного минимума $S_T = S(\beta_T)$ через абсолютный минимум $S(\beta)$:

$$S_T = S(\beta) + (T\beta - t_0)'D^{-1}(T\beta - t_0) = S(\beta) + Q_T \quad (5.23)$$

(Q_T определено этим равенством).

5. Оптимальный выбор матрицы плана. Рассмотрим «активный» эксперимент, т. е. когда значения факторов $z_1, \dots, z_n$ для каждого опыта выбирает исследователь. Покажем, как в этом случае оптимально задать матрицу плана $Z = [z^{(1)} \dots z^{(n)}]$, где i -й столбец $z^{(i)}$ — комбинация значений факторов для i -го опыта ($i = 1, \dots, n$). В качестве критерия оптимальности естественно использовать величину дисперсий оценок β , [см. (5.13)]; тогда задача сводится к выбору такой матрицы Z , чтобы диагональные элементы матрицы $A^{-1} = (ZZ')^{-1}$ были минимальны.

Если значения факторов выбирать произвольными, то все элементы матрицы A^{-1} можно сделать одновременно как угодно малыми, так как если Z заменить на aZ , то A^{-1} заменится на $a^{-2}A^{-1} \rightarrow 0$ при $a \rightarrow \infty$. Чтобы исключить этот случай, предположим, что значения факторов можно менять в ограниченных областях, именно: наложим ограничения вида

$$z_j' z_j = \sum_{i=1}^n (z_j^{(i)})^2 = a_j^2, \quad j=1, \dots, k, \quad (5.24)$$

где $z_1, \dots, z_k$ — столбцы матрицы Z' (наборы значений n уровней соответствующих факторов) и $a_1^2, \dots, a_k^2$ — заданные положительные константы.

При ограничениях (5.24) на допустимые значения n уровней каждого фактора следует подобрать комбинацию значений факторов (т. е. выбрать столбцы матрицы Z), так, чтобы дисперсии о. н. к. параметров $\beta_1, \dots, \beta_k$ приняли наименьшие возможные значения.

Теорема 5.3. При условиях (5.24) имеют место неравенства

$$D\hat{\beta}_j \geq \sigma^2/a_j^2, \quad j=1, \dots, k,$$

и минимум достигается тогда и только тогда, когда столбцы матрицы Z' ортогональны.

□ Пусть $j=1$. Запишем матрицу $A=ZZ'$ в виде

$$A = \begin{bmatrix} z_1' z_1 & b' \\ b & F \end{bmatrix},$$

где $b = (z_1' z_2, \dots, z_1' z_k)$. Тогда $a^{11} = |F|/|A|$.

Далее, умножая определитель $|A|$ справа на определитель, равный 1, получаем следующую формулу:

$$|A| = \begin{vmatrix} z_1' z_1 & b' \\ b & F \end{vmatrix} \begin{vmatrix} 1 & 0 \\ -F^{-1}b & E_{k-1} \end{vmatrix} = \begin{vmatrix} z_1' z_1 - b' F^{-1} b & b' \\ 0 & F \end{vmatrix} = \\ = |F| (z_1' z_1 - b' F^{-1} b).$$

Отсюда, учитывая соотношения (5.13) и (5.24), имеем

$$D\hat{\beta}_1 = \sigma^2 a^{11} = \sigma^2 / (a_1^2 - b' F^{-1} b) \geq \sigma^2 / a_1^2,$$

так как $b' F^{-1} b \geq 0$. Знак равенства имеет место только при $b=0$ (поскольку подматрица F положительно определена), т. е. при $z_j' z_1 = 0, j=2, \dots, k$ (первый столбец матрицы Z' ортогонален всем остальным столбцам). Доказательство для других j можно получить простой перенумерацией факторов. ■

В заключение отметим, что для оптимальной матрицы плана Z матрица A является диагональной с диагональными элементами $a_j^2 = z_j' z_j, j=1, \dots, k$, поэтому проблема обращения A для вычис-

ления о. н. к. и их вторых моментов отпадает; в этом случае

$$\beta_j = \frac{z_j' X}{z_j' z_j}, \quad D\hat{\beta}_j = \frac{\sigma^2}{z_j' z_j}, \quad \text{cov}(\hat{\beta}_j, \hat{\beta}_r) = 0, \quad j \neq r. \quad (5.25)$$

6. Примеры применения метода наименьших квадратов.

Пример 5.1 (простая регрессия, оценивание параметров). Проиллюстрируем общую теорию на примере важного для практических приложений случая простой регрессии, когда число параметров $k=2$, т. е. $\beta = (\beta_1, \beta_2)$, а векторы $z^{(i)}$ имеют вид $z^{(i)} = (1, t_i)$, $i=1, \dots, n$. Тогда

$$EX_i = \beta_1 + \beta_2 t_i, \quad i=1, \dots, n, \quad (5.26)$$

т. е. среднее значение наблюдений является линейной функцией одного фактора t . Так, t может быть температурой, при которой производится эксперимент, дозой лечебного препарата, возрастом обследуемых лиц и т. д., и речь идет об изучении связи между откликом (исходом эксперимента) и фактором t на основании выборки, при этом регистрируют n пар измерений (X_i, t_i) , $i=1, \dots, n$, где X_i наблюдается при значении t_i фактора t .

Прямую $\varphi(t) = \beta_1 + \beta_2 t$, соответствующую (5.26), называют *линией регрессии*, а коэффициент β_2 — ее *наклоном*.

В данном случае матрицы Z , A и столбец $Y=ZX$ равны:

$$Z = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}, \quad A = \begin{bmatrix} n & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}, \quad Y = \begin{pmatrix} \sum X_i \\ \sum t_i X_i \end{pmatrix}.$$

Будем предполагать, что не все t_i одинаковы (чтобы $\text{rang } Z=2$), тогда

$$|A| = n \sum_{i=1}^n t_i^2 - \left(\sum_{i=1}^n t_i \right)^2 = n \sum_{i=1}^n (t_i - \bar{t})^2 > 0$$

(черта сверху означает арифметическое среднее) и

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} \sum t_i^2 - n\bar{t} & -n\bar{t} \\ -n\bar{t} & n \end{bmatrix}, \quad A^{-1}Y = \frac{n}{|A|} \begin{pmatrix} X \sum t_i - \bar{t} \sum t_i X_i \\ \sum t_i X_i - n\bar{t} X \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix}.$$

В результате несложных преобразований запишем оценки $\hat{\beta}_1$ и $\hat{\beta}_2$ в следующем удобном для вычислений виде:

$$\hat{\beta}_2 = \sum (t_i - \bar{t})(X_i - \bar{X}) / \sum (t_i - \bar{t})^2, \quad \hat{\beta}_1 = \bar{X} - \bar{t} \hat{\beta}_2. \quad (5.27)$$

Вторые моменты этих оценок образуют матрицу $\sigma^2 A^{-1}$, поэтому

$$D\hat{\beta}_1 = \frac{\sum t_i^2}{n \sum (t_i - \bar{t})^2} \sigma^2, \quad D\hat{\beta}_2 = \frac{\sigma^2}{\sum (t_i - \bar{t})^2}, \quad \text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\frac{\bar{t}}{\sum (t_i - \bar{t})^2} \sigma^2.$$

Наконец, величина $S(\hat{\beta})$ (5.14) равна

$$S(\hat{\beta}) = \sum (X_i - \bar{X})^2 - \hat{\beta}_2^2 \sum (t_i - \bar{t})^2 \quad (5.28)$$

и несмещенная оценка для остаточной дисперсии σ^2 имеет вид $\hat{\sigma}^2 = S(\hat{\beta}) / (n-2)$.

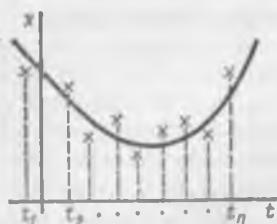


Рис. 5.1

Пример 5.2 (параболическая регрессия) Пусть j -й фактор z_j является полиномом $a_j(t)$ степени $j-1$ от общей переменной t ; тогда $z_j = (a_j(t_1), \dots, a_j(t_n))$, $j = 1, \dots, k$, — набор его значений для n уровней, а $z^{(i)} = (a_1(t_i), \dots, a_k(t_i))$, $i = 1, \dots, n$, — комбинация значений факторов $z_1, \dots, z_k$ для i -го опыта. Таким образом, матрица плана $Z = [z^{(1)} \dots z^{(n)}]$ определяется выбором n точек $t_1, \dots, t_n$ значений общего фак-

тора t для n опытов. В этом случае среднее значение i -го наблюдения X_i имеет вид

$$EX_i = \varphi(t_i; \beta) = \sum_{j=1}^k \beta_j a_j(t_i), \quad i = 1, \dots, n. \quad (5.29)$$

Определенная здесь функция $\varphi(t; \beta)$ представляет собой полином (параболу) степени $k-1$ и называется *кривой параболической регрессии*.

Приведем типичную ситуацию, когда имеет место схема параболической регрессии. Предположим, что имеется теоретическая зависимость вида $x = \varphi(t; \beta)$ между переменными t и x , причем значения t и x получают из наблюдений. Задача состоит в определении по экспериментальным данным неизвестных параметров $\beta_1, \dots, \beta_k$, входящих в это уравнение. Каждое измерение x при заданном t дает уравнение, связывающее неизвестные параметры, и если бы измерения величины x производились без погрешностей, то для определения параметров $\beta_1, \dots, \beta_k$ было бы достаточно k измерений. Однако точные измерения на практике чаще всего невозможны, поэтому для заданных значений t соответствующие значения x известны с какими-то погрешностями, т. е. реально вместо точного значения $x_i = \varphi(t_i; \beta)$ имеем случайный результат $X_i = \varphi(t_i; \beta) + \varepsilon_i$, где ε_i — ошибка измерения. Если условия эксперимента обеспечивают отсутствие систематической ошибки ($E\varepsilon_i = 0$), равнозначность ($D\varepsilon_i = \sigma^2$, $i = 1, \dots, n$) и некоррелированность ($\text{cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$) результатов измерений, то приходим к схеме регрессии вида (5.29).

Чтобы исключить влияние погрешностей измерений, нужна дополнительная информация. Для этого производят большое число измерений $n > k$. Полученные экспериментальные данные обрабатывают по методу наименьших квадратов. В результате находят не точные значения коэффициентов $\beta_1, \dots, \beta_k$, а определяют их о. н. к. $\hat{\beta}_1, \dots, \hat{\beta}_k$. В данном случае этому методу можно дать следующую наглядную геометрическую интерпретацию (рис. 5.1).

Нанесем на плоскости (t, x) наблюдавшиеся точки (t_i, X_i) , $i = 1, \dots, n$. Тогда значения β^* неизвестных коэффициентов β подбирают так, чтобы соответствующий *эмпирический* график $x = \varphi(t; \beta^*)$ наилучшим образом проходил около наблюдавшихся точек. Если в качестве критерия оптимальности взять

величину $\sum_{i=1}^n (X_i - \varphi(t_i; \beta^*))^2$, то приходим к методу наименьших квадратов

с решением $\beta^* = \hat{\beta}$.

Формально такой же вид имеет и классическая задача математического анализа приближения функций многочленами. В этом случае данные (t_i, X_i) , $i = 1, \dots, n$, интерпретируют как пары соответствующих абсцисс и ординат графика изучаемой функции и задача состоит в подборе *интерполяционного*

многочлена вида $\varphi(t; \beta)$, который давал бы наилучшее приближение в среднем, т. е. минимизировал бы величину $\sum_{i=1}^n (X_i - \varphi(t_i; \beta))^2$.

7. Ортогональные многочлены Чебышева. В описанной в примере 5.2 схеме предполагалось, что многочлены $a_j(t)$ известны. В частном случае $a_j(t) = t^{j-1}$, $j = 1, \dots, k$; тогда $\varphi(t; \beta) = \sum_{j=1}^k \beta_j t^{j-1}$, т. е. β_j — коэффициенты многочлена $\varphi(t; \beta)$ (это имеет место в примере 5.1). В других случаях (что особенно характерно для задач интерполяции) многочлены $a_j(t)$ можно выбрать достаточно произвольно; их следует взять такими, чтобы можно было упростить дальнейшие вычисления. Наиболее просто решается задача вычисления о. н. к. β при диагональной матрице $A = ZZ'$. В этом случае [см. (5.25)]

$$\beta_j = \sum_{i=1}^n a_j(t_i) \cdot X_i / \sum_{i=1}^n a_j^2(t_i), \quad j = 1, \dots, k. \quad (5.30)$$

Условием диагональности матрицы A является ортогональность векторов z_j :

$$z_j z_r = \sum_{i=1}^n a_j(t_i) a_r(t_i) = 0, \quad j \neq r. \quad (5.31)$$

Многочлены, удовлетворяющие условиям (5.31), называют *ортогональными многочленами Чебышева*.

Покажем, что при заданных $t_1, \dots, t_n$ такие многочлены всегда можно построить. Без потери общности будем предполагать старший коэффициент $a_j(t)$ равным единице, т. е. $a_1(t) \equiv 1$, $a_2(t) = t + c$, ... Из условия (5.31) при $j=1$, $r=2$ имеем

$$0 = \sum_{i=1}^n a_2(t_i) = \sum_{i=1}^n t_i + nc,$$

т. е. $c = -t$. Таким образом, $a_2(t) = t - t$.

Выведем рекуррентную формулу, позволяющую вычислять следующий многочлен по двум предыдущим — этим и доказывается существование всех многочленов $a_j(t)$. Предположим, что для некоторого $\tau \geq 2$ многочлены $a_1(t), \dots, a_\tau(t)$ уже построены. Рассмотрим произвольный многочлен $\psi(t)$ степени $< \tau$. Его можно однозначно выразить в виде линейной комбинации многочленов

$$a_1(t), \dots, a_\tau(t) : \psi(t) = \sum_{s=1}^{\tau} \alpha_s a_s(t). \text{ Действительно, из условия ортогональности (5.31) при } j=1, \dots, \tau \text{ имеем}$$

$$\sum_{i=1}^n \psi(t_i) a_j(t_i) = \alpha_j \sum_{i=1}^n a_j^2(t_i).$$

т. е. эти равенства однозначно определяют коэффициенты $\alpha_1, \dots, \alpha_\tau$. Отсюда, в частности, следует, что если $\psi(t)$ — многочлен степени $< \tau - 1$, то

$$\sum_{i=1}^n \psi(t_i) a_\tau(t_i) = 0. \quad (5.32)$$

Рассмотрим теперь многочлен степени τ следующего вида:

$$a(t) = (t + \alpha) a_{\tau}(t) + \beta a_{\tau-1}(t) \quad (5.33)$$

(старший коэффициент здесь равен 1). Если $j < \tau - 1$, то в силу (5.32)

$$\sum_{i=1}^n a_j(t_i) a(t_i) = \sum_{i=1}^n (t_i + \alpha) a_j(t_i) a_{\tau}(t_i) + \beta \sum_{i=1}^n a_j(t_i) a_{\tau-1}(t_i) = 0$$

(так как степень многочлена $(t + \alpha) a_j(t)$ равна $j < \tau - 1$), т. е. при любых постоянных α и β многочлен (5.33) ортогонален всем многочленам $a_1(t), \dots, a_{\tau-2}(t)$. Выберем эти постоянные так, чтобы многочлен (5.33) был ортогонален также $a_{\tau-1}(t)$ и $a_{\tau}(t)$. Для этого должны выполняться следующие условия:

$$\sum_{i=1}^n a_{\tau-1}(t_i) a(t_i) = \sum_{i=1}^n t_i a_{\tau-1}(t_i) a_{\tau}(t_i) + \beta \sum_{i=1}^n a_{\tau-1}^2(t_i) = 0,$$

$$\sum_{i=1}^n a_{\tau}(t_i) a(t_i) = \sum_{i=1}^n t_i a_{\tau}^2(t_i) + \alpha \sum_{i=1}^n a_{\tau}^2(t_i) = 0.$$

Отсюда

$$\alpha = - \frac{\sum_{i=1}^n t_i a_{\tau}^2(t_i)}{\sum_{i=1}^n a_{\tau}^2(t_i)}, \quad \beta = - \frac{\sum_{i=1}^n t_i a_{\tau-1}(t_i) a_{\tau}(t_i)}{\sum_{i=1}^n a_{\tau-1}^2(t_i)}. \quad (5.34)$$

При таком выборе постоянных α и β многочлен (5.33) ортогонален всем многочленам $a_1(t), \dots, a_{\tau}(t)$ и, следовательно, является следующим многочленом $a_{\tau+1}(t)$ системы Чебышева. Таким образом, приведен алгоритм построения системы ортогональных многочленов Чебышева. Первые два многочлена этой системы указаны выше; в частности, из (5.33) и (5.34) имеем общий вид третьего многочлена

$$a_3(t) = (t - t) \left(t - t - \frac{s_2(t)}{s_0(t)} \right) - s_2(t), \quad (5.35)$$

$$\text{где } t = (t_1, \dots, t_n), \quad s_k(t) = \frac{1}{n} \sum_{i=1}^n (t_i - t)^k.$$

Итак, если в схеме параболической регрессии (5.29) многочлены $a_j(t)$ являются ортогональными многочленами Чебышева, то о. н. к. $\hat{\beta}_j$ вычисляются по формулам (5.30), а соответствующее значение $S(\hat{\beta}) = \min_{\beta} S(\beta)$ таково:

$$\begin{aligned} S(\hat{\beta}) &= \sum_{i=1}^n \left(X_i - \sum_{j=1}^k \hat{\beta}_j a_j(t_i) \right)^2 = \sum_{i=1}^n X_i^2 - 2 \sum_{j=1}^k \hat{\beta}_j \sum_{i=1}^n a_j(t_i) X_i + \\ &+ \sum_{j=1}^k \hat{\beta}_j^2 \sum_{i=1}^n a_j^2(t_i) = \sum_{i=1}^n X_i^2 - \sum_{j=1}^k a_j^2 \hat{\beta}_j^2, \end{aligned} \quad (5.36)$$

$$\text{где } a_j^2 = \sum_{i=1}^n a_j^2(t_i), \quad j = 1, \dots, k.$$

Отметим следующее важное обстоятельство. Как видно из формулы (5.30), о. н. к. $\hat{\beta}_j$ определяется только многочленом $a_j(t)$ и не зависит от параметра k схемы (5.29). Это позволяет упростить задачу построения интерполяционного многочлена более высокой

степени, если требуется повысить точность интерполяции. Так, предположим, что для заданного k построен интерполяционный многочлен $\varphi(t; \hat{\beta}_1, \dots, \hat{\beta}_k) = \sum_{j=1}^k \hat{\beta}_j a_j(t)$, однако значение $S(\hat{\beta}_1, \dots, \hat{\beta}_k)$

[см. соотношение (5.36)] еще велико, что означает недостаточную точность приближения исследуемой функции многочленом степени $k-1$. Точность интерполяции можно повысить, приближая функцию

многочленом степени k вида $\varphi(t; \beta_1, \dots, \beta_{k+1}) = \sum_{j=1}^{k+1} \beta_j a_j(t)$,

т. е. добавляя следующий, $(k+1)$ -й многочлен системы Чебышева. При нахождении такого оптимального многочлена оценки коэффициентов $\beta_1, \dots, \beta_k$ остаются теми же $\hat{\beta}_1, \dots, \hat{\beta}_k$, необходимо вычислить только $\hat{\beta}_{k+1}$ по формуле (5.30). Далее имеем

$$S(\hat{\beta}_1, \dots, \hat{\beta}_{k+1}) = S(\hat{\beta}_1, \dots, \hat{\beta}_k) - a_{k+1}^2 \hat{\beta}_{k+1}^2, \quad (5.37)$$

и если точность приближения, достигнутая с помощью многочлена k -й степени, недостаточна, то можно подбирать далее многочлен $(k+1)$ -й степени и т. д.

Пример 5.3. В «Основах химии» Д. И. Менделеев приводит следующие данные о количестве (x) азотнатриевой соли NaNO_3 , которое можно растворить в 100 г воды в зависимости от температуры (t):

t_i	0	4	10	15	21	29	36	51	68
x_i	66,7	71,0	76,3	80,6	85,7	92,9	99,4	113,6	125,1

Построим по этим данным приближенную эмпирическую формулу вида $x = c_0 + c_1 t$, описывающую зависимость между рассматриваемыми величинами. Здесь имеет место место линейная зависимость, поэтому это схема простой регрессии, рассмотренная в примере 5.1. Используя формулы (5.27), получаем: $c_0 = 67,5$, $c_1 = 0,87$. Таким образом, искомая приближенная формула имеет вид

$$x = 67,5 + 0,87t. \quad (5.38)$$

При этом сумма квадратов отклонений, вычисленная по формуле (5.38), равна 11,1.

Установим, как повышается точность приближения, если в качестве интерполяционного многочлена использовать квадратичную параболу. Запишем искомый многочлен в виде

$$\varphi(t; \beta) = \beta_1 a_1(t) + \beta_2 a_2(t) + \beta_3 a_3(t),$$

где $a_1(t)$ — многочлены Чебышева [здесь $a_1(t) \equiv 1$, $a_2(t) = t - T = t - 26$, многочлен $a_3(t)$ определен в (5.35)]. Результат (5.38) можно записать в виде $x = 90,1 a_1(t) + 0,87 a_2(t)$; следовательно, $\hat{\beta}_1 = 90,1$; $\hat{\beta}_2 = 0,87$. Таким образом, достаточно вычислить по формуле (5.30) $\hat{\beta}_3$. В данном случае $a_3(t) = (t - 26)(t - 40) = 451,1$, поэтому значения $a_3(t)$ в точках t_i таковы: 588,9; 340,9; 28,9; -176,1; -356,1; -484,1 -491,1; -176,1; 724,9. Отсюда $a_3^2 = \sum_{i=1}^9 a_3^2(t_i) \approx 1,6 \cdot 10^6$,

$$\sum_{i=1}^9 a_3(t_i) x_i \approx -2,2 \cdot 10^5 \text{ и, по формуле (5.30), } \hat{\beta}_3 \approx -10^{-3}; \text{ поправка в (5.37) } a_3^2 \hat{\beta}_3^2 \approx 2,7.$$

Таким образом, учитывая многочлен второй степени, мы незначительно увеличиваем точность аппроксимации данных; в качестве удовлетворительной эмпирической зависимости можно принять формулу (5.38).

§ 5.3. Нормальная регрессия. Интервальное оценивание

1. Модель нормальной регрессии. До сих пор делались предположения только о первых и вторых моментах наблюдений [см. условия (5.1) и (5.2)]. Развитая при этих минимальных предположениях теория наименьших квадратов позволяет получать только точечные оценки для параметров $\beta_1, \dots, \beta_k$ и их линейных комбинаций. Чтобы получать более сильные утверждения (например, оценивать вероятности заданных отклонений о. н. к. от истинных значений рассматриваемых параметров), необходимо сделать дополнительные предположения о виде распределения случайного вектора $X = (X_1, \dots, X_n)$ или, что то же, вектора ошибок $e = (e_1, \dots, e_n)$, определенного в (5.1). Чаше всего задачи регрессионного анализа решают в предположении, что наблюдения подчиняются нормальному закону распределения; в этом случае к условиям (5.1) и (5.2) добавляют третье условие $\mathcal{L}(e) = \mathcal{N}(0, \sigma^2 E_n)$.

Если выполнены условия (5.1) и (5.39) [условие (5.2) вытекает из (5.39)], то говорят о *нормальной регрессии*.

Отметим, что условия (5.1) и (5.39) можно объединить и записать в виде

$$\mathcal{L}(X) = \mathcal{N}(Z'\beta, \sigma^2 E_n). \quad (5.40)$$

2. Оценки максимального правдоподобия параметров нормальной регрессии. Нормальная модель (5.40) определяется $(k+1)$ -мерным параметром $\theta = (\beta_1, \dots, \beta_k, \sigma^2)$, область возможных значений которого представляет собой евклидово полупространство $\Theta = \{\theta : -\infty < \beta_j < \infty, j = 1, \dots, k, \sigma^2 > 0\}$. Если $x = (x_1, \dots, x_n)$ — наблюдавшаяся реализация вектора X , то функция правдоподобия для этой модели имеет вид

$$L(x; \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} S(x; \beta)\right\}, \quad (5.41)$$

где квадратичная форма $S(x; \beta)$ определена в (5.4).

Вычислим оценки параметров θ с помощью метода максимального правдоподобия (см. § 2.4). Из (5.41) имеем, что при любом $\sigma^2 > 0$ максимизация $L(x; \theta)$ по β эквивалентна минимизации по β квадратичной формы $S(x; \beta)$. Таким образом, для нормальной модели оценки наименьших квадратов коэффициентов регрессии $\beta_1, \dots, \beta_k$ совпадают с оценками максимального правдоподобия этих параметров.

