МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского»

М.С. Тихов,
М.В. Котельникова

СОВРЕМЕННЫЕ МЕТОДЫ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ

Учебно-методическое пособие

Рекомендовано методической комиссией
Института экономики и предпринимательства ННГУ
для студентов, обучающихся по направлению подготовки
38.03.05 "Бизнес-информатика"(профиль подготовки
«Информатика и математика в анализе экономических систем и бизнеса»)

Нижний Новгород
2016

УДК 519.2
ББК 22.1
Т 46

Т 46 Тихов М.С., Котельникова М.В. СОВРЕМЕННЫЕ МЕТОДЫ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ: Учебно-методическое пособие. – Нижний Новгород, Нижегородский госуниверситет, 2016. – 120 с.

Рецензент: доктор физ.-мат. наук, проф. **В.Г. Гавриленко**

В учебно-методическом пособии рассматриваются современные методы теории статистического оценивания параметров распределений

Предназначено для знакомства студентов старших курсов бакалавриата, обучающихся по направлению 38.03.05 – Бизнес-информатика с современными методами оценивания в исследовании экономических показателей. Данное пособие может быть использовано как дополнительное при изучении курса "Теория вероятностей и математическая статистика"

Ответственный за выпуск:
председатель методической комиссии
Института экономики и предпринимательства ННГУ,
к.э.н., доцент **Е.Н. Летягина**

УДК 519.2
ББК 22.1

ВВЕДЕНИЕ

Закономерности, которым подчиняется исследуемая случайная величина X, физически обуславливается реальным комплексом условий ее наблюдения (или эксперимента), а математически задаются соответствующим вероятностным пространством $(\Omega, \mathcal{A}, \mathbf{P})$, или, соответствующим законом распределения вероятностей. В теории вероятностей функция распределения $F(x)$ исследуемой случайной величины X предполагается полностью известной и по $F(x)$ мы можем предсказать, куда более вероятно попадают значения с.в. X. В математической статистике рассматриваемая задача является, в известном смысле обратной задачей теории вероятностей: по наблюдениям $x_1, x_2, ..., x_n$ за каким либо процессом требуется указать неизвестную нам функцию распределения. При этом наблюдения $x_1, x_2, ..., x_n$ интерпретируются как значения случайных величин $X_1, X_2, \ldots, X_n$. Мы будем считать, что $X_1, X_2, \ldots, X_n$ независимы и одинаково распределены с неизвестной функцией распределения $F(x)$, принадлежащей известному классу $\mathcal{F}$. Степень неизвестности (информированности) о классе $\mathcal{F}$ тоже может быть различной. Так, если имеются общие сведения о классе $\mathcal{F}$ типа гладкости, т.е. множество $\mathcal{F}$ является *бесконечномерным* подпространством какого-либо функционального пространства и нет предпочтений одного распределения перед другим, то мы занимаемся оцениванием распределений; если же в этой ситуации есть предпочтения одних распределений перед другими, например, есть гипотеза о том что искомое распределение имеет заданный вид, то мы занимаемся проверкой гипотез о согласии. Класс $\mathcal{F}$ может иметь следующий вид: $\mathcal{F} = \{F(x;\theta), \theta \in \Theta \subset R^m\}$, функция $F(x;\theta)$ <u>известна</u>, а параметр $\theta-$ <u>неизвестен</u>, т.е. $\mathcal{F}$ является *конечномерным*, и речь идет об определении этого параметра. Если нет предпочтений одного значения параметра перед другими, то мы занимаемся оценкой параметра, если же есть предпочтения одних значений параметра перед другими, то в этом случае задача ставится как задача проверки гипотез о параметрах (сюда, в частности, относится и доверительное оценивание, т.е. доверительное оценивание – это раздел проверки гипотез о параметрах) . В многомерном случае, т.е. когда мы изучаем многомерные случайные величины, то, как правило, производится анализ взаимосвязей, который включает в себя как задачи оценивания, так и задачи проверки гипотез.

Рассматриваемая задача является задачей теории вероятностей и поэтому в качестве исходного измеримого пространства $(\Omega, \mathcal{A})$ обычно рассматривают пространство всех его элементарных исходов, т.е. $\omega = (x_1, x_2, ..., x_n)$, а в качестве $\sigma-$ алгебры $\mathcal{A}$ берется борелевская $\sigma-$алгебра. Качественное описание явления – пространство $(\Omega, \mathcal{A})$ – наблюдателю известно, а мера $\mathbf{P} \in \mathcal{P}$ неизвестна. Ее надо указать по выборке.

Кроме того, известно, что часто нетривиальные обратные задачи математической физики оказываются некорректными. Нужны дополнительные ограничения (например, требование гладкости ответа), чтобы задача получила строгое математическое содержание. В математической статистике необходимость такого ограничения более или менее узким семейством $\mathcal{P}$ была интуитивно понята на первых же этапах ее развития.

Определение. Пусть $\mathcal{P}$ – семейство вероятностных мер (распределений) на измеримом пространстве $(\Omega, \mathcal{A})$. *Статистической структурой* называется тройка $(\Omega, \mathcal{A}, \mathcal{P})$.

Довольно часто семейство $\mathcal{P}$ имеет следующий вид $\mathcal{P} = \{\mathbf{P}_\theta, \theta \in \Theta \subset \mathbf{R}^m\}$, в этом случае речь идет о параметрическом оценивании.

Настоящее учебное пособие "Современные методы статистического оценивания параметров"состоит из девяти глав. Главы 1 – 3 посвящены изложению основ теории параметрического оценивания. Здесь вводятся базовые понятия, иллюстрируемые различными примерами. В последующих главах излагаются методы, появившиеся в последнее время, в частности GMM-метод (обобщенный метод моментов, активно используемый при оценивании экономических показателей), а также методы оценивания по неполной информации (цензурированные выборки), поэтому здесь также излагаются элементарные методы теории порядковых статистик. Последние главы посвящены асимптотической теории оценивания, а также последовательному оцениванию.

Ознакомление с материалами пособия ориентировано на формирование у обучающегося следующих компетенций:

- способность самостоятельно приобретать и использовать в практической деятельности новые знания и умения, стремиться к саморазвитию (ОК-5);

-способность применять методы анализа прикладной области на концептуальном, логическом, математическом и алгоритмическом уровнях (ПК-17);

-способность применять системный подход и математические методы в формализации решения прикладных задач (ПК-21).

Глава 1.

Методы нахождения оценок

Всюду в пособии мы предполагаем, что функция распределения известна нам с точностью до неизвестного параметра θ, принадлежащего подмножеству пространства $\mathbf{R}^k$, т.е. статистику известно семейство $\mathcal{F} = F(x;\theta), \theta \in \Theta \subset \mathbf{R}^k\}$, но неизвестен параметр θ, и имеется повторная выборка $x_1, x_2, ..., x_n$. Требуется, используя эти наблюдения указать значение из параметрического пространства Θ, которое можно было бы принять в качестве «истинного» значения параметра θ (или известной функции от этого параметра $g(\theta)$).

Определение 1. *Оценкой параметра θ (функции от параметра $g(\theta)$) будем называть измеримую функцию от наблюдений $T_n(x_1, x_2, ..., x_n)$, значение которой принимается в качестве «истинного» значения параметра (функции от параметра).*

Так как наблюдения являются значениями случайных величин, то оценки тоже будут случайными величинами. Поэтому, если какая-то из них используется для оценивания параметра θ, то для отдельных выборок могут получиться значения, сильно отличающиеся от истинного. Наша задача будет состоять в нахождении функции $T_n(x_1, x_2, ..., x_n)$ (а не частных значений этой функции), «хорошей» в определенном смысле.

Перейдем к рассмотрению практически употребительных методов отыскания оценок: метод моментов (ММ) и метод максимального правдоподобия (ММП).

1.1. Метод моментов.

Этот метод был предложен К.Пирсоном (1894) и является исторически первым общим методом построения оценок. В простейшем варианте он заключается в следующем.

Пусть $x_1, x_2, ..., x_n$ – повторная выборка с общей функцией распределения $F(x;\theta), \theta \in \Theta$, причем $\theta = (\theta_1, ..., \theta_k)$. Предположим, что первые k начальных моментов распределения существуют и выражаются функциями от параметров:

$$\mu_j(\theta_1\,,\ \ldots\,,\ \theta_k) = \mathbf{E}_\theta(\,X^j) = \int_{-\infty}^{\infty} x^j\, dF(x\,;\theta_1\,,\ \ldots\,,\ \theta_k)\,,\ j = 1\,,\ \ldots\,,\ k\ ,$$

Составим выборочные аналоги соответствующих моментов:

$$m_j = \int_{-\infty}^{\infty} x^j\, dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^j\,,\ j = 1\,,\ \ldots\,,\ k\,,$$

где $F_n(x)$ – эмпирическая функция распределения.

Определение 1. В качестве оценки параметра $\theta = (\theta_1\,,\ \ldots\,,\ \theta_k)$ возьмем решение системы уравнений:

$$\{\,m_j = \mu_j\,(\theta_1\,,\ \ldots\,,\ \theta_k)\,,\ j = 1\,,\ \ldots\,,\ k\ ,$$

т.е. *метод моментов* состоит в приравнивании теоретических моментов выборочным и последующем решении полученной системы уравнений.

Замечание. Для нахождения оценок по методу моментов можно рассматривать не только начальные, но и центральные и другие моменты, квантили и другие функционалы.

Пример 1. *Геометрическое распределение и отрицательное биномиальное распределение.* Пусть случайная величина X имеет геометрическое распределение

$$f(k;p) = \mathbf{P}\,(X=k) = p\,q^k\,,\ k = 0,1,...;q = 1-p\,.$$

Найдем математическое ожидание с.в. X:

$$\mathbf{E}\,(X) = \sum_{j=0}^{\infty} k\,q^k p = pq\sum_{j=1}^{\infty} k\,q^{k-1} = pq\,\left(\sum_{j=1}^{\infty} q^k\right)' = pq\,\left(\frac{q}{1-q}\right)' = pq\,\left(-1+\frac{1}{1-q}\right)' =$$

$$= \frac{pq}{(1-q)^2} = \frac{q}{p}\,.$$

Таким образом, для нахождения оценки параметра p по методу моментов получаем: $\bar{x} = \dfrac{1-p}{p}$, откуда $\hat{p} = \dfrac{1}{\bar{x}+1}$, где $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$.

Если же с.в. X имеет отрицательное биномиальное распределение, т.е. $\mathbf{P}\,(X=k) = C_{m+k-1}^{k} p^m q^k,\ k = 0,1,...$, то поскольку случайная величина $X = Y_1 + Y_2 + ... + Y_m$, где $\{Y_i\}$ – независимы и каждая имеет геометрическое распределение и поскольку $\mathbf{E}(Y_i) = \dfrac{q}{p},\ q = 1-p$, и $\mathbf{D}(Y_i) = \dfrac{q}{p^2}$, поэтому $\mathbf{E}(X) = m\dfrac{q}{p}$, и $\mathbf{D}(X) = m\dfrac{q}{p^2}$, m и p – неизвестны. Имеем систему

$\begin{cases} mq/p = \bar{x}\,, \\ mq/p^2 = s^2\,. \end{cases}$ Решением системы является $\hat{p} = \dfrac{\bar{x}}{s^2},\ \hat{m} = \left[\dfrac{\bar{x}^2}{s^2-\bar{x}}\right]$.

Если же параметр m известен, то $\mathbf{E}\,(X) = \dfrac{mq}{p}$ и $\hat{p} = \dfrac{m}{\bar{x}+m}$.

Пример 2. *Гамма-распределение.*

Пусть плотность распределения равна (здесь $\theta = (\alpha,\lambda)$)

$$f\,(x;\alpha,\lambda) = \frac{x^{\alpha-1}}{\lambda^{\alpha}\,\Gamma\,(\alpha)}\,e^{-\,x/\lambda}\,,\ x > 0\,,$$

α, λ – неизвестны и их необходимо оценить по выборке $x_1, x_2, ..., x_n$. Находим: $\mathbf{E}_\theta(X) = \alpha\,\lambda$, $\mathbf{D}_\theta(X) = \alpha\,\lambda^2$ и $\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$, $s^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$.

Тогда $\begin{cases} \alpha\,\lambda = \bar{x}\,, \\ \alpha\,\lambda^2 = s^2, \end{cases}$ откуда $\hat{\lambda} = \dfrac{s^2}{\bar{x}},\ \hat{\alpha} = \dfrac{\bar{x}^2}{s^2}$.

Пример 3. *Составное распределение Пуассона, описывающее ветвящийся процесс:*

$$\mathbf{P}(X = r;\mu,\lambda) = \sum_{n=0}^{\infty}\frac{(n\mu)^r e^{-n\mu}\lambda^n e^{-\lambda}}{r!n!},\ r = 0,1,....$$

Проверим, что $\sum_{r=0}^{\infty}\mathbf{P}(X = r;\mu,\lambda) = 1$. Действительно,

$$\sum_{r=0}^{\infty}\mathbf{P}(X = r;\mu,\lambda) = \sum_{n=0}^{\infty}\frac{e^{-n\mu}\lambda^n e^{-\lambda}}{n!}\sum_{r=0}^{\infty}\frac{(n\mu)^r}{r!} = \sum_{n=0}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!} = 1.$$

Найдем теперь математическое ожидание:

$$\mathbf{E}(X) = \sum_{r=0}^{\infty} r\cdot\mathbf{P}(X = r;\mu,\lambda) = \sum_{n=0}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}\sum_{r=0}^{\infty} r\cdot\frac{(n\mu)^r}{r!}e^{-n\mu} = \sum_{n=0}^{\infty} n\mu\cdot\frac{\lambda^n e^{-\lambda}}{n!} = \mu\lambda$$

и второй начальный момент:

$$\mathbf{E}(X^2) = \sum_{r=0}^{\infty} r^2\cdot\mathbf{P}(X = r;\mu,\lambda) = \sum_{n=0}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}\sum_{r=0}^{\infty} r^2\cdot\frac{(n\mu)^r}{r!}e^{-n\mu} = \sum_{n=0}^{\infty}(n\mu + n^2\mu^2)\cdot\frac{\lambda^n e^{-\lambda}}{n!} =$$

$$= \mu\lambda + \mu^2\lambda + \mu^2\lambda^2.$$

Отсюда, $\mathbf{D}(X) = \mu\lambda(1+\mu)$.

Получаем систему

$$\begin{cases} \mu\lambda = \bar{x}\,, \\ \mu\lambda + \mu^2\lambda = s^2, \end{cases}$$

откуда $\hat{\mu} = \dfrac{s^2 - \bar{x}}{\bar{x}}$, $\hat{\lambda} = \dfrac{\bar{x}^2}{s^2 - \bar{x}}$.

Пример 4. *Бета-распределение.* Пусть плотность распределения равна

$$f(x) = \frac{\Gamma\,(a+b)}{\Gamma\,(a)\Gamma\,(b)} x^{a-1}(1-x)^{b-1}\,,\ 0 < x < 1\,,\ a > 0\,,\ b > 0\,.$$

Тогда математическое ожидание равно

$$\mu_{\mathbf{1}} = \mathbf{E}\,(X) = \frac{\Gamma\,(a+b)}{\Gamma\,(a)\,\Gamma\,(b)} \int_0^1 x^a(1-x)^{b-1}dx = \frac{\Gamma\,(a+b)}{\Gamma\,(a)\,\Gamma\,(b)} \cdot \frac{\Gamma\,(a+1)\,\Gamma\,(b)}{\Gamma\,(a+b+1)} = \frac{a}{a+b}\,,$$

а второй начальный момент равен

$$\mu_2 = \frac{\Gamma\,(a+b)}{\Gamma\,(a)\,\Gamma\,(b)} \cdot \frac{\Gamma\,(a+2)\,\Gamma\,(b)}{\Gamma\,(a+b+2)} = \frac{a\,(a+1)}{(a+b)\,(a+b+1)},$$

откуда находим дисперсию

$$D\,(X) = \mu_2 - \mu_1^2 = \frac{ab}{(a+b)^2(a+b+1)} = \frac{\mu_1(1-\mu_1)}{a+b+1}\,.$$

В таком случае имеем систему:

$$\begin{cases} \bar{x} = \dfrac{a}{a+b} \quad \Rightarrow \quad a = \dfrac{b\,\bar{x}}{1-\bar{x}}\,, \\ s^2 = \dfrac{a}{a+b} \cdot \dfrac{b}{(a+b)\,(a+b+1)} = \dfrac{\bar{x}\,(1-\bar{x})^2}{b+1-\bar{x}} \quad \Rightarrow \quad \hat{b} = \dfrac{(1-\bar{x}\,)\,(\bar{x}-m_2)}{s^2}\,. \end{cases}$$

Значит, $\hat{a} = \dfrac{\bar{x}\,(\bar{x}-m_2)}{(1-\bar{x})\,s^2}$, где $m_2 = s^2 + \bar{x}^2$.

Пусть $\bar{x} = 0.3007\,,\ s^2 = 0.01966$. Тогда $\hat{b} = 6.78\,,\ \hat{a} = 2.92\,.$

Пример 5. Пусть случайная величина X имеет функцию распределения

$$F(x) = (1-\varepsilon)\,\Phi(x) + \varepsilon\Phi(x/\sigma),$$

где будем считать, что $0 < \varepsilon < 1\,,\ \sigma > 1$.

Имеется выборка $x_1, ..., x_n$ из этого распределения. Требуется по этой выборке оценить неизвестные параметры ε и σ^2. Для этого воспользуемся методом моментов, именно, рассмотрим 2-й и 4-й начальный моменты, имея в виду, что

$$\int x^2 d\Phi(x/\sigma) = \sigma^2, \int x^4 d\Phi(x/\sigma) = 3\sigma^4.$$

В этом случае

$$\int x^2 dF(x) = (1-\varepsilon) + \varepsilon\sigma^2 = 1 + (\sigma^2-1)\varepsilon, \int x^4 dF(x) = 3(1-\varepsilon) + 3\varepsilon\sigma^4 = 3(1+(\sigma^4-1)\varepsilon).$$

Согласно методу моментов приравняем теоретические моменты выборочным, а именно,

$$\begin{cases} 1 + (\sigma^2-1)\varepsilon = m_2 = \dfrac{1}{n}\sum\limits_{i=1}^n x_i^2, \\ 3(1+(\sigma^4-1)\varepsilon) = m_4 = \dfrac{1}{n}\sum\limits_{i=1}^n x_i^4. \end{cases}$$

Решение данной системы:

$$\hat{\sigma}^2 = \frac{m_4 - 3m_2}{3(m_2 - 1)}\,, \quad \hat{\varepsilon} = \frac{3m_2^2 - 6m_2 + 3}{m_4 - 6m_2 + 3}\,.$$

Пример 6. *Распределение Вейбулла.* Пусть $F(x) = 1 - \exp\left(\left(\dfrac{x}{\alpha}\right)^{\beta}\right)$, $x > 0$, и имеется выборка $x_1, ..., x_n$ из этого распределения. Требуется по этой выборке оценить неизвестные параметры α и β. Обозначим через x_λ квантиль порядка $0 < \lambda < 1$ распределения $F(x)$, $d_\lambda = -\ln(1-\lambda)$ и возьмем $0 < p < q < 1$. Тогда имеем систему

$$\begin{cases} p = 1 - \exp\left(-\left(\dfrac{x_p}{\alpha}\right)^{\beta}\right), \\ q = 1 - \exp\left(-\left(\dfrac{x_q}{\alpha}\right)^{\beta}\right). \end{cases}$$

Поэтому

$$\hat{\beta} = \frac{\ln d_p - \ln d_q}{\ln x_p - \ln x_q}\,,\ \hat{\alpha} = \exp\left(\frac{\ln d_q \ln x_p - \ln d_p \ln x_q}{\ln x_p - \ln x_q}\right).$$

Замечание. 1) Если система, из которой определяются параметры имеет частные производные такие, что определитель, составленный из производных не равен нулю в точке θ_0 и которые непрерывны, то по теореме об обратной функции существует состоятельное решение системы.

2)Оценки метода моментов $\hat{\theta}_n$ есть решение системы
$\mathbf{P}_n f \equiv \dfrac{1}{n}\sum_{i=1}^{n} f(X_i) = e(\theta) \equiv \mathbf{P}_\theta f$, откуда $\hat{\theta}_n = e^{-1}(\mathbf{P}_n f)$ и
$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = \sqrt{n}\,(e^{-1}(\mathbf{P}_n f) - e^{-1}(\mathbf{P}_\theta f))$.
Если $\mathbf{P}_n f$ асимптотически нормально и e^{-1} дифференцируемо, то правая часть будет также асимптотически нормальна.

Пример 7. ***Техника рандомизированного ответа.***

Рассмотрим ***задачу***. Вам задают вопрос, на который вы должны ответить либо "да"(ставите 1 – событие A), либо "нет"(ставите 0 – событие $\bar{A}$). По определенным соображениям вам не хочется отвечать правдиво и тогда применяют технику рандомизированного ответа (как безопасную для вас). Именно, проводят случайный эксперимент с исходами B и $\bar{B}$, результат которого опрашиваемому не сообщают, но который известен отвечающему. Пусть вероятность $\mathbf{P}(A) = p$ ***неизвестна***, и $\mathbf{P}(B) = \lambda \neq \frac{1}{2}$, которая ***известна***. Если наступает событие B, то отвечающий дает правдивый ответ, если наступает событие $\bar{B}$, то опрашиваемый отвечает наоборот, т.е. ставит 1, если имеет место событие A, и 0, если имеет место событие $\bar{A}$. Обозначим вероятность исхода 1 через q. Тогда

$$q = \mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(\bar{A}|\bar{B})\mathbf{P}(\bar{B}) = p\lambda + (1-p)(1-\lambda) = p(2\lambda - 1) + 1 - \lambda.$$

Вместо q нам известна её оценка в виде частоты $\dfrac{m}{n}$, тогда

$$\hat{p} = \frac{\frac{m}{n} - (1-\lambda)}{2\lambda - 1}.$$

Пусть, например, $\dfrac{m}{n} = 0.35$, $\lambda = 0.75$. Тогда $\hat{p} = \dfrac{0.35 - 0.25}{0.5} = 0.2$.

Разновидностью этого является техника вынужденного ответа, когда при наступлении события $\bar{B}$ всегда ставится 1. Здесь λ может быть равна и 0.5. В последнем случае

$$\hat{p} = \frac{\frac{m}{n} - (1-\lambda)}{\lambda}.$$

Пусть, например, $\frac{m}{n} = 0.6$, $\lambda = 0.5$. Тогда $\hat{p} = \frac{0.6-0.5}{0.5} = 0.2$.

1.2. Метод максимального правдоподобия.

Пусть:

а) в непрерывном случае задана плотность $f(x;\ \theta)$,

б) в дискретном случае задана вероятность $\mathbf{P}\,(X = x;\ \theta) = f(x;\ \theta)$, которую мы также будем называть плотностью.

Пусть $x_1\,,\ x_2\,,\ \ldots\,,\ x_n$ – повторная выборка с общей функцией плотности $f\,(x\,;\ \theta\,)$. При фиксированном значении $x = (\,x_1\,,\ x_2\,,\ \ldots\,,\ x_n\,)$ составим функцию правдоподобия

$$L_x(\,\theta\,) = \prod_{i=1}^{n} f(x_{\,i}\,;\ \theta\,),$$

которая есть функция параметра θ при фиксированной выборке.

Предположим, что функция правдоподобия непрерывна, а множество Θ замкнуто.

Определение 2. *<u>Оценкой максимального правдоподобия</u> параметра θ назовем решение уравнения*

$$L_x(\hat{\theta}\,) = \max_{\theta\,\in\,\Theta}\ L_x(\theta\,).$$

Если функция $\ln\,L_x(\theta\,)$ *дифференцируема* и максимум достигается во внутренней точке, то $\hat{\theta}$ удовлетворяет уравнению

$$\left\{ \frac{\partial}{\partial \theta_{\,i}}\,\ln\,L_x(\theta\,) = 0\,,\ \ i = 1\,,\ 2\,,\ \ldots\,,\ \ k.\right.$$

Пример 8. *Гипергеометрическое распределение.* Пусть X имеет гипергеометрическое распределение (M – неизвестно, N – известно). Построим оценку максимального правдоподобия для M/N. Пусть наблюдалось x. Тогда

$$\mathbf{P}_M(x) = \frac{C_M^x \cdot C_{N-M}^{n-x}}{C_N^n}.$$

Функция правдоподобия равна $L_x(M) = \mathbf{P}_M(x)$. Найдем такое M, чтобы $L_x(M)$ было максимальным, т.е. чтобы одновременно выполнялись неравенства

$(\,1\,)\ \ \dfrac{\mathbf{P}_{M-1}(x)}{\mathbf{P}_M(x)} \le 1$ и $(\,2\,)\ \ \dfrac{\mathbf{P}_{M+1}(x)}{\mathbf{P}_M(x)} \le 1\,.$

Имеем:

$$\frac{\mathbf{P}_M(x)}{\mathbf{P}_{M+1}(x)} = \frac{C_M^x \cdot C_{N-M}^{n-x}}{C_{M+1}^x \cdot C_{N-M-1}^{n-x}} = \frac{M+1-x}{M+1} \cdot \frac{N-M}{N-M-n+x}.$$

Тогда (1) эквивалентно неравенству

$$\frac{M+1}{N} \ge \frac{x}{n} + \frac{x}{n\,N} \quad \Leftrightarrow \quad \frac{x}{n} \le \frac{M}{N} + \frac{1}{N}\,\frac{n-x}{n},$$

а (2) – неравенству $\dfrac{M}{N} - \dfrac{n-x}{n\,N} \le \dfrac{x}{n}$.

Длина между левой и правой частью, равна $\dfrac{x}{n\,N} - \dfrac{n-x}{n\,N} = \dfrac{1}{N} < \dfrac{1}{n}$, поэтому интервал $\left(\dfrac{M}{N} - \dfrac{x}{n\,N}\,,\ \dfrac{M}{N} + \dfrac{n-x}{n\,N}\right)$ «накроет» только одно значение $\dfrac{x}{n}$, значит, оценкой максимального правдоподобия будет $\dfrac{\hat{M}}{N} = \dfrac{x}{n}$.

Пример 9. *Распределение Бернулли.* Пусть X имеет распределение Бернулли и имеется выборка $x_1\,,\ x_2\,,\ \ldots\,,\ x_n$, где

$$\mathbf{P}(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}.$$

Тогда функция правдоподобия равна

$$L_x(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^m(1-p)^{n-m} \quad \text{где} \quad m = \sum_{i=1}^{n} x_i.$$

Следовательно,

$$\ln\ L_x(p) = m\ \ln\ p + (n-m)\ \ln\ 1-p), \quad \frac{d}{dp}\ln\ L_x(p) = \frac{m}{p} - \frac{n-m}{1-p} = 0,$$

откуда $\hat{p} = \dfrac{m}{n}$.

Задача. Пусть вектор $X = (X_1\,,\ X_2\,,\ \ldots\,,\ X_m)$ имеет полиномиальное распределение

$$\mathbf{P}\left(X_1 = k_1\,,\ \ldots\,,\ X_m = k_m\right) = \frac{n\,!}{k_1! \cdot \ldots \cdot k_m!}\, p_1^{k_1} \cdot\ \ldots\ \cdot p_m^{k_m}\,,$$
$$p_i \geq 0\,,\ k_i \geq 0\,,\ p_1 +\ \ldots\ + p_m = 1\,,\ k_1 +\ \ldots\ + k_m = n.$$

Покажите, что ОМП для p_i есть $\hat{p}_i = \dfrac{k_i}{n}\,,\ i = 1\,,\ 2\,,\ \ldots\,,\ m\,.$

1.3. Обобщенный метод максимального правдоподобия.

Weiss и Wolfowitz предложили метод, весьма близкий методу максимального правдоподобия. Пусть повторная выборка $\{x_i\}_{1\leq i\leq n}$ имеет плотность распределения $f(x;\theta)\,,\ \theta \in \Theta$. Выберем произвольное положительное число r и пусть $L_x(\theta)$ – функция правдоподобия, построенная по выборке $x = (x_1, ..., x_n)$. В этом случае статистика

$$\delta_r(x) = \max_t \int_{t-r}^{t+r} L_x(\theta)\, d\theta$$

называется *оценкой максимальной вероятности* параметра θ. Устремляя $r \to 0$ в пределе получим оценку максимального правдоподобия.

Пример 10. Пусть X_i подчиняется показательному распределению с параметром сдвига $\theta \in R$, т.е. $f(x;\theta) = \exp(-(x-\theta))$ для $x \geq \theta$, и нуль в противном случае. Тогда

$$L_x(\theta) = h(X^{(1)} - \theta)\ \exp\left(-\sum_{i=1}^{n} X_i + n\theta\right), \quad h(x) = 1, x \geq 0;\ h(x) = 0, x < 0.$$

При фиксированном $r > 0$ получим $\delta_r(x) = X^{(1)} - r$. Квадратичный риск этой оценки равен

$$\mathbf{E}_\theta((\delta_r(X) - \theta)^2) = \mathbf{E}_\theta((X^{(1)} - r - \theta)^2) = r^2 - \frac{2r}{n} + \frac{2}{n^2}.$$

Минимизируя его по r, получаем оценку $\delta_{1/n}(X) = X^{(1)} - 1/n$.

1.4. Метод максимального произведения спейсингов (MPS).

Рассмотрим с.в. X с плотностью $f(x;a,b)$, $x \in (a,b)$. Параметры a и b − неизвестны. Пусть $x_n^{(1)} < x_n^{(2)} < ... < x_n^{(n)}$ есть вариационный ряд, построенный по исходной выборке. Положим $x_n^{(0)} = a\,,\ x_n^{(n+1)} = b$. *Спейсингом* назовем величину

$$D_i = \int_{x_n^{(i-1)}}^{x_n^{(i)}} f(x;a,b)\,dx = F(x_n^{(i)};a,b) - F(x_n^{(i-1)};a,b)\,,\ i = 1,2,...,n+1.$$

Очевидно, что $D_i \geq 0$ и $\sum_{i=1}^{n+1} D_i = 1$. Идея метода состоит в том, чтобы в качестве оценок параметров a и b взять такое значение, которое **максимизирует** среднее геометрическое спейсингов

$G = \left(\prod_{i=1}^{n+1} D_i\right)^{1/(n+1)}$, или его логарифм $H = \ln G$.

Сравним этот метод с методом МП. На основании теоремы о среднем можно записать:

$$\ln\, D_i = \ln\left(f(X_n^{(i)};a,b)(X^{(i)} - X_n^{(i-1)})\right) + R(X_n^{(i)}, X_n^{(i-1)};a,b) =$$

$$= \ln\, f(X_n^{(i)};a,b) + \ln(X_n^{(i)} - X_n^{(i-1)}) + R(X_n^{(i)}, X_n^{(i-1)};a,b).$$

Если a и b известны, то остаточный член $R(X_n^{(i)}, X_n^{(i-1)};a,b) = O(|X_n^{(i)} - X_n^{(i-1)}|) \underset{n\to\infty}{\overset{p}{\to}} 0$, т.е. вклад остаточного члена пренебрежимо мал и, следовательно, поведение $\partial \ln D_i/\partial a(\partial b)$ мало отличается от поведения $\partial \ln f(X_n^{(i)};a,b)/\partial a(\partial b)$, т.е. от оценок максимального правдоподобия, поэтому обе эти оценки состоятельны и асимптотически эффективны. Если же a и b известны, то остаточный член $R(X_n^{(i)}, X_n^{(i-1)};a,b)$, уже может и не стремиться к нулю при $n \to \infty$. Это ведет к существенной разнице в поведении H и L, поскольку H ограничена, а L может оказаться неограниченно большой.

Пример 11. Пусть $f(x;a,b) = 1/(b-a)$ для $a \leq x \leq b$, и равна нулю вне интервала (a,b). Тогда $D_i = (x_n^{(i)} - x_n^{(i-1)})/(b-a)$ и

$$H = \frac{1}{n+1}\left(\ln(x_n^{(1)} - a) + \sum_{i=1}^{n} \ln(x_n^{(i)} - x_n^{(i-1)}) + \ln(b - x_n^{(n)})\right) - \ln(b-a).$$

В этом случае

$$\left\{\frac{\partial H}{\partial a} = 0\,,\ \ \frac{\partial H}{\partial b} = 0 \ \ \Leftrightarrow\ \ \left\{\frac{1}{b-a} = \frac{1}{(n+1)(x_n^{(1)} - a)}\,,\ \ \frac{1}{b-a} = \frac{1}{(n+1)(b - x_n^{(n)})}\right.\right.\,,$$

откуда получаем решение

$$\hat{a} = \frac{n x_n^{(1)} - x_n^{(n)}}{n-1}\,,\ \hat{b} = \frac{n x_n^{(n)} - x_n^{(1)}}{n-1}.$$

Пример 12. При ***оценивании рисков финансовых активов*** важной характеристикой является следующая условная вероятность $(y > 0)$

$$H_u(y) = \mathbf{P}\,(X - u < y\,|\,X \geq u) = \frac{\mathbf{P}\,(X - u < y\,,\ X > u)}{\mathbf{P}\,(X \geq u)} = \frac{\mathbf{P}\,(u \leq X < y + u)}{\mathbf{P}\,(X \geq u)} =$$

(1)
$$= \frac{F(y+u) - F(u)}{1 - F(u)}\,.$$

Pickands, Balkema, de Haan (1974, 1975) показали, что при больших объемах выборки и больших u

$$H_u(y) \approx G_{\xi,\beta}(y)\,,$$

где

$$(2) \qquad G_{\xi,\beta}(x) = \begin{cases} 1-(1+\xi\, x/\beta)^{-1/\xi},\ \xi \neq 0\,;\ x > 0\,(\xi > 0),\ 0 < x < -\beta/\xi\ (\xi < 0)\,, \\ 1-\exp\left(-x/\beta\right),\ \xi = 0\,,\ x > 0, \end{cases}$$

есть обобщенное распределение Парето (ОРП).

Из (1) получаем, что для оценки $F(x)$ при $x > u$ можно воспользоваться равенством: $F(x) = (1-F(u))\,G_{\xi,\beta}(x-u) + F(u)$, откуда $F(x) = 1-(1-F(u))\,(1-G_{\xi,\beta}(x-u))$.

Оценим вероятность $\mathbf{P}\,(X \ge u) = 1 - F(u)$ с помощью частоты $\dfrac{N_u}{N}$ по выборке объема N, где $N_u = \frac{1}{N}\sum_{i=1}^{N} I(X_i \ge u)$. Заметим, что если Y имеет ОПР с функцией распределения (2), то

$$\mathbf{E}(Y) = \frac{\beta}{1-\xi}\ \ (\xi < 1)\,,\ \mathbf{D}(Y) = \frac{\beta^2}{(1-\xi)^2(1-2\xi)}\ \ (\xi < 1/2).$$

Отсюда следует, что оценки по методу моментов будут равны

$$\hat{\xi} = \frac{s^2-\bar{x}^2}{2s^2},\ \hat{\beta} = \frac{\bar{x}\,(s^2+\bar{x}^2)}{2s^2}.$$

Тогда в качестве оценки $F(x)$ при $x > u$ можно взять :

$$\hat{F}(x) = 1 - \frac{N_u}{N}\left(1 + \frac{\hat{\xi}\,(x-u)}{\hat{\beta}}\right)^{1/\hat{\xi}},$$

и осуществить прогноз для $x > u$.

Глава 2.

Риск оценивания. Критерии качества оценок.

2.1. Несмещенность.

Хорошие ли оценки получаются при использовании метода моментов или метода максимального правдоподобия?

