А. А. БОРОВКОВ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ

А. А. БОРОВКОВ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ

*Допущено Министерством высшего
и среднего специального образования СССР
в качестве учебного пособия для студентов
математических и физических специальностей вузов*

МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1984

22.172
Б 83
УДК 519.2

Боровков А. А. Математическая статистика. Дополнительные главы: Учебное пособие для вузов.— М.: Наука. Главная редакция физико-математической литературы, 1984, 144 с.

Книга является продолжением учебного пособия «Математическая статистика» того же автора и содержит два важных раздела современной математической статистики, не вошедшие в названную книгу: 1) задачи с двумя и более выборками (в первой книге рассматривались лишь задачи с одной выборкой), 2) общий теоретико-игровой подход к задачам математической статистики.

В первом разделе излагаются задачи об однородности, задачи регрессии, задачи дисперсионного анализа и распознавания образов. Во втором — изучаются статистические игры и основные принципы отыскания оптимальных и асимптотически оптимальных решающих правил. Многие результаты названной выше основной книги обобщаются на случай произвольной функции потерь.

Илл. 4. Библ. 40 назв.

Рецензенты:

кафедра теории вероятностей и математической статистики Ленинградского государственного университета им. А. А. Жданова (заведующий кафедрой доктор физико-математических наук профессор *В. В. Петров*);

доктор физико-математических наук *Д. М. Чибисов*

$Б\frac{1702060000-127}{053(02)-84}50\text{-}84$

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Настоящая книга является продолжением книги автора «Математическая статистика» (М.: Наука, 1984). Она содержит изложение статистических задач, связанных с несколькими выборками (гл. 1), и изложение общего теоретико-игрового подхода к задачам математической статистики (гл. 2). Чтение книги предполагает знакомство с названной первой книгой [6] (в которой рассматривались лишь статистические задачи, связанные с одной выборкой), а также с университетским курсом теории вероятностей, например, в объеме [5].

Глава 1 книги, как уже отмечалось, посвящена задачам с двумя и более выборками. К ним относятся прежде всего задачи об однородности (полной или частичной, §§ 1, 2), задачи регрессии (§ 3) и дисперсионного анализа (§ 4). На основании результатов гл. 3 из [6] для задач однородности в параметрическом случае построены асимптотически оптимальные критерии в предположении, что альтернативные гипотезы являются близкими к основной гипотезе об однородности. Для задач регрессии (как для линейной регрессии, так и для регрессии по произвольным функциям) с помощью результатов глав 2, 3 из [6] найдены эффективные оценки неизвестных параметров и построены критерии для проверки основных гипотез. Рассмотрены также так называемые задачи распознавания образов (§ 5), которые появляются в учебной литературе, по-видимому, впервые.

Глава 2 посвящена общему теоретико-игровому подходу к задачам статистики. Он способствует выработке общего взгляда на предмет математической статистики и позволяет обобщить многие результаты глав 2, 3 из [6]. В § 2 излагаются основные понятия и результаты «обычной» теории игр (рассматриваются лишь игры двух лиц). В частности, устанавливаются связи между основными типами оптимальных стратегий — байесовскими, минимаксными, равномерно наилучшими в подклассах. В § 3 изучаются статистические игры. В § 4 формулируется и доказывается так называемый байесовский принцип, который позволяет свести задачу отыскания байесовского статистического решения к значительно более простой задаче построения байесовской стратегии для обычной игры двух лиц. В § 5 обсуждаются принципы достаточности, несмещенности и инвариантности для построения решений, равномерно наилучших в соответствующих подклассах. Параграфы 6—8 посвящены отысканию асимптотически оптимальных решающих правил. В § 6 изучаются асимптотически оптимальные оценки параметров при произвольной (не только квадратической) функции потерь. В этом случае удается установить результаты, близкие к результатам гл. 2 из [6] об асимптотической оптимальности оценок максимального правдоподобия. В §§ 7, 8 изучаются асимптотически оптимальные критерии при произвольной функции потерь. В § 7 доказана асимптотическая байесовость критерия отношения правдоподобия; в § 8 установлен предельный признак оптимальности критериев для проверки близких гипотез (обобщение результатов §§ 14, 15 гл. 3 из [6] на случай произвольной функции потерь).

Книга содержит также Приложение, в котором доказаны две фундаментальные теоремы теории статистических игр. Чтение этого раздела требует более высокой математической подготовки.

К предлагаемой книге могут быть почти целиком отнесены предисловие и введение из [6]. Здесь действует та же система обозначений, что и в [6], и те же правила нумерации параграфов, теорем, лемм и т. д. Нумерация формул, теорем, лемм, примеров в каждом параграфе самостоятельная. Если мы ссылаемся на формулы или теоремы другого параграфа или главы, то перед соответствующим номером появляются дополнительные цифры. Например, формула (1.3.12) означает формулу (12) § 3 гл. 1, а формула (3.12) означает формулу (12) § 3 читаемой главы. При ссылках на результаты и формулы из [6] будем добавлять соответствующий индекс, например, теорема 2.12.3[6] означает теорему 3 из § 12 гл. 2 из [6]. Значок $\lhd$ означает, как и в [6], окончание доказательства. Основные обозначения введены в [6] и часто используются без дополнительных пояснений.

Много ценных замечаний, направленных на улучшение книги, было сделано Д. М. Чибисовым и А. И. Саханенко. Им и всем тем, кто так или иначе помогал мне в работе над книгой, я хотел бы выразить свою искреннюю благодарность.

А. А. Боровков

ГЛАВА 1

СТАТИСТИЧЕСКИЕ ЗАДАЧИ С ДВУМЯ И БОЛЕЕ ВЫБОРКАМИ

В §§ 1, 2 рассмотрены задачи об однородности двух выборок.
В § 3 рассмотрены задачи регрессии.
В § 4 излагаются основные результаты дисперсионного анализа.
В § 5 рассмотрены задачи распознавания образов.

§ 1. Проверка гипотез об однородности (полной или частичной) в параметрическом случае

1. Рассматриваемый класс задач. В [6] основным объектом всех рассмотрений была выборка X объема n из полностью или частично неизвестного распределения $\mathbf{P}$. Мы переходим теперь к статистическим задачам, где фигурируют не одна, а *две или более* выборок.

Один из основных классов задач, которые при этом рассматриваются, составляют задачи о проверке *однородности* (полной или частичной) двух выборок.

Сюда относятся следующие основные три типа задач:

I. *Проверка «обычной» однородности.* Задача здесь состоит в проверке гипотезы о том, что две выборки X и Y извлечены из одного и того же неизвестного распределения. Такие задачи возникают, например, при сравнении двух способов обработки в каком-нибудь технологическом процессе или в сельском хозяйстве. Основой для сравнения обычно служат числовые характеристики конечного продукта (выборки), которые носят случайный характер. Такого же рода задачи возникнут, если мы по состоянию больных будем проверять воздействие нового лекарства, сравнивая подопытную группу пациентов с контрольной.

К задачам об однородности относится пример 3 из введения в [6].

В этом параграфе мы будем рассматривать *параметрический случай.* Пусть дано параметрическое семейство распределений $\{P_\theta\}_{\theta\in\Theta}$ и даны две независимые выборки $X=(x_1, \ldots, x_{n_1})$ и $Y=(y_1, \ldots, y_{n_2})$ объемов n_1 и n_2 соответственно, относительно которых заранее известно, что они относятся к семейству $\{P_\theta\}$:

$$X \in P_{\theta_1}, \quad Y \in P_{\theta_2} \tag{1}$$

при некоторых θ_1, θ_2. Обычная задача об однородности состоит здесь в проверке гипотезы $H_1=\{\theta_1=\theta_2\}$ против дополнительной альтернативы $H_2=\{\theta_1\neq\theta_2\}$. Очевидно, что здесь обе гипотезы H_1 и H_2 являются сложными.

II. *Проверка однородности при наличии мешающего параметра.* Здесь предполагается, что размерность k параметра θ больше 1. Запишем вектор θ в виде набора $\theta=(u, v)$ двух «подвекторов» u и v и обозначим u_j, v_j компоненты векторов θ_j в (1), $j=1, 2$.

Пусть нам заранее известно, что «подпараметр» v у обеих выборок хотя и неизвестен, но является общим: $v_1=v_2=v$. Проверяется гипотеза $H_1=\{u_1=u_2\}$ против $H_2=\{u_1\neq u_2\}$.

Это и есть задача об однородности при наличии мешающего параметра v. Она отличается от обычной задачи об однородности тем, что альтернатива для гипотезы $H_1=\{\theta_1=\theta_2\}$ имеет вид $H_2=\{u_1\neq u_2,\ v_1=v_2\}$.

Можно привести следующий пример возникновения такого рода задач. Допустим, что нас интересует состояние некоторого объекта, которое характеризуется вектором a. Но измерить a непосредственно нельзя. Нам доступны лишь измерения, при которых на a накладывается случайный шум. Природа шума при разных наблюдениях остается неизменной. Надо проверить гипотезу о неизменности a в двух сериях наблюдений X и Y.

Если, скажем, измерения имеют вид $x_i=a_1+\xi_i$, где $\xi_i\in\Phi_{\lambda,\sigma^2}$ определяют роль шума, и такой же характер имеют наблюдения y_i с заменой a_1 на a_2, то мы можем записать $X\in\Phi_{a_1+\lambda,\sigma^2}$, $Y\in\Phi_{a_2+\lambda,\sigma^2}$. Мы пришли к задаче о проверке равенства средних $\{\alpha_1=\alpha_2\}$ двух нормальных совокупностей Φ_{α_1,σ^2} и Φ_{α_2,σ^2} при общем неизвестном значении σ^2.

III. *Проверка частичной однородности.* Здесь проверяется гипотеза H_1 лишь о «частичном» совпадении θ_1 и θ_2. Именно, проверяется гипотеза $H_1=\{u_1=u_2\}$ (в обозначениях предыдущего раздела) против $H_2 = \{u_1 \neq u_2\}$. Значения v_1, v_2 для каждой из выборок X, Y могут быть свои.

Допустим, например, что в лаборатории оценивается результат воздействия нового способа обработки на урожайность какого-нибудь злака. Наблюдения представляют из себя суммарный вес зерен в отдельных колосках. Предположим, что $x_i \in \Phi_{\alpha_1, \sigma_1^2}$, $i = 1, \ldots, n_1$, для опытной партии колосков и $y_i \in \Phi_{\alpha_2, \sigma_2^2}$ для контрольной партии. Естественно допустить, что «разброс» σ^2 в результате изменения обработки мог измениться. Для нас же существенно, изменился ли главный показатель α, определяющий урожайность культуры. Мы приходим к задаче о проверке гипотезы $H_1 = \{\alpha_1 = \alpha_2\}$ против $H_2 = \{\alpha_1 \neq \alpha_2\}$ для нормальных совокупностей, дисперсии которых могут быть разными. Это задача, хорошо известная в литературе под названием *проблемы Беренса—Фишера* *).

В этом параграфе мы произведем редукцию всех трех типов задач для произвольных параметрических семейств к задаче, рассмотренной в § 3.15[6], о принадлежности одной выборки параметрическому подсемейству и найдем вид асимптотически минимаксных критериев в предположении близости проверяемых гипотез. Это будут критерии отношения правдоподобия, которые для нормальных совокупностей будут совпадать с критериями, построенными в поисках той или иной *точной* оптимальности (если таковые имеются; ср. с [18]).

Статистический критерий π для проверки H_1 против H_2 в нашем случае будет функцией $\pi = \pi(X, Y)$ двух выборок X и Y и будет, как и в [6], означать вероятность приема H_2 при данной объединенной выборке (X, Y) (см.

*) Поиску ее оптимальных решений посвящена довольно обширная литература. Значительный вклад в исследование проблемы Беренса — Фишера, оказавшейся весьма трудной, принадлежит Ю. В. Линнику и его ученикам. Эти исследования требуют введения новых понятий и использования весьма сложного математического аппарата. Это делает невозможным цитирование и доказательство (в рамках настоящей книги) полученных результатов. Положение в задачах об обычной однородности и об однородности при наличии мешающего параметра для нормальных совокупностей несколько лучше (в ряде задач удается найти равномерно наиболее мощные (р. н. м.) инвариантные несмещенные критерии, однако необходимые для этого построения также оказываются довольно сложными; подробнее об этом см. [18]).

гл. 3 [6]). Определения асимптотического уровня и асимптотической оптимальности критерия π здесь остаются теми же, что и в § 3.14[6].

Определение 1. Мы будем говорить, что критерий π *имеет асимптотический уровень* $1-\varepsilon$ (принадлежит классу $\mathcal{K}_\varepsilon$), если

$$\limsup_{n\to\infty} \sup_{(\theta_1,\theta_2)\in\overline{\Theta}_1} \mathsf{M}_{\theta_1\theta_2}\pi(X,\ Y)\leqslant \varepsilon,$$

где $\mathsf{M}_{\theta_1\theta_2}$ означает математическое ожидание по распределению $\mathbf{P}_{\theta_1}\times\mathbf{P}_{\theta_2}$, $\overline{\Theta}_1$ есть множество значений $(\theta_1,\ \theta_2)$, при которых выполнена гипотеза H_1 (например, множество всех точек $(\theta_1,\ \theta_2)$, лежащих на «биссектрисе» $\theta_1=\theta_2$ в задаче об обычной однородности).

Определение 2. Критерий $\pi_1\in\mathcal{K}_\varepsilon$ называется *асимптотически минимаксным в* $\mathcal{K}_\varepsilon$ для проверки H_1 против H_2, если для любого критерия $\pi\in\mathcal{K}_\varepsilon$ выполняется

$$\liminf_{n\to\infty}\left(\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_2} \mathsf{M}_{\theta_1\theta_2}\pi_1(X,Y)-\inf_{(\theta_1,\theta_2)\in\overline{\Theta}_2} \mathsf{M}_{\theta_1\theta_2}\pi(X,Y)\right)\geqslant 0,$$

где $\overline{\Theta}_2$ есть множество значений $(\theta_1,\ \theta_2)$, соответствующих альтернативам из H_2.

2. Асимптотически минимаксный критерий для проверки близких гипотез об обычной однородности. Введем новый параметр $\overline{\theta}=(\theta_1,\ \theta_2)$, характеризующий «объединенную» выборку $(X,\ Y)$. Функция правдоподобия этой выборки равна $f_{\overline{\theta}}(X,\ Y)=f_{\theta_1}(X)\,f_{\theta_2}(Y)$.

Допустим сначала для простоты, что объемы выборок совпадают: $n_1=n_2=n$. Тогда выборку $(X,\ Y)$ можно представлять себе как выборку объема n, образованную наблюдениями $(x_1,\ y_1),\ \ldots,\ (x_n,\ y_n)$ из распределения $\mathbf{P}_{\overline{\theta}}=\mathbf{P}_{\theta_1}\times\mathbf{P}_{\theta_2}$, имеющего плотность $f_{\theta_1}(x)\,f_{\theta_2}(y)$. Мы приходим к задаче, рассмотренной в § 3.15 [6], о проверке по выборке $(X,\ Y)$ гипотезы H_1 о том, что параметр $\overline{\theta}$ располагается на «кривой» $\theta_1=\theta_2$. В обозначениях § 3.15 [6] гипотеза H_1 в нашем случае имеет вид $\overline{\theta}=g(\alpha)$, где $\alpha=\theta_1$, $g(\alpha)=(\alpha,\ \alpha)$. Очевидно, что матрица $G=\left\|\frac{\partial g_i}{\partial\alpha_j}\right\|$, $i=1,\ldots,2k$, $j=1,\ldots,k$, имеет вид $\begin{pmatrix}E\\E\end{pmatrix}$, где E — единичная матрица k-го порядка, так что ранг G равен k.

Будем считать параметр $\bar{\theta}$ локализованным, т. е. будем считать, что значения θ_1 и θ_2 близки и, стало быть, возможные значения $\bar{\theta}$ располагаются в окрестности точки $\bar{\theta}_0 = (\theta_0, \theta_0)$ при некотором фиксированном θ_0. Если следовать § 3.15 [6], то нам будет удобнее ввести новый параметр $\tau = (\tau', \tau'') = (\gamma'/\sqrt{n}, \gamma''/\sqrt{n}) = \gamma/\sqrt{n}$, где $\tau' = \theta_1 - \theta_0$, $\tau'' = \theta_2 - \theta_1$, так что отображение $\bar{\theta} = \bar{\theta}(\tau)$ взаимно однозначно: $\theta_1 = \tau' + \theta_0$, $\theta_2 = \tau'' + \tau' + \theta_0$. В терминах параметров τ и γ гипотеза H_1 об однородности примет вид $H_1 = \{\tau'' = 0\} = \{\gamma'' = 0\}$. В качестве альтернативы мы рассмотрим «отделенную» гипотезу

$$H_2^b = \{\gamma'' I \gamma''^T \geqslant b^2\}, \qquad b > 0, \tag{2}$$

где $I = I(\theta_0)$ есть матрица Фишера для семейства $\{\mathbf{P}_\theta\}$ в точке θ_0.

Теорема 1. *Пусть в окрестности точки θ_0 семейство $\{\mathbf{P}_\theta\}$ удовлетворяет условиям (RR)* (см. § 2.28[6]). *Тогда критерий отношения правдоподобия*

$$R_1(X, Y) = \frac{\sup\limits_{\theta_1, \theta_2} f_{\theta_1}(X) f_{\theta_2}(Y)}{\sup\limits_{\theta} f_\theta(X) f_\theta(Y)} > e^{h_\varepsilon/2} \tag{3}$$

является асимптотически минимаксным критерием асимптотического уровня $1 - \varepsilon$ для проверки $H_1 = \{\theta_1 = \theta_2\}$ против $H_2^b = \{(\theta_1 - \theta_2) I (\theta_1 - \theta_2)^T \geqslant b^2/n\}$ при любом $b > 0$, где h_ε есть квантиль порядка $1 - \varepsilon$ распределения χ^2 с k степенями свободы (такое предельное распределение при гипотезе H_1 имеет статистика $2 \ln R_1(X, Y)$).

Пусть $\hat{\theta}_X^*, \hat{\theta}_Y^*, \hat{\theta}^*$ есть о. м. п. параметра $\theta = \theta_1 = \theta_2$ соответственно по выборкам X, Y, (X, Y). Тогда критерий

$$(\hat{\theta}_X^* - \hat{\theta}^*) I(\hat{\theta}^*) (\hat{\theta}_X^* - \hat{\theta}^*)^T + (\hat{\theta}_Y^* - \hat{\theta}^*) I(\hat{\theta}^*) (\hat{\theta}_Y^* - \hat{\theta}^*)^T > h_\varepsilon/n \tag{4}$$

будет асимптотически эквивалентным критерию (3).

Доказательство. Приведенное утверждение является прямым следствием теоремы 3.15.4[6]. Нам надо лишь выяснить, что из себя представляют матрица Фишера $\bar{I}(\bar{\theta}_0) = \bar{I}(\theta_0, \theta_0)$ для «объединенного» параметра $\bar{\theta} = (\theta_1, \theta_2)$ и матрица M_2 для параметрического семейства $\{\mathbf{P}_{(\theta_0, \theta_0 + \beta)}\}$ в точке $\beta = 0$. Имеем

$$\ln f_{\theta_1}(x) f_{\theta_2}(y) = l(x, \theta_1) + l(y, \theta_2).$$

Обозначим через t_i, $i = 1, \ldots, 2k$, координаты вектора $\bar{\theta}$. Тогда, если через $\mathsf{M}_{\bar{\theta}}$ обозначить математическое ожидание по распределению $\mathbf{P}_{\bar{\theta}}$, то элементы $\bar{I}_{ij}(\bar{\theta})$ матрицы $\bar{I}(\bar{\theta})$ будут равны

$$\bar{I}_{ij}(\bar{\theta}) = \mathsf{M}_{\bar{\theta}}\left(\frac{\partial l(x_1, \theta_1)}{\partial t_i} + \frac{\partial l(y_1, \theta_2)}{\partial t_i}\right)\left(\frac{\partial l(x_1, \theta_1)}{\partial t_j} + \frac{\partial l(y_1, \theta_2)}{\partial t_j}\right).$$

Отсюда в силу независимости x_1 и y_1 мы получаем

$$\bar{I}(\bar{\theta}) = \begin{pmatrix} I(\theta_1) & 0 \\ 0 & I(\theta_2) \end{pmatrix}.$$

Поэтому критерий (4) есть не что иное, как критерий (3.15.12)[6] в теореме 3.15.4[6].

Аналогичные вычисления показывают, что $M_2 = I(\theta_0)$, так как при $\beta = (\beta_1, \ldots, \beta_k) = 0$

$$\frac{\partial l(x_1, \theta_0)}{\partial \beta_i} + \frac{\partial l(y_1, \theta_0 + \beta)}{\partial \beta_i} = \frac{\partial l(y_1, \theta_0)}{\partial t_i}, \qquad i = 1, \ldots, k. \ \triangleleft$$

Замечание 1. Утверждение теоремы 1 мы получили в предположении, что $n_1 = n_2$. Однако это ограничение совершенно несущественно. Рассмотрим, например, случай, когда $n_1 \to \infty$, $n_2 \to \infty$ так, что отношение n_1/n_2 равно рациональному числу r_1/r_2 (r_1, r_2 — произвольные целые фиксированные числа, $n_i = nr_i$, $n \to \infty$). Мы введем опять новый параметр $\bar{\theta} = (\theta_1, \theta_2)$ и будем рассматривать объединенную выборку (X, Y) как выборку объема n с наблюдениями $(x_1, \ldots, x_{r_1}; y_1, \ldots, y_{r_2})$, $(x_{r_1+1}, \ldots, x_{2r_1}; y_{r_2+1}, \ldots, y_{2r_2})$, ... из распределения

$$\mathbf{P}_{\bar{\theta}} = \underbrace{\mathbf{P}_{\theta_1} \times \ldots \times \mathbf{P}_{\theta_1}}_{r_1 \text{ раз}} \times \underbrace{\mathbf{P}_{\theta_2} \times \ldots \times \mathbf{P}_{\theta_2}}_{r_2 \text{ раз}},$$

зависящего от параметра $\bar{\theta}$. Функция правдоподобия здесь вновь будет иметь вид

$$f_{\bar{\theta}}(X, Y) = f_{\theta_1}(X) f_{\theta_2}(Y).$$

Если ввести, как и прежде, новый параметр $\tau = (\tau', \tau'') = (\theta_1 - \theta_0, \theta_2 - \theta_1)$ и положить $\tau = \gamma/\sqrt{n} = (\gamma'/\sqrt{n}, \gamma''/\sqrt{n})$, то рассматриваемая задача состоит в проверке $H_1 = \{\gamma'' = 0\}$ против $H_2^b = \{\gamma'' M_2 \gamma''^T \geqslant b^2\}$, где M_2 — матрица Фишера для $\mathbf{P}_{(\theta_0, \theta_0+\beta)}$ в точке $\beta = 0$. Легко видеть, что в нашем случае $M_2 = r_2 I(\theta_0)$, так что множество альтернатив сохраняет свою форму (2):

$$H_2^b = \{\gamma'' I \gamma''^T > b^2/r^2\}.$$

Матрица Фишера $\bar{I}(\bar{\theta})$ будет иметь вид

$$\begin{pmatrix} r_1 I(\theta_1) & 0 \\ 0 & r_2 I(\theta_2) \end{pmatrix}.$$

Остается воспользоваться теоремой 3.15.4[6]. Мы получим тогда утверждение теоремы 1, в которой критерий (4) следует заменить на

$$n_1 (\hat{\theta}_X^* - \theta^*) I(\theta^*) (\hat{\theta}_X^* - \theta^*)^T + \\ + n_2 (\hat{\theta}_Y^* - \theta^*) I(\theta^*) (\hat{\theta}_Y^* - \theta^*) > h_\varepsilon. \quad (5)$$

С помощью теоремы 3.15.4[6] можно указать также гарантированную асимптотическую мощность критериев (3)—(5).

Утверждение теоремы остается справедливым и в общем случае, когда $n_1 \to \infty$, $n_2 \to \infty$, $n_1/n_2 \to c$, где c — произвольное число из $(0, 1)$. Однако доказательство этого факта требует дополнительных рассмотрений.

Замечание 2. Утверждение теоремы 1 сохранится, если гипотезу $H_1 = \{\theta_1 = \theta_2\}$ заменить гипотезой (см. [6])

$$H_1^a = \{(\theta_1 - \theta_2) I (\theta_1 - \theta_2)^T \leqslant a^2/n\}, \qquad 0 < a < b.$$

Замечание 3. Форма асимптотически минимаксных критериев в теореме 1 от θ_0 не зависит. Значение θ_0 входит только в определение гипотезы H_2^b через $I = I(\theta_0)$ (см. (2)), хотя и этого появления θ_0 можно было бы избежать, заменив I в (2) на $I((\theta_1 + \theta_2)/2)$. Это дало бы нам «асимптотически эквивалентную» H_2^b гипотезу $\widetilde{H}_2^b$, для которой утверждение теоремы 3 полностью сохраняется. Появление в (2) значения θ_0 есть следствие использования наиболее простого пути редукции рассматриваемой задачи к результатам § 3.15[6].

Пример 1. Пусть X и Y есть выборки объемов n_1, n_2 из полиномиальных распределений $X \in \mathbf{B}_{\theta_1}$, $Y \in \mathbf{B}_{\theta_2}$, $\theta_i \in R^k$, $\theta_i = (\theta_{i1}, \ldots, \theta_{ik})$, $i = 1, 2$. Векторы частот $\nu = (\nu_1, \ldots, \nu_k)$ и $\mu = (\mu_1, \ldots, \mu_k)$ появлений событий $A_1, \ldots$

$\ldots, A_k$ (см. § 2.2[6]) образуют достаточные статистики,

$$f_{\theta_1}(X) = \prod_{i=1}^{k} \theta_{1i}^{\nu_i}, \quad f_{\theta_2}(Y) = \prod_{i=1}^{k} \theta_{2i}^{\mu_i}.$$

О. м. п. имеют вид $\widehat{\theta}_X^* = \nu/n_1$, $\widehat{\theta}_Y^* = \mu/n_2$, $\widehat{\theta}^* = (\nu + \mu)/(n_1 + n_2)$. Матрица $I(\theta)$ определена в (3.16.5)[6], так что (см. (3.16.9)[6])

$$tI(\theta_0)\, t^T = \sum_{i=1}^{k} \frac{t_i^2}{\theta_{0i}}.$$

Итак, в силу теоремы 1 и замечания 1 асимптотически минимаксный критерий асимптотического уровня $1-\varepsilon$ для проверки $H_1 = \{\theta_1 = \theta_2\}$ против

$$H_2^b = \left\{\sum_{i=1}^{k} (\theta_{1i} - \theta_{2i})^2/\theta_{0i} \geqslant b^2/n_2\right\}$$

имеет вид

$$\ln R_1(X, Y) = \sum_{i=1}^{k} \nu_i \ln \frac{\nu_i}{n_1} + \sum_{i=1}^{k} \mu_i \ln \frac{\mu_i}{n_2} - \sum (\nu_i + \mu_i) \ln \frac{\nu_i + \mu_i}{n_1 + n_2} > \frac{h_\varepsilon}{2},$$

где h_ε — квантиль порядка $1-\varepsilon$ распределения χ^2 с $k-1$ степенями свободы. Согласно (4), (5) асимптотически эквивалентным будет критерий

$$n_1 \sum_{i=1}^{k} \left(\frac{\nu_i}{n_1} - \frac{\nu_i + \mu_i}{n_1 + n_2}\right)^2 \frac{n_1 + n_2}{\nu_i + \mu_i} + n_2 \sum_{i=1}^{k} \left(\frac{\mu_i}{n_2} - \frac{\nu_i + \mu_i}{n_1 + n_2}\right)^2 \frac{n_1 + n_2}{\nu_i + \mu_i} = \sum_{i=1}^{k} \left(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2}\right)^2 \frac{n_1 n_2}{\nu_i + \mu_i} > h_\varepsilon. \quad (6)$$

Пример 1А. В примере 2.26.3[6] был описан механизм наследования групп крови, которые были обозначены 0 (нуль), A, B, AB. Он управляется с помощью генов трех типов A, B, 0. Вероятности появления этих генов в данной популяции обозначим соответственно p, q, $r = 1 - p - q$. Вероятности $p_i(\alpha)$, $\alpha = (p, q)$, того, что человек будет иметь i-ю группу крови, выражаются через α по формулам, приведенным в табл. 1 § 2.26[6].

Нам даны две выборки X и Y с частотами ν_i и μ_i, $i=1, \ldots, 4$, появления i-й группы крови, полученные в результате обследования $n_1=353$ человек из общины I и $n_2=364$ человек из общины II. Распределение людей по группам крови приведено в табл. 1.

Таблица 1

	0	A	B	AB	Всего
Община I	121	120	79	33	353
Община II	118	95	121	30	364
Всего	239	215	200	63	717

Надо проверить гипотезу о принадлежности обследуемых общин одной популяции, т. е. гипотезу о равенстве вероятностей p и q для этих групп, или, что то же, о равенстве вероятностей $p_i(\alpha)$. Это есть, очевидно, задача об однородности, рассмотренная в примере 1.

Если проверять совпадение вероятностей четырех групп крови, то статистике (см. (6))

$$\chi_1^2 = \sum_{i=1}^{4} \left(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2}\right)^2 \frac{n_1 n_2}{\nu_i + \mu_i}$$

будет соответствовать распределение χ^2 с тремя степенями свободы. В нашем случае значение χ_1^2 составляет 11,74. Фактически достигаемый (см. § 3.4[61]) уровень полученного уклонения превосходит 0,99. Это значит, что гипотезу об однородности следует забраковать с точки зрения критерия $\chi_1^2 > h_{0,01}$ уровня 0,99.

Отметим, что примененный критерий не полностью отвечает природе рассмотренного явления, так как мы должны проверять совпадение вероятностей p и q, а не вероятностей p_i появлений групп крови. Если следовать в точности теореме 1, то мы должны методами § 2.26[61] вычислить о. м. п. $\widehat{\alpha}_X^*$, $\widehat{\alpha}_Y^*$ и $\widehat{\alpha^*}$ параметра $\alpha=(p, q)$ соответственно по выборкам X, Y, (X, Y) и воспользоваться статистикой

$$\chi_2^2 = 2\left[L(\widehat{\alpha}_X^*, X) + L(\widehat{\alpha}_Y^*, Y) - L(\widehat{\alpha}^*, (X, Y))\right] =$$
$$= 2\left[\sum_{i=1}^{4} \nu_i \ln p_i(\widehat{\alpha}_X^*) + \sum_{i=1}^{4} \mu_i \ln p_i(\widehat{\alpha}_Y^*) - \right.$$
$$\left. - \sum_{i=1}^{4} (\nu_i + \mu_i) \ln p_i(\widehat{\alpha}^*)\right],$$

имеющей при больших n распределение, близкое к распределению χ^2 с двумя степенями свободы. Если провести необходимые вычисления (см. пример 2.26.3[61]), то мы получим $\chi_2^2 \approx 11{,}04$, что для двух степеней свободы дает более значимое уклонение, чем 11,74 для трех.

О проверке самой гипотезы о принадлежности X и Y параметрическим подсемействам $\mathrm{B}_{p(\alpha)}$, где $p(\alpha) = (p_1(\alpha), \ldots, p_4(\alpha))$, см. пример 3.17.1[61]. Обе выборки хорошо согласуются с этой гипотезой.

Пример 2. Пусть $X \Subset \Phi_{\alpha_1, \sigma_1^2}$, $Y \Subset \Phi_{\alpha_2, \sigma_2^2}$, где точки $\theta_i = (\alpha_i, \sigma_i^2)$ располагаются в окрестности точки $\theta_0 = (\alpha_0, \sigma_0^2)$. Здесь

$$I(\theta_0) = \begin{pmatrix} \sigma_0^{-2} & 0 \\ 0 & \frac{1}{2}\sigma_0^{-4} \end{pmatrix}$$

(см. § 2.16[61]), и мы будем рассматривать задачу о проверке гипотезы $H_1 = \{\theta_1 = \theta_2\}$ против

$$H_2^b = \left\{\frac{(\alpha_1 - \alpha_2)^2}{\sigma_0^2} + \frac{(\sigma_1^2 - \sigma_2^2)^2}{2\sigma_0^4} \geqslant \frac{b^2}{n_2}\right\}, \qquad n = n_1 + n_2.$$

Имеем $\widehat{\theta}_X^* = (\bar{x}, S_X^2)$, $S_X^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2$, $f_{\widehat{\theta}_X^*}(X) = (2\pi e S_X^2)^{-n_1/2}$. Аналогичные формулы справедливы для выборки Y. Далее,

$$\widehat{\theta}^* = (\bar{z}, S_{X,Y}^2), \qquad \bar{z} = \frac{\left(\sum_{i=1}^{n_1} x_i + \sum_{i=1}^{n_2} y_i\right)}{n_1 + n_2} = a\bar{x} + (1-a)\,\bar{y},$$

$$S_{X,Y}^2 = \frac{1}{n_1 + n_2}\left[\sum_{i=1}^{n_1}(x_i - \bar{z})^2 + \sum_{i=1}^{n_2}(y_i - \bar{z})^2\right] = \tag{7}$$

$$= aS_X^2 + (1-a)\,S_Y^2 + (1-a)\,a\,(\bar{x} - \bar{y})^2,$$

где $a = n_1/(n_1 + n_2)$, $f_{\widehat{\theta}^*}(X)\, f_{\widehat{\theta}^*}(Y) = (2\pi e S_{X,Y}^2)^{-\frac{1}{2}(n_1 + n_2)}$.

Таким образом, асимптотически минимаксным для проверки H_1 против H_2^b будет критерий

$$\frac{S^2_{X,Y}}{S^{2a}_X S^{2(1-a)}_Y} > e^{h_\varepsilon/(n_1+n_2)},$$

где h_ε — квантиль распределения χ^2 с двумя степенями свободы. Мы предлагаем читателю в виде упражнения найти асимптотически эквивалентный критерий вида (5).

3. Асимптотически минимаксные критерии для задачи об однородности при наличии мешающего параметра. В этом и последующих разделах мы будем считать для простоты, что объемы выборок X и Y совпадают: $n_1 = n_2$. Это ограничение несущественно. В случае $n_1/n_2 = r_1/r_2$ (r_1, r_2 — целые) читатель может от него избавиться самостоятельно так же, как это было сделано в замечании 1 к теореме 1.

Итак, пусть даны две выборки $X \in \mathbf{P}_{\theta_1}$, $Y \in \mathbf{P}_{\theta_2}, \theta_i = (u_i, v_i)$, $i = 1, 2$, объемов $n_1 = n_2 = n$. Проверяется гипотеза $\{u_1 = u_2\}$ против $\{u_1 \neq u_2\}$ в предположении, что $v_1 = v_2 = v$ и v нам неизвестно. Размерность u_i мы обозначим через l, $l < k$.

Введем новый параметр $\overline{\theta} = (u_1, u_2, v)$. Объединенную выборку (X, Y) представим как выборку объема n с наблюдениями $(x_1, y_1), \ldots, (x_n, y_n)$, плотность распределения которых равна $f_{\overline{\theta}}(x, y) = f_{(u_1, v)}(x) f_{(u_2, v)}(y)$. Для этого параметрического семейства рассматриваемая нами задача эквивалентна задаче о проверке гипотезы H_1, состоящей в том, что значение $\overline{\theta}$ лежит на «кривой» $\overline{\theta} = g(\theta_1) = (u_1, u_1, v)$, против дополнительной альтернативы. Матрица $G = \left\| \frac{\partial g_i}{\partial \theta_{1j}} \right\|$, $i = 1, \ldots, k + l$, $j = 1, \ldots, k$, имеет форму $\begin{pmatrix} E_l & 0 \\ & E_k \end{pmatrix}$, где сверху стоит единичная матрица порядка l, а внизу — единичная матрица порядка k, так что ранг G равен k.

Как и в предыдущем разделе, мы будем считать параметр θ локализованным около точки $\theta_0 = (u_0, v_0)$. Введем параметр $\tau = \tau(\overline{\theta}) = (\tau', \tau'', \tau''') = (u_1 - u_0, u_2 - u_1, v - v_0)$. Обратное отображение $\overline{\theta} = \overline{\theta}(\tau)$ всегда существует и имеет координаты $u_1 = \tau' + u_0$, $u_2 = \tau'' + \tau' + u_0$, $v = \tau''' + v_0$. Положим $\tau = \gamma/\sqrt{n}$, $\gamma = (\gamma', \gamma'', \gamma''')$.

Для нового параметра τ (или γ) гипотеза об однородности имеет вид $H_1 = \{\gamma'' = 0\}$. В качестве альтернативы рассмотрим «отделенную» гипотезу $H_2^b = \{\gamma'' I_1(\theta_0)\gamma''^T \geqslant b^2\}$, где $I_1(\theta)$ есть подматрица исходной информационной матрицы Фишера $I(\theta)$, образованная ее первыми l строками и столбцами.

Теорема 2. *Пусть в окрестности точки θ_0 семейство $\{\mathbf{P}_\theta\}$ удовлетворяет условиям (RR). Тогда критерий отношения правдоподобия*

$$R_1(X, Y) = \frac{\sup\limits_{(u_1, u_2, v)} f_{(u_1, v)}(X) f_{(u_2, v)}(Y)}{\sup\limits_{\theta} f_\theta(X) f_\theta(Y)} > e^{h_\varepsilon/2} \tag{8}$$

является асимптотически минимаксным критерием асимптотического уровня $1-\varepsilon$ для проверки $H_1 = \{u_1 = u_2\}$ против

$$H_2^b = \{(u_1 - u_2) I_1(\theta_0)(u_1 - u_2)^T \geqslant b^2/n\} \tag{9}$$

при общем значении $v_1 = v_2 = v$ и при любом $b > 0$. Здесь h_ε — квантиль порядка $1-\varepsilon$ распределения χ^2_l с l степенями свободы. (Таким будет предельное распределение $2\ln R_1(X, Y)$ при гипотезе H_1.)

Обозначим $\overline{\theta}^*$ значение параметра $\overline{\theta}$, при котором достигается максимум числителя в (8), и $\theta^* = (u^*, v^*)$ — значение θ, при котором достигается максимум знаменателя. Представим матрицу $I(\theta)$ в виде

$$I(\theta) = \begin{pmatrix} I_1(\theta) & I_{21}(\theta) \\ I_{12}(\theta) & I_{22}(\theta) \end{pmatrix}.$$

Тогда критерий

$$(\overline{\theta}^* - (u^*, u^*, v^*))\overline{I}(\overline{\theta}^*)(\overline{\theta}^* - (u^*, u^*, v^*))^T > h_\varepsilon/n, \tag{10}$$

где

$$\overline{I}(\overline{\theta}) = \begin{pmatrix} I_1(\theta_1) & 0 & I_{21}(\theta_1) \\ 0 & I_1(\theta_2) & I_{21}(\theta_2) \\ I_{12}(\theta_1) & I_{12}(\theta_2) & I_{22}(\theta_1) + I_{22}(\theta_2) \end{pmatrix}, \tag{11}$$

будут асимптотически эквивалентным (8).

Доказательство. Эта теорема также является прямым следствием теоремы 3.15.4[6]. Нам надо лишь выяснить структуру матрицы $\overline{I}(\overline{\theta})$ для выборки (X, Y) и «объединенного» параметра $\bar\theta$ и матрицы M_2. Имеем

$$l \equiv \ln f_{\bar\theta}(x, y) = l(x, (u_1, v)) + l(y, (u_2, v)).$$

Обозначим через t_i, $i = 1, \ldots, k+l$, координаты вектора $\bar\theta$. Тогда

$$\frac{\partial l}{\partial t_i} = \begin{cases} \dfrac{\partial l(x, (u_1, v))}{\partial t_i}, & 0 < i \leqslant l, \\ \dfrac{\partial l(y, (u_2, v))}{\partial t_i}, & l < i \leqslant 2l, \\ \dfrac{\partial l(x, (u_1, v))}{\partial t_i} + \dfrac{\partial l(y, (u_2, v))}{\partial t_i}, & 2l < i \leqslant k + l; \end{cases}$$

отсюда без труда получается (11).

Матрица M_2 для параметрического семейства $\mathbf{P}_{\bar\theta(0,\beta,0)} = \mathbf{P}_{(u_0, u_0+\beta, v_0)}$ точке $\beta = 0$ вычисляется аналогично, равна $I_1(\theta_0)$ и соответствует средней подматрице матрицы $I(\bar\theta_0)$.◁

В примерах объемы выборок n_1 и n_2 мы будем считать произвольными.

Пример 3. Пусть $X \in \Phi_{\alpha_1, \sigma^2}$, $Y \in \Phi_{\alpha_2, \sigma^2}$. Надо проверить гипотезу $H_1 = \{\alpha_1 = \alpha_2\}$, когда σ^2 неизвестно. Для отыскания с помощью теоремы 2 асимптотически минимаксных критериев нам надо найти статистику $R_1(X, Y)$ в (8), где в нашем случае $u_i = \alpha_i$, $v = \sigma^2$, $\bar\theta = (\alpha_1, \alpha_2, \sigma^2)$. Имеем $\ln f_{(\alpha_1, \sigma^2)}(X) f_{(\alpha_2, \sigma^2)}(Y) = -\frac{1}{2}(n_1 + n_2) \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n_1} (x_i - \alpha_1)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n_2} (y_i - \alpha_2)^2$. Обращая в нуль производные этой функции по α_1, α_2 и σ^2 и решая полученные уравнения, находим (в обозначениях примера 2)

$$\bar\theta^* = (\bar{x}, \bar{y}, aS_X^2 + (1-a) S_Y^2), \quad a = \frac{n_1}{n_1 + n_2}, \tag{12}$$

$$f_{\bar\theta^*}(X, Y) = \left[2\pi e \left(aS_X^2 + (1-a) S_Y^2\right)\right]^{-(n_1+n_2)/2}.$$

Поступая таким же образом с функцией $\ln f_\theta(X) f_\theta(Y) = \ln f_{(\alpha, \sigma^2)}(X) f_{(\alpha, \sigma^2)}(Y)$, получим (см. пример 2).

$$\theta^* = (\bar{z}, S_{X,Y}^2),$$

$$f_{\theta^*}(X) f_{\theta^*}(Y) = \left(2\pi e S_{X,Y}^2\right)^{-\frac{1}{2}(n_1+n_2)}. \tag{13}$$

Таким образом, асимптотически оптимальный критерий имеет вид

$$\frac{S^2_{X,Y}}{aS^2_X + (1-a)\,S^2_Y} > e^{h_\varepsilon/(n_1+n_2)},$$

или (см. (7))

$$\frac{\sqrt{a(1-a)}\,|\bar{x}-\bar{y}|}{\sqrt{aS^2_X + (1-a)\,S^2_Y}} > \sqrt{\frac{h_\varepsilon}{n_1+n_2}},$$

где h_ε — квантиль порядка $1-\varepsilon$ распределения χ^2 с одной степенью свободы, так что $\sqrt{h_\varepsilon}$ можно заменить на значение $\lambda_{\varepsilon/2}$, для которого $\Phi_{0,1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) = 1-\varepsilon$. Нетрудно видеть, что левая часть неравенства

$$\frac{\sqrt{a(1-a)(n_1+n_2)}\,|\bar{x}-\bar{y}|}{\sqrt{aS^2_X + (1-a)\,S^2_Y}} > \lambda_{\varepsilon/2}, \qquad (14)$$

определяющего асимптотически минимаксный критерий, после замены $|\bar{x}-\bar{y}|$ на $\bar{x}-\bar{y}$ будет асимптотически нормальной с параметрами (0, 1) случайной величиной.

Но этот критерий можно сделать точным (т. е. имеющим в точности наперед заданный уровень). Действительно, в силу результатов § 2.32[6] при гипотезе H_1

$$\sqrt{\frac{n_1 n_2}{n_1+n_2}}\,\frac{\bar{x}-\bar{y}}{\sigma} \in \Phi_{0,1},$$

$$\frac{(n_1+n_2)\,aS^2_X}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n_1}(x_i-\bar{x})^2 \in \mathrm{H}_{n_1-1},$$

$$\frac{(n_1+n_2)(1-a)\,S^2_Y}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n_2}(y_i-\bar{y})^2 \in \mathrm{H}_{n_2-1}.$$

Так как все три случайные величины независимы, то отношение

$$(\bar{x}-\bar{y})\sqrt{\frac{n_1 n_2}{n_1+n_2}}\left[\frac{n_1+n_2}{n_1+n_2-2}\left(aS^2_X + (1-a)\,S^2_Y\right)\right]^{-1/2} =$$

$$= \frac{(\bar{x}-\bar{y})\sqrt{a(1-a)(n_1+n_2-2)}}{\sqrt{aS^2_X + (1-a)\,S^2_Y}} \in \mathrm{T}_{n_1+n_2-2}$$

имеет распределение Стьюдента с n_1+n_2-2 степенями свободы. Таким образом, критерий (ср. с (14))

$$\frac{(\bar{x}-\bar{y})\sqrt{a(1-a)(n_1+n_2-2)}}{\sqrt{aS_X^2+(1-a)S_Y^2}} > \tau_\varepsilon,$$

где τ_ε такое, что $T_{n_1+n_2-2}(-\tau_\varepsilon, \tau_\varepsilon) = 1-\varepsilon$, будет иметь уровень значимости, в точности равный $1-\varepsilon$, и его можно использовать при любых (а не только больших) значениях n_1, n_2. Этот критерий называется критерием Стьюдента. Он обладает также некоторыми свойствами точной (а не только асимптотической) оптимальности (см. [18]).