Из теоремы 5.2 следует, что о. н. к. являются оптимальными в классе линейных оценок. Для схемы нормальной регрессии справедливо более силь-

ное утверждение: о. н. к. являются оптимальными в классе всех (не только линейных) несмещенных оценок рассматриваемых параметрических функций.

Найдем оценку максимального правдоподобия $\hat{\sigma}^2$ для остаточной дисперсии σ^2 . Подставив в (5.41) вместо β оценку $\hat{\beta}$ и прологарифмировав, находим, что $\hat{\sigma}^2$ — это то значение σ^2 , которое минимизирует выражение $S(\hat{\beta})/\sigma^2 + n \ln \sigma^2$. Отсюда имеем

$$\hat{\sigma}^2 = S(\hat{\beta})/n. \quad (5.42)$$

Но, как было показано [см. (5.14)], в общем случае несмещенная оценка для σ^2 имеет вид $S(\hat{\beta})/(n-k)$, поэтому о. м. п. $\hat{\sigma}^2$ оказывается смещенной и ее смещение

$$E\hat{\sigma}^2 - \sigma^2 = E\left(\frac{n-k}{n} \sigma^2\right) - \sigma^2 = \left(\frac{n-k}{n} - 1\right) \sigma^2 = -\frac{k}{n} \sigma^2$$

убывает с ростом числа наблюдений n . Таким образом, оценка (5.42) является асимптотически несмещенной (что характерно для о. м. п.).

3. Основная теорема. Докажем следующую важную теорему теории нормальной регрессии, которая будет неоднократно использоваться в дальнейшем.

Теорема 5.4. *Случайные величины $\hat{\beta}$ и $S(\hat{\beta})$, а также $S(\hat{\beta})$ и $Q = S(\hat{\beta}) - S(\beta)$ независимы; при этом*

$$\begin{aligned} \mathcal{L}_0(\hat{\beta}) &= \mathcal{N}(\beta, \sigma^2 A^{-1}), \quad \mathcal{L}_0(S(\hat{\beta})/\sigma^2) = \\ &= \chi^2(n-k), \quad \mathcal{L}_0(Q/\sigma^2) = \chi^2(k). \end{aligned} \quad (5.43)$$

□ Введем нормированный вектор ошибок $e^* = e/\sigma$; $\mathcal{L}(e^*) = \mathcal{N}(0, E_n)$; тогда равенство (5.1) принимает вид

$$X = Z\beta + \sigma e^*, \quad (5.44)$$

а равенство (5.6) — следующий вид:

$$\hat{\beta} = \beta + \sigma A^{-1} Z' e^*. \quad (5.45)$$

Подставив вместо X его выражение (5.44) в (5.16), получим

$$S(\hat{\beta})/\sigma^2 = e^{*'} B e^*, \quad (5.46)$$

поскольку $ZB = BZ' = 0$ [матрица B определена в (5.15)]. Наконец, из тождества (5.8) имеем следующее представление для Q :

$$Q = S(\hat{\beta}) - S(\beta) = (\hat{\beta} - \beta)' A (\hat{\beta} - \beta). \quad (5.47)$$

Независимость $\hat{\beta}$ и $S(\hat{\beta})$ следует из представлений (5.45) и (5.46), легко устанавливаемого факта $A^{-1}ZB = 0$ и леммы 1.2.

Далее, из представления (5.47) имеем, что случайная величина Q зависит от выборки X лишь через $\hat{\beta}$; следовательно, в силу независимости $\hat{\beta}$ и $S(\hat{\beta})$ случайные величины Q и $S(\hat{\beta})$ также независимы. Первая часть теоремы доказана.

Докажем теперь утверждения (5.43) о распределениях рассматриваемых случайных величин. Так как $\hat{\beta}$ — линейная функция

нормального вектора [см. представление (5.45)], то закон распределения $\hat{\beta}$ также нормален. Первые и вторые моменты вектора $\hat{\beta}$ вычислены в п. 2 § 5.2.

Закон распределения Q/σ^2 устанавливается на основании представления (5.47), факта нормальности вектора $\hat{\beta}$ и теоремы 1.9. Наконец, закон распределения $S(\hat{\beta})/\sigma^2$ устанавливается на основании представления (5.46) и леммы 1.4, поскольку

$$\text{tr } B = \text{tr } E_n - \text{tr } (Z' A^{-1} Z) = n - \text{tr } (Z Z' A^{-1}) = n - \text{tr } E_k = n - k. \blacksquare$$

Следствие 1. Из первого соотношения в (5.43) имеем, что для любого $j=1, \dots, k$

$$\mathcal{L}_j \left(\frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{a/j}} \right) = \mathcal{N}(0, 1) \quad (5.48)$$

и $\hat{\beta}_j$ не зависит от $S(\hat{\beta})$. Поэтому из определения распределения Стюдента (см. п. 4 § 1.5) и из теоремы 5.4 следует, что при любом j

$$\mathcal{L}_j \left(t_j = \sqrt{\frac{n-k}{a/j S(\hat{\beta})}} (\hat{\beta}_j - \beta_j) \right) = S(n-k). \quad (5.49)$$

Следствие 2. Из определения распределения Снедекора (см. п. 5 § 1.5) и теоремы 5.4 имеем, что при любом j

$$\mathcal{L}_j \left(F = \frac{n-k}{k} \frac{Q}{S(\hat{\beta})} \right) = S(k, n-k). \quad (5.50)$$

4. Доверительное оценивание параметров нормальной регрессии. Найдем доверительный интервал для коэффициента регрессии β_j . Из соотношения (5.48) следует, что статистика $\hat{\beta}_j$ имеет распределение $\mathcal{N}(\beta_j, \sigma^2 a/j)$. Следовательно, имеет место задача оценивания неизвестного среднего нормального закона с неизвестной дисперсией (так как σ^2 неизвестно) по наблюдению над случайной величиной $\hat{\beta}_j$. На основании (5.49) стьюдентово отношение t_j является центральной статистикой для оценивания β_j (см. п. 2 § 2.6), поэтому γ -доверительный интервал для β_j строится по схеме примера 2.30 и имеет вид

$$\left(\hat{\beta}_j \pm t_{(1+\gamma)/2, n-k} \sqrt{\frac{a/j}{n-k} S(\hat{\beta})} \right). \quad (5.51)$$

Это симметричный интервал относительно точки $\hat{\beta}_j$ длины

$$2t_{(1+\gamma)/2, n-k} \sqrt{\frac{a/j}{n-k} S(\hat{\beta})}.$$

Чтобы построить доверительный интервал для остаточной дисперсии σ^2 , воспользуемся вторым из соотношений (5.43), из которого следует, что $S(\hat{\beta})/\sigma^2$ — нужная центральная статистика. Искомый γ -доверительный интервал строится здесь по схеме при-

мера 2.30 и имеет вид

$$S(\hat{\beta})/\chi^2_{(1+\gamma)/2, n-k} < \sigma^2 < S(\hat{\beta})/\chi^2_{(1-\gamma)/2, n-k}. \quad (5.52)$$

Итак, можно построить доверительный интервал для каждого из коэффициентов регрессии $\beta_1, \dots, \beta_k$. Если построить k таких интервалов с одним и тем же уровнем γ , то среднее значение числа интервалов, накрывающих соответствующие значения β_j , равно $k\gamma$. Оценим вероятность одновременного накрытия построенными интервалами соответствующих параметров.

Обозначим через A_j событие, состоящее в том, что интервал (5.51), в котором γ заменено на γ_j , накрывает параметр β_j , т. е. $P_1(A_j) = \gamma_j$, $j = 1, \dots, k$. Тогда вероятность совместного осуществления событий $A_1, \dots, A_k$ можно записать так:

$$P_1(A_1 \dots A_k) = 1 - P_1(\bar{A}_1 \cup \dots \cup \bar{A}_k).$$

Но

$$P_1(\bar{A}_1 \cup \dots \cup \bar{A}_k) \leq \sum_{j=1}^k P_1(\bar{A}_j) = \sum_{j=1}^k (1 - \gamma_j),$$

поэтому

$$P_1(A_1 \dots A_k) \geq 1 - \sum_{j=1}^k (1 - \gamma_j) = \sum_{j=1}^k \gamma_j - k + 1. \quad (5.53)$$

Если выбрать все γ_j равными $1 - (1 - \gamma)/k$ при некотором γ , то

$$P_1(A_1 \dots A_k) \geq \gamma. \quad (5.54)$$

Формулы (5.53) и (5.54) позволяют оценить вероятность правильного решения (одновременное накрытие доверительными интервалами всех параметров), однако желательно иметь более точный результат, т. е. уметь строить (случайную) доверительную область G_γ в евклидовом пространстве R^k , накрывающую неизвестную параметрическую точку $\beta = (\beta_1, \dots, \beta_k)$ с вероятностью γ . Такую область в данном случае можно построить, основываясь на результате (5.50).

Действительно, если $F_{\gamma, k, n-k}$ есть γ -квантиль распределения $S(k, n-k)$, то из (5.47) и (5.50) следует, что при любом θ

$$\gamma = P_1(F < F_{\gamma, k, n-k}) = P_1(\beta \in G_\gamma(X)), \quad \text{где} \quad (5.55)$$

$$G_\gamma(X) = \left\{ \beta : (\hat{\beta} - \beta)' A (\hat{\beta} - \beta) < \frac{k}{n-k} S(\hat{\beta}) F_{\gamma, k, n-k} \right\}. \quad (5.56)$$

Тем самым построена искомая γ -доверительная область для β . Это внутренность эллипсоида с центром в точке $\hat{\beta}$, граница которого задается уравнением

$$(\hat{\beta} - \beta)' A (\hat{\beta} - \beta) = a_\gamma^2(X) = \frac{k}{n-k} S(\hat{\beta}) F_{\gamma, k, n-k}. \quad (5.57)$$

Пример 5.4 (простая регрессия, доверительное оценивание параметров). Найдем доверительный интервал уровня γ для наклона β_1

простой регрессии (см. пример 5.1). Из (5.51) и (5.27) — (5.28) следует, что это интервал

$$\left(\hat{\beta}_2 \pm t_{(1+\gamma)/2, n-2} \sqrt{S(\hat{\beta}) / \left| (n-2) \sum_{i=1}^n (t_i - \bar{t})^2 \right|} \right).$$

Аналогично, из (5.55) — (5.57) и формул примера 5.1 имеем, что внутренность эллипса

$$\begin{aligned} (\beta_1 - \hat{\beta}_1)^2 + 2t(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + \frac{1}{n} \sum_{i=1}^n t_i^2 (\beta_2 - \hat{\beta}_2)^2 = \\ = \frac{2}{n(n-2)} S(\hat{\beta}) F_{\gamma, 2, n-2} \end{aligned}$$

в плоскости переменных (β_1, β_2) с вероятностью γ содержит неизвестную точку β .

5. Доверительная область для линейных комбинаций параметров $\beta_1, \dots, \beta_k$. Пусть требуется оценить одновременно $m \leq k$ линейных комбинаций $t = (t_1, \dots, t_m) : t = T\beta$, где T — заданная матрица размером $m \times k$ и $\text{rang } T = m$. Тогда из теоремы 5.4 следует, что о. н. к. $\hat{t} = T\hat{\beta}$ имеет распределение $\mathcal{N}(t, \sigma^2 D)$, где матрица D определена в (5.10). Отсюда по теореме 1.9 имеем

$$\mathcal{L}_0(Q_T/\sigma^2) = \chi^2(m), \text{ где}$$

$$Q_T = Q_T(X, t) = (\hat{t} - t)' D^{-1} (\hat{t} - t) \quad (5.58)$$

При этом Q_T как функция $\hat{\beta}$ не зависит от $S(\hat{\beta})$. Следовательно, отношение Снедекора F в данном случае имеет вид

$$F = \frac{n-k}{m} \frac{Q_T}{S(\hat{\beta})},$$

и при этом $\mathcal{L}_0(F) = S(m, n-k)$. Таким образом, как и в случае оценивания β , получаем следующий γ -доверительный эллипсоид для t :

$$G_{T\gamma}(X) = \left\{ t : (T\hat{\beta} - t)' D^{-1} (T\hat{\beta} - t) < \frac{m}{n-k} S(\hat{\beta}) F_{\gamma, m, n-k} \right\}. \quad (5.59)$$

При $T = E_k$ полученное решение сводится к (5.56).

Отметим некоторые частные случаи общего решения (5.59).

При $m=1$ речь идет об оценивании одной линейной комбинации

$\lambda' \beta = \sum_{j=1}^k \lambda_j \beta_j$. В этом случае эллипсоид (5.59) вырождается в интервал

$$\left(\lambda' \hat{\beta} \pm \left[\frac{1}{n-k} S(\hat{\beta}) F_{\gamma, 1, n-k} (\lambda' A^{-1} \lambda) \right]^{1/2} \right). \quad (5.60)$$

Распределение Снедекора $S(1, n-k)$ совпадает с распределением квадрата случайной величины, имеющей распределение Стьюдента $S(n-k)$, поэтому $F_{\gamma, 1, n-k} = t_{(1+\gamma)/2, n-k}^2$ и выражение (5.60) можно

записать в виде

$$\left(\lambda' \hat{\beta} \pm t_{(1+\gamma)/2, n-k} \sqrt{\frac{1}{n-k} S(\hat{\beta}) (\lambda' A^{-1} \lambda)} \right). \quad (5.61)$$

Пусть матрица T имеет следующую структуру: в каждой строке только один элемент отличен от нуля, при этом в r -й строке на месте j_r стоит единица, $r = 1, \dots, m$, $j_1 < j_2 < \dots < j_m$. Тогда $T\beta = (\beta_{j_1}, \dots, \beta_{j_m}) = \beta(m)$ и, следовательно, речь идет об одновременном оценивании части координат параметрического вектора β . В этом случае матрица D [см. (5.10)] представляет собой минор $A^{-1}(j_1, \dots, j_m) = A(m)$ матрицы A^{-1} , получающийся вычеркиванием всех строк и столбцов с номерами, отличными от $j_1, \dots, j_m$, и из (5.59) получаем, что γ -доверительный эллипсоид для параметров $\beta(m)$ имеет вид

$$G_{T\gamma}(X) = \left\{ \beta(m) : (\hat{\beta}(m) - \beta(m))' A^{-1}(m) (\hat{\beta}(m) - \beta(m)) < \frac{m}{n-k} S(\hat{\beta}) F_{\gamma, m, n-k} \right\}. \quad (5.62)$$

Пример 5.5 (доверительный интервал для ординаты линии регрессии). Построим доверительный интервал для ординаты $\varphi(t) = \varphi(t; \beta) = \beta_1 + \beta_2 t$ линии регрессии в произвольной точке t в случае простой регрессии (см. пример 5.1). Используя обозна-

чение $s^2(t) = \frac{1}{n} \sum_{i=1}^n (t_i - t)^2$, из формул примера 5.1 получаем, что при $\lambda = (1, t)$

$$\lambda' A^{-1} \lambda = \frac{1}{n} \left[1 + \left(\frac{t - \bar{t}}{s(t)} \right)^2 \right].$$

Кроме того, $\varphi(t; \hat{\beta}) = \lambda' \hat{\beta} = X + (t - \bar{t}) \hat{\beta}_2$. Отсюда и из (5.61) окончательно имеем, что искомым γ -доверительный интервал имеет вид

$$\left(X + (t - \bar{t}) \hat{\beta}_2 \pm t_{(1+\gamma)/2, n-2} \sqrt{\frac{1}{n(n-2)} S(\hat{\beta}) \left[1 + \left(\frac{t - \bar{t}}{s(t)} \right)^2 \right]} \right).$$

6. Совместные доверительные интервалы. В каждом конкретном случае построение эллипсоидов в (5.56) и (5.59) — трудная вычислительная задача. Поэтому предпочтительнее иметь [более точный, чем (5.54)] результат, позволяющий строить доверительные интервалы для отдельных компонент оцениваемого вектора, с заданной вероятностью одновременно накрывающие соответствующие координаты, или, другими словами, систему совместных доверительных интервалов. Решение этой задачи нетрудно получить из уже известных результатов. Для этого понадобится следующее утверждение из теории квадратичных форм.

Теорема 5.5. Пусть B — положительно определенная матрица и t, λ — вектор-столбцы. Тогда

$$t' B t = \max_{\lambda} \frac{(\lambda' t)^2}{\lambda' B^{-1} \lambda}. \quad (5.63)$$

□ Представим B в виде $B = HH'$. Это всегда можно сделать, выбрав $H = UA^{1/2}$, где U — ортогональная матрица, приводящая B к диагональному виду A . Положим теперь $X = H't$, $Y = H^{-1}\lambda$; тогда $X'Y = t'\lambda$, $X'X = t'Bt$, $Y'Y = \lambda'B^{-1}\lambda$. Но, по неравенству Коши — Буняковского, $(X'Y)^2 \leq (X'X)(Y'Y)$, причем знак равенства имеет место только когда векторы X и Y линейно зависимы. Отсюда и из предыдущих соотношений имеем $(\lambda't)^2 \leq (t'Bt)(\lambda'B^{-1}\lambda)$, или $t'Bt \geq (\lambda't)^2/(\lambda'B^{-1}\lambda)$. Отсюда следует (5.63). ■

Положим теперь в соотношении (5.63) $t = \hat{\beta} - \beta$, $B = A$; тогда соотношения (5.55) — (5.57) можно записать в виде

$$\gamma = P_0 \left(\max_{\lambda} \frac{|\lambda'(\hat{\beta} - \beta)|}{\sqrt{\lambda'A^{-1}\lambda}} < a_{\gamma}(X) \right) = P_0 (|\lambda'(\hat{\beta} - \beta)| < u_{\gamma}(X; \lambda), \forall \lambda),$$

где

$$u_{\gamma}(X; \lambda) = a_{\gamma}(X) \sqrt{\lambda'A^{-1}\lambda}. \quad (5.64)$$

Таким образом, для любого γ выполняется соотношение

$$P_0(\lambda'\hat{\beta} - u_{\gamma}(X; \lambda) < \lambda'\beta < \lambda'\hat{\beta} + u_{\gamma}(X; \lambda), \forall \lambda) = \gamma. \quad (5.65)$$

Соотношения (5.64) — (5.65) позволяют решить задачу о построении системы совместных доверительных интервалов для всех линейных функций $\lambda'\beta$. Если требуется построить систему совместных доверительных интервалов для конечного числа заданных линейных комбинаций $\lambda'_1\beta, \dots, \lambda'_m\beta$, то из (5.65) имеем следующий результат:

$$P_0(\lambda'_r\hat{\beta} - u_{\gamma}(X; \lambda_r) < \lambda'_r\beta < \lambda'_r\hat{\beta} + u_{\gamma}(X; \lambda_r), r = 1, \dots, m) \geq \gamma. \quad (5.66)$$

Рассмотрим несколько примеров использования соотношения (5.66).

Пример 5.6 (совместные доверительные интервалы для коэффициентов регрессии). Пусть $1 \leq m \leq k$ и $\lambda_r = (0 \dots 0 1 0 \dots 0)$, где 1 стоит на месте j_r , $r = 1, \dots, m$. Тогда $\lambda'_r\beta = \beta_{j_r}$, $\lambda'_rA^{-1}\lambda_r = a^{j_r j_r}$ и на основании (5.66) $\{(\hat{\beta}_{j_r} \pm a_{\gamma}(X) \sqrt{a^{j_r j_r}}), r = 1, \dots, m\}$ — система совместных доверительных интервалов уровня $\geq \gamma$ для координат $\beta_{j_1}, \dots, \beta_{j_m}$. В частности, при $m = k$ получаем систему совместных доверительных интервалов для всех координат $\beta_1, \dots, \beta_k$.

Пример 5.7 (оценивание средних). Положим $\lambda_i = z^{(i)}$, $i = 1, \dots, n$. Тогда с помощью формулы (5.66) можно найти систему совместных доверительных интервалов уровня $\geq \gamma$ для средних значений всех наблюдений $X_1, \dots, X_n$.

§ 5.4. Общая линейная гипотеза нормальной регрессии

1. Понятие линейной гипотезы. При применении схемы нормальной регрессии на практике часто возникает необходимость проверить те или иные гипотезы о значениях коэффициентов $\beta_1, \dots, \beta_k$. В общем виде задача формулируется

следующим образом: по наблюдениям $X = (X_1, \dots, X_n)$ требуется проверить гипотезу H_0 , согласно которой коэффициенты регрессии $\beta_1, \dots, \beta_k$ удовлетворяют некоторым заданным ограничениям, локализующим их допустимые значения в некотором подмножестве $\mathcal{B}_0 \subset R^k$. Если эти ограничения линейные, то говорят о *линейной гипотезе*. Для линейной гипотезы подмножество $\mathcal{B}_0$ является линейным подпространством вида

$$\mathcal{B}_0 = \{\beta: T\beta = t_0\}, \quad (5.67)$$

где T — заданная матрица коэффициентов ограничений, t_0 — вектор значений соответствующих комбинаций (предполагается, что система $T\beta = t_0$ совместна). В дальнейшем будем считать, что матрица T имеет размер $m \times k$ ($m \leq k$) и $\text{rang } T = m$. Таким образом, в общем случае линейная гипотеза записывается в виде $H_0: \beta \in \mathcal{B}_0$. В любом случае это сложная гипотеза, так как оставляет произвольным значение параметра σ^2 : в терминах общего параметра модели $\theta = (\beta, \sigma^2)$ гипотеза H_0 имеет вид

$$H_0: \theta \in \Theta_0 = \{\theta: \beta \in \mathcal{B}_0, \sigma^2 > 0\}.$$

2. F-критерий проверки линейной гипотезы. Чтобы построить критерий проверки гипотезы H_0 , воспользуемся уже известным решением задачи доверительного оценивания (см. п. 4 — 5, § 5.3) и принципом соответствия между теорией проверки гипотез и построением доверительных областей (см. п. 4 § 4.4). Напомним общую схему этого соответствия.

Пусть имеется система γ -доверительных областей $\{G_\gamma(x), x \in \mathcal{X}\}$ для некоторой параметрической функции $g = g(\theta)$. Тогда подмножество выборочного пространства, задаваемое условием $\mathcal{X}_{0, 1-\gamma} = \{x: g_0 \in G_\gamma(x)\}$, определяет область принятия гипотезы $H_0: g = g_0$ с уровнем значимости $\alpha = 1 - \gamma$.

В данном случае согласно (5.58) — (5.59) γ -доверительная область для $t = T\beta$ имеет вид

$$G_{T\gamma}(X) = \left\{ t: Q_T(X, t)/S(\hat{\beta}(X)) < \frac{m}{n-k} F_{\gamma, m, n-k} \right\},$$

поэтому область принятия гипотезы $H_0: t = t_0$ с уровнем значимости $\alpha = 1 - \gamma$ задается условием

$$\mathcal{X}_{0\alpha} = \left\{ x: Q_T(x, t_0)/S(\hat{\beta}(x)) < \frac{m}{n-k} F_{\gamma, m, n-k} \right\}.$$

Окончательно можно сформулировать следующий результат: *критерий уровня значимости α для проверки гипотезы $H_0: \beta \in \mathcal{B}_0$, где $\mathcal{B}_0$ определено в (5.67), задается критической областью*

$$\mathcal{X}_{1\alpha} = \left\{ x: \frac{n-k}{m} \frac{Q_T(x, t_0)}{S(\hat{\beta}(x))} \geq F_{1-\alpha, m, n-k} \right\}. \quad (5.68)$$

Статистикой этого критерия является величина

$$F = \frac{n-k}{m} \frac{Q_T(X, t_0)}{S(\hat{\beta})} = \frac{n-k}{m} (T\hat{\beta} - t_0)' D^{-1} (T\hat{\beta} - t_0)/S(\hat{\beta}), \quad (5.69)$$

большие значения которой в соответствии с (5.68) являются критическими для проверяемой гипотезы. Построенный критерий называют *F-критерием*.