Итак, нам надо выбрать функцию от наблюдений $T = T(x)$ в качестве оценки неизвестного параметра. Чтобы понять, как выбирать $T(x)$, необходимо сравнить последствия от использования той или иной оценки. Общепринятый ныне подход к этому вопросу, который установился после работ А.Вальда, заключается в следующем: если мы примем решение, что $T(x) = d$, то при условии, что распределение X равно $\mathbf{P}_\theta$, это приведет к потере $L(d, \theta)$. Средний убыток, возникающий от применения T в длинном ряду повторений эксперимента, приближенно равен математическому ожиданию $\mathbf{E}_\theta(L(T(X), \theta))$, вычисленному в предположении, что распределение X равно $\mathbf{P}_\theta$. Это математическое ожидание называется функцией риска и обозначается как $R(T, \theta) = \mathbf{E}_\theta(L(T(X), \theta))$.

Исторически потери сначала измерялись абсолютным отклонением $|\,T - \theta\,|$, но аналитически функция $y = |\,x\,|$ является не совсем удобной и тогда было предложено измерять потери с помощью квадратичной функции $L\,(T, \theta) = (T - \theta)^2$.

Заметим, что мы хотим строить «хорошие» оценки. Что это такое?

Определение 3. *Оценка T параметра θ называется* ***недопустимой*** *(по отношению к функции потерь L), если существует такая оценка T^*, что*
1) $R\,(T^*, \theta) \le R\,(T, \theta), \quad \forall \theta \in \Theta$,
2) *существует $\theta_0 \in \Theta$ такое, что $R\,(T^*, \theta_0) < R\,(T, \theta_0)$.*

Оценка, которая не является недопустимой, называется ***допустимой***.

Чтобы исключить из рассмотрения тривиальные оценки вида $T \equiv \theta_0$, которую большинство разумных оценок не доминируют, но которая бесполезна для статистического анализа, то требуется ограничиться **классом** каких-либо оценок и искать в этом классе наилучшую оценку в смысле рисков $R\,(T, \theta)$. Поэтому рассматривают класс несмещенных оценок, класс правильных оценок и т.д. Конечно описанные парадоксальные ситуации возникают потому, что на множестве всех функций рисков нельзя установить полного порядка и чтобы упорядочить множество всех рисков, а значит, и множество всех оценок, необходимо рассмотреть на множестве всех рисков какой-либо функционал $Q\,(R\,(T, \theta))$. Это приводит статистиков к понятию байесовских оценок, минимаксных оценок и их комбинаций. Однако мы не будем рассматривать эту проблему в общем виде, а ограничимся функциями риска.

Определение 4. *Оценку T_n для функции $g\,(\theta\,)$ будем называть <u>несмещенной</u>, если для всех n и θ среднее значение T_n равно $g\,(\theta\,)$, т.е.*

$$\mathbf{E}_\theta(T_n(X)) = g\,(\theta\,).$$

Понятие несмещенности как отсутствие «систематической ошибки», было введено Гауссом (1821) в работе по теории наименьших квадратов.

Пример 13. Пусть $X_i \in N(\theta\,,\ \sigma^2)$, $T_n = \frac{1}{n}\sum_{i=1}^n X_i$. Тогда

$$\mathbf{E}_\theta(T_n(X)) = \mathbf{E}_\theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbf{E}_\theta(X_i) = \theta,$$

т.е. статистика T_n является несмещенной оценкой параметра θ.

Пример 14. Рассмотрим статистику $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, где $X_i \in N(a\,,\ \sigma^2)$, $\theta = \sigma^2$ как оценку параметра σ^2. Тогда

$$\mathbf{E}_\theta(s^2) = \frac{1}{n}\,\mathbf{E}_\theta\left(\sum_{i=1}^{n}(X_i - \bar{X}\,)^2\right) = \frac{1}{n}\,\mathbf{E}_\theta\left(\sum_{i=1}^{n}(X_i - a)^2 - n\,(\bar{X} - a)^2\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}_\theta(X_i - a)^2 - \mathbf{E}_\theta(\bar{X} - a)^2 = \mathbf{D}_\theta(X_1) - \mathbf{D}_\theta(\bar{X}) =$$

$$= \sigma^2 - \frac{\sigma^2}{n} = \sigma^2 \cdot \frac{n-1}{n},$$

т.е. s^2 является смещенной оценкой σ^2, однако ее смещение $\left(-\frac{\sigma^2}{n}\right) \underset{n\to\infty}{\to} 0$.

Рассмотрим статистику $s_1^2 = \frac{n}{n-1}\,s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Для нее

$$\mathbf{E}_\theta(s_1^2) = \mathbf{E}_\theta\left(\frac{n}{n-1}\,s^2\right) = \frac{n}{n-1}\,\mathbf{E}_\theta(s^2) = \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 = \sigma^2,$$

т.е. s_1^2 является несмещенной оценкой дисперсии σ^2.

Замечание. В примерах 12 и 13 нормальность была не обязательна, лишь бы было $\mathbf{E}_\theta(X) = \theta$ – в примере 12 или $\mathbf{D}_\theta(X) = \sigma^2$ – в примере 13.

Пример 15. Пусть $X_i \in B\,(\,1\,, p)$, независимы, $i = 1\,,\ 2\,,\ \ldots\,,\ n$.
Поскольку $m = \sum_{i=1}^{n} X_i \in B\,(n\,, p)$, то

$$\mathbf{E}_\theta\left(\frac{1}{n}\ \sum_{i=1}^{n} X_i\right) = \mathbf{E}_\theta\left(\frac{m}{n}\right) = p, \quad \mathbf{E}_\theta\left(\frac{m}{n}\cdot\frac{m-1}{n-1}\right) = p^2,$$

$$\mathbf{E}_\theta\left(\frac{m}{n}\cdot\frac{n-m}{n-1}\right) = p\,(\,1 - p).$$

Имеются случаи, когда несмещенных оценок не существует.

Пример 16. Пусть случайная величина X имеет биномиальное распределение $B\,(n\,,\ p)$. В этом случае не существует несмещенной оценки для функции $g\,(p) = \dfrac{1}{p}$. Действительно, пусть $T\,(x)$ – несмещенная оценка функции $g\,(p)$, т.е.

$$\mathbf{E}_p(T\,(X)) = \frac{1}{p},$$

или

$$\sum_{m\,=\,0}^{n} T(m)\,C_n^m p^m(\,1 - p)^{n-m} = \frac{1}{p}.$$

Слева стоит ограниченная функция p, но, устремляя p к нулю, получаем, что правая часть не ограничена – противоречие. Значит, такой функции $T\,(x)$ не существует. Однако для геометрического распределения несмещенная оценка для $\dfrac{1}{p}$ есть $\bar{x}$.

Пример 17. Пусть X равномерно распределенная на множестве $\{1,\ 2,\ 3,\ \ldots\,,\ \theta\,\}$ случайная величина. Наблюдается выборкаповторная выборка (выбор с возвращением) $x_1,\ x_2,\ \ldots\,,\ x_n$ и пусть $x_n^{(n)} = \max\limits_{1\le\,i\,\le\,n} x_i = k$. В качестве оценки параметра θ возьмем

$t_n = \dfrac{k^{n+1} - (k-1)^{n+1}}{k^n - (k-1)^n}$. Покажем, что эта оценка является несмещенной (по утверждению Феллера [15], с.240, эта оценка использовалась во 2-й мировой войне для оценки объема выпуска военной продукции).

Найдем сначала распределение величины $Y = X_n^{(n)}$. Имеем:

$$\mathbf{P}\,(X_n^{(n)} \le k) = \mathbf{P}\,(\max_i X_i \le k) = \mathbf{P}^n(X_1 \le k) = \left(\frac{k}{\theta}\right)^n.$$

Значит,

$$\mathbf{P}\,(X_n^{(n)} = k) = \mathbf{P}\,(X_n^{(n)} \le k) - \mathbf{P}(\,X_n^{(n)} \le k-1) = \frac{k^n - (k-1)^n}{\theta^n}.$$

Отсюда,

$$\mathbf{E}(\,t_n) = \sum_{k=1}^{\theta} \frac{k^{n+1} - (k-1)^{n+1}}{k^n - (k-1)^n} \cdot \frac{k^n - (k-1)^n}{\theta^{\,n}} =$$

$$= \frac{1}{\theta^{\,n}} \sum_{k=1}^{\theta} (\,k^n - (k-1)^n) = \frac{\theta^{\,n+1}}{\theta^{\,n}} = \theta,$$

т.е. t_n – несмещенная оценка.

Аналогично показывается, что оценка $t_n^{(r)} = \dfrac{k^{n+r} - (k-1)^{n+r}}{k^n - (k-1)^n}$ будет несмещенной оценкой функции $\theta^{\,r}$.

Заметим также, что для достаточно больших θ и $x_n^{(n)} = k$ достаточно большом, $t_n = \dfrac{k^{n+1} - (k-1)^{n+1}}{k^n - (k-1)^n} = \dfrac{k^{n+1}(1-(1-1/k)^{n+1})}{k^n(1-(1-1/k)^n)} \approx k\dfrac{n+1}{n}$, поэтому для оценки параметра θ можно использовать статистику $\hat{t}_n = \dfrac{n+1}{n} X_n^{(n)}$.

Пусть $n = 10$ и наблюдается максимальное значение $x_{10}^{(10)} = 910$. Тогда $t_n = \dfrac{910^{11} - 909^{11}}{910^{10} - 909^{10}} = 1000.45$, а $\hat{t}_n = \dfrac{11}{10} \cdot 910 = 1001$.

Другой простой пример состоит в следующем: пусть вы выходите на платформу метро и наблюдаете номер подошедшего поезда и этот номер равен 5 (будем считать, что номера поездов 1, 2, …, θ). Тогда $n = 1$ и $t_n = \dfrac{5^2 - 4^2}{5 - 4} = 9$, т.е. в качестве оценки θ берем 9. Предположим, что мы наблюдаем еще один номер поезда (это может быть в другой день (тогда можно считать, что выборка повторная) и $x_2^{(2)} = 5$. В таком случае оценка равна $t_n = \dfrac{5^3 - 4^3}{5^2 - 4^2} = \dfrac{61}{9} = 6.78$, т.е. за оценку числа поездов можно взять 7.

Замечание. Для оценки количества поездов метро можно считать, что выборка повторная, но для оценки объема военной продукции, ясно, что выборка бесповторная. Насколько сильно будут отличаться результаты оценивания?

Для этого рассмотрим эту задачу при выборе *без возвращения.* Эта ситуация больше подходит для практических моделей (особенно оценки объема производимой военной продукции).

Итак, $f(x;\theta) = \dfrac{1}{\theta}$, $x = 1, 2, \dots, \theta$. Пусть $Y = X_n^{(n)}$, $n \le \theta$. Тогда

$$G(y) = \mathbf{P}(Y \le y) = \mathbf{P}(X_1 \le y, X_2 \le y, ..., X_n \le y) = \frac{y(y-1)\cdot ... \cdot(y-n+1)}{\theta(\theta-1)\cdot ... \cdot(\theta-n+1)} = \frac{C_y^n}{C_\theta^n}.$$

Действительно, имеем:

$$\mathbf{P}(X_1 \le y) = \frac{y}{\theta}\,,\ \mathbf{P}(X_1 \le y, X_2 \le y) = \sum_{k=1}^{y} \mathbf{P}(X_2 \le y | X_1 = k)\mathbf{P}(X_1 = k) = \frac{y-1}{\theta-1}\cdot\frac{y}{\theta}.$$

Применяя метод математической индукции, получим общий случай.

Отсюда следует, что

$$g(y) = \mathbf{P}(Y = y) = G(y) - G(y-1) = \frac{C_y^n - C_{y-1}^n}{C_\theta^n} = \frac{C_{y-1}^{n-1}}{C_\theta^n},\ \text{для } y \ge n+1,$$

и если $y = n$ (выбираются все номера 1, 2, ..., n), то

$$g(y) = \frac{1}{C_\theta^n} \quad \Rightarrow \quad g(y) = \frac{C_{y-1}^{n-1}}{C_\theta^n}\,,\ n \le y \le \theta\,.$$

Рассмотрим оценку $t_n = \dfrac{n+1}{n}Y - 1$. Имеем

$$S(n,\theta) = \mathbf{E}(t_n) = \sum_{y=n}^{\theta} \frac{(n+1)y}{n} g(y) - 1 = \sum_{y=n}^{\theta} \frac{(n+1)y}{nC_\theta^n}\frac{(y-1)!}{(n-1)!(y-n)!} - 1 = (n+1)\sum_{y=n}^{\theta}\frac{C_y^n}{C_\theta^n} - 1.$$

Обозначим $\theta = n + k$ и покажем, что $S(n,\theta) = \theta$, т.е. оценка *несмещенная*.

Именно, надо показать, что $(n+1)\sum\limits_{y=n}^{n+k} \dfrac{C_y^n}{C_\theta^n} = n + k + 1$.

Доказывать будем по индукции. Надо установить, что

$$(n+1)\frac{C_n^n + C_{n+1}^n + ... + C_{n+k}^n}{C_{n+k}^n} = n + k + 1.$$

Если $k = 0$, то $(n+1)\dfrac{C_n^n}{C_n^n} = n + 1$ – верно.

Если $k = 1$, то $(n+1)\dfrac{C_n^n + C_{n+1}^n}{C_{n+1}^n} = n + 2 \quad \Leftrightarrow \quad (n+1)\dfrac{1+n+1}{n+1} = n + 2$ – верно.

Пусть это равенство верно для k, т.е. $(n+1)\dfrac{C_n^n + C_{n+1}^n + ... + C_{n+k}^n}{C_{n+k}^n} = n+k+1$, и покажем, что $(n+1)\dfrac{C_n^n + C_{n+1}^n + ... + C_{n+k}^n + C_{n+k+1}^n}{C_{n+k+1}^n} = n + k + 2$. Но последнее равносильно

$$(n+k+1)\frac{C_{n+k}^n}{C_{n+k+1}^n} + (n+1)\frac{C_{n+k}^n}{C_{n+k+1}^n} = n+k+2 \quad \Leftrightarrow \quad (n+k+1)\frac{k+1}{n+k+1} + n+1 = k+1+n+1 = n+k+2$$

(здесь можно также воспользоваться соотношением $C_{\theta+1}^{n+1} = \sum_{y=n}^{\theta} C_y^n,\ \theta \ge n$). Доказательство завершено.

Для примера Феллера $y = 910, n = 10$ получаем $t_n = \frac{11}{10}\cdot 910 - 1 = 1000$.

Найдем дисперсию нашей оценки.

Из равенства $\mathbf{E}\left(\dfrac{n+1}{n}Y - 1\right) = \theta$ следует, что $\mathbf{E}(Y) = \dfrac{n\,(\theta+1)}{n+1}$. Тогда

$$\mathbf{D}(t_n) = \mathbf{D}\left(\frac{n+1}{n}Y - 1\right) = \mathbf{D}\left(\frac{n+1}{n}Y\right) = \frac{(n+1)^2}{n^2}\mathbf{E}(Y) - (\theta+1)^2.$$

Теперь

$$\mathbf{E}(Y^2) = \sum_{y=n}^{\theta} y^2 \frac{C_{y-1}^{n-1}}{C_\theta^n} = \frac{1}{C_\theta^n}\sum_{y=n}^{\theta} y^2 C_{y-1}^{n-1} = \frac{n}{C_\theta^n}\sum_{y=n}^{\theta} yC_y^n = \frac{n}{C_\theta^n}\sum_{y=n}^{\theta}(y+1-1)C_y^n =$$

$$= \frac{n(n+1)}{C_\theta^n}\sum_{y=n}^{\theta} C_{y+1}^{n+1} - \frac{n}{C_\theta^n}\sum_{y=n}^{\theta} C_y^n = \frac{n(n+1)}{C_\theta^n} C_{\theta+2}^{n+2} - \frac{n}{C_\theta^n} C_{\theta+1}^{n+1} = \frac{n(\theta+2)(\theta+1)}{n+2} - \frac{n(\theta+1)}{n}.$$

Значит,

$$b^2 = \mathbf{D}(t_n) = \frac{(n+1)^2(\theta+2)(\theta+1)}{n(n+2)} - \frac{(n+1)(\theta+1)}{n} = \frac{(\theta+2)(\theta+1)}{n(n+2)} - \frac{(\theta+1)}{n}.$$

Поскольку $\theta \geq n$, то $\mathbf{D}(t_n) \geq 0$ (равенство нулю при $n = \theta$).

Для повторной выборки дисперсия будет больше подсчитанной, поэтому мы ограничимся случаем бесповторной выборки.

При $\theta = 1000, n = 10$ получаем $b^2 = 8258.25$; при $\theta = 1000, n = 100$ получаем $b^2 = 19.8$, а при $\theta = 1000, n = 500$ получаем $b^2 = 2.0$, т.е. можно считать, что при $n = 100$ погрешность оценивания не превосходит 13, а при $n = 500$ – **четырех.**

2.2. Эффективность.

Рассмотрим повторную выборку $x_1, x_2, \dots, x_n$, где X_i имеют распределение Пуассона $P(\theta)$. Параметр θ можно несмещенно оценить: 1) с помощью $\bar{x}$, т.е. $\hat{\theta} = \bar{x}$; 2) с помощью $\tilde{\theta} = s_1^2$. Какая из этих двух оценок лучше? Чтобы ответить на этот вопрос, рассмотрим дисперсии этих оценок: чем меньше дисперсия оценки (т.е. чем меньше разброс), тем она лучше.

(1) $\mathbf{D}_\theta(\bar{X}) = \mathbf{D}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}_\theta(X_i) = \frac{\theta}{n}.$

(2) Используя характеристическую функцию, можно показать, что

$$\mathbf{D}_\theta(S_1^2) - \mathbf{D}_\theta(\bar{X}) = \frac{2\theta^2}{n-1} + \frac{\theta(3+2\theta)}{n(n-1)} + \frac{2\theta}{n^2(n-1)} > 0,$$

поэтому

$$\mathbf{D}_\theta(\bar{X}) < \mathbf{D}_\theta(S_1^2).$$

Можно ли найти оценку с еще меньшей дисперсией, более того, можно ли построить оценку с дисперсией равной нулю, т.е. чтобы $\mathbf{D}_\theta(T_n) = 0$, или, что равносильно, чтобы $\mathbf{P}_\theta(T_n) = 1$?

Мы сейчас рассмотрим неравенство, которое показывает, что в некоторых случаях (так называемый регулярный случай), существует нижняя граница для дисперсии несмещенных оценок, т.е. если объем выборки фиксирован, то нельзя сделать дисперсию меньше некоторой положительной границы, а уж тем более, равной нулю.

Через $f = f(x;\theta)$ будем обозначать плотность распределения в непрерывном случае и вероятность – в дискретном случае.

Предположим, что выполнены следующие условия (*регулярный случай*):

(1) Множество $\{x : f(x;\theta) > 0\}$ ***не зависит от параметра*** θ.

(В некоторых пособиях это условие почему-то игнорируется, хотя это тривиальное условие из математического анализа о возможности войти под знак интеграла; в случае его нарушения выводы могут оказаться неверными ! Можно сразу потребовать такую возможность).

(2) При каждом θ, принадлежащем некоторому невырожденному интервалу A, для почти всех x существуют производные $\frac{\partial \ln f}{\partial\theta}, \frac{\partial^2 \ln f}{\partial\theta^2}, \frac{\partial^3 \ln f}{\partial\theta^3}$.

(3) При каждом θ из A имеем $\left|\frac{\partial f}{\partial\theta}\right| < H_1(x)$, $\left|\frac{\partial^2 f}{\partial\theta^2}\right| < H_2(x)$, $\left|\frac{\partial^3 \ln f}{\partial\theta^3}\right| < H_3(x)$, где функции $H_1(x)$ и $H_2(x)$ интегрируемы на $\mathbf{R}$ и $\int_{-\infty}^{\infty} H_3(x)\, f(x;\theta)\, dx < M$, причем M не зависит от θ.

(4) При каждом θ из A ***информационное количество Фишера*** $I(\theta) = \int\limits_{-\infty}^{\infty} \left(\frac{\partial \ln f(x\,;\theta)}{\partial\theta} \right)^2 f(x\,;\theta)\,dx$ конечно и положительно, т.е. $0 < I(\theta) < \infty$.

ТЕОРЕМА 1. (Неравенство Крамера-Рао) *Пусть выполнены условия регулярного случая (1) – (4), $T_n(X)$ – оценка параметра θ, такая, что $\mathbf{E}_\theta(T_n(X)) = \theta + b(\theta)$, где функция $b(\theta)$ дифференцируема. Тогда среднее квадратичное отклонение оценки $T_n(X)$ от истинного значения параметра θ удовлетворяет неравенству*

$$\mathbf{E}_\theta(T_n - \theta)^2 \geq \frac{(1 + b'(\theta))^2}{n\,I(\theta)} + b^2(\theta).$$

Следствие 1. *Если $\hat\theta_n(X)$ – несмещенная оценка функции $g(\theta)$, т.е. такая, что $\mathbf{E}_\theta(\hat\theta_n(X)) = g(\theta)$, то*

$$\mathbf{E}_\theta(\hat\theta_n - \theta)^2 \geq \frac{(g'(\theta))^2}{n\,I(\theta)}.$$

Следствие 2. *Если $\hat\theta_n(X)$ – несмещенная оценка параметра θ, т.е. такая, что $\mathbf{E}_\theta(\hat\theta_n(X)) = \theta$, то*

$$\mathbf{E}_\theta(\hat\theta_n - \theta)^2 \geq \frac{1}{n\,I(\theta)}.$$

Д о к а з а т е л ь с т в о. Рассмотрим непрерывный случай (в дискретном случае доказательство аналогично) и для простоты записи будем доказывать следствие 2.

По условию теоремы имеем

$$\mathbf{E}_\theta(T_n(X)) = \theta \quad \Leftrightarrow \quad \int_{R^n} T_n(x_1, x_2, \ldots, x_n) \prod_{i=1}^{n} f(x_i;\theta)\,dx = \theta.$$

Возьмем от обеих частей производную по θ и поменяем знаки $\int$ и $\frac{\partial}{\partial\theta}$ (по условиям теоремы это можно сделать). Тогда получим

$$\int_{R^n} T_n(x_1, x_2, \ldots, x_n) \sum_{j=1}^{n} \frac{f'_\theta(x_j;\theta)}{f(x_j;\theta)} \prod_{i=1}^{n} f(x_i;\theta)\,dx = 1,$$

т.е.

$$\mathbf{E}_\theta \left(T_n \cdot \sum_{j=1}^{n} \frac{f'_\theta(X_j;\theta)}{f(X_j;\theta)} \right) = 1.$$

Лемма 1. *Если конечны приведенные ниже математические ожидания, то*

$$\mathbf{E}^2(X \cdot Y) \leq \mathbf{E}(X^2) \cdot \mathbf{E}(Y^2).$$

Д о к а з а т е л ь с т в о. Имеем:

$$\mathbf{E}(t\,X - Y)^2 \geq 0$$

для любого действительного t, или

$$t^2\,\mathbf{E}(X^2) - 2\,t\,\mathbf{E}(X \cdot Y) + \mathbf{E}(Y^2) \geq 0,$$

т.е. квадратичный трехчлен неотрицателен. Это будет тогда и только тогда, когда

$$\mathbf{E}^2(X \cdot Y) - \mathbf{E}(X^2) \cdot \mathbf{E}(Y^2) \leq 0.$$

Возвращаясь к доказательству теоремы, продифференцируем обе части равенства

$$\int_{R^n} \prod_{i=1}^{n} f(x_i; \theta)\, dx = 1,$$

по θ. Мы получим

$$\mathbf{E}_\theta \left(\sum_{j=1}^{n} \frac{f'_\theta(X_j;\theta)}{f(X_j;\theta)} \right) = 0.$$

Из предыдущего следует, что

$$\mathbf{E}_\theta \left((T_n - \theta) \cdot \sum_{j=1}^{n} \frac{f'_\theta(X_j;\theta)}{f(X_j;\theta)} \right) = 1.$$

Согласно неравенству, доказанному в лемме 1,

$$\mathbf{E}_\theta((T_n - \theta))^2 \cdot \mathbf{E}_\theta \left(\sum_{j=1}^{n} \frac{f'_\theta(X_j;\theta)}{f(X_j;\theta)} \right)^2 \geq 1.$$

Так как

$$\mathbf{E}_\theta \left(\sum_{j=1}^{n} \frac{f'_\theta(X_j;\theta)}{f(X_j;\theta)} \right)^2 = \mathbf{D}_\theta \left(\sum_{j=1}^{n} \frac{f'_\theta(X_j;\theta)}{f(X_j;\theta)} \right) = \sum_{j=1}^{n} \mathbf{D}_\theta \left(\frac{f'_\theta(X_j;\theta)}{f(X_j;\theta)} \right) =$$

$$= \sum_{j=1}^{n} \mathbf{E}_\theta \left(\frac{f'_\theta(X_j;\theta)}{f(X_j;\theta)} \right)^2 = n \cdot I(\theta),$$

то из предыдущего неравенства и этого замечания получаем неравенство Крамера-Рао.

Пример 18. Пусть задана повторная выборка $x_1, \; x_2, \; \dots, \; x_n$, где X_i имеют нормальное распределение $N(\theta, \; \sigma^2)$, параметр θ неизвестен и его требуется оценить по заданной выборке. В качестве оценки будем использовать $\hat{\theta} = \bar{x}$. Тогда имеем

$$\mathbf{D}_\theta \left(\hat{\theta}_n \right) = \mathbf{D}_\theta \left(\frac{1}{n} \sum_{j=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{j=1}^{n} \mathbf{D}_\theta (X_i) = \frac{1}{n^2} \cdot n\, \sigma^2 = \frac{\sigma^2}{n}.$$

Кроме того,

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \theta)^2}{2\sigma^2} \right), \quad -\infty < x < \infty,$$

$$I(\theta) = \mathbf{E}_\theta \left(\frac{\partial \ln f(X_1;\theta)}{\partial \theta} \right)^2 = \mathbf{E}_\theta \left(\frac{(X_1 - \theta)^2}{\sigma^4} \right) = \frac{1}{\sigma^2}, \quad \text{т.е.}$$

$$\frac{1}{n\, I(\theta)} = \frac{\sigma^2}{n} = \mathbf{E}_\theta((\hat{\theta}_n - \theta)^2),$$

значит, в неравенстве Крамера-Рао имеет место равенство и меньшей дисперсии достичь нельзя.

Рассмотрим функцию $g(\theta) = \theta^2$. В качестве несмещенной оценки функции θ^2 возьмем статистику $T(X) = \bar{X}^2 - \dfrac{\sigma^2}{n}$.

Действительно, $\mathbf{E}(T(X)) = \mathbf{E}(\bar{X}^2) - \dfrac{\sigma^2}{n}$. Но $\bar{X}$ имеет нормальное распределение с параметрами $\left(\theta, \dfrac{\sigma^2}{n}\right)$, поэтому

$$\mathbf{E}(T(X)) = \theta^2 + \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = \theta^2.$$

Поскольку $g'(\theta) = 2\theta$, то нижняя граница Крамера-Рао равна $\dfrac{4\theta^2\sigma^2}{n}$.

Найдем дисперсию оценки $T(X)$. Имеем

$$\mathbf{D}(T(X)) = \mathbf{D}\left(\bar{X}^2 - \frac{\sigma^2}{n}\right) = \mathbf{D}(\bar{X}^2) = \mathbf{E}(\bar{X}^4) - (\mathbf{E}(\bar{X}^2))^2.$$

Кроме того, $\mathbf{D}(\bar{X}^2) = \dfrac{2\sigma^4}{n^2} + \dfrac{4\theta^2\sigma^2}{n}$. Поэтому нижняя граница Крамера-Рао здесь не достигается, поскольку $\dfrac{2\sigma^4}{n^2} + \dfrac{4\theta^2\sigma^2}{n} > \dfrac{4\theta^2\sigma^2}{n}$.

Определение 5. *Несмещенную оценку θ_n^* будем называть* <u>*эффективной*</u> *для параметра θ распределения $f\,(x\,;\theta\,)$, если в неравенстве Крамера-Рао имеет место равенство, т.е.*

$$\mathbf{E}_\theta((\theta_n^* - \theta)^2 = \mathbf{D}_\theta(\theta_n^*) = \frac{1}{n\,I\,(\theta\,)}.$$

Так оценка $\bar{x}$ является эффективной для параметра θ нормального распределения. Точно так же $\bar{x}$ является эффективной оценкой параметра θ распределения Пуассона $P\,(\theta\,)$, а частота m/n – эффективная оценка для параметра p биномиального распределения $B\,(n\,,\ p\,)$. Однако для некоторых распределений эта граница является недостижимой (см. предыдущий пример). Еще один пример: несмещенная оценка $s_1^2 = \frac{1}{n-1}\,\sum_{i\,=\,1}^n (x_i - \bar{x}\,)^2$параметра σ^2 нормального распределения имеет дисперсию $\mathbf{D}_\theta(s_1^2) = \frac{2\,\sigma^4}{n-1}$, а информационное количество Фишера для σ^2 равно $I\,(\sigma^2\,) = \frac{1}{2\,\sigma^4}$. В книге [16], с.283, показано, что лучшей оценки найти нельзя. Доказательство основано на понятии достаточной статистики для параметра, т.е. такой статистики, которая содержит столько же информации о параметре, сколько и сами наблюдения. Для нормального распределения достаточной статистикой для пары $(\theta\ ,\ \sigma^2\,)$ является пара $(\bar{x}\,,\ s_1)$.

Пусть другая несмещенная оценка σ^2 отличается от s_1^2 слагаемым $g\,(\bar{x}\,,\ s_1)$, тогда

$$\int g\,(\bar{x}\,,\ s_1)\ \exp\{\,-\,[\,n\,(\bar{x}\ -\ \theta\,)^2 + (n-1)\,s_1^2\,]/\,2\sigma^2\}\,dv = 0.$$

Дифференцируя дважды по θ, получим

$$\mathbf{E}\,[(\bar{X} - \theta\,)^2\,g\,(\bar{X}\,,\ s_1\,)\,] = 0\,,$$

а дифференцируя по σ^2, будем иметь соотношение

$$\mathbf{E}\,[\,n\ (\bar{X} - \theta\,)^2\,g\,(\bar{X}\,,\ s_1\,) + (n-1)\,s_1^2\,g\,(\bar{X}\,,\ s_1^2\,)\,] = 0\,,$$

Отсюда $cov\,[\,s_1^2\,,\ g\,(\bar{X}\,,\ s_1^2\,)\,] = 0$, что дает

$$\mathbf{D}\,[\,s_1^2 + g\,(\bar{X}\,,\ s_1^2\,)\,] = \mathbf{D}\,(\,s_1^2) + \mathbf{D}\,[g\,(\bar{X}\,,\ s_1^2\,)\,].$$

Из последнего соотношения следует $\mathbf{D}\,(\,s_1^2) \le \mathbf{D}\,[\,s_1^2 + g\,(\bar{X}\,,\ s_1^2\,)\,]$ для любой несмещенной оценки нуля $g\,(\bar{x}\,,\ s_1^2)$, т.е. s_1^2 – несмещенная оценка с минимальной дисперсией параметра σ^2.

С другой стороны, если параметр θ нормального распределения известен, то несмещенная оценка $m_2 = \frac{1}{n}\sum_{i=1}^{n}(\, x_i - \theta\,)^2$ имеет дисперсию, в точности равную σ^2/n, т.е. в этом случае достигается равенство в информационном неравенстве.

Неравенство Крамера-Рао позволяет показать, что существование состоятельных оценок связано с неограниченным в некотором смысле притоком фишеровской информации. Рассмотрим случай, когда параметрическое множество Θ есть ограниченный интервал $(\, a\,,\ b\,)$ вещественной прямой. Последовательность состоятельных оценок $\{\, T_n\,\}$ для θ в Θ может существовать лишь, если для любого $[\, \alpha\,,\ \beta\,] \subset (\, a\,,\ b\,)$

$$\inf_{[\,\alpha\,,\,\beta\,]} I_n(\theta\,) \underset{n}{\to} \infty,$$

где $I_n(\theta\,)$ – количество информации в n наблюдениях.

Действительно, пусть для всех $\theta \in \Theta$ последовательность T_n сходится при $n \to \infty$ к θ по вероятности. Так как множество Θ ограничено, можно считать, что и $|\, T_n\,| \le C$. Поэтому $\mathbf{E}_\theta((T_n - \theta)^2) \to 0$ для всех θ. Допустим, что для какого-нибудь интервала $[\, \alpha\,,\ \beta\,]$ $I_n(\theta\,) < M$ для всех $\theta \in [\, \alpha\,,\ \beta\,]$. Из неравенству Крамера-Рао следует, что

$$\frac{(\,1 + d_n(\theta\,))^2}{M} + d_n^2(\theta\,) \to\ 0\,,\ \theta\, \in\, [\,\alpha\,, \beta\,].$$

Здесь $d_n(\theta\,)$ – смещение оценки T_n. Тогда $d_n(\theta\,) \to 0$, $\theta\, \in \Theta$, $d_n'(\theta\,) \to -1$ для всех $\theta \in [\, \alpha\,,\ \beta\,]$. Но

$$|d_n'(\theta\,)\,| \ = \left| \frac{\partial}{\partial\theta}\, \mathbf{E}_\theta T_n - \theta\, \right| = \left|\, \int T_n\, \frac{\partial \ln\ f(\,x\,;\theta\,)}{\partial\theta}\, f(\,x\,;\theta\,)\, dv - 1\, \right| \le$$

$$\le\ \mathbf{E}_\theta^{1/2} T_n^2 \cdot I^{1/2}\,(\theta\,) + 1 \le \sqrt{M}\, \cdot C^{1/2} + 1.$$

По теореме Лебега

$$\lim_n\ \int_\alpha^\tau d_n'(\,u\,)\, du = \int_\alpha^\tau \lim_n\ d_n'(u)\, du = -\,(\tau\, -\, \alpha\,)\,,\ \alpha \le \tau\, \le \beta.$$

Но тогда

$$d_n(\tau\,) = \int_\alpha^\tau d_n'(u)\ du + d_n(\alpha\,) \to\ -(\tau\, - \alpha\,) \ne 0.$$

Полученное противоречие доказывает наше утверждение.

Приведем и докажем еще одно неравенство – неравенство Бхаттачариа, частным случаем которого является неравенство Крамера-Рао.

Предположим в регулярном случае, что (мы ограничимся в записи непрерывным случаем):

(5) $\int |\, (\partial^i/\partial\theta^i) f(x;\theta)\,|\, dx < \infty$ для $i = 1\,,\ 2\,,\ \ldots\,, k$ и всех θ.

(6) $J_{ij} = \int (\partial^i/\partial\theta^i) \ln f(x;\theta) \cdot (\partial^j/\partial\theta^j) \ln f(x;\theta) \cdot f(x;\theta)\ dx < \infty$ для $i, j = 1\,,\ 2\,,\ \ldots\,, k$ и всех θ.

Предположим, что функция $g(\theta)$, допускающая оценку, k раз дифференцируема. Обозначим через $\hat{g}(\theta)$ несмещенную оценку функции $g(\theta)$, которая имеет конечную дисперсию и удовлетворяет следующему условию:

(7) $\int |\, \hat{g}(x) \cdot (\partial^i/\partial\theta^i) f(x;\theta)\,| / f(x;\theta)\, dx < \infty$ для $i = 1\,,\ 2\,,\ \ldots\,, k$ и всех θ.