Пример 4. Пусть $X \in \Phi_{\alpha,\sigma_1^2}$, $Y \in \Phi_{\alpha,\sigma_2^2}$. Проверяется гипотеза $\{\sigma_1=\sigma_2\}$, когда α неизвестно. Поступая аналогично предыдущему, мы придем к значению R_1 в (8), у которого знаменатель тот же, что и в предыдущем примере, а числитель равен

$$\sup_{(\alpha,\sigma_1,\sigma_2)} f_{(\alpha,\sigma_1^2)}(X)\, f_{(\alpha,\sigma_2^2)}(Y). \tag{15}$$

Выписывая уравнения для точки максимума, получим

$$\sigma_1^2 = \frac{1}{n_1}\sum_{i=1}^{n_1}(x_i-\alpha)^2 = S_X^2 + (\bar{x}-\alpha)^2,$$

$$\sigma_2^2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(y_i-\alpha)^2 = S_Y^2 + (\bar{y}-\alpha)^2,$$

$$\frac{n_1}{\sigma_1^2}(\bar{x}-\alpha) + \frac{n_2}{\sigma_2^2}(\bar{y}-\alpha) = 0.$$

Отсюда, полагая

$$p = \frac{a}{\sigma_1^2}\cdot\frac{1}{a/\sigma_1^2+(1-a)/\sigma_2^2} \in (0, 1), \tag{16}$$

находим

$$\alpha = p\bar{x} + (1-p)\bar{y},$$

$$\sigma_1^2 = S_X^2 + (1-p)^2\Delta^2, \quad \sigma_2^2 = S_Y^2 + p^2\Delta^2,$$

где мы для краткости положили $\Delta = \bar{x}-\bar{y}$; p можно рассматривать как решение уравнения (16) или

$$p = \frac{a\left(S_Y^2+p^2\Delta^2\right)}{a\left(S_Y^2+p^2\Delta^2\right)+(1-a)\left(S_X^2+(1-p)^2\Delta^2\right)}.$$

Так как максимум в (15) равен

$$(2\pi e)^{-(n_1+n_2)/2}\left(S_X^2+(1-p)^2\Delta^2\right)^{-n_1/2}\left(S_Y^2+p^2\Delta^2\right)^{-n_2/2}, \quad (17)$$

то, сравнивая это с (13), (7), мы получим асимптотически минимаксный критерий

$$\frac{aS_X^2+(1-a)S_Y^2+a(1-a)\Delta^2}{\left(S_X^2+(1-p)^2\Delta^2\right)^a\left(S_Y^2+p^2\Delta^2\right)^{1-a}}>e^{h_\varepsilon/(n_1+n_2)}, \quad (18)$$

или

$$\frac{aS_X^2+(1-a)S_Y^2}{S_X^{2a}S_Y^{2(1-a)}}>e^{h_\varepsilon/(n_1+n_2)}A^{-1}, \quad (19)$$

где $A=\dfrac{1+\dfrac{a(1-a)\Delta^2}{aS_X^2+(1-a)S_Y^2}}{\left(1+(1-p)^2\Delta^2/S_X^2\right)^a\left(1+p^2\Delta^2/S_Y^2\right)^{1-a}}$, h_ε есть квантиль распределения χ^2 с одной степенью свободы. Здесь $\Delta^2=\left(\sigma_1^2/n_1+\sigma_2^2/n_2\right)\xi^2$, $\xi\in\Phi_{0,1}$, $S_X^2/\sigma_1^2\xrightarrow[P]{}1$, $S_Y^2/\sigma_2^2\xrightarrow[P]{}1$, $\sigma_1^2/\sigma_2^2\to1$, $p\to a$ (можно считать для простоты, что $a=\dfrac{n_1}{n_1+n_2}$ фиксировано), $\ln A\xrightarrow[P]{}0$ для каждой из рассматриваемых близких гипотез. Стало быть, правая часть в (19) имеет вид

$$1+\frac{h_\varepsilon+\delta_n}{n_1+n_2},\qquad \delta_n\xrightarrow[P]{}0.$$

Левая часть (19) представляет собой отношение среднего арифметического к среднему геометрическому величин S_X^2 и S_Y^2. Если обозначить $S_X^2/S_Y^2=Z^2$, то неравенство, обратное к (19), можно записать в виде

$$\frac{aZ^2+(1-a)}{Z^{2a}}-1\leqslant\frac{h_\varepsilon+\delta_n}{n_1+n_2}. \quad (20)$$

В левой части здесь стоит выпуклая вниз функция от Z (можно для определенности считать $a\leqslant1/2$), имеющая в точке $Z=1$ кратный нуль. Так как правая часть этого неравенства мала, то решение удобно искать в виде $Z^2=1+\zeta$ при малом ζ. Пользуясь разложением в ряд по степеням ζ и отбрасывая члены третьего и более высоких порядков малости, мы получим для границ ζ_1, ζ_2 интервала, в котором справедливо (20), значения

$$\zeta_1 = \sqrt{\frac{2(h_\varepsilon + \delta_n')}{a(1-a)(n_1+n_2)}}, \quad \zeta_2 = \sqrt{\frac{2(h_\varepsilon + \delta_n'')}{a(1-a)(n_1+n_2)}},$$

$$\delta_n' \xrightarrow[P]{} 0, \quad \delta_n'' \xrightarrow[P]{} 0.$$

Это означает, если вернуться к исходным переменным, что область

$$\sqrt{\frac{1}{2}a(1-a)(n_1+n_2)}\,|S_X^2/S_Y^2 - 1| > \sqrt{h_\varepsilon} = \lambda_{\varepsilon/2} \qquad (21)$$

(λ_ε определено в примере 3) определяет критерий, асимптотически эквивалентный (18) и, стало быть, асимптотически минимаксный.

Здесь, как и в примере 3, мы можем сделать полученный критерий точным, поскольку нам известно точное распределение статистики S_X^2/S_Y^3. Действительно,

$$n_1 S_X^2/\sigma_1^2 \in \mathrm{H}_{n_1-1}, \quad n_2 S_Y^2/\sigma_2^2 \in \mathrm{H}_{n_2-1}$$

и при гипотезе $H_1 = \{\sigma_1 = \sigma_2\}$

$$\frac{n_1 S_X^2}{n_2 S_Y^2} \in \mathbf{F}_{n_1-1,n_2-1},$$

где $\mathbf{F}_{n_1-1,n_2-1}$ есть распределение Фишера, введенное нами в § 2.2[6] и табулированное в руководствах по математической статистике. Это означает, что можно вычислить точный уровень значимости критерия (21) и применять его при любых n_1 и n_2 (о свойствах точной оптимальности этого критерия см. [18]). Если n_1 и n_2 велики, то левая часть в (21) (без знака абсолютной величины) асимптотически нормальна с параметрами (0, 1).

4. Асимптотически минимаксный критерий для задачи о частичной однородности. Пусть $X \in \mathbf{P}_{\theta_1}$, $Y \in \mathbf{P}_{\theta_2}$, $\theta_i = (u_i, v_i)$, $i = 1, 2$. Проверяется гипотеза $\{u_1 = u_2\}$ против $\{u_1 \neq u_2\}$, когда значения v_1 и v_2 в выборках X и Y могут быть любыми. Размерность u_i, как и прежде, обозначим через l, $l < k$.

Введем новый параметр $\bar\theta = (\theta_1, \theta_2) = (u_1, v_1, u_2, v_2)$ размерности $2k$. Как и раньше, выборку (X, Y) (при $n_1 = n_2 = n$) представим как выборку с наблюдениями

$(x_1, y_1), \ldots, (x_n, y_n)$, имеющими плотность

$$f_{\bar\theta}(x, y) = f_{(u_1, v_1)}(x)\, f_{(u_2, v_2)}(y).$$

Для этого семейства исходная задача о частичной однородности эквивалентна задаче о проверке гипотезы H_1, состоящей в том, что $\bar\theta$ лежит на «кривой» $\theta = g(\alpha) = (u_1, v_1, u_1, v_2)$, где $\alpha = (u_1, v_1, v_2)$ есть «подпараметр» размерности $2k - l$. Мы предлагаем читателю, следуя рассуждениям двух предыдущих разделов, выписать матрицу $G = \left\| \frac{\partial g_i}{\partial \alpha_j} \right\|$, $i = 1, \ldots, 2k$; $j = 1, \ldots, 2k - l$. Ее ранг будет равен $2k - l$.

Как и в разделах 2, 3, мы будем считать задачу «локализованной» около точки $\theta_0 = (u_0, v_0)$. Введем наряду с $\bar\theta$ параметр $\tau = \tau(\bar\theta) = (\tau', \tau'', \tau''', \tau^{\mathrm{IV}}) = (u_1 - u_0, v_1 - v_0, u_2 - u_1, v_2 - v_0)$. Обратное преобразование $\bar\theta = \bar\theta(\tau)$ имеет координаты

$$u_1 = \tau' + u_0, \quad v_1 = \tau'' + v_0, \quad u_2 = \tau''' + \tau' + u_0,$$
$$v_2 = \tau^{\mathrm{IV}} + v_0.$$

Если положить $\tau = \gamma/\sqrt{n}$, $\gamma = (\gamma', \gamma'', \gamma''', \gamma^{\mathrm{IV}})$, то гипотеза H_1 будет иметь вид $H_1 = \{\gamma''' = 0\}$. В качестве альтернативы будем рассматривать «отделенную» гипотезу $H_2^b = \{\gamma''' I_1(\theta_0)\gamma'''^T \geqslant b^2\}$, где $I_1(\theta)$ имеет тот же смысл, что и в теореме 2.

Теорема 3. *Пусть в окрестности точки θ_0 семейство $\{P_\theta\}$ удовлетворяет условиям (RR). Тогда критерий отношения правдоподобия*

$$R_1(X, Y) = \frac{\sup\limits_{(\theta_1, \theta_2)} f_{\theta_1}(X)\, f_{\theta_2}(Y)}{\sup\limits_{(u, v_1, v_2)} f_{(u, v_1)}(X)\, f_{(u, v_2)}(Y)} > e^{h_\varepsilon/2} \qquad (22)$$

является асимптотически минимаксным критерием асимптотического уровня $1 - \varepsilon$ для проверки H_1 против гипотезы H_2^b, определенной в (9), при произвольных значениях v_1 и v_2. Значение h_ε здесь то же, что и в теореме 2.

Доказательство этой теоремы повторяет рассуждения предыдущих разделов и тоже полностью основано на теореме 3.15.4[6]. Отыскание информационной матрицы Фишера $I(\bar\theta)$ для параметра $\bar\theta$ и матрицы M_2 для семейства с плотностью $f_{\bar\theta(0,0,\beta,0)} = f_{(u_0,v_0,u_0+\beta,v_0)}$ в точке $\beta = 0$ мы предоставляем читателю.

С помощью матрицы $\bar I((\hat\theta_X^*, \hat\theta_Y^*))$ и вектора $(\hat\theta_X^*, \hat\theta_Y^*) - (u^*, v_1^*, u^*, v_2^*)$, где $(\hat\theta_X^*, \hat\theta_Y^*)$ и (u^*, v_1^*, v_2^*) есть векторы, в которых достигаются максимумы числителя и знаменателя в (22), можно, как и прежде, с помощью теоремы 3.15.4[6] (см. (3.15.12)[6]) построить асимптотически эквивалентный критерий, использующий квадратичную форму от введенных оценок. ◁

Пример 5. *Сравнение дисперсий нормальных совокупностей. Пусть* $X \in \Phi_{\alpha_1,\sigma_1^2}$, $Y \in \Phi_{(\alpha_2,\sigma_2^2)}$, $H_1 = \{\sigma_1 = \sigma_2\}$. Здесь вычисления значительно проще, чем в примере 4, поскольку значение числителя в (22) (как и вектор $(\hat\theta_X^*, \hat\theta_Y^*) = (\bar x, S_X^2, \bar y, S_Y^2)$) нам известно, а значение знаменателя найдено в примере 3 (см. (12)). Неравенство (22) здесь будет иметь вид

$$\frac{aS_X^2 + (1-a)\,S_Y^2}{S_X^{2a} S_Y^{2(1-a)}} > e^{h_\varepsilon/(n_1+n_2)}.$$

Сравнивая это с (19) и с последующими рассмотрениями, мы придем к тем же критериям и выводам, что и в примере 4.

Пример 6. *Проблема Беренса — Фишера о сравнении средних двух нормальных совокупностей.* Пусть $X \in \Phi_{\alpha_1,\sigma_1^2}$, $Y \in \Phi_{\alpha_2,\sigma_2^2}$, $H_1 = \{\alpha_1 = \alpha_2\}$, значения σ_1 и σ_2 произвольны. Для этого примера числитель в (22) тот же, что и в предыдущем примере, а знаменатель найден в примере 4 (см. (17); там это был числитель для (8)).

Асимптотически минимаксный критерий, стало быть, будет иметь вид

$$\left(\frac{S_X^2 + (1-p)^2\Delta^2}{S_X^2}\right)^a \left(\frac{S_Y^2 + p^2\Delta^2}{S_Y^2}\right)^{1-a} > e^{h_\varepsilon/(n_1+n_2)}; \qquad (23)$$

здесь $\Delta = \bar x - \bar y$ представимо в виде

$$\Delta = (\alpha_1 - \alpha_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\,\xi, \quad \xi \in \Phi_{0,1},$$

$$S_X^2/\sigma_1^2 \xrightarrow[p]{} 1, \quad S_Y^2/\sigma_2^2 \xrightarrow[p]{} 1,$$

так что $\Delta \underset{P}{\rightarrow} 0$ при гипотезе H_1. Это соотношение сохранится, очевидно, и для каждой из близких альтернатив. Чтобы найти более простой по форме критерий, асимптотически эквивалентный (23), выделим в обеих частях неравенства (23) главные части. Получим

$$\frac{a(1-p)^2\Delta^2}{S_X^2} + \frac{(1-a)p^2\Delta^2}{S_Y^2} + \Delta^4\rho_n > \frac{h_\varepsilon}{n_1+n_2} + O\left(\frac{1}{(n_1+n_2)^2}\right),$$

где $\rho_n \underset{P}{\rightarrow} \rho = \mathrm{const}$. Учитывая, что

$$p = \frac{aS_Y^2}{aS_Y^2 + (1-a)S_X^2} + \Delta^2\rho_n', \qquad \rho_n' \underset{P}{\rightarrow} \rho' = \mathrm{const},$$

мы получим

$$\frac{a(1-a)^2S_X^2\Delta^2(n_1+n_2) + a^2(1-a)S_Y^2\Delta^2(n_1+n_2)}{\left(aS_Y^2+(1-a)S_X^2\right)^2} + \\ + \Delta^4(n_1+n_2)\rho_n'' > h_\varepsilon + O\left(\frac{1}{n_1+n_2}\right);$$

где $\rho_n'' \rightarrow \rho'' = \mathrm{const}$, $\Delta^4(n_1+n_2) \underset{P}{\rightarrow} 0$. Это неравенство можно эквивалентным образом записать в форме

$$\frac{\Delta^2(n_1+n_2)}{S_X^2/a + S_Y^2/(1-a)} > h_\varepsilon + \delta_n, \quad \delta_n \underset{P}{\rightarrow} 0.$$

Отсюда вытекает, что критерий

$$\frac{|\bar{x}-\bar{y}|\sqrt{n_1+n_2}}{\sqrt{S_X^2/a + S_Y^2/(1-a)}} > \sqrt{h_\varepsilon} = \lambda_{\varepsilon/2} \tag{24}$$

является асимптотически эквивалентным (23) и, стало быть, асимптотически минимаксным для проблемы Беренса — Фишера. Здесь $\lambda_{\varepsilon/2}$ имеет тот же смысл, что и в примере 4. В отличие от примеров 2—4, здесь допредельное распределение статистики в левой части (24) зависит при гипотезе H_1 от неизвестных параметров σ_1^2, σ_2^2.

5. Некоторые другие задачи. Мы отметим здесь еще два класса задач, асимптотическое решение которых может быть найдено с помощью теоремы 3.15.4[6].

1) Первый класс задач составляют задачи, обобщающие задачи разделов 2—4 на случай, когда проверяются

гипотезы вида $\{\theta_1 = f(\theta_2)\}$ (например, $\{\theta_1 = a + b\theta_2\}$) в условиях п. 2 и вида $\{u_1 = f(u_2)\}$ в условиях пп. 3, 4. Нетрудно видеть, что проведенные в пп. 2—4 рассмотрения переносятся на этот более общий случай.

2) Второй класс задач составляют задачи с тремя и более выборками. Рассмотрим, например, задачу об однородности для трех выборок. Пусть $X \in \mathbf{P}_{\theta_1}$, $Y \in \mathbf{P}_{\theta_2}$, $Z \in \mathbf{P}_{\theta_3}$. Проверяется гипотеза $H_1 = \{\theta_1 = \theta_2 = \theta_3\}$ против дополнительной альтернативы. Предположим для простоты, что объемы выборок n_1, n_2, n_3 равны $n_1 = n_2 = n_3 = n$. Рассмотрим объединенную выборку (X, Y, Z) как выборку объема n с наблюдениями $(x_1, y_1, z_1), \ldots, (x_n, y_n, z_n)$, имеющими плотность $f_{\overline{\theta}}(x, y, z) = f_{\theta_1}(x) f_{\theta_2}(y) f_{\theta_3}(z)$, где $\overline{\theta} = (\theta_1, \theta_2, \theta_3)$. Гипотеза H_1 тогда будет эквивалентна тому, что $\overline{\theta}$ лежит на «кривой» $\overline{\theta} = g(\alpha)$, $\alpha \equiv \theta_1$, $g(\alpha) = (\alpha, \alpha, \alpha)$. Мы видим, что проблема вновь сводится к задаче, рассмотренной в теореме 3.15.4[6].

§ 2. Задачи об однородности в общем случае

1. Постановка задачи. В этом параграфе мы будем рассматривать две выборки X и Y объемов n_1 и n_2 соответственно, не предполагая, что они относятся к какому-либо параметрическому семейству.

Задача об однородности выборок X и Y в общем случае выглядит следующим образом. Обозначим $\mathbf{P}_1$ и $\mathbf{P}_2$ распределения выборок X и Y: $X \in \mathbf{P}_1$, $Y \in \mathbf{P}_2$. Проверяется гипотеза $H_1 = \{\mathbf{P}_1 = \mathbf{P}_2\}$ против $H_2 = \{\mathbf{P}_1 \neq \mathbf{P}_2\}$. Обе гипотезы, очевидно, являются сложными. Распределения $\mathbf{P}_1$, $\mathbf{P}_2$ могут выбираться из заданного семейства $\mathcal{P}$ или быть произвольными. Общий принцип построения статистического критерия для проверки H_1 против H_2 остается тем же, что и в гл. 3[6]. Разница, как и в § 1, состоит лишь в том, что здесь он основан на объединенной выборке (X, Y), так что $\pi = \pi(X, Y)$ есть вероятность принять H_2 при данной выборке (X, Y). В нерандомизированном случае ($\pi = 0$ или 1) критерий π определяется критической областью $\Omega \subset \mathscr{X}^{n_1+n_2}$ такой, что при $(X, Y) \in \Omega$ принимается H_2. Число

$$1 - \varepsilon = \inf_{\mathbf{P}_1 \in \mathcal{P}} \mathbf{P}_1 \times \mathbf{P}_1 ((X, Y) \notin \Omega)$$

называется *уровнем значимости*, а значение

$$\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) = \mathbf{P}_1 \times \mathbf{P}_2((X, Y) \in \Omega), \ \mathbf{P}_1 \in \mathscr{P}, \ \mathbf{P}_2 \in \mathscr{P},$$

— мощностью критерия π в «точке» $(\mathbf{P}_1, \mathbf{P}_2)$.

Критерий π называется *состоятельным*, если $\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) \to 1$ при $n_1 \to \infty$, $n_2 \to \infty$ и при любых $\mathbf{P}_1 \neq \mathbf{P}_2$, $\mathbf{P}_1 \in \mathscr{P}$, $\mathbf{P}_2 \in \mathscr{P}$.

Мы уже знаем, что с ростом n_1, n_2 эмпирические распределения $\mathbf{P}_X^*$, $\mathbf{P}_Y^*$, соответствующие выборкам X и Y, неограниченно сближаются с $\mathbf{P}_1$ и $\mathbf{P}_2$ соответственно. Поэтому естественной основой для построения критериев однородности является использование разного рода «расстояний» $d(\mathbf{P}_X^*, \mathbf{P}_Y^*)$ между $\mathbf{P}_X^*$ и $\mathbf{P}_Y^*$, где d удовлетворяет тем же общим условиям, которые описаны в § 3.12 [6]. При этом особый интерес представляют непараметрические и асимптотически непараметрические критерии, которые определяются следующим образом.

Пусть $d(\mathbf{P}, \mathbf{Q})$ — некоторое расстояние (не обязательно метрика) в пространстве распределений. Если вероятность

$$\mathbf{P}_1 \times \mathbf{P}_1 \left(d\left(\mathbf{P}_X^*, \mathbf{P}_Y^*\right) > c\right) = \varepsilon \tag{1}$$

не зависит от выбора $\mathbf{P}_1$, то критерий π, определенный равенствами

$$\pi(X, Y) = \begin{cases} 0, & \text{если } d\left(\mathbf{P}_X^*, \mathbf{P}_Y^*\right) \leqslant c, \\ 1 & \text{в противном случае,} \end{cases} \tag{2}$$

называется *непараметрическим*. Очевидно, построенный непараметрический критерий будет иметь уровень $1-\varepsilon$.

Аналогичным образом определяются асимптотически непараметрические критерии, когда (1) сохраняется асимптотически, при добавлении в левой части операции $\lim\limits_{n_1\to\infty, n_2\to\infty}$. В этом случае критерий (2) будет иметь асимптотический уровень $1-\varepsilon$. При отсутствии непараметричности (точной или асимптотической) строить критерии проверки однородности заданного уровня весьма трудно.

Рассмотрим некоторые основные критерии проверки однородности.

2. Критерий Колмогорова — Смирнова. Пусть $\mathbf{P}_1$, $\mathbf{P}_2$ принадлежат классу $\mathscr{P}$ всех непрерывных распределений на прямой, и пусть F_X^* и F_Y^* — эмпирические функции распределения, соответствующие $\mathbf{P}^*_X$ и $\mathbf{P}^*_Y$. Критерий Колмогорова — Смирнова в качестве расстояния $d(\mathbf{P}^*_X, \mathbf{P}^*_Y)$ рассматривает статистику

$$D_{n_1,n_2} = \sup_t \left| F^*_X(t) - F^*_Y(t) \right|.$$

Критерий $D_{n_1,n_2} > c$, построенный с помощью статистики D_{n_1,n_2}, *является непараметрическим*. Действительно, пусть верна гипотеза H_1 и $F(t)$ есть общая функция распределения для X и Y. Статистику D_{n_1,n_2} можно записать в виде

$$D_{n_1,n_2} = \sup_t \left| G^*_X(F(t)) - G^*_Y(F(t)) \right|, \tag{3}$$

где $G^*_X(u) = F^*_X(F^{-1}(u))$ есть эмпирическая функция распределения, соответствующая равномерному на [0, 1] распределению (см. §§ 1.6[6], 3.12[6]). Но в силу (3) $D_{n_1,n_2} = \sup_u \left| G^*_X(u) - G^*_Y(u) \right|$, так что распределение D_{n_1,n_2} от F никак не зависит.

Можно найти точное распределение статистики D_{n_1,n_2}. Например, при $n_1 = n_2 = n$

$$\mathbf{P}(nD_{n,n} \geqslant k) = 2\left(C^n_{2n}\right)^{-1} \sum_{j=1}^{[n/k]} (-j)^{j+1} C^{n-jk}_{2n}, \tag{4}$$

$k = 1, 2, \ldots, n$. Этот факт был установлен Гнеденко и Королюком сведением задачи к простой проблеме для случайных блужданий (см. [33]).

В § 1.6[6] мы видели, что распределение $n_1 G^*_X(u)$ совпадает с распределением пуассоновского процесса $\zeta_1(u)$ при условии $\zeta_1(1) = n_1$. Так как $G^*_X(u)$ и $G^*_Y(u)$ независимы, то распределение $G^*_X(u) - G^*_Y(u)$, $u \in [0, 1]$, совпадает с распределением сложного пуассоновского процесса $\zeta(u)$, в котором с интенсивностью n_1 происходят скачки величиной $1/n_1$ и с интенсивностью n_2 — скачки величиной $-1/n_2$; распределение надо брать при условии, что произошло $n_1 + n_2$ скачков и что $\zeta(1) = 0$. Поэтому

$$\mathbf{P}\left(D_{n_1,n_2} < x\right) =$$
$$= \mathbf{P}\left(\sup_{u<1} |\zeta(u)| < x / \zeta(1) = 0;\ \textit{произошло}\ n_1 + n_2\ \textit{скачков}\right).$$

На основании этого факта в Приложении II[6] наряду с теоремой 1.6.2[6] о сходимости процесса $w_n(u) = \sqrt{n_1}\left(G_X^*(u) - u\right)$ к броуновскому мосту $w^\circ(u)$ доказано также утверждение о сходимости к броуновскому мосту и процесса

$$w_{n_1,n_2}(u) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\left(G_X^*(u) - G_Y^*(u)\right).$$

Точнее, для любого измеримого и непрерывного в равномерной метрике функционала f распределение $f\left(w_{n_1,n_2}\right)$ сходится к распределению $f(w^\circ)$. Отсюда сразу вытекает следующее утверждение, называемое теоремой Смирнова.

Теорема 1.

$$\lim_{n_1\to\infty,\, n_2\to\infty} \mathbf{P}\left(\sqrt{\frac{n_1 n_2}{n_1 + n_2}}\, D_{n_1,n_2} < x\right) = \mathbf{P}\left(\sup_{u\leqslant 1} |w^\circ(u)| < x\right) = K(x),$$

где $K(x)$ — *функция Колмогорова* (см. §§ 1.8[6], 3.12[6]).

Так как функция $K(x)$ табулирована, то теорема 1 дает удобное средство для приближенного вычисления уровня значимости критерия Колмогорова — Смирнова.

Мы предоставляем читателю убедиться, что критерий Колмогорова — Смирнова является состоятельным.

3. Критерий знаков. Пусть $n_1 = n_2 = n$. Тогда из наблюдений выборок X и Y можно составить n разностей

$$\mathrm{x}_1 - \mathrm{y}_1, \ldots, \mathrm{x}_n - \mathrm{y}_n. \tag{5}$$

Если верна гипотеза H_1 и $\mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 = 0) = 0$ при всех $\mathbf{P}_1 \in \mathscr{P}$ (это, очевидно, всегда так, если $\mathscr{P}$ — множество непрерывных распределений), то

$$\mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 > 0) = \mathbf{P}_1 \times \mathbf{P}_1(\mathrm{x}_1 - \mathrm{y}_1 < 0) = 1/2.$$

Статистика ν критерия знаков представляет собой число положительных разностей в (5) *). Сам критерий можно построить, взяв в качестве критического множества

$$\Omega = \left\{(X, Y)\colon \left|\nu - \frac{n}{2}\right| > c\right\}.$$

*) Если в выборках X, Y из-за огрубления данных окажется, что некоторые разности $\mathrm{x}_i - \mathrm{y}_i = 0$, то их следует просто выбросить, а в качестве n взять число разностей, отличных от нуля.

Так как распределение ν от $\mathbf{P}_1$ не зависит,

$$\mathbf{P}_1 \times \mathbf{P}_1 (\nu = k) = C_n^k 2^{-n},$$

то этот критерий является непараметрическим.

Число c по заданному уровню $1-\varepsilon$ критерия выбирается из соотношения

$$\sum_{k:\ |2k-n|<2c} C_n^k 2^{-n} \geqslant 1-\varepsilon. \tag{6}$$

Так как левая часть здесь возрастает с ростом c дискретно, то в качестве решения следует взять наименьшее c, при котором левая часть в (6) превосходит $1-\varepsilon$.

Мы видим, что здесь используется критерий для проверки гипотезы о том, что вероятность успеха в схеме Бернулли равна 1/2. С точки зрения исходной задачи проверяется не гипотеза об однородности, а более широкая гипотеза о том, что

$$\mathbf{P}_1 \times \mathbf{P}_2 (\mathrm{x}_1 - \mathrm{y}_1 < 0) = \int F_1(t)\, dF_2(t) = 1/2, \tag{7}$$

где F_i соответствуют $\mathbf{P}_i$, $i = 1, 2$. Соотношение (7) означает, что медиана распределения $\mathrm{x}_1 - \mathrm{y}_1$ равна 0.

Критерий знаков асимптотического уровня $1-\varepsilon$ будет иметь вид

$$\begin{gathered} \pi(X, Y) = 1, \text{ если } \frac{2\left|\nu - \frac{n}{2}\right|}{\sqrt{n}} > \lambda_{\varepsilon/2}, \\ \Phi_{0,1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) = 1-\varepsilon. \end{gathered} \tag{8}$$

Этот критерий не является состоятельным, так как для $\mathbf{P}_1 \neq \mathbf{P}_2$, удовлетворяющих (7), $\beta_\pi(\mathbf{P}_1, \mathbf{P}_2) \to \varepsilon < 1$ при $n_1 \to \infty$, $n_2 \to \infty$.

4. Критерий Вилкоксона. Этот критерий весьма распространен при проверке гипотез об однородности.

Объединим выборки X и Y в одну выборку (X, Y) и построим из нее вариационный ряд, т. е. расположим все наблюдения в порядке возрастания. Получим последовательность вида

$$\mathrm{y}^{(1)}, \mathrm{y}^{(2)}, \mathrm{x}^{(3)}, \mathrm{y}^{(4)}, \mathrm{x}^{(5)}, \ldots, \tag{9}$$

где верхний индекс означает номер наблюдения в общем вариационном ряде, а буква указывает на принадлежность к выборке. Пусть $r_1, r_2, \ldots, r_{n_1}$ означают номера элементов выборки X в вариационном ряде (9). Для выписанной в (9) последовательности $r_1=3$, $r_2=5$. *Статистикой Вилкоксона* называется функция

$$U=U(X,Y)=\sum_{i=1}^{n_1}(r_i-i),$$

где r_i-i есть число элементов выборки Y, меньших чем $x_{(i)}$.

Так как порядок наблюдений в (9) инвариантен относительно монотонных преобразований над переменными (порядок для $F_X^*(t)$, $F_Y^*(t)$ будет тем же, что и для $F_X^*(F^{-1}(t))$, $F_Y^*(F^{-1}(t))$, где F — функция распределения), то критерий, построенный по статистике U, будет непараметрическим.

Теорема 2. *Пусть $X \Subset \mathbf{P}_1$, $Y \Subset \mathbf{P}_2$, $F_i \in \mathscr{F}$ есть функции распределения, соответствующие $\mathbf{P}_i$, $i=1, 2$; $\mathscr{F}$ есть класс всех непрерывных функций распределения. Предположим также, что $a=n_1/(n_1+n_2) \to a_0$ при $n_1 \to \infty$, $n_2 \to \infty$. Тогда*

$$\frac{U-n_1n_2\mathsf{M}F_2(x_1)}{\sqrt{n_1n_2(n_1+n_2)}} \Leftrightarrow \Phi_{0,\sigma^2}, \tag{10}$$

где $\sigma^2=(1-a_0)\mathsf{D}F_2(x_1)+a_0\mathsf{D}F_1(y_1)$.

Если $F_1=F_2=F$, *то* $F_2(x_1)\Subset \mathbf{U}_{0,1}$, $F_1(y_1)\Subset \mathbf{U}_{0,1}$ *и, стало быть,* $\mathsf{M}F_2(x_1)=1/2$, $\mathsf{D}F_2(x_1)=\mathsf{D}F_1(y_1)=1/12$.

Следовательно, критерий Вилкоксона асимптотического уровня $1-\varepsilon$ будет иметь вид

$$\left|U-\frac{n_1n_2}{2}\right| > \frac{\lambda_{\varepsilon/2}\sqrt{n_1n_2(n_1+n_2)}}{2\sqrt{3}}, \tag{11}$$

$$\Phi_{0,1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2})=1-\varepsilon.$$

Из (10) видно, что этот критерий нацелен главным образом на проверку гипотезы (ср. с (7))

$$\int F_2(t)\,dF_1(t)=1/2, \quad \text{или} \quad \int (F_2(t)-F_1(t))\,dF_1(t)=0. \tag{12}$$

Если принять, не ограничивая общности, что $F_1(t)\equiv t$, $t\in[0, 1]$, и предположить, что $F_2(0)=0$, $F_2(1)=1$, то в силу равенства

$$\int_0^1 (1-F_2(t))\,dt=\mathsf{M}y_1$$

проверяемая гипотеза примет вид $\mathsf{M}y_1=1/2$.

Это значит, что критерий Вилкоксона, как и критерий знаков, чувствителен главным образом к сдвигам распределений одно относительно другого. Для таких сдвинутых альтернатив мощность их бывает достаточно большой (см. пример 1). Если же $F_2 \neq F_1$ и выполнено (12), то согласно критерию Вилкоксона гипотеза $\{F_2 = F_1\}$ будет приниматься с вероятностью, близкой к $\Phi_{0,1}\left(-\frac{\lambda_{\varepsilon/2}}{2\sqrt{3}\sigma}, \frac{\lambda_{\varepsilon/2}}{2\sqrt{3}\sigma}\right)$. Это означает, что критерий Вилкоксона будет несостоятельным.

Доказательство теоремы 2. Статистику U можно записать в виде

$$U = \sum_{i=1}^{n_1} n_2 F_Y^*(x_i) = n_1 n_2 \int F_Y^*(t)\, dF_X^*(t).$$

Обозначим

$$w_X(t) = \sqrt{n_1}\left(F_X^*(t) - F_1(t)\right), \quad w_Y(t) = \sqrt{n_2}\left(F_Y^*(t) - F_2(t)\right).$$

Тогда, очевидно,

$$U = n_1 n_2 \int F_2(t)\, dF_1(t) + \sqrt{n_1 n_2 (n_1 + n_2)} \times$$
$$\times \left[\sqrt{a} \int w_Y(t)\, dF_1(t) + \sqrt{1-a} \int F_2(t)\, dw_X(t)\right] +$$
$$+ \sqrt{n_1 n_2} \int w_Y(t)\, dw_X(t). \qquad (13)$$

Так как здесь $\int F_2(t)\, dw_X(t) = \int w_X(t)\, dF_2(t)$ и, стало быть, второй и третий интегралы в (13) имеют одну и ту же форму и независимы, то для доказательства теоремы нам достаточно убедиться, что

$$\int w_Y(t)\, dF_1(t) \Leftrightarrow \Phi_{0,\sigma_1^2}, \quad \sigma_1^2 = \mathsf{D}F_1(y_1), \qquad (14)$$

и что

$$\frac{1}{\sqrt{n_1 + n_2}} \int w_Y(t)\, dw_X(t) \xrightarrow[\mathrm{P}]{} 0. \qquad (15)$$

В силу теоремы 1.6.2[6]

$$\int w_Y(t)\, dF_1(t) \Leftrightarrow \int w^\circ(F_2(t))\, dF_1(t), \qquad (16)$$

где $w^\circ(u)$ есть броуновский мост. Чтобы найти распределение последнего интеграла, заметим, что траектории винеровского процесса $w(u)$ с вероятностью 1 непрерывны [5], $w^\circ(u) = w(u) - uw(1)$, и что, следовательно, интеграл (16) есть по определению результат почти наверное сходимости при $N \to \infty$ сумм

$$\sum_{i=1}^{N} w(F_2(t_i))\,\Delta_i F_1 - m_1 w(1), \tag{17}$$

где $m_1 = \int F_2(t)\,dF_1(t)$, $\{t_i\}_{i=0}^{N}$ образуют разбиение вещественной оси, $\Delta_i g = g(t_i) - g(t_{i-1})$, $w(F_2(t_i)) = \sum_{l=1}^{i} \Delta_l w(F_2)$, $w(1) = \sum_{l=1}^{N} \Delta_l w(F_2)$.

В силу преобразования Абеля

$$\sum_{i=1}^{N}\left(\sum_{l=1}^{i} d_l\right) b_i = \sum_{i=1}^{N}\left(\sum_{l=i}^{N} b_l\right) d_i.$$

Поэтому (17) равно

$$\sum_{i=1}^{N} (1 - F_1(t_{i-1}) - m_1)\,\Delta_i w(F_2). \tag{18}$$

Здесь $1 - m_1 = \int F_1(t)\,dF_2(t) \equiv m_2$, а $\Delta_i w(F_2)$ есть независимые, нормально распределенные случайные величины с параметрами $(0, \Delta_i F_2)$. Поэтому распределение (17), (18) будет нормальным с нулевым средним и дисперсией

$$\sum_{i=1}^{N} (m_2 - F_1(t_{i-1}))^2\,\Delta_i F_2 \to \int (m_2 - F_1(t))^2\,dF_2(t) = \mathsf{D}F_1(y_1).$$

Соотношение (14) доказано.

Чтобы доказать (15) *), проще всего оценить дисперсию интеграла в (15). Приближая интеграл снова с помощью конечной суммы, можно убедиться, что дисперсия

$$D_{X,Y} = \mathsf{M}\left(\int w_Y(t)\,dw_X(t)\right)^2$$

ограничена при $n_1 \to \infty$, $n_2 \to \infty$. Отсюда и из неравенства Чебышева следует (15). Доказательство ограниченности $D_{X,Y}$ мы опускаем ввиду громоздкости и рутинного характера вычислений. ◁

По поводу критериев знаков и Вилкоксона см. также [10].

Пример 1. Мы отмечали, что критерии знаков и Вилкоксона наиболее чувствительны к сдвигам. Поэтому интересно сравнить их мощности с мощностью оптималь-

---

*) Интеграл в (15) сходится по распределению к $\int w^\circ(F_2(t)) \times$ $\times\, dw^\circ(F_1(t))$.

ного критерия в задаче, где однородность проверяется для семейства $\mathscr{P}$ распределений, отличающихся лишь сдвигами. Именно, пусть

$$\mathscr{P} = \{\Phi_{\alpha,1}\}, \quad \mathbf{P}_1 = \Phi_{\alpha_1,1}, \quad \mathbf{P}_2 = \Phi_{\alpha_2,1}, \quad n_1 = n_2 = n.$$

В этом случае, согласно теореме 1.1, для проверки гипотезы $H_1 = \{\mathbf{P}_1 = \mathbf{P}_2\} = \{\alpha_1 = \alpha_2\}$ против $H_2^b = \{|\alpha_1 - \alpha_2| \geqslant b/\sqrt{n}\}$ существует асимптотически минимаксный критерий π_0 уровня $1-\varepsilon$, имеющий вид

$$|\bar{\mathbf{x}} - \bar{\mathbf{y}}| > \lambda_{\varepsilon/2}\sqrt{2/n}, \quad \Phi_{0,1}((-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2})) = 1-\varepsilon$$

(то, что это неравенство в нашем примере эквивалентно (1.3), (1.4), читатель может проверить сам). Примем этот критерий за эталон для сравнения с другими критериями и рассмотрим альтернативу $(\mathbf{P}_1, \mathbf{P}_2)$, где $\alpha_2 = \alpha_1 + c/\sqrt{n}$ (мы рассматриваем близкие альтернативы, чтобы не иметь дела с задачей о больших уклонениях). Очевидно, что в этом случае $(\bar{\mathbf{x}} - \bar{\mathbf{y}}) \in \Phi_{-c/\sqrt{n},2/n}$. Стало быть,

$$\begin{aligned}\beta_{\pi_0}(\mathbf{P}_1, \mathbf{P}_2) &= \mathbf{P}_1\times\mathbf{P}_2\left(|\bar{\mathbf{x}} - \bar{\mathbf{y}}| > \lambda_{\varepsilon/2}\sqrt{2/n}\right) = \\ &= 1 - \Phi_{-c/\sqrt{2},1}(-\lambda_{\varepsilon/2}, \lambda_{\varepsilon/2}) = \\ &= 1 - \Phi_{0,1}\left(-\lambda_{\varepsilon/2} + c/\sqrt{2}, \lambda_{\varepsilon/2} + c/\sqrt{2}\right) \equiv \beta_0(c). \quad (19)\end{aligned}$$

Рассмотрим теперь критерий знаков (8) и обозначим его π_1. Пользуясь разложением в ряд по степеням $c/\sqrt{n}$, находим $\left(\Phi_{\alpha,\sigma^2}(x) = \Phi_{\alpha,\sigma^2}((-\infty, x))\right)$

$$\mathbf{P}_1\times\mathbf{P}_2(\mathbf{x}_1 - \mathbf{y}_1 < 0) = \Phi_{0,2}\left(\frac{c}{\sqrt{n}}\right) = \frac{1}{2} + \frac{c}{\sqrt{n}}\cdot\frac{1}{2\sqrt{\pi}} + O\left(\frac{1}{n}\right).$$

Поэтому в точке $(\mathbf{P}_1, \mathbf{P}_2)$

$$\frac{2}{\sqrt{n}}\left(\nu - \frac{n}{2} + \frac{c\sqrt{n}}{2\sqrt{\pi}}\right) \Leftarrow\Rightarrow \Phi_{0,1}.$$

Стало быть, для критерия знаков π_1 асимптотического уровня $1-\varepsilon$

$$\begin{aligned}\beta_{\pi_1}(\mathbf{P}_1, \mathbf{P}_2) &= \mathbf{P}_1\times\mathbf{P}_2\left(2\left|\nu - \frac{n}{2}\right| > \lambda_{\varepsilon/2}\sqrt{n}\right) \to \\ &\to 1 - \Phi_{0,1}\left(\left(-\lambda_{\varepsilon/2} + \frac{c}{\sqrt{\pi}}, \lambda_{\varepsilon/2} + \frac{c}{\sqrt{\pi}}\right)\right).\end{aligned}$$

Обратимся, наконец, ко критерию π_2 Вилкоксона (см. (11)), имеющему в нашем случае вид

$$\left|U-\frac{n^2}{2}\right|>\frac{\lambda_{\varepsilon/2}n^{3/2}}{\sqrt{6}}.$$

Очевидно, что статистика U инвариантна относительно преобразования сдвига элементов выборок X и Y. Поэтому можно считать, что $\mathrm{P}_1=\Phi_{0,1}$, $\mathrm{P}_2=\Phi_{c/\sqrt{n},1}$ и, стало быть,

$$\mathsf{M}F_2(x_1)=\int F_2(t)\,dF_1(t)=\int\Phi_{0,1}\left(t-\frac{c}{\sqrt{n}}\right)d\Phi_{0,1}(t)=$$
$$=\Phi_{0,2}\left(-\frac{c}{\sqrt{n}}\right)=\frac{1}{2}-\frac{c}{\sqrt{n}}\cdot\frac{1}{2\sqrt{\pi}}+O\left(\frac{1}{n}\right).$$

Так как $\mathsf{D}F_2(x_1)\to\mathsf{D}F_1(x_1)=1/12$, $\mathsf{D}F_1(y_1)\to\mathsf{D}F_1(x_1)=$ $=1/12$, то по теореме 2

$$\beta_{\pi_2}(\mathrm{P}_1,\mathrm{P}_2)=\mathrm{P}_1\times\mathrm{P}_2\left(\left|U-\frac{n^2}{2}\right|>\frac{\lambda_{\varepsilon/2}n^{3/2}}{\sqrt{6}}\right)=$$
$$=1-\mathrm{P}_1\times\mathrm{P}_2\left(-\lambda_{\varepsilon/2}+c\sqrt{\frac{3}{2\pi}}\leqslant\right.$$
$$\left.\leqslant\sqrt{6}\,n^{-3/2}\left(U-\frac{n^2}{2}+\frac{n^{3/2}c}{2\sqrt{\pi}}\right)\leqslant\lambda_{\varepsilon/2}+c\sqrt{\frac{3}{2\pi}}\right)=$$
$$=1-\Phi_{0,1}\left(-\lambda_{\varepsilon/2}+c\sqrt{\frac{3}{2\pi}},\ \lambda_{\varepsilon/2}+c\sqrt{\frac{3}{2\pi}}\right).$$

Заметим теперь, что $\beta_0(c)$ (см. (19)) есть монотонно возрастающая функция от c и что при больших n

$$\beta_{\pi_1}(\mathrm{P}_1,\mathrm{P}_2)\approx\beta_0\left(\sqrt{\frac{2}{\pi}}\,c\right),\quad\beta_{\pi_2}(\mathrm{P}_1,\mathrm{P}_2)\approx\beta_0\left(\sqrt{\frac{3}{\pi}}\,c\right).$$

Таким образом, при каждом $c>0$ наиболее мощным среди π_0, π_1, π_2 оказывается, как этого и следовало ожидать, критерий π_0. Затем идет критерий Вилкоксона и критерий знаков; при этом критерий Вилкоксона уступает критерию π_0 совсем немного, так как $\sqrt{3/\pi}\approx0{,}977$.

Если мы при том же смещении $\alpha_2-\alpha_1=c/\sqrt{n}$ рассмотрим выборки X' и Y' объема $n'>n$, то для того, чтобы с помощью проведенных вычислений получить мощность в точке $(\mathrm{P}_1,\ \mathrm{P}_2)$ критериев $\pi_j(X',\ Y')$, мы должны рассмотреть прежнюю задачу при новом значении c, равном

$c' = c\sqrt{n'}/\sqrt{n}$ (тогда $\alpha_2 - \alpha_1$ можно будет записать в виде $c'/\sqrt{n'}$). Стало быть, мощности $\pi_1(X', Y')$ и $\pi_2(X', Y')$ в той же точке $(\mathbf{P}_1, \mathbf{P}_2)$ будут равны приближенно

$$\beta_0\left(\sqrt{\frac{2}{\pi}}c'\right) = \beta_0\left(\sqrt{\frac{2n'}{\pi n}}c\right), \quad \beta_0\left(\sqrt{\frac{3}{\pi}}c'\right) = \beta_0\left(\sqrt{\frac{3n'}{\pi n}}c\right).$$

Приравнивая $\frac{2n'}{\pi n} = 1$, $\frac{3n'}{\pi n} = 1$, мы получим значения $n' = \frac{\pi}{2}n$, $n' = \frac{\pi}{3}n$ (они от c не зависят) для числа наблюдений, которые надо проделать, чтобы соответственно с помощью критериев π_1 и π_2 получить ту же мощность, что и для критерия π_0 с n наблюдениями. Например, для $n = 100$ наблюдений с критерием π_0 нам потребуется для получения тех же результатов $n' \approx 105$ наблюдений с критерием π_2 и $n' \approx 157$ наблюдений с критерием π_1.