Критерий (5.68) можно получить, исходя из других соображений. Рассмотрим о. н. к. $\hat{t} = T\hat{\beta}$ для $\hat{t}$ и составим вектор отклонений этих оценок от проверяемых значений $T\hat{\beta} - t_0$. Если гипотеза H_0 не верна, то весьма вероятно, что эти отклонения велики. Будем измерять степень отклонения с помощью величины $Q_T(X, t_0)$. Тогда, поскольку математическое ожидание квадратичной формы можно преобразовать следующим образом:

$$\begin{aligned} E(Y'BY) &= \sum_{i,j} b_{ij} E(Y_i Y_j) = \sum_{i,j} b_{ij} \sigma_{ij} + \sum_{i,j} b_{ij} E(Y_i) E(Y_j) = \\ &= \text{tr}(B\Sigma) + E(Y') B E(Y) \end{aligned}$$

[здесь $\Sigma = D(Y)$], учитывая (5.10), имеем

$$\begin{aligned} \frac{1}{m} E Q_T(X, t_0) &= \frac{\sigma^2}{m} \text{tr}(D^{-1}D) + \frac{1}{m} E(T\hat{\beta} - t_0)' D^{-1} E(T\hat{\beta} - t_0) = \\ &= \sigma^2 + \frac{1}{m} (T\hat{\beta} - t_0)' D^{-1} (T\hat{\beta} - t_0) \geq \sigma^2. \end{aligned}$$

знак равенства имеет место только при $T\hat{\beta} = t_0$, т. е. когда верна нулевая гипотеза. С другой стороны, $S(\hat{\beta})/(n-k)$ — несмещенная оценка для σ^2 всегда, т. е. безотносительно к истинному значению β . Поэтому у отношения $F =$

$$= \frac{Q_T(X, t_0)}{m} \bigg/ \frac{S(\hat{\beta})}{n-k}$$

числитель в среднем больше знаменателя, когда гипотеза

H_0 не верна, и поэтому большие значения F естественно рассматривать как свидетельствующие против гипотезы H_0 . Из этих рассуждений также следует, что статистику F можно считать отношением двух независимых и несмещенных (при гипотезе H_0) оценок для остаточной дисперсии σ^2 . С такой интерпретацией и связан термин *дисперсионное отношение*, используемый иногда для статистики F (см. п. 5 § 1.5).

Для вычисления статистики F можно использовать также формулу

$$F = \frac{n-k}{m} \frac{S_T - S(\hat{\beta})}{S(\hat{\beta})}, \quad (5.70)$$

которая следует из (5.23). Таким образом, F непосредственно можно выразить через условный S_T (т. е. при условии $T\hat{\beta} = t_0$) и абсолютный $S(\hat{\beta})$ минимумы исходной квадратичной формы $S(\beta)$. В конкретных задачах условный минимум S_T зачастую находят непосредственно, поэтому представление (5.70) более удобно для практических расчетов, чем (5.69).

Пример 5.8 (*простая регрессия, гипотеза о наклоне линии регрессии*). Требуется проверить гипотезу $H_0: \beta_2 = \beta_{20}$, фиксирующую значение наклона линии регрессии. В данном случае параметр $m = 1$ и $S_T = \min_{\beta_1} \sum_{i=1}^n (X_i - \beta_{20} t_i - \beta_1)^2$. Этот минимум достигается при $\beta_1 = X - \beta_{20} t$, и его значение, учитывая соотношения

(5.27) – (5.28), можно записать в виде

$$S_T = S(\hat{\beta}) + (\hat{\beta}_2 - \beta_{20})^2 \sum_{i=1}^n (t_i - t)^2.$$

По формуле (5.70) имеем, что в данном случае статистика

$$F = (n - 2) (\hat{\beta}_2 - \beta_{20})^2 \sum_{i=1}^n (t_i - t)^2 / S(\hat{\beta}).$$

а критическая граница в (5.68) выбирается равной $F_{1-\alpha, 1, n-2}$. Так как $F_{1-\alpha, 1, n-2} = t_{1-\alpha/2, n-2}^2$, где $t_{p, n}$ – p -квантиль распределения

Стьюдента $S(n)$ [см. замечание к формуле (5.60)], то F -критерий в данном случае сводится к критерию, критическая область которого определяется условием

$$|\hat{\beta}_2 - \beta_{20}| \geq t_{1-\alpha/2, n-2} \sqrt{S(\hat{\beta}) \left/ \left((n-2) \sum_{i=1}^n (t_i - t)^2 \right) \right|},$$

что соответствует результату примера 5.4.

§ 5.5. Применение теории линейной регрессии

Развитая в предыдущем параграфе теория проверки гипотез о коэффициентах регрессии имеет широкое применение. К схеме регрессии могут быть сведены многие гипотезы о математических ожиданиях нормальных совокупностей, которые иногда имеют форму, отличную от рассмотренной выше. Некоторые из таких задач рассматриваются в этом параграфе.

1. Гипотеза о параллельности линий регрессии. Предположим, что имеется $r \geq 2$ групп независимых наблюдений $X_1^{(i)}, \dots, X_{n_i}^{(i)}$, $i = 1, \dots, r$ (n_i – объем i -й группы), распределенных нормально с общей дисперсией σ^2 и средними

$$EX_j^{(i)} = \mu_j^{(i)} = \beta_1^{(i)} + \beta_2^{(i)} t_j^{(i)}, \quad j = 1, \dots, n_i; \quad i = 1, \dots, r. \quad (5.71)$$

Другими словами, имеется r различных схем простой регрессии и проверяемая гипотеза H_0 состоит в том, что все линии регрессии (5.71) имеют один и тот же наклон, т. е. $H_0: \beta_2^{(1)} = \dots = \beta_2^{(r)}$. Такая гипотеза может встретиться, например, при проверке равенства нескольких скоростей роста, при сравнении нескольких способов обработки и т. д.

Покажем, что рассматриваемый случай можно свести к схеме общей линейной гипотезы. Положим $X = (X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_{n_1+n_2}, \dots, X_{n_1+\dots+n_r}) = (X_1^{(1)}, \dots, X_{n_1}^{(1)}, X_1^{(2)}, \dots, X_{n_2}^{(2)}, \dots, X_1^{(r)}, \dots, X_{n_r}^{(r)})$, $n = n_1 + \dots + n_r$, $\beta = (\beta_1, \dots, \beta_{2r}) = (\beta_1^{(1)}, \beta_1^{(1)}, \beta_1^{(2)}, \dots, \beta_2^{(r)})$ и введем матрицу Z раз-

мером $2r \times n$, составленную из вектор-столбцов $z^{(i)}$ вида

$$Z = [z^{(1)} \dots z^{(n_1)} z^{(n_1+1)} \dots z^{(n)}] = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ t_1^{(1)} & \dots & t_{n_1}^{(1)} & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \\ 0 & \dots & 0 & t_1^{(2)} & \dots & t_{n_2}^{(2)} \\ & & & & \ddots & \\ 0 & & & & & 1 & \dots & 1 \\ & & & & & t_1^{(r)} & \dots & t_{n_r}^{(r)} \end{bmatrix}.$$

В этих обозначениях вектор X распределен по закону $N(Z'\beta, \sigma^2 E_n)$. Таким образом, имеем частный случай (при указанном выборе $z^{(i)}$) схемы нормальной регрессии. Гипотезу H_0 можно записать в виде $H_0: \beta_1 - \beta_2 = 0, \dots, \beta_r - \beta_2 = 0$, т. е. в данном случае H_0 задается $(r-1)$ -м линейно независимым соотношением между коэффициентами регрессии и, следовательно, для ее проверки можно применить F -критерий. Для этого надо найти абсолютный $S_1 = S(\beta)$ и условный S_T (т. е. при гипотезе H_0) минимумы по β квадратичной формы

$$S(\beta) = \sum_{i=1}^n (X_i - z^{(i)'}\beta)^2 = \sum_{i,j} (X_j^{(i)} - \mu_j^{(i)})^2. \quad (5.72)$$

Нахождение S_1 сводится здесь к минимизации выражения

$$S(\beta) = \sum_{i=1}^r S^{(i)},$$

$$\text{где } S^{(i)} = \sum_{j=1}^{n_i} (X_j^{(i)} - \beta_1^{(i)} - \beta_2^{(i)} t_j^{(i)})^2, \quad i = 1, \dots, r,$$

т. е. достаточно найти минимум каждой квадратичной формы $S^{(i)}$ отдельно. В других обозначениях эта задача решена в примере 5.1, поэтому из (5.27) и (5.28) имеем, что минимум $S^{(i)}$ достигается при $\beta_1^{(i)} = \hat{\beta}_1^{(i)}$, $\beta_2^{(i)} = \hat{\beta}_2^{(i)}$, где

$$\hat{\beta}_2^{(i)} = \frac{\sum_{j=1}^{n_i} (X_j^{(i)} - \bar{X}^{(i)})(t_j^{(i)} - \bar{t}^{(i)})}{\sum_{j=1}^{n_i} (t_j^{(i)} - \bar{t}^{(i)})^2}, \quad \hat{\beta}_1^{(i)} = \bar{X}^{(i)} - \hat{\beta}_2^{(i)} \bar{t}^{(i)} \quad (5.73)$$

(здесь $\bar{X}^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_j^{(i)}$, $\bar{t}^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} t_j^{(i)}$ и при этом предполагается, что для любого $i = 1, \dots, r$ не все числа $t_1^{(i)}, \dots, t_{n_i}^{(i)}$ одинаковы). Таким образом, в данном случае

$$S_1 = \sum_{i=1}^r \sum_{j=1}^{n_i} [(X_j^{(i)} - \bar{X}^{(i)}) - \hat{\beta}_2^{(i)} (t_j^{(i)} - \bar{t}^{(i)})]^2. \quad (5.74)$$

Введем следующие обозначения:

$$s_i^2(t) = \sum_{j=1}^{n_i} (t_i^{(j)} - \bar{t}^{(i)})^2, \quad s_i^2(X) = \sum_{j=1}^{n_i} (X_i^{(j)} - \bar{X}^{(i)})^2,$$

$$s_i(X, t) = \sum_{j=1}^{n_i} (X_i^{(j)} - \bar{X}^{(i)})(t_i^{(j)} - \bar{t}^{(i)}), \quad i = 1, \dots, r \quad (5.75)$$

[с точностью до множителей — это выборочные дисперсии и ковариация множеств $(X_1^{(i)}, \dots, X_{n_i}^{(i)})$ и $(t_1^{(i)}, \dots, t_{n_i}^{(i)})$]. Тогда из (5.73) и (5.74) следует, что S_1 можно записать в виде

$$S_1 = s_1^2(X) + \dots + s_r^2(X) - s_1^2(X, t)/s_1^2(t) - \dots - s_r^2(X, t)/s_r^2(t). \quad (5.76)$$

Вычислим S_T . Для этого минимизируем сумму $\sum_{i,j} (X_i^{(j)} - \beta_1^{(i)} - \beta_2 t_i^{(j)})^2$.

Непосредственным дифференцированием можно проверить, что при любом фиксировании β_2 минимум достигается при $\beta_1^{(i)} = b_1^{(i)}$, где $b_1^{(i)} = \bar{X}^{(i)} - \beta_2 \bar{t}^{(i)}$, $i = 1, \dots, r$, и сумма квадратов при этом равна $\sum_{i,j} [(\bar{X}^{(i)} - \bar{X}^{(i)}) - \beta_2 (\bar{t}^{(i)} - \bar{t}^{(i)})]^2$. Минимум последнего выражения достигается при $\beta_2 = b_2$, где

$$b_2 = \frac{\sum_{i,j} (X_i^{(j)} - \bar{X}^{(i)}) (\bar{t}^{(i)} - \bar{t}^{(i)})}{\sum_{i,j} (\bar{t}^{(i)} - \bar{t}^{(i)})^2} = \frac{s_1(X, t) + \dots + s_r(X, t)}{s_1^2(t) + \dots + s_r^2(t)}$$

[см. (5.75)], и он равен

$$S_T = s_1^2(X) + \dots + s_r^2(X) - \frac{(s_1(X, t) + \dots + s_r(X, t))^2}{s_1^2(t) + \dots + s_r^2(t)}. \quad (5.77)$$

В данном случае параметры k и m в (5.70) равны соответственно $2r$ и $r-1$, поэтому статистика F в силу соотношений (5.76) и (5.77) равна

$$F = \frac{n-2r}{r-1} \frac{\left(\sum_{i=1}^r s_i^2(t) \right) \left(\sum_{i=1}^r s_i^2(X, t)/s_i^2(t) \right) - \left(\sum_{i=1}^r s_i(X, t) \right)^2}{\left(\sum_{i=1}^r s_i^2(t) \right) \left(\sum_{i=1}^r s_i^2(X) - \sum_{i=1}^r s_i^2(X, t)/s_i^2(t) \right)}.$$

При гипотезе H_0 эта статистика имеет распределение Снедекора $S(r-1, n-2r)$; следовательно, критическая область уровня значимости α для гипотезы H_0 определяется условием $F \geq F_{1-\alpha, r-1, n-2r}$.

2. Критерий однородности. Пусть $X_1^{(i)}, \dots, X_{n_i}^{(i)}$ — независимые наблюдения над совокупностью $\mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \dots, k$. Рассмотрим гипотезу однородности $H_0: \mu_1 = \dots = \mu_k$, т. е. при гипотезе H_0 все наблюдения производятся над одной и той же нормальной совокупностью $\mathcal{N}(\mu, \sigma^2)$ с некоторыми (неизвестными) параметрами. Такая гипотеза возникает, например, при сравнении нескольких различных способов обработки или процедур, условий, размещений и т. п. с целью выяснения, влияют ли различия

между ними на интересующий исследователя исход. Вообще подобная схема имеет место каждый раз, когда интересуются влиянием на исход эксперимента какого-то одного фактора. Если число различных значений этого фактора (число уровней, на которых может находиться фактор) равно k , то n_i — число наблюдений, соответствующих i -му уровню фактора, а $X_j^{(i)}$ при $j = 1, \dots, n_i$ — сами наблюдения (результаты эксперимента). В этом случае говорят об *одинарной классификации исходов*.

Сведем рассматриваемую схему к схеме регрессии. Для этого положим $\mathbf{X} = (X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_n) = (X_1^{(1)}, \dots, X_{n_1}^{(1)}, X_1^{(2)}, \dots, X_{n_2}^{(2)}, \dots, X_{n_k}^{(k)})$, $n = n_1 + \dots + n_k$,

$$Z = \begin{vmatrix} z^{(1)} & \dots & z^{(n_1)} & z^{(n_1+1)} & \dots & z^{(n)} \end{vmatrix} = \begin{vmatrix} 1 & \dots & 1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & \dots & 0 & \dots & 1 & \dots & 1 \end{vmatrix},$$

где в j -й строке матрицы Z на местах $n_1 + \dots + n_{j-1} + 1, \dots, n_1 + \dots + n_j$ стоят единицы, а на остальных местах — нули, $j = 1, \dots, k$ ($n_0 = 0$). Тогда $X_1, \dots, X_n$ независимы и $\mathcal{L}(X_i) = \mathcal{N}(z^{(i)}, \mu, \sigma^2)$, $i = 1, \dots, n$ (здесь $\mu = (\mu_1, \dots, \mu_k)$). Таким образом, имеем частный случай схемы регрессии, для которой требуется проверить гипотезу $H_0: \mu_2 - \mu_1 = 0, \dots, \mu_k - \mu_1 = 0$ (в данном случае H_0 задается $(k-1)$ -м линейно независимым соотношением между коэффициентами регрессии). Следовательно, можно применить теорию F -критерия. Для этого вычислим сначала S_T и S_1 , т. е. минимальные значения квадратичной формы $S = \sum_{i,j} (X_j^{(i)} - \mu_i)^2$

при гипотезе H_0 и без ограничений на параметры μ соответственно. Так как

$$S = \sum_{i,j} (X_j^{(i)} - \bar{X}^{(i)})^2 + \sum_i n_i (\bar{X}^{(i)} - \mu_i)^2, \quad \bar{X}^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_j^{(i)},$$

то при отсутствии ограничений на $\mu_1, \dots, \mu_k$ минимум достигается при $\mu_i = \bar{X}^{(i)}$, $i = 1, \dots, k$, и равен

$$S_1 = \sum_{i,j} (X_j^{(i)} - \bar{X}^{(i)})^2 = \sum_{i=1}^k s_i^2, \quad s_i^2 = \sum_{j=1}^{n_i} (X_j^{(i)} - \bar{X}^{(i)})^2. \quad (5.78)$$

Далее имеем $S_T = \min_{\mu} \sum_{i,j} (X_j^{(i)} - \mu)^2$, но $\sum_{i,j} (X_j^{(i)} - \mu)^2 = \sum_{i,j} (X_j^{(i)} - \bar{X})^2 + n(\bar{X} - \mu)^2$, $\bar{X} = \frac{1}{n} \sum_{i,j} X_j^{(i)}$, поэтому

$$S_T = \sum_{i,j} (X_j^{(i)} - \bar{X})^2 = S_1 + \sum_{i=1}^k n_i (\bar{X}^{(i)} - \bar{X})^2 \quad (5.79)$$

и минимум достигается при $\mu = \bar{X}$.

Таким образом, в данном случае

$$F = \frac{n-k}{k-1} \frac{\sum_{i=1}^k n_i (\bar{X}^{(i)} - \bar{X})^2}{\sum_{i=1}^k s_i^2}$$

и при гипотезе H_0 эта статистика распределена по закону Снедекора $S(k-1, n-k)$. Отсюда следует, что критическая область уровня значимости α для гипотезы H_0 определяется условием $F \geq F_{1-\alpha, k-1, n-k}$.

Отметим, что соотношение (5.79) можно интерпретировать как разложение полной суммы квадратов отклонений наблюдений от общего среднего («полной изменчивости») на сумму квадратов отклонений каждой величины от соответствующего группового среднего значения («изменчивость внутри групп») и сумму квадратов отклонений групповых средних от общего среднего («изменчивость между группами»). При этом так как при гипотезе H_0

$$ES_T = E \sum_{i, l} (X_l^{(i)} - \bar{X})^2 = (n-1) \sigma^2,$$

$$ES_1 = \sum_{i=1}^k E \sum_{l=1}^{n_i} (X_l^{(i)} - \bar{X}^{(i)})^2 = \sum_{i=1}^k (n_i - 1) \sigma^2 = (n-k) \sigma^2,$$

$$E \sum_{i=1}^k n_i (\bar{X}^{(i)} - \bar{X})^2 = ES_T - ES_1 = (k-1) \sigma^2,$$

то каждая из величин $Q = \frac{1}{n-1} \sum_{i, l} (X_l^{(i)} - \bar{X})^2$, $Q_1 = \frac{1}{k-1} \sum_i n_i (\bar{X}^{(i)} - \bar{X})^2$, $Q_2 = \frac{1}{n-k} \sum_{i, l} (X_l^{(i)} - \bar{X}^{(i)})^2$ при нулевой ги-

потезе определяет несмещенную оценку для дисперсии σ^2 и критерий F можно считать критерием совместности независимых оценок Q_1 и Q_2 (статистику F можно записать в виде $F = Q_1/Q_2$). Такое разложение полной изменчивости исходных данных и его интерпретация характерны для *дисперсионного анализа* — раздела математической статистики, объединяющего совокупность статистических приемов, широко применяемых в самых разнообразных экспериментах, в которых наблюдения тем или иным способом классифицируются (разбиваются на группы) в соответствии со значениями некоторых факторов.

3. Двойная классификация. Рассмотренный в предыдущем пункте случай является простейшей схемой дисперсионного анализа (один фактор), в которой решение получают с помощью методов регрессионного анализа. Рассмотрим применение регрессионного анализа еще к одной имеющей большое практическое значение схеме дисперсионного анализа, когда число факторов, влияющих на исход эксперимента, равно двум.

Предположим, что исход эксперимента зависит от значений двух

факторов R и C , причем фактор R может находиться на r уровнях $R_1, \dots, R_r$, а фактор C — на s уровнях $C_1, \dots, C_s$. Пусть, далее, при каждой возможной комбинации уровней обоих факторов производится ровно одно наблюдение. Обозначим через X_{ij} ($i=1, \dots, r$; $j=1, \dots, s$) результат наблюдения, соответствующего случаю, когда фактор R находится на уровне R_i , а фактор C — на уровне C_j . Предполагаем, что наблюдения X_{ij} независимы и распределены нормально с одинаковыми дисперсиями σ^2 и средними значениями

$$EX_{ij} = \xi_{ij} = \mu + \alpha_i + \beta_j, \quad \text{где} \quad (5.80)$$

$$\sum_{i=1}^r \alpha_i = \sum_{j=1}^s \beta_j = 0. \quad (5.81)$$

Условия (5.80) и (5.81) означают, что оба фактора действуют независимо (в этом случае говорят, что они *аддитивны*). Действительно, если обозначать через α' и β' действия факторов R и C соответственно, то аддитивность означает, что $\xi_{ij} = \alpha'_i + \beta'_j$. Если положить теперь $\mu = \alpha' + \beta'$, $\alpha_i = \alpha'_i - \alpha'$ и $\beta_j = \beta'_j - \beta'$ (черта сверху означает среднее арифметическое), то получим соотношения (5.80) — (5.81).

Такую схему удобно представлять в виде прямоугольной таблицы из r строк, соответствующих уровням $R_1, \dots, R_r$, и s столбцов, соответствующих уровням $C_1, \dots, C_s$. Случайную величину X_{ij} называют *реагирующей*: она описывает *реакцию*, вызываемую комбинацией (R_i, C_j) факторов R и C .

Из (5.80) получаем, что EX_{ij} для любых i и j оказывается при условии (5.81) линейной функцией от $r+s-1$ неизвестных коэффициентов μ, α_i, β_j . Таким образом, здесь также имеем частный случай схемы линейной регрессии и, следовательно, можно применить развитую выше теорию.

Как известно, все положения теории нормальной регрессии основываются на анализе квадратичной формы

$$S = \sum_{i=1}^r \sum_{j=1}^s (X_{ij} - \xi_{ij})^2. \quad (5.82)$$

В данном случае удобно предварительно преобразовать эту форму следующим образом. Обозначим через $X_{..} = \frac{1}{rs} \sum_{i,j} X_{ij}$, $X_{i.} = \frac{1}{s} \sum_{j=1}^s X_{ij}$,

$$X_{.j} = \frac{1}{r} \sum_{i=1}^r X_{ij}. \quad \text{Тогда, учитывая (5.80), имеем}$$

$$\begin{aligned} S &= \sum_{i,j} (X_{ij} - \mu - \alpha_i - \beta_j)^2 = \sum_{i,j} [(X_{ij} - X_{i.} - X_{.j} + X_{..}) + \\ &+ (X_{i.} - X_{..} - \alpha_i) + (X_{.j} - X_{..} - \beta_j) + (X_{..} - \mu)]^2 = \\ &= \sum_{i,j} (X_{ij} - X_{i.} - X_{.j} + X_{..})^2 + s \sum_i (X_{i.} - X_{..} - \alpha_i)^2 + \\ &+ r \sum_j (X_{.j} - X_{..} - \beta_j)^2 + rs (X_{..} - \mu)^2, \end{aligned} \quad (5.83)$$

так как все попарные произведения после суммирования исчезают.

Оценки наименьших квадратов для коэффициентов регрессии — это значения параметров, которые обращают квадратичную форму (5.82) в минимум, поэтому из разложения (5.83) следует, что в данном случае

$$\hat{\mu} = X_{..}, \quad \hat{\alpha}_i = X_{i.} - X_{..}, \quad \hat{\beta}_j = X_{.j} - X_{..} \quad (5.84)$$

Из (5.83) и (5.84) также имеем, что минимальное значение формы S равно

$$S_1 = \sum_{i,j} (X_{ij} - \xi_{ij})^2 = \sum_{i,j} (X_{ij} - X_{i.} - X_{.j} + X_{..})^2 \quad (5.85)$$

Найдем дисперсии полученных оценок. Так как $X_{..} = \frac{1}{r} \sum_{i=1}^r X_{i.} =$

$$= \frac{1}{r} \sum_{i=1}^r X_{i.}, \quad \text{то}$$

$$\hat{\alpha}_i = \frac{r-1}{r} X_{i.} - \frac{1}{r} \sum_{l \neq i} X_{l.}, \quad \hat{\beta}_j = \frac{s-1}{s} X_{.j} - \frac{1}{s} \sum_{l \neq j} X_{.l}.$$

Но $X_{i.}$, $i=1, \dots, r$, независимы и $DX_{i.} = \sigma^2/s$; аналогично, $X_{.j}$, $j=1, \dots, s$, независимы и $DX_{.j} = \sigma^2/r$; следовательно,

$$D\hat{\alpha}_i = \frac{r-1}{rs} \sigma^2, \quad D\hat{\beta}_j = \frac{s-1}{rs} \sigma^2, \quad D\hat{\mu} = \frac{1}{rs} \sigma^2. \quad (5.86)$$

В соответствии с общей теорией (см. п. 4 § 5.3) отсюда найдем, что γ -доверительными интервалами для параметров μ, α_i и β_j являются соответственно [в данном случае параметр $n-k$ в (5.51) равен $rs - (r+s-1) = (r-1)(s-1)$]

$$(\hat{\mu} \pm t_{(1+\gamma)/2, (r-1)(s-1)} \sqrt{S_1/[rs(r-1)(s-1)]}),$$

$$(\hat{\alpha}_i \pm t_{(1+\gamma)/2, (r-1)(s-1)} \sqrt{S_1/[rs(s-1)]}),$$

$$(\hat{\beta}_j \pm t_{(1+\gamma)/2, (r-1)(s-1)} \sqrt{S_1/[rs(r-1)]}),$$

где $\hat{\mu}, \hat{\alpha}_i, \hat{\beta}_j$ и S_1 определены в (5.84) и (5.85).