Далее, пусть $\mathbf{J} = (J_{ij})$ – неотрицательно определенная матрица размера $k \times k$ с элементами J_{ij} и $\xi^T = (g'(\theta), g''(\theta), \ldots, g^{(k)}(\theta))$, и если $\mathbf{J}$ невырождена в точке θ, то (неравенство Бхаттачариа)

$$\mathbf{D}_\theta(\hat{g}(X)) \ge \xi^T \mathbf{J}^{-1} \xi.$$

Д о к а з а т е л ь с т в о. Для сокращения записи возьмем случай $k = 2$ и пусть $S_i(x) = (\partial^i/\partial\theta^i)\ln f(x;\theta)$. Рассмотрим ковариационную матрицу Σ случайного вектора $(\hat{g}(X), S_1(X), S_2(X))^T$. Покажем, что Σ – симметричная матрица, имеющая вид

$$\Sigma = \begin{pmatrix} \mathbf{D}_\theta(\hat{g}(X)) & g^{(1)}(\theta) & g^{(2)}(\theta) \\ g^{(1)}(\theta) & J_{11} & J_{12} \\ g^{(2)}(\theta) & J_{12} & J_{22} \end{pmatrix}.$$

В самом деле,

$$\mathbf{Cov}(\hat{g}(X), S_i(X)) = \int \hat{g}(x) f(x;\theta)\,dx = \frac{\partial^i}{\partial\theta^i}\int \hat{g}(x) f(x;\theta)\,dx = \ g^{(i)}(\theta).$$

Далее, так как Σ неотрицательно определена, то

$$\det(\Sigma) = \det(\mathbf{J})(\mathbf{D}(\hat{g}(\theta) - (g^{(1)}(\theta), g^{(2)}(\theta))\mathbf{J}^{-1}(g^{(1)}(\theta), g^{(2)}(\theta))^T) \geq 0.$$

Более того, поскольку $\mathbf{J}$ – положительно определенная матрица в точке θ, то $\det(\mathbf{J}) > 0$. Отсюда получаем неравенство Бхаттачариа.

Вернемся к примеру, когда $X_i \in N(\theta, \sigma^2)$, $i = 1, 2, ..., n$. Здесь $\hat{g}(x) = \bar{x}^2 - \dfrac{\sigma^2}{n}$, $g(\theta) = \theta^2$, $g'(\theta) = 2\theta$, $g''(\theta) = 2$.

Тогда

$$\frac{f'}{f} = \frac{d\ln f}{d\theta} = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \theta), f' = f \cdot \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \theta), f'' = f' \cdot \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \theta) - f \cdot \frac{n}{\sigma^2},$$

$$\frac{f''}{f} = \frac{f'}{f} \cdot \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \theta) - \frac{n}{\sigma^2} = \frac{1}{\sigma^4}\left(\sum_{i=1}^n (X_i - \theta)\right)^2 - \frac{n}{\sigma^2}.$$

Но $\sum_{i=1}^n (X_i - \theta) \in N(0, n\sigma^2)$, поэтому

$$J_{11} = \mathbf{E}\left(\frac{f'}{f}\right)^2 = \frac{n}{\sigma^2}, J_{22} = \mathbf{E}\left(\frac{f''}{f}\right)^2 = \frac{2n^2}{\sigma^4}, J_{12} = 0.$$

Значит,

$$\mathbf{J} = \begin{pmatrix} \dfrac{n}{\sigma^2} & 0 \\ 0 & \dfrac{2n^2}{\sigma^4} \end{pmatrix} \quad \Rightarrow \quad \mathbf{J}^{-1} = \begin{pmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{\sigma^4}{2n^2} \end{pmatrix}.$$

Тогда

$$\xi^T \mathbf{J}^{-1}\xi = (2\theta, 2)\mathbf{J}^{-1}\begin{pmatrix} 2\theta \\ 2 \end{pmatrix} = \left(\frac{2\theta\sigma^2}{n}, \frac{\sigma^4}{n^2}\right)\begin{pmatrix} 2\theta \\ 2 \end{pmatrix} = \frac{4\theta^2\sigma^2}{n} + \frac{2\sigma^4}{n^2}.$$

Отсюда следует, что оценка $\hat{g}(x) = \bar{x}^2 - \frac{\sigma^2}{n}$ является оптимальной, однако радость от этого известия немного тускнеет от замечания, что если рассматривать достаточно большие n и поскольку во второе слагаемое входит множитель $\dfrac{1}{n^2}$ то эта разность будет незначительна.

Приведем также пример распределения и несмещенной оценки параметра этого распределения, когда дисперсия оценки будет меньше, чем нижняя граница неравенства Крамера-Рао. Ясно, что это будет в том случае, когда будут нарушены условия применимости этого неравенства.

Пример 19. Пусть x_1, x_2, ... , x_n – повторная выборка с общей функцией распределения $F(x\,;\theta\,) = \dfrac{x}{\theta}$, если $0 < x < \theta$. Для нее функция правдоподобия равна $L_x(\theta\,) = \dfrac{1}{\theta^n}$ для всех $0 < x_i < \theta$, $i = 1, 2, \ldots, n$. Последнее неравенство будет выполнено тогда и только тогда, когда $0 < x_n^{(n)} < \theta$, поэтому оценкой максимального правдоподобия будет $\hat{\theta}_n = x_n^{(n)}$. Найдем распределение величины $X_n^{(n)}$. Имеем:

$F_{X_n^{(n)}}(x) = \mathbf{P}\,(X_n^{(n)} < x) = \dfrac{x^n}{\theta^{\,n}}$, а плотность распределения равна

$$f_n\,(x) = n\,\frac{x^{n-1}}{\theta^{\,n}}\,,\ \ 0 < x < \theta.$$

Тогда

$$\mathbf{E}_\theta(X_n^{(n)}) = \int_0^\theta x \cdot f_n(x)\,dx = \int_0^\theta x \cdot \frac{n\,x^{n-1}}{\theta^{\,n}}\,dx = \frac{n}{\theta^{\,n}} \cdot \frac{\theta^{\,n+1}}{n+1} = \frac{n}{n+1}\,\theta.$$

Отсюда несмещенной оценкой параметра θ будет

$$T_n = \frac{n+1}{n}\,X_n^{(n)}.$$

Найдем дисперсию оценки T_n. Имеем:

$$\mathbf{D}_\theta(T_n) = \mathbf{E}_\theta(T_n^2) - \theta^{\,2},$$

$$\mathbf{E}_\theta(T_n^2) = \left(\frac{n+1}{n}\right)^2 \mathbf{E}_\theta((X_n^{(n)})^2) = \frac{(n+1)^2}{n^2}\int_0^\theta x^2 \cdot \frac{n\,x^{n-1}}{\theta^{\,n}}\,dx =$$

$$= \frac{(n+1)^2}{n\,\theta^{\,n}} \cdot \frac{\theta^{\,n+2}}{n+2} = \frac{(n+1)^2\,\theta^{\,2}}{n\,(\,n+2\,)},$$

$$\mathbf{D}_\theta(T_n) = \theta^2\,\left(\frac{(n+1)^2}{n\,(n+2)} - 1\right) = \frac{\theta^{\,2}}{n\,(n+2)} < \frac{\theta^{\,2}}{n^2}\,,$$

что меньше, чем в регулярном случае (там граница имеет вид $\dfrac{c}{n}$). Здесь плотность нерегулярна (разрывна) на краях распределения, поэтому скорость сходимости дисперсии оценки T_n выше. Таким образом, мы превзошли границу, даваемую неравенством Крамера-Рао за счет нарушения регулярности.

Определение 6. *Эффективностью* $\mathbf{e} = \mathbf{e}\,(T_n\,,\ \hat{\theta}_n\,)$ *оценки* T_n *по отношению к оценке* $\hat{\theta}_n$ *назовем дробь*

$$\mathbf{e}\,(T_n,\,\hat{\theta}_n) = \frac{\mathbf{D}_\theta(\hat{\theta}_n\,)}{\mathbf{D}_\theta(T_n\,)}.$$

Пусть теперь $\hat{\theta}_n$ – эффективная оценка, а T_n – какая-нибудь регулярная несмещенная оценка с эффективностью $\mathbf{e}$. Покажем, что коэффициент корреляции между $\hat{\theta}_n$ и T_n равен $\rho = \rho\,(\hat{\theta}_n,\,T_n) = \sqrt{\mathbf{e}}$.

Действительно, регулярная несмещенная оценка $\theta_n^* = (\,1-k)\,\hat{\theta}_n + k\,T_n$ имеет дисперсию

$$\mathbf{D}_\theta(\theta_n^*\,) = \left((\,1-k)^2 + \frac{2\rho\,k\,(\,1-k\,)}{\sqrt{\mathbf{e}}} + \frac{k^2}{\mathbf{e}}\right)\,\mathbf{D}_\theta(\hat{\theta}_n\,)\ =$$

$$= \left(\ 1 + 2k\,\frac{\rho\ - \sqrt{\mathbf{e}}}{\sqrt{\mathbf{e}}} + k^2\,\frac{\mathbf{e} - 2\rho\,\sqrt{\mathbf{e}} + 1}{\mathbf{e}}\right)\,\mathbf{D}_\theta(\hat{\theta}_n\,)$$

и если $\rho \neq \sqrt{\mathbf{e}}$, то коэффициент при $\mathbf{D}_\theta(\hat{\theta}_n\,)$ можно сделать меньшим 1, придавая числу k достаточно малое положительное или отрицательное значение. Но тогда получим $\mathbf{D}_\theta(\theta_n^*\,) < \mathbf{D}_\theta(\hat{\theta}_n\,)$, и θ_n^* будет иметь эффективность больше 1, что невозможно.

В частности, при $\mathbf{e} = 1$ мы имеем $\rho\ = 1$ и можно показать, что две эффективные оценки одного и того же параметра «почти всегда» равны друг другу.

Пример 20. Пусть $X_1, \ldots, X_n$ – выборка из нормального распределения $N(a, \sigma^2)$ и $\hat{a} = X_n^{(k)}$, где $\sqrt{n}\left(\frac{k}{n} - 0.5\right) \underset{n\to\infty}{\to} 0$, есть оценка параметра a. Известно, что при $n \to \infty$, $\sqrt{n}\,(X_n^{(k)} - a) \to \xi \in N\left(0, \frac{1}{4\,f^2(x_{0.5})}\right)$. Учитывая, что $x_{0.5} = a$ – медиана, то $f(a) = \frac{1}{\sqrt{2\pi}\sigma}$, то эффективность выборочной медианы по сравнению с оценкой $\bar{x}$ равна $\frac{2}{\pi} = 0.637$.

Пример 21. Найдем информационное количество Фишера параметра σ^2 для нормального распределения $N(a, \sigma^2)$. Имеем последовательно

$$f(x;\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right) \;\Rightarrow\; \ln f(x;\sigma = -\ln\sqrt{2\pi} - \ln\sigma - \frac{(x-a)^2}{2\sigma^2}\,,$$

$$\begin{aligned}
&\frac{\partial \ln f(x;\sigma)}{\partial\sigma} = -\frac{1}{\sigma} + \frac{(x-a)^2}{\sigma^3} \;\Rightarrow \\
&\mathbf{E}\left(\frac{\partial \ln f(x;\sigma)}{\partial\sigma}\right)^2 = \mathbf{E}\left(\frac{(X-a)^4}{\sigma^6}\right) - \frac{2}{\sigma}\mathbf{E}\left(\frac{(X-a)^2}{\sigma^3}\right) + \frac{1}{\sigma^2} = \\
&= \frac{3\sigma^4}{\sigma^6} - \frac{2\sigma^2}{\sigma^4} + \frac{1}{\sigma^2} = \frac{2}{\sigma^2}\,.
\end{aligned}$$

Рассмотрим оценку

$$s' = \sqrt{\frac{n}{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}s\,,\quad s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \;\Rightarrow\; \mathbf{E}\,(s') = \sigma\,.$$

Имеем: $\mathbf{D}\,(s') = \left(\frac{n-1}{2}\frac{\Gamma^2\left(\frac{n-1}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)} - 1\right)\sigma^2 \approx \frac{\sigma^2}{2n} + O\left(\frac{1}{n^2}\right)$.

Так что если сравнивать эту дисперсию с границей, даваемой неравенством Крамера-Рао, то $\mathbf{e}(s') \to 1$ при $n \to \infty$. Однако при $n = 2$, $\mathbf{e}(s') = \frac{1}{2(\pi - 2)} = 0.4380$, а при $n = 3$, $\mathbf{e}(s') = \frac{\pi}{6(4-\pi)} = 0.61$.

Аналогично для оценки

$$s_0' = \sqrt{\frac{n}{2}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}s_0\,,\quad s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 \;\Rightarrow\; \mathbf{E}\,(s_0') = \sigma\,,$$

$$\mathbf{D}\,(s_0') = \left(\frac{n}{2}\frac{\Gamma^2\left(\frac{n}{2}\right)}{\Gamma^2\left(\frac{n+1}{2}\right)} - 1\right)\sigma^2 \approx \frac{\sigma^2}{2n} + O\left(\frac{1}{n^2}\right).$$

и $\mathbf{e}(s_0') \to 1$ при $n \to \infty$. Однако при $n = 2$, $\mathrm{eff}(s_0') = \frac{\pi}{4(4-\pi)} = 0.9151$, а при $n = 3$, $\mathbf{e}(s_0') = \frac{4}{3(3\pi - 8)} = 0.9358$.

Для *среднего абсолютного отклонения* $s_2 = \frac{1}{n}\sum_{i=1}^{n} \mid x_i - \theta \mid$ получаем после выкладок

$$\mathbf{E}\left(\sqrt{\frac{\pi}{2}}\,s_2\right) = \sigma\,,\quad \mathbf{D}\left(\sqrt{\frac{\pi}{2}}\,s_2\right) = \frac{(\pi-2)\sigma^2}{2\,n},$$

так что $\sqrt{\frac{\pi}{2}}\,s_2$ является несмещенной оценкой для σ с эффективностью $\frac{1}{\pi - 2} = 0.876$.

Для распределения Коши с плотностью $f(x;a) = \dfrac{1}{\pi\,(1+(x-a)^2)}$ информационное количество Фишера равно

$$\mathbf{E}\left(\frac{\partial \ln f}{\partial a}\right)^2 = \frac{4}{\pi}\int\limits_{-\infty}^{\infty} \frac{(x-a)^2 dx}{[1+(x-a)^2]^3} = \frac{1}{2}, \quad \text{поэтому} \quad D\,(T_n) \geq \frac{2}{n}.$$

Если оценивать параметр a с помощью $\bar{x}$, то она имеет то же распределение Коши и поэтому не является состоятельной оценкой, а для $\hat{a} = x_n^{(k)}$, $k = \left[\dfrac{n}{2}\right]$,
$\sqrt{n}\,(\hat{a}-a) \xrightarrow{d} N(0, \dfrac{\pi^2}{4})$ при $n \to \infty$, поэтому $\mathbf{e}(\hat{a}) = \dfrac{2}{n} : \dfrac{\pi^2}{4n} = \dfrac{8}{\pi^2} = 0.8106$.

Покажем также, что если две несмещенных оценки имеют минимальную дисперсию, то они совпадают. Для этого рассмотрим понятие несмещенной оценки нуля.

Определение 7. Статистика U называется *несмещенной оценкой нуля*, если

$$\mathbf{E}_\theta(U) = 0 \quad \text{для любого} \theta.$$

Конечно, $U \equiv 0$ тривиально будет несмещенной оценкой нуля, но нас будет интересовать класс $\mathcal{U}$ всех несмещенных оценок нуля.

ТЕОРЕМА 2. *Пусть $\mathcal{U} = \{U\}$ – множество всех несмещенных оценок нуля, а T – несмещенная оценка параметра θ и $\mathbf{E}_\theta(T^2) < \infty$. Необходимым и достаточным условием для того, чтобы T имела равномерно минимальную дисперсию является то, что*

$$\mathbf{E}_\theta(TU) = 0 \quad \text{для любой} \quad U \in \mathcal{U} \quad \text{и любого} \quad \theta.$$

Д о к а з а т е л ь с т в о. 1) Предположим, что T имеет минимальную дисперсию, а $U \in \mathcal{U}$. Тогда для любого $c \in R$ оценка $T_c = T + c\,U$ тоже будет несмещенной оценкой, причем $\mathbf{D}(T_c) \geq \mathbf{D}(T) \quad \Leftrightarrow \quad c^2\mathbf{D}(U) + 2c\,\mathbf{Cov}(T,U) \geq 0$, откуда следует, что $\mathbf{Cov}(T,U) = 0$. Действительно, если $\mathbf{D}(U) \neq 0$, то минимальное значение $c^2\mathbf{D}(U) + 2c\,\mathbf{Cov}(T,U)$ равно $-\dfrac{\mathbf{Cov}^2(T,U)}{\mathbf{D}(U)} \geq 0$ только для $\mathbf{Cov}(T,U) = 0$. Отсюда

$$\mathbf{Cov}(T,U) = \mathbf{E}_\theta((T-\theta)U) = \mathbf{E}_\theta(TU) = 0.$$

2) Предположим теперь, что $\mathbf{E}_\theta(TU) = 0$ для любой $U \in \mathcal{U}$ и любого θ. Пусть T_0 – другая несмещенная оценка параметра θ с $\mathbf{D}(T_0) < \infty$. Тогда $T - T_0 \in \mathcal{U}$ и, следовательно, $\mathbf{E}_\theta(T(T-T_0)) = 0 \quad \Leftrightarrow \quad \mathbf{E}_\theta(T^2) = \mathbf{E}_\theta(TT_0) \;\Leftrightarrow\; \mathbf{D}(T) = \mathbf{Cov}(T,T_0)$ для любого θ, поскольку $\mathbf{E}_\theta(T) = \mathbf{E}_\theta(T_0) = \theta$. Из неравенства $\mathbf{Cov}^2(T,T_0) \leq \mathbf{D}(T)\mathbf{D}(T_0) \quad \Leftrightarrow \quad \mathbf{D}^2(T) \leq \mathbf{D}(T)\mathbf{D}(T_0)$ следует, что $\mathbf{D}(T) \leq \mathbf{D}(T_0)$, т.е. T имеет минимальную дисперсию.

2.3. Состоятельность.

Определение 8. *Оценка T_n называется <u>состоятельной</u> оценкой для θ, если она сходится по $\mathbf{P}_\theta$–вероятности к «истинному» значению параметра θ, т.е.*
$T_n \underset{n\to\infty}{\xrightarrow{\mathbf{P}_\theta}} \theta$*, или* $\mathbf{P}_\theta\,(\,|\,T_n - \theta\,|\; > \varepsilon\,) \underset{n\to\infty}{\to} 0, \;\forall\,\varepsilon > 0.$

Пример 22. Пусть $X_1\,,\; X_2\,,\; \dots\,,\; X_n$ – независимы и одинаково распределены с математическим ожиданием $\mathbf{E}_\theta(X_1) = \theta$. По теореме Хинчина $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ – состоятельная оценка для θ.

Определение 9. *Оценка T_n называется* <u>*состоятельной*</u> *оценкой для $g(\theta)$, если она сходится по $\mathbf{P}_\theta-$вероятности к $g(\theta)$, т.е.*

$$T_n \underset{n\to\infty}{\overset{\mathbf{P}_\theta}{\to}} g(\theta).$$

ТЕОРЕМА 3. *Если T_n – состоятельная оценка $g(\theta)$, а $f(x)$ – непрерывная функция, то $f(T_n)$ – состоятельная оценка для $f(g(\theta))$.*

Д о к а з а т е л ь с т в о. Из определения непрерывности функции $f(x)$ следует, что для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что

$$(\,|\,f(T_n) - f(g(\theta))\,| < \varepsilon\,) \supset (\,|\,T_n - g(\theta)\,| < \delta\,)\,.$$

Отсюда

$$\mathbf{P}_\theta\,(\,|\,f(T_n) - f(g(\theta))\,| < \varepsilon) \geq \mathbf{P}_\theta\,(\,|\,T_n - g(\theta)\,| < \delta\,) \underset{n\to\infty}{\to} 1.$$

Значит, $\mathbf{P}_\theta\,(\,|\,f(T_n) - f(g(\theta))\,| < \varepsilon\,) \underset{n\to\infty}{\to} 1.$

Пример 23. Пусть $X_1,\, X_2,\, \dots,\, X_n$ – независимы и одинаково распределены с математическим ожиданием $\mathbf{E}_\theta(X_1) = \theta$ и дисперсией $\mathbf{D}_\theta(X_1) = \sigma^2$. Рассмотрим $S_n^2 = \dfrac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2$. По теореме Хинчина $\frac{1}{n}\sum_{i=1}^n X_i^2$ сходится по вероятности к $\mathbf{E}_\theta(X_1^2) = \sigma^2 + \theta^2$. Значит, S_n^2 сходится по вероятности к σ^2, по предыдущей теореме состоятельной оценкой для среднеквадратичного уклонения σ является $S_n = \sqrt{S_n^2}$, так как $\sigma = \sqrt{\sigma^2}$ – непрерывная функция.

ТЕОРЕМА 4. (Достаточное условие состоятельности). *Пусть T_n – несмещенная оценка параметра θ и $\mathbf{D}_\theta(T_n) \underset{n\to\infty}{\to} 0$. Тогда T_n – состоятельная оценка θ.*

Д о к а з а т е л ь с т в о следует из неравенства Чебышева.

Пример 24. Пусть $X_1,\, X_2,\, \dots,\, X_n$ – независимы и одинаково распределены $N\,(\theta,\, \sigma^2)$. Тогда $S_1^2 = \dfrac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ имеет $\chi^2_{n-1}-$распределение χ^2 с $(n-1)$ степенями свободы и $\mathbf{D}_\theta(S_1^2) = \dfrac{2\,\sigma^4}{n-1} \underset{n\to\infty}{\to} 0$, поэтому S_1^2 – состоятельная оценка σ^2.

ТЕОРЕМА 5. *Если*

$$\sqrt{n}\,\left(T_n - \theta\right) \underset{n\to\infty}{\overset{d}{\to}} N(\,0,\, \tau^2\,)\,,$$

то

$$\sqrt{n}\,(f(T_n) - f(\theta)\,) \underset{n\to\infty}{\overset{d}{\to}} N(\,0\,,\; \tau^2\,(f'(\theta\,))^2)\,,$$

при условии, что $f'(\theta\,)$ существует, непрерывна и не равна нулю.

Доказательство теоремы 4 использует следующую лемму.

Лемма 2. *Если $Y_n \overset{d}{\to} Y$ и A_n, B_n сходятся по вероятности к a и b, то $A_n + B_nY_n \overset{d}{\to} a + bY$.*

Д о к а т е л ь с т в о теоремы 4. Разложим $f(T_n)$ по формуле Тейлора в точке θ:

$$f(T_n) = f(\theta) + (\,T_n - \theta\,)\,(\,f'(\theta\,) + R_n),$$

где $R_n \to 0$ при $T_n \to \theta$. Из условия теоремы следует, что $T_n \to \theta$ по вероятности и, следовательно, $R_n \to 0$ по вероятности. Требуемый результат следует теперь из леммы 2, примененной к $\sqrt{n}\,(f(T_n) - f(\theta))$.

Замечание. Когда $f'(\theta)=0$, а $f''(\theta)\neq 0$, естественно взять еще один член разложения Тейлора:

$$f(T_n)=f(\theta)+(T_n-\theta)f'(\theta)+\frac{1}{2}(\,T_n-\theta\,)^2\,(f''(\theta)+R_n),$$

где $R_n\to 0$ по вероятности при $T_n\to\theta$, или, поскольку $f'(\theta\,)=0$,

$$2n(f\,(T_n)-f\,(\theta))=(\sqrt{n}\,(\,T_n-\theta))^2\,(\,f''(\theta)+R_n),$$

откуда

$$n(f(T_n)-f(\theta))\xrightarrow{d}\frac{1}{4}\tau^2(f''(\theta))^2\chi_1^2.$$

Пример 25. Пусть случайная величина X имеет распределение Вейбулла с параметром формы c и сдвигом θ. Используя формулу для дисперсии величины X при большом значении c, предложим следующую оценку для c:
$\hat{c}^{-1}=\dfrac{\sqrt{6}}{\pi}\sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(\ln\,x_i-\overline{\ln\,x})}$, которая является состоятельной оценкой, правда ее эффективность относительно границы Крамера-Рао равна 55%.

Введем понятие *преобразования, стабилизирующего дисперсию*. Пусть распределение величин Z_n, $n=1,\,2\,\ldots$, зависит от параметра θ, $\{a_n\}$ – неограниченная числовая последовательность, не зависящая от θ, при $n\to\infty$ последовательность $a_n(Z_n-\theta)$ сходится по распределению к $\zeta\in N(0,\sigma^2(\theta))$. Тогда

$$a_n(g(Z_n)-g(\theta))g'(\theta)\zeta$$

для любой непрерывно дифференцируемой функции $g(\theta)$.

Пусть функция $g(\theta)$ такова, что

$$g'(\theta)\sigma^2(\theta)=c,$$

c – постоянная, не зависящая от θ. В таком случае преобразование $g(Z_n)$ будем называть ***преобразованием, стабилизирующим дисперсию***.

Суть определения состоит в том, что если применить такое преобразование, то предельная дисперсия не будет зависеть от неизвестного параметра θ. При больших n это позволит построить доверительные интервалы длины, не зависящей от параметра.

Пример 26. Пусть $X_1,X_2,...,X_n$ – независимые и одинаково распределенные (распределение Бернулли $b(1,\theta)$) случайные величины, $\bar{X}=\sum_{i=1}^{n}X_i/n$ и $g(\theta)=\arcsin(\sqrt{\theta})$. Имеем:

$$\frac{\sigma^2(\theta)}{n}=\mathbf{D}(\bar{X})=\frac{\theta(1-\theta)}{n},\quad g'(\theta)=\frac{1}{2\sqrt{\theta(1-\theta)}},\quad \sigma^2(\theta)\cdot(g'(\theta))^2=\frac{1}{4},$$

поэтому преобразование $g(\theta)=\arcsin(\sqrt{\theta})$ есть преобразование, стабилизирующее дисперсию.

Глава 3.

Метод наименьших квадратов.

3.1. Оценки метода наименьших квадратов.

Рассмотрим вновь вопрос построения МП-оценок для среднего a по выборке из нормального распределения, когда дисперсия σ^2 известна. В этом случае плотность распределения равна

$$f = f(x;a) = \frac{1}{\sqrt{2\pi}}\ \exp\left\{-\frac{1}{2}\frac{(\,x-a\,)^2}{\sigma^2}\right\},$$

а функция правдоподобия равна

$$l_x(a) = \ln\ L_x(a) = \sum_{i=1}^{n}\frac{f_a'(x_i;a)}{f(x_i;a)} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\,x_i-a\,)^2.$$

Отсюда видно, что эта функция максимальна, когда минимальна сумма

$$\sum_{i=1}^{n}(\,x_i-a\,)^2.$$

Таким образом, принцип максимального правдоподобия сводится в данном случае к тому, чтобы выбирать оценку для параметра a так, чтобы выражение (1) обращалось в минимум.

Заметим, что функция $\dfrac{f_a'(x;a)}{f(x;a)} = -\dfrac{x-a}{2\sigma^2}$ – линейная функция x.

Предположим теперь, что n наблюдений получены не из одинаковых нормальных распределений, а из распределений с различными средними. Именно, пусть
$a_i = \sum_{j=1}^{k} b_{ij}\theta_j$, $i = 1,\,2\,,\,\ldots\,,\,n$; b_{ij} – известны, θ_j – неизвестны.

В этом случае принцип максимального правдоподобия приводит к тому, что надо минимизировать по θ_j сумму

$$\sum_{i=1}^{n}(\,x_i - \sum_{j=1}^{k} b_{ij}\theta_j\,)^2.$$

Откажемся теперь от предположения нормальности и рассмотрим некоррелированные наблюдения $(x_1,\,x_2,\,\ldots\,,\,x_n\,)^{\mathrm{T}}$ с условием

(3) $$\left\{\begin{array}{l}\mathbf{E}\,(X_i) = b_{i1}\theta_1 + b_{i2}\theta_2 +\ \ldots\ + b_{ik}\theta_k\,,\\ \mathbf{D}\,(X_i) = \sigma^2\,,\ \ i = 1\,,\,2\,,\,\ldots\,,\,n\,.\end{array}\right.$$

где $(\,\theta_1,\,\ldots\,,\,\theta_k)^{\mathrm{T}}$ – неизвестные параметры, а $\mathbf{B} = (b_{ij})_{n\times k}$ – матрица известных коэффициентов. В матричной форме (48) запишется в виде

(4) $$\left\{\begin{array}{l}\mathbf{E}\,(X) = \mathbf{B}\theta\,,\\ \mathbf{D}\,(X) = \sigma^2\mathbf{J}\,,\end{array}\right.$$

где θ – вектор-столбец, $\mathbf{J}$ – единичная матрица порядка $n\times n$, т.е.

$$\mathbf{X} = \mathbf{B}\theta + \varepsilon\ ,\ \ \mathbf{E}(\varepsilon) = \mathbf{0}\,,\ \mathbf{D}\,(\varepsilon) = \sigma^2\mathbf{J}.$$

Пример 27. Пусть $\mathbf{E}\,(X_i) = \theta_o + \theta_1 b_i +\ \ldots\ + \theta_{k-1}b_i^k$, $\mathbf{D}\,(X_i) = \sigma^2$,
$i = 1\,,\,2\,,\,\ldots\,,\,n$. Здесь $b_{ij} = b_i^{j-1}$, $j = 1\,,\,2\,,\,\ldots\,,\,k$.

Метод наименьших квадратов (в матричном виде) состоит в минимизации

$$Q = (\,\mathbf{X} - \mathbf{B}\theta\,)^{\mathrm{T}}(\,\mathbf{X} - \mathrm{B}\theta\,) = \sum_{i=1}^{n}(\,x_i - b_{i1}\theta_1 - \;\dots\; - b_{ik}\theta_k\,)^2 \tag{5}$$

Необходимым (а для квадратичной функции и достаточным) условием обращения (24) в минимум является условие

$$\frac{\partial Q}{\partial \theta_j} = 0\,,\; j = 1\,,\,2\,,\,\dots\,,\,k, \qquad (7)$$

или,

$$\left\{\; \sum_{i=1}^{n} b_{ij}\,(\,x_i - b_{i1}\theta_1 - \;\dots\; - b_{ik}\theta_k\,) = 0\,,\;\; j = 1\,,\,2\,,\,\dots\,,\;k\,. \right. \qquad (8)$$

В матричном виде система (8) принимает вид:

$\mathbf{B}^T(\,\mathbf{X} - \mathbf{B}\theta\,) = 0.$ (9)

Уравнение (9) относительно θ вида $\mathbf{B}^T\mathbf{B}\theta = \mathbf{B}^T\mathbf{X}$ называется нормальным уравнением. Его решение (при условии, что матрица $\mathbf{B}^T\mathbf{B}$ невырождена) есть

$\hat\theta_n = (\,\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}.$ (10)

Отметим, что нормальное уравнение имеет решение, если предполагается, что матрица $\mathbf{B}^T\mathbf{B}$ невырождена. Матрица $(\,\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$ является *псевдообратной* для прямоугольной матрицы $\mathbf{B}$.

Покажем напрямую, что оценки наименьших квадратов доставляют минимум выражению $(\,\mathbf{X} - \mathbf{B}\theta\,)^T(\,\mathbf{X} - \mathbf{B}\theta\,)$. Действительно,

$$(\mathbf{X} - \mathbf{B}\theta\,)^T(\mathbf{X} - \mathbf{B}\theta\,) = (\,\mathbf{X} - \mathbf{B}\theta + \mathbf{B}\,(\hat\theta - \theta)\,)^T(\,\mathbf{X} - \mathbf{B}\theta + \mathbf{B}\,(\hat\theta - \theta)\,) =$$

$$= (\,\mathbf{X} - \mathbf{B}\hat\theta\,)^T(\,\mathbf{X} - \mathbf{B}\hat\theta\,) + (\hat\theta - \theta)^T\mathbf{B}^T\mathbf{B}(\hat\theta - \theta\,) + (\,\mathbf{X} - \mathbf{B}\hat\theta\,)^T\mathbf{B}\,(\hat\theta - \theta\,) + (\hat\theta - \theta)^T\mathbf{B}^T(\mathbf{X} - \mathbf{B}\hat\theta).$$

Но

$$\mathbf{B}^T(\mathbf{X} - \mathbf{B}\hat\theta) = \mathbf{B}^T\mathbf{X} - \mathbf{B}^T\mathbf{B}\hat\theta = 0,$$

$$(\mathbf{X} - \mathbf{B}\hat\theta)^T\mathbf{B} = (\,\mathbf{B}^T(\mathbf{X} - \mathbf{B}\hat\theta))^T = 0,$$

откуда

$$(\,\mathbf{X} - \mathbf{B}\theta\,)^T(\,\mathbf{X} - \; bfB\theta\,) = (\,\mathbf{X} - \mathbf{B}\hat\theta\,)^T(\,\mathbf{X} - \mathbf{B}\hat\theta\,) + (\hat\theta - \theta)^T\mathbf{B}^T\mathbf{B}(\hat\theta - \theta\,) \geqslant$$

$$\geqslant\; (\,\mathbf{X} - \mathbf{B}\hat\theta\,)^T(\,\mathbf{X} - \mathbf{B}\hat\theta\,).$$

Пример 28. Вернемся к примеру 27. Здесь система нормальных уравнений имеет вид

$$\left\{\begin{array}{l}
\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n}(\;\theta_0 + \theta_1 b_i + \;\dots\; + \theta_k b_i^k\,)\,,\\
\sum\limits_{i=1}^{n} x_i b_i = \sum\limits_{i=1}^{n}\; b_i(\;\theta_0 + \theta_1 b_i + \;\dots\; + \theta_k b_i^k\,)\,,\\
.\quad\;.\quad\;.\quad\;.\quad\;.\quad\;.\quad\;.\quad\;.\\
\sum\limits_{i=1}^{n} x_i b_i^k = \sum\limits_{i=1}^{n}\; b_i^k(\;\theta_0 + \theta_1 b_i + \;\dots\; + \theta_k b_i^k\,)\,.
\end{array}\right.$$

Если $k = 1$, то система принимает вид

$$\left\{\begin{array}{l}
\bar x = \theta_0 + \theta_1 \bar b\,,\\
\dfrac{1}{n}\sum\limits_{i=1}^{n} x_i b_i = \theta_0 + \theta_1\,\dfrac{1}{n}\sum\limits_{i=1}^{n} b_i^2\,.
\end{array}\right.$$

Решение последней системы есть

$$\hat\theta_1 = \frac{S_{xb}}{S_b^2}\,, \quad \hat\theta_o = \bar x - \hat\theta_1 \bar b,$$

где $S_{xb} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i b_i - \bar x \cdot \bar b\,, \quad S_b^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} b_i^2 - \bar b^2.$

Пример 29. В табл. 1 приведены 16 наблюдений отклика y в зависимости от переменной x.