Совсем другие результаты мы получили бы, если бы проверяли однородность для семейства $\mathscr{P} = \{\Phi_{0,\sigma^2}\}$. В этом случае критерии знаков и Вилкоксона оказались бы несостоятельными. Более того, критерий знаков уровня $1 - \varepsilon$ был бы по существу равнозначен критерию $\pi = \varepsilon$, не зависящему от выборок, так как $\mathbf{M}(x_1 - y_1) = 0$ и $\mathbf{P}_1 \times \mathbf{P}_2(x_1 - y_1 > 0) = 1/2$ для любой пары распределений $\mathbf{P}_1$ и $\mathbf{P}_2$ из $\mathscr{P}$. Для этой задачи можно было бы рассмотреть другие непараметрические критерии, использующие статистики r_i, например критерий $\sum_{i=0}^{n_1}(r_{i+1} - r_i)^2$, $r_0 = 0$, $r_{n_1+1} = n_2$, напоминающий по своим свойствам критерий Морана (§ 3.12[61]).

5. **Критерий χ^2 как асимптотически оптимальный критерий проверки однородности по сгруппированным данным.** В этом разделе мы будем предполагать, что данные в обеих выборках X и Y объемов соответственно n_1 и n_2 сгруппированы (см. § 3.16[61]). В этом случае вместо выборок X, Y можно использовать векторы $\nu = (\nu_1, \ldots, \nu_r)$ и $\mu = (\mu_1, \ldots, \mu_r)$ частот наблюдений соответственно из выборок X и Y, попавших в интервалы $\Delta_1, \ldots, \Delta_r$, определяющие группировку. Обозначим $\theta_i = (\theta_{i1}, \ldots, \theta_{ir})$, $i = 1, 2$, векторы вероятностей попадания наблюдений соответственно из первой и второй выборок в интервалы $\Delta_1, \ldots, \Delta_r$, так что $\theta_{1l} = \mathbf{P}(x_j \in \Delta_l)$, $\theta_{2l} = \mathbf{P}(y_j \in \Delta_l)$. Огрубленные выборки X и Y можно рассматривать тогда как выборки из параметрических семейств $\mathbf{B}_{\theta_1}$ и $\mathbf{B}_{\theta_2}$

соответственно. Таким образом, задача становится параметрической, и мы можем использовать результаты, изложенные в примере 1 предыдущего параграфа. Из этого примера следует, что если мы проверяем гипотезу об однородности $H_1 = \{\theta_1 = \theta_2\}$ в случае, когда параметр θ локализован, т. е. значения θ_1, θ_2 располагаются в окрестности точки $\theta_0 = (\theta_{01}, \ldots, \theta_{0r})$, то асимптотически минимаксный критерий асимптотического уровня $1-\varepsilon$ для проверки H_1 против

$$H_2^b = \left\{\sum_{i=1}^{r} \frac{(\theta_{1i} - \theta_{2i})^2}{\theta_{0i}} \geqslant \frac{b^2}{n_2}\right\}$$

имеет вид

$$\sum_{i=1}^{r} \left(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2}\right)^2 \frac{n_1 n_2}{\nu_i + \mu_i} \geqslant h_\varepsilon,$$

где h_ε есть квантиль порядка $1-\varepsilon$ распределения χ^2 с $r-1$ степенями свободы. Это и есть критерий χ^2 для проверки однородности по сгруппированным данным.

В качестве асимптотически эквивалентного можно рассматривать критерий

$$\sum_{i=1}^{r} \nu_i \ln \frac{\nu_i}{n_1} + \sum_{i=1}^{r} \mu_i \ln \frac{\mu_i}{n_2} - \sum_{i=1}^{r} (\nu_i + \mu_i) \ln \frac{\nu_i + \mu_i}{n_1 + n_2} > \frac{h_\varepsilon}{2}.$$

§ 3. Задачи регрессии

1. Постановка задачи. В приложениях часто возникают задачи, относящиеся к наблюдениям, распределение которых меняется в разных экспериментах при изменении некоторых параметров, характеризующих эти эксперименты. Набор значений указанных параметров в i-м эксперименте, $i = 1, \ldots, n$, мы обозначим

$$x_i = (x_{i,1}, \ldots, x_{i,r})$$

(так что r есть размерность векторов x_i). Значения $x_{i,k}$ определяются либо экспериментатором, либо природой изучаемого явления. Вектор $(x_{1,k}, \ldots, x_{n,k})$ мы обозначим буквой X_k, а матрицу $\begin{pmatrix} X_1 \\ \cdots \\ X_r \end{pmatrix} = (x_1^T, \ldots, x_n^T)$ — буквой X. Таким образом, здесь, в отличие от предыдущего,

X является матрицей порядка $r \times n$ и может быть произвольным неслучайным набором чисел, природа которых нас интересовать не будет. Вектор наблюдений мы обозначим через $Y = (y_1, \ldots, y_n)$.

Задачи регрессии связаны с предположением, что наблюдения y_i в зависимости от набора параметров $x_i = (x_{i,1}, \ldots, x_{i,r})$ имеют вид

$$y_i = \alpha_1 x_{i,1} + \ldots + \alpha_r x_{i,r} + \xi_i, \quad i = 1, \ldots, n, \tag{1}$$

где $\alpha = (\alpha_1, \ldots, \alpha_r)$ — неизвестные нам постоянные, $\xi_i \in \Phi_{0,\sigma^2}$ и независимы.

Постоянная α_1 часто играет особую роль, так как в ряде случаев она выделяет в представлении (1) постоянное слагаемое, что соответствует тому, что в матрице X заранее полагается $X_1 = (1, \ldots, 1)$ $(x_{i,1} = 1)$. Мы этим предположением пользоваться не будем. Случайные величины ξ_i обуславливаются шумами, флуктуациями или ошибками измерений.

В матричной форме соотношения (1) можно записать в виде

$$Y = \alpha X + \xi. \tag{2}$$

Регрессия вида (1), (2) называется линейной (как по α, так и по X). В качестве задач регрессии могут рассматриваться как задача оценки неизвестных параметров α, σ^2, если известно, что справедливо (1), (2), так и задача проверки самой гипотезы о том, что представление (1), (2) имеет место. Исходными данными в обоих случаях служит «выборка» (X, Y). Термин «выборка» мы здесь используем уже в более широком, чем прежде, смысле, обозначая им набор результатов наблюдений, не обязательно имеющих одинаковую природу. Кроме того, напомним, что первая из двух «выборок» X и Y может быть неслучайной. Матрицу X называют иногда *регрессором*, а вектор Y — *откликом*.

Модель регрессии (1), (2) является весьма общей, если иметь в виду зависимость y_i от набора параметров. Полагая, например, $x_{i,k} = \psi_k(z_i)$, где $\psi_1, \ldots, \psi_r$ — заданный набор функций, а z_i — значения одномерного параметра, мы получим модель

$$y_i = \alpha_1 \psi_1(z_i) + \ldots + \alpha_r \psi_r(z_i) + \xi_i, \quad i = 1, \ldots, n, \tag{3}$$

регрессии по произвольным функциям $\psi_1, \ldots, \psi_r$ (и по-прежнему линейной по α). Если $\psi_1(z) \equiv 1$, $\psi_2(z) \equiv z$,

$r=2$, то мы получим модель *простейшей* (одномерной) *линейной* регрессии (рис. 1).

В отличие от простейшей, общую модель (1), (2) иногда называют *множественной* регрессией. В целом задачи регрессии, как мы видим, связаны с изучением (с наличием) функциональной зависимости $y=\varphi(x)$ для заданного класса функций φ в тех случаях, когда наблюдения над переменной y при данном x «зашумлены» случайными отклонениями.

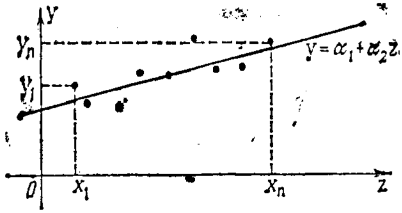

Рис. 1

Строки $X_1, \ldots, X_r$ матрицы X в (2) обычно выбирают таким образом, чтобы они были *линейно независимы* (иначе координаты α нам не удастся оценить). Мы тоже будем следовать этому соглашению, означающему, что ранг матрицы X равен r.

Иногда бывает удобнее иметь дело с ортогональными векторами $X_1, \ldots, X_r$, т. е. удовлетворяющими условию $(X_i, X_j)=0$, $i\neq j$, где (a, b) означает скалярное произведение. Если исходный набор линейно независимых векторов $\{X_k\}$ таким свойством не обладает, то его можно ортогонализировать, введя новые векторы

$$\begin{aligned} X_1' &= X_1, \\ X_2' &= X_2 + a_{2,1}X_1, \\ &\cdots\cdots\cdots \\ X_r' &= X_r + a_{r,r-1}X_{r-1} + \ldots + a_{r,1}X_1. \end{aligned} \tag{4}$$

Коэффициенты $a_{k,j}$ легко определяются из условий ортогональности $X_k' \perp X_j'$, $k\neq j$, так что, например, $a_{2,1} = -\dfrac{(X_1, X_2)}{(X_1, X_1)}$. Соотношения (4) можно записать в виде $X'=AX$, где A — треугольная обратимая матрица (с единицами по главной диагонали). Отсюда получаем $X=A^{-1}X'$, $Y=\alpha A^{-1}X'+\xi$. Мы пришли к задаче регрессии с коэффициентами $\beta=\alpha A^{-1}$. Вектор α очевидным образом восстанавливается по β с помощью равенства $\alpha=\beta A$.

Для простейшей линейной регрессии предположение об ортогональности $X_1=(1, \ldots, 1)$ и $X_2=(z_1, \ldots, z_n)$

означает предположение $\sum z_i = 0$, которое очевидным образом может быть удовлетворено путем изменения начала отсчета переменной z.

2. **Оценка параметров.** Везде в дальнейшем мы будем предполагать, что $r<n$ и что векторы X_k, $k=1, \ldots, r$, линейно независимы. Функция правдоподобия наблюдения Y при данном X для регрессии (1), (2) равна

$$f_{\alpha,\sigma^2}(Y) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \sum_{k=1}^{r}\alpha_k x_{i,k}\right)^2\right\} =$$
$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{|Y-\alpha X|^2}{2\sigma^2}\right\}. \quad (5)$$

Функция (5) зависит от параметра $\theta=(\alpha, \sigma^2)$. Отметим, что если (5) рассматривать как функцию правдоподобия не одного наблюдения Y (или (X, Y)), а n наблюдений $y_1, \ldots, y_n$, то она не будет соответствовать выборке из какого-нибудь одного параметрического семейства. Наблюдения y_i относятся к разным распределениям Φ_{γ_i,σ^2}, $\gamma_i = \sum_{k=1}^{r}\alpha_k x_{i,k}$, зависящим от x_i. Поэтому те рассмотрения из [6], в которых использовалась одинаковая распределенность элементов выборки, здесь непосредственно неприменимы.

Итак, мы будем рассматривать (5) как функцию правдоподобия наблюдения (X, Y). Воспользуемся методом максимального правдоподобия. Непосредственно из (5) видно, что оценка максимального правдоподобия $\alpha^*=\widehat{\alpha}^*$, максимизирующая $f_\theta(Y)$ по α, есть оценка, минимизирующая $|Y-\alpha X|^2$. Поэтому в нашем случае метод максимального правдоподобия совпадает с «методом наименьших квадратов».

Обозначим $\mathscr{L}[X]$ подпространство, натянутое на векторы $X_1, \ldots, X_r$. Это есть совокупность точек вида αX, когда α пробегает значения из R^r. Это пространство имеет размерность r и содержит единственную точку $\beta=\alpha^* X$, наименее удаленную от Y (рис. 2). Значение β однозначно определяется условием ортогональности $Y-\beta$ и $\mathscr{L}[X]$, или, что то же, r условиями

$$(Y-\alpha^* X, X_k) = (Y-\alpha^* X)X_k^T = 0, \quad k=1, \ldots, r.$$

В матричной форме эти условия можно записать в виде

$(Y-\alpha^*X)X^T=0$. Отсюда находим

$$\alpha^*=YX^T(XX^T)^{-1}. \tag{6}$$

Здесь обратная матрица $(XX^T)^{-1}$ (порядка $r\times r$) существует, так как матрица $D=XX^T$ является положительно определенной. Действительно, мы видели, что существует невырожденная матрица A такая, что строки

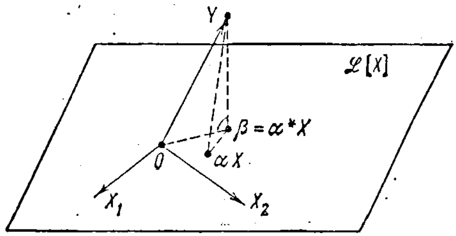

Рис. 2

матрицы $X'=AX$ ортогональны. Стало быть, матрицу D можно записать в виде

$$XX^T=A^{-1}X'(X')^T(A^{-1})^T=A^{-1}B(A^{-1})^T,$$

где $B=X'(X')^T$ — диагональная матрица с элементами

$$(X_i', X_j')=\begin{cases}|X_i'|^2>0 & \text{при} \quad i=j,\\ 0 & \text{при} \quad i\neq j.\end{cases}$$

Следовательно, B положительно определена, $aBa^T>0$ для любого $a\in R^r$, $a\neq 0$. Полагая $b=aA$, получим $bDb^T=aAXX^TA^Ta^T=aBa^T>0$ при любом $b\in R^r$, $b\neq 0$, что и требовалось доказать.

Если X_k ортогональны, то из (6) находим $\alpha_k^*=\dfrac{(Y, X_k)}{(X_k, X_k)}$.

Результат (6) мы могли бы получить и другим путем — дифференцируя (5) по α_k и приравнивая производные нулю.

Разность $Y-\alpha^*X$ называют иногда *остатком*. Эта разность ортогональна $\mathscr{L}[X]$, а вместе с этим и любому вектору $\gamma X\in\mathscr{L}[X]$, $\gamma\in R^r$. Если взять $\gamma=\alpha^*-\alpha$, то из равенства $Y-\alpha X=Y-\alpha^*X+(\alpha^*-\alpha)X$ будет следовать

$$|Y-\alpha X|^2=|Y-\alpha^*X|^2+|(\alpha^*-\alpha)X|^2. \tag{7}$$

Найдем теперь о. м. п. для σ^2. Из (5) видно, что это будет та же оценка, что и для нормального семейства (можно заново продифференцировать (5) по σ и приравнять производную нулю), так что

$$(\widehat{\sigma^2})^* = \frac{1}{n} |Y - \alpha^* X|^2. \tag{8}$$

Положим

$$(\sigma^2)^* = \frac{1}{n-r} |Y - \alpha^* X|^2 = \frac{n}{n-r} (\widehat{\sigma^2})^*. \tag{9}$$

E_l будет означать в дальнейшем единичную матрицу порядка l, $\sigma^* = \sqrt{(\sigma^2)^*}$.

Теорема 1. *Оценки* (6) *и* (9) *являются независимыми несмещенными эффективными оценками параметров α и σ^2. Кроме того,*

$$(\alpha^* - \alpha) D^{1/2} \in \Phi_{0,\sigma^2 E_r}, \quad D = XX^T, \tag{10}$$

$$(n-r)(\sigma^2)^*/\sigma^2 = |Y - \alpha^* X|^2/\sigma^2 \in \mathrm{H}_{n-r}. \tag{11}$$

Если X_k ортогональны, то α_k^* независимы,

$$(\alpha_k^* - \alpha_k) |X_k| \in \Phi_{0,\sigma^2}. \tag{12}$$

Следствие 1. *Из* (10), (11) *вытекает, что*

$$\frac{(\alpha^* - \alpha) D (\alpha^* - \alpha)^T}{(n-r)(\sigma^2)^*} = \frac{|(\alpha^* - \alpha) X|^2}{|Y - \alpha^* X|^2} \in \mathrm{F}_{r,\, n-r}. \tag{13}$$

Пусть $\bar{\alpha}$, $\bar{\alpha}^*$ — «подвекторы» размерности $l \leqslant r$ векторов α и α^*, составленные из координат с фиксированными номерами $k_1, \ldots, k_l$, и пусть $\bar{X}$ — матрица, составленная из строк $X_{k_1}, \ldots, X_{k_l}$. Тогда, если X_k, $k = 1, \ldots, r$, ортогональны, то

$$(\bar{\alpha}^* - \bar{\alpha})(\bar{X}\bar{X}^T)^{1/2} \in \Phi_{0,\sigma^2 E_l}, \quad (\alpha_k^* - \alpha_k) |X_k| / \sigma^* \in \mathrm{T}_{n-r}. \tag{14}$$

Доказательство теоремы 1. Так как $YX^T = \alpha XX^T + \xi X^T$, то

$$\alpha = (YX^T - \xi X^T) D^{-1}, \quad \alpha^* - \alpha = \xi X^T D^{-1}. \tag{15}$$

Матрица вторых моментов вектора $(\alpha^* - \alpha) D^{1/2}$ равна

$$\mathsf{M} D^{1/2} (\alpha^* - \alpha)^T (\alpha^* - \alpha) D^{1/2} =$$

$$= D^{1/2} D^{-1} X \mathsf{M} \xi^T \xi X^T D^{-1} D^{1/2} = \sigma^2 E_r.$$

Так как компоненты этого вектора нормальны, то они независимы и $\frac{1}{\sigma^2}|(\alpha^*-\alpha)D^{1/2}|^2 \in H_r$. Далее, в силу (7), (9)

$$(n-r)(\sigma^2)^* = |Y-\alpha^*X|^2 = |\xi|^2 - |(\alpha^*-\alpha)X|^2.$$

Убедимся теперь, что векторы α^* и $Y-\alpha^*X$ (а стало быть, α^* и σ^*) независимы. В силу их нормальности достаточно проверить, что коэффициенты корреляции между их компонентами равны нулю или, что то же, что матрица центральных вторых моментов $\mathsf{M}(\alpha^*-\alpha)^T(Y-\alpha^*X)$ равна нулю. Заметим, что в силу (6)

$$\alpha^*X = YX^T(XX^T)^{-1}X = YX^TD^{-1}X$$

и вектор α^*X получается из Y путем проектирования Y на $\mathscr{L}[X]$. Оператор проектирования, определяемый матрицей $\Pi = X^TD^{-1}X$, обладает очевидными свойствами: $\Pi^2 = \Pi$, $BX\Pi = BX$ для любой матрицы B, имеющей r столбцов. Поэтому в силу (15)

$$\mathsf{M}(\alpha^*-\alpha)^T(Y-\alpha^*X) = \mathsf{M}D^{-1}X\xi^T(\xi-\xi X^TD^{-1}X) =$$
$$= D^{-1}X\sigma^2(E_n - \Pi) = 0.$$

Докажем теперь (11). В силу (7)

$$|Y-\alpha^*X|^2 = |\xi|^2 - |(\alpha^*-\alpha)X|^2 = |\xi|^2 - |(\alpha^*-\alpha)D^{1/2}|^2,$$

где $\frac{1}{\sigma^2}|\xi|^2 \in H_n$, $\frac{1}{\sigma^2}|(\alpha^*-\alpha)D^{1/2}|^2 \in H_r$ (см. (10)). Утверждение (11) будет следствием этих соотношений и леммы 1.

Лемма 1. *Если* $\eta = \eta_1 + \eta_2$, *где* η_1 *и* η_2 *независимы*, $\eta \in H_n$, $\eta_1 \in H_r$, *то* $\eta_2 \in H_{n-r}$.

Доказательство. Если обозначить $\varphi(t)$ характеристическую функцию распределения H_1: $\varphi(t) = (1+2it)^{-1/2}$, то

$$\mathsf{M}e^{i\eta t} = \varphi(t)^n = \varphi(t)^r \cdot \mathsf{M}e^{i\eta_2 t}.$$

Так как $\varphi(t) \neq 0$ на вещественной оси, то $\mathsf{M}e^{i\eta_2 t} = \varphi(t)^{n-r}$. Лемма доказана.

Несмещенность оценок α^*, $(\sigma^2)^*$ с очевидностью следует из (10), (11) ($\mathsf{M}\eta = l$, если $\eta \in H_l$).

Нам осталось доказать эффективность оценки $\theta^* = (\alpha^*, (\sigma^2)^*)$. Для этого заметим, что параметрическое семейство (5) относится к экспоненциальному типу, так как (5) представимо в форме (см. (2.15.1)[6])

$$f_\theta(Y) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left(|Y|^2 - 2(Y, \alpha X) + |\alpha X|^2\right)\right\} = \\ = h(Y)\exp\left\{\sum_{k=1}^{r+1} a_k(\theta)\, U_k(Y) + V(\theta)\right\},$$

где

$$h(Y) = (2\pi)^{-n/2}, \quad V(\theta) = -n\ln\sigma - \frac{1}{2\sigma^2}|\alpha X|^2,$$

$$a_k(\theta) = \frac{\alpha_k}{\sigma^2}, \quad U_k(Y) = (Y, X_k), \quad k = 1, \ldots, r,$$

$$a_{r+1}(\theta) = -\frac{1}{2\sigma^2}, \quad U_{r+1}(Y) = |Y|^2.$$

Так как условия теорем 2.15.1[6], 2.15.2[6] здесь выполнены, то статистика $U = (U_1(X), \ldots, U_{r+1}(X))$ (а вместе с ней и θ^*) является минимальной полной достаточной статистикой. Отсюда следует (см. следствие 2.15.1[6]) эффективность θ^*.

Утверждение (12) с очевидностью следует из (10), так как для ортогональных X_k матрица $D^{1/2}$ диагональна с элементами $|X_k|$ по диагонали. Теорема доказана.

З а м е ч а н и е 1. Хотеллинг (см. [29]) доказал, что $\mathsf{D}\alpha_k^* \geqslant \sigma^2/|X_k|^2$ и равенство достигается только в том случае, когда X_k ортогональны. Таким образом, при планировании эксперимента при заданных значениях $|X_k|$ оптимальный выбор регрессора X состоит в том, чтобы сделать X_k ортогональными.

З а м е ч а н и е 2. Интересно сравнить матрицу вторых моментов оценки θ^* с нижней границей для несмещенных оценок, определяемой в силу многомерного неравенства Рао — Крамера матрицей $I^{-1}(\theta)$, где $I(\theta)$ — информационная матрица Фишера:

$$I(\theta) = \|I_{ij}(\theta)\|, \; I_{ij}(\theta) = \mathsf{M}_\theta \frac{\partial L}{\partial \theta_i}\cdot\frac{\partial L}{\partial \theta_j}, \; L = L(Y;\theta) = \ln f_\theta(Y).$$

Здесь мы приняли $\theta_k = \alpha_k$, $k = 1, \ldots, r$, $\theta_{r+1} = \sigma^2$. Пусть для простоты X_k ортогональны. Из независимости θ_k^* следует, что матрица $\mathsf{M}_\theta(\theta^* - \theta)^T(\theta^* - \theta)$ будет диагональной с диагональными элементами

$$\mathsf{M}_\theta(\alpha_k^* - \alpha)^2 = \frac{\sigma^2}{|X_k|^2}, \qquad k = 1, \ldots, r,$$

$$\mathsf{M}_\theta((\sigma^2)^* - \sigma^2)^2 = \mathsf{M}\left(\frac{\sigma^2\chi^2_{n-r}}{n-r} - \sigma^2\right)^2 = \frac{\sigma^4}{n-r}\mathsf{M}(\chi_1^2 - 1)^2 = \frac{2\sigma^4}{n-r},$$

где $\chi_l^2 \in \mathrm{H}_l$.

С другой стороны, для матрицы $I(\theta)$ в силу того, что

$$\frac{\partial L}{\partial \alpha_k} = \frac{1}{\sigma^2}\sum_{i=1}^n \left(y_i - \sum_{j=1}^r \alpha_j x_{ij}\right) x_{ik} = \frac{1}{\sigma^2}(Y - \alpha X) X_k^T,$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n \left(y_i - \sum_{j=1}^r \alpha_j x_{ij}\right)^2 =$$

$$= \frac{1}{2\sigma^2}\left(\frac{|Y - \alpha X|^2}{\sigma^2} - n\right),$$

находим при $k = 1, \ldots, r$:

$$I_{kk}(\theta) = \mathsf{M}_\theta \frac{1}{\sigma^4} X_k (Y - \alpha X)^T (Y - \alpha X) X_k^T =$$

$$= \frac{1}{\sigma^4}\mathsf{M} X_k \xi^T \xi X_k^T = \frac{1}{\sigma^4}\mathsf{M}(\xi, X_k)^2 = \frac{|X_k|^2 \mathsf{M}|\xi|^2}{\sigma^4} = \frac{|X_k|^2}{\sigma^2},$$

$$I_{r+1,\, r+1}(\theta) = \frac{1}{4\sigma^4}\mathsf{M}\left[\sum_{i=1}^n \left(\frac{\xi_i^2}{\sigma^2} - 1\right)\right]^2 = \frac{n}{2\sigma^4}, \quad I_{ij}(\theta) = 0$$

при $i \neq j$. Так что

$$I^{-1}(\theta) = \left(\begin{array}{ccc|c} \frac{\sigma^2}{|X_1|^2} & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & \frac{\sigma^2}{|X_r|^2} & 0 \\ \hline 0 & \ldots & 0 & \frac{2\sigma^4}{n} \end{array}\right).$$

Таким образом, в неравенстве Рао — Крамера

$$\mathsf{M}_\theta(\theta^* - \theta)^T(\theta^* - \theta) \geqslant I^{-1}(\theta) \tag{16}$$

для первых r компонент θ^* достигается равенство. Для $(r+1)$-й оно достигаться не может (хотя асимптотически при $n\to\infty$ обе части (16) ведут себя одинаково), так как необходимое и достаточное условие теоремы 2.16.1А[6] здесь не выполнено.

Замечание 3. Предположение о нормальности ξ_i становится малосущественным для утверждений (10)—(12), если n велико (в (11) лучше произвести нормировку и утверждать близость к нормальному закону).

Замечание 4. Сам термин «регрессия» относится к совместному распределению двух случайных величин ξ и η и означает кривую

$$g(x) = \mathsf{M}(\eta/\xi = x),$$

которую называют также регрессией η на ξ. Например, если $(\xi, \eta) \in \Phi_{\gamma,\sigma^2}$, $\gamma = (\gamma_1, \gamma_2)$, $\sigma^2 = \|\sigma_{ij}\|$, $i, j = 1, 2$, то, как мы видели в [6], $g(x) = \gamma_2 + \frac{\sigma_{12}}{\sigma_{22}}(x-\gamma_1)$. Это есть простейшая линейная регрессия.

Замечание 5. Предположение $\xi_i \in \Phi_{0,\sigma^2}$ об одинаковой распределенности ξ_i при известном σ^2 может быть ослаблено. Мы можем считать, что $\xi_i \in \Phi_{0,\sigma_i^2}$, если σ_i различны и известны. В этом случае, обозначив через σ диагональную матрицу $\sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_n \end{pmatrix}$ и введя новые переменные $\xi' = \xi\sigma^{-1}$, $X' = X\sigma^{-1}$, $Y' = Y\sigma^{-1}$ (так что $\xi_i' = \xi_i/\sigma_i$, $x_i' = x_i/\sigma_i$, $y_i' = y_i/\sigma_i$), мы придем к задаче регрессии

$$Y' = \alpha X' + \xi',$$

в которой вектор наблюдений Y' и регрессор X' нам известны, $\xi' \in \Phi_{0,E_n}$. Легко проверить (мы предоставляем это читателю), что справедлив следующий аналог теоремы 1.

Теорема 2. *Оценка*

$$\alpha^* = Y\sigma^{-2}X^T(D')^{-1}, \quad D' = X\sigma^{-2}X^T,$$

является несмещенной эффективной оценкой α,

$$(\alpha^* - \alpha)(D')^{1/2} \in \Phi_{0,E_r},$$

$$|Y' - \alpha^* X'|^2 = \sum_{i=1}^{n} \frac{\left(y_i - \sum_{k=1}^{r} \alpha_k^* x_{ik}\right)^2}{\sigma_i^2} \in H_{n-r}.$$

Обратимся снова к теореме 1. Соотношения (10)—(12), установленные в ней, дают возможность строить доверительные множества как для отдельных координат θ, так и для вектора θ в целом. Например,

$$P_\theta\left(\frac{(n-r)(\sigma^2)^*}{h_{\varepsilon/2}^{(2)}} < \sigma^2 < \frac{(n-r)(\sigma^2)^*}{h_{\varepsilon/2}^{(1)}}\right) = 1 - \varepsilon, \tag{17}$$

и если X_k ортогональны, то

$$P_\theta\left(|\alpha_k - \alpha_k^*| < \frac{t_{\varepsilon/2}\sigma^*}{|X_k|}\right) = 1 - \varepsilon, \tag{18}$$

где $T_{n-r}((-t_{\varepsilon/2}, t_{\varepsilon/2})) = 1 - \varepsilon$, $H_{n-r}\left(\left(h_{\varepsilon/2}^{(1)}, h_{\varepsilon/2}^{(2)}\right)\right) = 1 - \varepsilon$.

Пусть X_k ортогональны. Обозначим через $\overline{\alpha}$ «подвектор» вектора α, определенный в следствии 1. В силу теоремы 1 доверительное множество для $\overline{\alpha}$ естественно строить с помощью соотношения

$$\frac{|(\overline{\alpha} - \overline{\alpha}^*)\,\overline{X}\,|^2}{(n-r)(\sigma^2)^*} < f_\varepsilon. \tag{19}$$

Значение f_ε, соответствующее заданному уровню $1-\varepsilon$, определяется известным образом (см. гл. 3[6]) с помощью распределения Фишера $F_{l,\,n-r}$ с числом степеней свободы l, $n-r$.

Если σ^2 известно, то доверительный интервал будет определяться соотношением

$$|(\overline{\alpha} - \overline{\alpha}^*)\overline{X}|^2 < \sigma^2 h_\varepsilon, \tag{20}$$

где h_ε соответствует распределению H_l.

В задачах регрессии может понадобиться также оценить значение поверхности регрессии $y = \alpha z^T$ в новой, наперед заданной точке $z = (z_1, \ldots, z_r) \in R^r$. Положим $y^* = \alpha^* z^T$. Тогда, как и раньше, находим

$$y^* - y = (\alpha^* - \alpha) z^T = \xi X^T D^{-1} z^T \in \Phi_{0,d^2},$$

$$d^2 = \sigma^2 z D^{-1} z^T, \quad \frac{y^* - y}{d\sigma^*} \in T_{n-r}.$$

Это дает возможность строить доверительные интервалы для у.

Следует отметить, что отыскание доверительной области для поверхности регрессии «в целом» есть задача более трудоемкая (ср. с [29]). Совокупность поверхностей, входящих в доверительное множество, будет определяться доверительным множеством для θ, построенным, например, с помощью (10), (11) (см. § 3.8[6]). Подробнее об этом см. [29].

3. Проверка гипотез относительно линейной регрессии. Мы коснемся здесь двух типов задач.

1) Пусть известно, что представление (1), (2) имеет место. Требуется проверить гипотезу о том, что θ равно заданному значению θ' или что набор координат $\theta_{k_1}, \ldots, \theta_{k_l}$ равен набору $\theta'_{k_1}, \ldots, \theta'_{k_l}$, в то время как остальные координаты неизвестны.

Критерии для проверки таких гипотез удобно строить с помощью доверительных множеств (17)—(20) (см. § 3.8[6]). Пусть, например, требуется проверить гипотезу H_1 о независимости Y от X для простейшей линейной регрессии, т. е. гипотезу $H_1 = \{\alpha_2 = 0\}$. Тогда из (18) (или из (14)) мы получим критерий уровня $1-\varepsilon$, отвергающий H_1, если

$$|\alpha_2^*| \geqslant t_{\varepsilon/2}\sigma^*/|X_2|. \tag{21}$$

В общем случае регрессии (1) с ортогональными X_k гипотеза о независимости Y от X будет иметь вид $H_1 = \{\bar{\alpha} = 0\}$, где $\bar{\alpha} = (\alpha_2, \ldots, \alpha_r)$, $x_{i1} \equiv 1$, и для ее проверки можно воспользоваться критерием

$$\frac{|\bar{\alpha}^*\bar{X}|^2}{(n-r)(\sigma^2)^*} \geqslant f_\varepsilon, \tag{22}$$

где $\bar{X}$ и f_ε определены в (19) при $l = r-1$.

Можно использовать также подходы § 3.15[6], где рассматривалась проверка принадлежности выборки параметрическому подсемейству. Мы придем тогда к критерию отношения правдоподобия, который будет в известном смысле близок к (22). Если σ^2 известно, то к. о. п. для проверки $H_1 = \{\bar{\alpha} = 0\}$ будет иметь вид

$$\sigma^{-2}|\bar{\alpha}^*\bar{X}|^2 > h_\varepsilon,$$

где h_ε есть квантиль H_{r-1} порядка $1-\varepsilon$. Этот критерий будет минимаксным (см. §§ 3.9[6], 3.10[6]) для соответствующим образом отделенных альтернатив.

2) Проверка гипотезы о наличии в выборке (X, Y) самой регрессии (1), (2). Под этими словами мы понимаем гипотезу о том, что при каких-нибудь α, σ имеет место представление (1), (2), т. е. что при каких-нибудь α, σ справедливо $\sigma^{-1}(Y-\alpha X) \in \Phi_{0,E_n}$. Это есть задача о принадлежности Y параметрическому семейству. Но, как уже отмечалось, наблюдения в Y неодинаково распределены. Чтобы свести задачу к случаю одинаково распределенных наблюдений (см. § 3.17[6]), мы воспользуемся следующим утверждением, дополняющим теорему 1. Мы будем считать X_k ортогональными.

Теорема 3. *Пусть C — любая ортогональная матрица порядка $n\times n$, содержащая в качестве первых r столбцов столбцы матрицы $X^T D^{-1/2}$. Тогда вектор $\delta = (Y-\alpha^* X)C$ имеет независимые координаты, обладающие свойством $\delta_1 = \ldots = \delta_r = 0$, $\delta_i \in \Phi_{0,\sigma^2}$, $i=r+1, \ldots, n$.*

Таким образом, задача сводится к проверке гипотезы о принадлежности выборки $\delta_{r+1}, \ldots, \delta_n$ объема $n-r$ семейству Φ_{0,σ^2} (r наблюдений ушло, грубо говоря, на оценки для α). Эта задача рассматривалась в § 3.17[6]. Чтобы получить значения δ_i, надо по выборкам X, Y вычислить последовательно значения α^*, $Y-\alpha^* X$ и применить к $Y-\alpha^* X$ любое преобразование C, обладающее отмеченными в теореме 3 свойствами.

Если σ известно, то мы придем к задаче о проверке простой гипотезы о принадлежности Φ_{0,σ^2}. Однако в этом случае для проверки интересующей нас гипотезы можно использовать и теорему 1, в силу которой

$$(n-r)(\sigma^2)^*/\sigma^2 \in \mathrm{H}_{n-r}.$$

Доказательство теоремы 3. Если $Z \perp \mathscr{L}[X]$, то первые r координат вектора ZC образуют вектор $ZX^T D^{-1/2} = 0$. Так как $(Y-\alpha^* X) \perp \mathscr{L}[X]$ и $\delta = (Y-\alpha^* X)C$, то отсюда следует, что $\delta_1 = \ldots = \delta_r = 0$. Далее,

$$\delta = (Y-\alpha X)C - (\alpha^* - \alpha)XC = \eta - \bar{\eta} D^{-1/2} XC,$$

где $\eta = \xi C$, $\bar{\eta} = (\eta_1, \ldots, \eta_r) = (\alpha^* - \alpha) D^{1/2} = \xi X^T D^{-1/2}$ и, стало быть, δ есть результат линейного преобразования над η,

$$|\delta|^2 = |Y - \alpha^* X|^2 = |\xi|^2 - |(\alpha - \alpha^*) X|^2 =$$
$$= \sum_{i=1}^{n} \eta_i^2 - |\bar{\eta}|^2 = \sum_{i=r+1}^{n} \eta_i^2,$$

так что $\sum\limits_{i=r+1}^{n} \delta_i^2 = \sum\limits_{i=r+1}^{n} \eta_i^2$. Это возможно только в том случае, когда $(\delta_{r+1}, \ldots, \delta_n)$ есть результат поворота вектора $(\eta_{r+1}, \ldots, \eta_n)$, или, что то же, результат ортогонального преобразования над $(\eta_{r+1}, \ldots, \eta_n)$. Так как $\sigma^{-1}\eta \in \Phi_{0,E_n}$, то теорема доказана.

Пример 1. В этом примере мы опишем математическую сторону физического эксперимента, с помощью которого был обнаружен эффект распада φ-мезона на два π-мезона (см. [30]). Полученный результат носит статистический характер, и в нем по существу использовалась модель регрессии.

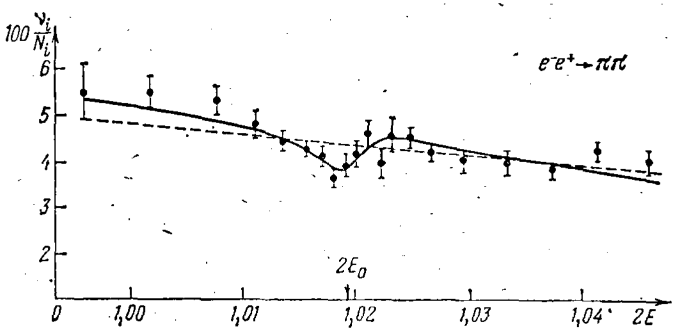

Рис. 3. Кривые изображают оценки линий регрессии при гипотезах H_1 и H_2

Исследование относится к изучению взаимодействия электронов (e^-) и позитронов (e^+) на встречных пучках. Если суммарная энергия этих частиц $2E$ находится в окрестности точки $2E_0 = 1019{,}6$ МэВ (рис. 3), то при их «столкновении» в результате взаимодействия образуются (наряду с другими) частицы двух типов: φ-мезоны и пары π-мезонов. Вероятность возникновения пары π-мезонов при взаимодействии e^+, e^- в зависимости от энергии E описывается с большой точностью линейной функцией, которую мы представим в виде (гипотеза H_1)

$$p_1^{\pi\pi}(E) = a_0 + a_1 x, \quad x = E - E_0, \tag{23}$$

где a_0, a_1 неизвестны.

Было выдвинуто предположение (гипотеза H_2) о том, что при распаде родившихся φ-мезонов также могут появляться пары π-мезонов. Обнаружить непосредственно этот эффект практически невозможно, так как было установлено, что это явление, если оно происходит, то происходит очень редко, не чаще чем один раз на 10^4 распадов φ-мезонов. Однако, благодаря эффекту интерференции этого дополнительного канала рождения π-мезонов с основным каналом, вероятность возникновения этих частиц будет равна не (23), а

$$p_2^{\pi\pi}(E) = [a_0 + a_1 x]\left[1 + \frac{b_0 + b_1 x + b_2 x^2}{x^2 + d^2}\right] \tag{24}$$

(как и (23), это есть весьма точное приближение несколько более сложной формулы, основанное на том, что рассматриваемый диапазон изменения $x = E - E_0$ мал по сравнению с E_0). В этом равенстве коэффициенты b_i, как и a_i, неизвестны, d известно.

Чтобы установить, какое из двух соотношений (23) или (24) на самом деле имеет место, было проведено $n = 20$ экспериментов при различных значениях энергии $E_1, \ldots, E_{20}$.

Таблица 1. Таблица экспериментальных данных

Номер эксперимента i	E_i, МэВ	N_i	ν_i	Номер эксперимента i	E_i, МэВ	N_i	ν_i
1	497,75	6 960	384	11	510,40	14 322	716
2	500,65	7 908	435	12	510,92	13 470	568
3	503,65	8 102	432	13	511,39	12 008	569
4	505,40	22 259	1080	14	512,17	23 951	1117
5	506,62	16 938	765	15	513,20	27 796	1185
6	507,66	21 728	951	16	514,62	37 771	1539
7	508,40	14 014	603	17	516,58	25 902	1036
8	508,90	13 793	545	18	518,64	27 857	1057
9	509,40	14 075	615	19	520,61	23 228	989
10	509,90	14 867	691	20	522,88	26 482	1066

Результатами экспериментов (см. табл. 1, рис. 3) являются количества N_i, $i = 1, \ldots, 20$, взаимодействий e^+, e^- и количества ν_i родившихся пар π-мезонов при энергии E_i. В каждом из проведенных экспериментов числа N_i и ν_i достаточно велики (N_i имеют порядок 10^4). Так как при фиксированном N_i число пар π-мезонов ν_i имеет распределение Бернулли $\mathbf{B}_{p_i}^{N_i}$ ($p_i = p_1^{\pi\pi}(E_i)$ при гипотезе H_1 и $p_i = p_2^{\pi\pi}(E_i)$ при гипотезе H_2), то, пользуясь нормальным приближением, мы вправе считать, что имеет место представление

$$y_i \equiv \frac{\nu_i}{N_i} = p_i + \xi_i, \quad \xi_i \in \Phi_{0,\sigma_i^2}$$

(в слагаемое ξ_i входят также случайные помехи (фон)). В силу (23), (24) мы будем иметь два возможных варианта регрессии:

$$p_i = \sum_{k=0}^{1} \alpha_k \psi_k(x_i), \quad \psi_k(x) = x^k, \qquad k = 0, 1 \tag{25}$$

(гипотеза H_1) и

$$p_i = \sum_{k=0}^{3} \alpha_k \psi_k(x_i), \quad \psi_k(x) = \frac{x^k}{x^2 + d^2}, \qquad k = 0, 1, 2, 3 \tag{26}$$

(гипотеза H_2).

Значения σ_i^2 при изменении гипотез меняются очень мало; они могут быть весьма точно оценены, и их можно считать известными. Тогда на основании теоремы 2 статистика

$$\chi^2 = \left| Y' - \sum_k \alpha_k^* \psi_k \right|^2 = \sum_{i=1}^{n} \left(y_i - \sum_k \alpha_k^* \psi_k(x_i) \right)^2 \Big/ \sigma_i^2 \tag{27}$$

будет иметь распределение H_{n-r}, где r — число оцениваемых параметров α_k ($r = 2$ при гипотезе H_1 и $r = 4$ при гипотезе H_2).

Проводя необходимые вычисления в соответствии с рекомендациями теоремы 2, мы получим для статистики (27) значения: в первом случае ($r = 2$) $\chi_1^2 = 36{,}8$, во втором ($r = 4$) $\chi_2^2 = 19{,}0$. Фактически достигаемые уровни значимости (см. § 3.4[6]) критерия $\chi^2 > c$ для проверки (в качестве основных) гипотез H_1 и H_2 составят соответственно $\mathrm{H}_{18}((0,\ 36{,}8)) = 0{,}9944$, $\mathrm{H}_{16}((0,\ 19{,}0)) = 0{,}731$.

Другими словами, предположение об отсутствии дополнительного канала рождения пар π-мезонов отвергается критерием, основанным на статистике χ^2, с уровнем значимости, равным, например, 0,99. В то же время предположение о наличии этого канала хорошо согласуется с экспериментальными данными.

Если быть более точным, то в этой задаче мы должны были бы проверять две сложные параметрические гипотезы, соответствующие предположениям (25), (26) для значений вероятностей появления пар π-мезонов. Если использовать критерий отношения правдоподобия, то он, как нетрудно проверить, будет основан на разности статистик χ^2, соответствующих моделям (25), (26), и, стало быть, результаты его будут примерно теми же.

4. Оценивание и проверка гипотез при наличии линейных связей. Рассмотрим, как и прежде, линейную регрессию (1), (2), но в предположении, что координаты вектора α связаны $s < r$ линейными соотношениями

$$\sum_{k=1}^{r} \alpha_k a_{kl} = c_l, \quad l = 1, \ldots, s.$$

В матричной форме эти соотношения можно записать в виде

$$\alpha A = c, \tag{28}$$

где A — матрица порядка $r \times s$. Мы будем предполагать, что A имеет ранг s. В этом случае мы могли бы выразить s переменных (скажем, $\alpha_{r-s+1}, \ldots, \alpha_r$) через остальные (т. е. через $\alpha_1, \ldots, \alpha_{r-s}$), подставить полученные значения в (1), (2) и получить снова стандартную задачу линейной регрессии (но с измененным регрессором).

Однако для дальнейшего нам будет удобнее подойти к решению этой задачи несколько иначе. Обратимся к доказательству теоремы 1. Подпространство $\mathscr{A}$ значений α, определяемое соотношениями (28), выделяет в $\mathscr{L}[X]$ подпространство размерности s значений αX, которое мы обозначим $\mathscr{L}_A[X]$. Очевидно, что оценивание $\alpha \in \mathscr{A}$ можно вести теперь теми же приемами, которые мы использовали в теореме 1. Нужная оценка $\alpha_A^* \in \mathscr{A}$ будет определяться, как и в теореме 1, с помощью проекции $\alpha_A^* X$ вектора Y на $\mathscr{L}_A[X]$. Таким образом, наряду с соотношением $(Y - \alpha^* X) \perp \mathscr{L}[X]$ мы будем иметь соотношение $(Y - \alpha_A^* X) \perp \mathscr{L}_A[X]$, которое однозначно определяет α_A^*. Для получения самого значения α_A^* удобнее воспользоваться аналитическим подходом — применить метод неопределенных множителей Лагранжа для отыскания $\min_\alpha |Y - \alpha X|^2$ при условии $\alpha A = c$. Для этого мы должны решить уравнения

$$\alpha A = c, \quad \frac{\partial}{\partial \alpha}\left[\,|Y - \alpha X|^2 + \lambda(\alpha A - c)^T\right] = 0 \tag{29}$$

(мы используем множители $\lambda_1, \ldots, \lambda_s$, образующие вектор λ и соответствующие условиям (28)). Так как $|Y - \alpha X|^2 = (Y - \alpha X)(Y - \alpha X)^T$, то второе из уравнений (29) примет вид

$$-2YX^T - 2\alpha XX^T + \lambda A^T = 0.$$

Отсюда находим

$$\alpha_A^* = YX^T D^{-1} - \frac{1}{2}\lambda A^T D^{-1} = \alpha^* - \frac{1}{2}\lambda A^T D^{-1}.$$

В силу (29) $c = \alpha_A^* A = \alpha^* A - \frac{1}{2}\lambda A^T D^{-1} A$. Так как матрица D положительно определена, A имеет ранг s, то $B = D^{-1/2}A$ также будет иметь ранг s, а матрица $B^T B = A^T D^{-1} A$ будет также положительно определенной (см. п. 1). Следовательно,

$$-\frac{1}{2}\lambda = (c - \alpha^* A) D_A,$$

$$\alpha_A^* = \alpha^* + (c - \alpha^* A) D_A A^T D^{-1}, \tag{30}$$

где мы положили для краткости $D_A = [A^T D^{-1} A]^{-1}$.