Отметим также, что $S_1/(r-1)(s-1)$ — несмещенная оценка σ^2 .

Рассмотрим теперь задачу проверки наиболее интересных в исследуемой ситуации гипотез. Пусть требуется проверить гипотезу

$$H_0^{(1)}: \alpha_1 = \dots = \alpha_r = 0. \quad (5.87)$$

Другими словами, гипотеза $H_0^{(1)}$ означает, что фактор R не влияет на исход испытаний. Так, например, главным интересующим фактором может быть фактор C , соответствующий, например, различным способам обработки, в то время как R соответствует

классификации по месту, где производятся наблюдения. В этом случае гипотеза H_0 соответствует предположению о том, что эта вспомогательная классификация не влияет на результаты эксперимента и, следовательно, может не приниматься во внимание.

Следуя общей теории F -критерия (см. § 5.4), вычислим сначала S_T — минимальное значение квадратичной формы S в (5.82) при гипотезе H_0 . Из (5.83) и (5.85) имеем

$$S_T = S_1 + s \sum_{i=1}^r (X_{i.} - X_{..})^2,$$

поэтому статистика F [см. (5.70)] имеет в данном случае следующий вид (при $m = r - 1$):

$$F = F(X) = s(s-1) \frac{\sum_{i=1}^r (X_{i.} - X_{..})^2}{\sum_{i,j} (X_{ij} - X_{i.} - X_{.j} + X_{..})^2}. \quad (5.88)$$

При гипотезе H_0 статистика (5.88) имеет распределение Снедекора $S(r-1, (r-1)(s-1))$, поэтому критическая область уровня значимости α для H_0 имеет вид

$$\mathcal{R}_{1-\alpha}^{(1)} = \{x: F(x) \geq F_{1-\alpha, r-1, (r-1)(s-1)}\}. \quad (5.89)$$

Аналогично строится критерий проверки гипотезы

$$H_0^{(2)}: \beta_1 = \dots = \beta_s = 0 \quad (5.90)$$

о несущественности влияния на исходы испытаний второго фактора C . Здесь критическая область уровня значимости α для $H_0^{(2)}$ имеет вид

$$\mathcal{R}_{1-\alpha}^{(2)} = \left\{x: r(r-1) \frac{\sum_{j=1}^s (\bar{x}_{.j} - \bar{x}_{..})^2}{\sum_{i,j} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2} \geq F_{1-\alpha, s-1, (r-1)(s-1)}\right\}. \quad (5.91)$$

Наконец, критическая область уровня значимости α для гипотезы $H_0^{(3)}: \alpha_1 = \dots = \alpha_r = \beta_1 = \dots = \beta_s = 0$ о независимости результатов испытаний от влияния обоих факторов имеет вид

$$\mathcal{R}_{1-\alpha}^{(3)} = \left\{x: \frac{(r-1)(s-1)}{r+s-2} \frac{s \sum_{i=1}^r (\bar{x}_{i.} - \bar{x}_{..})^2 + r \sum_{j=1}^s (\bar{x}_{.j} - \bar{x}_{..})^2}{\sum_{i,j} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2} \geq F_{1-\alpha, r+s-2, (r-1)(s-1)}\right\}.$$

Найдем, наконец, доверительные области для различных групп параметров схемы (5.80). Пусть требуется построить γ -доверительную область $R_\gamma^{(1)}$ для эффектов $\alpha_1, \dots, \alpha_r$ первого фактора. В соответствии с общим принципом, чтобы построить $R_\gamma^{(1)}$, надо рассмотреть соответствующую гипотезу $H_0: \alpha_1 = \alpha_1^0, \dots, \alpha_r = \alpha_r^0$

$\left(\sum_{i=1}^r \alpha_i^0 = 0\right)$. Критическую область уровня значимости $1 - \gamma$ для нее можно получить из (5.88) — (5.89), заменяя x_{ij} на $x_{ij} - \alpha_i^0$; она имеет вид

$$\sum_{i,j} (x_{ij} - \bar{x}_{..} - \alpha_i^0)^2 \geq F_{\gamma, r-1, (r-1)(s-1)} \frac{1}{s(s-1)} \sum_{i,j} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2.$$

Отсюда следует вывод, что

$$R_\gamma^{(1)} = \left\{(\alpha_1, \dots, \alpha_r): \sum_{i=1}^r \alpha_i = 0 \text{ и } \sum_{i=1}^r (\alpha_i - (\bar{x}_{i.} - \bar{x}_{..}))^2 < F_{\gamma, r-1, (r-1)(s-1)} \frac{1}{s(s-1)} \sum_{i,j} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2\right\}.$$

Здесь неравенство определяет внутренность шара в пространстве параметров $(\alpha_1, \dots, \alpha_r)$ с центром $(\bar{x}_{1.} - \bar{x}_{..}, \dots, \bar{x}_{r.} - \bar{x}_{..})$ в гиперплоскости $\sum \alpha_i = 0$. Таким образом, в данном случае доверительная область $R_\gamma^{(1)}$ состоит из внутренности гиперсферы, получающейся в пересечении r -мерной сферы с гиперплоскостью $\sum \alpha_i = 0$. Так, если $r = 2$, то доверительная область — отрезок биссектрисы $\alpha_2 = -\alpha_1$, являющийся диаметром некоторой окружности с центром на этой биссектрисе (рис. 5.2).

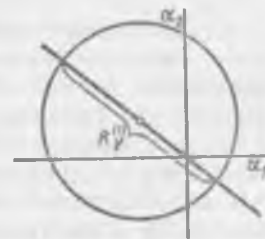


Рис. 5.2

Аналогично строится доверительная область для эффектов $\beta_1, \dots, \beta_s$ второго фактора.

В заключение дадим интерпретацию рассмотренной схемы, характерную для дисперсионного анализа. Положим в разложении (5.83) все α_i и β_j равными нулю, а $\mu = \bar{x}_{..}$. В результате получаем соотношение

$$\sum_{i,j} (x_{ij} - \bar{x}_{..})^2 = s \sum_i (\bar{x}_{i.} - \bar{x}_{..})^2 + r \sum_j (\bar{x}_{.j} - \bar{x}_{..})^2 + \sum_{i,j} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2,$$

которое можно интерпретировать как разложение полной изменчивости $S_0 = \sum_{i,j} (x_{ij} - \bar{x}_{..})^2$ на три компоненты:

$$S_0 = s \sum_i (\bar{x}_{i.} - \bar{x}_{..})^2, \quad S_0 = r \sum_j (\bar{x}_{.j} - \bar{x}_{..})^2,$$

$$S_0 = S_1 = \sum_{i,j} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2. \quad (5.92)$$

Компонента S_0 описывает изменчивость, обусловленную первым фактором R , компонента S_0 — вторым фактором C . Компонента же S_1 есть сумма квадратов величин с нулевыми средними:

$$E(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}) = \mu + \alpha_i + \beta_j - (\mu + \alpha_i) - (\mu + \beta_j) + \mu = 0,$$

поэтому она не может быть связана с факторами R и C . (Напомним, что величина $Q = S_{..}/(r-1)(s-1)$ определяет несмещенную оценку дисперсии σ^2 .) Поэтому $S_{..}$ часто называют «ошибкой», подчеркивая, что она связана со случайностью результатов наблюдений, а не с каким-либо расхождением в средних значениях наблюдений

Вычислим ES_0 и ES_0 . Из (5.84) и (5.86) имеем

$$E(X_i - X_{..})^2 = D\hat{\alpha}_i + (E\hat{\alpha}_i)^2 = \frac{r-1}{r} \sigma^2 + \alpha_i^2.$$

Отсюда и из (5.92) следует, что

$$ES_0 = (r-1) \sigma^2 + r \sum_{i=1}^r \alpha_i^2. \quad (5.93)$$

Аналогично находим

$$ES_0 = (s-1) \sigma^2 + r \sum_{j=1}^s \beta_j^2. \quad (5.94)$$

Из (5.93) имеем, что при гипотезе (5.87) величина $Q_1 = S_0/(r-1)$ может служить несмещенной оценкой для σ^2 . Аналогично, из (5.94) заключаем, что $Q_2 = S_0/(s-1)$ — несмещенная оценка для σ^2 , когда справедлива гипотеза (5.90).

Теперь F -критерий (5.89) для гипотезы (5.87) можно интерпретировать как критерий совместности двух независимых оценок Q_1 и Q_2 для σ^2 [статистика (5.88) в этих обозначениях равна отношению Q_1/Q_2]. Аналогично можно интерпретировать и критерий (5.91) для гипотезы (5.90) [статистикой этого критерия является отношение Q_2/Q_1].

Обычно составляющие этой двухфакторной модели объединяют в таблицу дисперсионного анализа (табл. 5.1)

Таблица 5.1

Источники дисперсии	Степень свободы	Сумма квадратов	Среднее суммы квадратов	Отношение Снедекора F
Строки	$r-1$	S_0	$Q_1 = S_0/(r-1)$	$F_{.0} = Q_1/Q$
Столбцы	$s-1$	S_0	$Q_2 = S_0/(s-1)$	$F_{0.} = Q_2/Q$
Ошибка	$(r-1)(s-1)$	$S_{..}$	$Q = S_{..}/[(r-1)(s-1)]$	

Первое отношение Снедекора $F_{.0}$ служит для проверки гипотезы о том, что все α_i равны нулю, второе — для проверки гипотезы о равенстве нулю всех β_j .

§ 5.6 Элементы теории статистической регрессии и корреляции

1. Задачи статистического прогноза. Предположим, что случайные величины Y и $X = (X_1, \dots, X_p)$ связаны некоторой статистической зависимостью, которую в общем случае можно выра-

знать их совместной функцией распределения $F_{XY}(x_1, \dots, x_p, y)$. Пусть, далее, случайная величина X доступна наблюдению, в то время как значение Y непосредственно измерить невозможно. Тогда возникает задача предсказания (прогноза, оценки) величины Y на основании информации, доставляемой измерением величин $X_1, \dots, X_p$, которые в этом случае называются *предсказывающими переменными*. Функция от предсказывающих переменных $\varphi(X)$, которую используют в качестве оценки для Y , называют *предиктором* величины Y по X . Задачей разработки методов построения оптимальных в том или ином смысле предикторов занимается *теория статистической регрессии*.

Прогноз необходим во многих практических ситуациях. Примерами могут быть прогнозирование погоды по результатам соответствующих атмосферных измерений, селекционирование новых видов растений и животных, определение возможностей индивидуумов в определенных областях с помощью соответствующей системы контрольных тестов и т. д. Во всех этих случаях речь идет о величинах, относящихся к будущему, недоступных наблюдению в данный момент, которые надо оценивать (прогнозировать) с помощью доступных измерению сопутствующих величин.

2. **Оптимальный предиктор и его свойства.** Научно обоснованный прогноз использует наличие статистической связи между переменными Y и X . (Если X и Y независимы, то предсказать Y по X нельзя.) Предположим сначала, что совместное распределение $\mathcal{L}(X, Y)$ известно. Тогда можно определить условное распределение $\mathcal{L}(Y | X = x)$. Так, если исходное распределение абсолютно непрерывно и $f_{XY}(x, y)$ — его плотность, то соответствующая условная плотность равна

$$f_{Y|X}(y|x) = f_{XY}(x, y) / \int_{-\infty}^{\infty} f_{XY}(x, t) dt.$$

Для дискретных распределений интеграл в последней формуле заменяют соответствующей суммой. Это условное распределение имеет среднее

$$M(x) = E(Y | X = x), \quad (5.95)$$

которое зависит от x и называется *функцией регрессии* Y на $X_1, \dots, X_p$. Например, для абсолютно непрерывного распределения $M(x) = \int y f_{Y|X}(y|x) dy$ (везде предполагается, что все соответствующие моменты существуют).

Пусть $\varphi(X)$ — произвольный предиктор Y по X . Назовем среднеквадратической ошибкой этого предиктора величину $E[Y - \varphi(X)]^2$. Предиктор $\varphi^*(X)$ называют *оптимальным* (в среднеквадратическом смысле), если

$$\Delta = E[Y - \varphi^*(X)]^2 = \inf_{\varphi} E[Y - \varphi(X)]^2. \quad (5.96)$$

Ответ на вопрос о существовании и виде оптимального предиктора дает следующее утверждение.

Теорема 5.6. *Оптимальный предиктор $\varphi^*(X)$ существует и имеет вид $\varphi^*(X) = M(X)$.*

□ По определению условного математического ожидания,

$$E[(Y - M(X))(M(X) - \varphi(X))] = E\{E[(Y - M(X))(M(X) - \varphi(X)) | X]\} = E\{(M(X) - \varphi(X))E(Y - M(X) | X)\} = 0,$$

поэтому

$$E[Y - \varphi(X)]^2 = E[(Y - M(X)) + (M(X) - \varphi(X))]^2 = \\ = E[Y - M(X)]^2 + E[M(X) - \varphi(X)]^2 \geq E[Y - M(X)]^2.$$

Знак равенства здесь имеет место при $\varphi = M$; следовательно, оптимальный предиктор есть определенная равенством (5.95) функция регрессии Y на X . ■

Заметим, что минимальную ошибку предсказания Δ в (5.96) можно записать в виде

$$\Delta = E\{E[(Y - M(X))^2 | X]\} = ED(Y | X) = \sigma_{Y|X}^2, \quad (5.97)$$

т. е. она представляет собой среднее значение условной дисперсии Y при данном X . Например, для абсолютно непрерывного распределения условная дисперсия

$$D(Y | X = x) = E[(Y - M(X))^2 | X = x] = \int (y - M(x))^2 f_{Y|X}(y | x) dy.$$

Оптимальный предиктор $M(X)$ обладает следующим важным свойством: он имеет максимальную корреляцию с Y среди всех предикторов. Для доказательства этого прежде всего заметим, что для произвольного предиктора $\varphi = \varphi(X)$

$$\text{cov}(\varphi, Y) = E[(\varphi - E\varphi)(Y - EY)] = E\{(\varphi - E\varphi)E(Y - EY | X)\} = \\ = E[(\varphi - E\varphi)(M - EM)] = \text{cov}(\varphi, M).$$

В дальнейшем используются обозначения $\rho(\xi_1, \xi_2) = \text{corr}(\xi_1, \xi_2)$, $\sigma_{\xi}^2 = D\xi$. При $\varphi = M$ из предыдущего равенства получаем $\text{cov}(M, Y) = \sigma_M^2 \geq 0$. Отсюда $\rho(M, Y) = \sigma_M^2 / (\sigma_M \sigma_Y) = \sigma_M / \sigma_Y$. Используя этот факт и свойства коэффициента корреляции, имеем

$$\rho^2(\varphi, Y) = \frac{\text{cov}^2(\varphi, Y)}{\sigma_{\varphi}^2 \sigma_Y^2} = \frac{\text{cov}^2(\varphi, M)}{\sigma_{\varphi}^2 \sigma_M^2} \frac{\sigma_M^2}{\sigma_Y^2} = \rho^2(\varphi, M) \rho^2(M, Y) \leq \rho^2(M, Y),$$

причем знак равенства имеет место только если φ — линейная функция M . Таким образом, $\rho(M, Y) \geq |\rho(\varphi, Y)|$ для любой φ .

Квадрат максимального значения коэффициента корреляции

$$\rho^2(M, Y) = \sigma_M^2 / \sigma_Y^2 = \eta_{YX}^2$$

обозначают η_{YX}^2 и называют *корреляционным отношением*. По определению, $0 \leq \eta_{YX}^2 \leq 1$, при этом $\eta_{YX}^2 = 1$ тогда и только тогда, когда $Y = M(X)$, т. е. когда Y функционально связано с X .

Выразим η_{YX}^2 через ошибку предсказания $\sigma_{Y|X}^2$, определенную в (5.97). Для этого запишем разложение дисперсии:

$$\sigma_Y^2 = E[Y - EY]^2 = E[(Y - M) + (M - EM)]^2 = \sigma_{Y|X}^2 + \sigma_M^2.$$

Отсюда

$$\eta_{YX} = 1 - \sigma_{YX}^2 / \sigma_Y^2. \quad (5.98)$$

Из этого представления следует, что $\eta_{YX} \rightarrow 1$, если ошибка прогноза $\sigma_{YX} \rightarrow 0$, и $\eta_{YX} = 0$, если, учитывая X , мы не уменьшаем ошибки прогноза ($\sigma_{YX} = \sigma_Y$). Таким образом, корреляционное отношение η_{YX} является мерой зависимости между Y и X (мерой точности прогноза) и с ее помощью можно сравнивать различные совокупности предсказывающих переменных в конкретных задачах.

Пример 5.9 (нормальное распределение, оптимальное прогнозирование для него). Пусть X имеет размерность $p=1$ и пара (X, Y) распределена нормально с параметрами $EX = \mu_X$, $EY = \mu_Y$, $DX = \sigma_X^2$, $DY = \sigma_Y^2$, $\text{corr}(X, Y) = \rho$, $|\rho| < 1$. Тогда совместная плотность распределения имеет вид

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \right] \right\}$$

и в результате несложных вычислений получаем, что условная плотность Y по X

$$f_{Y|X}(y|x) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho^2)}} \exp \left\{ -\frac{(y-m(x))^2}{2\sigma_Y^2(1-\rho^2)} \right\}, \text{ где}$$

$$m(x) = \mu_Y + \frac{\sigma_Y}{\sigma_X} \rho (x - \mu_X). \quad (5.99)$$

Таким образом, в данном случае условное распределение $\mathcal{L}(Y|X=x) = \mathcal{N}(m(x), \sigma_Y^2(1-\rho^2))$ и, следовательно,

$$M(x) = E(Y|X=x) = m(x), \quad D(Y|X=x) = \sigma_Y^2(1-\rho^2) \quad (5.100)$$

Условная дисперсия здесь не зависит от x , поэтому на основании (5.97)

$$\sigma_{YX}^2 = \sigma_Y^2(1-\rho^2).$$

Из (5.99) и (5.100) следует, что в данном случае функция регрессии Y на X является линейной, а оптимальный предиктор для Y можно записать в виде

$$M(X) = EY + \frac{\text{cov}(X, Y)}{\sigma_X^2} (X - EX). \quad (5.101)$$

Среднеквадратическая ошибка для этого предиктора равна

$$\sigma_{YX}^2 = DY - \frac{\text{cov}^2(X, Y)}{\sigma_X^2}. \quad (5.102)$$

Наконец, отсюда и из формулы (5.98) имеем

$$\eta_{YX} = \frac{\text{cov}^2(X, Y)}{DX DY} = \rho^2.$$

т. е. корреляционное отношение для нормально распределенной пары (X, Y) равно квадрату их коэффициента корреляции.

3. Прогнозирование в случае линейной функции регрессии. Пусть функция регрессии (5.95) является линейной, т. е. имеет вид

$$M(X) = \beta_0 + \beta'X = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p. \quad (5.103)$$

Найдем выражение коэффициентов β_i , $i = 0, 1, \dots, p$, через характеристики распределения $\mathcal{L}(X, Y)$. Как было показано в примере 5.9 для нормально распределенной пары (X, Y) , коэффициенты регрессии [см. (5.99) и (5.101)] выражаются только через первые и вторые моменты исходного распределения. Покажем, что это обстоятельство имеет общий характер.

Теорема 5.7. Пусть имеет место представление (5.103) и матрица вторых моментов вектора X , $\Sigma = D(X)$, не вырождена. Тогда оптимальный предиктор Y по X имеет вид

$$M(X) = EY + a' \Sigma^{-1} (X - EX), \quad (5.104)$$

где $a = (a_1, \dots, a_p)$, $a_i = \text{cov}(Y, X_i)$, $i = 1, \dots, p$.

□ Так как согласно теореме 5.6 оптимальный предиктор есть $M(X)$ и имеет место представление (5.103), то оптимальными являются те значения β^* коэффициентов β_i , которые минимизируют величину

$$E[Y - \beta_0 - \beta'X]^2 = E[(Y - EY) - b - \beta'(X - EX)]^2 = \\ = DY + b^2 + \beta' \Sigma \beta - 2\beta'a, \quad (5.105)$$

где $b = \beta_0 - EY + \beta'EX$. Из этого представления следует, что оптимальные значения b и β таковы:

$$b^* = 0, \quad \beta^* = \Sigma^{-1}a. \quad (5.106)$$

Действительно, при $\beta = \beta^* + \delta$

$$b^2 + \beta' \Sigma \beta - 2\beta'a = -\beta^{*'}a + \delta' \Sigma \delta + b^2 \geq -\beta^{*'}a \quad (5.107)$$

и равенство достигается только при $b = 0$, $\delta = 0$. Из (5.106) следует, что оптимальным значением коэффициента β_0 является

$$\beta_0^* = EY - \beta^{*'}EX. \quad (5.108)$$

Таким образом, из (5.106) и (5.108) имеем, что оптимальный предиктор

$$\varphi^*(X) = M(X) = \beta_0^* + \beta^{*'}X \quad (5.109)$$

имеет вид (5.104). ■

Отметим также, что в силу соотношений (5.105) — (5.107) среднеквадратическую ошибку предиктора (5.104) можно записать в виде

$$\sigma_{YX}^2 = E[Y - M(X)]^2 = DY - a'\beta^* = DY - a'\Sigma^{-1}a. \quad (5.110)$$

Отсюда имеем, что корреляционное отношение η_{YX} [см. (5.98)] совпадает с величиной

$$\rho_{YX} = a'\Sigma^{-1}a/\sigma_Y, \quad (5.111)$$

которая называется *множественным коэффициентом корреляции*. Эта величина зависит только от дисперсий и ковариаций переменных и является обобщением квадрата коэффициента корреляции двух величин на случай многих переменных (для случая двух переменных $\rho_{YX}^2 = \rho^2(X, Y)$). Отметим, что отношения (5.101) и (5.102) являются соответственно частными случаями соотношений (5.104) и (5.110).

Замечание. Разлагая определитель $\begin{vmatrix} \Sigma & a \\ a' & \sigma_Y^2 \end{vmatrix}$ по элементам последней строки, а затем и алгебраические дополнения этих элементов (за исключением алгебраического дополнения элемента σ_Y^2) — по элементам их последнего столбца, для σ_{YX}^2 можно получить следующее эквивалентное представление:

$$\sigma_{YX}^2 = \frac{\begin{vmatrix} \Sigma & a \\ a' & \sigma_Y^2 \end{vmatrix}}{\begin{vmatrix} \Sigma \end{vmatrix}}.$$

4. Линейное прогнозирование. Предположим теперь, что вид функции регрессии неизвестен и нужно построить оптимальный предиктор в классе линейных предикторов $\varphi(X)$, т. е. вместо всех функций от X будем рассматривать только функции вида $\varphi(X) = \beta_0 + \beta'X$. Тогда имеет место следующее важное утверждение.

Теорема 5.8. *Линейная функция $\varphi^*(X) = \beta_0^* + \beta^{*'}X$, определенная соотношениями (5.106), (5.108) и (5.109), является оптимальным линейным предиктором для Y . Эта функция имеет также максимальную корреляцию с Y среди всех линейных предикторов.*

□ Первая часть утверждения следует из того, что β_0^* и β^* определяются из условия минимума $E(Y - \beta_0 - \beta'X)^2$. Далее, из определения векторов a и β^* имеем следующие равенства: $\text{cov}(Y, \beta'X) = \beta'a = \beta'\Sigma\beta^*$, $\text{cov}(Y, \beta^{*'}X) = \beta^{*'}\Sigma\beta^* = D(\beta^{*'}X) \geq 0$. Из последнего равенства получаем

$$\sigma_Y^2 \rho^2(Y, \beta^{*'}X) = \text{cov}^2(Y, \beta^{*'}X) / D(\beta^{*'}X) = \beta^{*'}\Sigma\beta^*. \quad (5.112)$$

Воспользовавшись неравенством Коши — Буняковского $(\beta'\Sigma\beta^*)^2 = (\beta'\Sigma^{1/2}\Sigma^{1/2}\beta^*)^2 \leq (\beta'\Sigma\beta)(\beta^{*'}\Sigma\beta^*)$ и учитывая соотношение (5.112), получаем

$$\sigma_Y^2 \rho^2(Y, \beta'X) = \frac{\text{cov}^2(Y, \beta'X)}{D(\beta'X)} = \frac{(\beta'\Sigma\beta^*)^2}{\beta'\Sigma\beta} \leq \beta^{*'}\Sigma\beta^* = \sigma_Y^2 \rho^2(Y, \beta^{*'}X).$$

Таким образом, $\rho(Y, \beta^{*'}X) = \rho(Y, \varphi^*(X)) \geq |\rho(Y, \varphi(X))|$ для любой линейной функции $\varphi(X)$. ■

Отметим, что на основании соотношений (5.112), (5.106) и (5.111) квадрат максимальной корреляции

$$\rho^2(Y, \varphi^*(X)) = \beta^{*'}\Sigma\beta^* / \sigma_Y^2 = \rho_{YX}^2,$$

т. е. множественный коэффициент корреляции равен квадрату коэффициента корреляции между Y и оптимальным линейным предиктором для Y .