Таблица 1. *Объем продаж (y в единицах продукции) при различных затратах на рекламу (x в тыс. руб.)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x	7	19	22	63	70	80	105	126	145	155	165	170	190	250	270	320
y	18.2	7.0	5.2	33.6	28.5	58.8	60.7	65.4	61.1	57.8	82.7	89.3	47.5	76.9	108.6	92.7
$\hat x_t$	18.6	22.1	23	34.9	37.0	39.9	47.2	53.3	58.9	61.8	64.7	66.1	72.0	89.5	95.3	109.9

Уравнение прямой имеет вид

$$y = 0.292\,x + 16.534.$$

Коэффициент 0.302 при переменной x означает, что при увеличении затрат на рекламу на 1 тыс. руб., объем продаваемой продукции увеличивается на 0.302, т.е. если рекламу увеличить на 10 000 руб., объем увеличится в среднем на 3 единицы продукции.

$$\bar y = 55.875\,, \; S_{yy} = 891.267\,, \; \bar x = 134.8125\,, \; S_{xx} = 7959.277\,, \; S_{yx} = 2322.695,$$

$$\hat a_1 = \frac{S_{yx}}{S_{xx}} = 0.292\,, \; \hat a_0 = \bar y - \hat a_1 \bar x = 16.534$$

$$\alpha = 0.005\,, \; \nu = n - 2 = 16 - 2 = 14, \chi^2_{0.0025}(14) = 2.697\,, \; \chi^2_{0.9975}(14) = 38.109\,,$$

$$\gamma^2 = \frac{1}{S_{yy}}\sum_{i=1}^{16}(y_i - \hat y_i)^2 = \frac{n(S_{yy} - S_{yx}^2/S_{xx})}{S_{yy}} = n\left(1 - \frac{S_{yx}^2}{S_{xx}S_{yy}}\right) = \frac{3415.24}{S_{yy}} = 3.832,$$

$2.697 < 3.832 < 38.109$ – гипотеза $k = 1$ не отвергается.

Пример 30. Пусть θ – одномерный параметр и $\mathbf{X} = \mathbf{B}\theta + \varepsilon$, где $\mathbf{B}$ – вектор столбец размерности n, т.е.

$$\mathbf{B} = \begin{pmatrix} b_1 \\ b_2 \\ . \\ . \\ b_n \end{pmatrix}.$$

Оценка наименьших квадратов имеет вид

$$\hat\theta_n = (\,\mathbf{B}^T\mathbf{B}\,)^{-1}\mathbf{B}^T\mathbf{X} = \frac{\sum_{i=1}^{n} b_i x_i}{\sum_{i=1}^{n} b_i^2}.$$

Если $\mathbf{D}\,(X) = \Sigma = \mathbf{E}(\varepsilon \cdot \varepsilon^{\mathrm{T}})$ и Σ известна, то с помощью преобразования $\mathbf{Z} = \Sigma^{-1/2}\mathbf{X}$ получаем $\mathbf{E}(\mathbf{Z}) = \Sigma^{-1/2}\mathbf{B}\theta = \mathbf{U}\theta$, где $\mathbf{U} = \Sigma^{-1/2}\mathbf{B}$, и

$$\mathbf{D}\,(\mathbf{Z}) = \mathbf{E}((\mathbf{Z} - \Sigma^{-1/2}\mathbf{B}\theta)\cdot(\mathbf{Z} - \Sigma^{-1/2}\mathbf{B}\theta)^T) = \mathbf{E}((\,\Sigma^{-1/2}\varepsilon)\cdot(\Sigma^{-1/2}\varepsilon)^{\mathrm{T}}) =$$

$$= \Sigma^{-1/2}\,\mathbf{E}((\varepsilon\varepsilon^T)(\Sigma^{-1/2})^T) = \mathbf{J}.$$

Оценки наименьших квадратов здесь можно определить по формуле

$$\hat\theta_n = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{X}.$$

Пример 31. Пусть

$$\mathbf{E}\,(X)\ =\begin{pmatrix}\theta\\ \theta\\ .\\ .\\ \theta\end{pmatrix}=\theta^n\begin{pmatrix}1\\ 1\\ .\\ .\\ 1\end{pmatrix},\ \text{т.е. } \mathbf{B}=\begin{pmatrix}1\\ 1\\ .\\ .\\ 1\end{pmatrix},\ \mathbf{D}\,(X)=\begin{pmatrix}\sigma_1^2 & 0 & . & . & 0\\ . & \sigma_2^2 & . & . & 0\\ . & . & . & . & .\\ . & . & . & . & .\\ 0 & 0 & . & . & \sigma_n^2\end{pmatrix}=\Sigma,$$

т.е. величины $X_1,\ X_2,\ \dots,\ X_n$ – независимы, но имеют разные дисперсии (неравноточные измерения). Тогда

$$\Sigma^{-1/2}=(\Sigma^{-1/2})^{\mathrm{T}}=\begin{pmatrix}1/\sigma_1 & 0 & . & . & 0\\ . & 1/\sigma_2 & . & . & 0\\ . & . & . & . & .\\ . & . & . & . & .\\ 0 & 0 & . & . & 1/\sigma_n\end{pmatrix}.$$

Так как оценка наименьших квадратов равна

$$\hat{\theta}_n=(\mathbf{U}^{\mathrm{T}}\mathrm{U})^{-1}\mathrm{U}X \quad \text{и} \quad \mathbf{U}=\Sigma^{-1/2}\mathbf{B}=\begin{pmatrix}1/\sigma_1\\ 1/\sigma_2\\ .\\ 1/\sigma_n\end{pmatrix},$$

то

$$\hat{\theta}_n=\frac{\sum_{i=1}^n\frac{x_i}{\sigma_i}}{\sum_{j=1}^n\frac{1}{\sigma_i^2}}.$$

Пример 32. Пусть результаты измерения углов во всех комбинациях равны

$$\begin{cases}38.5=\theta_1 & +\,\varepsilon_1,\\ 46.1=\quad\theta_2 & +\,\varepsilon_2,\\ 17.7=\qquad\theta_3 & +\,\varepsilon_3,\\ 84.6=\theta_1+\theta_2 & +\,\varepsilon_4,\\ 63.8=\quad\theta_2+\theta_3 & +\,\varepsilon_5,\\ 102.3=\theta_1+\theta_2+\theta_3 & +\,\varepsilon_6,\end{cases}$$

где $\mathbf{E}(\varepsilon_i)=0,\ \mathbf{D}(\varepsilon_i)=\sigma^2,\ i=1,2,..,6.$

Тогда

$$\mathbf{B}:=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\\1&1&0\\0&1&1\\1&1&1\end{pmatrix},\ \mathbf{X}=\begin{pmatrix}36.5\\46.1\\17.75\\84.6\\63.85\\102.35\end{pmatrix},\ \mathbf{B}^T=\begin{pmatrix}1&0&0&1&0&1\\0&1&0&1&1&1\\0&0&1&0&1&1\end{pmatrix},\ \mathbf{B}^T\mathbf{B}=\begin{pmatrix}3&2&1\\2&4&2\\1&2&3\end{pmatrix},$$

$$(\mathbf{B}^T\mathbf{B})^{-1}=\frac{1}{4}\begin{pmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{pmatrix},\quad \mathbf{B}^T\mathbf{X}=\begin{pmatrix}223.4\\296.8\\183.8\end{pmatrix},\quad \hat{\theta}=(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}=\begin{pmatrix}37.5\\46.6\\17.7\end{pmatrix}.$$

Как и в случае любого другого систематического принципа оценивания, вопрос о возможности использования метода наименьших квадратов зависит от свойств оценок, к которым он приводит. Отметим некоторые свойства оценок метода НК.

3.2. Несмещенность оценок наименьших квадратов.

Представим оценку по методу НК в следующем виде:

$$\hat{\theta}_n = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T(\mathbf{B}\theta + \varepsilon) = \theta + (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon.$$

Так как матрица $\mathbf{B}$ неслучайна, то получаем

$$\mathbf{E}\,(\hat{\theta}_n) = \theta.$$

Найдем матрицу ковариаций оценок наименьших квадратов. Учитывая, что

$$\hat{\theta}_n - \theta = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon,$$

получаем

$$\mathbf{D}\,(\hat{\theta}_n) = \mathbf{E}\,((\hat{\theta}_n - \theta)(\hat{\theta}_n - \theta)^{\mathrm{T}}) = \mathbf{E}\,(((\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon)((\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\varepsilon)^T) =$$
$$= (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{E}\,(\varepsilon\varepsilon^T)\mathbf{B}\,(\mathbf{B}^T\mathbf{B})^{-1} = \sigma^2(\mathbf{B}^T\mathbf{B})^{-1}.$$

Пусть есть линейная оценка $t = \mathbf{A}X$. Чтобы она была несмещенной, должно быть

$$\mathbf{E}\,(t) = \mathbf{E}(\mathbf{A}\mathbf{X}) = \theta = \mathbf{J}\cdot\theta,$$

это дает($\mathbf{J}$ есть единичная матрица)

$$\mathbf{E}(\mathbf{A}\,(\mathbf{B}\theta + \varepsilon)) = \theta,$$

откуда

$$\mathbf{A}\cdot\mathbf{B} = \mathbf{J},$$

что является необходимым и достаточным условием того, чтобы оценка $\mathbf{A}\,\mathbf{X}$ была несмещенной оценкой θ.

Рассмотрим матрицу ковариаций. Так как
$t = \mathbf{A}\mathbf{X} = \mathbf{A}(\mathbf{B}\theta + \varepsilon) = \theta + \mathbf{A}\varepsilon$ и $t - \theta = \mathbf{A}\varepsilon$,
то

$$\mathbf{D}\,(t) - \mathbf{E}\,((t-\theta)(t-\theta)^T) = \mathbf{A}\cdot\mathbf{E}\,(\varepsilon\cdot\varepsilon^T)\cdot\mathbf{A}^T = \sigma^2\cdot\mathbf{A}\cdot\mathbf{A}^T.$$

Дисперсии оценок – диагональные элементы матрицы $\mathbf{A}\cdot\mathbf{A}^T$.
Прямой проверкой убеждаемся, что

$$\mathbf{A}\cdot\mathbf{A}^T = \mathbf{B}\cdot\mathbf{B}^T + (\mathbf{A} - (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)\cdot(\mathbf{A} - (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T)^T.$$

Следовательно, диагональные элементы минимальны, когда равны нулю элементы второго слагаемого, т.е. $\mathbf{A} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$.
В частности, для оценок наименьших квадратов

$$\mathbf{D}\,(\hat{\theta}_n) = \sigma^2\cdot(\mathbf{B}^T\mathbf{B})^{-1}.$$

Кроме того, можно получить, что несмещенная оценка для σ^2 дается формулой

$$S^2 = \frac{1}{n-k}\,(X - \mathbf{B}\cdot\hat{\theta})^T\,(\mathbf{X} - \mathbf{B}\cdot\hat{\theta}\,).$$

Пусть $\mathbf{E}\,(\mathbf{X}) = \mathbf{B}\theta\,,\quad \mathbf{D}\,(\mathbf{X}) = \Sigma$. Преобразование $\mathbf{Z} = \Sigma^{-1/2}\mathbf{B}\theta = \mathbf{U}\theta$, $\mathbf{U} = \Sigma^{-1/2}\mathbf{B}$, приводит к схеме, когда $\mathbf{D}\,(\mathbf{X}) = \mathbf{J}$. Если значение факторов (элементы матрицы $\mathbf{B}$) можно выбирать из неограниченной области, то величина $(\mathbf{B}^T\mathbf{B})^{-1}$ может быть сделана сколь у годно малой, так как если $\mathbf{B}$ заменить на $a\,\mathbf{B}$, то $(\mathbf{B}^{\mathrm{T}}\mathrm{B})^{-1}$ заменится на $a^{-2}(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1} \to 0$ при $a \to \infty$. Однако на практике принятая модель не применима, так как параметры нельзя менять в широкой области. Мы должны ограничить значения факторов небольшими областями, внутри которых и исследуется т.н. функция отклика.

3.3. Эффективность оценок наименьших квадратов.

Теперь мы наложим ограничения вида

$$\mathbf{B}_i^T\mathbf{B}_i = c_i^2, \qquad \left\{ \begin{array}{l} \mathbf{B}_i^T \ - \text{вектор-строка}, \\ \mathbf{B}_i \ - \text{вектор-столбец}. \end{array} \right. \tag{6}$$

где c_i^2 заданы, $i = 1\,,\ 2\,,\ \dots\,,\ m$, n уровней фактора $\mathbf{B}_i$.

ТЕОРЕМА 6 (линейный аналог неравенства Крамера-Рао). *Пусть* $\mathbf{B}$ *– матрица плана,* $\hat{\theta}_i$ *– оценки наименьших квадратов параметра* θ_i*. Тогда при условии* (6) *на* $\mathbf{B}$ *имеем неравенства*

$$\mathbf{D}\,(\hat{\theta}_i) \geq \frac{1}{c_i^2}, i = 1\,,\ 2\,,\ \dots\,,\ m,$$

и минимум достигается , когда $\mathbf{B}_i^T\mathbf{B}_j = 0$*,* $i \neq j$*.*

Д о к а з а т е л ь с т в о. Пусть $i = 1$. Матрицу $\mathbf{B}^T\mathbf{B}$ разобьем следующим образом

$$\mathbf{B}^T\mathbf{B} = \begin{pmatrix} \mathbf{B}_1^T\mathbf{B}_1 & \mathbf{G}^T \\ \mathbf{G} & \mathbf{F} \end{pmatrix},$$

где элементами матрицы $\mathbf{G}^T$ являются $\mathbf{B}_1^T\mathbf{B}_j$.

Заметим, что

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\,\mathbf{D}\,| \cdot \left|\,\mathbf{A} - \mathbf{B}\cdot\mathbf{D}^{-1}\cdot\mathbf{C}\,\right|,$$

где $\mathbf{A}$ и $\mathbf{D}$ – квадратные матрицы, $\mathbf{D}$ – невырожденная матрица.

Действительно,

$$\begin{vmatrix} 1 & -\mathbf{B}\cdot\mathbf{D}^{-1} \\ 0 & 1 \end{vmatrix} = 1,$$

поэтому

$$\begin{vmatrix} 1 & -\mathbf{B}\cdot\mathbf{D}^{-1} \\ 0 & 1 \end{vmatrix} \cdot \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = \begin{vmatrix} \mathbf{A} - \mathbf{B}\cdot\mathbf{D}^{-1}\mathbf{C} & 0 \\ \mathbf{B} & \mathbf{D} \end{vmatrix} = |\,\mathbf{D}\,| \cdot \left|\,\mathbf{A} - \mathbf{B}\cdot\mathbf{D}^{-1}\cdot\mathbf{C}\,\right|.$$

Отсюда следует, что

$$\left|\,\mathbf{B}^T\mathbf{B}\,\right| = |\,\mathbf{F}\,| \cdot (\,\mathbf{B}_1^T\mathbf{B}_1 - \mathbf{G}^T\mathbf{F}^{-1}\mathbf{G}\,),$$

поэтому угловой элемент обратной матрицы равен

$$\frac{|\,\mathbf{F}\,|}{|\,\mathbf{B}^T\mathbf{B}\,|} = \frac{1}{\mathbf{B}_1^T\mathbf{B}_1 - \mathbf{G}^T\mathbf{F}^{-1}\mathbf{G}} \ \geqslant\ \frac{1}{\mathbf{B}_1^T\mathbf{B}_1},$$

поскольку матрица $\mathbf{F}$, а, следовательно, и $\mathbf{F}^{-1}$ – положительно определенные.

Действительно, известно: 1) матрица $\mathbf{A}$ неотрицательно определенная тогда и только тогда, когда все главные миноры неотрицательны; 2) существует ортогональное преобразование, приводящее $\mathbf{A}$ к диагональному виду:

$$\mathbf{QAQ}^{-1} = \begin{pmatrix} d_1 & 0 & . & 0 \\ 0 & d_2 & . & . \\ . & . & . & 0 \\ 0 & . & 0 & d_m \end{pmatrix} = D$$

($\mathbf{A}$ –положительно определенная, если все $d_i > 0$). Тогда

$$D^{-1} = \begin{pmatrix} 1/d_1 & 0 & . & 0 \\ 0 & 1/d_2 & . & . \\ . & . & . & 0 \\ 0 & . & 0 & 1/d_m \end{pmatrix} = \mathbf{QA}^{-1}\mathbf{Q}^{-1},$$

т.е. $\mathbf{A}^{-1}$ – тоже положительно определенная матрица.

Значит, квадратичная форма $\mathbf{G}^T\mathbf{F}^{-1}\mathbf{G} \geqslant 0$. Но

$$\mathbf{D}\,(\hat{\theta}_1) = \frac{|\,\mathbf{F}\,|}{|\,\mathbf{B}^T\mathbf{B}\,|},$$

следовательно,

$$\mathbf{D}\,(\hat{\theta}_1) \geqslant \frac{1}{c_1^2}.$$

Равенство возможно тогда и только тогда, когда $\mathbf{G} = 0$, т.е. когда $\mathbf{B}_1^T\mathbf{B}_j = 0\,,\ j \neq 1$.

Пример 33. Пусть $X_i = \theta_o + \theta_1 b_i + \varepsilon_i\,,\ i = 1\,,\ 2\,,\ \dots\,,\ n,$
и число наблюдений n – четно, а область изменения величины b есть интервал $(\,a_1\,,\ a_2\,)$. Для того чтобы сделать дисперсии оценок наименьших квадратов наименьшими, надо половину наблюдений провести в точке $b = a_1$, а вторую половину в точке $b = a_2$.

3.4. Оценивание по методу наименьших квадратов при наличии ограничений.

Пусть

$$\begin{cases} \mathbf{E}\,(\mathbf{X}) = \mathbf{B}\theta\,, \\ \mathbf{D}\,(\mathbf{X}) = \sigma^2\mathbf{J}\,, \end{cases} \tag{12}$$

и предположим, кроме того, что параметры θ удовлетворяют совместной системе линейных связей:

$\mathbf{H}^T\theta = \xi$ при условии $rang\ \mathbf{H} = k - s.$ (13)

Известно, что общее решение системы (13) имеет вид:

$\tilde{\theta} = \theta_0 + \mathbf{A}\beta$, где θ_0 – частное решение (13), а $rang\ \mathbf{A} = s$, причем $\mathbf{H}^T\mathbf{A} = 0$, β – произвольный вектор из s элементов. Тогда

$$\mathbf{E}\,(\mathbf{X}) = \mathbf{B}\theta\ = \mathbf{B}\,(\theta_0 + \mathbf{A}\beta\,) = \mathbf{B}\,\theta_0 + \mathbf{B}\,\mathbf{A}\beta,$$

откуда

$$\mathbf{E}\,(\mathbf{X} - \mathbf{B}\,\theta_0\,) = \mathbf{B}\,\mathbf{A}\beta.$$

Мы имеем

$\begin{cases} \mathbf{E}\,(\mathbf{Z}) = \mathbf{B}\mathbf{A}\beta\,, \\ \mathbf{D}\,(\mathbf{Z}) = \sigma^2\mathbf{J}\,, \end{cases}$ где $\mathbf{Z} = \mathbf{X} - \mathbf{B}\theta_0$, и теперь можно действовать обычным МНК.

Можно также в этом случае идти следующим путем: минимизировать форму

$$(\,\mathbf{X} - \mathbf{B}\theta\,)^T(\,\mathbf{X} - \mathbf{B}\theta\,) - 2\,(\,\mathbf{H}^T\theta - \xi\,)^T\lambda,$$

т.е. с помощью неопределенных множителей Лагранжа λ – придем к тому же самому. Решение в обоих случаях имеет вид

$$\hat{\theta}_n = (\,\mathbf{B}^T\mathbf{B}\,)^{-1}\mathbf{B}^T\mathbf{X} + (\,\mathbf{B}^T\mathbf{B}\,)^{-1}\mathbf{H}\,(\mathbf{H}^T(\,\mathbf{B}^T\mathbf{B}\,\mathbf{B})^{-1}\mathbf{H})^{-1}(\xi - \mathbf{H}^T(\,\mathbf{B}^T\mathbf{B}\,\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}).$$

3.5. Проверка модели на адекватность.

Пусть имеем следующую полиномиальную модель

$X_i = \theta_0 + \theta_1 b_i + \ \dots\ + \theta_k b_i^k + \varepsilon_i$ и ε_i - погрешности измерений имеют нормальное распределение с нулевым средним и известной дисперсией σ^2, т.е. имеем гипотезу H_0 : $k = k_0$ (о степени полинома).

При нулевой степени полинома ($k = 0$)в качестве оценки МНК мы получим среднее арифметическое выборочных данных $\bar{x}$ (самое плохое совпадение), а при $k = n$ - объему выборочных данных, получится слишком хорошее совпадение регрессионной модели с выборочными данными (оценка пройдет точно через выборочные данные). Такое (слишком хорошее) совпадение неправдоподобно. Поэтому необходимо выбрать компромиссный вариант. Итак пусть

$$\phi(b;\theta) = \theta_0 + \theta_1 b + \ \dots \ + \theta_{k_0} b^{k_0} \ -$$

регрессионная зависимость и $\hat{\theta} = (\hat{\theta}_0\,,\ \hat{\theta}_1\,,\ \dots\,,\ \hat{\theta}_{k_0})$ – оценки наименьших квадратов, найденные из системы нормальных уравнений.

Пусть

$$\hat{x}_i = \phi(b_i\,;\ \hat{\theta}\,)\,,\ i = 1\,,\ 2\,,\ \dots\,,\ n,$$

так называемый вектор «кажущихся наблюдений».

Составим величину (σ^2 - известна)

$$\gamma^2 = \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n} (x_i - \hat{x}_i)^2.$$

Имеем:

$$\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 = n(S_x^2 - S_{xb}^2/S_b^2) = n(S_{xx} - S_{xb}^2/S_{bb}).$$

Действительно,

$$\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 = \sum_{i=1}^{n} ((x_i - \bar{x}) - \frac{S_{xb}}{S_{bb}}(b_i - \bar{b}))^2 =$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + \frac{S_{xb}^2}{S_{bb}^2} \sum_{i=1}^{n} (b_i - \bar{b})^2 - 2\frac{S_{xb}}{S_{bb}} \sum_{i=1}^{n} (x_i - \bar{x})(b_i - \bar{b}) = n(S_{xx} - S_{xb}^2/S_{bb}).$$

Если бы модель была адекватной, т.е. степень полинома k была бы той, которая есть на самом деле, то величина γ^2 имела бы хи-квадрат распределение с $n - k - 1 = n - (k+1)$ степенями свободы ($k\ +1$ – число неизвестных коэффициентов полинома $\phi(b;\theta)$).

Зададим уровень значимости α и определим из таблиц хи-квадрат распределения квантили порядков $\frac{\alpha}{2}$ и $1 - \frac{\alpha}{2}$ (соответственно $\chi^2_{\alpha/2}(n-k-1)$ и $\chi^2_{1-\alpha/2}(n-k-1)$), т.е. чтобы вероятность принять очень маленькие или очень большие значения была не больше α. Обычно α берется маленьким, чтобы события с вероятностью α можно было считать практически невозможными. Тогда, если

$$\chi^2_{\alpha/2}(n-k-1) < \gamma^2 < \chi^2_{1-\alpha/2}(n-k-1),$$

то принимаем решение: данная модель полинома не противоречит выборочным данным.

Если

$$\gamma^2 \le \chi^2_{\alpha/2}(n-k-1),$$

то принимаем решение: степень полинома k завышена и ее следует уменьшить.

Если

$$\chi^2_{1-\alpha/2}(n-k-1) \le \gamma^2,$$

то принимаем решение: степень полинома k занижена и ее следует увеличить.

В последних двух случаях если бы модель была верной, то этого бы практически не произошло.

Пример 34. Закон Хаббла в астрономии утверждает, что скорость космоголологического разбегания галактических скоплений пропорциональна расстоянию до них. Современная методика измерений скорости достаточно точна, поэтому будем считать, что расстояния r (млн. световых лет) до скоплений случайны, а скорость v (сотни миль в сек.) неслучайна.

№	Созвездие	r_i	v_i	$\hat{r}_i$
1.	Дева	22	7.5	20.1
2.	Пегас	68	24	66.7
3.	Персей	108	32	89.3
4.	Волосы Вероники	137	47	131.6
5.	Б. Медведица №1	255	93	261.5
6.	Лев	315	120	337.7
7.	Северная Корона	390	134	377.2
8.	Близнецы	405	144	405.5
9.	Волопас	685	245	690.6
10.	Б. Медведица №2	700	260	733.0
11.	Гидра	1100	380	1071.7

По приведенным ниже данным оценим линейную зависимость $r = \theta_o + \theta_1 \cdot v$ расстояний от скорости . Здесь:

$$\bar{r} = 380.45454, \bar{v} = 135.13636, S_v^2 = 12432, S_{vr} = 35097, \hat{\theta}_1 = \frac{S_{vr}}{S_v^2} = 2.823,$$

$$\sum_{i=1}^{11} (r_i - \hat{r}_i)^2 = 3022.4, \sigma^2 = 100\,00, \gamma^2 = \frac{\sum_{i=1}^{11} (r_i - \hat{r}_i)^2}{\sigma^2} = 3.022.$$

$$\hat{\theta}_0 = \bar{r} - \hat{\theta}_1 \bar{v} = -1.0533, \;\; \hat{r} = 2.823v - 1.053.$$

Зададим уровень значимости $\alpha = 0.01$. Из таблиц хи-квадрат распределения с 10 степенями свободы находим $C_1 = \chi^2_{0.005} = 2.156$, $C_2 = \chi^2_{0.995} = 25.188$.

Так как

$$C_1 < \gamma^2 < C_2,$$

то принимаем решение: *степень полинома, равная 1 не отвергается*, поэтому выборочные данные говорят о хорошем согласии с законом Хаббла.

Пример 35. Предположим, что имеется m предметов, веса которых надо определить. Один из методов состоит в том, чтобы взвешивать каждый предмет r раз и среднее значение результатов взвешиваний принять в качестве оценки веса этого предмета. Такая процедура требует всего mr взвешиваний и точность каждой оценки будет σ^2/r, где σ^2 – средний квадрат ошибки индивидуального взвешивания. Другой метод состоит в том, чтобы взвешивать комбинации предметов. Каждая операция (одно взвешивание) состоит в том, что несколько предметов кладут на одну чашку весов, несколько – на другую и добавляют разновес, чтобы привести чашки в равновесие. В результате получают уравнение вида

$$b_1\theta_1 + \; ... \; + b_m\theta_m = x,$$

в котором θ_1 , ... θ_m – неизвестные веса предметов, $b_i = +1\,,\; -1\,,\; 0$ в зависимости от того, лежит i–й предмет на левой чашке, на правой чашке или вообще не участвует в данном взвешивании, а x – требуемый разновес. n операциям (взвешиваниям) соответствует n уравнений, из которых неизвестные веса θ_1 , ... , θ_m могут быть оценены по методу наименьших квадратов.

Оценить веса четырех предметов по данным следующей таблицы:

θ_1	θ_2	θ_3	θ_4	Вес x_i
1	1	1	1	20.2
1	-1	1	-1	8.1
1	1	-1	-1	9.7
1	-1	-1	1	1.9
1	1	1	1	19.9
1	-1	1	-1	8.3
1	1	-1	-1	10.2
1	-1	-1	1	1.8

Тогда

$$\mathbf{B}^T\mathbf{B} = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{pmatrix}, \qquad \mathbf{B}^T\mathbf{X} = \begin{pmatrix} 80.1 \\ 39.9 \\ 32.9 \\ 7.5 \end{pmatrix},$$

$$(\mathbf{B}^T\mathbf{B})^{-1} = \begin{pmatrix} 1/8 & 0 & 0 & 0 \\ 0 & 1/8 & 0 & 0 \\ 0 & 0 & 1/8 & 0 \\ 0 & 0 & 0 & 1/8 \end{pmatrix}, \quad (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X} = \begin{pmatrix} 10.0125 \\ 4.9875 \\ 4.1125 \\ 0.9375 \end{pmatrix}$$

Таким образом, оценки НК весов предметов равны соответственно 10.0125, 4.9875, 4.1125, 0.9375. Дисперсии оценок равны все по 1/8 (при $\sigma^2 = 1$). План – ортогональный.

3.6. Задачи.

1. Пусть

$$a_1 = \frac{\theta_1}{\theta_1 + \theta_2}, \; a_2 = \frac{\theta_2}{\theta_1 + \theta_2}, \;\; \theta_1, \; \theta_2 > 0,$$

и

$$\mathbf{P}(X = m) = a_1 \frac{\theta_1^{-m}}{m!} e^{-1/\theta_1} + a_2 \frac{\theta_2^{-m}}{m!} e^{-1/\theta_2}, \; m = 0, \; 1, \; 2, \; \ldots \; .$$

Используя выборку $x_1, \; x_2, \; \ldots, \; x_n$ найти оценки параметров $\theta_1, \; \theta_2$ по методу моментов. Найти значение этой оценки, если
$n = 15$, а выборка равна 5, 4, 4, 2, 0, 0, 2, 1, 1, 3, 2, 4, 1, 2, 0.

2. Пусть $X_1, \; X_2, \; \ldots, \; X_N$ – независимые случайные величины, причем

$$\mathbf{P}(X_i = m) = C_{n_i}^m p^m (1-p)^{n_i - m}, \; m = 0, \; 1, \; \ldots, n_i; \; 0 < p < 1; \; i = 1, \; 2, \; \ldots, N.$$

Найдите оценку максимального правдоподобия для p по выборке
$(n_1, m_1), \; (n_2, m_2), \; \ldots, \; (n_N, m_N)$, n_i – неслучайны, $i = 1, 2, \; \ldots, \; N$.

3. По выборке $x_1, \; x_2, \; \ldots, \; x_n$ из распределения с плотностью

$$f(x; \; \theta_1, \; \theta_2) = \alpha\,\theta_1 e^{-\theta_1 x} + (1-\alpha)\,\theta_2 e^{-\theta_2 x}, \;\; \left(\frac{2}{\theta_1 + \theta_2} = \frac{\alpha}{\theta_1} + \frac{1-\alpha}{\theta_2} \right),$$

$$x > 0; \; \theta_1 > 0, \; \theta_2 > 0, \; \alpha > 0.$$

найти по методу моментов оценки параметров θ_1 и θ_2. Определить ее значение по выборке 0.01, 0.02, 0.04, 0.35, 0.52, 1.1 .

4. Имеются две независимые выборки $x_1, \; x_2, \; \ldots, \; x_n$ и $x_{n+1}, \; x_{n+2}, \; \ldots, \; x_{n+m}$ из двух нормальных распределений с одинаковыми средними μ и дисперсиями $\lambda\sigma^2$ и σ^2 соответственно. По объединенной выборке объема $n + m$ найти оценку максимального правдоподобия для μ и σ^2, считая, что λ – известно.

5. Пусть $x_1\,,\ x_2\,,\ \ldots\,,\ x_n$ – выборка из распределения с плотностью

$$f(x\,;\ \theta_1\,,\ \theta_2) = \frac{\theta_1\theta_2}{\theta_2-\theta_1}e^{-\,\theta_1 x} - \frac{\theta_1\theta_2}{\theta_2-\theta_1}e^{-\,\theta_2 x}\,,$$

$$x>0\ ;\ \theta_1>0\,,\ \theta_2>0\,.$$

Используя метод моментов найти оценки параметров θ_1 и θ_2.

6. Пусть $x_1\,,\ x_2\,,\ \ldots\,,\ x_n$ – выборка из нормального распределения $N(\mu\ ,\ \gamma\,\mu^2)$ с плотностью распределения

$$f(x\,;\ \mu\,) = \frac{1}{\sqrt{2\pi\,\gamma\,\mu^2}}\,e^{-\,\frac{(x\,-\mu)^2}{2\gamma\mu^2}}\,,$$

где γ – известно. Найти оценку максимального правдоподобия для параметра μ . Определить ее значение по выборке

3.23, 1.10, 0.06, 3.17, 2.22, 2.94, 4.15, 1.08, 4.25, 2.79, 1.04, 2.36, 0.11, 3.32 для $\gamma=2$.

7. Пусть $x_1\,,\ x_2\,,\ \ldots\,,\ x_n$ – выборка из дискретного распределения

$$\mathbf{P}(X=m) = \frac{\theta_1\theta_2}{\theta_2-\theta_1}\,(1-\theta_1)^{m\,-\,1} - \frac{\theta_1\theta_2}{\theta_2-\theta_1}\,(1-\theta_2)^{m\,-\,1},\ m=1\,,\ 2\,,\ \ldots\ \ ,$$

$$0<\theta_1\,,\ \theta_2<\ 1.$$

Найти по методу моментов оценки параметров θ_1 и θ_2.

Определить их значение по выборке 5, 8, 6, 2, 9, 14, 2, 3, 1, 8.

Указание. Воспользоваться значениями следующих сумм:

$$\sum_{k=1}^{\infty} k\cdot x^{k-1} = \frac{1}{(1-x)^2}\ ,\ \ \sum_{k=1}^{\infty} k^2\cdot x^{k-1} = \frac{1+x}{(1-x)^3}\ .$$

8. Пусть $x_1\,,\ x_2\,,\ \ldots\,,\ x_n$ – выборка из распределения с плотностью

$$f(x\,;\ \theta\,) = \sqrt{\theta}\ x^{\sqrt{\theta}\,-1},\ \ 0\le x\le 1\,,\ \theta\ >0.$$

Найти оценку максимального правдоподобия для параметра θ.

9. Пусть $x_1\,,\ x_2\,,\ \ldots\,,\ x_n$ – выборка из распределения с плотностью

$$f(x\,;\ a\,,\ b) = b\,\cdot\frac{1}{\sqrt{2\pi}}\ e^{-\,\frac{x^2}{2}} + (1-b\,)\cdot\frac{1}{\sqrt{2\pi}}\ e^{-\,\frac{(x-a)^2}{2}}\ ,\ 0\le b\le 1.$$

Найти по методу моментов оценки параметров a и b. Определить их значения по выборке 0.32, 7.09, −3.24, 1.26, 10.76, 0.93, 1.14, 1.77, −6.39, 2.48, 1.68, −1.13, −1.54, 3.92.

10. Пусть $x_1\,,\ x_2\,,\ \ldots\,,\ x_n$ – выборка из дискретного распределения

$$\mathbf{P}(X=m) = \left(\frac{1+\theta}{2}\right)^m\cdot\left(\frac{1-\theta}{2}\right)^{1\,-\,m}\ ,\ m=0,\ 1\,,$$

$$-1\le\ \theta\le\ 1.$$

Найти оценку максимального правдоподобия для параметра θ.

11. Пусть $x_1\,,\ x_2\,,\ \ldots\,,\ x_n$ – выборка из дискретного распределения

$$\mathbf{P}(X=m) = \frac{1}{2}C_2^m p_1^m(1-p_1)^{2\,-\,m} + \frac{1}{2}C_2^m p_2^m(1-p_2)^{2\,-\,m}\,,\ m=0\,,\ 1\,,\ 2\,;$$

$$0<p_1\,,\ p_2<1\,.$$

Найти по методу моментов оценки параметров p_1 и p_2.

12. Пусть x_1, x_2, ... , x_n – выборка из дискретного распределения

$$\mathbf{P}(X=1)=q\,,\quad \mathbf{P}(X=2)=p\,(1-q),\quad \mathbf{P}(X=3)=(1-q)\,(1-p),$$

$(0<p<1,\ 0<q<1)$, p 8 q – неизвестны,
в которой k_1 «единиц», k_2 «двоек», k_3 «троек».
Найти оценки максимального правдоподобия для p и q.
Определить значения оценок по выборке
1, 2, 1, 3, 3, 2, 1, 2, 3, 1, 2, 1, 2, 2, 1, 2, 2, 3, 2, 3.

13. Пусть x_1, x_2, ... , x_n – выборка из распределения с плотностью

$$f(x\,;\,p_1\,,\,p_2)=\begin{cases}\dfrac{p_1p_2}{p_1+p_2}\,e^{-p_1x}, & \text{если } x>0\,;\\[2mm] \dfrac{p_1p_2}{p_1+p_2}\,e^{p_2x}, & \text{если } x<0\,;\end{cases}\qquad (p_1\,,\,p_2>0).$$

Найти оценки параметров p_1 и p_2 по методу моментов.