Читатель может проверить, что это есть о. м. п. параметра α при условии $\alpha A = c$. Тот же результат (30) можно получить и из геометрических соображений, используя соотношения $\alpha_A^* X \in \mathscr{L}_A[X]$ и ортогональность

$$\begin{gathered}(Y - \alpha_A^* X) \perp \mathscr{L}_A[X], \\ (\alpha_A^* - \alpha^*) X = \left[(Y - \alpha^* X) - (Y - \alpha_A^* X)\right] \perp \mathscr{L}_A[X].\end{gathered} \tag{31}$$

Обратимся теперь к задаче *проверки линейных гипотез*. Гипотезу H_1 относительно параметра α мы будем называть *линейной*, если она имеет вид $H_1 = \{\alpha A = c\}$, где матрицы A и c определены выше.

Сразу же можно отметить, что, введя новый параметр $\beta = \alpha A_c$, где A_c — любая невырожденная матрица, у которой первые s столбцов совпадают с A, мы сведем задачу к регрессии

$$Y = \beta X' + \xi, \qquad X' = A_c^{-1} X, \tag{32}$$

и к проверке гипотезы $\bar{\beta} = c$, $\bar{\beta} = (\beta_1, \ldots, \beta_s)$ (см. п. 2).

Естественно также исходить из следующих соображений. Чем больше αA отличается от c, тем αX удаленнее от $\mathscr{L}_A[X]$ и тем больше точки αX, $\alpha^* X$ будут отличаться от $\alpha_A^* X \in \mathscr{L}_A[X]$. Поэтому в основу критерия для проверки H_1 естественно положить расстояние $\alpha_A^* X$ от $\alpha^* X$. Если гипотеза H_1 верна, то в силу (31)

$$|(\alpha_A^* - \alpha^*)X|^2 = |Y - \alpha_A^* X|^2 - |Y - \alpha^* X|^2. \qquad (33)$$

В силу (30) (с заменой c на αA) $\alpha_A^* - \alpha^*$ есть результат линейного преобразования над $\alpha - \alpha^*$. Поэтому $(\alpha_A^* - \alpha^*)X$ не зависит от $Y - \alpha^* X$ (см. теорему 1).

Далее, в силу (30)

$$|(\alpha_A^* - \alpha^*)X|^2 = (\alpha_A^* - \alpha^*)XX^T(\alpha_A^* - \alpha^*)^T =$$
$$= (c - \alpha^* A)D_A(c - \alpha^* A) = (\alpha^* - \alpha)AD_A A^T(\alpha^* - \alpha)^T. \qquad (34)$$

Так как

$$(\alpha^* - \alpha)A = \xi X^T D^{-1} A \in \Phi_{0,\sigma^2 A^T D^{-1} A} = \Phi_{0,\sigma^2 D_A^{-1}},$$

то в силу (34) и § 2.2[6] (п. 4)

$$\frac{1}{\sigma^2}|(\alpha_A^* - \alpha^*)X|^2 \in \mathrm{H}_s. \qquad (35)$$

Из сказанного и из теоремы 1 следует, что

$$\frac{|(\alpha_A^* - \alpha^*)X|^2}{|Y - \alpha^* X|^2} = \frac{|Y - \alpha_A^* X|^2}{|Y - \alpha^* X|^2} - 1 \in \mathbf{F}_{s,n-r}. \qquad (36)$$

Соотношения (35), (36) дают нам возможность строить критерии (основанные на использовании удаленности $\alpha^* X$ от $\alpha_A^* X$) для проверки гипотезы H_1 соответственно в случаях, когда σ^2 известно и неизвестно (см. гл. 3[6]).

Важно отметить, что гипотеза H_1 есть гипотеза о принадлежности α параметрическому подсемейству (при наличии мешающего параметра σ^2, если σ^2 неизвестно), а статистики (35), (36) есть не что иное, как статистики отношения правдоподобия (см. §§ 3.10[6], 3.15[6]). Действительно, пусть, например, σ^2 неизвестно. Тогда (см. (5), (8))

$$\sup_{\alpha,\sigma} f_\theta(Y) = \sup_{\alpha,\sigma}(\sqrt{2\pi}\sigma)^{-n}\exp\left\{-\frac{|Y - \alpha X|^2}{2\sigma^2}\right\} =$$
$$= (\sqrt{2\pi}\hat{\sigma}^*)^{-n}\exp\left\{-\frac{|Y - \alpha^* X|^2}{2(\sigma^2)^*}\right\} = \left(\sqrt{2\pi}\frac{|Y - \alpha^* X|}{n}\right)^{-n} e^{-n/2}.$$

Значение $\sup\limits_{\alpha\in\mathcal{A},\sigma} f_\theta(Y)$ вычисляется точно так же. Надо лишь заметить, что о. м. п. для α в случае $\alpha\in\mathcal{A}$ будет равна α_A^*, а о. м. п. для σ^2 будет равна, аналогично (8) $\frac{1}{n}|Y-\alpha_A^*X|^2$. Поэтому

$$\sup_{\alpha\in\mathcal{A},\sigma} f_\theta(Y)=\left(\sqrt{2\pi}\frac{|Y-\alpha_A^*X|}{n}\right)^{-n}e^{-n/2},$$

$$\frac{\sup\limits_{\alpha\in\mathcal{A},\sigma} f_\theta(Y)}{\sup\limits_{\alpha,\sigma} f_\theta(Y)}=\frac{|Y-\alpha^*X|^n}{|Y-\alpha_A^*X|^n}$$

и, стало быть, статистика критерия отношения правдоподобия эквивалентна (36).

Если σ^2 известно, то в основу критерия для проверки H_1 можно положить соотношение (35). Читатель может аналогично предыдущему убедиться, что это тоже есть критерий отношения правдоподобия. Так как этот критерий инвариантен относительно замены параметра (см. § 3.10[6]), то в силу замечания и утверждений §§ 3.9[6], 3.10[6] можно утверждать, что к. о. п.

$$|(\alpha_A^*-\alpha^*)X|^2 > \sigma^2 h_\varepsilon,$$

где h_ε — квантиль порядка $1-\varepsilon$ распределения $\mathbf{H}_\varepsilon$, будет минимаксным критерием уровня $1-\varepsilon$ для проверки H_1 против отделенных соответствующим образом альтернатив.

Сказанное выше и результаты гл. 2[6], 3[6] (см., в частности, § 3.15[6]) дают основания считать, что критерий (36), как и оценка (30), также будут обладать свойствами оптимальности. Подробнее на этом мы здесь останавливаться не будем. Весьма полное изложение задач регрессии можно найти в [29].

§ 4. Дисперсионный анализ

Задачи дисперсионного анализа, излагаемые ниже, по своему существу относятся к задачам регрессии. В последних изучалась зависимость наблюдений от числового фактора x, который мог принимать любые наперед заданные значения $x_1, \ldots, x_n$, и каждому из них соответствовало одно наблюдение. В задачах дисперсионного анализа обычно изучается воздействие лишь дискретных факторов (одного, двух или более), которые могут принимать лишь конечное число значений. При каждом из этих значений мы располагаем набором наблюдений (выборкой). Дисперсионный анализ объединяет группу статистических приемов, основанных на анализе среднеквадратических уклонений и предназначенных для проверки различных гипотез и оценки параметров, связанных с воздействием факторов. Основы дисперсионного анализа были заложены Фишером.

1. Задачи дисперсионного анализа как задачи регрессии. Случай одного фактора. Пусть дано r независимых выборок

$$Y_1 = (y_{11}, \ldots, y_{1n_1}), \ldots, Y_r = (y_{r1}, \ldots, y_{rn_r})$$

объемов $n_1, \ldots, n_r$ из нормальных совокупностей: $Y_k \in \Phi_{\alpha_k, \sigma^2}$. Предполагается, что наблюдения Y_k, $k = 1, \ldots, r$, проводились при разных значениях некоторого фактора, роль которого нас интересует, и что влияние этого фактора сказывается на значении среднего α_k. Значение дисперсии σ^2 предполагается одним и тем же для всех выборок и, как правило, неизвестным. Задачи дисперсионного анализа включают в себя проверку гипотез, относящихся к значениям $\alpha_1, \ldots, \alpha_r$, и, в частности, гипотезы об однородности $\alpha_1 = \ldots = \alpha_r = \alpha$ (эта последняя задача рассматривалась нами в § 1), а также оценки параметров α_k и их изменчивости.

Дисперсионный анализ, как и задачи регрессии вообще, имеет чрезвычайно широкое поле применений, особенно в социологии, сельском хозяйстве, биологии, медицине. В качестве весьма типичной задачи для применения методов дисперсионного анализа можно назвать, например, проблему о выяснении зависимости содержания холестерина в крови человека от его профессии.

Сформулированные выше задачи дисперсионного анализа есть частные случаи задач линейной регрессии. Действительно, наблюдения y_{ki} можно представить в форме

$$y_{ki} = \alpha_k + \xi_{ki}, \quad \xi_{ki} \in \Phi_{0,\sigma^2}, \quad k = 1, \ldots, r, \quad i = 1, \ldots, n_k. \tag{1}$$

Образуем вектор

$$Y = (y_{11}, \ldots, y_{1n_1}; y_{21}, \ldots, y_{2n_2}; \ldots; y_{r1}, \ldots, y_{rn_r})$$

и вектор ξ по тому же правилу. Тогда соотношения (1) можно записать в матричной форме $Y = \alpha X + \xi$, где X есть матрица размера $r \times n$, $n = n_1 + \ldots + n_r$, вида

$$X = \left(\begin{array}{c|c|c|c} 1\ \ 1 \ldots 1 & 0\ \ 0 \ldots 0 & & 0\ \ 0 \ldots 0 \\ 0\ \ 0 \ldots 0 & 1\ \ 1 \ldots 1 & & 0\ \ 0 \ldots 0 \\ 0\ \ 0 \ldots 0 & 0\ \ 0 \ldots 0 & & 0\ \ 0 \ldots 0 \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot\ \cdot & & \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ 0\ \ 0 \ldots 0 & 0\ \ 0 \ldots 0 & & 1\ \ 1 \ldots 1 \end{array}\right).$$

Очевидно, что строки этой матрицы (векторы X_j) ортогональны. Гипотезу $H_1 = \{\alpha_1 = \alpha_2 = \ldots = \alpha_r\}$ можно записать в виде

$$\alpha A = 0, \tag{2}$$

где A есть матрица размера $r \times (r-1)$:

$$A = \begin{pmatrix} 1 & 0 \ldots & 0 \\ 0 & 1 \ldots & 0 \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ 0 & 0 \ldots & 1 \\ -1 & -1 \ldots & -1 \end{pmatrix}.$$

Очевидно, что A имеет ранг $r-1$.

Мы видим, что проверка основной гипотезы H_1 дисперсионного анализа есть не что иное, как задача о проверке линейной гипотезы для регрессии.

Выясним, что из себя представляют эффективные оценки для α и σ^2, найденные в теореме 3.1. В нашем случае $|X_k|^2 = n_k$, матрица $D = XX^T$ порядка $r \times r$ имеет вид

$$D = \begin{pmatrix} n_1 & 0 & \ldots 0 \\ 0 & n_2 & \ldots 0 \\ \cdot & \cdot & \cdot\ \cdot\ \cdot \\ 0 & 0 & \ldots n_r \end{pmatrix},$$

$$\alpha_k^* = \frac{(Y, X_k)}{(X_k, X_k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} y_{ki} \equiv \bar{y}_{k\cdot}, \tag{3}$$

$$(n-r)(\sigma^2)^* = |Y - \alpha^* X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \bar{y}_{k\cdot})^2 \equiv Q_2(Y).$$

При этом $\alpha_1^*, \ldots, \alpha_r^*, (\sigma^2)^*$ независимы. Доверительные интервалы для параметров α, σ^2 и функций от них строятся так же, как в § 3.

Для проверки линейной гипотезы (2) нам нужно вычислить также о. м. п. α_A^* при наличии условия (2) (см.

п. 4 предыдущего параграфа). Простейший путь здесь — использовать подход, изложенный в начале п. 4 § 3 и выразить $\alpha_1, \ldots, \alpha_r$ через независимые переменные. В нашем случае независимая переменная лишь одна — пусть это будет $\alpha_r = \mu$, и $\alpha_A^* = (\mu^*, \ldots, \mu^*)$, где μ^* минимизирует

$$|Y - (\mu, \ldots, \mu) X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \mu)^2.$$

Очевидно, что

$$\mu^* = \bar{y} \equiv \frac{1}{n} \sum_{k=1}^{r} \sum_{i=1}^{n_k} y_{ki},$$

$$|Y - \alpha_A^* X|^2 = \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \bar{y})^2 \equiv Q(Y) =$$

$$= \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \bar{y} + \bar{y}_{k\cdot} - \bar{y}_{k\cdot})^2 =$$

$$= \sum_{k=1}^{r} \sum_{i=1}^{n_k} (y_{ki} - \bar{y}_{k\cdot})^2 + \sum_{k=1}^{r} n_k (\bar{y}_{k\cdot} - \bar{y})^2$$

(сумма смешанных произведений равна нулю, так как $\sum_{i=1} (y_{ki} - \bar{y}_{k\cdot}) = 0$). Если гипотеза H_1 верна, то в силу (3.33), (3) и только что полученного равенства

$$|(\alpha_A^* - \alpha^*) X|^2 = Q(Y) - Q_2(Y) = \sum_{k=1}^{r} n_k (\bar{y}_{k\cdot} - \bar{y})^2 \equiv Q_1(Y).$$

На основании (3.36) при выполнении H_1 мы получаем $Q_1(Y)/Q_2(Y) \Subset \mathrm{F}_{r-1, n-r}$, что дает нам возможность строить критерий $Q_1(Y)/Q_2(Y) > f_\varepsilon$ (f_ε — квантиль $\mathrm{F}_{r-1, n-r}$ порядка $1-\varepsilon$) для проверки H_1, который будет к. о. п. Если σ^2 известно, к. о. п. будет иметь вид

$$Q_1(Y) > \sigma^2 h_\varepsilon$$

(h_ε — квантиль H_{r-1}) и будет минимаксным критерием при соответствующим образом отделенных альтернативах (см. § 3.9[61]).

2. Влияние двух факторов. Элементарный подход. В задачах этого раздела исследуется влияние на результаты эксперимента факторов двух видов. Применительно, скажем, к сельскому хозяйству это может быть изучение влияния на урожайность состава почвы (фактор A принимает r значений) и способа обработки (фактор B принимает s значений).

Здесь наблюдения можно представить в форме

$$y_{kli}=\alpha_{kl}+\xi_{kli},\quad \xi_{kli}\in\Phi_{0,\sigma^2},\qquad (4)$$
$$k=1,\ldots,r,\ l=1,\ldots,s,\ i=1,\ldots,n_{kl},$$

и рассматриваемая модель по существу ничем не будет отличаться от модели (1), рассмотренной в п. 1. Стало быть, здесь также применимы все результаты § 3, но прямое их применение становится более громоздким. Громоздким является уже само присутствие тройных индексов. Чтобы несколько упростить задачу, мы положим $n_{kl}=1$; это избавит нас от одного индекса (индекса i в (4)). Кроме того, мы предложим в этом разделе несколько иной, элементарный подход, который независимо от теорем § 3 позволит получить утверждения, необходимые для проверки основных гипотез.

Итак, будем рассматривать выборки $Y_{kl}=y_{kl}$ единичного объема, так что набор экспериментальных данных Y здесь будет представлять собой матрицу $r\times s$ чисел y_{kl}, определяющих результат эксперимента под воздействием k-го значения фактора A и l-го значения фактора B. Эту матрицу можно трактовать как r выборок (строк) объема s, соответствующих разным значениям фактора A, или как s выборок (столбцов) объема r, соответствующих разным значениям фактора B. В соответствии с этим в дальнейшем и будет происходить группировка наблюдений. Положим

$$\bar{y}_{k\cdot}=\frac{1}{s}\sum_{l=1}^{s}y_{kl},\quad \bar{y}_{\cdot l}=\frac{1}{r}\sum_{k=1}^{r}y_{kl},\quad \bar{y}=\frac{1}{rs}\sum_{k,l}y_{kl}.$$

Справедливо тождество

$$Q(Y)\equiv\sum_{k,l}(y_{kl}-\bar{y})^2=Q_1(Y)+Q_2(Y)+Q_3(Y),\qquad (5)$$

где

$$Q_1(Y)=s\sum_{k}(\bar{y}_{k\cdot}-\bar{y})^2,\quad Q_2(Y)=r\sum_{l}(\bar{y}_{\cdot l}-\bar{y})^2,$$
$$Q_3(Y)=\sum_{k,l}(y_{kl}-\bar{y}_{k\cdot}-\bar{y}_{\cdot l}+\bar{y})^2.$$

Мы предположим, что влияние факторов аддитивно, т. е.

существуют a_k, b_l такие, что

$$\alpha_{kl} = a_k + b_l, \quad k = 1, \ldots, r, \quad l = 1, \ldots, s. \tag{6}$$

Очевидно, что Q_1 определяет изменчивость значений a_k (т. е. связано с фактором A), Q_2 определяет изменчивость b_l (фактор B), а Q_3 есть сумма, создаваемая всецело случайностью. Очевидно также, что

$$Q_i(Y + a) = Q_i(Y), \quad i = 1, 2, 3. \tag{7}$$

Теорема 1. 1)

$$Q_3(Y)/\sigma^2 \in \mathrm{H}_{(r-1)(s-1)}. \tag{8}$$

2) *Если верна гипотеза* $H_A = \{a_1 = \ldots = a_r = a\}$, *то* $Q_1(Y)$ *не зависит от* $Q_2(Y)$ *и* $Q_3(Y)$, $Q_1(Y)/\sigma^2 \in \mathrm{H}_{r-1}$. *Аналогичное утверждение имеет место относительно* Q_2 *и гипотезы* $H_B = \{b_1 = \ldots = b_s = b\}$.

3) *Если справедлива гипотеза* $H_1 = \{\alpha_{kl} = \alpha\}$, *то все три квадратичные формы* Q_1, Q_2, Q_3 *независимы.*

Доказательство. Положим, не ограничивая общности, $\sigma^2 = 1$. Тогда

$$\mathsf{M} y_{kl} y_{ij} = \begin{cases} \alpha_{kl}\alpha_{ij}, & \text{если } (i, j) \neq (k, l), \\ \alpha_{kl}^2 + 1, & \text{если } (i, j) = (k, l). \end{cases}$$

Отсюда следует, что

$$\mathsf{M}\left(\sum_{\mathrm{I}} y_{kl}\right)\left(\sum_{\mathrm{II}} y_{kl}\right) = \left(\sum_{\mathrm{I}} \alpha_{kl}\right)\left(\sum_{\mathrm{II}} \alpha_{kl}\right) + m,$$

где m — число одинаковых слагаемых в суммах $\sum\limits_{\mathrm{I}}$ и $\sum\limits_{\mathrm{II}}$. Пользуясь этим равенством, легко получить теперь, что

$$\mathsf{M}(\bar{y}_{k\cdot} - \bar{y})(\bar{y}_{\cdot l} - \bar{y}) = (\alpha_{k\cdot} - \bar{\alpha})(\alpha_{\cdot l} - \bar{\alpha}) = \\ = (a_k - \bar{a})(b_l - \bar{b}) \tag{9}$$

при естественных соглашениях относительно обозначений $\alpha_{k\cdot}, \alpha_{\cdot l}, \bar{\alpha}, \bar{a}, \bar{b}$. Если верна гипотеза $H_A = \{a_1 = \ldots \ldots = a_r = a\}$, то математическое ожидание в (9) равно нулю. Так как при этом $\mathsf{M}(\bar{y}_{k\cdot} - \bar{y}) = \alpha_{k\cdot} - \bar{\alpha} = 0$, то установленный факт означает, что набор случайных величин $\{\bar{y}_{k\cdot} - \bar{y}\}$ не зависит от $\{\bar{y}_{\cdot l} - \bar{y}\}$.

Аналогично устанавливаем, что при любых k, l, i

$$\mathsf{M}(y_{kl} - \bar{y}_{k\cdot})(\bar{y}_{i\cdot} - \bar{y}) = 0.$$

Это значит, что совокупность $\{\overline{y}_{k\cdot} - \overline{y}\}$ не зависит также от $\{y_{kl} - \overline{y}_{k\cdot} - \overline{y}_{\cdot l} + \overline{y}\}$. Это означает в свою очередь, что $Q_1(Y)$ при выполнении H_A не зависит от $Q_2(Y)$ и $Q_3(Y)$. Тот факт, что $Q_1(Y) \in \mathrm{H}_{r-1}$, вытекает из леммы Фишера (§ 2.32[6]).

Аналогично обстоит дело, если выполнена гипотеза H_B. Если же справедлива гипотеза H_1 (т. е. справедливы H_A и H_B), то, очевидно, все три названных выше набора случайных величин будут независимы. Это означает независимость $Q_1(Y)$, $Q_2(Y)$, $Q_3(Y)$.

Нам осталось найти распределение $Q_3(Y)$. Так как это распределение от a_k, b_l не зависит, то мы можем считать, что $a_k = b_l = 0$ для всех k, l и что, стало быть, выполнена H_1. Тогда из определения $Q(Y)$ следует, что $Q(Y) \in \mathrm{H}_{rs-1}$. Кроме того, справедливо (5), где $Q_1(Y) \in \mathrm{H}_{r-1}$, $Q_2(Y) \in \mathrm{H}_{s-1}$. Остается воспользоваться независимостью $Q_i(Y)$ и леммой 3.1. Теорема доказана.

Аналогичный подход может быть применен и к задачам п. 1.

Из теоремы 1 вытекает возможность построения следующих статистических процедур:

1) Оценка параметров $a_k - a_i$, $b_l - b_j$, σ^2 (числа a_k, b_l в (6) определены с точностью до постоянного слагаемого) с помощью оценок $\overline{y}_{k\cdot} - \overline{y}_{i\cdot}$, $\overline{y}_{\cdot l} - \overline{y}_{\cdot j}$, $(\sigma^2)^* = Q_3(Y)/(r-1)(s-1)$. Так как по существу проведенные выше рассмотрения совпадают с тем, что мы делали в § 3 и в п. 1 настоящего параграфа, то названные оценки будут эффективными. Доверительные интервалы для σ^2, $a_k - a_i$ можно строить с помощью соотношений (8),

$$\overline{y}_{k\cdot} - \overline{y}_{i\cdot} - (a_k - a_i) \in \Phi_{0,2\sigma^2/s},$$

$$\frac{\overline{y}_{k\cdot} - \overline{y}_{i\cdot} - (a_k - a_i)}{\sqrt{\dfrac{2Q_3(Y)}{s(r-1)(s-1)}}} \in \mathrm{T}_{(r-1)(s-1)}$$

(для $b_l - b_j$ все происходит аналогично).

2) Проверка гипотезы H_A с помощью критерия $Q_1/Q_3 > f_\varepsilon$. Критерий будет иметь уровень $1-\varepsilon$, если f_ε есть квантиль порядка $1-\varepsilon$ распределения $\mathrm{F}_{r-1,\,(r-1)(s-1)}$.

Аналогичный вид будет иметь критерий для проверки H_B: $Q_2/Q_3 > f_\varepsilon$, где f_ε — квантиль порядка $1-\varepsilon$ распределения $\mathrm{F}_{s-1,\,(r-1)(s-1)}$.

3) Проверка гипотезы H_1 с помощью критерия

$$\frac{Q_1+Q_2}{Q_3} > f_\varepsilon$$

уровня $1-\varepsilon$, где f_ε — квантиль порядка $1-\varepsilon$ распределения $\mathbf{F}_{r+s-2,\,(r-1)(s-1)}$.

Более подробно задачи дисперсионного анализа рассмотрены в [29], [39].

§ 5. Распознавание образов

В этом параграфе мы коснемся кратко круга задач, для обозначения которого наряду с названием «распознавание образов» иногда используются также термины «классификация» и «дискриминантный анализ» *).

В § 3.1[6] мы рассматривали следующую задачу о проверке r простых гипотез. Даны распределения $\mathbf{P}_1, \ldots, \mathbf{P}_r$ и выборка X объема n. Требуется определить, какая из гипотез

$$H_j = \{X \Subset \mathbf{P}_j\} \tag{1}$$

верна.

Однако в практических задачах распределения $\mathbf{P}_j$ часто неизвестны, и судить о них мы можем опять-таки лишь по выборкам.

Итак, пусть дано r выборок $X_i = (x_{i1}, \ldots, x_{in_i})$, $i = 1, \ldots, r$, объемов $n_1, \ldots, n_r$ соответственно, отвечающих r различным *неизвестным* нам распределениям $\mathbf{P}_1, \ldots, \mathbf{P}_r$, и пусть, кроме того, дана выборка X. Надо решить все ту же задачу: определить, какая из гипотез (1) верна. Другими словами, надо установить, продолжением какой из выборок $X_1, \ldots, X_r$ является выборка X. Это и есть задача распознавания образов.

Для простоты мы ограничимся рассмотрением случая $r = 2$.

1. Параметрический случай. Предположим сначала, что $\mathbf{P}_i$ принадлежат некоторому параметрическому семейству $\{\mathbf{P}_\theta\}$, удовлетворяющему условию (A_μ), т. е. $X_1 \Subset \mathbf{P}_{\theta_1}$, $X_2 \Subset \mathbf{P}_{\theta_2}$, $X \Subset \mathbf{P}_\theta$ при некоторых $\theta_1 \neq \theta_2$ и $\theta = \theta_1$

*) Следует отметить, что последние два термина употребляются также и для обозначения иных задач, например задач, в которых распределения $\mathbf{P}_i$ в (1) известны.

или $\theta=\theta_2$. Первое из этих равенств соответствует гипотезе $H_1=\{X\in\mathbf{P}_{\theta_1}\}$, второе—гипотезе $H_2=\{X\in\mathbf{P}_{\theta_2}\}$.

Предположим, далее, опять-таки ради простоты, что объемы выборок n_1, n_2, n равны: $n_1=n_2=n$.

Рассмотрим объединенную выборку (X_1, X_2, X) и представим ее как выборку объема n, образованную наблюдениями (x_{1i}, x_{2i}, x_i) и относящуюся к распределению $\mathbf{P}_{\theta_1}\times\mathbf{P}_{\theta_2}\times\mathbf{P}_{\theta}$, имеющему плотность $f_{\theta_1}(x_1)f_{\theta_2}(x_2)f_{\theta}(x)$, зависящую от параметра $\bar\theta=(\theta_1, \theta_2, \theta)$. Очевидно, что функция правдоподобия выборки (X_1, X_2, X) будет равна

$$f_{\bar\theta}(X_1, X_2, X)=f_{\theta_1}(X_1)f_{\theta_2}(X_2)f_{\theta}(X).$$

Мы приходим к задаче о проверке гипотезы H_1 о том, что параметр $\bar\theta$ лежит на «кривой» $\theta=\theta_1$, против гипотезы H_2 о том, что $\bar\theta$ располагается на другой «кривой» $\theta=\theta_2$. Это есть задача проверки гипотезы о принадлежности параметрическому подсемейству (см. § 3.15[6]), однако в том случае, когда конкурирующая гипотеза означает принадлежность другому параметрическому подсемейству. Рассмотрение этой задачи происходит аналогично § 3.15[6], но по своей технической трудности выходит за рамки настоящей книги. Мы ограничимся здесь тем, что для случая одномерного параметра θ опишем вкратце существо результата. Оно вполне аналогично содержанию § 3.15[6]: если параметр θ локализован, т. е. точки θ_1, θ_2 располагаются в окрестности некоторой точки θ_0, $|\theta_1-\theta_2|>b/\sqrt{n}$, и если семейство $\{\mathbf{P}_\theta\}$ удовлетворяет в точке θ_0 условиям регулярности (RR), то критерий отношения правдоподобия

$$\frac{\sup\limits_{\theta_1,\theta_2} f_{\theta_1}(X_1)\,f_{\theta_2}(X_2)\,f_{\theta_2}(X)}{\sup\limits_{\theta_1,\theta_2} f_{\theta_1}(X_1)\,f_{\theta_2}(X_2)\,f_{\theta_1}(X)}>c \qquad (2)$$

будет при $n\to\infty$ асимптотически минимаксным для проверки H_1 против H_2.

Ограничение $n_1=n_2=n$ несущественно. Оно устраняется так же, как и в рассмотрениях § 1.

2. Общий случай. В общем случае, когда у нас нет оснований предполагать, что X_i связаны с параметрическим семейством, возможен общий подход, в основе которого лежат те же соображения, которые использовались нами при построении в § 2 критериев однородности.

В этом случае критерий π для проверки H_1 против H_2 будет функцией трех выборок, так что $\pi=\pi(X_1, X_2, X)$ есть вероятность принять H_2 при данных (X_1, X_2, X). Как и прежде, нерандомизированный критерий определяется критической областью $\Omega\subset\mathscr{X}^{n_1+n_2+n}$ в пространстве значений (X_1, X_2, X). Уровнем значимости критерия естественно называть число

$$1-\varepsilon=\inf_{\mathbf{P}_1\in\mathscr{P},\mathbf{P}_2\in\mathscr{P}}\mathbf{P}_1\times\mathbf{P}_2\times\mathbf{P}_1((X_1, X_2, X)\notin\Omega),$$

где $\mathscr{P}$ есть класс допустимых распределений. Значение

$$\beta_\pi\mathbf{P}_1, \mathbf{P}_2=\mathbf{P}_1\times\mathbf{P}_2\times\mathbf{P}_2((X_1, X_2, X)\in\Omega),$$
$$\mathbf{P}_1\in\mathscr{P},\ \mathbf{P}_2\in\mathscr{P},$$

есть мощность критерия в точке $(\mathbf{P}_1, \mathbf{P}_2)$.

Критерий π называется состоятельным, если $\beta_\pi(\mathbf{P}_1, \mathbf{P}_2)\to 1$ при $n_1\to\infty$, $n_2\to\infty$, $n\to\infty$ и при любых $\mathbf{P}_1\neq\mathbf{P}_2$, $\mathbf{P}_1\in\mathscr{P}$, $\mathbf{P}_2\in\mathscr{P}$.

В качестве основы для построения состоятельных критериев можно использовать хорошо известный нам факт о сближении эмпирических распределений $\mathbf{P}^*_{X_1}$ и $\mathbf{P}^*_{X_2}$ для выборок X_1 и X_2 с $\mathbf{P}_1$ и $\mathbf{P}_2$ соответственно. Если $d(\mathbf{P}, \mathbf{Q})$ есть некоторое расстояние между распределениями, то при гипотезе H_2 расстояние $d(\mathbf{P}^*_{X_2}, \mathbf{P}^*_X)$ должно быть меньше, чем $d(\mathbf{P}^*_{X_1}, \mathbf{P}^*_X)$. Поэтому в качестве критерия можно использовать неравенство

$$d(\mathbf{P}^*_{X_2}, \mathbf{P}^*_X)-d(\mathbf{P}^*_{X_1}, \mathbf{P}^*_X)<c,$$

при выполнении которого принимается H_2. Расчет такого рода критериев (вычисление их уровней значимости и мощности) связан обычно с большими трудностями (ср. с более простыми задачами, изложенными в § 2).

Используя *группировку наблюдений*, мы можем в общем случае применять асимптотически оптимальный критерий (2). Предположим, что группировка по областям $\Delta_1, \ldots, \Delta_m$ произведена и $(\nu_{i1}, \ldots, \nu_{im})$ и $(\nu_1, \ldots, \nu_m)$ есть частоты попаданий в эти области наблюдений выборок X_i, $i=1, 2$, и X соответственно. Пусть, кроме того, $\theta_i=(\theta_{i1}, \ldots, \theta_{im})$ есть вероятности $(\mathbf{P}_i(\Delta_1), \ldots, \mathbf{P}_i(\Delta_m))$ попаданий в области $\Delta_1, \ldots, \Delta_m$ для распределений $\mathbf{P}_i$, $i=1, 2$. Так как для сгруппированной выборки X_i, $i=1, 2$, функция правдоподобия $f_{\theta_i}(X_i)$ равна

$f_{\theta_i}(X_i) = \prod\limits_{k=1}^{m} \theta_{ik}^{\nu_{ik}}$, то критерий (2) будет иметь вид

$$\sup_{\theta_2} \sum_{k=1}^{m} (\nu_{2k} + \nu_k) \ln \theta_{2k} + \sup_{\theta_1} \sum_{k=1}^{m} \nu_{1k} \ln \theta_{1k} - \\ - \sup_{\theta_1} \sum_{k=1}^{m} (\nu_{1k} + \nu_k) \ln \theta_{1k} - \sup_{\theta_2} \sum_{k=1}^{m} \nu_{2k} \ln \theta_{2k} > \ln c,$$

или

$$\sum_{k=1}^{m} (\nu_{2k} + \nu_k) \ln \frac{\nu_{2k} + \nu_k}{n_2 + n} + \sum_{k=1}^{m} \nu_{1k} \ln \frac{\nu_{1k}}{n_1} > \\ > \ln c + \sum_{k=1}^{m} (\nu_{1k} + \nu_k) \ln \frac{\nu_{1k} + \nu_k}{n_1 + n} + \sum_{k=1}^{m} \nu_{2k} \ln \frac{\nu_{2k}}{n_2}. \quad (3)$$

Аналогичные рассмотрения могут быть проведены и для $r > 2$.

ГЛАВА 2

ТЕОРЕТИКО-ИГРОВОЙ ПОДХОД К ЗАДАЧАМ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

В §§ 1—3 вводятся понятия обычной и статистической игр, выделяются основные классы оптимальных стратегий.

В §§ 4, 5 рассмотрены методы отыскания оптимальных статистических решений.

Материал §§ 6—8 посвящен построению асимптотически оптимальных решающих правил.

§ 1. Предварительные замечания

В [6] и в гл. 1 было рассмотрено большое количество разнообразных статистических задач. Всех их объединяет то обстоятельство, что статистик на основании экспериментальных данных должен принять некоторое решение. В теории оценок эти решения могут иметь форму точечных оценок θ^*, которые мы должны принять в качестве неизвестного параметра θ, в теории проверки статистических гипотез — форму утверждений, в которых говорится, какие предположения относительно природы исследуемого объекта являются верными, а какие — нет. Эти решения, если они ошибочны, как правило, сопряжены с последующими потерями. Например, ошибка в лабораторной оценке (производимой с помощью выборки) содержания различных компонент в руде может привести к нарушению оптимального режима плавки и ухудшению качества выплавляемого металла. Это означает, что мы понесем материальные потери, которые будут зависеть от величины ошибки. Ошибка, касающаяся эффективности действия медицинского препарата, который проверяется на выборочной группе больных, очевидно, тоже может привести к потерям, которые мы для единообразия подхода также будем считать возможным исчислять в некоторых единицах. Такое же соглашение мы примем и относительно других задач статистики, в которых потери не носят ясно выраженный материальный характер.

Сказанное позволяет нам в задачах математической статистики выделить следующие четыре общих элемента. Эти элементы, собственно, и определяют сущность каждой конкретной задачи. Для простоты в последующем изложении мы будем иметь в виду лишь *задачи с одной выборкой X фиксированного объема n.*

1) Множество Θ, элементы которого $\theta \in \Theta$ определяют состояние исследуемого объекта. Если θ известно, то необходимость построения статистического решения отсутствует. Множество Θ называют также множеством параметров, хотя θ могут допускать и более широкое толкование (например, множество Θ может быть очень богатым и совпадать с множеством всех распределений на некотором пространстве $\mathscr{X}$).

2) Чтобы получить информацию о неизвестном θ, статистик ставит эксперимент и производит наблюдения над некоторой случайной величиной, распределение которой зависит от θ. Другими словами, статистик располагает выборкой X из распределения $\mathbf{P}_\theta$. Как мы уже знаем, из выборки X можно извлечь информацию о $\mathbf{P}_\theta$ и, стало быть, о θ. Можно считать, что выполнено условие (A_0) (см. § 2.6[6I]) о взаимно однозначном соответствии между θ и $\mathbf{P}_\theta$.

3) В задачах статистики всегда определено множество $D=\{\delta\}$ решений, которые может принимать статистик. В теории оценивания множество D обычно совпадает с Θ, в задачах проверки гипотез множество D конечно, а число его элементов равно числу проверяемых гипотез. Если θ известно, то решение $\delta=\varphi(\theta)$ определяется однозначно. Если θ неизвестно, то решение δ желательно выбрать в известном смысле оптимально. Но оптимизация решений требует, чтобы мы имели возможность их сравнивать. Для этого мы будем предполагать, что задана функция потерь, определяющая количественно последствие принятия решений.

4) Функция потерь $w(\delta, \theta)$ определена на $D\times\Theta$ и указывает, какие потери мы будем нести, если примем решение δ, а исследуемый объект, к которому относится решение, находится в состоянии θ. Мы будем считать, что $w(\delta, \theta)>0$ при $\delta\neq\varphi(\theta)$, $w(\varphi(\theta), \theta)=0$.

Если из перечисленных четырех элементов убрать п. 2) об экспериментальных данных, то мы получим объект, представляющий из себя *обычную игру двух лиц*, игру, в которой роль первого игрока играет статистик (исследователь), а второго — природа.

§ 2. Основные понятия и теоремы, связанные с игрой двух лиц

1. Игра двух лиц.

Определение 1. *Игрой двух лиц* называется тройка (D, Θ, w), составленная из множеств D и Θ и функции w, отображающей $D\times\Theta$ в полупрямую $[0, \infty)$. Элементы δ множества D называются *стратегиями* (действиями) I *игрока*, элементы $\theta\in\Theta$ — *стратегиями* II *игрока*, а w есть *функция потерь* I *игрока* (или функция выигрыша второго игрока), определяющая потери $w(\delta, \theta)$, которые будет нести I игрок, если он выберет стратегию δ, а II игрок — стратегию θ.

Основная задача теории игр двух лиц состоит в выборе оптимальной стратегии I игрока, которого мы часто будем отождествлять с собой. Для этого множество стратегий нужно как-то упорядочить. Сделать это непросто, так как потери $w(\delta, \theta)$, с помощью которых мы должны производить упорядочение, зависят от двух аргументов, так что для каждого θ стратегия δ, минимизирующая $w(\delta, \theta)$, будет, вообще говоря, своя.

Определение 2. Мы будем говорить, что стратегия δ_1 *лучше, чем* δ_2, если

$$w(\delta_1, \theta)\leqslant w(\delta_2, \theta) \text{ при всех } \theta\in\Theta \tag{1}$$

и существует по крайней мере одно значение $\theta_1\in\Theta$, при котором $w(\delta_1, \theta_1)< w(\delta_2, \theta_1)$.

Если выполнено лишь (1), то мы будем говорить, что стратегия δ_1 *не хуже, чем* δ_2.

Стратегию δ_0, для которой

$$w(\delta_0, \theta)\leqslant w(\delta, \theta) \text{ при всех } \delta \text{ и } \theta,$$

мы будем называть *равномерно оптимальной* (или равномерно наилучшей).

Равномерно наилучшая стратегия обеспечивает наименьшие потери при всех θ. Однако, как правило, такие стратегии не существуют.

Мы отметим следующие три подхода к исследованию оптимальных стратегий I игрока:

— отыскание равномерно оптимальных стратегий в подклассах;

— отыскание байесовских и минимаксных стратегий;

— изучение совокупности всех неулучшаемых стратегий (так называемого полного класса стратегий).

2. Равномерно оптимальные стратегии в подклассах. Применительно к задачам математической статистики часто используется следующий прием (см. § 5). Из некоторых соображений, не связанных непосредственно с потерями (соображения симметрии, естественность процедуры, простота вычислений и др.), иногда оказывается возможным сузить класс рассматриваемых стратегий. Если это сужение оказывается таким, что после него существует равномерно оптимальная стратегия, то тем самым проблема выбора стратегии решается. Этот подход необходимо сопровождать исследованиями вопроса о том, не утратили ли мы путем сужения класса стратегий возможности получения существенно лучшего результата. Примеры использования такого подхода, относящиеся, правда, к более сложному объекту — статистическим играм, будут приведены в последующих двух параграфах. Читатель уже знает о них по главам 2[6], 3[6], где рассматривались наилучшие (эффективные) оценки в подклассе несмещенных оценок и равномерно наиболее мощные критерии в подклассах всех инвариантных или несмещенных критериев.

3. Байесовские стратегии. Они возникают в тех случаях, когда второй игрок выбирает свою стратегию случайным образом с некоторым распределением (известным или неизвестным) на Θ.

Чтобы иметь возможность рассматривать в дальнейшем «случайные» стратегии, мы будем предполагать, что на Θ и D выделены некоторые естественные σ-алгебры подмножеств $\mathfrak{F}_\Theta$ и $\mathfrak{F}_D$. Тогда на $(\Theta, \mathfrak{F}_\Theta)$, $(D, \mathfrak{F}_D)$ можно определить распределения Q и π соответственно, так что $(\Theta, \mathfrak{F}_\Theta, Q)$, $(D, \mathfrak{F}_D, \pi)$ будут вероятностными пространствами.

Задание распределений π и Q индуцирует вероятностное пространство $(D\times\Theta, \mathfrak{F}_{D\times\Theta}, \pi\times Q)$, где $\mathfrak{F}_{D\times\Theta}$ есть σ-алгебра, порожденная прямыми произведениями множеств из $\mathfrak{F}_D$ и $\mathfrak{F}_\Theta$. Выбор σ-алгебр $\mathfrak{F}_D$ и $\mathfrak{F}_\Theta$ должен быть таким, чтобы выполнялись следующие два условия:

а) $\mathfrak{F}_D$ *и* $\mathfrak{F}_\Theta$ *содержат одноточечные множества* $\{\delta\}$ *и* $\{\theta\}$.

б) *Функция потерь* $w(\delta, \theta)$ *измерима относительно* $\mathfrak{F}_{D\times\Theta}$.

Определение 3. Распределения π на $(D, \mathfrak{F}_D)$ и Q на $(\Theta, \mathfrak{F}_\Theta)$ мы будем называть *смешанными* или *рандомизированными стратегиями* соответственно I и II игроков.

Распределение Q мы часто будем называть *априорным*. Смысл этого термина должен быть ясен из глав 2[6], 3[6]. Он будет пояснен дополнительно в следующем параграфе.

Множества всех смешанных стратегий I и II игроков (т. е. множества всех распределений на $(D, \mathfrak{F}_D)$ и $(\Theta, \mathfrak{F}_\Theta)$) мы обозначим $\tilde{D}$ и $\tilde{\Theta}$. Так как $\mathfrak{F}_D$ и $\mathfrak{F}_\Theta$ содержат одноточечные множества, то $\tilde{D}$ и $\tilde{\Theta}$ будут содержать распределения, сосредоточенные в одной точке, и, стало быть, мы можем считать, что $\tilde{D}$ и $\tilde{\Theta}$ содержат в себе стратегии δ и θ, которые, чтобы иметь возможность их выделить, мы будем называть *чистыми стратегиями*. Соглашение, по которому мы будем распределения из $\tilde{D}$ и $\tilde{\Theta}$, сосредоточенные в одной точке δ или θ, обозначать соответственно теми же символами δ и θ, нигде к недоразумениям не приведет.

Определим теперь потери $\tilde{w}(\pi, Q)$ при использовании смешанных стратегий равенством

$$\tilde{w}(\pi, Q) = \mathsf{M}_{\pi\times Q} w(\delta, \theta) = \int w(u, t)\, \pi(du)\, Q(dt). \qquad (2)$$

Таким образом, наряду с исходной игрой мы можем рассматривать игру $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ с функцией потерь (2), которая называется *усреднением* или *рандомизацией игры* (D, Θ, w).

В соответствии с принятым соглашением мы будем писать

$$\tilde{w}(\pi_{(\delta)}, Q) = \tilde{w}(\delta, Q), \quad \tilde{w}(\pi, Q_{(\theta)}) = \tilde{w}(\pi, \theta),$$
$$\tilde{w}(\delta, \theta) = w(\delta, \theta),$$

если $\pi_{(\delta)}$ и $Q_{(\theta)}$ есть распределения, сосредоточенные соответственно в точках δ и θ.

Очевидно, что рандомизация игры (D, Θ, w) означает переход к игре с более богатыми множествами стратегий, по отношению к которой исходная пара является «вложенной» — она получается, если рассматривать лишь чистые стратегии обоих игроков. Задачи упорядочения стратегий в играх (D, Θ, w) и $(\tilde{D}, \tilde{\Theta}, \tilde{w})$, как мы увидим, тесно связаны.

Определение 4. Стратегия $\pi = \pi_Q$, для которой

$$\tilde{w}(\pi_Q, Q) = \inf_{\pi} \tilde{w}(\pi, Q),$$

называется *байесовской, соответствующей априорному распределению* Q.

Таким образом, байесовская стратегия есть не что иное, как наилучшая стратегия π при данном Q в усредненной игре.

Стратегия $\delta_Q \in D$, для которой $\tilde{w}(\delta_Q, Q) = \inf\limits_{\pi} \tilde{w}(\pi, Q)$, называется *чистой байесовской*.

Теорема 1. *Если для данного Q существует смешанная байесовская стратегия π_Q, то существует и чистая байесовская стратегия δ_Q такая, что*

$$\tilde{w}(\delta_Q, Q) = \tilde{w}(\pi_Q, Q).$$

Доказательство почти очевидно. Обозначим $a = \tilde{w}(\pi_Q, Q)$. Ясно, что

$$\tilde{w}(\delta, Q) \geqslant \inf_{\delta} \tilde{w}(\delta, Q) \geqslant a.$$

Если допустить, что $\tilde{w}(\delta, Q) > a$ при всех δ, то производя усреднение по δ с помощью π_Q, получим

$$a = \int \tilde{w}(u, Q)\, \pi_Q(du) > a.$$

Это противоречие доказывает теорему. $\triangleleft$

Таким образом, если $\inf\limits_{\pi} \tilde{w}(\pi, Q)$ достигается, то он достигается и на чистых стратегиях.