Так как η_{YX} — максимум корреляции между Y и любыми функциями X , а ρ_{YX} — максимум корреляции между Y и лишь линей-

ными функциями X , то $\eta_{YX}^2 \geq \rho_{YX}^2$. При этом, как было показано в п. 3, при линейной регрессии обе эти величины совпадают и *разность $\eta_{YX}^2 - \rho_{YX}^2$ может служить показателем отклонения регрессии от линейности*. Величины η_{YX}^2 и ρ_{YX}^2 принято записывать в виде $\eta_{0(1 \dots p)}^2$ и $\rho_{0(1 \dots p)}^2$.

5. Использование дополнительной информации. В ряде случаев бывает необходимо исследовать, насколько увеличивается точность прогноза при увеличении числа предсказываемых переменных. Среднеквадратическая ошибка предсказания Y по величинам $X = (X_1, \dots, X_p)$ равна [см. (5.98)] $\sigma_{YX}^2 = \sigma_{0(1 \dots p)}^2 = \sigma_Y^2 (1 - \eta_{0(1 \dots p)}^2)$. Если же учитывать какое-то число дополнительных предсказываемых переменных $X_{p+1}, \dots, X_k (k > p)$, то эта ошибка равна $\sigma_{0(1 \dots p \dots k)}^2 = \sigma_Y^2 (1 - \eta_{0(1 \dots p \dots k)}^2)$, что будет не больше предыдущей величины, и уменьшение ошибки определяется числом $\sigma_Y^2 (\eta_{0(1 \dots p \dots k)}^2 - \eta_{0(1 \dots p)}^2)$. Относительное уменьшение ошибки в результате использования дополнительных переменных, следовательно, равно

$$\eta_{0(p+1 \dots k)(1 \dots p)}^2 = (\eta_{0(1 \dots p \dots k)}^2 - \eta_{0(1 \dots p)}^2) / (1 - \eta_{0(1 \dots p)}^2). \quad (5.113)$$

Эту величину называют *частным корреляционным отношением*.

Если для прогнозирования используются только линейные предикторы, то соответствующее выражение для относительного уменьшения среднеквадратической ошибки прогноза можно получить из (5.113), заменяя η на ρ :

$$\rho_{0(p+1 \dots k)(1 \dots p)}^2 = (\rho_{0(1 \dots p \dots k)}^2 - \rho_{0(1 \dots p)}^2) / (1 - \rho_{0(1 \dots p)}^2). \quad (5.114)$$

Эту величину называют *частным множественным коэффициентом корреляции*; она измеряет корреляцию между Y и $X_{p+1}, \dots, X_k$, исключая зависимость от $X_1, \dots, X_p$.

Если необходимо исследовать целесообразность добавления отдельных переменных к уже имеющимся, то надо вычислить $\rho_{0(p)(1 \dots p-1)}^2$. Из (5.114) имеем

$$1 - \rho_{0(p)(1 \dots p-1)}^2 = (1 - \rho_{0(1 \dots p)}^2) / (1 - \rho_{0(1 \dots p-1)}^2). \quad (5.115)$$

Используя представление (5.111), можно показать, что

$$1 - \rho_{0(1 \dots p)}^2 = 1/\rho^{00}. \quad (5.116)$$

где $\rho^{ij} = \rho_{ij}/\sqrt{1 - \rho_{ij}^2}$ и $\rho_{ij} = \rho(X_i, X_j)$, $i, j = 0, 1, \dots, p$, ($X_0 = Y$). Из (5.115) — (5.116) легко получаем

$$\rho_{0(p)(1 \dots p-1)}^2 \equiv \rho_{0p}^2(1 \dots p-1) = (\rho^{0p})^2 / (\rho^{00} \rho^{pp}). \quad (5.117)$$

Определенный в соотношении (5.117) коэффициент $\rho_{0p}^2(1 \dots p-1)$ называют *частной корреляцией* между Y и X_p , исключаяющей влияние переменных $X_1, \dots, X_{p-1}$. Можно показать, что $\rho_{0p(1 \dots p-1)} = \rho(e_1, e_p)$, где $e_1 = Y_1 - (\beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1})$, $e_p = X_p - (\alpha_0 + \alpha_1 X_1 + \dots + \alpha_{p-1} X_{p-1})$, т. е. это обычная корреляция между остатками, получающимися вычитанием из Y и X_p их оптимальных линейных предикторов, основанных на $X_1, \dots, X_{p-1}$ (так исключают влияние общих факторов $X_1, \dots, X_{p-1}$ на переменные Y и X_p).

6. Эмпирические предикторы. Изложенная теория предсказания величины Y по сопутствующим переменным X предполагает известным совместный закон распределения $\mathcal{L}(X, Y)$; при линейном же прогнозировании достаточно знать только первые и вторые моменты этого распределения. В практических приложениях точный вид зависимости между Y и X чаще всего неизвестен, и поэтому все сведения, необходимые для построения оптимальных предикторов, получают в результате обработки статистических

данных, которые представляют собой выборку из распределения $\mathcal{L}(X, Y)$. Оценив по такой вспомогательной выборке (по результатам прошлых измерений X и Y) соответствующие характеристики распределения $\mathcal{L}(X, Y)$ (для построения линейных предикторов это первые и вторые моменты [см. (5.104)]) и заменив эти оценки теоретические характеристики, строят *эмпирический* предиктор, который и используют для предсказания в других случаях (в будущих измерениях). Более подробно теория статистической регрессии изложена, например, в [16, гл. 4] и в [10, гл. 26—28.].

7. Прогнозирование стационарных последовательностей. Теория линейного прогноза имеет наиболее широкое применение при прогнозировании *стационарных случайных процессов*. Под этим термином будем понимать бесконечную в обе стороны последовательность случайных величин $\{X_t\}$, $t = \dots, -n, \dots, -1, 0, 1, \dots, \dots, n, \dots$, удовлетворяющих следующим условиям:

$$E X_t = m = \text{const}, \text{cov}(X_{t+k}, X_t) = E(X_{t+k} - m)(X_t - m) = R_k. \quad (5.118)$$

Таким образом, для стационарной последовательности случайных величин их среднее постоянно, а ковариация любых двух величин зависит только от разности номеров этих величин. Последовательность чисел $\{R_k\}$, $k = \dots, -1, 0, 1, \dots$, называют *ковариационной функцией* последовательности $\{X_t\}$. Отметим следующие ее свойства: $R_0 = E(X_t - m)^2 = D X_t = \sigma^2 = \text{const}$;

$$\text{corr}(X_{t+k}, X_t) = \frac{R_k}{R_0} \Rightarrow |R_k| \leq R_0; R_{-k} = R_k; \quad (5.119)$$

$$\sum_{k, j=1}^n R_{k-j} c_k c_j = \sum_{k, j=1}^n c_k c_j \text{cov}(X_k, X_j) = D \left(\sum_{k=1}^n c_k X_k \right) \geq 0.$$

Последнее свойство называется *свойством положительной определенности*. Особый интерес представляет случай, когда ряд из R_k сходится абсолютно:

$$\sum_{k=-\infty}^{\infty} |R_k| = R_0 + 2 \sum_{k=1}^{\infty} |R_k| < \infty. \quad (5.120)$$

Это условие везде в дальнейшем будем предполагать выполненным.

Рассмотрим преобразование Фурье ковариационной функции $\{R_k\}$:

$$f(\lambda) = \frac{1}{2\pi} R_0 + \frac{1}{\pi} \sum_{k=1}^{\infty} R_k \cos k\lambda = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} R_k \cos k\lambda, \quad \lambda \in [-\pi, \pi]. \quad (5.121)$$

Условие (5.120) обеспечивает равномерную сходимость ряда (5.121) к непрерывной функции, и его сумму $f(\lambda)$ называют *спектральной плотностью* последовательности $\{X_t\}$. Коэффициенты сходящегося ряда Фурье однозначно определяются его суммой по формуле

$$R_k = \int_{-\pi}^{\pi} f(\lambda) \cos k\lambda \, d\lambda, \quad (5.122)$$

Эти соотношения показывают, что спектральная плотность $f(\lambda)$ (когда она существует) и ковариационная функция $\{R_k\}$ находятся во взаимно однозначном соответствии, поэтому стационарную последовательность $\{X_t\}$ можно описывать в терминах любой из этих функций.

Предположим теперь, что известны значения процесса $\{X_t\}$ в «прошлые» моменты $t = -n, -n+1, \dots, -1, 0$ и требуется предсказать X_1 , т. е. значение процесса в «будущий» момент $t = 1$. Обозначим оптимальный линейный предиктор для X_1 через X_{1n}^* ; тогда, по теореме 5.8,

$$X_{1n}^* = \sum_{t=-n}^0 \beta_{1n}^* X_t, \quad (5.123)$$

где β_{1n}^* , $t = -n, \dots, -1, 0$, — значения β_t , минимизирующие выражение $E \left[X_1 - \sum_{t=-n}^0 \beta_t X_t \right]^2$. Согласно соотношениям (5.106), величины β_{1n}^* определяются в данном случае ковариационной функцией процесса $\{X_t\}$ по следующим формулам: $\beta_{1n}^* = \sum_{t=-n}^0 R_{tj} R_{j-1}^{-1}$, $t = -n, \dots, -1, 0$, $|R_{tj}| = |R_{t-j}|^{-1}$ ($t, j = -n, \dots, -1, 0$). Обозначим

$$R(n+1) = \begin{vmatrix} R_0 & R_1 & \dots & R_n \\ R_1 & R_0 & \dots & R_{n-1} \\ \dots & \dots & \dots & \dots \\ R_n & R_{n-1} & \dots & R_0 \end{vmatrix};$$

тогда в силу четности функции R_k [см. (5.119)] $|R_{tj}| = R^{-1}(n+1)$. Наконец, принимая во внимание замечание к теореме (5.7), получаем, что средний квадрат ошибки предиктора (5.123)

$$\sigma^2(n) = E[X_1 - X_{1n}^*]^2 = \frac{|R(n+2)|}{|R(n+1)|}. \quad (5.124)$$

Ответ на вопрос, при каких условиях на процесс $\{X_t\}$ существует $\sigma^2 = \lim_{n \rightarrow \infty} \sigma^2(n)$ и когда $\sigma^2 = 0$ или $\sigma^2 > 0$, получен А. Н. Колмогоровым (1941 г.) и состоит в следующем. Обозначим через $\lambda_{j/n}$, $j = 1, \dots, n+1$, собственные значения матрицы $R(n+1)$. Тогда если

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{n+1} \ln \lambda_{j/n} = K, \text{ то } \sigma^2 = e^K. \quad (5.125)$$

Таким образом, если $K = -\infty$, то предел величины (5.124) при неограниченном увеличении числа наблюдений равен нулю, т. е. в этом случае прогноз является асимптотически точным.

Наконец, если предел в (5.125) конечен ($K > -\infty$), то он равен

$$K = \ln 2\pi + \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln f(\lambda) d\lambda.$$

Итак, если ковариационная функция $\{R_k\}$ (или спектральная плотность $f(\lambda)$) процесса $\{X_t\}$ известны, то линейный предиктор (5.123) и его среднеквадратическая ошибка (5.124) определяются однозначно. Если же $\{R_k\}$ неизвестна, то ее предварительно следует оценить по вспомогательной выборке. Рассмотрим кратко вопросы оценивания среднего m и ковариационной функции $\{R_k\}$ по выборке $(X_1, \dots, X_n)$ из процесса $\{X_t\}$ (таким образом, здесь имеем дело с зависимыми наблюдениями).

Среднее арифметическое $\bar{X} = (X_1 + \dots + X_n)/n$ на основании условия (5.118) является несмещенной оценкой для m : $E\bar{X} = m$. Вычислим дисперсию этой оценки. Имеем

$$\begin{aligned} D\bar{X} &= \frac{1}{n^2} \sum_{k,s=1}^n \text{cov}(X_k, X_s) = \frac{1}{n^2} \sum_{k,s=1}^n R_{k-s} = \\ &= \frac{1}{n} \left[R_0 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) R_k \right]. \end{aligned} \quad (5.126)$$

Отсюда и из условия (5.120) следует, что $D\bar{X} \rightarrow 0$ при $n \rightarrow \infty$, имея порядок $1/n$, т. е. $\bar{X}$ — состоятельная оценка m . Нетрудно показать, что $\lim_{n \rightarrow \infty} \sum_{k=1}^{n-1} (1 - k/n) R_k = \sum_{k=1}^{\infty} R_k$, поэтому из (5.126) и (5.121) следует, что $D\bar{X} \sim \frac{2\pi}{n} f(0)$.

Рассмотрим теперь задачу оценивания R_k (в силу четности R_k достаточно рассмотреть только $k \geq 0$). Если среднее m известно, то, полагая

$$C_k(n) = \frac{1}{n-k} \sum_{t=1}^{n-k} (X_t - m)(X_{t+k} - m), \quad 0 \leq k < n, \quad (5.127)$$

по формуле (5.118) имеем

$$EC_k(n) = \frac{1}{n-k} \sum_{t=1}^{n-k} R_k = R_k,$$

т. е. статистика $C_k(n)$ является несмещенной оценкой R_k . Если же m неизвестно, то в формуле (5.127) вместо m надо подставить его оценку $\bar{X}$; тогда в качестве оценки R_k можно рассматривать статистику

$$\bar{C}_k(n) = \frac{1}{n-k} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}). \quad (5.128)$$

Эта статистика уже не является несмещенной оценкой R_k , но можно показать, что при выполнении условия (5.120) смещение $E\bar{C}_k(n) - R_k = O(1/n)$, т. е. $\bar{C}_k(n)$ — асимптотически несмещенная оценка R_k для любого фиксированного k .

Исследование вторых моментов оценок $C_h(n)$ и $\bar{C}_h(n)$ более сложно и требует более сильных предположений о структуре процесса $\{X_t\}$. При соответствующих условиях дисперсии и ковариации этих оценок стремятся к нулю при $n \rightarrow \infty$ (для ограниченных k), т. е. они являются состоятельными оценками ковариационной функции R_h (по крайней мере, при ограниченных значениях k). Например, для *гауссовских* последовательностей $\{X_t\}$ (т. е. когда распределение любой конечной совокупности величин из $\{X_t\}$ нормальное) необходимое и достаточное условие состоятельности этих оценок имеет вид

$$\frac{1}{n} \sum_{h=1}^n R_h^2 \rightarrow 0, \quad n \rightarrow \infty.$$

Задачи

1. Вывести нормальное уравнение (5.5) метода наименьших квадратов.
2. Доказать формулу (5.21), определяющую обобщенную оценку наименьших квадратов $\hat{\beta}_T$, и проверить, что матрица D в ней положительно определена.

3. По данным независимых равнооточных измерений (X_i, t_i) , $i=1, \dots, n$, значений некоторой линейной функции $x = \beta_1 + \beta_2 t$ (погрешности измерений подчиняются нормальному распределению $\mathcal{N}(0, \sigma^2)$ с неизвестной дисперсией) построить доверительный интервал для интеграла от этой функции на отрезке $-a \leq t \leq a$ (a задано). Произвести соответствующие вычисления для следующих данных: (2,96; -2), (3,20; -1), (3,41; 0), (3,63; 1), (3,79; 2) при $a=2$ и доверительном уровне $\gamma=0,95$.

4. В четырехугольнике $ABCD$ результаты независимых и равнооточных измерений углов ABD , DBC , ABC , BCD , CDB , BDA , CDA и DAB (в градусах) соответственно таковы: 50,78; 30,25; 78,29; 99,57; 50,42; 40,59; 88,87; 89,86. Считая, что ошибки измерений распределены нормально $\mathcal{N}(0, \sigma^2)$, найти о. н. к. углов $\beta_1 = ABD$, $\beta_2 = DBC$, $\beta_3 = CDB$ и $\beta_4 = BDA$. Построить 0,95-доверительный интервал для σ^2 .

5. Результаты замеров координаты $a(t)$ движущейся равномерно и прямолинейно точки в моменты $t=1, 2, 3, 4, 5$ оказались соответственно равны: 12,98; 13,05; 13,32; 14,22; 13,97. Предполагая погрешности измерений независимыми и нормальными $\mathcal{N}(0, \sigma^2)$, построить 0,95-доверительный эллипс для точки $(a(0); v)$, где v — скорость точки.

6. Значения независимых случайных величин $X_j^{(i)}$ ($i=1, 2, 3, 4; j=1, 2$) приведены в следующей таблице:

$i \backslash j$	1	2	3	4
1	8,67	9,71	10,16	13,65
2	10,03	10,23	9,26	13,79

Предполагая, что $\mathcal{L}(X_j^{(i)}) = \mathcal{N}(\mu_j, \sigma^2)$ (все параметры неизвестны), построить оценки для $\mu_1, \mu_2, \mu_3, \mu_4$ и σ^2 и проверить гипотезу однородности $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ (уровень значимости принять равным 0,1).

7. Пусть имеются k предметов, веса которых $\beta_1, \dots, \beta_k$ неизвестны. Для определения этих весов взвешивают комбинации предметов: каждая операция (одно взвешивание) состоит в том, что несколько предметов кладут на одну чашу весов, несколько — на другую и добавляют равновес для приведения весов в равновесие. В результате получают соотношения

$$z_1^{(i)} \beta_1 + \dots + z_k^{(i)} \beta_k = y_i$$

(для i -го взвешивания, $i = 1, \dots, n$), где $x_j^{(i)} = 1, -1, 0$ в зависимости от того, лежит j -й предмет на левой чаше весов, на правой или вообще не участвует в данном взвешивании, а y_i — добавляемый равновес. Считая погрешности измерений независимыми и нормальными $N(0, \sigma^2)$, оценить веса четырех предметов по данным следующей таблицы восьми взвешиваний:

β_1	1	1	1	1	1	1	1	1
β_2	1	-1	1	-1	1	-1	1	-1
β_3	1	1	-1	-1	1	1	-1	-1
β_4	1	-1	-1	1	1	-1	-1	1
Вес	20,2	8,1	9,7	1,9	19,9	8,3	10,2	1,8

Найти матрицу ковариаций оценок, а также оценку для σ^2 . Сравнить точность этих оценок с точностью оценок, получаемых обычным способом, когда каждый предмет взвешивают несколько раз и в качестве оценки его веса принимают среднее арифметическое результатов взвешиваний. У к а з а н и е. Использовать теорему 5.3.

8. Для данных задачи 7 построить систему совместных доверительных интервалов для $\beta_1, \dots, \beta_4$ уровня $\geq 0,95$.

9. Убедиться в том, что интервалы

$$(\hat{\beta}_i - \hat{\beta}_j \pm \left[\frac{k}{n-k} S(\hat{\beta}) F_{\gamma, k, n-k} (a^{ii} - 2a^{ij} + a^{jj}) \right]^{1/2}), \quad 1 \leq j < i \leq k,$$

образуют систему совместных доверительных интервалов уровня $\geq \gamma$ для разностей $\beta_i - \beta_j$, $i > j$.

10. Доказать, что последовательности $\{X_t = \xi \cos t + \eta \sin t, \quad t = 0, \pm 1, \pm 2, \dots\}$ и $\left\{X_t = \sum_{j=0}^k \alpha_j \xi_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots\right\}$, где ξ, η и $\xi_j, t = 0, \pm 1, \pm 2, \dots$, — некоррелированные случайные величины с одинаковыми средними и дисперсиями, — являются стационарными, вычислить их ковариационные функции.

11. Доказать соотношение (5.116).

12. Доказать асимптотическую несмещенность оценки $\hat{C}_k(n)$ ковариационной функции R_k , определенной в (5.128).

Элементы теории решений. Дискриминантный анализ

Эта глава является кратким введением в одно из важных направлений современных статистических исследований — теорию статистических решающих функций. Здесь излагаются некоторые основные идеи и результаты этой теории в рамках двух традиционных подходов к принятию решения — байесовского и минимаксного. Общие идеи рассматриваются на примере решения практически важной задачи классификации наблюдений по нескольким категориям (классам). Подробно исследуется ряд ситуаций, характеризующихся предположением о нормальном законе распределения наблюдений.

§ 6.1. Статистические решающие функции. Байесовское и минимаксное решения]

1. Понятие решающей функции. Во многих случаях конечную цель статистического анализа можно выразить в форме решения о том или ином поведении или действии. Так, при выборочном контроле продукции следует принять одно из двух решений: принять партию изделий или забраковать ее; врач, анализируя симптомы болезни, должен отнести ее к одному из конечного числа стандартных видов, т. е. принять одно из конечного числа возможных решений; анализируя данные наблюдения за некоторым случайным процессом с постоянным, но неизвестным средним, требуется принять решение о величине воздействия на процесс (его коррекции) для приведения среднего, например, к нулю; это воздействие может быть выражено некоторым действительным числом t и, следовательно, здесь число решений бесконечно. Во всех этих случаях решение принимается на основании анализа наблюдения x над соответствующей случайной величиной X и, следовательно, представляет собой некоторую функцию $\delta(x)$ на выборочном пространстве $\mathcal{X} = \{x\}$, область значений которой — множество возможных в данной ситуации решений $D = \{d\}$. Таким образом, $\delta(x)$ — это правило, ставящее в соответствие каждому результату наблюдения $x \in \mathcal{X}$ решение $d = \delta(x) \in D$, которое должно быть принято. Функцию $\delta(x)$ называют *решающей функцией* (или *решающим правилом*, или *процедурой*), и ее следует выбирать в соответствии с некоторыми требованиями оптимальности. Принципы решения этой задачи развивает теория статистических решающих функций (теория решений), разрабо-

танная А. Вальдом (1950 г.). Кратко изложим некоторые основные идеи и результаты этой теории.

2. **Функция риска и допустимые решающие правила.** Пусть заданы класс распределений $\mathcal{F} = \{F(x; \theta), \theta \in \Theta\}$, которому, по предположению, принадлежит распределение наблюдаемой случайной величины X , и множество решений $D = \{d\}$, которые можно принимать на основании наблюдения над X . Чтобы получить критерий выбора решающей функции, необходимо сравнить результаты использования различных правил δ . Введем для этого неотрицательную *функцию потерь* $L(\theta, d)$, определенную на прямом произведении $\Theta \times D$, где для каждого $\theta \in \Theta$ и $d \in D$ число $L(\theta, d) \geq 0$ интерпретируется как убыток или потеря от принятия решения d при условии, что распределение X есть $F(x; \theta)$. Тогда для всякого правила δ можно определить функцию $R(\theta, \delta)$, называемую *функцией риска*, как среднее значение функции потерь $L(\theta, \delta(X))$:

$$R(\theta, \delta) = E_{\theta} L(\theta, \delta(X)). \quad (6.1)$$

Таким образом, $R(\theta, \delta)$ — это средние потери, которые имеют место в случае применения решающего правила δ , когда истинным распределением наблюдений является распределение $F(x; \theta)$.

Функция риска дает критерий сравнения различных решающих правил. Действительно, если имеются два правила δ' и δ такие, что

$$R(\theta, \delta') \leq R(\theta, \delta), \quad \forall \theta \in \Theta, \quad (6.2)$$

со строгим неравенством хотя бы для одного θ , то, очевидно, правило δ' предпочтительнее δ , так как использование δ' приводит в среднем к меньшим потерям. С другой стороны, решающие правила δ_1 и δ_2 могут оказаться несравнимыми по критерию (6.2), если $R(\theta, \delta_1) < R(\theta, \delta_2)$ для некоторых значений θ , а для остальных значений θ имеет место обратное неравенство. Чтобы выбрать в такой ситуации одно из двух правил, необходимо привлечь дополнительные соображения.

Решающее правило δ называют *недопустимым*, если существует правило δ' , предпочтительнее δ в смысле (6.2). Решающее правило, не являющееся недопустимым, называется *допустимым*. Все недопустимые решающие правила должны быть отвергнуты, так как для каждого из них можно найти более оптимальное (в указанном смысле) правило. Если класс допустимых решающих правил состоит из единственного правила, то имеет место оптимальное решение. Но обычно этот класс достаточно широк и никакие два решающие правила из множества допустимых несравнимы. Поэтому для дальнейшего упорядочения допустимых правил и выбора среди них лучшего привлекают дополнительные соображения. Для решения этой задачи в статистике традиционно применяют два подхода: *байесовский* и *минимаксный*.