14. Пусть x_1, x_2, ... , x_n – выборка из дискретного распределения

$$\mathbf{P}(X=1)=1-p\,,\quad \mathbf{P}(X=2)=p\,(1-p)\,,\quad \mathbf{P}(X=3)=p^2\,,\qquad 0<p<1\,,$$

в которой $m_1=168$ «единиц», $m_2=243$ «двоек» и $m_3=127$ «троек».
Найти оценку максимального правдоподобия для параметра p.

15. Пусть x_1, x_2, ... , x_n – выборка из распределения с плотностью

$$f(x;\theta)=\begin{cases}\dfrac{p}{\theta}\,, & \text{если } 0<x<\theta\,;\\[2mm] \dfrac{1-p}{1-\theta}\,, & \text{если } \theta<x<1\,;\end{cases}\qquad 0\le p\le 1\,;$$

p, θ – неизвестны. Найти оценки параметров $p\,,\theta$ по методу моментов.
Определить значения оценок по выборке 0.05, 0.045, 0.595, 0.125,
0.165, 0.64, 0.28, 0.005, 0.175, 0.79.

16. Пусть x_1, x_2, ... , x_n – выборка из дискретного распределения

$$\mathbf{P}(X=1)=p\,,\quad \mathbf{P}(X=2)=p\,(1-p),\quad \mathbf{P}(X=3)=p\,(1-p)^2,$$

$\mathbf{P}(X=4)=(1-p)^3$, $(0<p<1)$, p – неизвестно,
в которой k_1 «единиц», k_2 «двоек», k_3 «троек» и k_4 «четверок».
Найти оценку максимального правдоподобия.

17. Найти по методу моментов оценки параметров λ_1 и λ_2 двойного распределения Пуассона

$$\mathbf{P}(X=k)=\frac{1}{2\,k!}\,\lambda_1^k\,e^{-\lambda_1}+\frac{1}{2\,k!}\,\lambda_2^k\,e^{-\lambda_2},\ k=0\,,\,1\,,\,\dots\,,$$

по выборке объема $n=327$ (x_i – число появлений события, m_i – количество испытаний, в которых появилось это значение). Найти значение этих оценок.

x_i	0	1	2	3	4	5	6	7	8	9	10
m_i	28	47	81	67	53	24	13	8	3	2	1

18. Случайная величина X имеет распределение:

$$p_1=\frac{2\,(1-\theta)}{3}\,,\quad p_2=\frac{(1+\theta)}{3}\,,\quad p_3=\frac{\theta}{3}\qquad (0<\theta<1)\,,$$

k_i $(i=1\,,\,2\,,\,3)$ – значения, которые наблюдались в выборке.
Оценить θ по методу МП. Найти значение оценки $\hat{\theta}_n$, если $k_1=2044$, $k_2=928$, $k_3=345$.

19. Для распределения с плотностью $f(x)=\frac{1}{2\lambda_1}\,e^{-x/\lambda_1}+\frac{1}{2\lambda_2}\,e^{-x/\lambda_2}$, $x>0$,
найти методом моментов оценки параметров λ_1 и λ_2 по повторной выборке x_1, x_2, ... , x_n.

20. В таблице приведены частоты азимутов тех точек на горизонте, в которых наблюдатель, выпустивший немигрирующих английских крякв, видел их в последний раз (такие азимуты называют углами исчезновения). Утки выпускались в различное время года в окрестности Стиммбриджа (графство Глочектершир) из различных точек на расстоянии от 30 до 250 *км*.

Направление	0^0	20^0	40^0	60^0	80^0	100^0
Кол-во птиц	40	22	20	9	6	3

120^0	140^0	160^0	180^0	200^0	220^0	240^0
3	1	6	3	11	22	24
260^0	280^0	300^0	320^0	340^0	Всего	
58	136	138	143	69	714	

Для распределения с плотностью

$$f(x) = \frac{1}{2\pi I_0(2)} \exp\left(2\cos(x-\mu)\right), \quad 0 \le \mu < 360^0,\ 0 \le x < 360^0,$$

где $I_0(2) = 2.2795851$, найдите оценку МП параметра μ.

21. Пусть $X_1, \ldots, X_n$ независимы и одинаково распределены,

$$\mathbf{P}(X_1 = k) = \theta^{k-1}(1-\theta),\ k = 1,\ 2,\ \ldots,\ r,\ \mathbf{P}(X_1 = r+1) = \theta^r,\ 0 < \theta < 1.$$

Пусть $M-$ число индексов i, для которых $X_i = r+1$. Показать, что ОМП для θ есть

$$\hat{\theta} = \frac{\bar{X} - 1}{\bar{X} - M/n}.$$

22. Выборка $x_1,\ x_2,\ \ldots,\ x_n$ взята из распределения Парето с функцией распределения

$$F(x) = 1 - \left(\frac{\lambda}{x}\right)^a,\ x \ge \lambda;\lambda > 0,\ a > 0.$$

Найти оценки параметров λ и a по методу моментов.

23. Пусть $X_i = \theta_0 + \theta_1 b_i + \varepsilon_i$, где $\{\varepsilon_i\}-$ независимы и одинаково нормально $N(0,\sigma^2)$ распределены, $\{b_i\}-$ неслучайны и известны. Найдите МП-оценку параметра $\theta = (\theta_0, \theta_1, \sigma^2)$.

24. Функция распределения случайной величины X равна

$$G(x;\xi,\beta) = 1 + \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi},\ x > 0;\ \ 0 < \xi < 1/2,\ \beta > 0.$$

Используя метод моментов найти оценки ξ, β по выборке $x_1,\ x_2,\ \ldots,\ x_n$.

25. Пусть

$$X_i = \theta\, t_i + \varepsilon_i, \quad i = 1,\ \ldots,\ n,$$

где $\theta \in R$ есть неизвестный параметр, $t_i \in (a,b)-$ известны, $a, b-$ известные положительные константы, ε_i- независимые случайные величины, имеющие нормальное распределение $N(0, \sigma^2 t_i)$ с неизвестным $\sigma^2 > 0$. Найти оценку $\hat{\theta}$ параметра θ по методу максимального правдоподобия.

26. Случайная величина X имеет распределение

$$\mathbf{P}(X = 0) = 1 - \varepsilon + \varepsilon \cdot q^N,$$

$$\mathbf{P}(X = j) = \varepsilon\, C_N^j p^j q^{N-j},\ j = 1,\ 2,\ \ldots, N,\ 0 < \varepsilon < 1,\ 0 < p < 1,\ p + q = 1,\ N > 1.$$

Используя метод моментов найти по выборке $x_1,\ x_2,\ \ldots,\ x_n$ из этого распределения оценки для параметров ε и p.

27. Пусть $X_1, \dots, X_n$ выборка независимых случайных величин из распределения Парето

$$\mathbf{P}(X_i \geq x) = \exp(-(x-\theta)), \quad x \geq \theta.$$

Найти МП-оценку для параметра θ.

28. Случайная величина X имеет распределение Накагами с плотностью

$$f(x; \mu, \omega) = \frac{2}{\Gamma(\mu)} \left(\frac{\mu}{\omega}\right)^{\mu} x^{2\mu-1} \exp\left(-\frac{\mu}{\omega} x^2\right), \ x > 0; \ \Gamma(\mu) = \int_0^{\infty} t^{\mu-1} e^{-t} dt\,.$$

Используя метод моментов и $\mathbf{E}(X^2)$, $\mathbf{E}(X^4)$ найти по выборке x_1, x_2, ... , x_n из этого распределения оценки для параметров μ и ω.

29. Пусть $X_1, \dots, X_n$ выборка независимых случайных величин из распределения Парето

$$\mathbf{P}(X_i \geq x) = \left(\frac{\lambda}{x}\right)^{a}, \ \lambda > 0, \ a > 0, \ x \geq \lambda.$$

Считая, что $\lambda-$ известно, найти оценку метода МП для параметра a.

30. Случайная величина X имеет распределение

$$\mathbf{P}(X = j) = \varepsilon\, p\, (1-p)^{j-1} + (1-\varepsilon)\, 2^{-j}, \ j = 1, \ 2, \ \dots, \ 0 < \varepsilon < 1, \ 0 < p < 1.$$

Используя метод моментов найти по выборке x_1, x_2, ... , x_n из этого распределения оценки для параметров ε и p.

31. Дискретная случайная величина имеет распределение:

$$\mathbf{P}(Y_t = 1) = F(\alpha + \beta d_t), \mathbf{P}(Y_t = 0) = 1 - F(\alpha + \beta d_t), t = 1, 2, \dots, n,$$

$$F(x) = \frac{e^x}{1+e^x}, \ -\infty < x < \infty.$$

Ниже представлены результаты 100 наблюдений из данного распределения.

		y	
		0	1
d	0	20	32
	1	36	12

Оцените по методу максимального правдоподобия параметры α, β.

32. Случайная величина X имеет распределение

$$\mathbf{P}(X = 1) = 1 - \varepsilon + \varepsilon p, \mathbf{P}(X = j) = \varepsilon\, p q^{j-1}, \ j = 2, 3, \ \dots, \ 0 < \varepsilon < 1, \ 0 < p < 1, \ p + q = 1.$$

Используя метод моментов найти по выборке x_1, x_2, ... , x_n из этого распределения оценки для параметров ε и p.

33. Рассматривается гауссовский процесс авторегрессии первого порядка $X_t = \rho X_{t-1} + \varepsilon_t$, $t = 1, 2, \dots, n$, $|\rho| < 1$, $\{\varepsilon_t\}_{1 \leq t \leq n}-$ независимы, одинаково распределены и имеют нормальное распределение $N(0, \sigma^2)$, $X_0 = 0$. Найти оценку максимального правдоподобия $\hat{\theta}$ для $\theta = (\rho, \sigma^2)$.

Указание. Использовать то, что процесс $\{X_t\}_{t \geq 1}$ является марковским, и выписать совместную плотность распределения.

34. Пусть случайная величина X имеет характеристическую функцию

$$\varphi(t) = \exp\left(\frac{\lambda}{\mu} - \frac{\lambda}{\mu}\left(1 - \frac{2i\mu^2 t}{\lambda}\right)^{1/2}\right), \ i = \sqrt{-1}.$$

По выборке x_1 , x_2 , ... , x_n из этого распределения найти оценки для параметров λ и μ.

35. При изучении размножения яблонь Martin, Jones and Hadlow измерили число корней, репродуцированных культивируемыми яблонями. Получены следующие данные:

Число корней	0	1	2	3	4	5	6	7	8	9	Объем выборки
Частоты	19	2	2	4	3	1	4	3	0	2	40

Рассматривается следующая модель

$$\mathbf{P}(X=0) = p + (1-p)\,e^{-\lambda}; \mathbf{P}(X=k) = (1-p)\,\frac{\lambda^k}{k!}e^{-\lambda},\ k = 1, 2,\ ....$$

1) Используя метод моментов найдите оценки неизвестных параметров p и λ. Определите значения построенных оценок.

2) Предполагая, что параметр λ известен и равен 4.5 покажите, что МП-оценка параметра p имеет вид: $\hat{p} = \dfrac{n_0 - n\,e^{-\lambda}}{n\,(1-e^{-\lambda})}$, если $n_0 > n\,e^{-\lambda}$; $\hat{p} = 0$ если $n_0 \le n\,e^{-\lambda}$, где n_0- число нулей в выборке.

Подсчитайте значение этой оценки. Сравните со значением ММ-оценки.

36. Пусть $X_1, \ldots, X_n$ выборка независимых случайных величин из распределения Парето

$$\mathbf{P}(X_i \ge x) = \left(\frac{\lambda}{x}\right)^a,\ \lambda > 0\,,\ a > 0\,,\ x \ge \lambda.$$

Считая, что $a-$ известно, найти МП-оценку для параметра λ.

37. Случайная величина X имеет распределение

$$\mathbf{P}(X=k) = \frac{\theta^k e^{-\theta}}{2k!} + \frac{1}{2}\cdot C_n^k p^k (1-p)^{n-k},\ k = 0,\ 1,\ \ldots,\ \theta > 0,\ 0 < p < 1,\ C_n^k = 0\ \text{ для }\ k > n.$$

Найти оценки метода моментов для θ и p.

38. Пусть $X_1, \ldots, X_n$ выборка независимых случайных величин из распределения с плотностью

$$f(x) = \frac{a^{p-1}x^{-p}}{\Gamma(p-1)}\exp\left(-\frac{a}{x}\right),\ a > 0\,,\ p > 0,\ x \ge 0.$$

Найти МП-оценку для параметра a, считая $p > 1$ известным.

39. Найти оценки по методу моментов для параметров a и p.
Указание. ***Найти*** $\boldsymbol{E}\left(\frac{1}{X}\right)$ ***и*** $\boldsymbol{E}\left(\frac{1}{X^2}\right)$.

40. Случайная величина X имеет распределение

$$\mathbf{P}(X=1) = 1 - \varepsilon + \varepsilon \cdot e^{-\lambda},$$

$$\mathbf{P}(X=j) = \varepsilon\,\frac{\lambda^{j-1}}{(j-1)\,!}e^{-\lambda},\ j = 2,\ 3, \ldots,\ 0 < \varepsilon < 1\,,\ 0 < \lambda.$$

Используя метод моментов найти по выборке x_1 , x_2 , ... , x_n из этого распределения оценки для параметров ε и λ.

41. Пусть $X_1, \ldots, X_n$ выборка независимых случайных величин из распределения Парето

$$\mathbf{P}(X_i \ge x) = \left(\frac{\lambda}{x}\right)^a,\ \lambda > 0\,,\ a > 0\,,\ x \ge \lambda.$$

Считая, что $\lambda-$ известно, найти оценку метода МП для параметра a.

Глава 4.

Обобщенный и корректный методы моментов.

4.1. Обобщенный метод моментов.

Обычный метод моментов требует, чтобы число уравнений равнялось числу моментов. Если число уравнений будет меньше, чем число моментов, то мы не сможем однозначно найти неизвестные параметры; если же число моментов будет больше, чем число параметров, то система, как правило, будет противоречивой. Для нахождения неизвестных параметров в этом случае мы применим ***обобщенный метод моментов***.

Пусть имеется векторная функция $\boldsymbol{g}(\boldsymbol{x})$, размерности l и $\theta \in \Theta \subset R^k$, $l \geq k$, и $\mathbf{E}_\theta(\boldsymbol{g}(\boldsymbol{X});\theta)) = \mu(\theta)$, или $\mathbf{E}_\theta(\boldsymbol{g}(\boldsymbol{X})) - \mu(\theta)) = 0$.

Найдем их выборочные аналоги

$$\boldsymbol{m} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{x}_i).$$

Оценкой обобщенного метода моментов (GMM) называется статистика $\hat{\theta} = \arg\min_\theta \boldsymbol{m}^T W(\theta)\boldsymbol{m}$, где $W(\theta)$ некоторая симметрическая положительно определенная матрица.

Наиболее эффективными являются GMM-оценки с весовой матрицей, равной обратной ковариационной матрице моментных функций $W(\theta) = \boldsymbol{V}_{\boldsymbol{gg}}^{-1}(\theta)$. Это, так называемый, *эффективный GMM*. Однако, поскольку на практике эта ковариационная матрица неизвестна, то применяют *двухшаговую* процедуру:

(1) Оцениваются параметры модели с помощью GMM с единичной весовой матрицей.
(2) По выборочным данным и найденным на первом шаге значениям параметров оценивают ковариационную матрицу моментных функций $\hat{\boldsymbol{V}}_{\boldsymbol{gg}}(\theta) = \overline{\boldsymbol{g}(\boldsymbol{x})\boldsymbol{g}^T(\boldsymbol{x})}$ и используют полученную оценку в эффективном GMM

Эту двухшаговую процедуру можно продолжить (*итеративный GMM*) используя оценки параметров на втором шаге и т.д. до достижения требуемой точности. Также возможен подход к численный минимизации $\boldsymbol{m}^T \boldsymbol{V}_{\boldsymbol{gg}}^{-1}(\theta)\boldsymbol{m}$ с неизвестным параметром θ. Это так называемый *непрерывно обновляемый* GMM.

Пример 36. Пусть $\boldsymbol{y}_t = \boldsymbol{x}_t^T\theta + \varepsilon_t$ удовлетворяет условию $\mathbf{E}(\boldsymbol{x}_t^T\varepsilon_t) = 0$. Тогда условия на моменты выглядят следующим образом:

$\boldsymbol{X}^T\varepsilon = 0 \;\Rightarrow\; \boldsymbol{X}^T(\boldsymbol{Y} - \boldsymbol{X}\theta) = 0 \;\Rightarrow\; \boldsymbol{X}^T\boldsymbol{X}\theta = \boldsymbol{X}^T\boldsymbol{Y} \;\Rightarrow\; \hat{\theta} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ – оценка МНК.

Таким образом, МНК является частным случаем метода моментов.

Рассмотрим другой случай, когда имеются некоторые переменные z, ортогональные случайным ошибкам линейно регрессионной модели, т.е. $\mathbf{E}(z_t^T\varepsilon_t) = 0$. Тогда имеем выборочный аналог этого условия:

$\mathbf{Z}^T\varepsilon = 0 \;\Rightarrow\; \mathbf{Z}^T(\mathbf{Y} - \mathbf{X}\theta) = 0 \;\Rightarrow\; \mathbf{Z}^T\mathbf{X}\theta = \mathbf{Z}^T\mathbf{Y} \;\Rightarrow\; \hat{\theta} = (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{Y}$ – оценка *метода инструментальных переменных.*

Пример 37. Пусть известно, что распределение z симметрично, т.е. $\mathbf{E}((z-\theta)^3)=0$, и мы хотим оценить среднее $\theta=\mathbf{E}(z)$. Тогда условие на моменты запишется так:

$$g(z)-\mu(\theta)=\begin{pmatrix} z-\theta \\ (z-\theta)^3 \end{pmatrix}.$$

Построим эффективную GMM-оценку.

Здесь

$$V_{gg}=\mathbf{E}\left(\begin{pmatrix} (z-\theta)^2 & (z-\theta)^4 \\ (z-\theta)^4 & (z-\theta)^6 \end{pmatrix}\right)=\begin{pmatrix} \mu_2 & \mu_4 \\ \mu_4 & \mu_6 \end{pmatrix}.$$

Вместо μ_j подставим их выборочные аналоги $m_j=\frac{1}{n}\sum_{i=1}^{n}(z_i-\theta)^j$.

Обратная матрица равна

$$V_{gg}^{-1}=\frac{1}{\mu_2\mu_6-\mu_4^2}\begin{pmatrix} \mu_6 & -\mu_4 \\ -\mu_4 & \mu_2 \end{pmatrix},$$

поэтому необходимо минимизировать

$$\frac{1}{m_2m_6-m_4^2}(m_1,m_3)\begin{pmatrix} m_6 & -m_4 \\ -m_4 & m_2 \end{pmatrix}\begin{pmatrix} m_1 \\ m_3 \end{pmatrix}=\frac{m_1^2m_6-2m_1m_3m_4+m_2m_3^2}{m_2m_6-m_4^2}.$$

Отметим также следующий факт, что если $\hat{\theta}$ является GMM-оценкой, то

$$\sqrt{n}(\hat{\theta}-\theta)\underset{n\to\infty}{\xrightarrow{d}}N(0,V_b),$$

где (в случае эффективного GMM)

$V_b=G^TV_{gg}^{-1}G$, $G-$математическое ожидание первых производных g по параметрам.

В нашем случае

$G=\mathbf{E}\begin{pmatrix} -1 \\ -3(z-\theta)^2 \end{pmatrix}=-\begin{pmatrix} 1 \\ 3\mu_2 \end{pmatrix}$, поэтому (показать самим)

$$\hat{V}_b=\frac{m_2m_6-m_4^2}{m_6-6m_2m_4+9m_2^3}\underset{n\to\infty}{\xrightarrow{p}}V_b=\frac{\mu_2\mu_6-\mu_4^2}{\mu_6-6\mu_2\mu_4+9\mu_2^3}.$$

Для нормального распределения $\mu_2=\sigma^2$, $\mu_4=3\sigma^4$, $\mu_6=15\sigma^6$, поэтому $V_b=\sigma^2$, т.е. использование третьего момента не дает выигрыша в эффективности, поскольку для нормального распределения $\bar{x}$ является эффективной оценкой.

Если же распределение не является нормальным, т.е. $\mu_4\neq 3\mu_2^2$, то вновь построенная оценка является более эффективной (можно показать), что

$$\mu_2>\frac{\mu_2\mu_6-\mu_4^2}{\mu_6-6\mu_2\mu_4+9\mu_2^3}.$$

4.2. Метод разделяющих разбиений.

Понятие разделяющих разбиений было введено А.М.Каганом в работе Каган А.М. Семейства распределений и разделяющие разбиения. – ДАН СССР, 1963, т.153, №3., а в работе Каган А.М. Замечания о разделяющих разбиениях. – Труды МИ АН СССР, 1965, т.79, с.26-31, был дан пример построения асимптотически эффективной оценки для параметра Вейбулла. Заметим, что ни метод МП, ни метод моментов не дает аналитически вычислимой оценки.

Пусть задано семейство вероятностных распределений $\mathcal{P}=\{\mathbf{P}_\theta\}_{\theta\in\Theta}$ на (R,B).

Определение 10. Разбиение прямой $\mathbf{R}$ на непересекающиеся множества $A_1,...,A_q$; $A_i\in\mathbf{R}$ $(i=1,...,q)$; $A_1\bigcup...\bigcup A_q=\mathbf{R}$ будем называть *разделяющим* семейство $\mathcal{P}$, если из условия $\mathbf{P}_\theta(A_i)=\mathbf{P}_t(A_i)$ для $i=1,...,q$ следует $\theta=t$.

Пример 38. В качестве такого семейства рассмотрим семейство распределений Вейбулла, которые зависят от параметров $\alpha > 0$ и $\mu > 1$, задаваемое плотностью распределения

$$f(x;\alpha,\mu) = \frac{\mu}{\alpha} x^{\mu-1} e^{-\frac{x^\mu}{\alpha}}, \ \ x > 0, \quad \text{и} \quad 0 \ \text{в противном случае.}$$

Пусть $A_1 = (0,1)$, $A_2 = [1,2]$, $A_3 = (2,\infty)$, $\theta = (\alpha,\mu)$.
Тогда

$$\left\{ \mathbf{P}_\theta(A_1) = 1 - e^{-\frac{1}{\alpha}}, \ \mathbf{P}_\theta(A_2) = e^{-\frac{1}{\alpha}} - e^{-\frac{2^\mu}{\alpha}}, \ \mathbf{P}_\theta(A_3) = e^{-\frac{2^\mu}{\alpha}}. \right.$$

Отсюда видно, что разбиение (A_1, A_2, A_3) является разделяющим и

$$\alpha = -\left(\ln(1 - \mathbf{P}_\theta(A_1))\right)^{-1}, \ \mu = \log_2 \frac{\ln \mathbf{P}_\theta(A_3)}{\ln\left(1 - \mathbf{P}_\theta(A_1)\right)}.$$

Если $\nu_1 = \frac{m_1}{n}$, $\nu_3 = \frac{m_3}{n}$ – частоты попадания выборки $x_1, ..., x_n$ соответственно в интервалы A_1 и A_3 соответственно, то в качестве оценок можно взять (считая, что $\nu_i < 1$)

$$\tilde{\alpha} = -\left(\ln(1 - \nu_1)\right)^{-1}, \ \tilde{\mu} = \log_2 \frac{\ln(\nu_3)}{\ln(1-\nu_1)}.$$

Поскольку функции от частот непрерывны, то мы получим состоятельные оценки.
Матрица информации Фишера имеет вид

$$I(\theta) = \begin{pmatrix} \alpha^{-2} & -\mu\alpha^{-1-1/\mu}\Gamma(2-1/\mu) \\ -\mu\alpha^{-1-1/\mu}\Gamma(2-1/\mu) & \mu^{-2} + \mu(\mu-1)\alpha^{-2/\mu}\Gamma(2-2/\mu \end{pmatrix}.$$

Найдем обратную к ней

$$I^{-1}(\theta) = \begin{pmatrix} J^{11}(\theta) & J^{12}(\theta) \\ J^{21}(\theta) & J^{22}(\theta) \end{pmatrix}.$$

Можно показать, что оценка $\hat{\theta} = (\hat{\alpha}, \hat{\mu})$, где

$$\hat{\alpha} = \tilde{\alpha} + \frac{1}{n}\left(J^{11}(-n\tilde{\alpha} + \tilde{\alpha}^{-2}\sum_{i=1}^n x_i^{\tilde{\mu}}) + J^{12}(n\tilde{\mu}^{-1} - \sum_{i=1}^n \ln x_i - \tilde{\mu}\tilde{\alpha}^{-1}\sum_{i=1}^n x_i^{\tilde{\mu}-1}) \right),$$

$$\hat{\mu} = \tilde{\mu} + \frac{1}{n}\left(J^{21}(-n\tilde{\alpha} + \tilde{\alpha}^{-2}\sum_{i=1}^n x_i^{\tilde{\mu}}) + J^{22}(n\tilde{\mu}^{-1} - \sum_{i=1}^n \ln x_i - \tilde{\mu}\tilde{\alpha}^{-1}\sum_{i=1}^n x_i^{\tilde{\mu}-1}) \right),$$

будет асимптотически нормальной и асимптотически эффективной оценкой параметра θ.

Пример 39. В качестве такого семейства рассмотрим семейство распределений Фреше (обратное распределение Вейбулла), которые зависят от параметров $\alpha > 0$ и $\sigma > 0$, задаваемое функцией распределения

$$F(x) = \exp\left(-\frac{x^{-\alpha}}{\sigma}\right), x > 0.$$

Пусть $A_1 = (0,1)$, $A_2 = [1,2]$, $A_3 = (2,\infty)$, $\theta = (\alpha,\mu)$.
Тогда

$$\left\{ p_1 = e^{-\frac{1}{\alpha}}, \ p_2 = -e^{-\frac{1}{\alpha}} + e^{-\frac{2^{-\alpha}}{\sigma}}, \ p_3 = e^{-\frac{2^{-\alpha}}{\sigma}}. \right.$$

Если $\nu_1 = \frac{m_1}{n}$, $\nu_3 = \frac{m_3}{n}$ – частоты попадания выборки $x_1, ..., x_n$ соответственно в интервалы A_1 и A_3 соответственно, то в качестве оценок можно взять (считая, что $\nu_i < 1$)

$$\tilde{\sigma} = -\frac{1}{\ln \nu_1},\ \tilde{\alpha} = \log_2 \frac{\ln\,\nu_1}{\ln(1-\nu_3)}.$$

4.3. Преобразование переменных.

Пример 40. Рассмотрим распределение Вейбулла

$$F(x) = 1 - \exp\left(-\left(\frac{x}{\sigma}\right)^{\alpha}\right),\ x > 0, \sigma > 0, \alpha > 0.$$

Дважды беря логарифм, получим соотношение $(V):\quad z = \alpha y - \lambda$, где

$$z = \ln(-\ln(1 - F(x)),\ y = \ln\,x,\ \lambda = \alpha \ln\,\sigma.$$

Соотношение (V) есть уравнение прямой и мы можем применить метод наименьших квадратов. В таком случае получим следующие оценки:

$$\hat{\alpha} = \frac{\hat{\nu}_1\hat{\tau}_1 - \hat{\nu}_2}{\hat{\tau}_1^2 - \hat{\tau}_2},\ \hat{\lambda} = \frac{\hat{\nu}_1\hat{\tau}_2 - \hat{\nu}_2\hat{\tau}_1}{\hat{\tau}_1^2 - \hat{\tau}_2},$$

где

$$\hat{\tau}_j = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i)^j,\ \ \hat{\nu}_j = \frac{1}{n}\sum_{i=1}^{n-1}(\ln x_i)^{j-1}\ln(-\ln\left(1 - \frac{i}{n}\right),\ j = 1, 2.$$

Этим же приемами могут быть найдены оценки параметров для следующих распределений:

(1) Распределение Фреше (обратное распределение Вейбулла) $F(x) = \exp\left(-\left(\frac{x}{\sigma}\right)^{-\alpha}\right)$, $x > 0,\ \sigma > 0,\ \alpha > 0.$

(2) Распределение Гнеденко-Гумбеля $F(x) = -\exp(-\exp(\alpha(x - \ln\sigma))),\ x \in R.$

(3) Логистическое распределение $F(x) = \dfrac{1}{1 + \exp\left(\left(\frac{|x|}{\sigma}\right)^{\alpha}\cdot \operatorname{sign}(x)\right)}$, $x \in R.$

4.4. Корректный вариант метода моментов.

Мы рассмотрели различные методы оценивания параметров распределений: метод моментов (ММ, 1894) , метод максимального правдоподобия (ММП, 1912), байесовские оценки Нужно отметить, что это методы были придуманы достаточно давно. Относительно недавно (1982, Хэнсен Л.) был предложен обобщенный метод моментов. Следует отметить, что некоторые методы и алгоритмы мы не рассматривали, например, ЕМ-алгоритм. Отметим также, что здесь рассматривались, в основном, повторные выборки и сюда не вошли оценки по цензурированным выборкам.

Мы изложим подход, предложенный в статье Кагана А.М. [18]. Рассмотрим следующий вопрос: связаны ли, каким-либо образом метод моментов и метод максимального правдоподобия? Чтобы ответить на этот вопрос рассмотрим следующую функцию

$$J = J(x;\theta) = \frac{\partial \ln f(x;\theta)}{\partial\theta},$$

где $f(x;\theta)$ есть плотность по некоторой мере μ (для простоты рассуждений будем считать ее лебеговой мерой на прямой, а параметр $\theta \in \Theta \subset R^1$).

Рассмотрим линейное пространство L_θ^2 и определим скалярное произведение

$$(\varphi, \psi)_\theta = \int_{-\infty}^{\infty} \varphi(x)\,\psi(x)\,f(x;\theta)\,dx = \int \varphi\,\psi f\,dx.$$

Информационное количество Фишера тогда определится как

$$I(\theta) = \|J(x;\theta)\|^2 = (J, J)_\theta = \mathbf{E}_\theta \left(\frac{\partial \ln f(x;\theta)}{\partial \theta} \right)^2,$$

а оценка максимального правдоподобия параметра θ по выборке $x_1, x_2, ..., x_n$ есть решение уравнения

$$\sum_{i=0}^{n} J(x_i;\theta) = 0 \tag{1}$$

Пусть $\mathcal{H}$ – конечномерное подпространство и

$$J(x, \theta; \mathcal{H}) = J(\mathcal{H}, \theta) = \hat{\mathbf{E}}_\theta(J|\mathcal{H})$$

– оператор условного математического ожидания в L^2_θ. Тогда Фишеровская информация о θ, содержащаяся в $\mathcal{H}$ есть

$$I(\theta; \mathcal{H}) = \|J(\mathcal{H}, \theta)\|^2_\theta.$$

Пусть $\mathcal{H}$ – линейное пространство, порожденное элементами $\varphi_1, ..., \varphi_m$, т.е.

$$\mathcal{H} = \{c_1\varphi_1 + ... + c_m\varphi_m; c_1, ..., c_m - \text{действительные числа}\}$$

$$\pi_{ij}(\theta) = (\varphi_i, \varphi_j)_\theta, \quad \pi_i(\theta) = (\varphi_i, \mathbb{I})_\theta, \quad \text{где}, \quad (\mathbb{I}, \mathbb{I})_\theta = \|\mathbb{I}\|^2 = 1.$$

Условие (A). *Функции $\pi_{ij}(\theta)$ дифференцируемы.*
Условие (B). *Потребуем, чтобы матрица Грама элементов $\varphi_1, ..., \varphi_m$*

$$\left\| \begin{array}{cccc} \pi_{11}(\theta) & . \ . \ . & \pi_{1m}(\theta) \\ . & . \ . \ . \ . & . \\ \pi_{m1}(\theta) & . \ . \ . & \pi_{mm}(\theta) \end{array} \right\|$$

была невырожденной.

Введем матрицу

$$\Lambda(\theta) = \| \lambda_{ij}(\theta) \| = \left\| \begin{array}{cccc} \theta & \pi'_1(\theta) & ... & \pi'_m(\theta) \\ \pi'_1(\theta) & \pi_{11}(\theta) & ... & \pi_{1m}(\theta) \\ . & . & ... & . \\ \pi'_m(\theta) & \pi_{m1}(\theta) & ... & \pi_{mm}(\theta) \end{array} \right\|$$

и пусть $\Lambda_{ij}(\theta)$ есть алгебраическое дополнение элемента $\lambda_{ij}(\theta)$.

ТЕОРЕМА 6. *Если выполнены условия* (A), (B), *то*

$$J(\mathcal{H}, \theta) = -\sum_{j=1}^{m} \frac{\Lambda_{1,j+1}(\theta)}{\Lambda_{11}(\theta)} \varphi_j. \tag{2.4}$$

Д о к а з а т е л ь с т в о. Имеем равенство

$$0 = \left\| \begin{array}{cccc} \pi'_j(\theta) & \pi_{j1}(\theta) & ... & \pi_{jm}(\theta) \\ \pi'_1(\theta) & \pi_{11}(\theta) & ... & \pi_{1m}(\theta) \\ . & . & ... & . \\ \pi'_m(\theta) & \pi_{m1}(\theta) & ... & \pi_{mm}(\theta) \end{array} \right\| = \pi'_m(\theta)\Lambda_{11}(\theta) + \pi_{j1}\Lambda_{12}(\theta) + ... + \pi_{jm}\Lambda_{1,m+1}(\theta),$$

откуда

$$\pi'_j = -\sum_{k=1}^{m} \frac{\Lambda_{1,k+1}(\theta)}{\Lambda_{11}(\theta)} \pi_{jk}, \quad j = 1, ..., m. \tag{2.5}$$

Определенный равенством элемент (2.4) удовлетворяет (2.5), отсюда следует теорема.

Заметим, что $(J, \varphi)_\theta = \dfrac{d}{d\theta}(\varphi, \mathbb{I})_\theta$.