Если $\inf\limits_{\delta} \tilde{w}(\delta, Q)$ не достигается, то байесовские стратегии не существуют. В этом случае полезным оказывается понятие *ε-байесовской стратегии*, которая существует всегда и которая определяется как стратегия δ_Q, для которой

$$\tilde{w}(\delta_Q, Q) \leqslant \inf_{\delta} \tilde{w}(\delta, Q) + \varepsilon \tag{3}$$

для заданного $\varepsilon > 0$. Однако в дальнейшем для упрощения изложения мы будем обычно ограничиваться теми задачами, в которых байесовские стратегии существуют.

Вопрос о практическом использовании байесовских стратегий является довольно тонким. Если наличие априорного распределения обусловлено некоторым реальным физическим механизмом, то этот подход бесспорен. Но байесовский подход можно оправдывать и в тех случаях, когда его связывают с наличием некоторых, быть может, даже субъективных и не всегда достаточно полных представлений, которые тем не менее сбрасывать со счета нельзя. Более подробное обсуждение вопроса об использовании байесовского подхода см. в п. 4.

4. **Минимаксные стратегии.** Если априорная информация относительно θ отсутствует, то при упорядочении стратегий можно ориентироваться на «наихудшую» стратегию противника. Если мы выберем стратегию δ, то максимальные потери составят

$$\sup_{\theta} w(\delta, \theta) \doteq w(\delta, \uparrow). \tag{4}$$

Это количество зависит лишь от δ и так же, как и значения $w(\delta, Q)$, позволяет упорядочить δ.

Определение 5. Стратегия $\bar{\delta}$ называется *минимаксной*, если

$$w(\bar{\delta}, \uparrow) = \inf_{\delta} w(\delta, \uparrow) \equiv w^*. \tag{5}$$

Термин «минимаксный» образован из соединения наименований операций в правой части соотношения

$$w(\bar{\delta}, \uparrow) = \min_{\delta} \max_{\theta} w(\delta, \theta).$$

Очевидно, что минимаксные стратегии, так же как и байесовские, могут, вообще говоря, не существовать. В этом случае аналогично (3) можно ввести понятие ε-*минимаксной стратегии*. В последующих рассмотрениях мы будем исходить из того, что sup и inf в (4), (5) достигаются.

Так как при любом θ

$$w(\bar{\delta}, \theta) \leqslant w(\bar{\delta}, \uparrow) = w^*,$$

то минимаксная стратегия $\bar{\delta}$ характеризуется тем, что обеспечивает потери I игрока в размере, не большем чем w^*.

Определение 6. Значения

$$w^* = \inf_\delta w(\delta, \uparrow) \quad \left(w(\delta, \uparrow) = \sup_\theta w(\delta, \theta)\right),$$
$$w_* = \sup_\theta w(\downarrow, \delta) \quad \left(w(\downarrow, \theta) = \inf_\delta w(\delta, \theta)\right)$$

называются соответственно *верхней* и *нижней ценой игры*. Если $w^* = w_*$, то говорят, что *существует цена игры*, равная общему значению w^* и w_*.

Из сказанного выше и из соображений симметрии ясно, что II игрок, действуя аналогично первому и выбирая свою стратегию $\overline{\theta}$ из тех же минимаксных соображений, может всегда обеспечить себе выигрыш не меньше, чем $\widetilde{w}_*$. (Такую стратегию $\overline{\theta}$ было бы правильнее называть максиминной, но мы будем использовать для нее тот же термин: минимаксная стратегия.) Стало быть, *если существует цена игры*, то, выбирая минимаксную стратегию $\overline{\delta}$, мы обеспечим себе *неулучшаемый результат* в том смысле, что если противник выберет $\overline{\theta}$, то никакая другая стратегия не даст нам потери, меньшие чем $w_* = w^*$. Очевидно, что $w(\overline{\delta}, \overline{\theta}) = w^* = w_*$.

В общем случае всегда $w^* \geqslant w_*$, так как при всех δ и θ $w(\delta, \uparrow) \geqslant w(\delta, \theta) \geqslant w(\downarrow, \theta)$ и, следовательно,

$$w^* = \inf_\delta w(\delta, \uparrow) \geqslant \sup_\theta w(\downarrow, \theta) = w_*. \tag{6}$$

Если $w^* > w_*$, то минимаксную стратегию $\overline{\delta}$ *можно улучшить*, вводя смешанные стратегии. В этом состоит одно из главных назначений последних.

Минимаксные стратегии для усредненной игры (если они существуют) мы обозначим соответственно $\overline{\pi}$ и $\overline{Q}$ и положим

$$\widetilde{w}^* = \inf_\pi \sup_Q \widetilde{w}(\pi, Q), \quad \widetilde{w}_* = \sup_Q \inf_\pi \widetilde{w}(\pi, Q).$$

Мы покажем сначала, что при усреднении игры верхняя и нижняя цены игры сближаются.

Теорема 2. $w^* \geqslant \widetilde{w}^* \geqslant \widetilde{w}_* \geqslant w_*$.

Доказательство этой теоремы, как и теоремы 1, очень просто. Так как усреднение игры можно производить в два этапа: сначала по множеству D, а затем по Θ, то для доказательства достаточно рассмотреть лишь частичное усреднение $(\widetilde{D}, \Theta, \widetilde{w})$ игры (D, Θ, w). Имеем

$$\widetilde{w}^* = \inf_{\pi} \sup_{\theta} \widetilde{w}(\pi, \theta) \leqslant \inf_{\delta} \sup_{\theta} w(\delta, \theta) = w^*.$$

Так как при всех π

$$\widetilde{w}(\pi, \theta) = \int w(u, \theta)\, \pi(du) \geqslant \inf_{\delta} w(\delta, \theta) = w(\downarrow, \theta),$$

то $\inf\limits_{\pi} \widetilde{w}(\pi, \theta) \geqslant w(\downarrow, \theta)$,

$$\widetilde{w}_* = \sup_{\theta} \inf_{\pi} \widetilde{w}(\pi, \theta) \geqslant \sup_{\theta} w(\downarrow, \theta) = w_*.$$

Неравенство $\widetilde{w}^* \geqslant \widetilde{w}_*$ доказано в (6). ◁

Фундаментальным фактом теории игр является так называемая теорема о минимаксе, которая утверждает, что при весьма широких предположениях *усредненные игры имеют цену* $\widetilde{w}^* = \widetilde{w}_*$ *и для них существуют минимаксные стратегии.*

Более точно это утверждение будет сформулировано в следующем параграфе в более общей ситуации применительно к статистическим играм.

Исходная игра (D, Θ, w), особенно в случае, когда D и Θ конечны, цены, как правило, не имеет.

Пример 1. Рассмотрим простейшую игру, когда множества D и Θ двухточечные, $D = \{\delta_1, \delta_2\}$, $\Theta = \{\theta_1, \theta_2\}$. Значения функции потерь $w(\delta, \theta)$ определяются матрицей $\|w(\delta_i, \theta_j)\|$, $i, j = 1, 2$, которую мы положим равной $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix}$. Это соответствует, например, игре с угадыванием, когда I игрок угадывает, в какой руке II игрок спрятал монету. Угадывание означает нулевой проигрыш ($w(\delta_1, \theta_1) = w(\delta_2, \theta_2) = 0$), ошибка — проигрыш в размере 1 рубль ($w(\delta_1, \theta_2) = w(\delta_2, \theta_1) = 1$). Очевидно, что здесь $w(\delta_i, \uparrow) = 1$, $w^* = 1$, $w(\downarrow, \theta_i) = 0$, $w_* = 0$, так что игра не имеет цены, а I игрок не может себе гарантировать проигрыш меньше чем 1 рубль. Само понятие минимаксной стратегии здесь бесполезно.

Рассмотрим теперь усреднение этой игры. Здесь классы стратегий $\widetilde{D}$ и $\widetilde{\Theta}$ представляют собой совокупность всех распределений на двухточечном множестве. Очевидно, каждое из распределений на D и Θ описывается одной вероятностью p и q соответственно выбора стратегий δ_1 и θ_1. Поэтому можно считать, что $\widetilde{D} = [0, 1]$,

$\widetilde{\Theta} = [0, 1]$. Потери I игрока в этой игре равны

$$\widetilde{w}(p, q) = p(1-q) + q(1-p) = p + q - 2pq,$$

$$\widetilde{w}(p, \hat{q}) = \begin{cases} p + 1 - 2p = 1 - p & \text{при } 2p < 1, \\ p & \text{при } 2p \geqslant 1, \end{cases}$$

$$\widetilde{w}^* = 1/2.$$

Аналогично находим, что $\widetilde{w}_* = 1/2$. Таким образом, усредненная игра уже имеет цену, и первый игрок, выбирая δ_1 и δ_2 с вероятностью $p = 1 - p = 1/2$, может гарантировать себе проигрыш не больше 1/2. Улучшить эту стратегию нельзя, так как такой же выигрыш может гарантировать себе II игрок, выбирая $q = 1/2$.

Если все же окажется, что усредненная игра не имеет цены (это может быть только в играх специальной сложной конструкции), то повторное усреднение не даст никаких результатов, поскольку это повторное усреднение по существу будет совпадать с одинарным.

Байесовский и минимаксный подходы к решению игровых проблем весьма распространены в повседневной человеческой деятельности. Байесовский подход ориентирован на наличие некоторых представлений, хотя бы приближенных, о поведении второго игрока. Минимаксный подход оправдан в тех случаях, когда необходимо гарантировать себя от большого проигрыша.

Пример 2. Студент готовится к экзамену. Будем считать, что это не идеальный студент и что у него не оказалось достаточно времени для того, чтобы хорошо подготовить к сдаче весь материал. Кроме того, целью этого студента является получение возможно более высокой оценки.

В описанных условиях студент может выучить на отлично лишь часть материала. Поэтому для него возможны по крайней мере два пути: 1) выучить на отлично, но лишь те разделы, которые по имеющимся сведениям чаще всего спрашивает экзаменатор; 2) выучить все понемногу, чтобы гарантировать себе хорошую или удовлетворительную оценку. Первый вариант будет соответствовать байесовскому, второй — минимаксному подходу.

Ясно, что равномерно оптимальной стратегией здесь было бы выучить на отлично весь материал, но по условию задачи она невозможна.

Минимаксные стратегии в конкретных ситуациях выглядят не всегда разумными.

Пример 3. Пусть $\Theta = [0, 1]$, а множество $D = \{\delta_1, \delta_2\}$ состоит из двух элементов. Функция потерь определяется соотношениями (рис. 4)

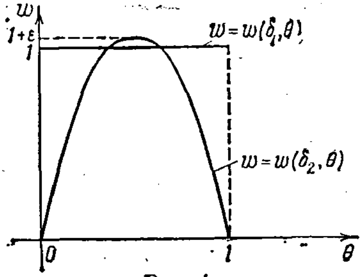

Рис. 4

$$w(\delta_1, \theta) \equiv 1,$$
$$w(\delta_2, \theta) = 4(1+\varepsilon)\theta(1-\theta).$$

Здесь $w(\delta_1, \uparrow) = 1$, $w(\delta_2, \uparrow) = 1 + \varepsilon$, $w^* = 1$, и минимаксной стратегией будет δ_1, хотя при малых $\varepsilon > 0$ для «большинства» значений θ стратегия δ_2 будет лучше: $w(\delta_2, \theta) < 1$ при θ из области $\left|\theta - \frac{1}{2}\right| > \frac{1}{2}\sqrt{\frac{\varepsilon}{1+\varepsilon}}$. Для «большинства» распределений Q на $\Theta = [0, 1]$ (чья масса не сосредоточена в окрестности точки $\theta = 1/2$) байесовские стратегии также будут совпадать с δ_2.

Понятия байесовской и минимаксной стратегии связаны между собой. Следующее утверждение дает способ отыскания минимаксных стратегий с помощью байесовских.

Определение 7. Стратегия $\bar{\pi}$ называется *уравнивающей* на множестве $\Theta_0 \subset \Theta$, если

1) $\tilde{w}(\bar{\pi}, \theta) = c = \mathrm{const}$, $\theta \in \Theta_0$,

2) $\tilde{w}(\bar{\pi}, \theta) \leqslant c$ при всех θ.

Теорема 3. *Пусть существуют априорное распределение $\bar{Q}$ и соответствующая ему байесовская стратегия $\pi_{\bar{Q}}$, которая является уравнивающей на носителе $N_{\bar{Q}}$ распределения $\bar{Q}$. Тогда $\bar{\pi} = \pi_{\bar{Q}}$ является минимаксной стратегией.*

Если $N_{\bar{Q}} = \Theta$, то уравнивающая стратегия $\bar{\pi}$ делает игру второго игрока «безразличной» — от него не зависящей (ср. с примером 1).

Доказательство теоремы 3. Обозначим $\sup\limits_{\theta} \tilde{w}(\pi, \theta) = \tilde{w}(\pi, \uparrow)$, $\inf\limits_{\delta} \tilde{w}(\delta, Q) = \tilde{w}(\downarrow, Q)$. Нам надо убедиться, что

$$\tilde{w}(\pi_{\bar{Q}}, \uparrow) = \inf_{\pi} \tilde{w}(\pi, \uparrow).$$

Это вытекает из следующих неравенств, справедливых для любого π:

$$\tilde{w}(\pi, \uparrow) \geqslant \tilde{w}(\pi, \bar{Q}) \geqslant \tilde{w}(\pi_{\bar{Q}}, \bar{Q}) =$$
$$= \int \tilde{w}(\pi_{\bar{Q}}, t)\,\bar{Q}(dt) = c \geqslant \tilde{w}(\pi_{\bar{Q}}, \uparrow). \quad \triangleleft$$

Иногда бывает полезным следующее небольшое обобщение теоремы 3.

Теорема 3А. *Пусть существуют последовательности* Q_n, π_{Q_n} *такие, что* $\tilde{w}(\pi_{Q_n}, Q_n) \to c$. *Пусть, кроме того, существует стратегия* $\bar{\pi}$, *обладающая свойством* $w(\bar{\pi}, \theta) \leqslant \leqslant c$ *при всех* θ. *Тогда* $\bar{\pi}$ — *минимаксная стратегия.*

Доказательство столь же просто:

$$\tilde{w}(\pi, \uparrow) \geqslant \tilde{w}(\pi, Q_n) \geqslant \tilde{w}(\pi_{Q_n}, Q_n) \to c.$$

Это может быть только тогда, когда $\inf\limits_{\pi} \tilde{w}(\pi, \uparrow) \geqslant c$. Так как $c \geqslant w(\bar{\pi}, \uparrow)$, то теорема доказана.

Распределение $\bar{Q}$ в теореме 3, определяющее байесовскую минимаксную стратегию $\pi_{\bar{Q}}$, обладает одним замечательным свойством: оно будет наихудшим в том смысле, что для него байесовские потери $\tilde{w}(\pi_Q, Q)$ будут максимальными.

Определение 8. Распределение $\bar{Q}$ называется *наименее благоприятным* или *наихудшим*, если

$$\tilde{w}(\pi_{\bar{Q}}, \bar{Q}) = \sup_{Q} \tilde{w}(\pi_Q, Q),$$

или, другими словами, $\tilde{w}(\downarrow, \bar{Q}) = \sup\limits_{Q} \tilde{w}(\downarrow, Q)$.

Теорема 4. *Пусть игра* $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ *имеет цену, а оба игрока имеют минимаксные стратегии* $\bar{\pi}$ *и* $\bar{Q}$. *Тогда распределение* $\bar{Q}$ *является наихудшим, а* $\bar{\pi}$ *является байесовской стратегией* $\bar{\pi} = \pi_{\bar{Q}}$, *отвечающей* $\bar{Q}$.

Замечание 1. Из того, что в силу теоремы 1 наряду с $\pi_{\bar{Q}}$ существует чистая байесовская стратегия $\delta_{\bar{Q}}$, вовсе не следует, что последняя также будет минимаксной.

Замечание 2. В силу фундаментальной теоремы о минимаксах условие теоремы 4 о существовании цены усредненной игры и минимаксных стратегий не следует рассматривать как существенное ограничение.

Нам понадобится следующее вспомогательное утверждение, которое мы сформулируем в терминах исходной (не усредненной) игры.

Лемма 1. *Пусть игра* (D, Θ, w) *имеет цену и минимаксные стратегии* $\bar{\delta}$, $\bar{\theta}$ *обоих игроков*:

$$w(\bar{\delta}, \uparrow) = \inf_{\delta} w(\delta, \uparrow), \quad w(\downarrow, \bar{\theta}) = \sup_{\theta} w(\downarrow, \theta).$$

Тогда

$$w(\bar{\delta}, \uparrow) = w(\bar{\delta}, \bar{\theta}) = w(\downarrow, \bar{\theta}), \tag{7}$$

$$w^* = w(\bar{\delta}, \bar{\theta}) = w_*. \tag{8}$$

Обратно, если при некоторых $\bar{\delta}$, $\bar{\theta}$ *выполнено* (7), *то справедливо* (8), *а* $\bar{\delta}$, $\bar{\theta}$ *являются минимаксными стратегиями.*

Доказательство. При всех δ и θ имеем

$$w(\delta, \uparrow) \geqslant w(\delta, \theta) \geqslant w(\downarrow, \theta).$$

Отсюда следует, что

$$w^* = w(\bar{\delta}, \uparrow) \geqslant w(\bar{\delta}, \bar{\theta}) \geqslant w(\downarrow, \bar{\theta}) = w_*. \tag{9}$$

Так как по условию $w^* = w_*$, то в (9) все знаки неравенства надо заменить равенствами. Это доказывает (7), (8).

Обратно, если справедливо (7), то

$$w^* = \inf_{\delta} w(\delta, \uparrow) \leqslant w(\bar{\delta}, \uparrow) = w(\downarrow, \bar{\theta}) \leqslant \sup_{\theta} w(\downarrow, \theta) = w_*.$$

Так как всегда $w^* \geqslant w_*$, то приведенные неравенства означают, что $w^* = w_*$, а стратегии $\bar{\delta}$, $\bar{\theta}$ являются минимаксными. Лемма доказана.

Точка $(\bar{\delta}, \bar{\theta})$, обладающая свойством (7), называется *седловой точкой*, а лемма 1 — критерием наличия седловой точки неулучшаемых минимаксных стратегий.

Доказательство теоремы 4. Применим лемму 1 к усредненной игре $(\tilde{D}, \tilde{\Theta}, \tilde{w})$. Мы получим тогда, что

$$\tilde{w}(\bar{\pi}, \bar{Q}) = \tilde{w}(\downarrow, \bar{Q}) = \tilde{w}_* = \sup_{Q} \tilde{w}(\downarrow, Q).$$

Отсюда следует, что распределение $\bar{Q}$ является наихудшим и что $\bar{\pi}$ есть байесовская стратегия, соответствующая $\bar{Q}$. Теорема доказана.

Содержание приведенных выше утверждений можно резюмировать теперь в виде следующего критерия минимаксности, весьма полно описывающего связь минимаксных и байесовских стратегий.

Теорема 5. *Пусть игра* $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ *имеет цену и минимаксные стратегии. Тогда следующие три условия эквивалентны:*

1) *Стратегия* $\bar{\pi}$ *является минимаксной.*

2) *Стратегия* $\bar{\pi}$ *является байесовской и уравнивающей.*

3) *Стратегия* $\bar{\pi}$ *является байесовской и соответствует наихудшему распределению* $\bar{Q}$: $\bar{\pi} = \pi_{\bar{Q}}$.

Доказательство. Соотношение 2) $\Rightarrow$ 1) доказано в теореме 3 (условие теоремы 5 для этого не требуется). Соотношение 1) $\Rightarrow$ 3) установлено в теореме 4. Нам осталось убедиться в том, что 3) $\Rightarrow$ 2), т. е. что байесовская стратегия, соответствующая наихудшему распределению, является уравнивающей. Имеем

$$\tilde{w}_* = \tilde{w}(\bar{\pi}, \bar{Q}) = \int \tilde{w}(\bar{\pi}, t)\bar{Q}(dt) \leqslant \sup_t \tilde{w}(\bar{\pi}, t) = \tilde{w}^*.$$

Это означает, что $\int \tilde{w}(\bar{\pi}, t)\bar{Q}(dt) = \sup_t \tilde{w}(\bar{\pi}, t)$ и, стало быть,

$$\tilde{w}(\bar{\pi}, t) = \tilde{w}(\bar{\pi}, \mathfrak{t}) \text{ п. в. } [\bar{Q}].$$

Так как, кроме того, всегда $\tilde{w}(\bar{\pi}, t) \leqslant \tilde{w}(\bar{\pi}, \mathfrak{t})$, то $\bar{\pi}$ является уравнивающей стратегией. Теорема доказана.

Вернемся теперь к вопросу о применении рассматриваемых классов стратегий. Допустим, что мы не можем выделить удовлетворяющий нас подкласс стратегий, среди которых существовала бы равномерно наилучшая. Допустим, далее, что мы располагаем некоторыми представлениями о поведении второго игрока (т. е. об ожидаемых значениях θ), недостаточными, однако, для применения байесовского подхода в его чистом виде. Минимаксный подход в этих условиях будет означать пренебрежение имеющейся у нас информацией. В такой ситуации можно использовать промежуточный подход, который выглядит следующим образом:

1. Сначала нужно оградить себя от высоких потерь, т. е. рассматривать лишь те стратегии δ, для которых $w(\delta, \theta) \leqslant w^* + a$ при подходящем значении $a > 0$ и при всех θ. Множество стратегий, удовлетворяющих этому неравенству, обозначим D_a.

2. В этом подмножестве (т. е. в игре (D_a, Θ, w)) уже можно применять байесовский подход с использованием доступных нам приближений для априорного распределения Q.

Такой смешанный подход также постоянно используется в повседневной человеческой деятельности. В условиях примера 2 этот подход будет означать, что студент совсем немного (чтобы избежать неудовлетворительной оценки) выучит весь материал, а затем получше выучит то, что чаще спрашивают.

Математически использование смешанного подхода должно сопровождаться исследованиями устойчивости байесовских потерь в игре (D_a, Θ, w) при допустимых изменениях Q.

5. **Полный класс стратегий.** Если все описанные выше подходы не дают возможность однозначно выбрать стратегию, то решение задачи ограничивают описанием так называемого полного класса стратегий.

Определение 9. Класс стратегий $D^c \subset \tilde{D}$ называется *полным*, если для любого $\pi \notin D^c$ существует стратегия $\pi_0 \in D^c$, которая лучше, чем π.

Класс D_0^c называется *минимальным полным классом*, если D_0^c есть полный класс, но никакой его собственный подкласс полным классом не является.

Другими словами, минимальный полный класс состоит из одних неулучшаемых стратегий.

Польза построения минимального полного класса или полного класса, который существенно меньше, чем D, очевидна. Это позволяет редуцировать игру $(\tilde{D}, \tilde{\Theta}, \tilde{w})$ к игре $(D^c, \tilde{\Theta}, \tilde{w})$, которая может иметь более простую структуру.

Вторая фундаментальная теорема теории игр состоит в том, что при широких предположениях *класс всех байесовских стратегий* $\{\pi_Q\}$, $Q \in \tilde{\Theta}$, *является полным классом.* Точная формулировка этой теоремы будет приведена в следующем параграфе. В некоторых случаях полные классы можно строить и непосредственно, используя структуру игры. Допустим, например, что существует разбиение пространства D на подмножества D_b, $D = \bigcup_{b \in B} D_b$, $D_{b_1} \neq D_{b_2}$ при $b_1 \neq b_2$, такое, что в каждом из этих подмножеств (т. е. для игр (D_b, Θ, w)) существует равномерно оптимальная стратегия $\delta_b \in D_b$. Ясно, что в этом случае класс $D^c = \{\delta_b\}_{b \in B}$ будет полным. Такой подход к построению полного класса будет проиллюстрирован в § 3.

§ 3. Статистические игры

1. Описание статистических игр. Основные элементы статистической игры образуются той же тройкой (D, Θ, w), которая рассматривалась нами в предыдущем параграфе. Однако к ним добавляется следующее:

1) В статистических играх роль I игрока играет *статистик* (исследователь), роль II игрока — *природа* (точнее, природа того явления, которое исследуется). Последняя выбирает (или «загадывает») параметр (стратегию) θ, который нам неизвестен и который определяет состояние исследуемого объекта. Большинство задач математической статистики так или иначе связано с принятием таких решений δ, которые как можно более точно «угадывали» бы это неизвестное θ. При этом следует иметь в виду, что природа как игрок не стремится к максимальному выигрышу (т. е. не стремится причинить нам максимальные потери) и в этом смысле является игроком, «беспристрастным» к выбору своих стратегий.

2) В статистических играх мы имеем возможность «разведывать» стратегию природы с помощью экспериментов, которые дают нам в виде выборки $X \in \mathbf{P}_\theta$ «наводящие» указания на то, каким может быть значение θ. Итак, элементом статистической игры является выборка X объема*) n из распределения $\mathbf{P}_\theta$, зависящего от θ.

В этих условиях наше решение δ мы должны выбирать, очевидно, в зависимости от X. Стало быть, стратегиями статистика становятся теперь все функции $\delta(X)$,

---

*) В построениях этого параграфа мы могли бы, не ограничивая общности, считать, что $n = 1$. Однако мы сохраним понятие выборки объема n, с тем чтобы сохранить простые связи с результатами [6] и с последующими рассмотрениями (§§ 6—8).

Более общая концепция статистической игры имеет дело с неограниченной выборкой $X_\infty = (\mathrm{x}_1, \mathrm{x}_2, \ldots)$, использование элемента x_n которой сопряжено с затратами $c_n \geqslant 0$ (см. [20]).

отображающие $\mathscr{X}^n$ в D. Эти функции $\delta(X)$ называются *решающими функциями*, или *решающими правилами*, или просто *решениями*. Мы ограничимся рассмотрением лишь функций $\delta(X)$, осуществляющих *измеримое* отображение $(\mathscr{X}^n, \mathfrak{B}^n_{\mathscr{X}})$ в $(D, \mathfrak{F}_D)$. Множество всех таких функций мы обозначим через $\mathscr{D}$.

Множество стратегий II игрока (природы) Θ остается прежним.

Если мы воспользуемся решением $\delta(X)$, а природа выберет θ, то наши потери составят $w(\delta(X), \theta)$. Это есть случайная величина. Чтобы избежать этого неудобства, естественно в качестве потерь для стратегий $\delta = \delta(\cdot) \in \mathscr{D}$ и $\theta \in \Theta$ принять значение математического ожидания

$$W(\delta(\cdot), \theta) = \mathsf{M}_\theta w(\delta(X), \theta) = \int w(\delta(x), \theta)\, \mathbf{P}_\theta(dx), \qquad (1)$$

которое называется *функцией риска* (появление слова «риск» здесь естественно, так как применение $\delta(\cdot)$ дает случайный результат). Если выполнено условие (A_μ) о наличии плотности $f_\theta(x)$ распределения $\mathbf{P}_\theta$ относительно некоторой σ-конечной меры μ, то функцию риска можно записать в виде

$$W(\delta(\cdot), \theta) = \int w(\delta(x), \theta) f_\theta(x) \mu^n(dx).$$

Мы можем дать теперь следующее

Определение 1. *Статистической игрой* называется тройка $(\mathscr{D}, \Theta, W)$, где Θ есть множество стратегий природы, $\mathscr{D}$ — множество всех измеримых отображений пространства $\mathscr{X}^n$ в множество D, W определено в (1). Для более полной характеризации статистической игры вместе с тройкой $(\mathscr{D}, \Theta, W)$ можно считать также заданной пару $(X, \mathbf{P}_\theta)$, где $X \in\!\!\!= \mathbf{P}_\theta$.

Пример 1. Пусть $\theta \in [0, 1]$ определяет содержание некоторой химической компоненты руды, приготовленной для плавки. Если мы примем решение, что доля этой компоненты равна $\delta \neq \theta$, и в соответствии с этим решением будет организован весь процесс плавки, то в результате качество выплавленного металла будет хуже, чем при $\delta = \theta$, а расходы энергии выше. Другими словами, мы будем нести потери $w(\delta, \theta)$, которые будут тем больше, чем больше δ отличается от θ. Предположим для простоты, что $w(\delta, \theta)$ пропорциональна квадрату отклонения δ от θ:

$$w(\delta, \theta) = c(\delta - \theta)^2.$$

(Если функция $w(\delta, \theta)$ гладкая и если рассматривать окрестность прямой $\delta = \theta$, то упрощающим предположением здесь будет лишь независимость c от θ.) В результате мы получим игру (D, Θ, w), у которой $D = [0, 1]$, $\Theta = [0, 1]$,

$$w(\delta, \uparrow) = \sup_\theta w(\delta, \theta) = \begin{cases} c\delta^2 & \text{при } \delta > 1/2, \\ c(1-\delta)^2 & \text{при } \delta \leqslant 1/2, \end{cases}$$

$$w^* = \inf_\delta w(\delta, \uparrow) = w(1/2, \uparrow) = c/4.$$

Таким образом, стратегия $\delta = 1/2$ является минимаксной и гарантирует потери $\leqslant c/4$. Так как $w_* = 0$, то эта игра цены не имеет. Рандомизация игры минимаксную стратегию $\delta = 1/2$ не улучшает (она дает $\tilde{w}_* = c/4$). Мы предоставляем читателю убедиться, что байесовская стратегия δ_Q здесь имеет вид $\delta_Q = \mathsf{M}_Q\theta = \int t\mathrm{Q}(dt)$ (это вытекает из равенств $\tilde{w}(\delta, \mathrm{Q}) = c\mathsf{M}_Q(\delta - \theta)^2 = c\mathsf{M}_Q(\theta - \mathsf{M}_Q\theta)^2 + c\mathsf{M}_Q(\delta - \mathsf{M}_Q\theta)^2)$ и что наихудшее распределение $\bar{\mathrm{Q}}$ будет иметь вид $\bar{\mathrm{Q}}(\{0\}) = \bar{\mathrm{Q}}(\{1\}) = 1/2$. Очевидно, что соответствующая ему байесовская стратегия есть $\delta_{\bar{Q}} = 1/2$.

Предположим теперь, что руда неоднородна и что мы имеем возможность брать n проб руды. Эти пробы организованы так, что результаты лабораторных анализов для содержания упомянутой компоненты в пробах будут случайны и дадут нам независимые значения $(\mathrm{x}_1, \ldots, \mathrm{x}_n) = X$, относительно которых известно, что $\mathsf{M}\mathrm{x}_i = \theta$, $\mathsf{D}\mathrm{x}_i = b(\theta)$. В этом случае решениями $\delta(X)$ будут всевозможные оценки $\theta^* = \delta(X)$ параметра θ по выборке X. Риск решающей функции $\delta(X)$ будет равен

$$W(\delta, \theta) = c\mathsf{M}_\theta(\delta(X) - \theta)^2,$$

и мы приходим к задаче отыскания оценки $\theta^* = \delta(X)$, минимизирующей в том или ином смысле этот риск. Если положить, например, $\delta_1(X) = \bar{\mathrm{x}}$, то мы получим

$$W(\delta_1, \theta) = \frac{cb(\theta)}{n}. \tag{2}$$

Наибольшее значение $b(\theta)$ равно $\theta(1-\theta)$ и достигается на распределении x_1, сосредоточенном в точках 0 и 1.

Так как такую возможность можно исключить, то

$$b(\theta) < \theta(1-\theta) \leqslant 1/4, \quad W(\delta_1, \theta) < c/4n.$$

Таким образом, даже при $n=1$ и при использовании, возможно, не наилучшей стратегии мы получим результат, лучший чем для минимаксной стратегии в игре без выборки. Соотношение (2) показывает также, что риск сходится к нулю при $n \to \infty$. ◁

Из приведенного выше определения статистической игры видно, что последняя обладает значительно более богатым множеством стратегий $\mathscr{D}$ по сравнению с исходной игрой (D, Θ, w).

Как и в § 2, наряду с игрой $(\mathscr{D}, \Theta, W)$, стратегии которой мы будем называть *чистыми*, можно рассматривать *рандомизированные* или *смешанные игры* $(\widetilde{\mathscr{D}}, \widetilde{\Theta}, \widetilde{W})$. Здесь множество $\widetilde{\mathscr{D}}$ является множеством отображений $\pi(X)$: $\mathscr{X}^n \to \widetilde{D}$. Эти отображения должны быть такими, чтобы значения

$$\widetilde{w}(\pi(X), \theta) = \int\limits_D w(u, \theta)\, \pi(X, du)$$

были случайными величинами ($\pi(X, A)$ есть вероятность множества $A \subset D$ в соответствии с решающим правилом π). Тогда по определению полагаем

$$\widetilde{W}(\pi(\cdot), Q) = \int\limits_\Theta \int\limits_{\mathscr{X}^n} \int\limits_D w(u, t)\, \pi(x, du)\, \mathbf{P}_t(dx)\, Q(dt).$$

Стратегия $\pi(X)$ называется *рандомизированным решающим правилом*.

Отношения частичного порядка между стратегиями, равномерно наилучшие стратегии, байесовские и минимаксные стратегии, полные классы стратегий для статистических игр определяются точно так же, как для игр обычных (с заменой множества D на $\mathscr{D}$, а функций w, $\widetilde{w}$ — на W, $\widetilde{W}$).

На статистические игры полностью переносятся утверждения теорем 2.1—2.5, так как последние с природой множества D никак не связаны.

2. Классификация статистических игр. С природой множеств D и Θ связана следующая классификация, выделяющая основные виды статистических игр:

1) Если $\Theta = A$, $D = A$, где A есть «телесное» подмножество в R^k (например, параллелепипед), $w(t, t) = 0$, $w(t, u) > 0$ при $t \neq u$, то мы получим задачи теории точечного оценивания неизвестного параметра θ.

2) Если множества $\Theta = \{\theta_1, \ldots, \theta_r\}$, $D = \{\delta_1, \ldots, \delta_r\}$ конечны и содержат одинаковое число элементов, $w(\delta_i, \theta_i) = 0$, $w(\delta_i, \theta_j) > 0$ при $i \neq j$, то мы получим задачи проверки конечного числа простых гипотез.

3) Если Θ есть «телесная» область в R^k, $D = \{\delta_1, \delta_2\}$ состоит из двух элементов, $w(\delta_1, \theta) = 0$ при $\theta \in \Theta_1$, $w(\delta_2, \theta) = 0$ при $\theta \in \Theta_2$ ($\Theta_1 \cap \Theta_2$ пусто) и $w(\delta_i, \theta) > 0$ в остальных случаях, то мы приходим к задаче проверки гипотез $\{\theta \in \Theta_1\}$ и $\{\theta \in \Theta_2\}$.

Возможны, конечно, и другие классы задач. Мы выделили эти три типа, поскольку они рассматривались нами в гл. 2[6], 3[6]. Мы рассматривали эти задачи с чисто «статистических» позиций, что соответствует специальному выбору функций $w(\delta, \theta)$: в первой группе задач потери определялись среднеквадратическим уклонением, что соответствует функции потерь $w(\delta, \theta) = (\delta - \theta)^2$; во второй группе задач потери определялись вероятностью ошибиться, что соответствует функции

$$w(\delta_i, \theta_j) = \begin{cases} 0, & i = j, \\ 1, & i \neq j. \end{cases}$$

То же относится и к третьей группе задач, в которой мы использовали функцию потерь

$$w(\delta_1, \theta) = \begin{cases} 0 \text{ при } \theta \in \Theta_1, \\ 1 \text{ при } \theta \in \Theta_2, \end{cases}$$

$$w(\delta_2, \theta) = \begin{cases} 1 \text{ при } \theta \in \Theta_1, \\ 0 \text{ при } \theta \in \Theta_2. \end{cases}$$

Эти функции потерь, соответствующие чисто статистическому подходу к проблемам, мы будем называть *статистическими*.

Приведенная классификация показывает, что никакой принципиальной разницы между задачами теории оценивания и проверкой статистических гипотез нет. Все дело лишь в природе множеств Θ и D и виде функций потерь.

На примере этой классификации можно отметить еще одно своеобразие статистических игр (в дополнение к пп. 1), 2) в начале этого параграфа); оно состоит в том, что множество D в статистических играх либо совпадает с Θ, либо представляет собой множество, более бедное, чем Θ.

3. **Две фундаментальные теоремы теории статистических игр.** Сформулируем теперь основные результаты теории статистических игр. Мы уже отмечали, что утверждения теорем 2.1 — 2.5 для статистических игр полностью сохраняются, так как они с природой игр не связаны. Чтобы получить две фундаментальные теоремы, упоминавшиеся в § 2, введем некоторые предположения. Это далеко не самые общие предположения (иначе формулировки и доказательства чрезвычайно усложнились бы), но они достаточно широкие, чтобы охватить наиболее интересный и содержательный круг задач и, в частности, задачи, рассмотренные в гл. 2[6], 3[6].

Условие (А). *Каждое из множеств Θ и D либо конечно, либо представляет из себя компактное множество в R^k.*

Как уже отмечалось, случай, когда Θ конечно, а $D \subset R^k$, можно не рассматривать. В остальных трех случаях мы будем предполагать, что функция потерь $w(\delta, \theta)$ удовлетворяет следующему условию.

Условие (В).

1) *Если $D \subset R^k$, $\Theta \subset R^k$, то функция $w(\delta, \theta)$ непрерывна на $D \times \Theta$.*

2) *Если $\Theta \subset R^k$, а $D = \{\delta_1, \ldots, \delta_r\}$ конечно, то каждая из r функций $w(\delta_i, \theta)$, $i = 1, \ldots, r$, непрерывна на Θ.*

Если $\Theta = \{\theta_1, \ldots, \theta_r\}$ и $D = \{\delta_1, \ldots, \delta_r\}$ конечны, то значения $w(\delta_i, \theta_j)$, $i, j = 1, \ldots, r$, могут быть произвольными.

Кроме того, мы потребуем, чтобы выполнялось

Условие (С). *Мы располагаем выборкой $X \in \mathbf{P}_\theta$ из распределения $\mathbf{P}_\theta$, абсолютно непрерывного при всех θ относительно некоторой σ-конечной меры μ. Если $\Theta \subset R^k$, то плотность $\frac{d\mathbf{P}_\theta}{d\mu}(x) = f_\theta(x)$ непрерывна в $L_1(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mu)$ относительно θ, т. е. при $\theta_m \to \theta$*

$$\int \left| f_{\theta_m}(x) - f_\theta(x) \right| \mu(dx) \to 0. \qquad (3)$$

Нетрудно проверить, что обычная непрерывность $f_\theta(x)$ по θ для $[\mu]$ п. в. x влечет за собой непрерывность (3).

Теорема 1. *Если выполнены условия* (A), (B), (C), *то усредненная игра* $(\widetilde{\mathscr{D}}, \widetilde{\Theta}, \widetilde{W})$ *имеет цену и минимаксные стратегии* $\bar{\pi}(X)$ *и* $\bar{\mathbf{Q}}$:

$$\widetilde{W}(\bar{\pi}(\cdot), \uparrow) \doteq \inf_{\pi} \widetilde{W}(\pi(\cdot), \uparrow), \quad \widetilde{W}(\downarrow, \bar{\mathbf{Q}}) = \sup_{\mathbf{Q}} \widetilde{W}(\downarrow, \mathbf{Q}).$$

Из теорем 2.4, 2.5 предыдущего параграфа мы знаем, что $\bar{\mathbf{Q}}$ есть наихудшее распределение,

$$\widetilde{W}(\pi_{\bar{Q}}(\cdot), \bar{\mathbf{Q}}) = \sup_{\mathbf{Q}} \widetilde{W}(\pi_Q(\cdot), \mathbf{Q}) = \sup_{\mathbf{Q}} \widetilde{W}(\downarrow, \mathbf{Q}),$$

и $\bar{\pi}(X) = \pi_{\bar{Q}}(X)$ является байесовской стратегией, соответствующей $\bar{\mathbf{Q}}$.

Мы знаем также (см. теорему 2.5), что для того чтобы стратегия $\bar{\pi}(X)$ была минимаксной, необходимо и достаточно, чтобы она была байесовской: $\bar{\pi}(X) = \pi_Q(X)$ для некоторого априорного распределения $\mathbf{Q}$, и

$$\widetilde{W}(\bar{\pi}(\cdot), \bar{\theta}) = c = \mathrm{const} \quad \text{п. в. } [\mathbf{Q}],$$
$$\widetilde{W}(\bar{\pi}(\cdot), \bar{\theta}) \leqslant c.$$

Этот последний критерий минимаксности мы уже неоднократно использовали в различных частных ситуациях (см. §§ 2.11[6], 3.1[6], 3.5[6], 3.9[6]).

Теорема 2. *При выполнении условий* (A), (B), (C) *класс всех байесовских стратегий является полным.*

В приложении мы приводим доказательства теорем 1, 2 в их более общей форме, когда D и Θ — произвольные компактные метрические пространства (условие (A)); функция $w(\delta, \theta)$: $D \times \Theta \to R$ непрерывна по δ и θ в соответствующих метриках (условие (B)); распределение $\mathbf{P}_\theta$ непрерывно относительно θ по вариации (условие (C)).

Доказательства теорем 1, 2 при некоторых дополнительных предположениях можно извлечь из [9]. Доказательства для случая конечных D и Θ можно найти в [3], [32]. В этих же монографиях можно найти сравнительно полное изложение элементов общей теории статистических игр (и, в частности, обсуждение для некоторых случаев конструкции минимального полного класса; см. [32]).

Теоремы 1, 2 показывают, насколько важной является задача описания класса всех байесовских решающих правил. Этой задаче посвящен следующий параграф.

§ 4. Байесовский принцип. Полный класс решающих функций

Мы видели, что по своей конструкции статистическая игра является объектом, более сложным, чем исходная игра (D, Θ, w). Для последней, особенно в случае простых множеств D и Θ (например, конечных), отыскание байесовских и минимаксных стратегий может оказаться делом сравнительно простым. В то же время даже простейшие статистические игры имеют весьма сложную природу множества $\mathscr{D}$, и это заметно усложняет их изучение, если подходить к ним как к обычным играм.

Пример 1. Пусть множества $D=\{\delta_1, \delta_2\}$, $\Theta=\{\theta_1, \theta_2\}$ — двухточечные, $w(\delta_i, \theta_j)=w_{ij}$, $w_{ii}=0$, $i, j=1, 2$. Пусть $Q=(q, 1-q)$ есть априорное распределение на Θ. Тогда

$$\widetilde{w}(\delta_i, Q)=qw_{i1}+(1-q)w_{i2}.$$

Стало быть, байесовская стратегия π_Q имеет вид

$$\pi_Q(\delta_2)=\begin{cases}0, & \text{если } \widetilde{w}(\delta_1, Q)<\widetilde{w}(\delta_2, Q) \quad (qw_{21}>(1-q)w_{12}),\\ 1, & \text{если } \widetilde{w}(\delta_2, Q)<\widetilde{w}(\delta_1, Q) \quad (qw_{21}<(1-q)w_{12})\end{cases} \tag{1}$$

($\pi_Q(\delta_i)$ есть вероятность принятия δ_i).

Если

$$\widetilde{w}(\delta_1, Q)=\widetilde{w}(\delta_2, Q), \tag{2}$$

или, что то же, $q=\bar{q}=w_{12}/(w_{12}+w_{21})$, то в качестве π_Q можно взять любое распределение π на множестве (δ_1, δ_2). Точно так же всегда можно найти распределение $\pi=(p, 1-p)$ такое, что

$$\widetilde{w}(\pi, \theta_1)=\widetilde{w}(\pi, \theta_2), \text{ или } pw_{12}=(1-p)w_{21}.$$

Решение этого уравнения $\bar{p}=w_{21}/(w_{21}+w_{12})$, очевидно, отвечает байесовской уравнивающей стратегии $\pi_{\bar{Q}}$, $\bar{Q}=(\bar{q}, 1-\bar{q})$, которая в силу теорем 2.4, 2.5 будет минимаксной. Распределение $\bar{Q}$ будет наихудшим.

Мы видим, что «решение» этой игры осуществляется довольно просто. Если же перейти к статистической игре, то даже в простейшем случае $w_{12}=w_{21}=1$ мы получим задачу о байесовских и минимаксных критериях, на рассмотрение которой нам потребовалось два параграфа 3.1[6], 3.2[6].

Замечательный факт, которому посвящен настоящий параграф, состоит в том, что задача отыскания байесовских стратегий (а стало быть, полного класса и минимаксных стратегий) для статистических игр может быть редуцирована в известном смысле к той же задаче для исходных игр (D, Θ, w). В основе этой редукции лежит следующее утверждение, которое мы назовем *байесовским принципом*. Пусть, как и прежде,

$$f_0(X) = \prod_{i=1}^{n} f_\theta(x_i)$$

есть функция правдоподобия выборки X; она же — плотность X в $\mathscr{X}^n$ относительно μ^n. Пусть, кроме того, априорное распределение Q на $(\Theta, \mathfrak{F}_\theta)$ имеет плотность $q(t)$ относительно некоторой меры λ (очевидно, это не есть ограничение). Тогда, согласно § 2.11[6], функция $f(x, t) = q(t) f_t(x)$ будет плотностью совместного распределения (X, θ) в $\mathscr{X}^n \times \Theta$. Это означает, что функция

$$q(t/x) = \frac{q(t) f_t(x)}{f(x)}, \tag{3}$$

$$f(x) = \int q(t) f_t(x) \lambda(dt),$$

определяет условную плотность распределения θ при условии $X = x$. Эта плотность соответствует *апостериорному распределению* Q_x случайной величины θ при условии $X = x$. Соотношение (3) носит название *формулы Байеса* (см. §§ 2.10[6], 2.11[6]).

Теорема 1 (байесовский принцип). *Пусть выполнено условие* (A_μ), *априорное распределение на* Θ *имеет плотность* $q(t)$ *и* Q_X *означает апостериорное распределение с плотностью* (3), *соответствующее априорному распределению* Q. *Пусть, далее, исходная игра* (D, Θ, w) *для любого априорного распределения* Q *имеет байесовскую стратегию* π_Q. *Тогда статистическая игра* $(\mathscr{D}, \Theta, W)$ *имеет байесовскую стратегию* $\pi_Q(X)$, *соответствующую распределению* Q, *которая совпадает с* π_{Q_X} — *байесовской стратегией исходной игры, соответствующей апостериорному распределению* Q_x.

Утверждение этой теоремы можно выразить одним равенством

$$\pi_Q(X) = \pi_{Q_X}.$$

Оно сводит поставленную задачу к нахождению апостериорного распределения Q_x и к отысканию байесовских стратегий для исходной игры.