3. **Байесовское решение.** При байесовском подходе дополнительно предполагается, что параметр θ — это случайная величина

с некоторым распределением, задаваемым плотностью распределения (или вероятностью в дискретном случае) $\pi(\theta)$. (В таких случаях будем использовать одно и то же обозначение θ как для случайной величины, так и для принимаемого ею значения, полагая, что это не вызовет затруднений при чтении.) Это распределение называют *априорным распределением параметра* и считают известным. В этом случае можно подсчитать полную среднюю потерю от применения решающего правила δ . Эту потерю обозначают $r(\delta)$ и называют *байесовским риском*:

$$r(\delta) = \int R(\theta, \delta) \pi(\theta) d\theta. \quad (6.3)$$

Если θ принимает дискретные значения, то вместо интеграла пишут соответствующую сумму. Таким образом, в такой (байесовской) ситуации каждое решающее правило характеризуется одним числом r , следовательно, все решающие правила могут быть упорядочены в соответствии со значениями этой характеристики. Оптимальной является процедура δ^* , минимизирующая байесовский риск $r(\delta)$; она называется *байесовским решением*. Отметим, что δ^* зависит от априорного распределения π , поэтому для различных априорных распределений параметра байесовские решения, вообще говоря, различны.

Байесовское решение δ^* находят на основании *теоремы Байеса*, согласно которой *апостериорное распределение параметра* при условии $X=x$ задается плотностью распределения (или вероятностью в дискретном случае)

$$\pi(\theta | x) = f(x; \theta) \pi(\theta) / f(x), \quad (6.4)$$

где $f(x) = E f(x; \theta) = \int f(x; \theta) \pi(\theta) d\theta$, или $f(x) = \sum_i f(x; \theta_i) \pi(\theta_i)$,

если θ принимает дискретные значения $\{\theta_i\}$. Тогда равенство (6.3) можно записать в виде

$$r(\delta) = \int f(x) E [L(\theta, \delta(X)) | X = x] dx \quad (6.5)$$

(если случайная величина X дискретна, то вместо интеграла пишут соответствующую сумму), где математическое ожидание вычисляется относительно апостериорного распределения (6.4), т. е., например, в дискретном случае

$$E [L(\theta, \delta(X)) | X = x] = \sum_i L(\theta_i, \delta(x)) \pi(\theta_i | x). \quad (6.6)$$

Из соотношения (6.5) следует, что *оптимальной является такая процедура, которая при каждом x минимизирует средние потери относительно апостериорного распределения $\pi(\theta | x)$.*

Таким образом, для заданного априорного распределения параметра $\pi(\theta)$ алгоритм нахождения байесовского решения имеет следующий вид:

а) для $X=x$ по формуле (6.4) находят апостериорное распределение $\pi(\theta | x)$;

б) после этого для каждого возможного решения $d \in D$ вычисляют среднюю потерю относительно этого апостериорного распределения;

в) в качестве искомого выбирают решение с минимальной средней потерей.

(Здесь предполагается, что такое решение входит в множество D , что всегда, в частности, имеет место для конечных множеств D .)

Отметим, наконец, что байесовское решение, построенное для любого распределения $\pi(\theta) > 0$, является допустимым. Действительно, пусть δ^* — соответствующее байесовское решение и предположим, что существует процедура δ , для которой $R(\theta, \delta) < R(\theta, \delta^*)$ со строгим неравенством на множестве значений θ , имеющем положительную (относительно распределения π) вероятность. Тогда, очевидно, для соответствующих байесовских рисков имеет место [см. (6.3)] неравенство $r(\delta) < r(\delta^*)$. Но это противоречит факту, что байесовское правило минимизирует байесовский риск.

4. Минимаксное решение. При отсутствии априорной информации о θ применяют прием упорядочения допустимых решающих правил, в котором в качестве основной характеристики функции риска $R(\theta, \delta)$ используется ее максимальное значение (или *максимальный риск*) $m(\delta) = \sup_{\theta \in \Theta} R(\theta, \delta)$. Тогда из двух решающих

правил предпочтительным считают то, которому соответствует меньший максимальный риск. Правило $\bar{\delta}$, минимизирующее $m(\delta)$, называется *минимаксным решающим правилом*. Таким образом, минимаксное правило избавляет от чрезмерных потерь: наихудший ожидаемый ущерб, связанный с использованием этого правила, настолько мал, насколько это возможно. Принцип минимакса не всегда является разумным (рис. 6.1). Здесь δ_2 имеет худшие свойства по сравнению с δ_1 для большинства значений θ , но предпочтительнее δ_1 по принципу минимакса.

В общем случае вопрос о существовании и строении минимаксного правила достаточно трудный и здесь не рассматривается. Отметим один случай, когда удастся просто установить минимаксность некоторого решающего правила.

Именно: предположим, что существует априорное распределение параметра $\pi(\theta) > 0$, для которого функция риска соответствующего байесовского правила δ^* постоянна: $R(\theta, \delta^*) = a = \text{const}$ (такое распределение π называют *наименее благоприятным априорным распределением*). Тогда δ^* — минимаксное решение.

Действительно, в противном случае существовала бы процедура δ , для которой максимальный риск $m(\delta) < a$. Но это означало бы, что $R(\theta, \delta) < R(\theta, \delta^*)$, $\forall \theta$, в противоречии с допустимостью байесовского решения.

Пример 6.1 (бернуллиевская модель, решающие правила для нее). Пусть X — бернуллиевская случайная величина, причем вероят-

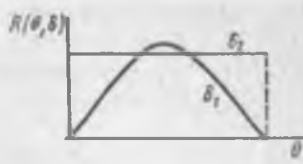


Рис. 6.1

ность «успеха» θ может быть либо $\theta_1 = 1/3$, либо $\theta = 1/2$. Таким образом, здесь $\mathcal{X} = \{0, 1\}$, $\Theta = \{1/3, 1/2\}$ и $f(x; \theta) = \theta^x (1 - \theta)^{1-x}$, $x = 0, 1$. Пусть, далее, множество решений D состоит из двух элементов d_1 и d_2 , а функция потерь $L(\theta, d_j)$ определяется таблицей

	d_1	d_2
θ_1	0	2
θ_2	3	1

В данном случае для каждого $x \in \mathcal{X}$ возможны только два решения, а множество $\mathcal{X}$ содержит две точки, поэтому всего имеются четыре решающие функции δ_k , $k = 1, 2, 3, 4$, а именно: $\delta_1(0) = d_1$, $\delta_1(1) = d_1$; $\delta_2(0) = d_1$, $\delta_2(1) = d_2$; $\delta_3(0) = d_2$, $\delta_3(1) = d_1$; $\delta_4(0) = d_2$, $\delta_4(1) = d_2$. По формуле (6.1), которая в данном случае имеет вид

$$R(\theta, \delta_k) = L(\theta, \delta_k(0))(1 - \theta) + L(\theta, \delta_k(1))\theta, \quad \theta = \theta_1, \theta_2,$$

находим четыре вектора риска $(R(\theta_1, \delta_k), R(\theta_2, \delta_k))$, $k = 1, \dots, 4$, числовые значения которых соответственно равны $(0, 3)$; $(2/3, 2)$; $(4/3, 2)$; $(2, 1)$. Здесь процедура δ_3 является недопустимой, так как процедура δ_2 предпочтительнее, а среди допустимых процедур δ_1 , δ_2 и δ_4 две последние обладают минимаксным свойством: $m(\delta_2) = m(\delta_4) = 2 < m(\delta_1) = 3$. Следовательно, в данном случае в качестве минимаксной процедуры можно выбрать либо δ_2 , либо δ_4 .

5. Оценивание параметров и проверка гипотез с позиций теории решений. В заключение отметим, что рассмотренные в предыдущих главах задачи оценивания параметров и проверки гипотез также можно сформулировать в терминах принятия решений. Рассмотрим задачу точечной оценки скалярного параметра θ . Выбор статистики $T(X)$, оценивающей θ , можно трактовать как решающее правило, предписывающее принимать решение d_t о том, что оцениваемое значение параметра равно $t = T(x)$, если наблюдается $X = x$. В этом случае функция потерь может быть, например, $L(\theta, d_t) = \omega(|t - \theta|)$, где ω — строго возрастающая функция ошибки $|t - \theta|$. Если, в частности, выбрать $\omega(z) = z^2$, то функция риска $R(\theta, \delta) = R(\theta, T) = E_\theta(T(X) - \theta)^2$ совпадает со среднеквадратической ошибкой оценки T . В гл. 2 была рассмотрена задача отыскания оптимальных оценок (минимизирующих эту функцию риска) в классе несмещенных оценок. Рассмотрим пример, иллюстрирующий возможности подхода с позиций теории решений.

Пример 6.2 (минимаксная оценка вероятности успеха в схеме Бернулли). Пусть X — число успехов в n испытаниях Бернулли с постоянной вероятностью успеха θ (наблюдаемая случайная величина). Как известно (см. теорему 2.2), оптимальной несме-

щенной оценкой θ является статистика $T = X/n$. Предположим теперь, что параметр θ — случайная величина с априорным распределением типа *бета-распределения* с плотностью

$$\pi(\theta) = \theta^{a-1} (1-\theta)^{b-1} / B(a, b)$$

при известных постоянных $a, b > 0$ [здесь $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$]. Тогда по формуле (6.4) плотность апостериорного распределения θ при $X = x \in \{0, 1, \dots, n\}$ равна

$$\pi(\theta | x) = \theta^{x+a-1} (1-\theta)^{n-x+b-1} / B(x+a, n-x+b).$$

Это другое бета-распределение, с новыми параметрами $(x+a, n-x+b)$.

Вычислим средние потери для каждого возможного решения d , относительно этого апостериорного распределения. Имеем

$$\int_0^1 (t-\theta)^2 \pi(\theta | x) d\theta = t^2 - 2t \frac{x+a}{n+a+b} + \frac{(x+a)(x+a+1)}{(n+a+b)(n+a+b+1)}.$$

Минимизируя полученное выражение по t , получаем, что байесовской (при заданном априорном распределении π) оценкой θ является статистика

$$\delta^*(X) = (X+a)/(n+a+b). \quad (6.7)$$

Функция риска этой оценки равна [см. (2.2)] сумме квадрата смещения

$$b(\theta) = E_{\theta} \delta^*(X) - \theta = (n\theta + a)/(n+a+b) - \theta = [a - \theta(a+b)]/(n+a+b)$$

и дисперсии

$$D_{\theta} \delta^*(X) = [n\theta(1-\theta)]/(n+a+b)^2, \text{ т. е.}$$

$$R(\theta, \delta^*) = \frac{[a - \theta(a+b)]^2 + n\theta(1-\theta)}{(n+a+b)^2}. \quad (6.8)$$

Чтобы получить минимаксное решение, надо найти такое априорное распределение, для которого эта функция риска постоянна. Из (6.8) следует, что это имеет место при $(a+b)^2 = n$ и $2a(a+b) = n$, откуда $a = b = \sqrt{n}/2$. Следовательно, минимаксная оценка θ в данной задаче имеет вид

$$\tilde{\delta}(X) = (X + \sqrt{n}/2)/(n + \sqrt{n}), \quad (6.9)$$

а ее риск равен

$$R(\theta, \tilde{\delta}) = m(\tilde{\delta}) = a^2/(n+a+b)^2 = 1/[4n(1+1/\sqrt{n})^2]. \quad (6.10)$$

Для сравнения рассмотрим риск оптимальной несмещенной оценки $T = X/n$. Имеем

$$R(\theta, T) = D_{\theta} T = \theta(1-\theta)/n.$$

Ее максимальный риск

$$m(T) = \sup_{\theta} \frac{\theta(1-\theta)}{n} = \frac{1}{4n} > m(\delta),$$

[см. (6.10)], но в данном случае имеет место ситуация, подобная изображенной на рис. 6.1: минимаксная оценка (6.9) точнее оценки T только для значений параметра $\theta \in (1/2 \pm \varepsilon_n)$, где

$$\varepsilon_n = \frac{1}{2} \sqrt{1 - (1 + 1/\sqrt{n})^{-2}} \rightarrow 0 \text{ при } n \rightarrow \infty.$$

Рассмотрим теперь интерпретацию задачи проверки гипотез в терминах теории решений. Предположим, что требуется проверить гипотезу $H_0: \theta \in \Theta_0$ при альтернативе $H_1: \theta \in \Theta_1 = \Theta \setminus \Theta_0$. Тогда любой критерий можно интерпретировать как решающее правило с двумя решениями: d_0 (принимать H_0) и d_1 (принимать H_1). Здесь естественно считать, что потеря равна нулю, если выбрано правильное решение. Тогда функция потерь должна удовлетворять условиям $L(\theta, d_0) = 0, \forall \theta \in \Theta_0; L(\theta, d_1) = 0, \forall \theta \in \Theta_1$. Если дополнительно принять, что потери от неправильного решения в любом случае равны единице, т. е. $L(\theta, d_0) = 1, \forall \theta \in \Theta_1; L(\theta, d_1) = 1, \forall \theta \in \Theta_0$, то при такой простой функции потерь для любого решающего правила δ функция риска имеет вид [см. (6.1)]

$$R(\theta, \delta) = \begin{cases} P_{\theta}(\delta(X) = d_1) = P(H_1 | H_0) & \text{при } \theta \in \Theta_0, \\ P_{\theta}(\delta(X) = d_0) = P(H_0 | H_1) & \text{при } \theta \in \Theta_1. \end{cases}$$

Таким образом, значения функции риска совпадают в данном случае с вероятностями ошибок первого и второго рода (см. § 4.1).

§ 6.2. Задача классификации наблюдений

1. Постановка задачи классификации. Будем рассматривать один частный, но представляющий большой практический интерес случай, когда параметрическое множество модели состоит из конечного числа точек: $\Theta = \{\theta_1, \dots, \theta_k\}$, т. е. имеется всего k распределений $F_i(x) = F(x; \theta_i)$, $i = 1, \dots, k$, одно из которых является истинным всякий раз, когда производится наблюдение над X . Пусть по наблюдению над X требуется решить, какое из распределений истинно. Таким образом, множество решений имеет в данном случае вид $D = \{d_1, \dots, d_k\}$, где решение d_i означает, что в качестве истинного следует выбрать распределение F_i , $i = 1, \dots, k$.

Типичный случай, когда возникает подобная задача, можно описать следующим образом. Пусть множество исследуемых объектов разбито на k классов или групп $H_1, \dots, H_k$. Каждый объект характеризуется набором X числовых параметров, которые непосредственно могут быть измерены. Предполагается, что X — случайная величина, а принадлежность объектов к разным

классам выражается в том, что для объектов из класса H_i эта случайная величина имеет распределение F_i , $i = 1, \dots, k$. Задача состоит в том, чтобы по наблюдению над X определить тот класс, к которому принадлежит соответствующий объект, или, что то же самое, какое из распределений $F_1, \dots, F_k$ истинно. Такие задачи называют задачами *дискриминации, классификации или идентификации*. Рассмотрим общие принципы решения таких задач с позиций теории решений и проиллюстрируем их на примерах наиболее важных моделей таких ситуаций.

2. Функция риска в задаче классификации. Пусть $\delta = \delta(x)$ — любое решающее правило в рассматриваемой задаче. Тогда оно порождает разбиение выборочного пространства $\mathcal{X}$ на k взаимно непересекающихся областей $W_1, \dots, W_k$, где

$$W_i = \{x : \delta(x) = d_i\}, \quad i = 1, \dots, k. \quad (6.11)$$

Таким образом, множество W_i включает все такие точки x , когда при наблюдении $X = x$ в качестве истинного выбирается распределение F_i .

Пусть, далее, задана функция потерь $L(\theta, d)$, определяющая потери от неправильной классификации, т. е. заданы числа $L(\theta_i, d_j) = l(j|i)$, $i, j = 1, \dots, k$, где $l(j|i)$ — убыток, который имеет место в случае, когда объект i -го класса отнесен к j -му ($j \neq i$). В данной задаче естественно считать, что потеря равна нулю, если выбрано правильное решение, т. е. $l(i|i) = 0$, $i = 1, \dots, k$. Тогда функция риска (6.1) представляет собой k -мерный вектор риска $R(\delta) = (R_1(\delta), \dots, R_k(\delta))$, где

$$R_i(\delta) = R(\theta_i, \delta) = \sum_{j=1}^k l(j|i) p(j|i) \quad (6.12)$$

и $p(j|i) = P_{\theta_i}(X \in W_j)$ — вероятность того, что объект i -го класса отнесен к j -му классу. Можно сказать, что $R_i(\delta)$ — средние потери, которые имеют место при классификации по правилу δ произвольного объекта i -го класса. Задача состоит в построении оптимального (т. е. с наименьшими потерями) решающего правила δ .

Найдем байесовское и минимаксное правила в данной ситуации.

3. Байесовское решение. Предположим, что известно априорное распределение $\pi = (\pi_1, \dots, \pi_k)$, т. е. произвольный наблюдаемый объект принадлежит i -му классу с вероятностью π_i , $i = 1, \dots, k$. Тогда байесовский риск (6.3) на основании (6.12) равен

$$r(\delta) = \sum_{i=1}^k R_i(\delta) \pi_i = \sum_{j=1}^k \sum_{i=1}^k l(j|i) p(j|i) \pi_i. \quad (6.13)$$

В соответствии с общим алгоритмом построения байесовского решения (см. п. 3 § 6.1), т. е. правила δ , минимизирующего риск (6.13), по формуле (6.4) находим апостериорные вероятности

классов при условии $X = x$:

$$\pi_i(x) = f_i(x) \pi_i / \sum_{s=1}^k \pi_s f_s(x), \quad i = 1, \dots, k. \quad (6.14)$$

Далее, если принять решение отнести объект с характеристикой x к j -му классу (т. е. если $\delta(x) = d_j$), то для такого правила средние условные потери равны [см. (6.6)]

$$\sum_{i=1}^k L(\theta_i, d_j) \pi_i(x) = \sum_{i=1}^k l(j|i) \pi_i f_i(x) / \sum_{s=1}^k \pi_s f_s(x), \quad j = 1, \dots, k. \quad (6.15)$$

Наконец, решение d_j должно быть выбрано так, чтобы минимизировать правую часть (6.15).

Итак, оптимальный способ действия в данном случае состоит в следующем: если наблюдалось $X = x$, то надо определить минимальную сумму

$$h_j(x) = \sum_{i=1}^k l(j|i) \pi_i f_i(x), \quad j = 1, \dots, k. \quad (6.16)$$

Номер j этой суммы и определяет класс, к которому относится наблюдаемый объект. Если при этом минимум достигается при нескольких значениях j , то можно взять любое из них (например, наименьшее).

Итак, имеет место следующее утверждение.

Теорема 6.1. *Оптимальное (байесовское) решающее правило δ^* в задаче классификации определяется следующим разбиением выборочного пространства $\mathcal{X} = W_1^* \cup \dots \cup W_k^*$:*

$$W_i^* = \{x : h_i(x) = \min_{1 \leq j \leq k} h_j(x)\}, \quad i = 1, \dots, k, \quad (6.17)$$

где функции $h_j(x)$ определены в (6.16) и i — минимальное значение индекса, удовлетворяющее указанному условию.

Пусть $l(j|i) = 1$, $j \neq i$; тогда

$$h_j(x) = \sum_{i=1}^k \pi_i f_i(x) = -\pi_j f_j(x) + \sum_{i=1}^k \pi_i f_i(x)$$

и условие в (6.17) принимает вид

$$\pi_i f_i(x) = \max_{1 \leq j \leq k} \pi_j f_j(x). \quad (6.18)$$

Величины $\pi_j f_j(x)$ пропорциональны апостериорным вероятностям классов [см. (6.14)], поэтому в данном случае байесовский принцип сводится к следующему: *относить объект с характеристикой x к тому классу, апостериорная вероятность которого максимальна.* Такой принцип называют *принципом максимума апостериорной вероятности*. Этот принцип используют во всех случаях, когда потери $l(j|i)$ либо неизвестны, либо их трудно оценить числом.

Выделим случай двух классов ($k=2$), т. е. когда объекты классифицируют по альтернативному признаку. Здесь $h_1(x) = I(1|2) \pi_2 f_2(x)$, $h_2(x) = I(2|1) \pi_1 f_1(x)$ и, следовательно, решение d_1 (отнести объект с характеристикой x к первому классу) принимается тогда и только тогда, когда $I(1|2) \pi_2 f_2(x) \leq I(2|1) \pi_1 f_1(x)$. Другими словами, байесовское правило δ^* имеет в данном случае следующий вид:

$$\delta^*(x) = \begin{cases} d_1 & \text{при } x \in W_1^*, \\ d_2 & \text{при } x \in W_2^* = W_1^*, \end{cases}$$

где

$$W_1^* = \left\{ x : \frac{f_2(x)}{f_1(x)} \leq \frac{\pi_1 I(2|1)}{\pi_2 I(1|2)} \right\}. \quad (6.19)$$

а соответствующий ему вектор риска $(R_1(\delta^*), R_2(\delta^*))$ [см. (6.12)] равен

$$I(2|1) P_0, (X \in W_2^*), I(1|2) P_0, (X \in W_1^*). \quad (6.20)$$

4. Минимаксное решение. Если априорные вероятности классов неизвестны, то для построения решающего правила используют минимаксный подход (см. п. 4 § 6.1), в соответствии с которым ищется правило $\bar{\delta}$, минимизирующее $\max_{1 \leq i \leq k} R_i(\delta)$ [см. (6.12)].

В ряде случаев минимаксное правило $\bar{\delta}$ удается построить, определив наименее благоприятное априорное распределение $\pi = (\pi_1, \dots, \pi_k)$, т. е. такое распределение, при котором для соответствующего байесовского решения δ^* компоненты вектора риска $R(\delta^*)$ все одинаковы — в этом случае $\bar{\delta}$ совпадает с решением δ^* . Например, в случае двух классов из (6.20) следует, что $\pi_1(\pi_2 = 1 - \pi_1)$ надо определить из условия

$$I(2|1) P_0, (X \in W_2^*) = I(1|2) P_0, (X \in W_1^*), \quad (6.21)$$

где область W_1^* определена в (6.19) ($W_2^* = W_1^*$). Отсюда следует, что даже в этом частном случае для общих распределений F_i задача построения минимаксного решения достаточно сложна. В следующих параграфах эти вопросы рассмотрены для некоторых конкретных распределений.

§ 6.3. Классификация наблюдений в случае двух нормальных классов

1. Байесовский подход. Предположим, что X — нормальный вектор размерности r , распределенный для объектов из класса H_1 по закону $\mathcal{N}(\mu^{(1)}, A)$, а для объектов из класса H_2 — по закону $\mathcal{N}(\mu^{(2)}, A)$. Таким образом, здесь имеет место случай двух нормальных классов, различающихся только средними значениями (общая ковариационная матрица A далее предполагается невырожденной). В данном случае функции $f_i(x)$

имеют вид

$$f_i(x) = \frac{1}{(2\pi)^{r/2} |A|} \exp \left\{ -\frac{1}{2} (x - \mu^{(i)})' A^{-1} (x - \mu^{(i)}) \right\}, \quad i = 1, 2;$$

а их отношение —

$$\frac{f_2(x)}{f_1(x)} = \exp \left\{ \frac{1}{2} (x - \mu^{(1)})' A^{-1} (x - \mu^{(1)}) - \frac{1}{2} (x - \mu^{(2)})' A^{-1} (x - \mu^{(2)}) \right\}.$$

Введем вектор

$$a = A^{-1} (\mu^{(1)} - \mu^{(2)}) \quad (6.22)$$

и константу $c = \ln \frac{\pi_2 I(1|2)}{\pi_1 I(2|1)}$. Простые преобразования позволяют записать область наилучшей классификации W_1^* , определенную в (6.19), в виде

$$W_1^* = \{x : a'x - (1/2) a'(\mu^{(1)} + \mu^{(2)}) \geq c\}. \quad (6.23)$$

Линейную функцию наблюдений $\varphi(x) = a'x$ называют *дискриминантной функцией*.

Таким образом, области наилучшей классификации W_1^* и W_2^* определяются в данном случае дискриминантной функцией $\varphi(x)$: наблюдение x относится к первому классу тогда и только тогда, когда $\varphi(x) \geq c_1$, где константа $c_1 = \frac{1}{2} a'(\mu^{(1)} + \mu^{(2)}) + c$ определяется параметрами модели, априорным распределением и заданными потерями. Если потери неизвестны или их трудно оценить числом, то применяют указанное правило классификации, где $I(1|2) = I(2|1) = 1$.

2. Минимаксный подход. Найдем минимаксное решение $\bar{\delta}$ (которое используют в случае, когда априорные вероятности неизвестны). Для этого вычислим вероятности ошибочных классификаций $p(j|i)$ для произвольного байесовского правила (6.23) и найдем из условия (6.21) наименьшее благоприятное априорное распределение. Соответствующее этому априорному распределению правило (6.23) и является искомым минимаксным решением $\bar{\delta}$.