ТЕОРЕМА 7. *Имеет место явная формула для информации*

$$I(\theta, \mathcal{H}) = -\frac{|\Lambda(\theta)|}{\Lambda_{11}(\theta)}.$$

Д о к а з а т е л ь с т в о. Имеем:

$$I(\theta) = (J, J)_\theta = -\sum_{j=1}^{m} \frac{\Lambda_{1,j+1}(\theta)}{\Lambda_{11}(\theta)}(J, \varphi_j)_\theta = -\sum_{j=1}^{m} \frac{\Lambda_{1,j+1}(\theta)}{\Lambda_{11}(\theta)}\pi_j'(\theta) = -\frac{|\Lambda(\theta)|}{\Lambda_{11}(\theta)}.$$

Если к тому же функции $\pi_{ij}(\theta)$ дифференцируемы, то

$$I(\theta; \mathcal{H}) = -(J, \mathbb{I})_\theta.$$

Действительно,

$$J = \sum_{j=1}^{m} \lambda_j(\theta)\varphi_j, \quad \text{где} \quad \lambda_j(\theta) = -\frac{\Lambda_{1,j+1}(\theta)}{\Lambda_{11}(\theta)},$$

так что $J' = \sum\limits_{j=1}^{m} \lambda_j'(\theta)\varphi_j$ и мы имеем:

$$-(J', \mathbb{I})_\theta = \sum_{j=1}^{m} \frac{\Lambda_{1,j+1}'(\theta)\Lambda_{11}(\theta) - \Lambda_{11}'(\theta)\Lambda_{1,j+1}(\theta)}{\Lambda_{11}^2(\theta)}\pi_j(\theta) = \sum_{j=1}^{m} \frac{\Lambda_{1,j+1}'(\theta)}{\Lambda_{11}(\theta)}\pi_j(\theta).$$

Так как

$$(J, \mathbb{I})_\theta = -\sum_{j=1}^{m} \frac{\Lambda_{1,j+1}(\theta)}{\Lambda_{11}^2(\theta)}\pi_j(\theta) = 0,$$

то дифференцируя это равенство, получаем

$$\sum_{j=1}^{m} \frac{\Lambda_{1,j+1}'(\theta)}{\Lambda_{11}^2(\theta)}\pi_j(\theta) = -\sum_{j=1}^{m} \frac{\Lambda_{1,j+1}(\theta)}{\Lambda_{11}^2(\theta)}\pi_j'(\theta) = -\frac{|\Lambda(\theta)|}{\Lambda_{11}(\theta)}.$$

Пример 41. Пусть $\mathcal{H} = \overline{\{\varphi_1, ..., \varphi_m\}}$, причем

$$\pi_{ij}(\theta) = (\varphi_i, \varphi_j)_\theta,\ i \doteq j; \quad \pi_{ii}(\theta) > 0,\ i = 1, 2, ..., m; \quad \pi_{11}(\theta) + ... + \pi_{mm}(\theta) = 1. \quad (3.17)$$

Условие (3.17) позволяет выбрать в качестве $\mathbb{I}$ элемент $\mathbb{I} = \varphi_1 + ... + \varphi_m$.
Тогда

$$\Lambda(\theta) = \left\| \begin{array}{cccc} 0 & \pi_1'(\theta) & ... & \pi_m'(\theta) \\ \pi_1'(\theta) & \pi_1(\theta) & ... & 0 \\ . & . & ... & . \\ \pi_1'(\theta) & 0 & ... & \pi_m(\theta) \end{array} \right\|,$$

так что

$$J(\mathcal{H}, \theta) = \sum_{j=1}^{m} \frac{\pi_j'(\theta)}{\pi_1(\theta)}\varphi_j \quad \text{и} \quad I(\theta; \mathcal{H}) = \sum_{j=1}^{m} \left(\frac{\pi_j'(\theta)}{\pi_1(\theta)}\right)^2 \pi_j(\theta)$$

в полном соответствии с формулой для фишеровской информацией, содержащейся в наблюдении, имеющем конечное число исходов с вероятностями $\pi_1(\theta), ..., \pi_m(\theta)$.

Аналог неравенства Крамера-Рао.

Для любой несмещенной оценки φ

$$\| \varphi - \theta \cdot \mathbb{I} \|^2 \geqslant \frac{1}{I(\theta, \mathcal{H})}.$$

Д о к а з а т е л ь с т в о. Пусть $\varphi = \sum\limits_{j=1}^{m} c_j\varphi_j$.

Из условия несмещенности

$$\theta = \mathbf{E}_\theta(\varphi) = (\varphi, \mathbb{I})_\theta = \sum_{j=1}^{m} c_j \pi_j(\theta) \quad \Rightarrow \quad 1 = \sum_{j=1}^{m} c_j \pi_j'(\theta).$$

Тогда

$$(\varphi - \theta \cdot \mathbb{I}, J)_\theta = \sum_{j=1}^{m} c_j \pi_j'(\theta)$$

и по неравенству Коши-Буняковского

$$1 = |\, (\varphi - \theta \cdot \mathbb{I}, J)_\theta \,|^2 \leqslant \|\, \varphi - \theta \cdot \mathbb{I} \,\|_\theta^2 \cdot \|\, J \,\|_\theta^2,$$

откуда следует неравенство Крамера-Рао.

Пример 42. Пусть $\int x^{2k} f(x;\theta)\,dx < \infty$. Положим $\alpha_j(\theta) = \int x^j f(x;\theta)\,dx,\ j = 0, 1, 2, ..., k$. Тогда матрица $\Lambda(\theta)$ принимает вид

$$\Lambda(\theta) = \left\| \begin{array}{ccccc} 0 & 0 & \alpha_1'(\theta) & ... & \alpha_k'(\theta) \\ 0 & 1 & \alpha_1(\theta) & ... & \alpha_k(\theta) \\ \alpha_1'(\theta) & \alpha_1(\theta) & \alpha_2(\theta) & ... & \alpha_{k+1}(\theta) \\ . & . & . & ... & . \\ \alpha_k'(\theta) & \alpha_k(\theta) & \alpha_{k+1}(\theta) & ... & \alpha_{2k}(\theta) \end{array} \right\|$$

и мы имеем

$$J(\mathcal{H}, \theta) = -\sum_{j=1}^{k+1} \frac{\Lambda_{1,k+1}(\theta)}{\Lambda_{11}(\theta)} x^{j-1} \quad I_k(\theta) = I(\theta; \mathcal{H}) = -\frac{|\,\Lambda(\theta)\,|}{\Lambda_{11}(\theta)}.$$

В частности для $k = 1$,

$$I_1(\theta) = \frac{(\alpha_1'(\theta))^2}{\alpha_2(\theta) - \alpha_1^2(\theta)}.$$

Имеет место следующая теорема.

ТЕОРЕМА 8. *Если выполнены указанные выше условия, то*

$$\sqrt{n}\,(\hat{\theta}_n - \theta) \xrightarrow[n \to \infty]{d} N(0, (I(\theta; \mathcal{H}))^{-1}),$$

где $\hat{\theta}_n$ *есть решение уравнения*

$$\sum_{j=1}^{m} \lambda_j(\theta) \bar{\varphi}_j = 0.$$

В случае $\mathcal{H} = \overline{\{1, x, ..., x^k\}}$ уравнение принимает вид

$$\sum_{j=1}^{k+1} \lambda_j(\theta) m_{j-1} = 0,$$

где

$$\lambda_j(\theta) = -\frac{\Lambda_{1,k+1}(\theta)}{\Lambda_{11}(\theta)}, \quad m_j = \frac{1}{n} \sum_{i=1}^{n} x_i^j.$$

Заметим также, что если $\alpha_1(\theta), ..., \alpha_k(\theta)$ трижды, а $\alpha_{k+1}(\theta), ..., \alpha_{2k}(\theta)$ дважды непрерывно дифференцируемы, то существует состоятельное решение уравнения

$$\sum_{j=1}^{k+1} \lambda_j(\theta) m_{j-1} = 0.$$

Глава 5.

Байесовские оценки.

5.1. Байесовские оценки.

Пусть $\pi(\theta) \ge 0$ измеримая функция θ, заданная на Θ и нуль вне Θ. Если $\int_\Theta \pi(\theta)\,d\theta = 1$, то $\pi(\theta)$ можно интерпретировать как плотность распределения с.в. θ. Итак, пусть $\pi(\theta)$ плотность распределения с.в. θ. Имеем

$$r(\pi, \delta) = \int_\Theta R(\theta, \delta)\,\pi(\theta)\,d\theta = \int_\Theta \int_X W(\theta - \delta(x))\,p(x;\theta)\,\pi(\theta)\,d\theta dx =$$

$$= \int_X q(x)\,dx \int_\Theta W(\theta - \delta(x))\,\pi(\theta\,|\,x)\,d\theta = \int_X q(x) R(\delta\,|\,x)\,dx\,,$$

где

$$q(x) = \int_\Theta p(x;\ \theta)\,\pi(\theta)\,d\theta.$$

Определение 11. Оценка θ_n^* называется *байесовской относительно априорного распределения* $\pi(\theta)$, если

$$r(\pi,\ \theta_n^*) = \inf_\delta r(\pi,\ \delta).$$

Оказывается, что задача минимизации безусловного риска $r(\pi,\ \delta)$ и условного риска $\int_\Theta W(\theta - \delta(x))\,\pi(\theta\,|\,x)\,d\theta$ – эквивалентны.

ТЕОРЕМА 9. (Де Грот, М.Рао, 1963) *Оценка θ_n^* является байесовской тогда и только тогда, когда*

$$\int_{\theta \ge \theta_n^*} W'(\theta - \theta_n^*)\,\pi(\theta\,|\,x)\,d\theta \ge \int_{\theta < \theta_n^*} W'(\theta - \theta_n^*)\,\pi(\theta\,|\,x)\,d\theta$$

и

$$\int_{\theta > \theta_n^*} W'(\theta - \theta_n^*)\,\pi(\theta\,|\,x)\,d\theta \le \int_{\theta \le \theta_n^*} W'(\theta - \theta_n^*)\,\pi(\theta\,|\,x)\,d\theta\,.$$

Если $W(t)$ –строго выпуклая и $W'(\mathbf{??}) = 0$, то θ_n^* единственная и

$$\int_{\theta > \theta_n^*} W'(\theta - \theta_n^*)\,\pi(\theta\,|\,x)\,d\theta = \int_{\theta \le \theta_n^*} W'(\theta - \theta_n^*)\,\pi(\theta\,|\,x)\,d\theta \quad \Rightarrow$$

$\theta_n^* = \int_\Theta \theta\,\pi(\theta\,|\,x)\,d\theta = \mathbf{E}(\theta\,|\,x)$. (46)
Д о к а з а т е л ь с т в о (без доказательства).

В случае $W(t) = t^2$ равенство (46) показывается непосредственно. Именно, пусть $\delta(x)$ – произвольная оценка с конечным риском. Тогда

$$\mathbf{E}(\delta(x) - \theta)^2 = \mathbf{E}(\delta(x) - \theta_n^*(x) + \theta_n^*(x) - \theta)^2 =$$

$$= \mathbf{E}(\delta(x) - \theta_n^*(x))^2 + \mathbf{E}(\theta_n^*(x) - \theta)^2 + 2\,\mathbf{E}((\delta(x) - \theta_n^*(x))(\theta_n^*(x) - \theta))\,.$$

Покажем, что $\mathbf{E}((\delta(x) - \theta_n^*(x))(\theta_n^*(x) - \theta)) = 0$. Действительно,

$$\mathbf{E}((\delta(x) - \theta_n^*(x))(\theta_n^*(x) - \theta)) = \mathbf{E}(\mathbf{E}((\delta(x) - \theta_n^*(x))(\theta_n^*(x) - \theta)|x) =$$

$$= \mathbf{E}((\delta(x) - \theta_n^*(x))\mathbf{E}((\theta_n^*(x) - \theta)|x))\,.$$

Но

$$\mathbf{E}((\theta_n^*(x) - \theta)|x) = \theta_n^*(x) - \mathbf{E}(\theta|x) = \theta_n^*(x) - \theta_n^*(x) = 0\,.$$

Значит,

$$\mathbf{E}(\delta(x) - \theta)^2 = \mathbf{E}(\delta(x) - \theta_n^*(x))^2 + \mathbf{E}(\theta_n^*(x) - \theta)^2 \ge \mathbf{E}(\theta_n^*(x) - \theta)^2\,,$$

откуда следует утверждение.

Пример 43. Рассмотрим задачу оценки параметра θ в биномиальной схеме, т.е.

$$p\,(x\,;\ \theta) = C_n^x \theta^x (1-\theta)^{n-x}\,,\ \ x = 0\,,\ 1\,,\ \dots\,,\ n,$$

априорное распределение задано плотностью бета-распределения

$$\pi\,(\theta) = \frac{\theta^{p-1}(1-\theta)^{q-1}}{\beta\,(p\,,q)}\,,\ \ 0 < \theta < 1\,;\ p > 0\,,\ q > 0.$$

Тогда совместное распределение X и θ задается выражением

$$p\,(x\,,\ \theta)\,\pi\,(\theta) = C_n^x \theta^x (1-\theta)^{n-x} \frac{\theta^{\alpha-1}(1-\theta)^{\gamma-1}}{\beta\,(\alpha\,,\gamma)} = C_n^x \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\gamma-1}}{\beta\,(\alpha\,,\gamma)}.$$

Отсюда маргинальное распределение X равно

$$q(x) = \int_0^1 C_n^x \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\gamma-1}}{\beta\,(\alpha\,,\gamma)}\,d\theta = \ C_n^x \frac{\beta\,(x+\alpha\,, n-x+\gamma)}{\beta\,(\alpha\,,\gamma)},$$

а апостериорное распределение величины θ задается плотностью

$$\pi\,(\theta\,|\,x) = \frac{p\,(x\,;\ \theta)\,\pi\,(\theta)}{q\,(x)} = \frac{\theta^{x+\alpha-1}(1-\theta)^{n-x+\gamma-1}}{\beta\,(x+\alpha\,, n-x+\gamma)},$$

т.е. имеет бета-распределение, но уже с другими параметрами.

Апостериорное среднее равно

$$\mathbf{E}\,(\theta\,|\,x) = \int_0^1 \theta\pi(\theta|x)\,d\theta = \int_0^1 \frac{\theta^{x+\alpha}(1-\theta)^{n-x+\gamma-1}}{\beta\,(x+\alpha\,, n-x+\gamma)}\,d\theta =$$

$$= \frac{\beta\,(x+\alpha\ +1, n-x+\gamma)}{\beta\,(x+\alpha\,, n-x+\gamma)} = \frac{x+\alpha}{n+\alpha+\gamma} = \theta_n^*.$$

Формально, при $\alpha = \gamma = 0$ эта оценка равна $\theta_n^* = x/n$.

5.2. Эмпирический байесовский подход.

Недостатком байесовского подхода является необходимость постулировать как существование априорного распределения для неизвестного параметра, так и знание его формы. Заметим, что асимптотическое распределение байесовских оценок не зависит от вида априорной плотности. В определенной степени последнее обстоятельство преодолевается в рамках эмпирического байесовского подхода.

По теореме Байеса для заданной функции $\phi\,(\theta)$ условное математическое ожидание $\mathbf{E}\,(\,\phi(\theta)\,|x)$ выражается отношением

$$\psi(x) = \frac{\int \phi(\theta)\,p(x\,;\ \theta)\,\pi\,(\theta)\,d\theta}{q(x)},$$

где

$q(x) = \int p\,(x\,;\ \theta)\,\pi\,(\theta)\,d\theta$ – плотность безусловного распределения x.

Бернштейн (1941) предложил заменить $q\,(x)$ ее оценкой $\hat{q}\,(x)$ и найти решение интегрального уравнения

$$\hat{q}\,(x) = \int p\,(x\,;\ \theta)\,\hat{\pi}\,(\theta)\,d\theta,$$

а затем подставить $\hat{\pi}\,(\theta)$ в формулу для $\psi\,(x)$. Однако этот путь затруднителен, так как решение указанного уравнения принадлежит к числу некорректно поставленных задач вычислительной математики.

Г.Роббинс (1956) предложил оценку не только $q\,(x)$, но и $\psi\,(x)$ в следующей ситуации.

Пусть справедливо тождество

(7) (8) $$\phi(\theta)\,p(x;\theta) = \lambda(x)\,r(z(x);\,\theta)\,,$$

где $\lambda(x)$ и $z(x)$ – функции, зависящие только от x, а $r(z;\,\theta)$ как функция z является плотностью распределения. Если (8) справедливо, то

$$\psi(x) = \frac{\lambda(x)\,s(z(x))}{q(x)}, \quad \text{где} \quad s(z(x)) = \int r(z(x);\,\theta)\,\pi(\theta)\,d\theta,$$

и поэтому $\hat{\psi}(x) = \frac{\lambda(x)\,\hat{s}(z(x))}{\tilde{q}(x)}$, $\hat{q}(x)$ и $\tilde{s}(z(x))$ - оценки плотностей по наблюдениям x и $z(x)$. Если x и $z = z(x)$ - дискретны, то

$$\hat{q}(x) = \frac{\text{число прошлых наблюдений, равных } x}{n},$$

$$\hat{s}(z) = \frac{\text{число прошлых наблюдений, равных } z}{n}.$$

Пример 44. Пусть $\varphi(\theta) = \theta^h$ (h – целое > 0), причем

$$p(x;\,\theta) = \frac{\theta^x}{x!}e^{-\theta} \quad (\,x = 0,\,1,\,...\,;\,\theta > 0\,),$$

тогда

$$\phi(\theta)\,p(x;\,\theta) = \frac{\theta^{x+h}}{x!}e^{-\theta} = \frac{(x+h)\,!}{x\,!}\cdot\frac{\theta^{x+h}}{(x+h)!}e^{-\theta}$$

и поэтому

$$\lambda(x) = \frac{(x+h)\,!}{x\,!}\,,\; r(x+h;\,\theta) = \frac{\theta^{x+h}}{(x+h)!}e^{-\theta}\,,\; z(x) = x+h\,.$$

Так как
$r(z;\,\theta) = p(z;\,\theta)$, то $s(z) = q(z)$ и поэтому

$$\hat{\psi}(x) = \frac{\lambda(x)\,\hat{q}(x+h)}{\hat{q}(x)}\,,\; \hat{q}(x) = \frac{\text{число прошлых наблюдений, равных } x}{n}.$$

Пример 45. Пусть

$$p(x;\,\theta) = C_n^x\theta^x(1-\theta)^{n-x}\,,\; x = 0\,,\,1\,,\,...\,,\,n\,;\,n \in N\,,\;0 \le \theta \le 1\,.$$

Тогда (здесь $\phi(\theta) = \theta^h$)

$$\phi(\theta)\,p(x;\,\theta) = C_n^x\theta^{x+h}(1-\theta)^{n-x} = \frac{C_n^x}{C_{n+h}^{x+h}}C_{n+h}^{x+h}\theta^{x+h}(1-\theta)^{n-x},$$

$$n - x = (n+h) - (x+h).$$

Поэтому

$$\lambda(x) = \frac{C_n^x}{C_{n+h}^{x+h}}\,,\quad r(z;\,\theta) = p(z;\,\theta)\,,\quad z = x+h\,,$$

и при построении $\hat{\psi}(x)$ нужно, как и предыдущем случае иметь две последовательности эмпирических значений x_i и z_i.

Указанная методика э.б.п. применима к весьма узкому классу плотностей, удовлетворяющих условию (8).

5.3. Минимаксные оценки.

Пусть $W(\theta,\,d)$ – заданная функция потерь, где $\theta \in \Theta$, $d = d(x)$, $d:\; X \to \Theta$ – оценка параметра θ. Обозначим через $R(\theta,\,d)$ соответствующую функцию риска, т.е.

$$R(\theta,\,d) = \int W(\theta,\,d(x))\,dF_\theta(x).$$

Без потери общности можно предположить, что $R(\theta,\ d) \leq M < \infty$ для каждых $(\theta,\ d)$.

Минимаксная оценка определяется на множестве возможных оценок D так, что

$$\sup_{\theta\in\Theta} R(\theta,\ d^*) = \inf_{d\in D}\sup_{\theta\in\Theta} R(\theta,\ d).$$

Покажем сначала, что

$$\sup_{\theta\in\Theta}\inf_{d\in D} R(\theta,\ d) \leq \inf_{d\in D}\sup_{\theta\in\Theta} R(\theta,\ d).$$

Действительно,

$$R(\theta,\ d) \leq \sup_{\theta\in\Theta} R(\theta,\ d) \quad\Rightarrow\quad \inf_{d\in D} R(\theta,\ d) \leq \inf_{d\in D}\sup_{\theta\in\Theta} R(\theta,\ d) \quad\Rightarrow$$

$$\Rightarrow \sup_{\theta\in\Theta}\ \inf_{d\in D} R(\theta,\ d) \leq \inf_{d\in D}\sup_{\theta\in\Theta} R(\theta,\ d).$$

Пусть H – класс всех априорных распределений на Θ. Для каждого $d \in D$ и любого ξ из H,

$$R(\xi, d) = \int R(\theta,\ d)\, d\xi(\theta) \leq \sup_{\theta\in\Theta} R(\theta,\ d).$$

Следовательно,

$$\sup_{\xi\in H} R(\xi,\ d) \leqslant \sup_{\theta\in\Theta} R(\theta,\ d).$$

Поскольку

$$\sup_{\theta\in\Theta} R(\theta,\ d) \leq \sup_{\xi\in H} R(\xi,\ d),$$

то

$$\sup_{\theta\in\Theta} R(\theta,\ d) = \sup_{\xi\in H} R(\xi,\ d).$$

ТЕОРЕМА 10. *Пусть $P = \{F_\theta,\ \theta \in \Theta\}$ –семейство функций распределения, а D – класс оценок параметра θ и пусть оценка $d^* \in D$ является байесовской при априорном распределении $\xi^*(\theta)$ на Θ, а риск $R(\theta, d^*)$ постоянен на Θ. Тогда d^* – минимаксная оценка параметра θ.*

Д о к а з а т е л ь с т в о. Поскольку $R(\theta, d^*) = \rho^*$ для всех θ и d^* – байесовская оценка для ξ^*, то имеем

$$\rho^* = \int R(\theta, d^*)\, d\xi^*(\theta) = \inf_{d\in D}\int R(\theta, d)\, d\xi^*(\theta).$$

Кроме того,

$$\rho^* = \inf_{d\in D}\int R(\theta, d)\, d\xi^*(\theta) \leq \sup_{\xi\in H}\ \inf_{d\in D}\int R(\theta, d)\, d\xi(\theta) \leq$$

$$\leq \inf_{d\in D}\sup_{\xi\in H}\int R(\theta, d)\, d\xi(\theta) \leq \inf_{d\in D}\sup_{\theta\in\Theta} R(\theta, d).$$

С другой стороны, поскольку $R(\theta, d^*) = \rho^*$ для всех $\theta \in \Theta$, то

$$\rho^* = \sup_{\theta\in\Theta} R(\theta, d^*) \geq \inf_{d\in D}\ \sup_{\theta\in\Theta} R(\theta, d).$$

Таким образом,

$$\sup_{\theta\in\Theta} R(\theta, d^*) = \inf_{d\in D}\ \sup_{\theta\in\Theta} R(\theta, d),$$

т.е. d^* – минимаксная оценка.

Пример 46. (Ходжес, Леман) Пусть X – случайная величина, имеющая биномиальное распределение $B(n\,;\,\theta)$, $0<\theta<1$; θ неизвестно. Покажем, что

$$d^*(X)=\frac{X}{n}\cdot\frac{\sqrt{n}}{1+\sqrt{n}}+\frac{1}{2\,(1+\sqrt{n})}$$

является минимаксной оценкой параметра θ при квадратичной функции потерь. Чтобы показать это, рассмотрим линейную оценку

$$d(X)=\alpha\,\frac{X}{n}+\beta\,,\ \alpha\neq 0.$$

Функция риска этой оценки имеет вид

$$R\,(\theta\,,\ d)=\mathbf{E}_\theta\left(\left[\alpha\,\frac{X}{n}+\beta-\theta\right]^2\right)=$$

$$=\beta^2+\frac{\theta}{n}\left[1-2(1-\alpha)+(1-\alpha)^2-2n\beta(1-\alpha)\right]-$$

$$-\frac{\theta^2}{n}\left[1-2(1-\alpha)+(1-\alpha)^2(1-n)\right].$$

Выберем α, β так, чтобы риск $R\,(\theta\,,\ d)$ был постоянен для всех $0<\theta<1$. Это будет, если коэффициенты при θ и θ^2 будут равны нулю. Значит, необходимо решить систему уравнений (заметим, что риск для этих значений α, β будет равен β^2):

$$\left\{\begin{array}{ll} 1-2(1-\alpha)+(1-\alpha)^2-2n\beta(1-\alpha) & =0\,,\\ 1-2(1-\alpha)+(1-\alpha)^2(1-n) & =0\,. \end{array}\right.$$

Решением этой системы являются значения

$$\alpha^*=\frac{\sqrt{n}}{1+\sqrt{n}}\,,\quad \beta^*=\frac{1}{2\,(1+\sqrt{n})}\,.$$

В таком случае,
$R\,(\theta\,,\ d^*)=(\beta^*)^2=\dfrac{1}{4(1+\sqrt{n})^2}$ для всех $0\le\theta\le 1$.

Покажем теперь, что d^* – байесовская оценка. Если априорным распределением параметра θ является бета-распределение $\beta\,(p\,,\ q)$, $0<p,q<\infty$, то, как мы уже показывали в теме «Байесовские оценки», апостериорным распределением параметра θ при заданном X будет $\beta\,(X+p\,,n-X+\ q)$. Следовательно, апостериорное ожидаемое значение параметра θ равно

$$\mathbf{E}_\xi(\,\theta\,|\,X\,)=\frac{X+p}{n+p+q}.$$

Если взять $p=q=\sqrt{n/2}$, байесовская оценка имеет вид

$$\frac{X+\sqrt{n/2}}{n+\sqrt{n/2}}=\frac{X}{n}\cdot\frac{\sqrt{n}}{1+\sqrt{n}}+\frac{1}{2\,(1+\sqrt{n})}=d^*(X).$$

Сравним риск этой оценки с риском оценки МП, равной $\widehat{\theta}=\dfrac{X}{n}$, которая является оценкой с минимальной дисперсией. Имеем:

$$E_\theta(\widehat{\theta}-\theta)^2=\frac{\theta\,(1-\theta)}{n}.$$

В таком случае среднеквадратичная ошибка оценки d^* меньше, чем у $\widehat{\theta} = \dfrac{X}{n}$, если $\left|\theta - \dfrac{1}{2}\right| < \dfrac{\sqrt{1+2\sqrt{n}}}{2(1+\sqrt{n})}$.

Например, для $n = 20$ правая часть равна 0.288, а при $n = 50$ она равна 0.241.
При $n = 1000$ правая часть равна 0.123.

5.4. Оценки Питмена.

Так как в классе всех оценок нельзя найти абсолютно наилучшую, то приходится ограничиваться классом каких-либо оценок.

Пусть $x_1, x_2, \ldots, x_n$ – повторная выборка с общей плотностью распределения $f(x - \theta)$, $-\infty < \theta < \infty$, $\mathbf{E}_\theta(X^2) < \infty$, $\mathbf{E}_\theta(X) = \theta$, $\theta \in R^1$.

Определение 12. Оценку $\tilde{\theta}_n(x_1, x_2, \ldots, x_n)$ параметра θ, удовлетворяющую при всех $c \in R^1$ условию

$$\tilde{\theta}_n(x_1 + c, x_2 + c, \ldots, x_n + c) = \tilde{\theta}_n(x_1, x_2, \ldots, x_n) + c$$

будем называть *эквивариантной* (или *правильной*) *оценкой.*

Пример 47. Пусть $\tilde{\theta}_n(x_1, x_2, \ldots, x_n) = \dfrac{x_1 + x_2 + \ldots + x_n}{n} = \bar{x}$.

Задание. Покажите, что $\bar{x}$ – правильная оценка.

Класс правильных оценок обозначим через R. Если $\tilde{\theta}'_n$ и $\tilde{\theta}''_n$ – две правильные оценки, то $\tilde{\theta}'_n - \tilde{\theta}''_n = \psi(x_2 - x_1, \ldots, x_n - x_1)$.

Если $\tilde{\theta}_n \in R$, то

$$\mathbf{E}_\theta(\tilde{\theta}_n - \theta)^2 = \mathbf{E}_0(\tilde{\theta}_n^2) = const.$$

Положим $y = (x_2 - x_1, \ldots, x_n - x_1)$ и рассмотрим оценку

$$t_n = \bar{x} - \mathbf{E}_0(\bar{x} \,|\, y).$$

Заметим, что

$$t_n = \bar{x} - \mathbf{E}_0(\bar{x} \,|\, y) = \bar{x} - \mathbf{E}_0(x_1 + \frac{(x_2 - x_1) + \ldots + (x_n - x_1)}{n} \,|\, y) =$$

$$= \bar{x} - \mathbf{E}_0(x_1|y) + \frac{(x_2 - x_1) + \ldots + (x_n - x_1)}{n} = x_1 - \mathbf{E}_0(x_1|y).$$

ТЕОРЕМА 11. *Оценка t_n оптимальна в классе R, т.е.*

1) $\mathbf{E}_\theta(t_n - \theta)^2 = \min\limits_{\tilde{\theta}_n \in R} E_0(\tilde{\theta}_n)^2$ *и*

2) *если $f(x)$ – плотность распределения по мере Лебега, то*

$$t_n = \frac{\int_{-\infty}^{\infty} \xi \prod_{i=1}^n f(x_i - \xi)\, d\xi}{\int_{-\infty}^{\infty} \prod_{i=1}^n f(x_i - \xi)\, d\xi}.$$

Доказательство. 1) Если $\mathbf{E}_0(\tilde{\theta}_n^2) = \infty$ для любой оценки, то теорема очевидна, поэтому рассмотрим случай, когда найдется оценка $\mathbf{E}_0(\tilde{\theta}_n^2) < \infty$. Имеем:

$$\mathbf{E}_\theta(\tilde{\theta}_n - \theta)^2 = \mathbf{E}_\theta(\tilde{\theta}_n - t_n + t_n - \theta)^2 = \mathbf{E}_\theta(t_n - \theta)^2 + \mathbf{E}_\theta(\tilde{\theta}_n - t_n)^2 +$$

$$+ 2\mathbf{E}_\theta((\tilde{\theta}_n - t_n)(t_n - \theta)).$$

Но

$$\mathbf{E}_\theta((\tilde{\theta}_n - t_n)(t_n - \theta)) = \mathbf{E}_0((\tilde{\theta}_n - t_n)\, t_n) = \mathbf{E}_0(\mathbf{E}_0((\tilde{\theta}_n - t_n)\, t_n|y)) =$$

$$= \mathbf{E}_0((\tilde{\theta}_n - t_n)\, \mathbf{E}_0(t_n|y)) = 0.$$

Перейдем к переменным

$$\left\{\begin{array}{l} x_1 \qquad\;\; = u_1\,, \\ x_2 - x_1 = u_2\,, \\ \quad .\;\;.\;\;. \\ x_n - x_1 = u_n\,, \end{array}\right. \qquad \Rightarrow \qquad \left\{\begin{array}{l} x_1 \;= u_1\,, \\ x_2 = u_2 + u_1\,, \\ \quad .\;\;.\;\;. \\ x_n = u_n + u_1\,. \end{array}\right.$$

Якобиан преобразования равен 1 и поэтому

$$p\,(u_1\,,\; u_2\,,\; \dots\,,\; u_n) = f(u_1)\,f(u_1+u_2)\;\dots f(u_1+u_n)\,.$$

Отсюда

$$\mathbf{E}_0(u_1\,|\,u_2\,,\;\dots\,,\;u_n) = \frac{\int u_1\,f(u_1)\,f(u_1+u_2)\;\dots f(u_1+u_n)\,du_1}{\int f(u_1)\,f(u_1+u_2)\;\dots f(u_1+u_n)\,du_1}.$$

Возвращаясь к старым переменным и обозначая $\xi = -\,u_1 + x_1 \;\Rightarrow\; u_1 = x_1 - \xi$, будем иметь

$$\mathbf{E}_0(x_1\,|\,x_2 - x_1\,,\;\dots\,,\;x_n - x_1) = \frac{\int u_1\,f(u_1)\,f(u_1 + x_2 - x_1)\;\dots f(u_1 + x_n - x_1)\,du_1}{\int\; f(u_1)\,f(u_1 + x_2 - x_1)\;\dots f(u_1 + x_n - x_1)\,du_1} =$$

$$= \frac{\int_{-\infty}^{\infty}(x_1 - \xi)\,\prod_{i=1}^n f(x_i - \xi)\,d\xi}{\int_{-\infty}^{\infty}\prod_{i=1}^n f(x_i - \xi)\,d\xi} = x_1 - \frac{\int_{-\infty}^{\infty}\xi\,\prod_{i=1}^n f(x_i - \xi)\,d\xi}{\int_{-\infty}^{\infty}\prod_{i=1}^n f(x_i - \xi)\,d\xi},$$

Откуда следует результат части 2 теоремы.

Пример 48. Пусть с.в. X имеет плотность

$$f(x-\theta) = \left\{\begin{array}{l} \exp\left(-(x-\theta)\right),\;\; \theta < x < \infty\,, \\ 0\,,\; \theta \geq x\,, \qquad \theta \in R^1\,. \end{array}\right.$$

Найти оценку Питмена для параметра θ по выборке $x_1\,,\; x_2\,,\; \dots\,,\; x_n$.

Решение. Плотность не равна нулю, если $0 < x_i - \xi < \infty$ для всех $i = 1\,,\; 2\,,\; \dots\,,\; n$, откуда $-\infty < \xi < x_n^{(1)} = \min\limits_{1 \leq\, i\, \leq n} x_i$. Значит,

$$t_n = \frac{\int_{-\infty}^{x_n^{(1)}}\xi\,\prod_1^n e^{-\,(x_i - \xi)}\,d\xi}{\int_{-\infty}^{x_n^{(1)}}\prod_1^n e^{-\,(x_i - \xi)}\,d\xi} = \frac{\int_{-\infty}^{x_n^{(1)}}\xi\,e^{n\xi}\,d\xi}{\int_{-\infty}^{x_n^{(1)}}e^{n\xi}\,d\xi} = x_n^{(1)} - \frac{1}{n}\,, \quad \mathbf{E}_\theta(t_n - \theta)^2 = \frac{1}{n^2}.$$

Задача 1. Построить оценку Питмена для параметра сдвига θ нормального распределения $N(\theta\,,\;\sigma^2)$. (σ^2 может быть как известным, так и неизвестным).

Ответ: $t_n = \bar{x}$.

Задача 2. Построить оценку Питмена для параметра сдвига θ равномерного распределения на интервале $(\theta - \frac{1}{2}\,,\;\theta + \frac{1}{2})$.

Ответ: $t_n = (x_n^{(1)} + x_n^{(n)})/n$.

Глава 6.

Порядковые статистики.

6.1. Порядковые статистики. Распределение порядковых статистик. Ранги и антиранги. Обобщение представления Реньи для показательных неодинаково распределенных величин. Сокращение времени испытания за счет цензурирования.

Пусть $o_i(x)$ есть значение $i-$й по величине координаты вектора $x = (x_1\,,\ x_2\,,\ \ldots\,,\ x_n)$. Функция $o_i(x)$ определена и в том случае, когда некоторые значения совпадают. Полагая $x_n^{(i)} = o_i(x_1\,,\ x_2\,,\ \ldots\,,\ x_n)$, мы имеем $x_n^{(1)} \le x_n^{(2)} \le\ \ldots\ \le x_n^{(n)}$.

Статистика $X_n^{(i)} = o_i(X) = o_i(X_1\,,\ X_2\,,\ \ldots\,,\ X_n)$ будет называться ***$i-$й порядковой статистикой***, а вектор порядковых статистик $(X_n^{(1)}, X_n^{(2)},\ \ldots\,, X_n^{(n)})$ кратко обозначаться $X_n^{(\bullet)}$.

Далее, для вектора $x = (x_1\,,\ x_2\,,\ \ldots\,,\ x_n)$, у которого никакие две координаты не совпадают, пусть $r_i = r_i(x)$ – число координат, не превосходящих x_i, т.е. номер x_i в последовательности

$$x_i = x_n^{(r_i)}\,, \quad 1 \le i \le n\,.$$

Статистику $R_i = r_i(X)$ будем называть ***рангом*** X_i и R будет обозначать вектор рангов $(\,R_1\,,\ R_2\,,\ \ldots\,,\ R_n)$. Определение рангов корректно, только если вероятность совпадения для любой пары равна 0. Это условие, например, выполняется, когда распределение задано плотностью.