Теорема 1 очень существенна для понимания механизма влияния информации, полученной от выборки, на выбор оптимальной стратегии. Априорная информация, представленная распределением Q на Θ, под влиянием экспериментальных данных постоянно меняется. Оптимальной стратегией будет та, которая учитывает эти изменения следующим образом: надо взять оптимальную стратегию в исходной игре, но соответствующую уже не Q, а Q_x.

Доказательство теоремы 1. Имеем

$$\widetilde{W}(\pi(\cdot), Q) = \int_{\Theta}\int_{\mathscr{X}^n} \widetilde{w}(\pi(x), t) f_t(x)\, \mu^n(dx)\, q(t)\, \lambda(dt) =$$

$$= \int_{\mathscr{X}^n} f(x)\, \mu^n(dx) \int_{\Theta} \widetilde{w}(\pi(x), t)\, q(t/x)\, \lambda(dt). \tag{4}$$

Здесь мы воспользовались (3). Изменение порядка интегрирования законно в силу неотрицательности подынтегральной функции. Второй интеграл в правой части (4) есть не что иное, как $\widetilde{w}(\lambda(x), Q_x)$. Но при любом x

$$\widetilde{w}(\pi(x), Q_x) \geqslant \widetilde{w}(\pi_{Q_x}, Q_x) = \int_{\Theta} \widetilde{w}(\pi_{Q_x}, t)\, q(t/x)\, \lambda(dt).$$

Подставляя это неравенство в (4) и возвращаясь к исходному порядку интегрирования, получим

$$\widetilde{W}(\pi(\cdot), Q) \geqslant \int_{\mathscr{X}^n} f(x)\, \mu^n(dx) \int_{\Theta} \widetilde{w}(\pi_{Q_x}, t)\, q(t/x)\, \lambda(dt) =$$

$$= \widetilde{W}(\pi_{Q_\cdot}, Q).$$

Так как $\pi(x)$ здесь произвольна, то это означает, что

$$\pi_Q(x) = \pi_{Q_x}. \ \triangleleft$$

Замечание 1. Аккуратности ради мы должны в проведенных рассмотрениях оговаривать измеримость относительно $\mathfrak{B}^n \times \mathfrak{F}_\Theta$ функции $\widetilde{w}(\pi_{Q_x}, t)$. Эти оговорки мы опускаем, так как они носят чисто технический характер, а при выполнении условий (А), (В), (С) § 3 заведомо излишни. Последнее утверждение читатель может проверить самостоятельно, пользуясь тем, что для дискретных D и Θ эта измеримость устанавливается очевидным образом, и тем, что произвольная игра при условиях (A), (B) может быть сколь угодно точно «приближена» дискретной.

Возвращаясь к примеру 1, мы можем теперь на основании теоремы 1 сразу указать вид байесовских стратегий для соответствующей статистической игры. Именно, из (1) получаем

$$\pi_{Q_X}(\delta_2) = \begin{cases} 0, \text{ если } q_X \equiv \dfrac{q f_{\theta_1}(X)}{q f_{\theta_1}(X) + (1-q) f_{\theta_2}(X)} > \dfrac{w_{12}}{w_{12}+w_{21}}, \\ 1, \text{ если } q_X < \dfrac{w_{12}}{w_{12}+w_{21}}. \end{cases} \tag{5}$$

Если

$$q_X = \frac{w_{12}}{w_{12}+w_{21}}, \tag{6}$$

то в качестве π_{Q_X} можно брать любое распределение на (δ_1, δ_2). Неравенство (5) можно переписать в виде

$$\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}, \qquad a = \frac{w_{12}}{w_{12}+w_{21}}. \tag{7}$$

Это знакомый нам критерий отношения правдоподобия.

Далее,

$$\widetilde{W}(\pi_{Q_.}, \theta_j) = w_{1j}\mathsf{M}_{\theta_j}\pi_{Q_X}(\delta_1) + w_{2j}\mathsf{M}_{\theta_j}\pi_{Q_X}(\delta_2), \qquad j = 1, 2.$$

Допустим для простоты, что равенство (6) имеет место с $\mathbf{P}_{\theta_j}$-вероятностью 0, так что байесовская стратегия с $\mathbf{P}_{\theta_j}$-вероятностью 1 будет чистой, $j = 1, 2$. Тогда

$$\mathsf{M}_{\theta_j}\pi_{Q_X}(\delta_1) = \mathbf{P}_{\theta_j}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}\right),$$

$$\widetilde{W}(\pi_{Q_.}, \theta_1) = w_{21}\mathbf{P}_{\theta_1}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} < \frac{a(1-q)}{q(1-a)}\right),$$

$$\widetilde{W}(\pi_{Q_.}, \theta_2) = w_{12}\mathbf{P}_{\theta_2}\left(\frac{f_{\theta_1}(X)}{f_{\theta_2}(X)} > \frac{a(1-q)}{q(1-a)}\right).$$

Отсюда уже нетрудно найти значение $\bar{q}$, соответствующее наихудшему распределению $\overline{Q}$, при котором $\pi_{\overline{Q}_X}$ будет уравнивающей стратегией, т. е. стратегией, при которой

$$\widetilde{W}\left(\pi_{\overline{Q}_{\cdot}}, \theta_1\right) = \widetilde{W}\left(\pi_{\overline{Q}_{\cdot}}, \theta_2\right).$$

Эта стратегия, согласно теоремам 2.4, 2.5, будет минимаксной.

Распространение описанной процедуры отыскания минимаксной стратегии на общий случай, когда P_{θ_1}- или P_{θ_2}-распределения $f_{\theta_1}(X)/f_{\theta_2}(X)$ содержат дискретную компоненту, мы предоставляем читателю.

Пользуясь теоремой 1, мы можем аналогичным образом получить обобщение результатов §§ 3.1[6], 3.2[6] на случай произвольных конечных D и Θ и произвольной функции потерь $w(\delta_i, \theta_j) = w_{ij}$, которую в этом случае можно называть также матрицей потерь $\|w(\delta_i, \theta_j)\|$. (В §§ 3.1[6], 3.2[6] мы рассматривали частный случай $w_{ij} = 1$ при $i \neq j$.) При произвольных w_{ij} байесовское решающее правило будет иметь следующий вид. Пусть $Q = (q(\theta_1), \ldots, q(\theta_r))$, $Q_X = (q_X(\theta_1), \ldots, q_X(\theta_r))$,

$$q_X(\theta_j) = \frac{q(\theta_j)\, f_{\theta_j}(X)}{\sum_i q(\theta_i)\, f_{\theta_i}(X)}.$$

Тогда $\widetilde{w}(\delta_i, Q_X) = \sum_{j=1}^{r} w_{ij} q_X(\theta_j)$ и, стало быть,

$$\pi_{Q_X}(\delta_k) = 1, \text{ если } \widetilde{w}(\delta_k, Q_X) \leqslant \widetilde{w}(\delta_i, Q_X) \quad \text{при всех } i,$$

или, что то же, если

$$\sum_{j=1}^{r} w_{kj} f_{\theta_j}(X)\, q(\theta_j) \leqslant \sum_{j=1}^{r} w_{ij} f_{\theta_j}(X)\, q(\theta_j).$$

Если существует несколько значений k, обладающих этим свойством (обозначим их $k_1, \ldots, k_s$), то байесовской стратегией π_{Q_X} будет также любое распределение на $\delta_{k_1}, \ldots, \delta_{k_s}$.

Отыскание минимаксной стратегии происходит следующим образом. Предположим опять для простоты, что P_{θ_j}-распределения $\widetilde{w}(\delta_i, Q_X)$ не имеют дискретных компонент. Тогда

$$\widetilde{W}(\pi_{Q_\cdot}, \theta_j) = \sum_{i \neq j} w_{ij} \mathrm{P}_{\theta_j}\left(\widetilde{w}(\delta_i, \mathrm{Q}_X) < \min_{l \neq i} \widetilde{w}(\delta_l, \mathrm{Q}_X)\right).$$

В силу теоремы 3.1 существует $\overline{\mathrm{Q}} = (\overline{q}(\theta_1), \ldots, \overline{q}(\theta_r))$, при котором стратегия $\pi_{\overline{Q}_X}$ будет уравнивать значения $\widetilde{W}\left(\pi_{\overline{Q}_X}, \theta_j\right)$ при всех j. Эта стратегия и будет минимаксной.

Из проведенных рассмотрений и теоремы 3.2 нетрудно получить также вид полного класса стратегий статистической игры $(\mathscr{D}, \Theta, \widetilde{W})$ в случае конечных D и Θ.

Рассмотрим стратегии π_{Q_X}, представляющие из себя произвольное распределение на тех $\delta_{k_1}, \ldots, \delta_{k_s}$, для которых

$$\min_i \left(\sum_{j=1}^{r} \left(w_{k_l, j} - w_{ij}\right) f_{\theta_j}(X)\, q(\theta_j) \right) = 0.$$

Класс таких стратегий (байесовских), который получится, если $q(\theta_1), \ldots, q(\theta_r)$ будут пробегать все возможные значения, и будут полным классом. Мы видели, что в случае $r = 2$ этот класс оказывается весьма простым и узким (см. (7)): он состоит из решающих функций $\pi(X) = (\pi(X, \delta_1), \pi(X, \delta_2))$, где $\pi(X, \delta_i)$ суть вероятности принятия решения δ_i,

$$\pi(X, \delta_1) = \begin{cases} 1, & \text{если } R(X) > c, \\ p \in [0, 1], & \text{если } R(X) = c, \\ 0, & \text{если } R(X) < c, \end{cases}$$

$$R(X) = \frac{f_{\theta_1}(X)}{f_{\theta_2}(X)}, \quad 0 \leqslant c \leqslant \infty. \tag{8}$$

В играх с континуальными множествами D и Θ для некоторых важных конкретных функций потерь также можно найти в явном виде форму байесовских решений. Пусть, например, D и Θ — области из R^k, а функция потерь является квадратичной:

$$w(\delta, \theta) = c\,|\delta - \theta|^2 = c \sum_{i=1}^{k} |\delta_i - \theta_i|^2, \tag{9}$$

где δ_i, θ_i — координаты δ и θ. Тогда

$$\widetilde{w}(\delta, \mathrm{Q}) = c \int |\delta - t|^2 \mathrm{Q}(dt) = c\mathrm{M}_Q |\delta - \theta|^2.$$

Мы знаем, что минимум этого выражения достигается при $\delta = M_Q\theta = \int tQ(dt)$. Это и есть, очевидно, байесовская стратегия $\delta_Q = M_Q\theta$. Отсюда и из байесовского принципа следует, что байесовская стратегия $\delta_Q(X) = \theta_Q^*$ в статистической игре будет иметь вид

$$\theta_Q^* = \delta_{Q_X} = \int\limits_{R^k} tQ_X(dt) = \int\limits_{R^k} tq(t/X)\,\lambda(dt). \tag{10}$$

Этот результат нам уже известен по главе 2[6].

Риск стратегии θ_Q^* равен $W(\theta_Q^*, \theta) = cM_\theta|\theta_Q^* - \theta|^2$. Априорное распределение $\overline{Q}$, для которого $M_\theta|\theta_Q^* - \theta|^2 = \mathrm{const}$, даст нам минимаксную оценку $\overline{\theta}^* = \delta_{\overline{Q}}(X)$. Примеры построения минимаксных оценок на этом пути содержатся в § 2.11[6].

Класс оценок (10), где Q пробегает значения в классе всех распределений на Θ, представляет собой полный класс.

Рассмотрим теперь другой частный случай функции потерь

$$w(\delta, \theta) = c|\delta - \theta| \tag{11}$$

и предположим, что $\Theta = R$, $D = R$. Тогда

$$\widetilde{w}(\delta, Q) = cM_Q|\delta - \theta| = c\int|\delta - t|\,Q(dt) =$$

$$= c\int\limits_{-\infty}^{\delta}(\delta - t)\,Q(dt) + c\int\limits_{\delta}^{\infty}(t - \delta)\,Q(dt).$$

Пользуясь интегрированием по частям и обозначая $F(t) = Q((-\infty, t))$, находим

$$\widetilde{w}(\delta, Q) = c\int\limits_{-\infty}^{\delta}(\delta - t)\,dF(t) - c\int\limits_{\delta}^{\infty}(t - \delta)\,d(1 - F(t)) =$$

$$= c\left[\int\limits_{-\infty}^{\delta}F(t)\,dt + \int\limits_{\delta}^{\infty}(1 - F(t))\,dt\right].$$

Производная этого выражения по δ существует п. в. и равна $c[2F(\delta) - 1]$. Эта функция монотонно возрастает и меняет знак в точке $\overline{\delta}$, равной медиане распределения F: $F(\overline{\delta} - 0) \leqslant 1/2$, $F(\overline{\delta} + 0) \geqslant 1/2$. Отсюда следует, что $\widetilde{w}(\delta, Q)$ будет выпуклой по δ и в точке $\overline{\delta}$ будет иметь минимум.

В силу байесовского принципа это означает, что *байесовской оценкой* $\theta^*_Q = \delta_Q(X)$ *для априорного распределения* Q *и функции потерь* (11) *будет медиана апостериорного распределения* Q_x. Это дает возможность, как и в случае (9), найти минимаксную решающую функцию и полный класс.

Аналогичным образом можно рассмотреть случай

$$w(\delta, \theta) = c|\delta - \theta|^\alpha, \quad \alpha > 0.$$

В заключение этого параграфа отметим, что квадратичная функция потерь (9) при $c = 1$ для континуальных множеств D и Θ и функция потерь

$$w(\delta_i, \theta_j) = \begin{cases} 0, & i = j, \\ 1, & i \neq j, \end{cases} \tag{12}$$

для конечных D и Θ играют в теории статистических игр особую роль. Для них функции риска превращаются соответственно в сумму дисперсии и квадрата смещения оценки для континуальных D и Θ и в вероятность ошибиться — для конечных D и Θ. Эти характеристики, являющиеся естественными сами по себе, служили нам основой для выбора оптимальных правил в гл. 2[6], 3[6], 1. Если статистическая задача не содержит прямых указаний на форму функции $w(\delta, \theta)$, то чаще всего в качестве $w(\delta, \theta)$ выбираются именно эти две функции (9) или (12). Мы условились называть их *статистическими функциями потерь*.

§ 5. Достаточность, несмещенность, инвариантность

Принципы достаточности, несмещенности, инвариантности служат для сужения класса решающих правил. Они состоят в том, чтобы использовать в качестве решающих функций лишь соответственно достаточные, несмещенные и инвариантные решающие правила. Использование одного из этих принципов, двух из них или всех трех (если это оказывается возможным) позволяет в ряде случаев настолько сузить класс рассматриваемых стратегий, что его пересечение с полным классом оказывается состоящим из одной-единственной решающей функции. Это означает, что в выделенном подклассе существует равномерно наилучшая стратегия (ср. с п. 1 § 2), и это решает проблему выбора решения.

Все три принципа представляются довольно естественными и уже обсуждались нами в разных конкретных задачах в гл. 2[6], 3[6].

Наиболее бесспорным из них является принцип достаточности, который часто преставляет собой не что иное, как способ описания полного класса.

1. Достаточность. Предположим, что выполнено условие (A_μ) и что существует достаточная статистика S, т. е. (см. § 2.12[6])

$$f_\theta(X) = \psi(\theta, S)\cdot h(X).$$

Пусть, далее, априорное распределение Q имеет плотность $q(t)$ относительно некоторой меры λ. Тогда в силу байесовского принципа байесовская стратегия будет полностью определяться апостериорной плотностью

$$q(t/X) = \frac{q(t)\, f_t(X)}{\int q(u)\, f_u(X)\, \lambda(du)} = \frac{q(t)\,\psi(t, S)}{\int q(u)\,\psi(u, S)\,\lambda(du)}, \qquad (1)$$

зависящей лишь от S. Так как любое распределение Q имеет плотность относительно подобранной соответствующим образом меры λ (можно положить, например, $\lambda = Q$, $q(t) \equiv 1$), то сказанное означает, что все байесовские решающие правила $\pi_Q(X)$ будут функциями лишь от S:

$$\pi_Q(X) = p_Q(S).$$

Другими словами, любая байесовская стратегия $\pi_Q(X)$ не зависит от X при фиксированной S.

Пусть теперь выполнены условия (A), (B), (C) § 3. Тогда высказанное утверждение будет относиться и к минимаксным стратегиям. Оно будет означать также, что все решающие правила, построенные как функции лишь от S (т. е. все измеримые отображения $\mathscr{S} \to \tilde{D}$, где $\mathscr{S}$ есть пространство, в котором лежат значения S), образуют полный класс $\mathscr{D}_{\mathscr{S}}$. Это следует из того, что $\mathscr{D}_{\mathscr{S}}$ содержит в себе все байесовские стратегии, образующие, как мы знаем, полный класс. Очевидно, класс $\mathscr{D}_{\mathscr{S}}$ будет наименьшим для минимальной достаточной статистики S.

Ясно, что в минимальный полный класс входят далеко не все функции от S (со значениями в $\tilde{D}$), а лишь узкая их часть. Это подтверждает формула (1), из которой следует, например, что для двухточечных множеств D и Θ (см. (4.8)) полный класс образуют функции $\pi(X)$, у которых вероятность $\pi(X, \delta_1)$ принятия решения δ_1 имеет форму индикатора множества $\{R(X) > c\}$, где $R(X) = \psi(\theta_1, S)/\psi(\theta_2, S)$ (точнее см. (4.8)).

Если $D \subset R^k$, $\Theta \subset R^k$, а функция потерь $w(\delta, \theta)$ имеет вид $w(\delta, \theta) = w(\delta - \theta)$, где $w(u)$ — выпуклая функция в R^k, то принципу достаточности можно придать очень конструктивную форму, позволяющую эффективно характеризовать полный класс. Именно, имеет место следующее обобщение теоремы 2.14.1[6].

Теорема 1 (Блекуэлл). *Для любой решающей функции (оценки)* $\theta^* = \delta(X)$ *существует оценка*

$$\theta_S^* = \mathsf{M}_\theta(\theta^*/S)$$

(θ_S^* от θ не зависит, так как S — достаточная статистика), которая не хуже, чем θ^*. Именно, для всех $\theta \in \Theta$

$$\mathsf{M}_\theta w\left(\theta_S^* - \theta\right) \leqslant \mathsf{M}_\theta w(\theta^* - \theta).$$

Доказательство. Имеет место следующее неравенство Иенсена (см. § 2.9[6]): если g — выпуклая функция в R^k, ξ — случайная величина со значениями в R^k, а $\mathfrak{F}$ — какая-нибудь σ-подалгебра основной σ-алгебры, то

$$\mathsf{M}\left(g(\xi)/\mathfrak{F}\right) \geqslant g\left(\mathsf{M}(\xi/\mathfrak{F})\right).$$

В соответствии с этим неравенством

$$\mathsf{M}_\theta w(\theta^* - \theta) = \mathsf{M}_\theta\{\mathsf{M}_\theta(w(\theta^* - \theta)/S)\} \geqslant$$
$$\geqslant \mathsf{M}_\theta w(\mathsf{M}_\theta(\theta^* - \theta/S)) = \mathsf{M}_\theta w\left(\theta_S^* - \theta\right). \quad \triangleleft$$

Если достаточная статистика S является полной, то теорема 1 вместе с принципом несмещенности позволяет однозначно определить равномерно наилучшую оценку. Действительно, рассмотрим класс K_0 всех несмещенных оценок $\theta^* = \delta(X)$:

$$\mathsf{M}_\theta\theta^* = \theta \text{ для } \theta^* \in K_0.$$

Тогда, следуя в точности рассуждениям § 2.14[6] (теорема 3), убеждаемся, что $\theta_S^* = \mathsf{M}_\theta(\theta^*/S)$ для всех $\theta^* \in K_0$ совпадают и, стало быть, пересечение K_0 и полного класса состоит из одной-единственной оценки $\varphi(S)$, которую естественно называть *эффективной*.

Из сказанного видно, что *эффективные оценки, если они существуют, будут одни и те же для произвольной выпуклой функции потерь* $w(\delta - \theta)$. Это позволяет использовать для любой такой функции потерь все утверждения соответствующих теорем гл. 2[6], полученных для $w(u)=u^2$.

Приведенные рассуждения иллюстрируют совместное применение принципов достаточности и несмещенности.

2. Несмещенность. Мы только что видели, какую роль может играть принцип несмещенности в теории оценок. В § 3.6[6] было установлено, что аналогичный эффект (существование равномерно наиболее мощных несмещенных критериев) может быть получен при использовании несмещенных критериев в теории проверки статистических гипотез.

В общем случае определение несмещенности выглядит следующим образом. Допустим, что проблема статистического решения состоит в «определении» неизвестного значения θ и что, следовательно, множества D и Θ совпадают. Функция потерь $w(\delta, \theta)$ может быть произвольной.

Определение 1. Решающая функция $\delta(X)$ называется *несмещенной*, если

$$\mathsf{M}_\theta w(\delta(X), \theta) \leqslant \mathsf{M}_\theta w(\delta(X), \theta')$$

при любых θ, $\theta' \neq \theta$.

Другими словами, при $v=\theta$ достигается $\min\limits_v \mathsf{M}_\theta w(\delta(X), v)$. Это означает, что $\delta(X)$ в среднем находится ближе всего к неизвестному θ, чем к какой-нибудь другой точке.

Легко видеть, что данное ранее определение несмещенных оценок является частным случаем этого определения.

Если проверяются две сложные гипотезы $H_1=\{\theta \in \Theta_1\}$ и $H_2=\{\theta \in \Theta_2\}$, то множество $D=\{\delta_1, \delta_2\}$ может существенно отличаться от Θ. В этом случае определение несмещенности будет формально несколько иным, хотя смысл его остается тем же. Именно, определение 1 можно модифицировать так (см. [18]), что оно перейдет в следующее определение.

Определение 1А. Решающая функция $\delta(X)$ называется *несмещенной*, если

$$\mathsf{M}_\theta w(\delta(X), \theta) \leqslant \mathsf{M}_\theta w(\delta(X), \theta')$$

при любых $\theta \in \Theta_1$, $\theta' \in \Theta_2$ или $\theta \in \Theta_2$, $\theta' \in \Theta_1$.

Пусть для простоты $w(\delta_1, \theta)=w_1=\text{const}$ при $\theta \in \Theta_2$; $w(\delta_2, \theta)=w_2=\text{const}$ при $\theta \in \Theta_1$; $\delta_1=0$, $\delta_2=1$ и $\delta(X)$

означает вероятность (1 или 0) принятия H_2. Тогда

$$\mathsf{M}_\theta w(\delta(X), \theta) = \begin{cases} w_2 \mathrm{P}_\theta(\delta(X) = 1) & \text{при } \theta \in \Theta_1, \\ w_1 \mathrm{P}_\theta(\delta(X) = 0) & \text{при } \theta \in \Theta_2, \end{cases}$$

$$\mathsf{M}_\theta w(\delta(X), \theta') = \begin{cases} w_1 \mathrm{P}_\theta(\delta(X) = 0) & \text{при } \theta \in \Theta_1, \theta' \in \Theta_2, \\ w_2 \mathrm{P}_\theta(\delta(X) = 1) & \text{при } \theta \in \Theta_2, \theta' \in \Theta_1, \end{cases}$$

и неравенства в определении 1А означают, что

$$w_2 \mathrm{P}_{\theta_1}(\delta(X) = 1) \leqslant w_1 \mathrm{P}_{\theta_1}(\delta(X) = 0) \quad \text{при } \theta_1 \in \Theta_1,$$

$$w_1 \mathrm{P}_{\theta_2}(\delta(X) = 0) \leqslant w_2 \mathrm{P}_{\theta_2}(\delta(X) = 1) \quad \text{при } \theta_2 \in \Theta_2,$$

или, что то же,

$$\mathrm{P}_{\theta_1}(\delta(X) = 1) \leqslant \frac{w_1}{w_1 + w_2}, \quad \mathrm{P}_{\theta_2}(\delta(X) = 1) \geqslant \frac{w_1}{w_1 + w_2}.$$

Отсюда следует, что

$$\sup_{\theta \in \Theta_1} \mathsf{M}_\theta \delta(X) \leqslant \inf_{\theta \in \Theta_2} \mathsf{M}_\theta \delta(X)$$

и что, стало быть, критерий δ будет несмещенным в смысле определений § 3.6[6]. Наоборот, если справедливо последнее неравенство, то критерий δ будет несмещенным в смысле определения 1А при подходящем выборе функции потерь $w(\delta, \theta)$, например, при $w_1/(w_1 + w_2) = \sup_{\theta \in \Theta_1} \mathsf{M}_\theta \delta(X)$.

Дополнительные примеры использования принципа несмещенности (помимо результатов § 3.6[6]) можно найти в [18].

3. Инвариантность. Мы видели что пересечение полного класса, порожденного «достаточными» решениями, с классом несмещенных решений может состоять из единственной стратегии. Еще одним естественным классом стратегий, в котором может оказаться единственное неулучшаемое решение, является класс инвариантных решающих правил (ср. с §§ 2.18[6], 2.19[6], 3.7[6]).

Определение инвариантной проблемы статистического решения связано с группами преобразований во всех трех пространствах, которые участвуют в определении статистической игры: в пространствах D, Θ и выборочном пространстве $\mathscr{X}^n$. В основе определения лежат измеримые взаимно однозначные преобразования g пространства $\mathscr{X}^n$ в себя, образующие некоторую группу G с групповой операцией, определенной как композиция: если $g_1 \in G$, $g_2 \in G$, то $g_2 g_1$ определяется как преобразование $x \to g_2(g_1 x)$, которое снова обязано принадлежать G. Тождественное преобразование мы обозначим через e. Обратное к g преобразование g^{-1} определяется как преобразование, для которого $g^{-1}g = e$. Измеримость $g \in G$ означает, что gX вместе с X будет случайной величиной в $\mathscr{X}^n$.

С введенной группой G тесно связано понятие инвариантности семейства $\mathbf{P}_\theta$, которое было определено нами в §§ 2.19[6] и 3.7[6]. Оно означает, что для каждых $g \in G$ и $\theta \in \Theta$ найдется элемент $\theta_g \in \Theta$ такой, что

$$\mathbf{P}_\theta(gX \in A) = \mathbf{P}_{\theta_g}(X \in A). \tag{2}$$

Преобразования $\bar{g}$ пространства Θ в себя, определенные равенством $\bar{g}\theta = \theta_g$, при выполнении условия (Λ_0) образуют группу $\bar{G}$ (см. § 2.19[6]).

В терминах математических ожиданий условие (2) означает, что для любой интегрируемой функции φ

$$\mathsf{M}_\theta \varphi(gX) = \mathsf{M}_{\bar{g}\theta}\varphi(X). \tag{3}$$

Определение 2. Проблема статистического решения, связанная со статистической игрой $(\mathscr{D}, \Theta, w)$, $(X, \mathbf{P}_\theta)$, называется *инвариантной относительно группы* G, если семейство $\mathbf{P}_\theta$ инвариантно относительно G и функция потерь w инвариантна относительно G в следующем смысле: для любых $\delta \in D$, $g \in G$ найдется единственное $\delta' \in D$ такое, что

$$w(\delta, \theta) = w(\delta', \bar{g}\theta) \quad \text{при всех } \theta \in \Theta. \tag{4}$$

Значение δ', однозначно определенное по g, мы обозначим $g'\delta$.

Лемма 1. *Преобразования g' пространства D в себя, порожденные группой G, образуют группу G'.*

Доказательство. Мы покажем, что совокупность G' всех преобразований g' замкнута относительно композиции, причем справедливо равенство $g_2'g_1' = (g_2g_1)'$.

Действительно,

$$w(\delta, \theta) = w(g_1'\delta, \bar{g}_1\theta) = w(g_2'g_1'\delta, \bar{g}_2\bar{g}_1\theta) = \\ = w((g_2g_1)'\delta, (\overline{g_2g_1})\theta).$$

Так как $(\overline{g_2g_1}) = \bar{g}_2\bar{g}_1$, то в силу единственности $(g_2g_1)' = g_2'g_1'$. Лемма доказана.

Итак, с основной группой G преобразований g пространства $\mathscr{X}^n$ в себя связаны еще две группы $\bar{G}$ и G' преобразований пространств Θ и D в себя. Применение одновременно всех трех преобразований g, $\bar{g}$ и g' оставляет проблему решения неизменной (инвариантной). Поэтому естественно выбирать такие решающие правила, которые не менялись бы при переходе от одной эквивалентной проблемы решения к другой. Естественность такого подхода нами обсуждалась весьма детально в §§ 2.18[6], 2.19[6], 3.7[6].

Определение 3. Решающая функция $\delta(X)$ инвариантной проблемы решения называется *инвариантной*, если

$$\delta(gX) = g'\delta(X).$$

Рандомизированное инвариантное правило $\pi(X)$ определяется как любое распределение, сосредоточенное на инвариантных решающих правилах.

Примеры использования принципа инвариантности содержатся в уже цитированных нами §§ 2.18[6], 2.19[6], 3.7[6], где рассматривались эквивариантные оценки и инвариантные критерии. Отметим некоторое своеобразие с точки зрения общего подхода этих двух частных случаев.

В *проблеме оценивания* группа преобразований G' не вводилась вовсе. В этом случае множества D и Θ совпадают, и с самого начала предполагалось, что $g'\delta = \bar{g}\delta$. Поэтому эквивариантные оценки определялись с помощью равенства $\theta^*(gX) = \bar{g}\theta^*(X)$.

В *теории проверки гипотез* преобразование g' полагалось равным тождественному $g' = e$, так что инвариантный критерий π определялся соотношением $\pi(gX) = \pi(X)$.

В этом случае для инвариантности проблемы проверки двух гипотез $\{\theta \in \Theta_1\}$ и $\{\theta \in \Theta_2\}$ нужно предполагать также (см. (4)), что $g\Theta_i = \Theta_i$.

Наличием некоторой разницы в этих двух подходах и объясняется в какой-то мере использование двух разных терминов: «эквивариантность» (для оценок) и «инвариантность» (для проверки гипотез) для обозначения инвариантных решающих правил. Дополнительно к примерам инвариантных проблем решения, рассмотренным в гл. 2[6] и 3[6], приведем еще один.

Пример 1. Пусть $X \in \Phi_{\alpha,\sigma^2}$. Здесь Θ есть полуплоскость $\{\theta = (\alpha, \sigma): \sigma \geqslant 0\}$. Пусть D есть вещественная прямая R, и пусть $w(\delta, \theta) = (\delta - \alpha)^2/\sigma^2$.

Рассмотрим группу G преобразований $g_{a,b}X = a + bX = (a + bx_1, \ldots, a + bx_n)$, где $b \neq 0$. Случайную величину $g_{a,b}X$ в $\mathscr{X}^n$ можно рассматривать, очевидно, как выборку из $\Phi_{a+b\alpha,b^2\sigma^2}$. Следовательно, семейство Φ_{α,σ^2} инвариантно относительно G, если положить $\bar{g}_{a,b}\theta = (a + b\alpha, |b|\sigma)$. Функция потерь будет инвариантной, если мы положим $g'_{a,b}\delta = a + b\delta$, поскольку

$$w\left(g'_{a,b}\delta, \bar{g}_{a,b}\theta\right) = \frac{(a + b\delta - a - b\alpha)^2}{b^2\sigma^2} = w(\delta, \theta).$$

Таким образом, мы имеем инвариантную относительно G проблему решения. Инвариантные решающие функции $\delta(X)$: $\mathscr{X}^n \to R$ должны обладать свойством

$$\delta(a + bX) = \delta(g_{a,b}X) = g'_{a,b}\delta(X) = a + b\delta(X). \qquad (5)$$

Далее, нетрудно установить, что рассматриваемая проблема решения является инвариантной и относительно группы F всех перестановок f координат вектора X; при этом $\bar{f}$ и f' будут тождественными преобразованиями. Поэтому, если потребовать, чтобы функция $\delta(X)$ была инвариантным решением и относительно F, то должно выполняться также

$$\delta(fX) = \delta(X). \qquad (6)$$

Отметим, что класс функций, удовлетворяющих (5), (6), все еще достаточно широк: в него входят, например, все линейные формы

$$\delta(X) = \sum_{k=1}^{n} a_k x_{(k)}, \qquad \sum_{k=1}^{n} a_k = 1,$$

где $x_{(1)}, \ldots, x_{(n)}$ — вариационный ряд выборки X. Если использовать принцип несмещенности, то мы получим еще одно условие на коэффициенты a_k:

$$\sum_{k=1}^{n} a_k \mathrm{M}_\theta (x_{(k)} - \alpha) = 0. \ \triangleleft$$

При построении оптимальных инвариантных решений в теории оценивания и теории проверки статистических гипотез важную роль играют понятия, в известном смысле близкие друг другу: понятие *орбиты* в теории оценок и понятие *инварианта* в теории проверки гипотез.

Напомним, что орбитой в пространстве Θ называется множество $\{\bar{g}\theta_0,\ \bar{g}\in\bar{G}\}$, где θ_0 есть некоторая точка из Θ. Другими словами, θ_1 и θ_2 принадлежат одной орбите, если существует $\bar{g}\in\bar{G}$ такое, что $\theta_1=\bar{g}\theta_2$.

Аналогично можно определить орбиты в $\mathscr{X}^n$. Тогда инварианты по определению есть статистики, постоянные на орбитах в $\mathscr{X}^n$.

Понятие орбиты сохраняет свое значение и в общем случае.

Лемма 2. *Функция риска инвариантной проблемы решения для инвариантного решающего правила постоянна на орбите:*

$$W(\delta(\cdot),\ \theta)=W(\delta(\cdot),\ \bar{g}\theta)$$

при всех $\theta\in\Theta$, $\bar{g}\in\bar{G}$.

Доказательство. В силу инвариантности соответственно функции потерь, решающего правила и семейства P_θ (см. (3), (4)) имеем

$$W(\delta(\cdot),\theta)=\mathsf{M}_\theta w(\delta(X),\theta)=\mathsf{M}_\theta w\left(g'\delta(X),\bar{g}\theta\right)=$$

$$=\mathsf{M}_\theta w\left(\delta(gX),\bar{g}\theta\right)=\mathsf{M}_{\bar{g}\theta} w\left(\delta(X),\bar{g}\theta\right)=W\left(\delta(\cdot),\bar{g}\theta\right). \quad \triangleleft$$

Постоянство на орбите риска для рандомизированных инвариантных решающих правил следует из их определения и леммы 2.

Из леммы 2 вытекает, что в том случае, когда все пространство Θ является орбитой (т. е. $\Theta=\{\bar{g}\theta_0,\ \bar{g}\in\bar{G}\}$ при каком-нибудь θ_0; это имеет место, например, для преобразований сдвига), инвариантное решающее правило становится уравнивающим. Поэтому из леммы 2 и теорем 2.3, 2.5 мы получаем немедленно следующее утверждение, устанавливающее важную связь между инвариантностью и минимаксностью.

Теорема 2. *Пусть пространство Θ является орбитой и существует априорное распределение $\mathbf{Q}$, для которого байесовская стратегия $\pi_Q(X)$ является инвариантной. Тогда $\pi_Q(X)$ — минимаксная стратегия.*

Из теоремы 3.3 вытекает также, что имеет место следующее обобщение теоремы 2.

Теорема 2А. *Пусть существует априорное распределение $\mathbf{Q}$, сосредоточенное на одной из орбит Θ_0, такое, что байесовская стратегия $\pi_Q(X)$ является инвариантной.*

Тогда, если при всех θ

$$W(\pi_Q(\cdot), \theta) \leqslant W(\pi_Q(\cdot), \theta_0), \qquad \theta_0 \in \Theta_0,$$

то $\pi_Q(X)$ *является минимаксной.*

Этим критерием мы пользовались в § 3.9[6].

§ 6. Асимптотически оптимальные оценки при произвольной функции потерь

Многие из результатов гл. 2[6] об асимптотически оптимальных оценках и гл. 3[6] об асимптотически оптимальных критериях допускают обобщения на функции потерь весьма общего вида.

В этом параграфе мы остановимся на задачах теории оценивания и предположим, что $w(\delta, \theta) = w(\delta - \theta)$.

Сделаем сначала одно общее замечание. В гл. 2[6] мы видели, что в регулярном случае ($X \in \mathbf{P}_\theta$, $\mathbf{P}_\theta$ удовлетворяет условиям (RR); см. §§ 2.24[6], 2.28[6]) все разумные оценки $\theta^* = \delta(X)$ параметра θ «сосредоточены» в $1/\sqrt{n}$-окрестности точки θ. Так, например, для асимптотически нормальных оценок $(\theta^* - \theta)\sqrt{n} \Leftarrow\!\!\!\!\!\!\Rightarrow \Phi_{0,\sigma^2(\theta)}$. Отсюда следует, что при широких предположениях относительно функции $w(t)$ асимптотическое поведение риска $\mathsf{M}_\theta w(\theta^* - \theta)$ будет определяться свойствами функции $w(t)$ в окрестности точки $t = 0$. Если $w(t)$ дважды непрерывно дифференцируема в нуле, $w'' > 0$, то при $t \to 0$

$$w(t) = \frac{w''(0)}{2} t^2 + o(t^2). \qquad (1)$$

Это значит, что в интересующей нас области значений t (порядка $1/\sqrt{n}$) функция $w(t)$ будет вести себя так же, как квадратичная функция потерь $w_0(t) = ct^2$ при $c = w''(0)/2$, для которой были установлены результаты гл. 2[6]. Если к тому же $w(t) < e^{\alpha|t|^2}$ при достаточно малом $\alpha > 0$ (см. теорему 2.28.6[6]), то все эти результаты останутся в силе — перенесение их на случай функции $w(t)$ вида (1) — дело несложной техники, вполне доступной читателю.

В этом параграфе мы рассмотрим значительно более содержательное обобщение. Мы предположим, что функция потерь $w(\delta, \theta)$ зависит от n и представима в виде

$$w(\delta, \theta) = w_n(\delta - \theta) = w(\sqrt{n}(\delta - \theta)), \qquad (2)$$

где функция $w(t) \geqslant 0$ определена во всем пространстве R^k. Очевидно, что в этом случае будут существенными значения $w(t)$ во всей области значений t.

Мы будем предполагать, что функция w в (2) удовлетворяет следующим условиям:

1) $w(t) \leqslant e^{c|t|}$ *при некотором* $c > 0$.

Такая форма условия 1) несколько упрощает выкладки. На самом деле все результаты сохранятся, если потребовать, чтобы $w(t) \leqslant c_1 e^{\alpha |t|^2}$ при достаточно малом $\alpha > 0$.

В дальнейшем существенную роль будет играть функция

$$V_{\sigma^2}(s) = \int w(s-u)\, e^{-\frac{1}{2} u \sigma^2 u^T}\, du,$$

где σ^2 — некоторая положительно определенная матрица вторых моментов. Функцию $V_{\sigma^2}(s)$ можно интерпретировать как

$$V_{\sigma^2}(s) = \frac{(2\pi)^{k/2}}{\sqrt{|\sigma^2|}} \mathrm{M} w(s-\xi), \qquad \xi \in \Phi_{0,\sigma^{-2}}.$$

Так как

$$V_{\sigma^2}(s) = \int w(v)\, e^{-\frac{1}{2}(s-v)\sigma^2(s-v)^T}\, dv,$$

то эта функция будет аналитической функцией переменных s и σ^2.

Нам понадобятся также условия:

2) *Функция* $V_{\sigma^2}(s)$ *достигает своего минимального значения по* s *в единственной точке, которую мы обозначим* b_w.

3) $b_w = 0$.

4) *Функция* $w(t)$ *непрерывна.*

Условие 2) будет заведомо выполнено, если $w(s) \neq \mathrm{const}$ есть выпуклая вниз функция. В этом случае, очевидно, $V_{\sigma^2}(s)$ также будет выпуклой и не будет содержать «линейных» участков (т. е. матрица вторых производных будет всюду положительно определенной).

Условие 3) будет выполнено, если

$$V'_{\sigma^2}(0) = -\int u w(u)\, e^{-\frac{1}{2} u \sigma^2 u^T}\, du = 0,$$

что всегда имеет место для симметричных функций $w(u) = w(-u)$.

Значение b_w можно было бы назвать *смещением* функции потерь w. Оно удовлетворяет уравнению $V'_{\sigma^2}(b_w) = 0$. Условие 3) о том, что $b_w = 0$, не является существенным и лишь упрощает изложение, которое читатель может без труда перенести и на случай $b_w \neq 0$. Изменения в формулировках теорем, которые при этом произойдут, будут проиллюстрированы в замечании 2 к теореме 1.

Напомним теперь, во что перейдут здесь определения оптимальных стратегий §§ 2, 3. Оценка θ_Q^* будет *байесовской* относительно априорного распределения Q с плотностью q относительно меры Лебега (и функции потерь w_n), если

$$\int W\left(\theta_Q^*, t\right) q(t)\, dt = \min_{\theta^*} \int W(\theta^*, t)\, q(t)\, dt, \tag{3}$$

где $W(\theta^*, t) = \mathsf{M}_t w_n(\theta^* - t)$. Здесь интеграл в правой части (3) можно записать в виде безусловного математического ожидания $\mathsf{M} w_n(\theta^* - \theta)$, где усреднение берется по распределению с плотностью $f_t(x) q(t)$.

Оценка $\overline{\theta}^*$ будет *минимаксной*, если для любой другой оценки θ^*

$$\sup_t W\left(\bar{\theta}^*, t\right) \leqslant \sup_t W(\theta^*, t).$$

Сказанное делает естественными следующие определения, совершенно аналогичные определениям § 2.11[6].

Определение 1. Оценку θ^* мы будем называть *асимптотически байесовской*, если

$$\limsup_{n \to \infty} \left[\mathsf{M} w_n(\theta^* - \theta) - \mathsf{M} w_n\left(\theta_Q^* - \theta\right)\right] \leqslant 0, \tag{4}$$

где θ_Q^* — байесовская оценка.

Определение 2. Оценку θ_1^* мы будем называть *асимптотически минимаксной*, если для любой другой оценки θ^*

$$\limsup_{n \to \infty} \left[\sup_{t \in \Theta_0} W\left(\theta_1^*, t\right) - \sup_{t \in \Theta_0} W(\theta^*, t)\right] \leqslant 0, \tag{5}$$

где Θ_0 — любое замкнутое подмножество, лежащее внутри Θ.

При изучении асимптотически оптимальных оценок в этом параграфе мы будем оперировать лишь понятиями, введенными в определениях 1 и 2. Это составляет некоторое отличие от гл. 2[6], где присутствовали также асимптотически эффективные оценки. Их отсутствие здесь объясняется тем, что для произвольных функций потерь w мы не располагаем неравенствами типа Рао — Крамера для $\inf\limits_{\theta^* \in K_0} W(\theta^*, \theta)$ (K_0 есть класс несмещенных оценок), с помощью которого можно было по значению $W(\theta^*, \theta)$ судить о качестве θ^* и выделять, в частности, эффективные (и асимптотически эффективные) оценки, т. е. оценки, равномерно наилучшие в классе K_0.

Следующие утверждения устанавливают, что оценка максимального правдоподобия является, как и в условиях гл. 2[6], асимптотически байесовской и асимптотически минимаксной. Кроме того, мы получим *асимптотическую* нижнюю границу для функции риска при произвольной функции потерь w (неравенство Рао — Крамера дает *точную* нижнюю границу). Во всех последующих трех теоремах мы предполагаем, что выполнено условие (RR).

Теорема 1. *Пусть* $X \in \mathbf{P}_\theta$, $\hat{\theta}^*$ *есть о. м. п., а* θ_Q^* *есть байесовская оценка, соответствующая функции потерь* w (см. (2)), *удовлетворяющей условиям* 1) — 3), *и априорному распределению* $\mathbf{Q}$ *с ограниченной плотностью* q *относительно меры Лебега. Тогда*

$$|\theta_Q^* - \hat{\theta}^*| \sqrt{n} \xrightarrow[P_\theta]{} 0, \qquad (6)$$

$$(\theta_Q^* - \theta)\sqrt{n} \Leftarrow\!\!\Rightarrow \Phi_{0, I^{-1}(\theta)} \qquad (7)$$

равномерно по $\theta \in \Theta_0$, Θ_0 — *любое замкнутое подмножество, лежащее внутри* Θ, *на котором* $q(\theta) > q_0 > 0$ *непрерывна.*

Если, кроме того, функция w *удовлетворяет условию* 4), *то*

$$\mathsf{M} w_n(\theta_Q^* - \theta) = \mathsf{M} w\left(\sqrt{n}(\theta_Q^* - \theta)\right) \to \mathsf{M} w(\eta_\theta) = \\ = \mathsf{M}\frac{\sqrt{I(0)}}{(2\pi)^{k/2}} V_{I(\theta)}(0), \qquad (8)$$

где $\eta_\theta \in \Phi_{0, I^{-1}(\theta)}$, $\theta \in \mathbf{Q}$; M, *как и прежде, означает безусловное математическое ожидание с плотностью* $f_t(x)\, q(t)$ $(X \in \mathbf{P}_\theta, \theta \in \mathbf{Q})$.

Замечание 1. Наряду со сходимостью (6) можно установить также сходимость почти наверное относительно $\mathbf{P}_\theta$.

Замечание 2. Если w таково, что смещение $b_w \neq 0$, то утверждение теоремы 1 полностью сохранится, если θ_Q^* в (6), (7), (8) заменить на $\theta_Q^* - b_w/\sqrt{n}$. Таким образом, b_w имеет смысл *асимптотического смещения величины* $(\theta_Q^* - \theta)\sqrt{n}$.

Теорема 2. *Пусть функция w удовлетворяет условиям 1) — 4). Тогда для любой оценки θ^**

$$\liminf_{n\to\infty} \sup_{t\in\Theta_0} \mathsf{M}_t w_n(\theta^* - t) \geqslant \sup_{t\in\Theta_0} \mathsf{M} w(\eta_t), \tag{9}$$

$$\eta_t \in \Phi_{0, I^{-1}(t)}.$$

Любая оценка θ^*, для которой

$$\mathsf{M}_t w_n(\theta^* - t) \to \mathsf{M} w(\eta_t) \tag{10}$$

равномерно по t, является асимптотически минимаксной.