Введем случайную величину $Y = a'X - \frac{1}{2} a'(\mu^{(1)} + \mu^{(2)})$ и найдем ее распределение при «гипотезах» H_1 и H_2 . Так как X — нормальный вектор при обеих гипотезах, то Y , как линейная функция от нормального вектора, в любом случае также нормальная случайная величина; следовательно, достаточно вычислить только ее первые и вторые моменты. Имеем

$$E(Y|H_i) = a'\mu^{(i)} - \frac{1}{2} a'(\mu^{(1)} + \mu^{(2)}) = \begin{cases} \alpha/2 & \text{при } i=1, \\ -\alpha/2 & \text{при } i=2, \end{cases}$$

где [см. (6.22)]

$$\alpha = a'(\mu^{(1)} - \mu^{(2)}) = (\mu^{(1)} - \mu^{(2)})' A^{-1} (\mu^{(1)} - \mu^{(2)}); \quad (6.24)$$

$$D(Y|H_i) = D(a'X|H_i) = a'Aa = \alpha, \quad i = 1, 2.$$

Таким образом,

$$\mathcal{L}(Y|H_1) = \phi^*\left(\frac{\alpha}{2}, \alpha\right), \quad \mathcal{L}(Y|H_2) = \phi^*\left(-\frac{\alpha}{2}, \alpha\right),$$

где α определено в (6.24).

Вычислим вероятности ошибочной классификации. Из (6.23) имеем:

$$p(2|1) = P(Y < c | H_1) = \frac{1}{\sqrt{2\pi\alpha}} \int_{-\infty}^c e^{-\frac{1}{2\alpha}\left(t - \frac{\alpha}{2}\right)^2} dt = \Phi\left(\frac{c - \alpha/2}{\sqrt{\alpha}}\right),$$

$$p(1|2) = P(Y \geq c | H_2) = \frac{1}{\sqrt{2\pi\alpha}} \int_c^{\infty} e^{-\frac{1}{2\alpha}\left(t + \frac{\alpha}{2}\right)^2} dt = \Phi\left(-\frac{c + \alpha/2}{\sqrt{\alpha}}\right).$$

Следовательно, уравнение (6.21) для определения наименее благоприятного априорного распределения $(\pi_1, 1 - \pi_1)$, или, что эквивалентно, константы c , имеет вид

$$l(2|1) \Phi\left(\frac{c - \alpha/2}{\sqrt{\alpha}}\right) = l(1|2) \Phi\left(-\frac{c + \alpha/2}{\sqrt{\alpha}}\right). \quad (6.25)$$

Итак, если априорные вероятности классов H_1 и H_2 неизвестны, то минимаксные области классификации определяются с помощью формулы (6.23), где константа c выбирается из условия (6.25).

Отметим, что если $l(1|2) = l(2|1)$, то решением уравнения (6.25) является $c = 0$. В этом случае вероятность ошибочной классификации произвольного объекта равна $\Phi(-\sqrt{\alpha/2})$.

Величину α , определенную в (6.24), называют *расстоянием Махаланобиса* между распределениями $\phi^*(\mu^{(1)}, A)$ и $\phi^*(\mu^{(2)}, A)$. Из предыдущего следует, что чем более далекими (в метрике α) являются гипотезы H_1 и H_2 , тем меньше вероятность ошибочной классификации правила (6.23).

§ 6.4. Классификация нормальных наблюдений. Общий случай

1. Байесовский подход. В этом параграфе будет применена изложенная выше (см. § 6.2) теория к общему случаю нескольких классов, заданных многомерными нормальными распределениями. Предположим, что эти распределения различаются только своими средними, и пусть $\phi^*(\mu^{(i)}, A)$ — распределение наблюдений X для объектов из класса H_i ($i = 1, \dots, k$). Кроме того, будем считать, что все цены ошибочных классификаций равны. Тогда байесовское решение (при заданных априорных вероятностях $\pi_i = P(H_i)$, $i = 1, \dots, k$) можно получить с помощью принципа максимума апостериорных вероятностей [см. (6.18)]; оно соответствует случаю, когда области классификации (6.11) имеют вид $W_i = \{x : f_i(x)/f_j(x) \geq \pi_j/\pi_i, j = 1, \dots, k, j \neq i\}$, $i = 1, \dots, k$, или

$$W_i = \{x : u_{ij}(x) \geq c_i - c_j, j = 1, \dots, k, j \neq i\}, \quad (6.26)$$

где $c_i = \ln(1/\pi_i)$, $i = 1, \dots, k$ и

$$u_{ij}(x) = \ln(f_i(x)/f_j(x)) = a'_{ij}x - (1/2) a'_{ij}(\mu^{(i)} + \mu^{(j)}), \quad (6.27)$$

$$a_{ij} = A^{-1}(\mu^{(i)} - \mu^{(j)}), \quad i, j = 1, \dots, k, i \neq j. \quad (6.28)$$

Отметим, что каждая классификационная функция $u_{ij}(x)$ связана только с i -й и j -й совокупностями: $u_{ij}(x) = -u_{ji}(x)$, при этом все функции являются линейными относительно результатов измерений x . Следовательно, области W_i ограничены гиперплоскостями.

2. Минимаксный подход. Если априорные вероятности классов неизвестны, то области классификации будем искать в виде (6.26), где неопределенные константы $c_i > 0$, $i = 1, \dots, k$, следует выбирать из условия равенства всех компонент вектора риска (6.12), которое в данном случае (при одинаковых $l(j|i)$) приводит к равенствам

$$p(1|1) = p(2|2) = \dots = p(k|k), \quad (6.29)$$

где $p(i|i) = \int_{W_i} f_i(x) dx = P(u_{ij}(X) \geq c_i - c_j, \quad j = 1, \dots, k, \quad j \neq i | H_i)$ — веро-

ятность правильной классификации объектов i -го класса. Области $W_1, \dots, W_k$ с определенными таким образом константами $c_1, \dots, c_k$ задают минимаксное правило классификации.

Из (6.29) следует, что задача сводится к отысканию распределения случайного вектора $U_i = (u_{ij}(X))$, $j = 1, \dots, k, \quad j \neq i$ при гипотезе H_i . Введем матрицу A_i , составленную из вектор-столбцов a_{ij} , $j = 1, \dots, k, \quad j \neq i$, определенных в (6.28). Тогда из формулы (6.27) следует, что вектор U_i можно получить из X с помощью линейного преобразования вида $U_i = A_i' X + b_i$. Следовательно,

$$\mathcal{L}(U_i | H_i) = \frac{1}{(2\pi)^{r/2}} \exp\left\{-\frac{1}{2} (A_i' \mu^{(i)} + b_i, A_i' A A_i)\right\}. \quad (6.30)$$

Отсюда имеем, что j -я координата вектора средних равна

$$a_{ij}' \mu^{(i)} - (1/2) a_{ij}' (\mu^{(i)} + \mu^{(j)}) = (1/2) a_{ij}' (\mu^{(i)} - \mu^{(j)}) = \alpha_{ij}/2,$$

где $\alpha_{ij} = (\mu^{(i)} - \mu^{(j)})' A^{-1} (\mu^{(i)} - \mu^{(j)})$ — расстояние Махаланобиса между i -м и j -м классами, а (j, s) -й элемент α_{ij} матрицы ковариаций, равный скалярному произведению j -й строки матрицы A_i' (т. е. вектора a_{ij}') на s -й столбец матрицы AA_i (т. е. $Aa_{is} = \mu^{(i)} - \mu^{(s)}$), имеет вид $\alpha_{ij/s} = (\mu^{(i)} - \mu^{(j)})' A^{-1} (\mu^{(i)} - \mu^{(s)})$. Тем самым распределение (6.30) полностью определено. Оно является невырожденным ($|A_i' A A_i| \neq 0$) тогда и только тогда, когда A_i — матрица полного ранга, т. е. когда $\text{rang } A_i = k - 1$. Это имеет место, в частности, если векторы средних значений $\mu^{(1)}, \dots, \mu^{(k)}$ линейно независимы: $r = \dim X \geq k - 1$. В этом случае, согласно (6.29),

$$p(i|i) = \int_{c_i - c_1}^{\infty} \dots \int_{c_i - c_k}^{\infty} g_i(u) du,$$

где $g_i(u)$ — плотность распределения, определенного в (6.30).

В качестве примера рассмотрим случай трех классов ($k=3$), задаваемых двумерными нормальными распределениями ($r=2$). В данном случае области W_i , $i=1, 2, 3$, имеют соответственно вид $W_1 = \{x: u_{12}(x) \geq c_1 - c_2, u_{13}(x) \geq c_1 - c_3\}$, $W_2 = \{x: u_{12}(x) < c_1 - c_2, u_{23}(x) \geq c_2 - c_3\}$, $W_3 = \{x: u_{13}(x) < c_1 - c_3, u_{23}(x) < c_2 - c_3\}$. Эти области должны исчерпывать все пространство R^2 , поэтому линии, задаваемые уравнениями $u_{12}(x) = c_1 - c_2$, $u_{13}(x) = c_1 - c_3$ и $u_{23}(x) = c_2 - c_3$, должны пересечься в точке, а равенство вероятностей $p(1|1) = p(2|2) = p(3|3)$ однозначно определяет разности $c_1 - c_j$. Для их определения можно воспользоваться таблицами двумерного нормального распределения. Соответствующие области минимаксного правила классификации изображены на рис. 6.2.

3. Классификация наблюдений при наличии неизвестных параметров. Выше предполагалось, что все допустимые распределения наблюдений полностью известны. Однако в приложениях эти распределения часто известны лишь с точностью до значений некоторых параметров (например, $\mu^{(i)}$, $i=1, \dots, k$, или A , или одновременно всех этих параметров). В таких случаях

можно решать задачи классификации, если дополнительно известно, что произвольно взятые n_i объектов из класса H_i имели характеристики $x_1^{(i)}, \dots, x_{n_i}^{(i)}$ ($i=1, \dots, k$). Другими словами, предполагается, что имеются обучающие выборки $(x_1^{(i)}, \dots, x_{n_i}^{(i)})$, $i=1, \dots, k$ из соответствующих распределений $\mathcal{N}(\mu^{(i)}, A)$, $i=1, \dots, k$. Эти выборки можно использовать для оценки соответствующих неизвестных параметров распределений и, заменив неизвестные параметры их оценками, поступать далее, как и в случае полностью известных распределений.

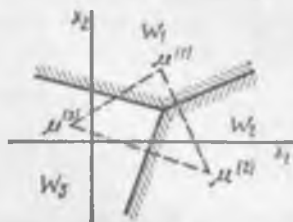


Рис. 6.2

Если неизвестны средние $\mu^{(i)}$, то их заменяют оценками $\hat{\mu}^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j^{(i)}$ —

средними арифметическими выборок (см. пример 2.18). Ковариационная матрица A (когда она неизвестна) оценивается выборочной ковариационной матрицей; при этом так как матрица A — общая для всех классов, то для ее оценивания следует использовать информацию, доставляемую всеми выборками. Эту информацию объединяют следующим образом. Введем выборочные матрицы

$$\hat{A}^{(i)} = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_j^{(i)} - \hat{\mu}^{(i)})(x_j^{(i)} - \hat{\mu}^{(i)})', \quad i=1, \dots, k,$$

каждая из которых является несмещенной оценкой матрицы A (см. задачу 2.29). Тогда

$$E \sum_{i=1}^k (n_i - 1) \hat{A}^{(i)} = \left(\sum_{i=1}^k n_i - k \right) A,$$

поэтому матрица $\hat{A} = \sum_{i=1}^k (n_i - 1) \hat{A}^{(i)} / \left(\sum_{i=1}^k n_i - k \right)$, построенная с учетом всех данных, также является несмещенной оценкой A .

Построив эти оценки параметров распределений, далее можно ввести оценки

$$\hat{u}_{ij}(x) = \hat{a}_{ij}'x - \frac{1}{2} \hat{a}_{ij}'(\hat{\mu}^{(i)} + \hat{\mu}^{(j)}), \quad \hat{a}_{ij} = \hat{A}^{-1}(\hat{\mu}^{(i)} - \hat{\mu}^{(j)})$$

для классификационных функций (6.27) и использовать их для построения приближенных областей классификации $\hat{W}_i = \{x: \hat{u}_{ij}(x) \geq c_j - c_i, j=1, \dots, k, j \neq i\}$, заменяющих в данном случае области (6.26).

В качестве обоснования этого метода приведем следующие «асимптотические» рассуждения. Предположим, что объемы обучающих выборок велики ($n_i \rightarrow \infty$, $i=1, \dots, k$). Тогда оценки $\hat{\mu}^{(i)}$, $i=1, \dots, k$, и $\hat{A}$ как арифметические средние сходятся по вероятности соответственно к $\mu^{(i)}$, $i=1, \dots, k$ и A . Отсюда следует, что $\hat{a}_{ij}$ сходится по вероятности к a_{ij} , а $\hat{a}_{ij}'(\hat{\mu}^{(i)} + \hat{\mu}^{(j)})$ — к $a_{ij}'(\mu^{(i)} + \mu^{(j)})$. Следовательно, предельное распределение $\hat{u}_{ij}(X)$ совпадает с распределением $u_{ij}(X)$, поэтому для достаточно больших обучающих выборок функции $\hat{u}_{ij}(x)$ можно использовать так же, как если бы были точно известны распределения совокупностей.

Задачи

1. Пусть в примере 6.1 параметр θ имеет априорное распределение $\pi(\theta_1) = 2/3$, $\pi(\theta_2) = 1/3$. Найти байесовское решение.

2. Пусть $\mathcal{L}(X) = Bi(3, \theta)$, $\Theta = \{1/100, 1/10\}$, $D = \{d_1, d_2\}$, а функция потерь $L(\theta_i, d_j)$ задана таблицей

	d_1	d_2
θ_1	0	2
θ_2	1	0

Рассмотрев решающие функции

$$\delta_k(x) = \begin{cases} d_1 & \text{при } x=0, 1, \dots, k-1, \\ d_2 & \text{в остальных случаях,} \end{cases} \quad k=1, 2, 3,$$

показать, что функция $\delta_1(x)$ дает минимаксное решение задачи. Построить байесовское решение для заданного априорного распределения.

3. Пусть $\mathcal{L}(X) = Bi(1, \theta)$, $\Theta = \{1/10, 2/10\}$, $D = \{d_1, d_2\}$ и функция потерь задана таблицей

	d_1	d_2
θ_1	0	1
θ_2	2	0

Определить минимаксную решающую функцию среди функций

$$\delta_k(x) = \begin{cases} d_1 & \text{при } x=0, 1, \dots, k-1, \\ d_2 & \text{при } x=k, k+1, \dots, \end{cases} \quad k=1, 2, 3, 4.$$

Построить байесовское решение для заданного априорного распределения.

4. Проверить, что полученная в примере 6.2 байесовская оценка параметра θ совпадает с апостериорным средним этого параметра.

5. Пусть $\mathcal{L}_\theta(X) = \Pi(\alpha\theta)$ и $\mathcal{L}(\theta) = \Gamma(\gamma, \lambda)$. Показать, что апостериорное распределение θ при $X=x$ есть $\mathcal{L}(\theta|x) = \Gamma\left(\frac{\gamma}{1+\alpha\gamma}, \lambda+x\right)$.

6. Пусть $\mathcal{L}_\theta(X) = \bar{Bi}(r, \theta)$, а параметр θ имеет бета-распределение с параметрами (a, b) (см. пример 6.2). Показать, что апостериорное распределение θ при условии $X=x$ является также бета-распределением с параметрами $(a+x, b+r)$.

7. Пусть $X=(X_1, \dots, X_N)$ — выборка из распределения $R(0, \theta)$ и при этом θ — случайная величина с плотностью $\pi(\theta) = \alpha\theta^\alpha/\theta^{\alpha+1}$, $\theta \geq a$ ($a, \alpha > 0$) [такое распределение называют *распределением Парето* с параметрами (a, α)]. Показать, что апостериорная плотность $\pi(\theta|x)$ при условии, что $X=x=(x_1, \dots, x_N)$, есть также плотность Парето с параметрами $(\max(a, x_1, \dots, x_N), \alpha+n)$.

8. Пусть случайный вектор $v=(v_1, \dots, v_N)$ имеет полиномиальное распределение $M(n, p=(p_1, \dots, p_N))$, где n — известное целое число, а компоненты вектора p случайны и их совместное распределение является *распределением Дирихле* с параметрами $\alpha=(\alpha_1, \dots, \alpha_N)$, $\alpha_i > 0$, $i=1, \dots, N$, т. е. их совместная плотность распределения

$$f(p_1, \dots, p_N; \alpha) = \begin{cases} \frac{\Gamma(\alpha_1 + \dots + \alpha_N)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_N)} p_1^{\alpha_1-1} \dots p_N^{\alpha_N-1} & \text{при } p_1 + \dots + p_N = 1, \\ 0 & \text{в остальных точках.} \end{cases}$$

Показать, что апостериорное распределение p при условии $v=h=(h_1, \dots, h_N)$ — также распределение Дирихле с параметрами $\alpha+h$.

ПРИЛОЖЕНИЯ

1. Нормальное распределение

Квантили распределения: $p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_p} e^{-x^2/2} dx$

p	u_p	p	u_p	p	u_p
0,50	0,000	0,68	0,468	0,86	1,080
0,51	0,025	0,69	0,496	0,87	1,126
0,52	0,050	0,70	0,524	0,88	1,175
0,53	0,075	0,71	0,553	0,89	1,227
0,54	0,100	0,72	0,583	0,90	1,282
0,55	0,126	0,73	0,613	0,91	1,341
0,56	0,151	0,74	0,643	0,92	1,405
0,57	0,176	0,75	0,674	0,93	1,476
0,58	0,202	0,76	0,706	0,94	1,555
0,59	0,228	0,77	0,739	0,95	1,645
0,60	0,253	0,78	0,772	0,96	1,751
0,61	0,279	0,79	0,806	0,97	1,881
0,62	0,305	0,80	0,842	0,98	2,054
0,63	0,332	0,81	0,878	0,99	2,326
0,64	0,358	0,82	0,915	0,999	3,090
0,65	0,385	0,83	0,954	0,9999	3,720
0,66	0,412	0,84	0,994	0,99999	4,265
0,67	0,440	0,85	1,036		

2. Распределение Пуассона

Значения функции $\sum_{k=x}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$

$\lambda \backslash x$	0.1	0.2	0.3	0.4	0.5	0.6
0	1,000	1,000	1,000	1,000	1,000	1,000
1	0,095	0,181	0,259	0,330	0,394	0,451
2	005	018	037	062	090	122
3		001	003	008	014	023
4				001	002	003

$\lambda \backslash x$	0.7	0.8	0.9	1.0	2.0	3.0
0	1,000	1,000	1,000	1,000	1,000	1,000
1	0,503	0,551	0,593	0,632	0,865	0,950
2	156	191	228	264	594	801
3	034	047	063	080	323	577
4	006	009	014	019	143	353
5	001	001	002	004	053	185
6				001	018	084
7					005	034
8					001	012

$\lambda \backslash x$	4.0	5.0	6.0	7.0	8.0	9.0
0	1,000	1,000	1,000	1,000	1,000	1,000
1	0,982	0,993	0,998	0,999	1,000	1,000
2	908	960	983	924	0,997	0,999
3	762	875	938	970	986	994
4	567	735	849	918	958	979
5	371	560	715	827	900	945
6	215	384	554	699	809	884
7	111	238	394	550	687	793
8	051	133	256	401	547	676
9	021	068	153	271	408	544
10	008	032	084	170	283	413
11	003	014	043	099	184	294
12	001	005	020	053	112	197
13		002	008	027	068	124
14		001	004	013	034	074
15			001	006	017	042
16			001	002	008	022
17				001	004	011
18					002	005
19					001	002
20						001

3. Биномиальное распределение

95 %-ные доверительные пределы (θ_1 , θ_2) для параметра θ :

$$P_0(\theta_1 < \theta < \theta_2) = 0,95$$

$\begin{matrix} n-k \\ k \end{matrix}$	0	1	2	3	4	5	6	7	8	9	10	11	12
0	— —	0,98 0,00	0,84 0,00	0,71 0,00	0,60 0,00	0,52 0,00	0,46 0,00	0,41 0,00	0,37 0,00	0,34 0,00	0,31 0,00	0,29 0,00	0,27 0,00
1	1,00 0,03	0,99 0,01	0,91 0,01	0,81 0,01	0,72 0,01	0,64 0,00	0,58 0,00	0,53 0,00	0,48 0,00	0,45 0,00	0,41 0,00	0,39 0,00	0,36 0,00
2	1,00 0,16	0,99 0,09	0,93 0,07	0,85 0,05	0,78 0,04	0,71 0,04	0,65 0,03	0,60 0,03	0,56 0,03	0,52 0,02	0,48 0,02	0,45 0,02	0,43 0,02
3	1,00 0,29	0,99 0,19	0,95 0,15	0,88 0,12	0,82 0,10	0,76 0,09	0,70 0,08	0,65 0,07	0,61 0,06	0,57 0,06	0,54 0,05	0,51 0,05	0,48 0,04
4	1,00 0,40	1,00 0,29	0,96 0,22	0,90 0,18	0,84 0,16	0,79 0,14	0,74 0,12	0,69 0,11	0,65 0,10	0,61 0,09	0,58 0,08	0,55 0,08	0,52 0,07
5	1,00 0,48	1,00 0,36	0,96 0,29	0,92 0,25	0,86 0,21	0,81 0,19	0,77 0,17	0,72 0,15	0,68 0,14	0,65 0,13	0,62 0,12	0,59 0,11	0,56 0,10
6	1,00 0,54	1,00 0,42	0,97 0,35	0,93 0,30	0,88 0,26	0,83 0,23	0,79 0,21	0,75 0,19	0,71 0,18	0,68 0,16	0,65 0,15	0,62 0,14	0,59 0,13
7	1,00 0,59	1,00 0,47	0,97 0,40	0,93 0,35	0,89 0,31	0,85 0,28	0,81 0,25	0,77 0,23	0,73 0,21	0,70 0,20	0,67 0,18	0,64 0,17	0,62 0,16
8	1,00 0,63	1,00 0,52	0,98 0,44	0,94 0,39	0,90 0,35	0,86 0,32	0,82 0,29	0,79 0,27	0,75 0,25	0,72 0,23	0,69 0,22	0,67 0,20	0,64 0,19
9	1,00 0,66	1,00 0,56	0,98 0,48	0,95 0,43	0,91 0,39	0,87 0,35	0,84 0,32	0,80 0,30	0,77 0,28	0,74 0,26	0,71 0,24	0,69 0,23	0,66 0,22
10	1,00 0,69	1,00 0,59	0,98 0,52	0,95 0,46	0,92 0,42	0,88 0,38	0,85 0,35	0,82 0,33	0,79 0,31	0,76 0,29	0,73 0,27	0,70 0,26	0,68 0,24
11	1,00 0,72	1,00 0,62	0,98 0,57	0,95 0,49	0,92 0,45	0,89 0,41	0,86 0,38	0,83 0,36	0,80 0,34	0,77 0,32	0,74 0,30	0,72 0,28	0,69 0,27
12	1,00 0,74	1,00 0,64	0,98 0,57	0,96 0,52	0,93 0,48	0,90 0,44	0,87 0,41	0,84 0,38	0,81 0,36	0,78 0,34	0,76 0,32	0,73 0,31	0,71 0,29

П р и м е ч а н и е. Значения θ_1 набраны в первых строках, значения θ_2 — по вторым.