Пусть $(d_1\,,\ \ldots\,,\ d_n)$ – обратная перестановка по отношению к $(r_1\,,\ \ldots\,,\ r_n)$, т.е.

$$r_{d_i} = d_{r_i} = 1\,,\ \ 1 \le i \le n\,.$$

Пусть $D = (D_1\,,\ \ldots\,,\ D_n)$ обратна в этом смысле к $R = (\,R_1\,,\ R_2\,,\ \ldots\,,\ R_n)$. Соответствие $(d_1\,,\ \ldots\,,\ d_n)\ \to (r_1\,,\ \ldots\,,\ r_n)$ взаимно однозначно.

ТЕОРЕМА 12. *Если X распределена с плотностью q, то $X_N^{(\bullet)}$ имеет распределение с плотностью*

$$\bar q\,(x_N^{(1)},\ x_N^{(2)},\ \ldots\,,\ x_N^{(N)}) = \sum_{r\,\in R} q\,(x_N^{(r_1)},\ x_N^{(r_2)},\ \ldots\,,\ x_N^{(r_N)})\,,\ x_N^{(\bullet)} \in X_N^{(\bullet)}. \tag{8}$$

Кроме того,

$$Q\,(R = r\,|\,X_N^{(\bullet)} = x_N^{(\bullet)}) = \frac{q\,(x_N^{(r_1)},\ x_N^{(r_2)},\ \ldots\,,\ x_N^{(r_N)})}{\bar q\,(x_N^{(1)},\ x_N^{(2)},\ \ldots\,,\ x_N^{(N)})}\,,\ \ r \in R\,,\ \ x_N^{(\bullet)} \in X_N^{(\bullet)}\,, \tag{9}$$

где Q – распределение вероятностей, отвечающее плотности q.

Д о к а з а т е л ь с т в о. Для любого $A \in A^{(\bullet)}$ выполняется соотношение

$$\int_{X_N^{(\bullet)} \in A} \ldots \int q\,(x_1\,,\ \ldots\,,\ x_N)\,dx_1 \ldots dx_N =$$

$$= \sum_{r\,\in R} \int_{R=r} \ldots \int_{X_N^{(\bullet)} \in A} q\,(x_1\,,\ \ldots\,,\ x_N)\,dx_1 \ldots dx_N =$$

$$= \sum_{r\,\in R} \int \ldots \int_A q\,(x_N^{(r_1)}\,,\ \ldots\,,\ x_N^{(r_N)})\,dx_N^{(1)} \ldots dx_N^{(N)} =$$

$$= \int .. \int_A \bar{q}\,(x_N^{(1)}\,,\ \dots\,,\ x_N^{(N)})\,dx_N^{(1)} \dots dx_N^{(N)}, \tag{10}$$

при выводе которого мы воспользовались тем фактом, что соответствие между $(x_1\,,\ \dots\,,\ x_N)$ и $(x_N^{(1)}\,,\ \dots\,,\ x_N^{(N)})$ взаимно однозначно и линейно с якобианом, равным 1, на каждом подмножестве $(\,R = r\,)$. Очевидно, (48) доказывает (46). Рассуждая таким же образом, имеем

$$Q\,(R = r\,,\, X_N^{(\bullet)} \in A) = \int_{R=r} .. \int_{X_N^{(\bullet)} \in A} q\,(x_1\,,\ \dots\,,\ x_N)\,dx_1 \dots dx_N =$$

$$= \int .. \int_A q\,(x_N^{(r_1)}\,,\ \dots\,,\ x_N^{(r_N)})\,dx_N^{(1)} \dots dx_N^{(N)} =$$

$$= \int .. \int_A \frac{q\,(x_N^{(r_1)}\,,\ \dots\,,\ x_N^{(r_N)})}{\bar{q}\,(x_N^{(1)}\,,\ \dots\,,\ x_N^{(N)})}\ \bar{q}\,(x_N^{(1)}\,,\ \dots\,,\ x_N^{(N)})\,dx_N^{(1)} \dots dx_N^{(N)}\,, \tag{11}$$

где мы использовали тот факт, что из $\bar{q}\,(x_N^{(1)}\,,\ \dots\,,\ x_N^{(N)}) = 0$ следует $q\,(x_N^{(r_1)}\,,\ \dots\,,\ x_N^{(r_N)}) = 0$ для любого $r \in R$. Ясно, что (11) доказывает (47).

Из теоремы 1 непосредственно получается следующая теорема.

ТЕОРЕМА 13. *Если X распределение с плотностью $f(x)$ для повторной выборки, то случайные векторы $R = (R_1\,,\ \dots\,,\ R_N)$ и $X_N^{(\bullet)} = (X_N^{(1)},\ \dots\,, X_N^{(N)})$ независимы,*

$$P\,(\,R = r) = \frac{1}{N!}\,,\ r \in R\,, \tag{12}$$

и плотность распределения $X_N^{(\bullet)}$ равна

$$N!\,f(x_1)\,\cdot f(x_2)\,\cdot\ \dots\ \cdot f(x_N)\,, \quad x_1 < x_2 <\ \dots\ < x_N\,. \tag{13}$$

ТЕОРЕМА 14. *Пусть $X_N^{(1)} <\ \dots\ < X_N^{(N)}$ – порядковые статистики из повторной выборки с плотностью $f(x)$. Тогда плотность распределения $X_N^{(i)}$ существует и равна*

$$f_N^{(i)}(x) = \frac{N!}{(i-1)!(N-i)!}(F(x))^i(1-F(x))^{N-i}f(x) \quad -\infty < x < \infty, 1 \leqslant i \leqslant N,$$

где $F(x) = \int_{-\infty}^{x} f(y)\,dy\,.$

Д о к а з а т е л ь с т в о. Из (49) получаем совместную плотность, производя необходимые интегрирования, получим маргинальную плотность распределения

$$f_N^{(i)}(x_i) = N! \int_{-\infty}^{x_2} .. \int_{-\infty}^{x_i} \int_{x_i}^{+\infty} .. \int_{x_{N-1}}^{+\infty} \prod_{j\,=\,1}^{N} f(x_j)\,dx_j =$$

$$= \frac{N!}{(i-1)!(N-i)!} F^i(x_i)(1-F(x_i))^{N-i}\,f(x_i)\,.$$

ТЕОРЕМА 15. *Пусть $U_N^{(1)} <\ \dots\ < U_N^{(N)}$ – порядковые статистики из повторной выборки с равномерной плотностью распределения на интервале $(\,0\,,\ 1)$. Тогда плотность распределения $U_N^{(i)}$ существует и равна*

$$f_N^{(i)}(x) = N \cdot C_{N-1}^{i-1} x^i (1-x)^{N-i}\ ,\ 0 < x < 1\,,\ 1 \le i \le N\,. \tag{14}$$

Кроме того,

$$\mathbf{E}(U_N^{(i)}) = \frac{i}{N+1}\,, \tag{15}$$

$$\mathbf{D}(U_N^{(i)}) = \frac{i\,(N-i+1)}{(N+1)^2(N+2)}\,, \tag{16}$$

$$\mathbf{Cov}(U_n^{(i)}\,,\;U_n^{(j)}) = \frac{i}{n+1}\left(1-\frac{j+1}{n+2}\right)\,,\;1\le i<j\le n\,.$$

Д о к а з а т е л ь с т в о. Равенства следуют из определения плотности и из соотношений

$$\int_0^1 x^a(1-x)^b\,dx = \frac{a!b!}{(a+b+1)!}\,,\;\;a,b\in N,$$

$$\mathbf{E}((U_n^{(i)})^a(U_n^{(j)})^b) = \frac{(i-1+a)!(j=1+b+a)!n!}{(i-1)!(j-1+a)!(n+a+b)!}\,,\;i<j\,.$$

Ввиду важности формулы (8) выведем ее несколькими способами.

Пусть

$G_{k,n}(x)=P(X_n^{(k)}<x)$ и $S_n(x)=\#\,\{X_i<x\}\,,\;i=1,\;\dots\,,\;n\,,$

где символ $\#$ означает число элементов из n, для которых выполнено условие $(X_i<x)$.

Тогда

$$P\,(X_n^{(k)}<x) = P\,(S_n(x)\ge k)$$

и

$$P\,(S_n(x)=m) = C_n^m F^m(x)(1-F(x))^{n-m}.$$

Отсюда

$$G_{k,n}(x) = \sum_{m=k}^{n}\frac{n!}{m!\,(n-m)!}\,F^m(x)(1-F(x))^{n-m} =$$

$$= \frac{n!}{(k-1)!\,(n-k)!}\,\int_0^{F(x)} x^{k-1}(1-x)^{n-k}\,dx.$$

Действительно,

$$\frac{1}{(k-1)!\,(n-k)!}\,\int_0^{F(x)} x^{k-1}(1-x)^{n-k}\,dx =$$

$$= \frac{1}{k!\,(n-k)!}\,x^k(1-x)^{n-k}\,\Big|_0^{F(x)} - \frac{1}{k!\,(n-k)!}\,\int_0^{F(x)} x^k(1-x)^{n-k}\,dx\,,$$

и т.д.

Беря производную, получим формулу (8).

Элементарные методы нахождения распределения порядковых статистик состоят в следующем.

1) $F_n^{(n)}(x)=P(X_n^{(n)}<x)=P\left(\bigcap_{i=1}^n(X_i<x)\right)=\prod_{i=1}^n P\,(X_i<x)=F^n(x)$.

2) $F_n^{(1)}(x)=P(X_n^{(1)}<x)=1-P(X_n^{(1)}\ge x)=$

$$= 1-P\left(\bigcap_{i=1}^{n}(X_i\ge x)\right) = 1-\prod_{i=1}^{n}P\,(X_i\ge x) = 1-(1-F(x))^n. \tag{17}$$

Событие $x<X_n^{(k)}<x+dx$ может быть реализовано следующим образом:

$$\begin{array}{ccc} k-1 & 1 & n-k \\ & & \\ & x & x+dx \end{array}$$

$X_i < x$ для $k-1$ величин X_i, $x \le X_i < x + dx$ для одной из X_i и $X_i \ge x + dx$ для остальных $n-k$ величин X_i. Число способов, на которые n наблюдений можно разбить на три такие группы, равно

$$\frac{n!}{(k-1)!\,(n-k)!},$$

и каждый из них имеет вероятность

$$F^{k-1}(x)(1 - F(x+dx))^{n-k} f(x)\,dx,$$

поэтому получим

$$f_n^{(k)}(x)\,dx \approx \frac{n!}{(k-1)!\,(n-k)!} F^{k-1}(x)(1-F(x))^{n-k} f(x)\,dx.$$

Поскольку dx бесконечно мало, имеем точное равенство.

Совместную плотность порядковых статистик $X_n^{(k)}$ и $X_n^{(s)}$ $(1 \le k < s \le n)$ обозначим $f_n^{k,s}(x,y)$, где мы будем рассматривать случай $x < y$, поскольку для $x \ge y$ плотность $f_n^{k,s}(x,y)$ равна нулю. Итак:

$k-1$	1	$s-k$	1	$n-s$
	x	$x+dx$	y	$y+dy$

Рассуждая как и выше, получаем плотность распределения пары порядковых статистик $X_n^{(k)}$ и $X_n^{(s)}$:

$$f_n^{k,s}(x,y) = \frac{n!}{(k-1)!\,(s-k-1)!\,(n-s)!}\cdot$$

$$F^{k-1}(x)(F(y)-F(x))^{s-k-1}(1-F(y))^{n-k} f(x) f(y)\,,\ x < y\,. \tag{18}$$

Совместная плотность распределения величин $(X_n^{(k_1)},\ X_n^{(k_2)},\ \dots\ ,\ X_n^{(k_m)})$, $(1 \le k_1 < k_2 <\ \dots\ <\ k_m \le n)$, $(1 \le m \le n)$, для $x_1 < x_2 <\ \dots\ < x_m$ имеет вид

$$f_n^{k_1,k_2,\dots,k_m}(x_1\,,\ x_2\,,\ \dots\ ,\ x_m) = \frac{n!}{(k_1-1)!\,(k_2-k_1-1)!\cdot\ \dots\ \cdot(n-k_m)!}\cdot$$

$$\cdot F^{k_1-1}(x_1)(F(x_2)-F(x_1))^{k_2-k_1-1}\cdot\ \dots\ \cdot(1-F(x_m))^{n-k_m}\cdot$$

$$\cdot f(x_1)\cdot f(x_2)\cdot\ \dots\ \cdot f(x_m)\,. \tag{19}$$

При испытаниях изделий на продолжительность эксплуатации в случаях высоконадежных изделий такие испытания могут продолжать достаточно долго. В теории надежности для сокращения времени испытаний иногда применяют метод форсированных испытаний. Но в этом случае встает проблема пересчета времени форсированных испытаний во времена испытаний в нормальных условиях. Другой способ, который можно применить – это использование цензурированных выборок. Рассмотрим случай, когда испытания независимы, а закон распределения наработки до отказа $i-$го изделия показательный с плотностью распределения

$f_i(x) = \exp\left(-\lambda_i x\right)\lambda_i,\ x > 0\,;\quad \lambda_i = 1/(\beta_i\theta)\,;\ \theta\,,\ \beta_i > 0\,,\ 1 \le i \le n$, (28)
где θ – неизвестный параметр, а β_i –известные коэффициенты. Параметры β_i могут трактоваться как коэффициенты форсирования испытаний по сравнению с обычными условиями эксплуатации. Задача состоит в оценке параметра θ по наблюдениям за отказом k первых отказавших изделий, поставленных на испытания.

В случае, когда $x_1\,,\ x_2\,,\ \dots\,,\ x_n$ – независимые и одинаково распределенные по показательному закону случайные величины ($\beta_i = 1\,,\ i = 1\,,\ 2\,,\ \dots\,,\ n$), наилучшей в определенном смысле оценкой для параметра θ является статистика

$$t_n^{(k)} = \frac{1}{k} \cdot \left(\sum_{j=1}^{k} x_n^{(j)} + (n-k)\, x_n^{(k)} \right) ,$$

где $x_n^{(1)} < x_n^{(2)} < \dots < x_n^{(n)}$ – вариационный ряд, построенный по $x_1\,,\ x_2\,,\ \dots\,, x_n$. Она обладает следующими свойствами: $t_n^{(k)}$ есть несмещенная оценка параметра θ с дисперсией $\mathbf{D}_\theta(t_n^{(k)}) = \theta^2/k$, не зависящей от n. Последнее обстоятельство дает возможность сравнить ее с оценкой $t_k^{(k)}$ по полной выборке объема k, дисперсия которой равна $\mathbf{D}_\theta(t_k^{(k)}) = \theta^2/k$, т.е. $\mathbf{D}_\theta(t_n^{(k)}) = \mathbf{D}_\theta(t_k^{(k)})$, но средняя длительность испытания для оценки $t_n^{(k)}$ равна

$$\tau_1 = \mathbf{E}_\theta(X_n^{(k)}) = \theta\, \left(\frac{1}{n} + \frac{1}{n-1} + \dots + \frac{1}{n-k+1} \right),$$

в то время как для оценки $t_k^{(k)}$ среднее время испытания равно

$$\tau_2 = \mathbf{E}_\theta(X_k^{(k)}) = \theta\, \left(\frac{1}{k} + \frac{1}{k-1} + \dots + \frac{1}{2} + 1 \right).$$

Следовательно, если $n \geq k$, то $\tau_1 \leq \tau_2$. Так при $n = 10\,,\ k = 5\,,\ \tau_1/\tau_2 = 0.283$. Эти выводы легко получить используя представление Реньи для показательного распределения, которое имеет вид

$$x_n^{(k)} = \frac{\theta\xi_1}{n} + \frac{\theta\xi_2}{n-1} + \dots + \frac{\theta\xi_k}{n-k+1},$$

$1 \leq k \leq n$,где $\xi_1\,,\ \xi_2\,,\ \dots\,,\ \xi_n$ –независимые одинаково распределенные с.в. с функцией распределения $F(x) = \max(0\,,\ 1 - \exp(-x))$. Это представление было обобщено в 1991г. автором (Тихов М.С.)

Основной результат (обобщение представления Реньи) состоит в следующем. Если $X = (X_1\,,\ X_2\,,\ \dots\,,\ X_n)$ имеют распределение (46), то

$$X_n^{(k)} \overset{d}{=} \frac{\xi_1}{\lambda_{D_1} + \lambda_{D_2} + \dots + \lambda_{D_n}} + \frac{\xi_2}{\lambda_{D_2} + \dots + \lambda_{D_n}} + \dots + \frac{\xi_k}{\lambda_{D_k} + \dots + \lambda_{D_n}}.$$

В том случае, когда распределения одинаковые ($\lambda_1 = \lambda_2 = \dots = \lambda_n = \lambda$), то

$$X_n^{(k)} = \frac{\xi_1}{n\lambda} + \frac{\xi_2}{(n-1)\lambda} + \dots + \frac{\xi_k}{(n-k+1)},$$

т.е. имеем представление Реньи. Так как $\mathbf{E}(\xi_i) = 1$ и ожидание суммы равно сумме ожиданий, то имеем выражения для количеств τ_1 и τ_2.

Пусть $X_k\ (1 \leq k \leq n)$ – независимые случайные величины, имеющие соответствующие функции распределения $F_k(x) = \max(0\,,\ 1 - \exp(-\lambda_k x))$, $\lambda_k > 0\,,\ (1 \leq k \leq n)$. Совместная плотность распределения случайных величин $(X_1\,,\ X_2\,,\ \dots\,,\ X_n)$ имеет вид

$$(20)\qquad f_{X_1, X_2, \dots, X_n}(x_1\,,\, x_2\,,\, \dots, x_n) = \lambda_1\lambda_2 \dots \lambda_n \exp\,(-\sum_{i=1}^{n} \lambda_i x_i)\,,\ \ x_i > 0\,,\ 1 \leq i \leq n\,,$$

а совместная плотность распределения порядковых статистик $X_N^{(\bullet)} = (X_N^{(1)},\ \dots\,, X_N^{(N)})$ задается формулой

$$f_{X_n^{(\bullet)}}(x_1\,,\, x_2\,,\, \dots, x_n) = \lambda_1\lambda_2 \dots \lambda_n \sum_{(i_1, i_2, \dots, i_n)} \exp\left(- \sum_{j=1}^{n} \lambda_{i_j} x_j \right)\,,\ x_1 < x_2 < \dots < x_n\,,$$

В которой суммирование производится по всем $n!$ перестановкам $(i_1\,,\ i_2\,,\ \dots,i_n)$ чисел $(\,1\,,\ 2\,,\ \dots\,,\ n\,)$. Тогда

$$f_{X_n^{(1)},X_n^{(2)}-X_n^{(1)},\dots,X_n^{(n)}-X_n^{(n-1)}}(y_1\,,\ y_2\,,\ \dots,y_n)=$$

$$=\lambda_1\lambda_2...\lambda_n\sum_{(i_1,i_2,\dots,i_n)}\exp\left(-\left(y_1(\lambda_{i_1}+\lambda_{i_2}+\ \dots\ +\lambda_{i_n})+y_2(\lambda_{i_2}+\ \dots\ +\lambda_{i_n})+y_n\lambda_{i_n}\right)\right)=$$

$$=\sum_{(i_1,i_2,\dots,i_n)}\frac{\lambda_1\lambda_2...\lambda_n}{(\lambda_{i_1}+\lambda_{i_2}+\ \dots\ +\lambda_{i_n})(\lambda_{i_2}+\ \dots\ +\lambda_{i_n})...\lambda_{i_n}}f_{\eta_1}(y_1)f_{\eta_2}(y_2)\dots f_{\eta_n}(y_n),$$

$$y_i>0\,,\ 1\le i\le n\,,$$

где

$$\eta_1=\frac{\xi_1}{\lambda_{i_1}+\lambda_{i_2}+\ \dots\ +\lambda_{i_n}}\,,\ \eta_2=\frac{\xi_2}{\lambda_{i_2}+\ \dots\ +\lambda_{i_n}}\,,\ \dots\,,\ \eta_n=\frac{\xi_n}{\lambda_{i_n}}\,,$$

а величины $\xi_1\,,\ \xi_2\,,\ \dots\,,\ \xi_n$ – независимы и имеют стандартное показательное распределение. Отсюда выводим представление

$$X_n^{(k)}=\sum_{(i_1,i_2,\dots,i_n)}\frac{\lambda_1\lambda_2...\lambda_n}{(\lambda_{i_1}+\lambda_{i_2}+\ \dots\ +\lambda_{i_n})(\lambda_{i_2}+\ \dots\ +\lambda_{i_n})...\lambda_{i_n}}\times$$

$$\times\left(\frac{\xi_1}{\lambda_{i_1}+\lambda_{i_2}+\ \dots\ +\lambda_{i_n}}+\frac{\xi_2}{\lambda_{i_2}+\ \dots\ +\lambda_{i_n}}+\ \dots\ +\frac{\xi_k}{\lambda_{i_k}+\ \dots\ +\lambda_{i_n}}\right).$$

Пусть есть вектор антирангов, построенных по $X=(X_1\,,\ X_2\,,\ \dots\,,\ X_n)$.

Для плотности распределения (20) вектора $X=(X_1\,,\ X_2\,,\ \dots\,,\ X_n)$ совместное распределение антирангов задается формулой

$$P(D_1=i_{\,1}\,,\ D_2=i_{\,2}\,,\dots\,,\ D_n=i_{\,n})=P(X_{i_{\,1}}<X_{i_{\,2}}<\ \dots\ <X_{i_{\,n}})=$$

$$=\sum_{(i_1,i_2,\dots,i_n)}\frac{\lambda_1\lambda_2...\lambda_n}{(\lambda_{i_1}+\lambda_{i_2}+\ \dots\ +\lambda_{i_n})(\lambda_{i_2}+\ \dots\ +\lambda_{i_n})...\lambda_{i_n}}.$$

В таком случае из предыдущего получаем представление

$$X_n^{(k)}\stackrel{d}{=}\frac{\xi_1}{\lambda_{D_1}+\lambda_{D_2}+\ \dots\ +\lambda_{D_n}}+\frac{\xi_2}{\lambda_{D_2}+\ \dots\ +\lambda_{D_n}}+\ \dots\ +\frac{\xi_k}{\lambda_{D_k}+\ \dots\ +\lambda_{D_n}}.$$

6.2. Асимптотические распределения максимальных и крайних членов вариационного ряда.

Начнем с примера, взятого из книги *Тихов М.С.* Асимптотические методы в теории порядковых статистик и их приложения, 1985, с.25.

Пример 49. Пусть $X_1\,,\ X_2\,,\ \dots\,,\ X_{l(n)},\ X_{l(n)+1},\ \dots\,,\ X_n$ – последовательность независимых случайных величин, причем $X_1\,,\ X_2\,,\ \dots\,,\ X_{l(n)}$ распределены с функцией распределения $F(x)$, а $X_{l(n)+1},\ \dots\,,\ X_n$ – с ф.р. $F_1(x)$. Пусть:

1) $F(0)=0\,,\ F_1(0)=0$ и для всякого $\varepsilon>0$ имеет место $F(\varepsilon)>0\,,\ F_1(\varepsilon)>0$;

2) для любого $\tau>0$

$$\lim_{x\downarrow0}\frac{F(\tau x)}{F(x)}=\tau^{\alpha}\,,\ \alpha>0\,;\lim_{x\downarrow0}\frac{F_1(\tau x)}{F_1(x)}=\tau^{\alpha+\beta}\,,\ \beta>0\,.$$

Возьмем $a_n=n^{-\frac{1}{\alpha+\beta}}$, $l(n)=n^{\frac{\alpha}{\alpha+\beta}}$. Тогда

$$P(X_n^{(1)}<a_nx)=1-(1-F(a_nx))^{l(n)}(1-F_1(a_nx))^{n-l(n)}.$$

Но

$$(1-F(a_nx))^{l(n)}\underset{n\to\infty}{\to}\exp(-x^{\alpha}),(1-F_1(a_nx))^{n-l(n)}\underset{n\to\infty}{\to}\exp(-x^{\alpha+\beta}),x>0\,,$$

следовательно, предельная ф.р. п.с. $X_n^{(1)}$ имеет вид

$$\Phi_1(x) = 1 - \exp(-x^{\alpha} - x^{\alpha+\beta})\,,\ x > 0\,.$$

Замечание. Если при x, близких к нулю $F(x) = x^{\alpha}$, а $F_1(x) = x^{\alpha+\beta}$, то

$$(1 - F(a_n x))^{l(n)} = \left(1 - \frac{x^{\alpha}}{n^{\frac{\alpha}{\alpha+\beta}}}\right)^{n^{\frac{\alpha}{\alpha+\beta}}} \to \exp(-x^{\alpha})\,,$$

а

$$(1 - F_1(a_n x))^{n-l(n)} = \left(1 - \frac{x^{\alpha+\beta}}{n}\right)^{n-n^{\frac{\alpha}{\alpha+\beta}}} \sim \left(1 - \frac{x^{\alpha+\beta}}{n}\right)^{n} \to \exp(-x^{\alpha+\beta})\,.$$

Пусть $\{\,X_i\,,\ i \geq 1\,\}$ – стационарная последовательность зависимых случайных величин. Обозначим через $\mathcal{F}_s^u = \sigma(X_s, X_{s+1}, ..., X_{u-1}, X_u)$ – σ–алгебру, порожденную величинами $\{\,X_i\,,\ s \leq i \leq u\,\}$.

Определение 13. Последовательность $\{\,X_i\,,\ i \geq 1\,\}$ будем называть удовлетворяющей условию *сильного перемешивания (с.п.)*, если

$$\sup_{i\geq 1}\ \sup_{A\in\mathcal{F}_1^i}\ \sup_{B\in\mathcal{F}_{i+j}^{\infty}}\ |\,\mathbf{P}(AB) - \mathbf{P}(A)\mathbf{P}(B)\,|\ \leq \alpha(j) \xrightarrow[j\to\infty]{} 0.$$

Ясно, что $\alpha(1) \geq \alpha(2) \geq\ ...\ \geq \alpha(n) \geq\ ...\ .$

Определение 14. Если в определении 1 $\alpha\,(j) = 0$ для $j > m$, то $\{\,X_i\,,\ i \geq 1\}$ называется последовательностью m*–зависимых* случайных величин.

Пусть $\xi \geq 0$ задано. Для $n \in N$ определим константы $C_n(\xi)$ из равенства

$$\xi = n \cdot \mathbf{P}(X_1 \geq C_n(\xi)).$$

ТЕОРЕМА 16. *Пусть* $\{\,X_i\,,\ i \geq 1\,\}$ *– стационарная последовательность* m*–зависимых случайных величин со свойством*

$$\lim_{n\to\infty}\ n \cdot \max_{2\leq j\,\leq m}\ P(X_1 \geq C_n(\xi)\,,\ X_j \geq C_n(\xi)) = 0.$$

Тогда

$$\lim_{n\to\infty}\ P\,(X_n^{(n-k)} < C_n(\xi)) = 1 - \int_0^{\xi} \frac{x^{k-1}}{k!}\, e^{-x}\, dx\ \ (\,k = const\,).$$

Д о к а з а т е л ь с т в о. Обозначим через $\mathbf{P}(k;n)$ вероятность одновременного осуществления в точности k из n событий $A_1\,,\ A_2\,,\ ...\ ,\ A_n$. Тогда (см. Феллер В. Введение в теорию вероятностей и ее приложения, т.1, М.:Мир, 1984, с.125) имеет место формула включения-исключения

$$\mathbf{P}(k;n) = \sum_{j=k+1}^{n} (-1)^{j-k-1} C_{j-1}^{k} S_j,$$

где $S_m = \sum_{i_1<i_2<\,...\,<i_m} P\,(\bigcap_{t=1}^{m} A_{i_t})$. Если $A_i = (X_i \geq C_n(\xi))$, то для четного $(l-k-1)$ $(l \leq n)$

$$\sum_{j=k+1}^{l-1} (-1)^{j-k-1} C_{j-1}^{k} S_j \leq P\left(X_n^{(n-k)} \geq C_n(\xi)\right) \leq \sum_{j=k+1}^{l} (-1)^{j-k-1} C_{j-1}^{k} S_j.$$

Ясно, что

$$\sum P\,(X_i \geq C_n(\xi)) = n \cdot P\,(X_1 \geq C_n(\xi)) = \xi.$$

Далее,

$$\sum P\,(X_i \geq C_n(\xi)\,,\, X_j \geq C_n(\xi)) = \sum_{i=1}^{m} (n-i) P\,(X_i \geq C_n(\xi)\,,\, X_{i+1} \geq C_n(\xi)) +$$

$$+P^2\left(X_1 \geq C_n(\xi)\right)(n-m-1)(n-m)/2.$$

Но

$$\sum_{i=1}^{m}(n-i)P\left(X_i \geq C_n(\xi)\,,\, X_{i+1} \geq C_n(\xi)\right) \leq$$

$$\leq nm\left[1-\frac{m+1}{2n}\right]\cdot \max_{|\,i-j\,|\,\leq m} P(X_i \geq C_n(\xi)\,,\ X_j \geq C_n(\xi)).$$

Следовательно, когда $n \to \infty$, а ξ – фиксировано, то

$$\lim_{n\to\infty}\sum P\left(X_i \geq C_n(\xi)\,,\, X_j \geq C_n(\xi)\right) = \frac{\xi^2}{2}.$$

Общая сумма $\sum P\left(X_{i_1} \geq C_n(\xi)\,,\,\dots\,,X_{i_q} \geq C_n(\xi)\right)$ содержит C_n^q членов. Разобьем её на классы слагаемых. В первый класс отнесем слагаемые, в которых индексы не отличаются друг от друга больше, чем на m (их будет $C_m^q + C_m^{q-1}C_{n-m}^1$, порядка $(n-m)$ членов); во второй класс отнесем слагаемые только с одним X_i, отличающимся в индексе больше, чем на m (этих членов будет $C_m^{q-2}C_{n-m}^2$, порядка $\frac{(n-m)^2}{2}$ членов) и т.д. до $q-$го класса, в котором q величин $X_{i_1}\,,\,\dots\,,X_{i_q}$ отличаются в индексах больше, чем на m (их будет C_{n-m}^q, порядка $\dfrac{(n-m)^q}{q!}$ членов). Сумма членов в первом классе будет меньше константы, умноженной на $n\cdot \max\limits_{|\,i-j\,|\,\leq m} P(X_i \geq C_n(\xi)\,,\ X_j \geq C_n(\xi))$, сумма членов второго класса будет меньше константы, умноженной на $\frac{1}{2}n^2P(X_i \geq C_n(\xi))\cdot \max\limits_{|\,i-j\,|\,\leq m} P(X_i \geq C_n(\xi)\,,\ X_j \geq C_n(\xi))$, и так далее, до $q-$го класса, где сумма равна $[n^q/q! + O(n^{q-1})]\cdot P(X_i \geq C_n(\xi))$. В таком случае имеем:

$$\sum_{q=k+1}^{l-1}(-1)^{q-k-1}C_{q-1}^k\frac{\xi^q}{q!} \leq P\left(X_n^{(n-k)} \geq C_n(\xi)\right) \leq \sum_{q=k+1}^{l}(-1)^{q-k-1}C_{q-1}^k\frac{\xi^q}{q!}.$$

Так как $\dfrac{\xi^l}{l!}\xrightarrow[l\to\infty]{} 0$ и $\dfrac{\xi^l}{l!}C_l^k \xrightarrow[l\to\infty]{} 0$, то

$$\lim_{n\to\infty} P\left(X_n^{(n-k)} \geq C_n(\xi)\right) = \sum_{q=k+1}^{\infty}(-1)^{q-k-1}C_{q-1}^k\frac{\xi^q}{q!}.$$

Рассмотрим произведение $\left(\sum_{j=k+1}^{\infty}\frac{\xi^j}{j!}\right)\left(\sum_{i=0}^{\infty}\frac{(-\xi)^i}{i!}\right)$ и слагаемое

$$\sum_{i+j=k+q}\frac{(-1)^i\xi^{i+j}}{i!j!} = \frac{\xi^{k+q}}{(k+q)!}\left[C_{k+q}^0 - C_{k+q}^1 + \ \dots\ + (-1)^qC_{k+q}^q\right]$$

в этом произведении. Заметим, что (см. Феллер, с.81, (12.7))

$$C_{k+q}^0 - C_{k+q}^1 + \ \dots\ + (-1)^qC_{k+q}^q = (-1)^{q-1}C_{k+q-1}^{q-1}.$$

Из последних соотношений получаем

$$P\left(X_n^{(n-k)} \geq C_n(\xi)\right) \sim \sum_{q=k+1}^{\infty}(-1)^{q-k-1}C_{q-1}^k\frac{\xi^q}{q!} = \left(\sum_{j=k+1}^{\infty}\frac{\xi^j}{j!}\right)\left(\sum_{i=0}^{\infty}\frac{(-\xi)^i}{i!}\right) =$$

$$= \left(1-\sum_{j=0}^{k}\frac{\xi^j}{j!}\right)\cdot e^{-\xi} = \int_{\xi}^{\infty}\frac{x^k}{k!}\,e^{-x}\,dx.$$

ТЕОРЕМА 17. *Если $\{X_i\}$ –последовательность н.о.р. (стандартных) нормальных случайных величин, то асимптотическое распределение максимума $M_n = \max\{X_1\,,\ \dots\,,\ X_n\}$ является распределением типа I. Более точно,*

$$\mathbf{P}\,(\,a_n(M_n - b_n) < x) \to \exp(-\,e^{-x})\,,$$

где

$$a_n = (2\,\ln\,n)^{1/2},\ b_n = (2\,\ln\,n)^{1/2} - \tfrac{1}{2}(2\,\ln\,n)^{-1/2}(\ln\ln\,n + \ln\,4\pi)\,.$$

Д о к а з а т е л ь с т в о. Представим ξ из равенства $\xi = e^{-x}$. Тогда $1 - \Phi(u_n) = \frac{1}{n}\,e^{-x}$, откуда при $n \to \infty$, $u_n \to \infty$, и мы воспользуемся тем, что $1 - \Phi(u) \sim \frac{\varphi(u)}{u}$ при $u \to \infty$. Из последнего получаем, что

$$\Phi^n(u_n) = \left(1 - \frac{1}{n}\,e^{-x}\right)^n \to \exp(-e^{-x}) \quad \text{при} \quad n \to \infty.$$

Далее,

$$\frac{1}{n}\,e^{-x} \cdot \frac{\varphi\,(u_n)}{u_n} \to 1 \tag{21}$$

и поэтому логарифмируя обе части, получим

$$-\ln\,n - x + \ln\,u_n - \ln\,\varphi(u_n) \to 0,$$

или

$$-\ln\,n - x + \ln\,u_n + \frac{1}{2}\,\ln\,2\pi + \frac{u_n^2}{2} \to 0. \tag{22}$$

Деля обе части на $\ln\,n$ и учитывая, что $\ln\,u_n/u_n \to 0$, получим $\dfrac{u_n^2}{2\,\ln\,n} \to 1.$

Логарифмируя его обе части еще раз, получим

$$2\,\ln\,u_n - \ln\,2 - \ln\ln\,n \to 0 \quad\Rightarrow\quad \ln\,u_n = \frac{1}{2}\,(\,\ln\,2 + \ln\ln\,n) + o(1).$$

Тогда

$$\frac{u_n^2}{2} = x + \ln\,n - \frac{1}{2}\,\ln\,4\pi - \frac{1}{2}\,\ln\ln\,n + o(1),$$

или

$$u_n^2 = 2\,\ln\,n\,\left\{1 + \frac{x - \frac{1}{2}\,\ln\,4\pi - \frac{1}{2}\,\ln\ln\,n}{\ln\,n} + o\left(\frac{1}{\ln\,n}\right)\right\}$$

и поэтому

$$u_n = (2\,\ln\,n)^{1/2}\,\left\{1 + \frac{x - \frac{1}{2}\,\ln\,4\pi - \frac{1}{2}\,\ln\ln\,n}{2\ln\,n} + o\left(\frac{1}{\ln\,n}\right)\right\},$$

Так что

$$u_n = x\,a_n + b_n + o\left((\ln\,n)^{-1/2}\right) = x\,a_n + b_n + o\left(a_n^{-1}\right).$$

Отсюда

$$P(M_n < u_n) \to \exp(-e^{-x}),$$

или

$$P(M_n < x\,a_n + b_n + o(a_n^{-1})) \to \exp(-e^{-x}),$$

или

$$P\left(\frac{M_n - b_n}{a_n} + o(1) < x\right) \to \exp(-e^{-x}).$$

Пример 50. Показательное распределение:

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

Возьмем $a_n = 1\,,\; b_n = \ln\, n$. Тогда

$$P\left(\frac{M_n - b_n}{a_n} < x\right) = P\,(M_n < b_n + a_n x) = P\,(X_n^{(n)} < b_n + a_n x) =$$

$$= P^n(X_1 < b_n + x) = (1 - e^{-\ln\, n - x})^n = \left(1 - \frac{1}{n} e^{-x}\right)^n \to \exp(-e^{-x}).$$

Пример 51. Предельное распределение экстремальных значений типа I:

$$F(x) = \exp(-e^{-x})\,,\; x \in R.$$

Здесь $a_n = 1\,,\; b_n = \ln\, n$. Тогда

$$P\,(X_n^{(n)} < b_n + a_n x) = P^n(X_1 < b_n + x) = \exp\left(n\,(-e^{-\ln\, n - x})\right) = \exp(-e^{-x}) \;\;\Rightarrow$$

$$P\left(\frac{M_n - b_n}{a_n} < x\right) = \exp(-e^{-x}).$$

Пример 52. Распределение Парето

$$F(x) = 1 - k\, x^{-\alpha}\,,\; \alpha > 0\,,\; k > 0\,,\; x \ge k^{1/\alpha}.$$

Возьмем $a_n = (k\, n)^{1/\alpha}\,,\, b_n = 0$. Тогда

$$P\left(\frac{M_n - b_n}{a_n} < x\right) = F^n(a_n x) = \left(1 - k \cdot ((k\, n)^{1/\alpha} x)^{-\alpha}\right)^n =$$

$$= \left(1 - \frac{1}{n}\, x^{-\alpha}\right)^n \to \exp\,(-\, x^{-\alpha}).$$

Пример 53. Полиномиальный рост в концевой точке

$$F(x) = 1 - k\,(x_0 - x)^{\alpha}\,,\; \alpha > 0\,,\; k > 0\,,\; x_0 - k^{-1/\alpha} \le x \le x_0\,.$$

В частности, если $x_0 = 1\,,\; k = 1\,,\; \alpha = 1\,$, то имеем равномерное распределение с функцией распределения $F(x) = x\,,\; 0 \le x \le 1\,$.