Теорема 3. *Пусть $X \in \mathbf{P}_\theta$ и функция w удовлетворяет 1)—4). Тогда оценка максимального правдоподобия $\hat{\theta}^*$ является асимптотически минимаксной и асимптотически байесовской для любого априорного распределения $\mathbf{Q}$ с непрерывной положительной в точке θ плотностью q.*

Все эти утверждения вполне аналогичны соответствующим утверждениям гл. 2[6]. Они делают правдоподобным предположение, что и для произвольной функции потерь w, удовлетворяющей 1) — 4), о. м. п. является асимптотически равномерно наилучшей в классе асимптотически несмещенных оценок (ср. с §§ 2.25[6], 2.28[6]).

Доказательство теоремы 1. Байесовская оценка определяется в силу байесовского принципа как значение θ_Q^*, обладающее свойством

$$\int w_n(\theta_Q^* - t)\, q(t/X)\, dt = \min_{u\in\Theta} \int w_n(u - t)\, q(t/X)\, dt =$$
$$= \min_{u\in\Theta} \int w\left(\sqrt{n}(u-\theta) - \sqrt{n}(t-\theta)\right) \frac{q(t) f_t(X)}{\int q(v) f_v(X)\, dv}\, dt.$$

Это значит, что в качестве $(\theta_Q^* - \theta)\sqrt{n} \equiv u_Q^*$ можно взять любое значение s, при котором достигается $\min_s U(s)$,

$$U(s)=\int w(s-v)\,q\left(\theta+\frac{v}{\sqrt{n}}\right)Z\left(\frac{v}{\sqrt{n}}\right)dv, \tag{11}$$

где, как и прежде, $Z(t)=\dfrac{f_{\theta+t}(X)}{f_\theta(X)}$.

Нам понадобятся утверждения об асимптотическом поведении $U(s)$. В §§ 2.28[6], 2.29[6] было установлено (теорема 2.28.5[6]), что при выполнении условий (RR)

$$U(u^*)=e^{Y(u^*)}q(\hat\theta^*)\left(V_{I(\theta)}(0)+\varepsilon_n(X,\theta)\right), \tag{12}$$

где $\varepsilon_n(X,\theta)\xrightarrow[P_\theta]{}0$ равномерно по θ (мы заменили здесь $\dfrac{(2\pi)^{k/2}}{\sqrt{I(\theta)}}\mathsf{M}w(\xi)$ на $V_{I(\theta)}$, $q(\theta)$ на $q(\hat\theta^*)$).

Заметим теперь, что

$$\mathbf{P}\left(\sqrt{n}\,|\theta_Q^*-\hat\theta^*|\geqslant\varepsilon\right)=\mathbf{P}\left(|u_Q^*-u^*|\geqslant\varepsilon\right)\leqslant$$
$$\leqslant\mathbf{P}\left(\min_{|s-u^*|\geqslant\varepsilon}U(s)\leqslant U(u^*)\right). \tag{13}$$

Поскольку асимптотическое представление для $U(u^*)$ мы имеем, то нам надо оценить здесь значение $U(s)$. Из теорем 2.28.4[6], 2.29.3[6] следует, что для произвольной последовательности $\delta_n\to 0$ при $|v|<\delta_n\sqrt{n}$

$$\ln Z\left(\frac{v}{\sqrt{n}}\right)=Y(u^*)-\frac{1}{2}(v-u^*)I(\theta)(v-u^*)^T(1+\varepsilon_n(X,\theta,u)),$$

$|\varepsilon_n(X,\theta,u)|\leqslant\varepsilon_n^{(1)}(X,\theta)\xrightarrow[P_\theta]{}0$ равномерно по θ. Но

$$U(s)\geqslant U_n(s)\equiv\int\limits_{|v-u^*|<\delta_n\sqrt{n}}w(s-v)\,q\left(\theta+\frac{v}{\sqrt{n}}\right)Z\left(\frac{v}{\sqrt{n}}\right)dv.$$

Рассмотрим множество

$$A_n=\left\{\varepsilon_n^{(1)}(X,\theta)<\rho,\ \inf_{|v-u^*|<\delta_n\sqrt{n}}q\left(\theta+\frac{v}{\sqrt{n}}\right)>q(\hat\theta^*)(1-\rho)\right\},$$
$$\rho>0,$$

обладающее, очевидно, свойством

$$\mathbf{P}_\theta(A_n)\to 1. \tag{14}$$

Равномерно по θ на этом множестве

$$U_n(s) \geqslant (1-\rho)\, q(\hat\theta^*)\, e^{Y(u^*)} \times$$
$$\times \int\limits_{|v-u^*|<\delta_n\sqrt{n}} w(s-v)\exp\left\{-\frac{1}{2}(v-u^*)I(\theta)(v-u^*)^T(1+\rho)\right\}dv =$$
$$= (1-\rho)\, q(\hat\theta^*)\, e^{Y(u^*)}\left[V_{I(\theta)(1+\rho)}(s-u^*) - r_n(s)\right], \quad (15)$$

где по условию 1)

$$r_n(s) = \int\limits_{|v-u^*|>\delta_n\sqrt{n}} w(s-v)\exp\left\{-\frac{1}{2}(v-u^*)I(\theta)(v-u^*)^T \times\right.$$
$$\left.\times (1+\rho)\right\}dv \leqslant e^{c\sqrt{n}d}\frac{(2\pi)^{k/2}}{\sqrt{|I(\theta)|}c}\mathbf{P}\left(|\eta| > \delta_n\sqrt{n}\right),$$
$$\eta \in \Phi_{0,I(\theta)(1+\rho)},$$

d — диаметр области Θ. Аналогично лемме 2.23.1[6] нетрудно убедиться, что

$$\mathbf{P}\left(|\eta| > \delta_n\sqrt{n}\right) \leqslant e^{-\alpha n\delta_n^2}, \quad \alpha > 0.$$

Выбирая $\delta_n = n^{-1/9}$, мы получим, что при всех s и достаточно больших n

$$r_n(s) \leqslant e^{-n^{3/4}}. \quad (16)$$

Воспользуемся теперь условиями 2), 3), в силу которых

$$\min_{|s-u^*|\geqslant\varepsilon} V_{I(\theta)}(s-u^*) \geqslant V_{I(\theta)}(0) + 4\tau, \quad \tau = \tau(\varepsilon) > 0.$$

В силу аналитических свойств $V_{\sigma^2}(s)$ мы получим, что при достаточно малых ρ

$$\min_{|s-u^*|\geqslant\varepsilon} V_{I(\theta)(1+\rho)}(s-u^*) \geqslant V_{I(\theta)}(0) + 3\tau,$$

и в силу (15), (16) при $X \in A_n$ и при достаточно больших n

$$\min_{|s-u^*|\geqslant\varepsilon} U_n(s) \geqslant (1-\rho)\, q(\hat\theta^*)\, e^{Y(u^*)}\left[V_{I(\theta)}(0) + 2\tau\right].$$

Используя (12), (13), окончательно находим

$$\mathbf{P}_\theta\left(\sqrt{n}\,|\theta_Q^* - \hat\theta^*| \geqslant \varepsilon\right) \leqslant \mathbf{P}_\theta\left(\min_{|s-u^*|\geqslant\varepsilon} U_n(s) \leqslant U(u^*)\right) \leqslant$$
$$\leqslant \mathbf{P}_\theta(X \notin A_n) + \mathbf{P}_\theta\left((1-\rho)\left[V_{I(\theta)}(0) + 2\tau\right] \leqslant\right.$$
$$\left.\leqslant V_{I(\theta)}(0) + \varepsilon_n(X, \theta)\right),$$

Выбирая дополнительно ρ столь малым, чтобы выполнялось $(1-\rho)2\tau-\rho V_{I(\theta)}(0)\geqslant\tau$, мы получим

$$\mathbf{P}(\sqrt{n}|\theta_Q^*-\hat{\theta}^*|\geqslant\varepsilon)\leqslant\mathbf{P}_\theta(X\notin A_n)+\mathbf{P}_\theta(\varepsilon_n(X,\theta)>\tau)\to 0$$

при $n\to\infty$. В силу (12), (14) утверждение (6) доказано.

Из (6) и из теорем § 2.29[6] вытекает (7). Докажем теперь соотношение (8). В силу (7) и свойства 4)

$$w(\sqrt{n}(\theta_Q^*-\theta))\Rightarrow w(\eta_\theta),\quad \eta_\theta\in\Phi_{0,I^{-1}(\theta)}.$$

По лемме Фату

$$\liminf_{n\to\infty}\mathsf{M}_t w(\sqrt{n}(\theta_Q^*-t))\geqslant\mathsf{M}w(\eta_t),$$

$$\liminf_{n\to\infty}\mathsf{M}w(\sqrt{n}(\theta_Q^*-\theta))\geqslant\int q(t)\,\mathsf{M}w(\eta_t)\,dt=\mathsf{M}w(\eta_\theta).$$

С другой стороны, по определению θ_Q^*

$$\mathsf{M}w(\sqrt{n}(\theta_Q^*-(\theta))\leqslant\mathsf{M}w(\sqrt{n}(\hat{\theta}^*-\theta))\to\mathsf{M}w(\eta_\theta).$$

Последнее соотношение вытекает из равномерной сходимости $\mathsf{M}_t w(\sqrt{n}(\hat{\theta}^*-t))\to\mathsf{M}w(\eta_t)$, доказанной в § 2.29[6]. Теорема доказана.

Доказательство теоремы 2. Возьмем распределение Q, сосредоточенное на Θ_0, с ограниченной плотностью $q(t)>0$ при $t\in\Theta_0$, и пусть θ_Q^* — байесовская оценка, отвечающая Q. Тогда для любой оценки θ^*

$$\sup_{t\in\Theta_0}\mathsf{M}_\theta w_n(\theta^*-t)\geqslant\int_{\Theta_0}\mathsf{M}_t w_n(\theta^*-t)\,q(t)\,dt\geqslant$$
$$\geqslant\int_{\Theta_0}\mathsf{M}_t w_n(\theta_Q^*-t)\,q(t)\,dt=\mathsf{M}w_n(\theta_Q^*-\theta).$$

По лемме Фату в силу (8)

$$\liminf_{n\to\infty}\sup_{t\in\Theta_0}\mathsf{M}_t w_n(\theta^*-t)\geqslant\liminf_{n\to\infty}\mathsf{M}w_n(\theta_Q^*-\theta)\geqslant\mathsf{M}w(\eta_\theta)=$$
$$=\int_{\Theta_0}\mathsf{M}w(\eta_t)q(t)\,dt.$$

Так как функция $\mathsf{M}w(\eta_t)=\frac{\sqrt{I(t)}}{(2\pi)^{k/2}}V_{I(t)}(0)$ непрерывна относительно t, то выбором $q(t)$ интеграл

$$\int\limits_{\Theta_0} \sqrt{I(t)}\, V_{I(t)}(0)\, q(t)\, dt$$

может быть сделан сколь угодно близким к $\sup\limits_{t\in\Theta_0} \sqrt{I(t)}\, V_{I(t)}(0) = \sup\limits_{t\in\Theta_0} \mathsf{M} w(\eta_t)$. Это доказывает (9).

Пусть теперь оценка θ_1^* обладает свойством (10) и θ^* — любая другая оценка. Тогда в силу (9) и равномерности сходимости (10)

$$\limsup_{n\to\infty}\left[\sup_{t\in\Theta_0} \mathsf{M}_t w_n\left(\theta_1^* - t\right) - \sup_{t\in\Theta_0} \mathsf{M}_t w_n\left(\theta^* - t\right)\right] \leqslant$$

$$\leqslant \sup_{t\in\Theta_0} \lim_{n\to\infty} \mathsf{M}_t w_n\left(\theta_1^* - t\right) - \sup_{t\in\Theta_0} \mathsf{M} w(\eta_t) = 0.$$

Неравенство (5) определения асимптотической минимаксности, а вместе с ним и теорема 2 доказаны.

Доказательство теоремы 3. Асимптотическая минимаксность $\hat{\theta}^*$ следует из того, что для о. м. п. $\hat{\theta}^*$ по теореме 2.29.4[6] справедливо (10).

Асимптотическая байесовость $\hat{\theta}^*$ следует из того, что при $\theta^* = \hat{\theta}^*$ выполнено (4), ибо для $\hat{\theta}^*$ имеет место равномерная сходимость (10) и, стало быть,

$$\lim_{n\to\infty} \mathsf{M} w_n\left(\hat{\theta}^* - \theta\right) = \lim_{n\to\infty} \int \mathsf{M}_t w_n\left(\hat{\theta}^* - \theta\right) q(t)\, dt =$$

$$= \mathsf{M} w(\eta_0) = \lim_{n\to\infty} \mathsf{M} w_n\left(\theta_Q^* - \theta\right).$$

Последнее равенство вытекает из (8). Теорема доказана.

Утверждение теоремы 1 можно усилить, если потребовать дополнительно, чтобы функция $w(t)$ росла достаточно быстро. Именно, обозначим $w_N = \min\limits_{|t|>N} w(t)$, $W_M = \max\limits_{|t|\leqslant M} w(t)$ и рассмотрим условие

5) *Существует $\gamma < 1$ такое, что $w_N > 2W_{\gamma N}$ при всех достаточно больших N.*

Если $w(t)$ растет при $|t| \to \infty$ как степенная или экспоненциальная функция, то условие 5) выполнено.

Теорема 4. *Если выполнены условия* 1), 5), *$q(t) > q_0 > 0$ на замкнутом множестве Θ_0 и $q(t) \leqslant q_m < \infty$,* *то при некоторых* $c < \infty$, $\alpha > 0$, *не зависящих от* t,

$$\mathrm{P}_t\left(\sqrt{n}\left(\theta_Q^* - t\right) > N\right) \leqslant ce^{-\alpha N^2}, \qquad t \in \Theta_0.$$

Отсюда и из теоремы 1 следует, что для любой непрерывной функции $v(t)$ такой, что $|v(t)| \leqslant e^{-\alpha N^2/2}$, справедливо

$$\mathsf{M}_t v\left(\sqrt{n}\left(\theta_Q^* - t\right)\right) \to \mathsf{M} v(\eta_t), \quad t \in \Theta_0.$$

Обозначим

$$u(r) = \int\limits_{|v| \geqslant r} w(-v)\, q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv$$

(это есть часть интеграла $U(0)$ по области $|v| \geqslant r$). Для доказательства теоремы 4 нам понадобится

Лемма 1. *Если* $w(t)$ *удовлетворяет условию* 1), $q_m = \max\limits_u q(u) < \infty$, *то при некоторых* $\beta > 0$, $a < \infty$, *не зависящих от* θ, *и всех* $0 < \delta < 1$

$$\mathrm{P}_\theta(u(r) > \delta) \leqslant \frac{a}{\delta} e^{-\beta r^2}.$$

Это неравенство сохраняется при $w(t) \equiv 1$.

Доказательство. Имеем

$$\mathrm{P}_\theta(u(r) > \delta) \leqslant \mathrm{P}_\theta\left(\sup_{|v| \geqslant r} Z\left(\frac{v}{\sqrt{n}}\right) > 1\right) + \mathrm{P}_\theta\left(u(r) > \delta, \sup_{|v| \geqslant r} Z\left(\frac{v}{\sqrt{n}}\right) \leqslant 1\right).$$

Оценка для первого слагаемого дана в теореме 2.23.2[6], в силу которой это слагаемое не превосходит $c_1 e^{-r^2\beta'}$, $\beta > 0$. Второе слагаемое не превосходит

$$\mathrm{P}_\theta\left(\int\limits_{|v| \geqslant r} w(-v)\, q\left(\theta + \frac{v}{\sqrt{n}}\right) Z^{1/2}\left(\frac{v}{\sqrt{n}}\right) dv > \delta\right). \tag{17}$$

Так как в силу теоремы 2.23.1[6]

$$\mathsf{M}_\theta Z^{1/2}\left(\frac{v}{\sqrt{n}}\right) \leqslant e^{-2|v|^2\beta}, \quad \beta > 0,$$

то математическое ожидание интеграла в (17) не будет превосходить (см. лемму 2.23.1[6])

$$q_m \int\limits_{|v| \geqslant r} e^{c|v|} e^{-2|v|^2\beta} dv \leqslant c_2 e^{-r^2\beta}.$$

Поэтому в силу неравенства Чебышева вероятность (17) не превосходит $c_2 e^{-r^2\beta}/\delta$. Лемма доказана.

Обозначим через $u_1(r)$ значение интеграла $u(r)$ при $w(t) \equiv 1$:

$$u_1(r) = \int\limits_{|v| \geqslant r} q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv.$$

Лемма 2. *Если $q(\theta) > 0$ на замкнутом множестве Θ_0, то при некотором $b < \infty$, не зависящем от θ, любом $\varepsilon > 0$ и всех достаточно больших n*

$$\mathrm{P}_\theta(u_1(0) < \varepsilon) \leqslant b\varepsilon^2, \ \theta \in \Theta_0.$$

Доказательство. При всех достаточно больших n имеем

$$u_1(0) \geqslant \int\limits_{|v| \leqslant 1} q\left(\theta + \frac{v}{\sqrt{n}}\right) Z\left(\frac{v}{\sqrt{n}}\right) dv \geqslant$$

$$\geqslant q_0 \int\limits_{|v| \leqslant 1} \exp\left\{L\left(X, \theta + \frac{v}{\sqrt{n}}\right) - L(X, \theta)\right\} dv =$$

$$= q_0 \int\limits_{|v| \leqslant 1} \exp\left\{(v, \zeta_n) + \frac{1}{2} v\gamma_n v^T\right\} dv,$$

где

$$q_0 = \min_{t \in \Theta_0} q(t) > 0, \quad \zeta_n = \frac{1}{\sqrt{n}} L'(X, t), \quad \gamma_n = \frac{1}{n} \left\| L''_{ij}(X, \tilde{\theta}) \right\|,$$

$\tilde{\theta} = \theta + \rho v n^{-1/2}$, $|\rho| \leqslant 1$. (Здесь L' есть вектор производных логарифмической функции правдоподобия, L''_{ij} — частные производные второго порядка.) Так как $|(v, \zeta_n)| \leqslant |v|\,|\zeta_n|$ и так как в силу условий (RR)

$$\left| v\gamma_n v^T \right| \leqslant \frac{1}{n} \sum_{i=1}^{n} l(x_i) \sum_{l,\, j=1}^{k} |v_l v_j| \leqslant \frac{k\,|v|^2}{n} L_n,$$

где $L_n = \sum\limits_{i=1}^{n} l(x_i)$, то на множестве $A = \{|\zeta_n| \leqslant 1/\varepsilon,\ L_n \leqslant \leqslant n/\varepsilon^2 k\}$ справедливо

$$u_1(0) \geqslant q_0 \int\limits_{|v| \leqslant 1} \exp\left\{-\frac{|v|}{\varepsilon} - \frac{|v|^2}{2\varepsilon^2}\right\} dv \geqslant$$

$$\geqslant q_0 \varepsilon \int\limits_{|s| \leqslant \varepsilon^{-1}} \exp\left\{-|s| - \frac{|s|^2}{2}\right\} ds \geqslant c_1 \varepsilon.$$

Это означает, что имеет место вложение $\{u_1(0) < c_1\varepsilon\} \subset \bar{A}$. Так как

$$\mathbf{P}_\theta(\bar{A}) \leqslant \mathbf{P}_\theta\left(|\zeta_n| \geqslant \varepsilon^{-1}\right) + \mathbf{P}_\theta\left(L_n > \frac{n}{\varepsilon^2 k}\right) \leqslant$$

$$\leqslant \varepsilon^2 \mathsf{M}_\theta |\zeta_n|^2 + \frac{\varepsilon^2 k}{n} \mathsf{M}_\theta L_n,$$

$$\mathsf{M}_\theta |\zeta_n|^2 = \sum_{i=1}^{k} I_{ii}(\theta), \quad \mathsf{M}_\theta L_n = n \mathsf{M}_\theta l(x_1),$$

то

$$\mathbf{P}_\theta(\bar{A}) \leqslant c_2 \varepsilon^2.$$

Лемма доказана.

Доказательство теоремы 4. Обозначим через M_n множество точек s, в которых достигается $\min U(s)$ (т. е. множество точек $(\theta_Q^* - \theta)\sqrt{n}$; см. (11)) *). Тогда

$$\{M_n \subset D\} = \left\{\min_{s \in D} U(s) < \min_{s \notin D} U(s)\right\}. \tag{18}$$

Стало быть,

$$\left\{\sqrt{n}\,|\theta_Q^* - \theta| > 2N\right\} =$$

$$= \left\{\min_{|s| > 2N} U(s) < \min_{|s| \leqslant 2N} U(s)\right\} \subset \left\{\min_{|s| \geqslant 2N} U(s) < U(0)\right\}.$$

Здесь

$$\min_{|s| \geqslant 2N} U(s) \geqslant w_N \int_{|u| < N} q\left(\theta + \frac{u}{\sqrt{n}}\right) Z\left(\frac{u}{\sqrt{n}}\right) du =$$

$$= w_N (u_1(0) - u_1(N)),$$

$$w_N \equiv \min_{\substack{|s| > 2N \\ |u| < N}} w(s - u) = \min_{|t| > N} w(t).$$

Далее,

$$U(0) = \int w(-u)\, q\left(\theta + \frac{u}{\sqrt{n}}\right) Z\left(\frac{u}{\sqrt{n}}\right) du \leqslant$$

$$\leqslant (u_1(0) - u_1(M)) W_M + u(M),$$

где $W_M = \max_{|t| \leqslant M} w(t)$.

Отсюда получаем

$$\left\{\sqrt{n}\,|\theta_Q^* - \theta| > 2N\right\} \subset \{w_N (u_1(0) - u_1(N)) <$$

$$< W_M (u_1(0) - u_1(M)) + u(M)\} \subset$$

$$\subset \left\{\left(\frac{w_N}{W_M} - 1\right) u_1(0) < \frac{u(M)}{W_M} + \frac{u_1(N)\, w_N}{W_M} + u_1(M)\right\}.$$

---

*) Вместо M_n можно было бы рассматривать, например, наименьшую по норме точку, в которой достигается $\min U(s)$.

В силу условия 5) выберем $M = \gamma N$, $\gamma < 1$, так, чтобы $w_N > 2W_M$ при всех достаточно больших N. Кроме того, воспользуемся неравенствами $W_M > 2$ (при достаточно больших M), $w_N < w(N) < e^{cN}$. Тогда, очевидно,

$$\{\sqrt{n}\,|\theta_Q^* - \theta| > 2N\} \subset \{u_1(0) < u(\gamma N) + u_1(N)\,e^{cN}\}. \tag{19}$$

В силу леммы 1 находим

$$\mathbf{P}_\theta\left(u(\gamma N) > \frac{1}{2}\,e^{-\alpha N^2}\right) \leqslant 2ae^{-\beta N^2\gamma^2+\alpha N^2},$$

$$\mathbf{P}_\theta\left(u_1(N) > \frac{1}{2}\,e^{-cN-\alpha N^2}\right) \leqslant 2ae^{-\beta N^2+\alpha N^2+cN}.$$

Выбирая $\alpha < \frac{1}{2}\beta\gamma^2$, мы получим, что при достаточно больших N из (19) следует

$$\mathbf{P}_\theta\left(\sqrt{n}\,|\theta_Q^* - \theta| > 2N\right) \leqslant 4ae^{-\alpha N^2} + \mathbf{P}_\theta\left(u_1(0) < e^{-\alpha N^2}\right).$$

Нам остается лишь воспользоваться леммой 2, в силу которой

$$\mathbf{P}_\theta\left(u_1(0) < e^{-\alpha N^2}\right) \leqslant be^{-2\alpha N^2}.$$

Теорема доказана.

§ 7. Оптимальные статистические критерии при произвольной функции потерь. Критерий отношения правдоподобия как асимптотически байесовское решение

1. Свойства оптимальности статистических критериев при произвольной функции потерь. Мы видели в предыдущих двух параграфах, что многие главные результаты теории оценивания качественно сохраняются при переходе к более общим задачам статистического решения с потерями $w(\delta, \theta)$, $\delta \in D \subset R^k$, $\theta \in \Theta \subset R^k$, отличными от квадратичных.

Такая же картина имеет место и в теории проверки гипотез. Мы видели в § 4, что оптимальные решающие правила для игр с конечными множествами D и Θ и произвольной функцией потерь имеют ту же форму, что и оптимальные критерии для проверки конечного числа простых гипотез, рассмотренные в § 3.1. Результаты §§ 3.5[6]—3.7[6], 3.9[6], 3.11[6], 3.13[6]—3.15[6] также в основном сохранятся. В частности, теоремы о р. н. м. к. в

§§ 3.5[6]—3.7[6] перейдут в утверждения о равномерно наилучших стратегиях в соответствующих статистических играх ($\Theta \subset R^k$, $D=\{\delta_1, \delta_2\}$ двухточечно), в которых, однако, функция потерь $w(\delta_i, \theta)=w_i(\theta)$, $w_i(\theta)=0$ при $\theta \in \Theta_i$, $i=1, 2$, уже будет не обязательно статистической ($w_i(\theta)=1$ при $\theta \notin \Theta_i$), а будет лишь удовлетворять некоторым весьма общим условиям (например, свойствам монотонного возрастания $w_i(\theta)$ при удалении θ от Θ_i). Роль классов K_ε, в которых мы искали р. н. м. к., будут играть теперь классы решающих функций $\pi(X)$ с фиксированным максимальным значением ε «потерь первого рода»:

$$\varepsilon = \sup_{\theta \in \Theta_1} W(\pi(\cdot), \theta) = \sup_{\theta \in \Theta_1} w_2(\theta) \mathsf{M}_\theta \pi(X, \delta_2). \tag{1}$$

Минимизироваться будет значение «потерь второго рода»:

$$W(\pi(\cdot), \theta) = w_1(\theta) \mathsf{M}_\theta \pi(X, \delta_1) \text{ при } \theta \in \Theta_2. \tag{2}$$

Здесь $\pi(X, \delta_i)$ означает вероятность принятия решения δ_i критерием π. Для сокращения записи мы положим, следуя гл. 3[6], $\pi(X, \delta_2)=\pi(X)$, так что $\pi(X, \delta_1)=1-\pi(X)$. Обозначение одним символом $\pi(X)$ критерия и числа $\pi(X, \delta_2)$ удобно и, как мы видели в [6], к недоразумениям не приводит.

В (1), (2) ищутся экстремумы выражений, которые отличаются от соответствующих выражений для статистических функций потерь лишь не зависящими от $\pi(X)$ множителями. Если эти множители обладают естественным свойством монотонности, то изложение §§ 3.5[6]—3.7[6], 3.9[6], 3.11[6] при переходе к задаче, определяемой (1), (2), существенно не изменится.

По существу мало изменятся и результаты асимптотического характера в §§ 3.13[6]—3.15[6]. В этом параграфе мы остановимся более подробно на обобщении на случай произвольной функции потерь результатов § 3.13[6] и убедимся, что это обобщение действительно не требует никаких дополнительных усилий.

2. К. о. п. как асимптотически байесовский критерий. Рассмотрим статистическую игру $(\mathscr{D}, \Theta, W)$, в которой множество Θ континуально и есть выпуклое компактное множество в R^k, а множество D стратегий статистика двухточечно: $D=\{\delta_1, \delta_2\}$. Функция потерь $w(\delta, \theta)$

имеет вид

$$w(\delta_1, \theta) = \begin{cases} 0, & \theta = \theta_1, \\ w_1(\theta), & \theta \neq \theta_1, \end{cases}$$

$$w(\delta_2, \theta) = \begin{cases} w_2, & \theta = \theta_1, \\ 0, & \theta \neq \theta_1, \end{cases}$$

где θ_1 — фиксированная внутренняя точка Θ. При $w_2 = w_1(\theta) \equiv 1$ это соответствует задаче проверки простой гипотезы $H_1 = \{\theta = \theta_1\}$ против дополнительной альтернативы $H_2 = \{\theta \neq \theta_1\}$.

Чтобы, пользуясь байесовским принципом, найти вид байесовского решения, рассмотрим обычную игру (D, Θ, w) и предположим, что на Θ задано распределение $\mathbf{Q}$ такое, что $q = \mathbf{Q}(\{\theta_1\}) > 0$ (полная байесовская постановка задачи). Обозначим $\mathbf{Q}_2 = \dfrac{\mathbf{Q} - q\mathbf{I}_{\theta_1}}{1-q}$, где $\mathbf{I}_\theta$ есть вырожденное распределение, сосредоточенное в точке θ. Тогда

$$\tilde{w}(\delta_1, \mathbf{Q}) = (1-q)\int w_1(t)\,\mathbf{Q}_2(dt), \qquad \tilde{w}(\delta_2, \mathbf{Q}) = qw_2.$$

Это означает, что байесовская стратегия $\pi_{\mathbf{Q}}(\delta_2) = 1$, если

$$(1-q)\int w_1(t)\,\mathbf{Q}_2(dt) > qw_2, \tag{3}$$

и $\pi_{\mathbf{Q}}(\delta_1) = 1$, если имеет место обратное неравенство. Соотношение (3) можно записать в виде

$$\int w(t)\,\mathbf{Q}(dt) > 0,$$

где

$$w(t) = \begin{cases} w_1(t) & \text{при } t \neq \theta_1, \\ -w_2 & \text{при } t = \theta_1. \end{cases}$$

В силу байесовского принципа байесовское решающее правило $\pi_{\mathbf{Q}}(X)$ имеет вид $\pi_{\mathbf{Q}}(X) = 1$, если

$$\int w(t)\,\mathbf{Q}_X(dt) > 0,$$

где $\mathbf{Q}_X$ — апостериорное распределение. Пусть $\lambda(dt) = dt$ при $t \neq \theta_1$, $\lambda(\{\theta_1\}) = 1$ и распределение $\mathbf{Q}_2$ имеет плотность $q_2(t)$ относительно меры Лебега. Тогда априорное распределение $\mathbf{Q}$ имеет плотность $q(t)$ относительно λ, равную $(1-q)q_2(t)$ при $t \neq \theta_1$ и $q(t) = q$ при $t = \theta_1$. Это означает, что апостериорная плотность относительно меры λ будет равна

$$q(t/X) = \frac{f_t(X)\,q(t)}{f(X)},$$

$$f(X) = \int f_u(X)\,q(u)\,\lambda(du).$$

Следовательно, байесовское решающее правило $\pi_Q(X)$ имеет вид $\pi_Q(X) = 1$, если

$$(1-q)\int w_1(t)\,f_t(X)\,q_2(t)\,dt > w_2 q f_{\theta_1}(X). \tag{4}$$

Риск этого правила равен

$$\widetilde{W}(\pi_Q(\cdot), Q) = qw_2 P_{\theta_1}(\pi_Q(X)=1) + \\ + (1-q)\int w_1(u)\,q_2(u)\,P_u(\pi_Q(X)=0)\,du.$$

Сравнивая эти соотношения с содержанием § 3.13[6], мы видим, что область (4) приема решения δ_2 имеет здесь тот же вид, что и область $\Omega(c)$ в (3.13.3)[6] при $c = w_2 q/(1-q)$ и при замене функции $q(t)$ в (3.13.3)[6] на $w_1(t)q_2(t)$. Другими словами,

$$\pi_Q(X) = \begin{cases} 1, & \text{если } r_{Q_2}(X) > c, \\ \gamma, & \text{если } r_{Q_2}(X) = c, \\ 0, & \text{если } r_{Q_2}(X) < c, \end{cases} \tag{5}$$

где

$$r_{Q_2}(X) = \frac{\int w_1(t)\,q_2(t)\,f_t(X)\,dt}{f_{\theta_1}(X)}, \quad c = \frac{w_2 q}{1-q}.$$

Далее, мы можем, следуя рассуждениям § 3.13[6], поступить следующим образом. Из совокупности байесовских правил (5) путем изменения числа q выбрать такое решение $\pi_Q(X)$, которое бы имело фиксированное значение «потерь первого рода»:

$$w_2\left[P_{\theta_1}(\pi_Q(X)=1) + \gamma P_{\theta_1}(\pi_Q(X)=\gamma)\right] = \alpha.$$

Тогда среди всех правил $\pi(X)$, для которых

$$\alpha_1(\pi) = w_2 M_{\theta_1}\pi(X) \leqslant \alpha, \tag{6}$$

решение $\pi_Q(X)$ будет минимизировать «потери второго рода», равные

$$\alpha_2(\pi) = \int w_1(u)\, q_2(u)\, \mathrm{M}_u (1 - \pi(X))\, du. \tag{7}$$

Это есть прямое следствие байесовости решения π_Q. Сравнение значений (6), (7) с величинами вероятностей ошибок первого и второго рода (3.13.4)[6] показывает, что мы снова имеем дело с несущественными отличиями, главное из которых состоит в замене функции $q(u)$ в (3.13.4)[6] на функцию $w_1(u)q_2(u)$. Числа c и γ в (5) определяются по α.

Сказанное позволяет нам, в точности следуя рассуждениям § 3.13[6], сформулировать следующие определения и утверждения.

Определение 1. Решающее правило $\pi(X)$ принадлежит классу $\bar{K}_\varepsilon$ (имеет асимтотический уровень $1-\varepsilon$), если

$$\limsup_{n\to\infty} \mathrm{M}_{\theta_1}\pi(X) \leqslant \varepsilon.$$

Это определение по существу ничем не отличается от определения 3.13.1[6].

Покажем теперь, что выбором q мы можем добиться того, чтобы $\pi_Q \in \bar{K}_\varepsilon$. Положим

$$r_{Q_2}(X) \equiv \frac{\int w_1(t)\, q_2(t)\, f_t(X)\, dt}{f_{\theta_1}(X)} = \left(\frac{2\pi}{n}\right)^{k/2} \frac{w_1(\theta_1)\, q_2(\theta_1)}{\sqrt{|I|}}\, e^{T(X)},$$

где $I = I(\theta_1)$ есть информационная матрица Фишера в точке θ_1. Предположим, далее, что выполнены условия (RR) и что θ_1 есть внутренняя точка Θ, функция $w_1(t)q_2(t)$ непрерывна в точке θ_1 и положительна,

$$c = \left(\frac{2\pi}{n}\right)^{k/2} \frac{w_1(\theta_1)\, q_2(\theta_1)}{\sqrt{|I|}}\, e^{z}. \tag{8}$$

Тогда в силу леммы 3.13.1[6] для функции $p_t(c) = \mathrm{P}_t(r_{Q_2}(X) > c)$ получаем

$$p_{\theta_1}(c) = \mathrm{P}_{\theta_1}(T(X) > z) \to \mathrm{H}_k((2z, \infty)).$$

Следовательно, полагая $q = c/(c + w_2)$, где c определено в (8), $z = h_\varepsilon/2$, h_ε есть квантиль порядка $1-\varepsilon$ распределения χ^2 с k степенями свободы, мы получим

$$\lim_{n\to\infty} p_{\theta_1}\left(\frac{w_2 q}{1-q}\right) = \varepsilon,$$

и, стало быть, $\pi_Q(X) \in \bar{K}_\varepsilon$.

Определение 2. Для данного априорного распределения Q решающее правило $\pi(X)$ называется *асимптотически байесовским в* $\bar{K}_\varepsilon$, если $\pi_Q \in \bar{K}_\varepsilon$,

$$\limsup_{n\to\infty} \frac{\alpha_2(\pi)}{\alpha_2(\pi_Q)} = 1.$$

Теорема 1. *Пусть выполнены условия* (RR), *точка* θ_1 *является внутренней точкой* Θ. *Тогда в* $\bar{K}_\varepsilon$ *существует асимптотически байесовское решающее правило* $\hat{\pi}(X)$ — *одно и то же для любых распределений* Q_2 *и любых функций* $w_1(t)$ *таких, что функция* $w_1(t)q_2(t)$ *непрерывна и положительна в точке* θ_1 *и ограничена на* Θ. *Критерий* $\hat{\pi}$ *определяется соотношением*

$$\hat{\pi}(X) = 1, \text{ если } \frac{f_{\hat{\theta}^*}(X)}{f_{\theta_1}(X)} > e^{h_\varepsilon/2}. \tag{9}$$

Теорема доказывается точно так же, как теорема 3.13.1[6], с точностью до замены функции $q(t)$ на $w_1(t)q_2(t)$. Теорема 3.13.1[6] позволяет также найти значение «потерь второго рода» (см. (7)) критерия $\hat{\pi}$.

Критерий (9) есть не что иное, как критерий отношения правдоподобия.

§ 8. Асимптотически оптимальные решения при произвольной функции потерь в случае близких гипотез

В этом параграфе мы рассмотрим обобщение на случай произвольной функции потерь результатов § 3.14[6]. Это обобщение будет более содержательным, чем в предыдущем параграфе, так как функции потерь будут зависящими от n (ср. с § 6).

Пусть $(\mathscr{D}, \Theta, W)$ — статистическая игра, в которой $\Theta \subset R^k$, множество $D = \{\delta_1, \delta_2\}$ двухточечно, $w(\delta_i, \theta) = w_i(\theta)$, где $w_i(\theta) = 0$ при $\theta \in \Theta_i$, $i = 1, 2$, пересечение $\Theta_1 \cap \Theta_2$ пусто.

Если $w_i(\theta) = 1$ при $\theta \notin \Theta_i$, то мы получим задачу о проверке гипотез $H_i = \{\theta \in \Theta_i\}$, $i = 1, 2$.

Найдем байесовскую стратегию для игры (D, Θ, w). Пусть Q_i — распределения на Θ_i,

$$Q = q_1Q_1 + q_2Q_2, \quad q_1 + q_2 = 1.$$

Тогда, очевидно, $\widetilde{w}(\delta_i, Q) = \int w_i(t)\,Q(dt)$ и $\pi_Q(\delta_2) = 1$, если

$$\int w_2(t)\,Q(dt) < \int w_1(t)\,Q(dt),$$

или

$$q_1 \int w_2(t)\,Q_1(dt) < q_2 \int w_1(t)\,Q_2(dt).$$

Следовательно, в силу байесовского принципа байесовское решающее правило $\pi_Q(X)$ будет иметь вид $\pi_Q(X) = 1$, если

$$\int w_2(t)\,Q_X(dt) < \int w_1(t)\,Q_X(dt). \qquad (1)$$

Допустим, что распределения Q_i имеют плотности $q_i(t)$, $i = 1, 2$, относительно меры λ. Тогда Q и апостериорное распределение Q_X будут иметь плотности соответственно $q(t) = q_1 q_1(t) + q_2 q_2(t)$ и

$$q(t/X) = \frac{q(t)\,f_t(X)}{f(X)}, \quad f(X) = \int q(u)\,f_u(X)\,\lambda(du).$$

Это значит, что соотношение (1) можно переписать в виде

$$q_1 \int_{\Theta_1} w_2(t)\,q_1(t)\,f_t(X)\,\lambda(dt) < q_2 \int_{\Theta_2} w_1(t)\,q_2(t) f_t(X)\,\lambda(dt). \qquad (2)$$

Риск байесовского правила $\pi_Q(X)$ равен

$$W(\pi_Q(\cdot), \theta) = w_1(\theta)\mathsf{M}_\theta \pi_Q(X) + w_2(\theta)(1 - \mathsf{M}_\theta \pi_Q(X)),$$

$$\widetilde{W}(\pi_Q(\cdot), Q) = \int W(\pi_Q(\cdot), t)\,q(t)\,\lambda(dt).$$

Перейдем теперь к рассмотрению *близких альтернатив*. Пусть θ_1 — какое-нибудь фиксированное значение параметра θ. Как и в § 3.14[6], мы будем предполагать, что множества Θ_i имеют вид

$$\Theta_i = \theta_1 + \Gamma_i/\sqrt{n}, \qquad (3)$$

где Γ_i от n не зависят. Относительно Q_i мы будем предполагать, что они индуцированы некоторыми распределениями Π_i, сосредоточенными на Γ_i и от n не зависящими. Если множества Γ_i ограничены, то стратегии природы θ располагаются в $1/\sqrt{n}$-окрестности точки θ_1. Поэтому, если $w_1(t)$, $w_2(t)$ непрерывны, $w_i(t) > c > 0$, $i = 1, 2$, соответственно на множествах Θ_2, Θ_1, то статистическая игра $(\mathscr{D}, \Theta, W)$ для такой функции потерь практически не будет отличаться по своим свойствам от игры со статистической функцией потерь $w_i(t) = 1$ при $t \notin \Theta_i$, рассмотренной нами в §§ 3.14[6], 3.15[6].

Мы рассмотрим здесь более содержательное обобщение, аналогичное тому, которое проводилось нами в § 6. Будем считать, что функция потерь $w(\delta_i, \theta) = w_i(\theta)$ зависит от n так, что

$$w_i(\theta) = w_{i,\,n}(\theta) = v_i(\sqrt{n}(\theta - \theta_1)), \tag{4}$$

где $v_i(t)$ — ограниченные измеримые функции, не зависящие от n.

Следуя § 3.14[6], мы будем задачу отыскания по выборке $X \in \mathbf{P}_\theta$ решения игры $(\mathscr{D}, \Theta, W)$, описанной выше, называть *задачей* А. Если выполнено (3), (4), то мы будем говорить о задаче А для близких гипотез с функциями потерь $v_i(t)$.

Рассмотрим теперь другую статистическую игру $(\mathscr{D}_B, \Gamma, V)$, относящуюся к выборке $Y \in \Phi_{\gamma, I^{-1}}$ единичного объема, где $I = I(\theta_1)$ есть информационная матрица Фишера для семейства $\mathbf{P}_\theta$ в точке θ_1. Эта игра имеет двухточечное множество решений $D_B = (d_1, d_2)$ и множество стратегий природы (параметрическое множество) $\Gamma = \Gamma_1 \cup \cup \Gamma_2$. Функция потерь $v(d, \gamma)$: $D_B \times \Gamma \to R$ определяется соотношениями

$$v(d_i, \gamma) = \bar{v}_i(\gamma), \quad v_i(\gamma) = 0 \text{ при } \gamma \in \Gamma_i.$$

Таким образом, в этой игре $\mathscr{D}_B$ есть класс всех решений $d(Y)$: $\mathscr{Y} = R^k \to D_B$,

$$V(d(\cdot), \gamma) = v_1(\gamma)\, \Phi_{\gamma, I^{-1}}(d(Y) = d_1) + v_2(\gamma)\, \Phi_{\gamma, I^{-1}}(d(Y) = d_2)$$

(одно из слагаемых в правой части равно нулю). Аналогично выписываются потери для рандомизированных стратегий $\pi(Y)$ в терминах $\mathsf{M}\pi(Y)$, $Y \in \Phi_{\gamma, I^{-1}}$. Сформулированную задачу мы будем называть *задачей B*.

Между задачами А и В здесь существует то же соотношение, которое было установлено между этими задачами в § 3.14[6]. Пусть $\pi(Y)$ есть решение задачи B, оптимальное в том или ином смысле (байесовское, минимаксное). И пусть $\hat{\theta}^*$ есть о. м. ц. в задаче А, $\gamma^* = (\hat{\theta}^* - \theta_1)\sqrt{n}$. Тогда решение $\pi(\gamma^*)$ будет асимптотически оптимальным (в том же смысле) решением задачи А.

Сформулированный «предельный признак оптимальности» позволяет редуцировать задачу А к более простой задаче В.

Чтобы придать сказанному точный смысл, дадим следующие определения. Пусть на Γ_i заданы распределения Π_i. Положим $\Pi = q_1\Pi_1 + q_2\Pi_2$, $q_1 + q_2 = 1$, и обозначим через Q распределение на Θ, индуцированное распределением Π и преобразованием $\theta = \theta_1 + \gamma/\sqrt{n}$.

Определение 1. Решение $\pi_1(X)$ называется *асимптотически байесовским*, если

$$\limsup_{n\to\infty} [\widetilde{W}(\pi_1(\cdot), Q) - \widetilde{W}(\pi_Q(\cdot), Q)] \leqslant 0.$$

Здесь, как и прежде,

$$\widetilde{W}(\pi(\cdot), \theta) = w_1(\theta) \mathsf{M}_\theta \pi(X) + w_2(\theta)(1 - \mathsf{M}_\theta \pi(X)),$$

$$\widetilde{W}(\pi(\cdot), \theta) = \int \widetilde{W}(\pi(\cdot), t) Q(dt),$$

π_Q — байесовское решающее правило.

Определение 2. Решение $\pi_1(X)$ называется *асимптотически минимаксным*, если для любого другого решения $\pi(X)$

$$\limsup_{n\to\infty} \left[\sup_{\theta\in\Theta} \widetilde{W}(\pi_1(\cdot), \theta) - \sup_{\theta\in\Theta} \widetilde{W}(\pi(\cdot), \theta)\right] \leqslant 0.$$

Здесь можно было бы сравнивать π_1 лишь с минимаксным правилом $\bar{\pi}$ (ср. с определением 1).

Аналогично § 3.14[6], мы могли бы рассматривать также асимптотически байесовские и минимаксные решения в классе $\bar{K}_\varepsilon$ решений фиксированных асимптотических «потерь первого рода»:

$$\varepsilon = \limsup_{n\to\infty} \sup_{\theta\in\Theta_1} w_1(\theta) \mathsf{M}_\theta \pi(X).$$

Чтобы получить соответствующие результаты, достаточно будет сравнить содержание этого параграфа с § 3.14[6].

Обозначим $\pi_\Pi(Y)$ байесовское решение игры $(\mathscr{D}_B, \Gamma, V)$ (т. е. задачи В), соответствующее априорному распределению Π, и предположим для простоты, что множества Γ_i ограничены.

Теорема 1. *Пусть в окрестности точки θ_1 выполнены условия (RR), а функции v_i и распределение Π_i таковы, что* $0 < \int v_1(u) \Pi_2(du) < \infty$, $0 < \int v_2(u) \Pi_1(du) < \infty$. *Тогда* *во введенных обозначениях критерий*

$$\pi_1(X) = \pi_\Pi(\gamma^*); \quad \gamma^* = (\widehat{\theta}^* - \theta_1)\sqrt{n},$$

будет асимптотически байесовским решением игры $(\mathscr{D}, \Theta, W)$ (т. е. *задачи* A), *соответствующим априорному распределению* Q.

Теорема 2. *Пусть в окрестности точки* θ_1 *выполнены условия* (*RR*) *и в задаче* B *существуют минимаксное решение* $\overline{\pi}(Y)$ *и соответствующее ему наихудшее распределение* $\overline{\Pi}$. *Тогда критерий* $\pi_1(X) = \overline{\pi}(\gamma^*)$ *будет асимптотически минимаксным решением задачи* A.