4. Распределение $\chi^2(n)$

Квантили распределения: $p = \int_0^{\chi_{p,n}} k_{\chi^2}(x) dx = \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^{\chi_{p,n}} x^{n/2-1} e^{-x/2} dx$

$\begin{array}{c} p \\ n \end{array}$	0.1	0.3	0.5	0.7	0.9	0.95	0.999	0.9999
1	0,016	0,148	0,455	1,07	2,71	3,84	6,63	10,8
2	0,211	0,713	1,39	2,41	4,61	5,99	9,21	13,8
3	0,584	1,42	2,37	3,67	6,25	7,82	11,3	16,3
4	1,06	2,20	3,36	4,88	7,78	9,49	13,3	18,5
5	1,61	3,00	4,35	6,06	9,24	11,1	15,1	20,5
6	2,20	3,83	5,35	7,23	10,6	12,6	16,8	22,5
7	2,83	4,67	6,35	8,38	12,0	14,1	18,5	24,3
8	3,49	5,53	7,34	9,52	13,4	15,5	20,1	26,1
9	4,17	6,39	8,34	10,7	14,7	16,9	21,7	27,9
10	4,87	7,27	9,34	11,8	16,0	18,3	23,2	29,6
11	5,58	8,15	10,3	12,9	17,3	19,7	24,7	31,3
12	6,30	9,03	11,3	14,0	18,5	21,0	26,2	32,9
13	7,04	9,93	12,3	15,1	19,8	22,4	27,7	34,5
14	7,79	10,08	13,3	16,2	21,1	23,7	29,1	36,1
15	8,55	11,7	14,3	17,3	22,3	25,0	30,6	37,7
16	9,31	12,6	15,3	18,4	23,5	26,3	32,0	39,3
17	10,09	13,5	16,3	19,5	24,8	27,6	33,4	40,8
18	10,9	14,4	17,3	20,6	26,0	28,9	34,8	42,3
19	11,7	15,4	18,3	21,7	27,2	30,1	36,2	43,8
20	12,4	16,3	19,3	22,8	28,4	31,4	37,6	45,3
21	13,2	17,2	20,3	23,9	29,6	32,7	38,9	46,8
22	14,0	18,1	21,3	24,9	30,8	33,9	40,3	48,3
23	14,8	19,0	22,3	26,0	32,0	35,2	41,6	49,7
24	15,7	19,9	23,3	27,1	33,2	36,4	43,0	51,2
25	16,5	20,9	24,3	28,2	34,3	37,7	44,3	52,6
26	17,3	21,8	25,3	29,2	35,6	38,9	45,6	54,1
27	18,1	22,7	26,3	30,3	36,7	40,1	47,0	55,5
28	18,9	23,6	27,3	31,4	37,9	41,3	48,3	56,9
29	19,8	24,6	28,3	32,5	39,1	42,6	49,6	58,3
30	20,6	25,5	29,3	33,5	40,3	43,8	50,9	59,7

5. Распределение Стьюдента $S(n)$

Значения функции $t_{\gamma, n}$:

$$\frac{1+\gamma}{2} = \int_{-\infty}^{t_{\gamma, n}} s_n(x) dx = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \int_{-\infty}^{t_{\gamma, n}} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} dx$$

$n \backslash \gamma$	0,9	0,95	0,98	0,99
1	6,314	12,706	31,821	63,657
2	2,920	4,303	6,965	9,925
3	2,353	3,182	4,541	5,841
4	2,132	2,776	3,747	4,604
5	2,015	2,571	3,365	4,032
6	1,943	2,447	3,143	3,707
7	1,895	2,365	2,998	3,499
8	1,860	2,306	2,896	3,355
9	1,833	2,262	2,821	3,250
10	1,812	2,228	2,764	3,169
12	1,782	2,179	2,681	3,055
14	1,761	2,145	2,625	2,977
16	1,746	2,120	2,584	2,921
18	1,734	2,101	2,552	2,878
20	1,725	2,086	2,528	2,845
22	1,717	2,074	2,508	2,819
24	1,711	2,064	2,492	2,797
26	1,706	2,056	2,479	2,779
28	1,701	2,048	2,467	2,763
30	1,697	2,042	2,457	2,750
∞	1,645	1,960	2,326	2,576

6. Распределение Снедекора $S(n_1, n_2)$

Значения функции F_{p, n_1, n_2} :

$$p = \int_0^{F_{p, n_1, n_2}} f_{n_1, n_2}(x) dx =$$

$$= \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \int_0^{F_{p, n_1, n_2}} x^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} dx$$

при $p=0,95$ и $p=0,99$. «Левые» границы доверительных интервалов находятся из условия $F_{1-p, n_1, n_2} = F_{p, n_1, n_2}^{-1}$

$n_1 \backslash n_2$	1	2	3	4	6	8	10	12	20	50	100
1	161 4052	200 4999	216 5403	225 5625	234 5859	239 5981	242 6056	244 6106	248 6208	252 6302	254 6334
2	18,51 98,49	19,00 99,01	19,16 99,17	19,25 99,25	19,33 99,33	19,37 99,36	19,39 99,40	19,41 99,42	19,44 99,45	19,47 99,48	19,49 99,49
3	10,13 34,12	9,55 30,81	9,28 29,46	9,12 28,71	8,94 27,91	8,84 27,49	8,78 27,23	8,74 27,05	8,66 26,69	8,58 26,35	8,56 26,23
4	7,71 21,20	6,94 18,00	6,59 16,69	6,39 15,98	6,16 15,21	6,04 14,80	5,96 14,54	5,91 14,37	5,80 14,02	5,70 13,69	5,66 13,57
5	6,61 16,26	5,79 13,27	5,41 12,06	5,19 11,39	4,95 10,67	4,82 10,27	4,74 10,05	4,68 9,89	4,56 9,55	4,44 9,24	4,40 9,13
6	5,99 13,74	5,14 10,92	4,76 9,78	4,53 9,15	4,28 8,47	4,15 8,10	4,06 7,87	4,00 7,72	3,87 7,39	3,75 7,09	3,71 6,99
8	5,32 11,26	4,46 8,65	4,07 7,59	3,84 7,01	3,58 6,37	3,44 6,03	3,34 5,82	3,28 5,67	3,15 5,36	3,03 5,06	2,98 4,96
10	4,96 10,04	4,10 7,56	3,71 6,55	3,48 5,99	3,22 5,39	3,07 5,06	2,97 4,85	2,91 4,71	2,77 4,41	2,64 4,12	2,59 4,01
12	4,75 9,33	3,88 6,93	3,49 5,95	3,26 5,41	3,00 4,82	2,85 4,50	2,76 4,30	2,69 4,16	2,54 3,86	2,40 3,56	2,35 3,46
20	4,35 8,10	3,49 5,85	3,10 4,94	2,87 4,43	2,60 3,87	2,45 3,56	2,35 3,37	2,28 3,23	2,12 2,94	1,96 2,63	1,90 2,53
30	4,17 7,56	3,32 5,39	2,92 4,51	2,69 4,02	2,42 3,47	2,27 3,17	2,16 2,98	2,09 2,84	1,93 2,55	1,76 2,24	1,69 2,13
50	4,03 7,17	3,18 5,06	2,79 4,20	2,56 3,72	2,29 3,18	2,13 2,88	2,02 2,70	1,95 2,56	1,78 2,26	1,60 1,94	1,52 1,82
100	3,94 6,90	3,09 4,82	2,70 3,98	2,46 3,51	2,19 2,99	2,03 2,69	1,92 2,51	1,85 2,36	1,68 2,06	1,48 1,73	1,39 1,59
200	3,89 6,76	3,04 4,71	2,65 3,88	2,41 3,41	2,14 2,90	1,98 2,60	1,87 2,41	1,80 2,28	1,62 1,97	1,42 1,62	1,32 1,48
1000	3,85 6,66	3,00 4,62	2,61 3,80	2,38 3,34	2,10 2,82	1,95 2,53	1,84 2,34	1,76 2,20	1,58 1,89	1,36 1,54	1,26 1,38

Примечания: 1. Значения $F_{0,95; n_1, n_2}$ набраны в первых строках, значения $F_{0,99; n_1, n_2}$ — во вторых. 2. Здесь n_1 — степени свободы для большей дисперсии; n_2 — степени свободы для меньшей дисперсии.

7. Критерий Колмогорова

Значения функции λ_p : $p = P(\mathcal{D}_n = \sup_x |F_n(x) - F(x)| > \lambda_p)$

$n \backslash p$	0,10	0,05	0,01	$n \backslash p$	0,10	0,05	0,01
1	0,950	0,975	0,995	19	0,271	0,301	0,361
2	776	842	929	20	265	294	352
3	636	708	829	25	238	264	317
4	565	624	734	30	218	242	290
5	509	563	669	35	202	224	269
6	468	519	617	40	189	210	252
7	436	483	576	45	179	198	238
8	410	454	542	50	170	188	226
9	387	430	513	55	162	180	216
10	369	409	489	60	155	172	207
11	352	391	468	65	149	166	199
12	338	375	449	70	144	160	192
13	325	361	432	75	139	154	185
14	314	349	418	80	135	150	179
15	304	338	404	85	131	145	174
16	295	327	392	90	127	141	169
17	286	318	381	95	124	137	165
18	279	309	371	100	121	134	161

8. Критерий Смирнова

Значения вероятности $P(\mathcal{D}_{n\lambda} \leq k/n)$, где $\mathcal{D}_{n\lambda} = \sup_x |F_{1n}(x) - F_{2n}(x)|$.

$n \backslash k$	1	2	3	4	5	6	7	8	9	10	11
1	1,000										
2	0,667	1,000									
3	400	0,900	1,000								
4	229	771	0,971	1,000							
5	127	643	921	0,992	1,000						
6	069	526	857	974	0,998	1,000					
7	037	425	788	947	992	0,999	1,000				
8	020	340	717	913	981	998	1,000	1,000			
9	011	270	648	874	966	993	0,999	1,000	1,000		
10	006	213	582	832	948	988	998	1,000	1,000	1,000	
11	003	167	521	789	925	979	996	0,999	1,000	1,000	1,000
12	002	131	464	744	900	969	992	999	1,000	1,000	1,000
13	001	102	412	700	874	956	987	997	1,000	1,000	1,000
14	000	079	365	657	845	941	981	995	0,999	1,000	1,000
15	000	062	322	614	816	925	974	992	998	1,000	1,000
16	000	048	284	574	785	907	965	989	997	1,000	1,000
17	000	037	249	535	755	888	955	984	995	0,999	0,999
18	000	028	219	497	725	868	944	979	993	998	999
19	000	022	192	462	694	847	932	973	991	997	999
20	000	017	168	429	664	825	919	966	988	996	999
21	000	013	147	397	635	804	905	959	984	995	998
22	000	010	128	368	606	782	891	951	980	993	998
23	000	008	112	340	578	759	876	942	975	991	997
24	000	006	098	314	551	737	860	932	970	988	996
25	000	004	085	290	525	715	844	922	964	985	994
26	000	003	074	267	499	693	828	911	958	982	993
27	000	003	064	246	474	671	811	900	952	978	991

1. Андерсон Т. Введение в многомерный статистический анализ. М., 1963.
2. Болышев Л. Н., Смирнов Н. В. Таблицы математической статистики М., 1965.
3. Вальд А. Последовательный анализ. М., 1960.
4. Гаек Я., Шидак Э. Теория ранговых критериев. М., 1971.
5. Де Гроот М. Оптимальные статистические решения. М., 1974.
6. Джейвид Г. Порядковые статистики. М., 1979.
7. Закс Ш. Теория статистических выводов. М., 1975.
8. Ивченко Г. И., Медведев Ю. И. Разделимые статистики и проверка гипотез для группированных данных. — Теория вероятностей и ее применения, т. XXV, 1980, вып. 3.
9. Кендалл М., Стьюарт А. Теория распределений. М., 1966.
10. Кендалл М., Стьюарт А. Статистические выводы и связи. М., 1973.
11. Кокс Д., Хинкли Д. Теоретическая статистика. М., 1978.
12. Кокс Д., Хинкли Д. Задачи по теоретической статистике с решениями. М., 1981.
13. Крамер Г. Математические методы статистики. М., 1975.
14. Леман Э. Проверка статистических гипотез. М., 1964.
15. Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов. М., 1974.
16. Рао С. Р. Линейные статистические методы и их применения. М., 1968.
17. Севастьянов Б. А., Чистяков В. П., Zubkov A. M. Сборник задач по теории вероятностей. М., 1980.
18. Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений. М., 1980.
19. Уилкс С. Математическая статистика. М., 1967.
20. Феллер В. Введение в теорию вероятностей и ее приложения. М., 1967, т. 1.
21. Чибисов Л. М. Некоторые критерии типа χ^2 для непрерывных распределений. — Теория вероятностей и ее применения, т. XVI, 1971, вып. 1.
22. Чистяков В. П. Курс теории вероятностей. М., 1982.
23. Шеффе Г. Дисперсионный анализ. М., 1980.
24. Ширяев А. Н. Статистический последовательный анализ. М., 1976.
25. Billingsley P. Statistical Inference for Markov Processes. Univ. of Chicago, 1961.
26. Davidson R., Levr W. The limiting distribution of the likelihood ratio statistic under a class of local alternatives. Sankhya, ser. A, v. 32, p. 2 (1970), 209 — 224.
27. Wald A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. Amer. Math. Soc., v. 54. (1943), 426 — 482.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Аддитивные факторы 206
Альтернатива (альтернативная гипотеза) 108, 138
— односторонняя (право- и левосторонняя) 154

Байесовский подход 223
Бернуллиевская модель 11
—, оценивание параметра 43—44, 92
—, — минимальное 226—227
—, — параметрических функций 60
—, проверка простых гипотез 147—148
—, — сложных односторонних гипотез 158

—, решающее правило 225
—, функция информации 46

Вариационный ряд 13
—, его члены (крайние и средние) 25
Выборка 7
—, ее вклад 45
Выборочные блоки 127
Выборочный момент 18
— пространство 8
— распределение 19
— среднее 18, 21
— теория 13
— характеристика 18, 20

Гамма-модель 10
 —, свойство воспроизводимости 28
 —, метод моментов 78
 —, проверка гипотез 156—157
 —, функция информации 47
 —, оценка параметра 47
 Гауссовская последовательность 220
 Генеральная совокупность 8
 Гипотеза линейная 173
 — независимости 103, 130
 — о виде распределения 103, 107
 — однородности 103, 123, 203
 — о параллельности линий регрессии 201—203
 — основная (нулевая) 103
 — параметрическая 104
 — простая 104
 — сложная 104
 — случайности 104, 133
 Гистограмма 16

Двойная классификация 205
 Дисперсионный анализ 205
 Дисперсия асимптотическая 75
 — выборочная 18, 22
 — остаточная 180
 Доверительные границы 82
 Доверительный интервал 81
 — асимптотический 82
 — односторонний 82
 — центральный 90
 Доверительная область 94
 — максимального правдоподобия 176

Инверсия 134
 Интеграл Стильеса 9

Квантиль выборочная 23
 Корреляционное отношение 212
 — частное 216
 Корреляция частная 216
 Коэффициент корреляции множественный 215
 — асимметричный 21
 — эксцесса 21
 Критерий Бхаттачария оптимальности оценки 50
 — Вальда (последовательный критерий отношения вероятностей) 149
 — Вилкоксона 128
 — знаков 117
 — квантилей 117
 — Кендалла 133
 — локальный наиболее мощный 161
 — независимости хи-квадрат 130—131
 — Неймана—Пирсона 142—145
 — несмещенный 107, 141
 — однородности Смирнова 124
 — хи-квадрат 124—127
 — отношения правдоподобия 167
 — пустых блоков 127
 — пустых ящиков 120—122
 — равномерно наиболее мощный 141
 — Рао—Крамера оптимальности оценки 48
 — ранговый 128
 — рандомизированный 139
 — серий 128
 — симметрический 122
 — согласия 104
 — Колмогорова 107—108
 — хи-квадрат Пирсона 109—111
 — состоятельный 113
 — Спирмена 132
 — статистический 103, 139
 — факторизации 55
 Критическая область 105, 139

Матрица идемпотентная 30
 — информационная 52
 — плана 180
 — ее след 30
 — ее спектральное представление 29

Медиана выборочная 24
 Мера хи-квадрат Пирсона 79
 Метод группировки наблюдений 80
 — максимального правдоподобия 65
 — моментов 77
 — наименьших квадратов 181
 — накопления 69
 — отношения правдоподобия 166
 Минимальный подход 223
 Минимальная ошибка предсказания 212
 Многочлен интерполяционный 188
 Многочлены ортогональные Чебышева 189

Множество параметрическое 9
 Модель абсолютно-непрерывная 9
 — биномиальная 10
 —, функция информации 47
 —, оценка параметра 49
 — геометрическая 11
 — дискретная 9
 — Коши 10, 57
 —, метод накопления 70
 —, проверка гипотезы 162
 —, функция информации 47
 — отрицательная биномиальная 10, 61
 —, оценивание параметра 40, 49, 63
 —, функция информации 47
 — параметрическая 9
 — регулярная 46
 — с монотонным отношением правдоподобия 155
 — статистическая 6, 8
 — экспоненциальная 48, 155

Неравенство Бхаттачария 50

— Пенсена 121
 — Рао—Крамера 48, 51
 Нормальная модель 10
 —, гипотеза о равенстве двух средних 166
 —, доверительный интервал для среднего в $\mathcal{N}(\theta, \sigma^2)$ 84, 163
 —, — — в $\mathcal{N}(\theta_1, \theta_2)$ 87
 —, — — для дисперсии в $\mathcal{N}(\mu, \sigma^2)$ 65, 164
 —, — — в $\mathcal{N}(\theta_1, \theta_2)$ 87
 —, — — для разности двух средних 87, 166
 —, — — для отношения двух дисперсий 88
 —, доверительная область для (θ_1, θ_2) 94—96
 —, достаточные статистики 56
 —, задача классификации 231—235
 —, информационная матрица 53
 —, критерий отношения правдоподобия для среднего в $\mathcal{N}(\theta_1, \theta_2)$ 167
 —, — — для дисперсии в $\mathcal{N}(\theta_1, \theta_2)$ 175—176
 —, локальный наиболее мощный критерий в $\mathcal{N}(\theta, \sigma^2)$ 162
 —, односторонние альтернативы в $\mathcal{N}(\mu, \sigma^2)$ 155
 —, оптимальное прогнозирование 213
 —, о. м. п. 66—67
 —, оценивание θ в $\mathcal{N}(\theta, \sigma^2)$ 49, 77
 —, — θ^1 в $\mathcal{N}(\theta, \sigma^2)$ 51
 —, — θ^2 в $\mathcal{N}(\mu, \theta^1)$ 49
 —, — θ_1 в $\mathcal{N}(\theta_1, \theta_2)$ 52
 —, — θ_2 в $\mathcal{N}(\theta_1, \theta_2)$ 40, 53—54
 —, — $\Phi\left(\frac{x_0 - \theta_1}{\theta_2}\right)$ 68
 —, — σ^2 и ρ для $\mathcal{N}\left(0, \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}\right)$ 68—69

—, последовательный критерий 152—153
 —, проверка гипотезы в $\mathcal{N}(\theta, \sigma^2)$ 145—146
 —, сверхэффективная оценка для θ в $\mathcal{N}(\theta, 1)$ 76—77
 —, сложные двусторонние альтернативы в $\mathcal{N}(\theta, \sigma^2)$ и $\mathcal{N}(\mu, \theta^2)$ 158, 160, 162, 164

—, — — гипотезы для θ_1 и θ_2 в $\mathcal{H}(\theta_1, \theta_2)$

165

—, — односторонние альтернативы 154

—, функции информации 47

—, эффективные оценки 49

Обратный биномиальный выбор 156

Остаточный вектор 184

Оценивание интервальное 81

— точечное 38

Оценка 38

— асимптотически несмещенная 71

— эффективная 75

—, ее асимптотическая дисперсия 75

—, ее асимптотическая эффективность 75

—, ее смещение 30

—, ее среднеквадратическая ошибка 39

—, максимального правдоподобия 65

—, свойство инвариантности 68

—, состоятельность, асимптотическая нормальность и асимптотическая эффективность 73—75

—, мультиномиальная 81

—, наименьших квадратов 181

—, обобщенная 185

—, несмещенная 39

—, оптимальная (несмещенная оценка с минимальной дисперсией) 38, 42

—, по методу минимума хи-квадрат 79—80

—, сверхэффективная 76

—, состоятельная 72

—, эффективная 48, 52

—, ядерная 17

Ошибки 1-го и 2-го рода 139

Плотность спектральная 217

Подмножество цилиндрическое 178

Полигон частот 17

Порядковые статистики 13

Последовательный анализ 148

Предиктор оптимальный 211

— эмпирический 217

Предсказывающие переменные 211

Принцип максимума апостериорной вероятности 224

Пуассоновская модель 10, 61

—, доверительное оценивание параметра 93

—, критерий согласия хи-квадрат 116

—, оценивание параметрических функций 39, 62

—, о. м. п. параметра 65, 76

—, функции информации 47

—, эффективная оценка параметра 49

Равномерная модель 10

—, доверительный интервал для параметра 88

—, достаточные статистики 56—57

—, оценивание параметра 63

—, о. м. п. параметра 67—68, 76

Ранги наблюдений 128

Распределение альтернативное 108

— апостериорное 224

— априорное 224

—, — наименее благоприятное 225

— бета 227

— биномиальное 10

— Вейбула 97

— гамма 10

— гипотетическое 107

— Дирихле 236

— Коши 10

— Лапласа 99

— логарифмическое 61

— логистическое 96

— логнормальное 99

— нормальное 10

— отрицательное биномиальное 10

— Парето 236

— полиномиальное (мультиномиальное) 38,

81

— пуассоновское 10

— равномерное 10

— Снедекора (F -распределение) 33

— Стьюдента (t -распределение) 31

— типа степенного ряда 61, 70, 93

— усеченное 61

— хи-квадрат 27, 28

— экспоненциальное (показательное) 28

Расстояние Махаланобиса 233

Регрессия 180

—, ее коэффициенты 180

— нормальная 192

—, —, общая линейная гипотеза 199—200

—, —, совместные доверительные интервалы 197

—, параболическая 180, 188

—, простая 187

—, —, доверительное оценивание параметров 195

—, —, проверка гипотез 200

Решающее правило (функция, процедура)

222

— допустимое 223

Решение байесовское 224

— минимаксное 225

Риск байесовский 224

— максимальный 225

Серия 128

Статистика 38

— достаточная 54

— минимальная 55

— полная 58

— Кендалла 133

— критерия 105, 141

— отношения правдоподобия 142

— симметрическая 119

— Спирмена 132

— центральная 83

Статистический аналог 15

Стационарный случайный процесс 217

Сходимость по вероятности 10

— по распределению (слабая) 10

Таблица дисперсионного анализа 210

— сопряженности двух признаков 130

Теорема Байеса 224

— Глиенко 15

— Колмогорова 15

— Неймана—Пирсона 142

— Рао—Блекузлла—Колмогорова 58

— Смирнова 16

— Фишера 31

Тождество Вальда 151

Точки сверхэффективности 77

Уравнение несмещенности 59

— нормальное метода наименьших квадратов 181

— правдоподобия 65

Уровень значимости 105, 140

Функция вклада 45

— дискриминантная 232

— информации 46

— классификационная 234

— ковариационная стационарного процесса 217

— критическая 139

— мощности 108, 139

— правдоподобия 45

— частотная 81

— параметрическая 39

— потеря 223

— распределения теоретическая 14

—, —, доверительная зона для нее 16

— эмпирическая 14, 44

— регрессии 211

— риска 223

— Хевисайда 14

ОГЛАВЛЕНИЕ

Предисловие 3
Введение 5

Глава 1

Основные понятия
и элементы выборочной
теории

§ 1.1. Вариационный ряд выборки и эмпирическая функция распределения 13. § 1.2. Выборочные характеристики 18. § 1.3. Асимптотическое поведение выборочных моментов 20. § 1.4. Порядковые статистики 23. § 1.5. Распределения некоторых функций от нормальных случайных величин 27. Задачи 34.

Глава 2

Оценивание неизвестных
параметров распределений

§ 2.1. Статистические оценки и общие требования к ним. Несмещенные оценки с минимальной дисперсией 37. § 2.2. Критерии оптимальности оценок, основанные на неравенстве Рао — Крамера и его обобщениях 45. § 2.3. Принцип достаточности и оптимальные оценки 54. § 2.4. Оценки максимального правдоподобия 65. § 2.5. Метод моментов и другие методы, основанные на группированных данных 77. § 2.6. Интервальное оценивание 81. Задачи 95.

Глава 3

Проверка
статистических
гипотез

§ 3.1. Понятие статистической гипотезы и статистического критерия. Критерии согласия 102. § 3.2. Проверка гипотезы о виде распределения 107. § 3.3. Симметрические критерии в схеме группировки с растущим числом интервалов. Критерий пустых ящиков 118. § 3.4. Гипотеза однородности 123. § 3.5. Гипотеза независимости 130. § 3.6. Гипотеза случайности 133. Задачи 136.

Глава 4

Параметрические
гипотезы

§ 4.1. Общие положения 138. § 4.2. Выбор из двух простых гипотез. Критерий Неймана — Пирсона 141. § 4.3. Выбор из двух простых гипотез. Понятие о последовательном анализе 148. § 4.4. Сложные гипотезы 153. § 4.5. Критерий отношения правдоподобия 166. Задачи 177.

Глава 5

Линейная регрессия и метод наименьших квадратов

§ 5.1. Модель линейной регрессии 179. § 5.2. Оценивание неизвестных параметров модели 181. § 5.3. Нормальная регрессия. Интервальное оценивание 192. § 5.4. Общая линейная гипотеза нормальной регрессии 198. § 5.5. Применение теории линейной регрессии 201. § 5.6. Элементы теории статистической регрессии и корреляции 210. Задачи 220.

Глава 6

Элементы теории решений. Дискриминантный анализ

§ 6.1. Статистические решающие функции. Байесовское и минимаксное решения 222. § 6.2. Задача классификации наблюдений 228. § 6.3. Классификация наблюдений в случае двух нормальных классов 231. § 6.4. Классификация нормальных наблюдений. Общий случай 233. Задачи 235.

Приложения 237.

Литература 244.

Предметный указатель 244.