Замечание 1. Условия теоремы о существовании $\overline{\pi}$ и $\overline{\Pi}$ в силу теорем § 3 будут выполнены, если v_i — непрерывные функции.

Доказательство теоремы 1 вполне аналогично доказательству теоремы 3.14.1[6]. Из (2) следует, что байесовское решающее правило π_Q имеет вид $\pi_Q(X) = 1$, если

$$\frac{\int w_1(t)\, q_2(t)\, f_t(X)\, \lambda(dt)}{\int w_2(t)\, q_1(t)\, f_t(X)\, \lambda(dt)} > \frac{q_1}{q_2}. \tag{5}$$

Полагая $Z_1(t) = \dfrac{f_{\theta_1+t}(X)}{f_{\theta_1}(X)}$ и учитывая, что

$$q_i(t)\lambda(dt) = Q_i(dt), \quad Q_i(\theta_1 + du/\sqrt{n}) = \Pi_i(du),$$
$$w_i(\theta_1 + u/\sqrt{n}) = v_i(u),$$

мы можем неравенство (5) с помощью замены $t = \theta_1 + u/\sqrt{n}$ преобразовать к виду

$$\frac{\int v_1(u)\, Z_1(u/\sqrt{n})\, \Pi_2(du)}{\int v_2(u)\, Z_1(u/\sqrt{n})\, \Pi_1(du)} = \frac{\int Z_1(u/\sqrt{n})\, \Pi_2'(du)}{\int Z_1(u/\sqrt{n})\, \Pi_1'(du)} > c, \quad c = \frac{q_1}{q_2}, \tag{6}$$

где обобщенные распределения $\Pi_i'(A) = \int\limits_A v_{i+1}(u)\, \Pi_i(du)$ ($v_3(u) \equiv v_1(u)$, $i = 1, 2$) можно превратить в вероятностные перенормировкой, введя распределения $\Pi_i''(A) = \Pi_i'(A)/\Pi_i'(\Gamma_i)$ (по условию $0 < \Pi_i'(\Gamma_i) < \infty$). Мы получим тогда в качестве (5) неравенство точно того же вида, как в 3.14[6].

Последующие рассуждения доказательства отличаются от соответствующих рассуждений в § 3.14[6] лишь упрощениями. Мы предоставляем провести их читателю. В основе этих рассуждений лежит равномерная по γ сходимость ($\theta = \theta_1 + \gamma/\sqrt{n}$)

$$\widetilde{W}(\pi_Q(\cdot), \theta) \to \widetilde{V}(\pi_\Pi(\cdot), \gamma), \quad \widetilde{W}(\pi_1(\cdot), \theta \to \widetilde{V}(\pi_\Pi(\cdot), \gamma), \tag{7}$$

где $\pi_1(X) = \pi_\Pi(\gamma^*)$. $\triangleleft$

Для доказательства теоремы 2 нам понадобится

Лемма 1. *Пусть* Q — *априорное распределение и* π_1 — *соответствующее ему асимптотически байесовское решение такое, что*

$$\limsup_{n\to\infty} \widetilde{W}(\pi_1(\cdot), Q) = c, \quad \limsup_{n\to\infty} \sup_{\theta\in\Theta} \widetilde{W}(\pi_1(\cdot), \theta) \leqslant c. \tag{8}$$

Тогда π_1 — *асимптотически минимаксное решение.*

Доказательство. Обозначим, как и прежде, через π_Q байесовское решение. Тогда для любого решения π имеем

$$\limsup_{n\to\infty} \sup_{\theta\in\Theta} \widetilde{W}(\pi, \theta) \geqslant \limsup_{n\to\infty} \widetilde{W}(\pi, Q) \geqslant$$
$$\geqslant \limsup_{n\to\infty} \widetilde{W}(\pi_Q, Q) \geqslant \limsup_{n\to\infty} \widetilde{W}(\pi_1, Q) =$$
$$= c \geqslant \limsup_{n\to\infty} \sup_{\theta\in\Theta} \widetilde{W}(\pi_1, \theta). \ \triangleleft$$

Доказательство теоремы 2. Пусть $\overline{\Pi}$ — наихудшее распределение на Γ, так что $\bar{\pi}(Y) = \pi_{\overline{\Pi}}(Y)$ есть минимаксное решающее правило в игре $(\mathscr{D}_B, \Gamma, V)$. Тогда по теореме 1 $\pi_1(X) = \pi_{\overline{\Pi}}(\gamma^*)$ будет асимптотически байесовским решением для распределения $\overline{Q}$, соответствующего $\overline{\Pi}$, и для доказательства теоремы нам достаточно убедиться, что $\overline{Q}$ и π_1 удовлетворяют условиям леммы 1.

Обозначим $N_{\overline{\Pi}}$ носитель распределения $\overline{\Pi}$. Тогда в силу теорем § 3

$$\begin{gathered} \widetilde{V}\left(\pi_{\overline{\Pi}}(\cdot), \gamma\right) = c \text{ при } \gamma \in N_{\overline{\Pi}}, \\ \sup_{\gamma\in\Gamma} \widetilde{V}\left(\pi_{\overline{\Pi}}(\cdot), \gamma\right) \leqslant c. \end{gathered} \tag{9}$$

Но для $\theta = \theta_1 + \gamma/\sqrt{n}$ имеет место (см. (7)) равномерная по γ сходимость $\widetilde{W}(\pi_1(\cdot), \theta) \to \widetilde{V}\left(\pi_{\overline{\Pi}}(\cdot), \gamma\right)$. Отсюда и из (9) следует (8). Теорема доказана.

ПРИЛОЖЕНИЕ

ДОКАЗАТЕЛЬСТВО ДВУХ ФУНДАМЕНТАЛЬНЫХ ТЕОРЕМ ТЕОРИИ СТАТИСТИЧЕСКИХ ИГР

Мы будем предполагать здесь, что выполнены следующие условия.

Условие (A). *Множество решений D и множество параметров (чистых стратегий природы) Θ представляют собой компактные метрические пространства соответственно с метриками ρ_D и ρ_θ.*

Условие (B). *Функция потерь $w(\delta, \theta)$: $D \times \Theta \to R$ непрерывна по δ и θ в метриках ρ_D и ρ_θ соответственно.*

Свойство $w(\delta, \theta) \geqslant 0$ нам не потребуется, и мы не будем предполагать, что оно имеет место.

Далее, мы располагаем выборкой $X \in \mathbf{P}_\theta$ из распределения $\mathbf{P}_\theta$. Ее объем n, не ограничивая общности, можно считать равным 1.

Условие (C). *Распределения $\mathbf{P}_\theta$ непрерывны по вариации относительно θ, т. е.*

$$\sup_{B \in \mathfrak{B}_{\mathscr{X}}} \left| \mathbf{P}_{\theta_m}(B) - \mathbf{P}_\theta(B) \right| \to 0,$$

если $\rho_\theta(\theta_m, \theta) \to 0$ *при* $m \to \infty$.

Если выполнено условие (A_μ), т. е. $\mathbf{P}_\theta$ имеет плотность $f_\theta(x)$ относительно некоторой σ-конечной меры μ на $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}})$:

$$f_\theta(x) = \frac{d\mathbf{P}_\theta}{d\mu}(x),$$

то условие (C) будет эквивалентно непрерывности $f_\theta(x)$ в $L_1(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mu)$:

$$\int \left| f_{\theta_m}(x) - f_\theta(x) \right| \mu(dx) \to 0,$$

если $\rho_\theta(\theta_m, \theta) \to 0$ при $m \to \infty$.

Условия (A), (B), (C) допускают, конечно, возможность множествам D или Θ быть конечными.

Если D конечно и состоит из точек $\delta_1, \ldots, \delta_r$, то условие (A) относительно D будет выполненным (выбор ρ_D не играет роли), а условие (B) будет означать непрерывность относительно ρ_θ функций $w(\delta_1, \theta), \ldots, w(\delta_r, \theta)$.

Если оба множества D и Θ конечны, то условия (A), (B) и (C) автоматически выполнены.

Обозначим через σ_D, σ_Θ соответственно σ-алгебры борелевских множеств из D и из Θ. Следуя § 2.3, обозначим $(\widetilde{\mathscr{D}}, \widetilde{\Theta}, \widetilde{W})$ усредненную статистическую игру, где элементами $\widetilde{\Theta}$ являются распределения Q на (Θ, σ_Θ), а элементами $\widetilde{\mathscr{D}}$ являются распределения $\pi(x) = \pi(x, \cdot)$ на (D, σ_D) (для каждого $x \in \mathscr{X}$), где $\pi(x, A)$ для каждого $A \in \sigma_D$ есть функция, измеримая по x.

Функция риска $\widetilde{W}(\pi, Q)$ определяется равенством

$$\widetilde{W}(\pi, Q) = \int\limits_{\Theta} \int\limits_{\mathscr{X}} \int\limits_{D} w(u, t)\, \pi(x, du)\, f_t(x)\, \mu(dx)\, Q(dt).$$

Если вместо аргумента Q подставить θ, то $\widetilde{W}(\pi, \theta)$ будет означать $\widetilde{W}(\pi, I_\theta)$, где I_θ есть распределение, сосредоточенное в точке θ. Такое же соглашение будет действовать относительно замены $\pi \in \widetilde{\mathscr{D}}$ на $\delta \in \mathscr{D}$. Нам будет удобнее также вместо $\widetilde{W}$ писать W. Это нигде не приведет к недоразумениям.

Лемма 1. *Если выполнены условия* (A), (B), (C), *то функция* $W(\pi, \theta)$ *непрерывна по* θ *для любой стратегии* $\pi(x)$.

Доказательство. Имеем при $\theta_n \to \theta$:

$$\begin{aligned} |W(\pi, \theta_n) - W(\pi, \theta)| &\leqslant |\mathsf{M}_\theta \mathsf{M}[w(\pi(X), \theta) - w(\pi(X), \theta_n)/X]| + \\ &+ |\mathsf{M}_\theta \mathsf{M}[w(\pi(X), \theta_n)/X] - \mathsf{M}_{\theta_n}\mathsf{M}[w(\pi(X), \theta_n)/X]| \leqslant \\ &\leqslant \int |w(\pi(x), \theta) - w(\pi(x), \theta_n)|\, P_\theta(dx) + \\ &+ \sup_{\delta, \theta} |w(\delta, \theta)| \int |P_{\theta_n}(dx) - P_\theta(dx)|. \quad (1) \end{aligned}$$

Первый интеграл здесь сходится к 0 в силу непрерывности функции w по θ. Сходимость к нулю второго интеграла вытекает из условия (C). Действительно, пусть $f_{\theta_n}(x)$ есть плотность P_{θ_n} относительно меры

$$\mu = P_\theta + \sum_{j=1}^{\infty} 2^{-j} P_{\theta_j},$$

и пусть $B_n = \{x\colon f_{\theta_n}(x) \geqslant f_\theta(x)\}$. Тогда второй интеграл в (1) равен

$$\begin{aligned} \int |f_{\theta_n}(x) - f_\theta(x)|\, \mu(dx) &= 2 \int\limits_{B_n} \left(f_{\theta_n}(x) - f_\theta(x)\right) \mu(dx) = \\ &= 2\left(P_{\theta_n}(B_n) - P_\theta(B_n)\right) \to 0. \end{aligned}$$

Лемма доказана.

Теорема 1 (первая фундаментальная теорема). *Если выполнены условия* (A), (B), (C), *то игра* $(\widetilde{\mathscr{D}}, \widetilde{\Theta}, W)$ *имеет цену и минимаксные стратегии обоих игроков. Другими словами, существуют наименее благоприятное распределение* $\overline{Q}$ *и минимаксное решающее правило* $\overline{\pi}(x)$:

$$W_* = \sup_Q \inf_\pi W(\pi, Q) = W(\overline{\pi}, \overline{Q}) = \inf_\pi \sup_Q W(\pi, Q) = W^*. \quad (2)$$

В силу леммы 2.1 утверждение (2) эквивалентно тому, что

$$W(\bar{\pi}, \uparrow) = \sup_{Q} W(\bar{\pi}, Q) = W(\bar{\pi}, \bar{Q}) = \inf_{\pi} W(\pi, \bar{Q}) = W(\downarrow, \bar{Q}). \quad (3)$$

Теорема 2 (вторая фундаментальная теорема). *Если выполнены условия* (A), (B), (C), *то байесовские решения* $\pi_Q(x)$ *образуют полный класс. Другими словами, для любого* $\pi_0 \in \widetilde{\mathscr{D}}$ *найдутся* $Q \in \widetilde{\Theta}$, $\pi_Q \in \widetilde{\mathscr{D}}$ *такие, что*

1) $W(\pi_Q, Q) = W(\downarrow, Q)$,

2) $W(\pi_Q, \theta) \leqslant W(\pi_0, \theta)$ *при всех* θ.

Доказательство теоремы 2. Вторая фундаментальная теорема является следствием первой. Рассмотрим произвольную стратегию $\pi_0 \in \widetilde{\mathscr{D}}$ и игру $(\widetilde{\mathscr{D}}, \widetilde{\Theta}, W_0)$, где W_0 построена по функции $w_0(\delta, \theta) = w(\delta, \theta) - W(\pi_0, \theta)$, так что

$$W_0(\pi, \theta) = W(\pi, \theta) - W(\pi_0, \theta). \quad (4)$$

Функция $v(\theta) = W(\pi_0, \theta)$ в силу леммы 1 непрерывна по θ, и, стало быть, функция потерь $w_0(\delta, \theta) = w(\delta, \theta) - v(\theta)$ вместе с $w(\delta, \theta)$ удовлетворяет условию (B). Это означает, что к игре $(\widetilde{\mathscr{D}}, \widetilde{\Theta}, W_0)$ применима теорема 1. Так как $W_0(\pi_0, \uparrow) = 0$ (см. (4)), то верхняя цена этой игры удовлетворяет неравенству $W_0^* \leqslant 0$. Из (2), (3) следует тогда, что существуют $\bar{\pi}$, $\bar{Q}$ такие, что

$$\sup_{P} W_0(\bar{\pi}, P) = \sup_{\theta} W_0(\bar{\pi}, \theta) \leqslant 0, \ \bar{\pi} = \pi_{\bar{Q}}.$$

Эти два соотношения эквивалентны утверждениям 2), 1) теоремы 2, если положить $\bar{Q} = Q$, $\bar{\pi} = \pi_Q$. Теорема доказана.

Доказательство теоремы 1 будет вытекать из следующих двух лемм.

Лемма 2. *При выполнении условий* (A), (B), (C) *существует распределение* $\bar{Q}$ *такое, что* $W(\downarrow, \bar{Q}) \geqslant \inf_{\pi} W(\pi, \uparrow) = W^*$.

Лемма 3. *При выполнении условий* (A), (B), (C) *существует стратегия* $\bar{\pi}$ *такая, что* $W(\bar{\pi}, \uparrow) \leqslant W^*$.

Из неравенств лемм 2, 3 следует соотношение

$$W^* \geqslant W(\bar{\pi}, \uparrow) \geqslant W(\bar{\pi}, \bar{Q}) \geqslant W(\downarrow, \bar{Q}) \geqslant W^*,$$

эквивалентное (3) и, стало быть, (2). Это доказывает теорему 1. ◁

Леммы 2, 3 разделяют доказательство теоремы 1 на две части. Первая из них (лемма 2) очень мало связана с тем, что игра является статистической. Эта часть доказательства проводится примерно так же, как для обычных игр (ср. с [40]).

Доказательство леммы 2. Пусть V есть множество функций $\Theta \to R$, представимых в виде $v(\theta) = W(\pi, \theta)$, $\pi \in \widetilde{\mathscr{D}}$. В силу леммы 1 все функции из V непрерывны, так что $V \subset C(\Theta)$, где $C(\Theta)$ — пространство всех непрерывных функций на Θ. Далее, пусть $v_1(\theta) = W(\pi_1, \theta)$, $v_2(\theta) = W(\pi_2, \theta)$. Так как при $p \in [0, 1]$

$$v(\theta) = pv_1(\theta) + (1-p)v_2(\theta) = W(p\pi_1 + (1-p)\pi_2, \theta),$$
$$\pi = p\pi_1 + (1-p)\pi_2 \in \widetilde{\mathscr{D}},$$

то $v \in V$ и, стало быть, множество V выпукло.

Заметим теперь, что $W^* = \inf_\pi W(\pi, \dagger) = \inf_{v \in V} \sup_\theta v(\theta)$. Нам будет удобнее вместо исходной функции $w(\delta, \theta)$ рассматривать функцию $\frac{w(\delta, \theta) - v_0 + 1}{W^* - v_0 + 1}$, $v_0 = \inf_{v \in V} \inf_\theta v(\theta)$. Обозначая новую функцию снова через $w(\delta, \theta)$ (задача при этом остается неизменной), мы получим, что для нее

$$W^* = 1, \quad v_0 > 0. \tag{5}$$

Пусть теперь U есть множество непрерывных функций $v(\theta)$: $\Theta \to R$ таких, что $\sup_\theta v(\theta) < 1$. Очевидно, что U — выпуклое открытое множество из $C(\Theta)$. Кроме того, из (5) следует, что пересечение $V \cap U$ пусто. Поэтому в силу теоремы Хана — Банаха (см., например, [40], с. 171, 200—206) существует линейный функционал $L(v)$: $C(\Theta) \to R$ такой, что

$$L(v) < 1 \text{ при } v \in U, \; L(v) \geqslant 1 \text{ при } v \in V. \tag{6}$$

Этот функционал с необходимостью обладает свойством $L(v) \geqslant 0$, если $v(\downarrow) \equiv \inf_\theta v(0) \geqslant 0$. Действительно, допустив существование элемента $v_0 \in C(\Theta)$, $v_0(\downarrow) \geqslant 0$, для которого $L(v_0) < 0$, мы получим, что $v_s = -sv_0 \in U$ при любом $s > 0$, $L(v_s) = -sL(v_0) > 1$ при достаточно большом s. Это приводит к противоречию с (6).

Но неотрицательный функционал L в силу теоремы Рисса ([36], с. 240) допускает представление в виде интеграла

$$L(v) = \int_\Theta v(0)\, \lambda(d0),$$

где λ — конечная мера. Так как $1 \geqslant \sup_{v \in U} L(v) = \lambda(\Theta)$, то, полагая $\overline{Q}(A) = \lambda(A)/\lambda(\Theta)$, мы получим при $v \in V$:

$$L(v) = \int W(\pi, 0)\, \lambda(d\theta) = \lambda(\Theta)\, W(\pi, \overline{Q}),$$

$$W(\downarrow, \overline{Q}) = \frac{1}{\lambda(\Theta)} \inf_{v \in V} L(v) \geqslant 1 = W^*.$$

Лемма доказана.

Доказательство леммы 3. Так как функция $W(\pi, \theta)$ при каждом $\pi \in \widetilde{\mathscr{D}}$ непрерывна по θ (см. лемму 1), то нам достаточно построить стратегию $\overline{\pi}$, для которой при всех $k = 1, 2, \ldots$

$$W(\overline{\pi}, \theta_k) \leqslant W^*, \tag{7}$$

где θ_k суть точки некоторого счетного всюду плотного в D множества $T = \{\theta_1, \theta_2, \ldots\}$. По определению верхней цены W^* существует последовательность стратегий $\pi_n = \pi_n(x, \cdot)$ такая, что

$$W(\pi_n, \theta_k) < W^* + 1/n \tag{8}$$

при всех k.

Построим теперь с помощью распределений π_n последовательность специально подобранных случайных элементов ζ_n и выделим из нее сходящуюся подпоследовательность. Для этого обозначим через $f_{\theta_k}(x)$ плотность распределения $\mathbf{P}_{\theta_k}$ относительно вероятностной меры $\mu=\sum_{j=1}^{\infty} 2^{-k}\mathbf{P}_{\theta_k}$, так что

$$W(\pi_n, \theta_k) = \int \int w(u, \theta_k)\, \pi_n(x, du)\, f_{\theta_k}(x)\, \mu(dx). \tag{9}$$

Рассмотрим пространство $D\times R^T$, где R^T есть пространство значений элементов $f(x)=\{f_{\theta_1}(x), f_{\theta_2}(x), \ldots\}$ с σ-алгеброй $\mathfrak{B}^T$, порожденной цилиндрическими множествами. Каждой стратегии π поставим в соответствие вероятностное пространство $(D\times\mathscr{X}, \sigma_D\times\times\mathfrak{B}_{\mathscr{X}}, \mathbf{P})$, где распределение $\mathbf{P}$ определяется равенством

$$\mathbf{P}(\delta\in A, X\in B) = \int_B \mu(dx)\, \pi(x, A), \quad A\in\sigma_D, \quad A\in\mathfrak{B}_X. \tag{10}$$

На этом пространстве определим случайные элементы $\zeta=\zeta(\delta; X)= =(\delta; f_{\theta_1}(X), f_{\theta_2}(X), \ldots)=(\delta; f(X))$ и обозначим через ζ_n элементы, соответствующие π_n, так что ζ_n есть случайные величины на выборочном вероятностном пространстве $(D\times R^T, \sigma_D\times\mathfrak{B}^T, \Pi_n)$, а распределение Π_n порождено π_n, формулой (10) и отображением $\zeta(\delta, x)$: $D\times\mathscr{X}\to D\times R^T$.

Обозначим $\Pi_n^{(k)}$ сужения распределения Π_n на $D\times R^k$ (это есть совместное распределение $(\delta; f_{\theta_1}(X), \ldots, f_{\theta_k}(X))$ и через λ — распределение $f(X)$ на $(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mu)$. Нам понадобится

Лемма 4. *Существуют такое распределение $\overline{\Pi}$ на измеримом пространстве $(D\times R^T, \sigma_D\times\mathfrak{B}^T)$ и такая подпоследовательность $\{\pi_{n*}\}\subset\{\pi_n\}$, что*

$$\Pi_{n*}^{(k)}\Rightarrow\overline{\Pi}^{(k)} \tag{11}$$

при любом k ($\overline{\Pi}^{(k)}$ суть сужения $\overline{\Pi}$),

$$\overline{\Pi}(D\times C)=\lambda(C), \quad C\in\mathfrak{B}^T. \tag{12}$$

Доказательство леммы 4 мы проведем позже.

Обозначим $\overline{\zeta}=(\overline{\delta}; \overline{f})$ случайный элемент с распределением $\overline{\Pi}$. Соотношение (12) означает, что распределение $\overline{f}$ совпадает с λ (вторая «координата» ζ_n при изменении n распределение не меняет). Так как пространство D есть метрический компакт, то оно сепарабельно и, стало быть (см. [11], с. 191), существует условное (регулярное) распределение $\overline{\delta}$ относительно $\overline{f}(X)$, которое мы обозначим $\Pi(\cdot/\overline{f}(x))$.

Рассмотрим стратегию $\overline{\pi}(x, A)=\Pi(\overline{\delta}\in A/\overline{f}(X))$ и докажем, что для нее выполняется (7).

Заметим предварительно, что

$$\mathsf{M}w(\overline{\delta}, \theta_k)\overline{f}_{\theta_k}=\mathsf{M}\overline{f}_{\theta_k}\mathsf{M}(w(\overline{\delta}, \theta_k)/X)=$$

$$=\int f_{\theta_k}(x)\int w(u, \theta_k)\,\overline{\pi}(u, dx)\,\mu(dx)=W(\overline{\pi}, \theta_k). \tag{13}$$

Далее, в силу леммы 4 распределение $(\delta_{n*}, f_{\theta_k}(X))$ слабо сходится к распределению $(\bar{\delta}, \bar{f}_{\theta_k}(X))$. Поскольку функция w непрерывна, то совместное распределение $(w(\delta_{n*}, \theta_k), f_{\theta_k}(X))$ слабо сходится к распределению $(w(\bar{\delta}, \theta_k), \bar{f}_{\theta_k}(X))$. Но функция $g(u, v) = w(u, \theta_k)v$ непрерывна по u и v и мажорируется функцией $g(v) = cv$, $c = \max_u w(u, \theta_k)$, такой, что $\mathsf{M}g(f_{\theta_k}(X)) = c\int f_{\theta_k}(x)\,\mu(dx) = c$. Поэтому по теореме непрерывности для моментов (см. теорему 1.5.4[61])

$$\lim_{n*\to\infty} \mathsf{M}g\left(\delta_{n*}, f_{\theta_k}(X)\right) = \mathsf{M}g\left(\bar{\delta}, \bar{f}_{\theta_k}(X)\right),$$

или, что то же, $\lim\limits_{n*\to\infty} \mathsf{M}w(\delta_{n*}, \theta_k) f_{\theta_k}(X) = \mathsf{M}w(\bar{\delta}, \theta_k)\bar{f}_{\theta_k}(X)$.

В силу (9), (13) это дает нам сходимость

$$\lim_{n*\to\infty} W(\pi_{n*}, \theta_k) = W(\bar{\pi}, \theta_k).$$

Так как левая часть этого равенства (см. (8)) не превосходит W^*, то лемма 3 доказана.

Доказательство леммы 4. Фиксируем какое-нибудь $k \geqslant 1$ и рассмотрим $D \times R^k$ как полное метрическое сепарабельное пространство относительно метрики, порожденной евклидовой метрикой в R^k и метрикой ρ_D. Для любого $\varepsilon > 0$ в R^k найдется компакт K_ε такой, что $\mathsf{P}\left((f_{\theta_1}(X), \ldots, f_{\theta_k}(X)) \in K_\varepsilon\right) \geqslant 1-\varepsilon$. Так как $D \times K_\varepsilon$ есть компакт в $D \times R^k$ и так как

$$\mathsf{P}\left(\delta_n \in D, (f_{\theta_1}(X), \ldots, f_{\theta_k}(X)) \in K_\varepsilon\right) \geqslant 1-\varepsilon,$$

то последовательность распределений $\Pi_n^{(k)}$ является плотной (см. [1]). Следовательно, по теореме Прохорова [1] существуют распределение $\bar{\Pi}^{(k)}$ и подпоследовательность $n^{(k)} = \left(n_1^{(k)}, n_2^{(k)}, \ldots\right)$ такие, что $\Pi_{n(k)}^{(k)} \Rightarrow \bar{\Pi}^{(k)}$. Но распределения $\bar{\Pi}^{(k)}$, очевидно, согласованы, и, следовательно, по теореме Колмогорова на $(D \times R^T, \sigma_D \times \mathfrak{B}^T)$ существует распределение $\bar{\Pi}$, для которого $\bar{\Pi}^{(k)}$ являются сужениями на $(D \times R^k, \sigma_D \times \mathfrak{B}^k)$.

С другой стороны, можно считать, что $n^{(k+1)} \subset n^{(k)}$. Положив $n^* = \left(n_1^{(1)}, n_2^{(2)}, n_3^{(3)}, \ldots\right)$, мы получим подпоследовательность, для которой $\Pi_{n*}^{(k)} \Rightarrow \bar{\Pi}^{(k)}$ при всех k.

Докажем теперь (12). Пусть $C \in \mathfrak{B}^T$ есть цилиндрическое множество такое, что $\bar{\Pi}$-мера его границы равна нулю. Обозначим через $C^{(k)} = C \cap R^k \in \mathfrak{B}^k$ множество из R^k, образованное первыми k координатами точек из C, и положим $\bar{C}^{(k)} = C^{(k)} \times R^{T-k} \in \mathfrak{B}^T$. Тогда $\lambda(\bar{C}^{(k)}) = \Pi_{n*}^{(k)}(D \times C^{(k)}) \to \bar{\Pi}^{(k)}(D \times C^{(k)})$. Так как $\bar{C}^{(k+1)} \subset \bar{C}^{(k)}$, $C = \bigcap\limits_{k=1}^{\infty} \bar{C}^{(k)}$, то

$$\lambda(C) = \lim_{k\to\infty} \lambda(\bar{C}^{(k)}) = \lim_{k\to\infty} \bar{\Pi}^{(k)}(D \times C^{(k)}) = \lim_{k\to\infty} \bar{\Pi}(D \times \bar{C}^{(k)}) = \bar{\Pi}(D \times C).$$

Лемма 4 доказана.

БИБЛИОГРАФИЧЕСКИЕ ЗАМЕЧАНИЯ

Предлагаемые ниже библиографические комментарии следует рассматривать как продолжение библиографических замечаний, приведенных в [6]. Эти комментарии не претендуют на полноту и часто будут содержать ссылки не на оригинальные труднодоступные статьи, а на учебники, монографии или обзорные статьи, в которых нужные результаты найти проще.

ГЛАВА 1

§ 1. Критерий χ^2 в задаче примера 1.1.1, критерий Стьюдента в задаче примера 1.1.3 и критерий Фишера в задачах примеров 1.1.4, 1.1.5 используются очень часто. О других свойствах оптимальности этих критериев см. книгу Лемана [18]. Пример 1.1.1А заимствован из [27]. По проблеме Беренса — Фишера (пример 1.1.6) имеется обширная литература (см. [18]).

§ 2. Точное распределение статистики $D_{n,n}$ найдено Гнеденко и Королюком (см. [33]); предельное распределение статистики D_{n_1,n_2}—Смирновым. Теорема 1.2.2 доказана впервые в [21] методом моментов. О критериях знаков и Вилкоксона см. также [10].

§§ 3, 4. Подробнее задачи регрессии и дисперсионного анализа изложены в специальных монографиях Себера [29] и Шеффе [39]. См. также [17], [18], [27].

§ 5. Замечание об асимптотической оптимальности критерия (1.5.3) содержится в [4].

ГЛАВА 2

В математике теоретико-игровое направление возникло с появлением работ Бореля 1921 года и фон Неймана 1928 года. В математической статистике исходной работой, подготовившей использование теории игр, можно считать классическую работу Неймана и Пирсона [23], в которой были высказаны многие основные идеи теории статистических решений. Фундаментальный вклад в развитие общей теории статистических решений принадлежит Вальду. Основные положения этой теории изложены в его завершающей книге [9]. Общематематическая теория игр получила свое полное развитие в книге фон Неймана и Моргенштерна [24].

Доступное изложение основ теории статистических игр можно найти в книгах Гиршика и Блекуэла [3] и Фергюсона [34],

§ 2. Сравнительно полным введением в обычную теорию игр является книга Мак-Кинси [20].

§§ 3, 4. Более полное описание основ теории статистических игр см. в [3], [34]. В этих книгах две фундаментальные теоремы теории статистических игр доказаны лишь в частном случае для дискретных множеств D и Θ. Объясняется это большой сложностью изложения в общем случае (см. [9]). В приложении приведено наиболее простое из известных нам доказательств этих теорем. Оно было найдено А. И. Саханенко.

Роль байесовского подхода в разное время оценивалась по-разному. Он широко использовался в прошлом веке Лапласом. Затем байесовский подход был подвергнут критике со стороны Фишера, и в 20-х, 30-х годах нашего столетия центр тяжести переместился на эффективные оценки и на наиболее мощные критерии. Затем, по мере того, как стала осознаваться фундаментальная роль байесовского подхода, интерес к нему стал вновь возрастать. Эта фундаментальная роль выясняется в теоремах 2.3.1, 2.3.2.

§ 5. Фундаментальное понятие достаточной статистики было введено Р. Фишером в [35] в 1922 году. Р. Фишер [35] и позже Ю. Нейман [22] предложили простой критерий, обнаруживающий существование и вид достаточной статистики. Этот критерий носит название факторизационной теоремы Неймана — Фишера и представлен в теореме 2.12.1[6]. Строгое теоретико-множественное доказательство теоремы Неймана — Фишера было получено лишь в 1949 году Халмошем и Сэвиджем [37].

Понятие достаточной σ-алгебры шире, чем понятие достаточной статистики. Необходимые и достаточные условия для их совпадения приведены в [13]. Теорема 2.5.1 (сначала для квадратической функции потерь) была получена независимо Блекуэллом [2] (1947), Рао [25] (1945), [26] (1949) и Колмогоровым [15] (1950). Обобщения на случай произвольной функции потерь связаны с именами Лемана и Шеффе [13].

Идея использования инвариантных соображений принадлежит Хотеллингу и Питмену. Значительный вклад в разработку теории внес Ч. Стейн (см. [13], [14]).

О несмещенности подробнее см. [13].

§ 6. Результаты, близкие к теоремам этого параграфа, можно найти в книге Ибрагимова и Хасьминского [14].

§ 7. Асимптотическая байесовость к. о. н. установлена в работе автора [4]. Результаты о предельном распределении отношения правдоподобия при основной гипотезе принадлежат Уилксу [31] и Вальду [8] (см. также книгу Уилкса [32]). Идея замены сложной гипотезы усредненной использовалась Вальдом. Асимптотический вид байесовских критериев содержится в [19].

§ 8. Основные идеи, связанные с отысканием асимптотически оптимальных тестов для близких гипотез, изложены в работах Вальда [8], Ле-Кама, Русаса (см. книгу Русаса [28]), Чибисова [38]. О возможности перенесения основных результатов на случай бесконечномерного параметра (на случайные процессы) см. [7]. Форма изложения § 8 и §§ 14[6], 15[6] с цитированными работами связана мало. Редукция исходной задачи A к задаче B (для параметра нормального распределения) при отыскании оптимальных критериев для основных типов задач содержится в работе Вальда [8].

Приложение

Доказательство двух фундаментальных теорем теории статистических игр можно найти в [9] и, при более частных предположениях, в [3], [34]. В настоящей книге излагается подход к доказательству, предложенный А. И. Саханенко. Центральной его частью являются леммы 2, 3. Лемма 2 по существу не связана со статистическим характером игры, основана на теоремах Хана — Банаха и Рисса и близка по своей идее к рассуждениям, использованным, например, в [40]. Доказательство леммы 3 основано на теоремах Колмогорова [16] и Прохорова [1].

СПИСОК ЛИТЕРАТУРЫ

1. Биллингсли П. Сходимость вероятностных мер.— М.: Наука, 1977.
2. Блекуэлл Д. (Blackwell D.). Conditional expectation and unbiased sequential estimation.— Ann. Math. Statist., 1947, 18, 105—110.
3. Блекуэлл Д., Гиршик М. А. Теория игр и статистических решений.— М.: ИЛ, 1958.
4. Боровков А. А. Асимптотически оптимальные тесты для проверки сложных гипотез.— Теория вероятностей и ее применения, 1975, 20, 3, 463—487.
5. Боровков А. А. Теория вероятностей.— М.: Наука, 1976.
6. Боровков А. А. Математическая статистика.— М.: Наука, 1984.
7. Боровков А. А., Саханенко А. И. Об асимптотически оптимальных тестах для проверки сложных близких гипотез.— Труды Института математики СО АН СССР, 1982, т. 1.
8. Вальд А. (Wald A.). Tests of statistical hypothess concerning several parameters when the number of observations is large.— Trans. Amer. Math. Soc. 1943, 54, 3, 426—482.
9. Вальд А. Статистические решающие функции.— В кн.: «Позиционные игры».— М.: Наука, 1967, 300—522.
10. Гаек Я., Шидак З. Теория ранговых критериев.— М.: Наука, 1971.
11. Гихман И. И., Скороход А. В. Введение в теорию случайных процессов.— М.: Наука, 1977.
12. Дуб Дж. Л. Вероятностные процессы.— М.: ИЛ, 1956.
13. Закс Ш. Теория статистических выводов.— М.: Мир, 1975.
14. Ибрагимов И. А., Хасьминский Р. З. Асимптотическая теория оценивания.— М.: Наука, 1979.
15. Колмогоров А. Н. Несмещенные оценки.— Изв. АН СССР, сер. мат., 1950, 14, 303.
16. Колмогоров А. Н. Основные понятия теории вероятностей.— М.: Наука, 1974.
17. Крамер Г. Математические методы статистики.— М.: Мир, 1975.
18. Леман Э. Проверка статистических гипотез.— М.: Наука, 1964.
19. Линдли (Lindley D.). The use of prior probability distributions in statistical inference and decision, Proc. 4-th Berkeley Sympos. Math. Statist. Prob.— Berkeley — Los Angeles, v. 1, 1960, 453—468.
20. Мак Кинси Дж. Введение в теорию игр. М.: Физматгиз, 1960.

21. Манн, Уитни (Mann H. B., Whitney D. R.). On a test whether one of two random variables is stochastically larger than the other.— Ann. Math. Statist. 1947, 18, 50.
22. Нейман Ю. (Neyman J.). Sur un teorems concerente le cosidette statistische sufficienti.— Inst. Ital. Atti. Giorn., 1935, 6, 320—334.
23. Нейман Ю., Пирсон Э. (Neyman J., Pearson E. S.). The testing of statistical hypotheses in relation to probabilities a priori.— Proc. Camb. Phil. Soc. 1933, 24, 492—510.
24. Фон Нейман Дж., Моргенштерн О. Теория игр и экономическое поведение.— М.: Наука, 1970.
25. Рао С. (Rao C. R.). Information and accuracy attainable in estimation of statistical parameters.— Bull. Calcutta Math Soc. 1945, 37, 81—91.
26. Рао С. (Rao C. R.). Sufficient statistics and minimum variance estimates.— Proc. Camb. Phil. Soc., 1949, 45, pp. 213—218.
27. Рао С. Линейные статистические методы и их применения.— М.: Наука, 1968.
28. Русас Дж. Контигуальность вероятностных мер.— М.: Мир, 1975.
29. Себер Дж. Линейный регрессионный анализ.— М.: Мир, 1980.
30. Сидоров В. А. и др. (Sidorov V. A., ...) Measurement of the $\varphi \to \pi^+\pi^-$ branching ration.— Physics Letters, 1981, 99 B, 1, 62—65.
31. Уилкс С. (Wilks S. S.). The large sample distribution of the likelihood ratio for testing composite hypotheses.— Ann. Math. Statist., 1938, 9, 60—62.
32. Уилкс С. Математическая статистика.— М.: Наука, 1967.
33. Феллер Э. Введение в теорию вероятностей и ее приложения, т. 1, 2.— М.: Мир, 1967.
34. Фергюсон Т. (Ferguson T. S.). Mathematical statistics. A decision theoretic approach.— New Jork and London: Academic Press, 1967.
35. Фишер Р. (Fischer R. A.). On the mathematical foundations of theoretical statistics.— Philos. Trans. Roy. Soc. A. 1922, 222, 309—368.
36. Халмош П. *Теория меры*, М.: ИЛ, 1953.
37. Халмош П., Сэвидж Л. (Halmos P. R., Savage L. J.). Application of the Radon — Nikodym theorem to the theory of sufficient statistics.— Ann. Math. Statist., 1949, 20, 225—241.
38. Чибисов Д. М. (Chibisov D. M.). Transition to the limiting process for deriving asymptotically optimal tests.— Sankhya, 1969, A31, 3, 241—258.
39. Шеффе Г. Дисперсионный анализ.— М.: Физматгиз, 1963.
40. Эдвардс Р. Функциональный анализ.— М.: Мир, 1969.

СПИСОК ОСНОВНЫХ ОБОЗНАЧЕНИЙ

Многие обозначения в настоящей книге заимствованы из основной части курса [6]. В приводимом списке такие обозначения сопровождаются указанием номера той страницы из [6], на которой они были введены.

к. о. п.— критерий отношения правдоподобия 349[6]
о. м. п.— оценка максимального правдоподобия 87[6]

(A_0) — условие взаимно однозначного соответствия между параметрическим множеством Θ и семейством распределений $\mathscr{P} = \{\mathrm{P}_\theta\}_{\theta\in\Theta}$ $\left(\mathrm{P}_{\theta_1} \neq \mathrm{P}_{\theta_2}\text{, если } \theta_1 \neq \theta_2\right)$ 83[6]

(A_c) — условие, состоящее в том, что параметрическое множество Θ компактно 199[6]

(A_μ) — условие, в силу которого все распределения семейства $\mathscr{P} = \{\mathrm{P}_\theta\}$ доминируются мерой μ (существует плотность $f_\theta = d\mathrm{P}_\theta/d\mu$) 85[6]

(R) — условие регулярности параметрического семейства, в силу которого функция $\sqrt{f_\theta(x)}$ непрерывно дифференцируема по θ, а информация Фишера положительна и непрерывна 150[6], 158[6]

(RR) — условие регулярности параметрического семейства, требующее выполнения условий (A_0), (A_c), (R), а также непрерывной дифференцируемости второго порядка по θ функции $l(x,\theta)$ и существования мажоранты $l(x) \geqslant |l''(x, t)|$, для которой интеграл $\mathrm{M}_\theta l(x_1)$ сходится равномерно в Θ 213[6], 238[6]

$\mathfrak{B}_{\mathscr{X}}$ — σ-алгебра в фазовом пространстве $\mathscr{X}$ 17[6].

B_p — полиномиальное распределение 73[6]

D — пространство стратегий I игрока (в гл. 2) 70

$\mathscr{D}$ — пространство решающих функций в статистической игре 84

D_θ — дисперсия по распределению P_θ 103[6]

E — единичная матрица 158[6]

$f_\theta(x)$ — плотность распределения P_θ относительно меры μ 85[6]

$f_\theta(X) = \prod_{i=1}^{n} f_\theta(x_i)$ — функция правдоподобия 89[6]

$F(x)$ — функция распределения, соответствующая распределению P 18[6]

$F_X^*(x)$ — эмпирическая функция распределения, соответствующая выборке X 28, 22[6], 29[6]

F_{k_1,k_2} —распределение Фишера 66[6]

G — группа преобразований $\mathscr{X}$ в себя, соответствующая инвариантному семейству 101, 180[6]

h_ε — квантиль распределения χ^2 11, 265[6]

H_i — гипотеза 270[6]

H_k — распределение χ^2 65[6]

I_x — распределение, сосредоточенное в точке x 21[6], 72[6]

$I(\theta) = \| I_{ij}(\theta) \|$, $I_{ij}(\theta) = \mathsf{M}_\theta \frac{\partial}{\partial \theta_i} l(x_1, \theta) \cdot \frac{\partial}{\partial \theta_j} l(x_1, \theta)$ — информационная матрица Фишера 167[6]

$l(x, \theta) = \ln f_\theta(x)$ 89[6]

$L(X, \theta) = \ln f_\theta(X)$ — логарифмическая функция правдоподобия 89[6]

M_θ — математическое ожидание по распределению P_θ 103[6]

$\mathsf{M}(\xi/\eta)$ — условное математическое ожидание ξ относительно случайной величины η 114[6]

P — символ распределения, употребляемый в разных смыслах, указанных на с. 19[6]

P_X^* — эмпирическое распределение, соответствующее выборке X 28, 21[6]

$\mathscr{P}$ — семейство распределений 62[6]

Q — рандомизированная стратегия «природы» (априорное распределение θ) 72

Q_X — апостериорное распределение θ 91

$\bar{\mathrm{Q}}$ — наихудшее распределение θ (минимаксная стратегия «природы») 75, 91

$q(t/X)$ — апостериорная плотность θ 91

R^m — m-мерное евклидово пространство 17[6]

S_X^2 — эмпирическая дисперсия, соответствующая выборке X 16, 25[6]

T_k — распределение Стьюдента 67

$\mathrm{U}_{a,b}$ — равномерное распределение на $[a, b]$ 69[6]

$w(\delta, \theta)$ — функция потерь I игрока 70

$W(\delta(\cdot), \theta) = \mathsf{M}_\theta w(\delta(X), \theta)$ — функция риска 84

$w^\circ(t)$ — броуновский мост 39[6]

x_i — элемент выборки 18[6]

$\mathrm{x}_{(i)}$ — i-й элемент вариационного ряда 22[6]

$\bar{\mathrm{x}}$ — эмпирическое среднее 16, 25[6]

$X = (\mathrm{x}_1, \ldots, \mathrm{x}_n)$ — выборка 18[6]

$\mathscr{X}$ — фазовое пространство выборки 17[6]

$(\mathscr{X}, \mathfrak{B}_{\mathscr{X}}, \mathrm{P})$ — выборочное вероятностное пространство 19[6]

$x = (x_1, \ldots, x_n)$ — элемент $\mathscr{X}^n$ 20[6]

$\Gamma_{\alpha,\lambda}$ — гамма-распределение 64[6]

δ — стратегия I игрока 70

$\delta = \delta(X)$ — решающая функция 84

θ — параметр (стратегия «природы») 70, 61[6]

θ^* — оценка параметра θ 62[6]

θ_Q^* — байесовская оценка параметра θ, соответствующая априорному распределению Q 121[6]

$\hat{\theta}^*$ — оценка максимального правдоподобия параметра θ 87[6]

Θ — множество возможных значений параметра θ (пространство стратегий «природы») 70, 62[6], 73[6]

λ_ε — квантиль нормального распределения 256[6], 31

π — рандомизированная стратегия I игрока 72

$\pi = \pi(X)$ — рандомизированное решающее правило (критерий) 86, 275[6], 300[6]

π_Q — байесовская стратегия 73

$\bar{\pi}$ — минимаксная стратегия 75

Φ_{α,σ^2} — нормальное распределение 63[6], 64[6]

$\Phi(x)$ — функция стандартного нормального распределения 66[6]

$\xrightarrow[P]{}$ — знак сходимости по вероятности 31[6]

$\Subset$ — знак, употребляемый между обозначениями выборки (или случайной величины) и распределения и означающий, что выборка извлечена из данного распределения (случайная величина имеет данное распределение) 18[6]

$\Leftarrow\!\!\Rightarrow$ — знак слабой сходимости. Соотношение $\xi_n \Leftarrow\!\!\Rightarrow \mathbf{P}$ означает, что распределения случайных величин ξ_n слабо сходятся к $\mathbf{P}$ при $n \to \infty$ 32[6]

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Александр Алексеевич Боровков

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА.
ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ

Редактор *М. М. Горячая*
Техн. редактор *Л. В. Лихачева*
Корректоры *Г. В. Подвольская, И. Я. Кришталь*

ИБ № 12519

Сдано в набор 12.01.84. Подписано к печати 13.07.84. Формат 84×108¹/₃₂. Бумага тип. № 3. Гарнитура обыкновенная. Высокая печать. Усл. печ. л. 7,56. Усл. кр.-отт. 7,77. Уч.-изд. л. 7,69. Тираж 16 000 экз. Заказ № 38. Цена 25 коп.

Издательство «Наука»
Главная редакция физико-математической литературы
117071 Москва В-71, Ленинский проспект, 15

4-я типография издательства «Наука»
630077 г. Новосибирск-77, Станиславского, 25