Возьмем $a_n = (nk)^{1/\alpha}\,,\; b_n = x_0\,$. Тогда

$$P\,(M_n < b_n + a_n x) = F^n(a_n x + x_0) = \left(1 - k \cdot (-(k\, n)^{-1/\alpha} x)^{\alpha}\right)^n =$$

$$= \left(1 - \frac{1}{n}\,(-x)^{\alpha}\right)^n \to \exp\,(-\,(-x)^{\alpha})\,,\; x < 0.$$

Для равномерного распределения предельная функция распределения равна

$$e^x\,,\; x < 0\,.$$

6.3. Предельные распределения центральных и промежуточных членов вариационного ряда.

Пусть $\lambda_{k,n} = \dfrac{k}{n} = \lambda + o\left(\frac{1}{\sqrt{n}}\right)\,,\; 0 < \lambda < 1\,$. Рассмотрим распределение $X_n^{(k)}$ при $F(x) = x\,,\; (\,0 < x < 1\,)$.

Плотность распределения этой порядковой статистики равна

$f_n(x) = \frac{n!}{(k-1)!\,(n-k)!}\, x^{k-1}(1-x)^{n-k},\; 0 < x < 1\,$, и равна нулю в остальных случаях. Тогда

$$P\left(\frac{\sqrt{n}}{\sqrt{\lambda\,(1-\lambda)}}\,(X_n^{(k)} - \lambda) < x\right) = P\left(X_n^{(k)} < \lambda + x\frac{\sqrt{\lambda\,(1-\lambda)}}{\sqrt{n}}\right).$$

Беря производную по x от правой части, получим

$$\frac{n!}{(k-1)!\,(n-k)!}\cdot\frac{\sqrt{\lambda\,(1-\lambda)}}{\sqrt{n}}\cdot\left(\lambda+x\,\frac{\sqrt{\lambda\,(1-\lambda)}}{\sqrt{n}}\right)^{k-1}\left(1-\lambda-x\,\frac{\sqrt{\lambda\,(1-\lambda)}}{\sqrt{n}}\right)^{n-k}=$$

$$=\frac{n!}{(k-1)!\,(n-k)!}\cdot\frac{\lambda^{\,k-1/2}(1-\lambda)^{n-k+1/2}}{\sqrt{n}}\cdot\left(1+x\,\sqrt{\frac{1-\lambda}{\lambda n}}\right)^{k-1}\left(1-x\,\sqrt{\frac{\lambda}{(1-\lambda)n}}\right)^{n-k}\equiv A.$$

По формуле Стирлинга $n!\sim n^n e^{-n}\sqrt{2\pi n}$ $(n\to\infty)$. Поскольку $k\to\infty$, $n-k\to\infty$, то

$$\frac{n!}{(k-1)!\,(n-k)!}\sim\frac{n^n e^{-n}\sqrt{2\pi n}}{(k-1)^{k-1}e^{-(k-1)}\sqrt{2\pi(k-1)}\cdot(n-k)^{n-k}e^{-(n-k)}\sqrt{2\pi(n-k)}}=$$

$$=\frac{n^{k-1}\cdot n^{n-k}\cdot n\cdot e\sqrt{2\pi n}}{k^{k-1}(1-\frac{1}{k})^{k-1}\sqrt{2\pi(k-1)}(n-k)^{n-k}\sqrt{2\pi(n-k)}}\sim$$

$$\sim\left(\frac{n}{k}\right)^{k-1}\left(\frac{n}{n-k}\right)^{n-k}\cdot\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{\frac{k-1}{n}}}\cdot\frac{1}{\sqrt{\frac{n-k}{n}}}\cdot\sqrt{n}\sim$$

$$\sim\frac{1}{\lambda^{k-1}}\cdot\frac{1}{(1-\lambda)^{n-k}}\cdot\frac{1}{\sqrt{2\pi}\,\lambda^{1/2}(1-\lambda)^{1/2}}\cdot\sqrt{n}.$$

Отсюда

$$A=\frac{1}{\sqrt{2\pi}}\cdot B_n\,,\ B_n=\left(1+x\,\sqrt{\frac{1-\lambda}{\lambda n}}\right)^{k-1}\left(1-x\,\sqrt{\frac{\lambda}{(1-\lambda)n}}\right)^{n-k},$$

$$\ln\,B_n=(k-1)\,\ln\,\left(1+x\,\sqrt{\frac{1-\lambda}{\lambda n}}\right)+(n-k)\,\ln\left(1-x\,\sqrt{\frac{\lambda}{(1-\lambda)n}}\right)\sim$$

$$\sim k\,\left(x\,\sqrt{\frac{1-\lambda}{\lambda n}}-\frac{1}{2}x^2\,\frac{1-\lambda}{\lambda n}+o\left(\frac{1}{n}\right)\right)+$$

$$+(n-k)\left(-x\,\sqrt{\frac{\lambda}{(1-\lambda)n}}-\frac{1}{2}x^2\,\frac{\lambda}{(1-\lambda)n}+o\left(\frac{1}{n}\right)\right)\sim$$

$$\sim x\,\sqrt{\lambda\,(1-\lambda)}\cdot\sqrt{n}-x\,\sqrt{\lambda\,(1-\lambda)}\cdot\sqrt{n}-\frac{1}{2}x^2((1-\lambda)+\lambda)+o(1)\sim$$

$$\sim\ -\frac{1}{2}\,x^2\quad\Rightarrow\quad B_n\to e^{-\frac{1}{2}x^2}\quad\Rightarrow\quad A\to\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}.$$

Таким образом,

$$\sqrt{n}\,(X_n^{(k)}-\lambda)\ \ \sim\ \ N\,(0\,,\ \lambda\,(1-\lambda))\ .$$

Пусть теперь случайная величина X имеет непрерывную функцию распределения $F(x)$. Совершим преобразование Смирнова: $Y=F(X)$. Тогда случайная величина Y будет иметь равномерное на $(0,\,1)$ и $X=F^{-1}(Y)$ – обратное преобразование. Кроме того, $X_n^{(k)}=F^{-1}(Y_n^{(k)})$, поэтому

$$\sqrt{n}\,(X_n^{(k)}-\lambda)\ \ \sim\ \ N\left(0\,,\ \frac{\lambda\,(1-\lambda)}{f^2(x_\lambda)}\right).$$

Пример 54. Пусть случайная величина X имеет нормальное распределение $N(\theta\,,\ 1)$ и $\lambda = 0.5$. Тогда $x_{0.5} = me = \theta$. Если $\frac{k}{n} = \frac{1}{2} + o\left(\frac{1}{\sqrt{n}}\right)$, то

$$\sqrt{n}\,(X_n^{(k)} - \theta) \ \ \sim N\left(0\,,\ \frac{1}{4\,f^2(\theta)}\right).$$

В качестве оценки параметра θ можно взять выборочную медиану:

$$\tilde{m}_e = \begin{cases} X_n^{((n+1)/2)}\,, \text{ если } n \ - \text{ нечетно,} \\ X_n^{(n/2+1)}\,, \text{ если } n \ - \text{ четно,} \end{cases}$$

Но

$$f^2(\theta) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 = \frac{1}{2\pi}\,,\ \frac{1}{4\,f^2(\theta)} = \frac{\pi}{2} \approx \frac{1}{0.636619772}.$$

Кроме того,

$$\sqrt{n}\,(\bar{X} - \theta) \ \ \sim N(0\,, 1\,).$$

Получаем, что эффективность выборочной медианы по сравнению с эффективной оценкой – выборочное среднее составляет 63,66%.

Обозначим через $\$(x)$ функцию распределения стандартного нормального распределения. Нас будет интересовать параметр сдвига θ. Пусть функция распределения имеет вид

$F(x) = (1-\varepsilon)\,\Phi(x) + \varepsilon\,\Phi\left(\frac{x}{\tau}\right)$. (47)

Мы будем сравнивать две оценки: выборочную медиану me и выборочное среднее $\bar{x}$. Имеем

$$f(0) = \frac{1}{\sqrt{2\pi}}\left(1-\varepsilon+\frac{\varepsilon}{\tau}\right)\ ,\ \ f^2(0) = \frac{1}{2\pi}\left(1-\varepsilon+\frac{\varepsilon}{\tau}\right)^2\ ,$$

$$\sigma^2 = D(X) = (1-\varepsilon) + \varepsilon\tau^2.$$

Эффективность равна

$$e_{me,\bar{x}} = \frac{D(\bar{X})}{D(me)} = 4f^2(0)\,\sigma^2 = \frac{2}{\pi}\left(1-\varepsilon+\frac{\varepsilon}{\tau}\right)^2\left(1-\varepsilon+\varepsilon\tau^2\right).$$

Таблица 2. *Эффективность выборочной медианы по сравнению с выборочным средним.*

τ $\qquad\varepsilon$	0.01	0.05	0.1
2	0.65	0.70	0.75
3	0.68	0.83	1.00
4	0.72	1.03	1.36

ТЕОРЕМА 18. *Для унимодальных симметричных плотностей, удовлетворяющих соотношению* $f(x) \le f(0)$ *для всех* x,

$$e_{me,\bar{x}} \ge \frac{1}{3}\,.$$

Нижняя граница достигается для равномерного распределения (оценка параметра масштаба).

Д о к а з а т е л ь с т в о. Эффективность не зависит от масштаба, поэтому будем считать, что $f(0) = 1$. Тогда

$$\sigma^2 = E(X^2) = \int x^2 f(x)\,dx \tag{23}$$

при условии $0 \le f(x) \le 1\,,\ f(0) = 1\,,$ и $\int f(x)\,dx = 1$. (11)

Вместо (48) будем минимизировать

$$\int (x^2 - a^2)\, f(x)\, dx\,. \tag{24}$$

Но (24) будет минимально, если $f(x)$ будет сколь угодно большой при $x^2 < a^2$ и сколь угодно малой при $x^2 > a^2$, т.е.

$$f(x) = \begin{cases} 1 & \text{при} \quad |\,x\,| \;<\; a\,, \\ 0 & \text{при} \quad |\,x\,| \;>\; a\,. \end{cases}$$

Положим $a = 1/2$, чтобы (11) было выполнено. Тогда $f(x)$ – плотность равномерного распределения и $\sigma^2 = 1/12$. Значит

$$e_{me,\bar{x}} = 4 \cdot 1 \cdot \frac{1}{12} = \frac{1}{3}.$$

Пусть $\varsigma_1\,,\ \varsigma_2\,,\ \ldots\,,\ \varsigma_n$ – н.о.р. с.в. с ф.р. показательного закона

$F(x) = P(\varsigma_k < x) = 1 - e^{-\lambda x}$ для $x > 0\,,\quad \lambda > 0\,.$ (15)

Мы используем следующее свойство показательного распределения: для каждого $x > 0\,,\ y > 0$

$$P(\varsigma > x + y\,|\,\varsigma > y) = P(\varsigma > x + y\,,\, \varsigma > y)/P(\,\varsigma > y) = e^{-\lambda\,(x+y)}/e^{-\lambda\, y} = e^{-\lambda\, x}. \tag{25}$$

Свойство (25) характеризует показательное распределение. Действительно, пусть $S(x) = 1 - F(x)$. Тогда (25) эквивалентно уравнению

$$S(x+y) = S(x)\, S(y).$$

Пусть $G(x) = \ln\, S(x)$. Ввиду этого последнее преобразуется к уравнению

$$G(x+y) = G(x) + G(y).$$

Беря $x = y = 0$, получим $G(0) = 2\, G(0) \;\;\Rightarrow\;\; G(0) = 0$.

Если $x = n \in N\,,\ y = 1\,,$ то $G(n+1) = G(n) + G(1)$. Отсюда $G(2) = 2\, G(1)$ и по индукции $G(n) = n\, G(1)$. Мы также получаем, что $G(nx) = n\, G(x)$, откуда полагая $x = \dfrac{1}{n}$, получаем $G\left(\dfrac{1}{n}\right) = \frac{1}{n} G(1)$ и, следовательно, $G\left(\dfrac{m}{n}\right) = \dfrac{m}{n} G(1)$. Если мы предположим, что $G(x)$ – непрерывная функция, то переходя к пределу, получаем, что $G(x) = k\, x$, где $k = G(1)$; последнее равенство будет верно для всех действительных x. Итак, решение (12) на классе непрерывных функций исчерпывается функциями

$$G(x) = k\, x\,,\ x \in R.$$

Значит, $S(x) = e^{kx}$, а поскольку $S(x)$ убывает, то $k < 0$, т.е. $S(x) = e^{-\lambda x}\,,\ x > 0$.

Вычислим теперь распределение разности $\varsigma_n^{(k+1)} - \varsigma_n^{(k)}$. Ясно, что

$$P(\varsigma_n^{(k+1)} - \varsigma_n^{(k)} > x\,|\,\varsigma_n^{(k)} = y) = P(\varsigma_n^{(k+1)} > y + x\,|\,\varsigma_n^{(k)} = y)\,.$$

Из $(n-k)$ событий в момент времени y ни одно не должно произойти до момента $x + y$; согласно (25) вероятность этого равна

$$[\,P\,(\varsigma > x)]^{n-k} = e^{-\,\lambda(n-k)x},$$

Поэтому условная функция распределения св $\varsigma_n^{(k+1)} - \varsigma_n^{(k)}$, при условии $\varsigma_n^{(k)} = y$ равна

$$P(\varsigma_n^{(k+1)} - \varsigma_n^{(k)} > x\,|\,\varsigma_n^{(k)} = y) = e^{-\,\lambda(n-k)x}\,,\ x > 0\,..$$

Эта ф.р. не зависит от y, поэтому она равна безусловной фр св $\varsigma_n^{(k+1)} - \varsigma_n^{(k)}$, которая имеет п.р.

$$\lambda\,(n-k)\, e^{-\,\lambda(n-k)x}\,,\ x > 0\,. \tag{26}$$

Для показательного распределения совместная плотность распределения вектора $\varsigma_n^{(\bullet)}$ равна

$$n!\ \exp\left(-\sum_{i=1}^{n} x_n^{(i)}\right)\ ,\ x_n^{(1)} < \ ...\ < x_n^{(n)}.$$

Последнее выражение можно переписать в виде

$$n!\ \exp\left(-\sum_{i=1}^{n}(n-i+1)(x_n^{(i)} - x_n^{(i-1)})\right)\ , \tag{27}$$

где $x_n^{(0)} = 0$. Положив $\delta_i = (n-i+1)\,(\varsigma_n^{(i)} - \varsigma_n^{(i-1)})\ ,\ i = 1\ ,\ 2\ ,\ ...\ ,\ n$, мы замечаем, что $\delta_i\ ,\ i = 1\ ,\ 2\ ,\ ...\ ,\ n$, имеют показательное распределение с плотностью $\lambda\,\exp\,(-\lambda\,x)$ для $x > 0$. Из (26) и (27) следует, что с.в. независимы и

$$\varsigma_n^{(k)} = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} +\ ...\ + \frac{\delta_k}{n-k+1}\,. \tag{28}$$

Соотношение (28) доказывает марковскую зависимость величин $\{\,\varsigma_n^{(k)}\,\}$.

Пусть теперь $\xi_1\ ,\ \xi_2\ ,\ ...\ ,\ \xi_n$ – независимые одинаково распределенные случайные величины с общей строго монотонной функцией распределения $F(x)$, $\xi_n^{(i)} = o_i(\xi_1\ ,\ \xi_2\ ,\ ...\ ,\ \xi_n)$. Пусть

$\varsigma_k = -\ \ln\ F(\xi_k)$ и $\varsigma_n^{(k)} = o_k(\varsigma_1\ ,\ \varsigma_2\ ,\ ...\ ,\ \varsigma_n)$, $k = 1\ ,\ 2\ ,\ ...\ ,\ n$.

Так как $-\ \ln\ F(x)$ не возрастает, то $\varsigma_n^{(k)} = -\ \ln\ F(\xi_n^{(n-k+1)})$ и $\{\,\xi_k\,\}$ будучи независимыми, дают независимость ς_k.

Рассмотрим теперь распределение с.в. ς_k. Пусть $y = F^{-1}(x)$ ($0 \le x \le 1$), где $F^{-1}(x)$ – функция, обратная функции $F(x)$, точнее

$$F^{-1}(x) = \inf\,\{\,y:\ F(y) > x\,\}.$$

Из соотношения

$$P(\varsigma_k < x) = P(F(\xi_k) > e^{-x}) = P(\xi_k > F^{-1}(e^{-x}))$$

следует, что

$$P(\varsigma_k < x) = 1 - F(F^{-1}(e^{-x})) = 1 - e^{-x},$$

для $0 < x < \infty$. Следовательно, с.в. $\{\,\varsigma_k\,\}$ распределены по показательному закону с параметром $\lambda = 1$. Поэтому $\xi_n^{(k)}$ может быть записано в виде

$\xi_n^{(k)} = F^{-1}(\exp(-\,\varsigma_n^{(n-k+1)})) = F^{-1}(\exp(-\,\sum_{j=1}^{n+1-k}\delta_j/(n+1-j))$, (20)

где $\delta_1\ ,\ \delta_2\ ,\ ...\ ,\ \delta_n$ – н.о.р. по стандартному показательному закону с.в.

Из (20) следует, что

$$Y_k = \frac{F(\xi_n^{(k+1)})}{F(\xi_n^{(k)})} = \exp\left(\frac{\delta_{n-k+1}}{k}\right)\ ,\ k = 1\ ,\ 2\ ,\ ...\ ,\ n\,,$$

(полагая $F(\xi_n^{(n+1)}) = 1$) и, следовательно, величины $\{\,Y_k\,\}$ независимы.

Отметим также, что

$$P\,(F(\xi_k) < x) = P\,(\xi_k < F^{-1}(x)) = F(F^{-1}(x)) = x$$

для $0 < x < 1$, т.е. с.в. $F(\xi_k)$ равномерно распределены на ($0\ ,\ 1$). Обозначив $U_k = F(\xi_k)$, $k = 1\ ,\ 2\ ,\ ...\ ,\ n$, получаем также, что

$$F(\xi_n^{(k)}) = o_k(U_1\ ,\ U_2\ ,\ ...\ ,\ U_n) = U_n^{(k)}.$$

ТЕОРЕМА 19. *Если $k \geq 1$ – фиксированное целое, то*

$$\lim_{n\to\infty} P\,(n\varsigma_n^{(k)} < x) = \int_0^x \frac{t^{k-1}e^{-t}}{(k-1)!}\,dt \qquad (x>0)\,,$$

т.е. $n\varsigma_n^{(k)}$ имеет в пределе Γ_k-распределение.

Д о к а з а т е л ь с т в о. Плотность распределения п.с. $\varsigma_n^{(k)}$ равна

$$g_k(t) = k\,C_n^k e^{-nt}(e^t-1)^{k-1}$$

и, следовательно, п.р. с.в. $n\varsigma_n^{(k)}$ равна

$$\frac{1}{n}g_k\left(\frac{t}{n}\right) = C_{n-1}^{k-1}e^{-t}(e^{\frac{t}{n}}-1)^{k-1} = \left[n\left(e^{\frac{t}{n}}-1\right)\right]^{k-1}e^{-t}\prod_{j=1}^{k}\frac{\left(1-\frac{j}{n}\right)}{(k-1)!}.$$

Так как $\lim\limits_{n\to\infty} n\,(e^{t/n}-1) = t$, то $n\varsigma_n^{(k)}$ сходится при $n\to\infty$ к $\dfrac{t^{k-1}e^{-t}}{(k-1)!}$, т.е. к плотности Γ_k-распределения.

Этот результат может быть получен по-другому. Очевидно, что

$$n\varsigma_n^{(k)} = \delta_1+\delta_2+\,\dots\,+\delta_k+\sum_{j=2}^{k}\frac{(j-1)\,\delta_j}{n-j+1}.$$

П.р. с.в. $\delta_1+\delta_2+\,\dots\,+\delta_k$ равна $\dfrac{t^{k-1}e^{-t}}{(k-1)!}$. Остается показать, что остаток по вероятности стремится к нулю, а далее результат теоремы. Для этого нам понадобится следующая лемма.

Лемма 3. *Пусть $\gamma_n = \alpha_n+\beta_n$, где α_n и β_n – случайные величины, $F_n(x) = P\,(\alpha_n < x)$ и $\lim\limits_{n\to\infty} E(\beta_n) = 0$, $\lim\limits_{n\to\infty} D(\beta_n) = 0$. Пусть $\lim\limits_{n\to\infty} F_n(x) = F(x)$ в каждой точке непрерывности $F(x)$. Если $G_n(x) = P\,(\gamma_n < x)$, то*

$$\lim_{n\to\infty} G_n(x) = F(x).$$

Д о к а з а т е л ь с т в о. По неравенству Чебышева

$$P\,(\,|\beta_n|\ > \varepsilon) \leq \frac{E(\beta_n^2)}{\varepsilon^2} = \frac{D(\beta_n)+[E(\beta_n)]^2}{\varepsilon^2}\ \ (\varepsilon>0\,)\,.$$

Отсюда, для любых $\varepsilon > 0$ и $\delta > 0$, существует натуральное $n_0(\varepsilon\,,\ \delta)$ такое, что $P\,(\,|\beta_n|\ > \varepsilon) < \delta$ если $n \geq n_0(\varepsilon\,,\ \delta)$. Но

(29) $$G_n(x) = P\,(\gamma_n < x) \leq P\,(\alpha_n < x+\varepsilon) + P\,(\beta_n < -\varepsilon).$$

Действительно, если $\alpha_n+\beta_n < x$, то либо $\alpha_n < x+\varepsilon$, либо $\alpha_n \geq x+\varepsilon$. Но в последнем случае должно быть $\beta_n < -\varepsilon$, отсюда следует (29). С другой стороны,

(30) $$G_n(x) = P\,(\gamma_n < x) \leq P\,(\alpha_n < x-\varepsilon) + P\,(\beta_n > \varepsilon).$$

Действительно, если $\alpha_n < x-\varepsilon$, то либо $\alpha_n+\beta_n < x$, либо $\alpha_n+\beta_n \geq x$. Но в последнем случае должно быть $\beta_n > \varepsilon$, отсюда следует (30). Из (29) – (30) следует, что

$$F_n(x-\varepsilon)-\delta \leq G_n(x) \leq F_n(x+\varepsilon)+\delta.$$

Устремляя $n\to\infty$, ввиду произвольности ε и δ получим, что

$$F(x-0) \leq \varliminf_{x\to\infty} G_n(x) \leq \varlimsup_{x\to\infty} G_n(x) \leq F(x+0),$$

и если x – точка непрерывности $F(x)$, то $\lim\limits_{n\to\infty} G_n(x) = F(x)$.

Теорема 3 следует непосредственно из леммы 1, так как

$$\mathbf{E}\left(\sum_{j=2}^{k}\frac{(j-1)\,\delta_j}{n+1-j}\right)=\sum_{j=2}^{k}\frac{(j-1)}{n+1-j},\quad \mathbf{D}\left(\sum_{j=2}^{k}\frac{(j-1)\,\delta_j}{n+1-j}\right)=\sum_{j=2}^{k}\frac{(j-1)^2}{(n+1-j)^2}\,,$$

и условия леммы 1 выполнены. Имеем:

$$P(\,n\,(1-\eta_n^{(n+1-k)})<x)=P(\,\varsigma_n^{(k)}<-\ln\,(1-x/n))\,.$$

Но

$$\lim_{n\to\infty}\frac{-\ln(1-x/n)}{x/n}=1\,,$$

откуда следует, что с.в. $n\,(1-\eta_n^{(n+1-k)})$ имеет предельное распределение также Γ_k. Отметим, что с.в. $\eta_k=\exp(-\varsigma_k)$ равномерно распределены на (0,1), независимы и имеют такое же распределение, что и $1-\eta_k$.

Таким образом, мы получили следующий результат.

ТЕОРЕМА 20. *Если $k\geq 1$ – фиксированное целое, то распределения случайных величинп $\eta_n^{(k)}$ и $n\,(1-\eta_n^{(n+1-k)})$ сходятся при $n\to\infty$ к Гк-распределению с п.р.*

$$\frac{x^{k-1}e^{-x}}{(k-1)!}\,,\ x>0\,.$$

В следующих двух теоремах указаны предельные распределения для центральных членов (теорема 5) и промежуточных членов (теорема 6) вариационного ряда.

ТЕОРЕМА 21. *Пусть $\xi_1\,,\ \xi_2\,,\ \dots\,,\ \xi_n$ – нор св с абсолютно непрерывной фр $F(x)$. Предположим, что пр $f(x)$ непрерывна и положительна в интервале $[a\,,\ b]$. Если $0<F(a)<\lambda<F(b)<1$, и если $k(n)$ – последовательность натуральных чисел такая, что*

$$\lim_{n\to\infty}\left|\frac{k(n)}{n}-\lambda\right|\cdot\sqrt{n}=0\,,$$

тогда $\xi_n^{(k(n))}$, где $\xi_n^{(k)}$ –k-ая порядковая статистика, имеет в пределе нормальное распределение, именно,

$$\lim_{n\to\infty}P\left(\frac{\xi_n^{(k(n))}-b_n}{a_n}<x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\,e^{-t^2/2}dt\,,$$

где

$$b_n=F^{-1}(\lambda)\,,\ a_n=\frac{\sqrt{\lambda\,(1-\lambda)}}{f(F^{-1}(\lambda))\sqrt{n}}.$$

Д о к а з а т е л ь с т в о. Рассмотрим сначала предельное распределение св
(23) $\varsigma_n^{(n+1-k)}=-\ln\,F(\xi_n^{(k(n))})=\sum_{j=1}^{n+1-k(n)}\frac{\delta_j}{n+1-j}$, где δ_j – независимые и экспоненциально распределенные с.в. с п.р. $e^{-x}\,,\ x>0$. Следовательно, $E(\delta_j)=D(\delta_j)=1$, и

$$E(|\delta_j-1\,|^3)=\int_0^1(1-x)^3e^{-x}\,dx+\int_1^\infty(x-1)^3e^{-x}\,dx=$$

$$=e^{-1}\int_0^1x^3e^{-x}\,dx+e^{-1}\int_1^\infty x^3e^{-x}\,dx<(1-e^{-1})+6e^{-1}=1+5e^{-1}<3\,.$$

Из хорошо известной формулы

$$\sum_{j=1}^{N}\frac{1}{j}=\ln\ N+C+O\left(\frac{1}{N}\right),$$

где $C = 0,5772156649...$ – постоянная Эйлера, и из (22) следует, что

$$M_n = E\,(\varsigma_n^{(n+1-k(n))}) = \ln\left(\frac{1}{\lambda}\right) + O\left(\frac{1}{\sqrt{n}}\right).$$

Так как

$$\sum_{j=N_1}^{N_2}\frac{1}{j^2} = \frac{1}{N_1} - \frac{1}{N_2} + O\left(\frac{1}{N_1^2}\right), \quad \sum_{j=N_1}^{N_2}\frac{1}{j^3} = \frac{1}{2N_1^2} - \frac{1}{2N_2^2} + O\left(\frac{1}{N_1^3}\right),$$

то мы получаем

$$B_n^2 = D\,(\varsigma_n^{(n+1-k(n))}) = \frac{1-\lambda}{n\lambda} + O\left(\frac{1}{n^{3/2}}\right)$$

$$K_n^3 = \sum_{j=1}^{n+1-k(n)}\frac{E|\delta_j - 1\,|^3}{(n+1-j)^3} \le 3\sum_{j=1}^{n+1-k(n)}\frac{1}{(n+1-j)^3} = O\left(\frac{1}{n^2}\right),$$

Таким образом,
$\left(\frac{K_n}{B_n}\right)^3 = O\left(\frac{1}{\sqrt{n}}\right) = o(1)$ при $n \to \infty$ и по ц.п.т. Ляпунова, примененной к последовательности сумм (23), получаем результат.

Из (23) следует, что

$$P\left(\frac{\varsigma_n^{(n+1-k(n))} + \ln\lambda}{\sqrt{(1-\lambda)/n\lambda}} < x\right) = P\left(\xi_n^{(k(n))} > F^{-1}\left(\lambda\exp\left(-x\sqrt{\frac{1-\lambda}{n\lambda}}\right)\right)\right).$$

По теореме о среднем получаем

$$F^{-1}\left(\lambda\exp\left(-x\sqrt{\frac{1-\lambda}{n\lambda}}\right)\right) = F^{-1}(\lambda) + \lambda\left(\exp\left(-x\sqrt{\frac{1-\lambda}{n\lambda}}\right) - 1\right)/f(b_n\theta_n),$$

где $\lim\limits_{n\to\infty}\theta_n = 1$. Кроме того,

$$\lambda\left(\exp\left(-x\sqrt{\frac{1-\lambda}{n\lambda}}\right) - 1\right) = -x\sqrt{\frac{1-\lambda}{n\lambda}} + O\left(\frac{1}{n}\right).$$

Это влечет соотношение

$$\lim_{n\to\infty} P\left(\xi_n^{(k(n))} - F^{-1}(\lambda) < x\sqrt{\frac{1-\lambda}{n\lambda}}\cdot\frac{1}{f(F^{-1}(\lambda))}\right) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt\ .$$

Пусть $\varsigma_n^{(n+1-k(n))} = o_k(\varsigma_1\,,\ \varsigma_2\,,\ \dots\,, \varsigma_n)$, где ς_i имеет показательное распределение с п.р. $e^{-x}\,,\ x > 0$. Вновь напишем соотношение

$$\varsigma_n^{(n+1-k(n))} = \frac{\delta_1}{n} + \frac{\delta_2}{n-1} + \ \dots\ + \frac{\delta_{n+1-k(n)}}{k(n)}\,,$$

где будем предполагать, что $k = k(n) \to \infty$, $k(n)/n \to 0$, при $n \to \infty$.

Тогда

$$M_n = E\,(\varsigma_n^{(n+1-k(n))}) = \sum_{j=k+1}^{n}\frac{1}{j} = \sum_{j=1}^{n}\frac{1}{j} - \sum_{j=1}^{k}\frac{1}{j} = \ln\left(\frac{n}{k}\right) + O\left(\frac{1}{k}\right),$$

$$B_n^2 = D\,(\varsigma_n^{(n+1-k(n))}) = \sum_{j=k+1}^{n}\frac{1}{j^2} = \frac{n-k}{nk} + O\left(\frac{1}{k^2}\right)$$

$$L_n^3 = \sum_{j=k+1}^{n} \frac{E|\delta_j - 1|^3}{(n+1-j)^3} \le 3 \sum_{j=k+1}^{n} \frac{1}{(n+1-j)^3} \le \frac{3}{k^2}(1+o(1)), \quad \text{т.е.} \quad L_n^3 = O(k^{-2}).$$

Таким образом,
$\dfrac{L_n}{B_n} = O\left(k^{-1/6}\right) = o(1)$ при $n \to \infty$ и по ц.п.т. Ляпунова

$$\lim_{n\to\infty} P\left(\frac{\varsigma_n^{(n+1-k)} - M_n}{B_n} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt\ .$$

При этом

$$M_n = \ln\left(\frac{n}{k}\right) + O\left(\frac{1}{k}\right), B_n = \sqrt{\frac{n-k}{nk}(1+o(1))}.$$

ТЕОРЕМА 22. *Пусть $\xi_1\,,\ \xi_2\,,\ \ldots\,,\ \xi_n$ – нор св с абсолютно непрерывной фр $F(x)$. Предположим, что пр $f(x)$ непрерывна и положительна в интервале $[a\,,\ b]$, $F(x) = 0\,,\ x < a$. Если $k(n)$ – последовательность натуральных чисел такая, что*

$$k(n) \underset{n\to\infty}{\to} \infty\,, \quad \lim_{n\to\infty} \frac{k(n)}{n} = 0\,,$$

тогда $\xi_n^{(k(n))}$, где $\xi_n^{(k)}$ –k-ая порядковая статистика, имеет в пределе нормальное распределение, именно,

$$\lim_{n\to\infty} P\left(\frac{\xi_n^{(k(n))} - b_n}{a_n} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt\,,$$

где

$$b_n = F^{-1}\left(\frac{k}{n}\right)\ ,\ a_n = F^{-1}\left(\frac{k}{n} + \frac{\sqrt{k}}{n}\right) - F^{-1}\left(\frac{k}{n}\right).$$

Д о к а з а т е л ь с т в о. Используем соотношение $\varsigma_n^{(n+1-k)} = -\ln\ F(\xi_n^{(k)})$. Имеем

$$\lim_{n\to\infty} P\left(\frac{\varsigma_n^{(n+1-k)} - \ln\ \frac{n}{k}}{\sqrt{(n-k)/nk}} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt\,.$$

Из последнего следует, что

$$\lim_{n\to\infty}\ P\left(\xi_n^{(k)} > F^{-1}\left(\frac{k}{n}\exp\left(-\frac{x}{\sqrt{k}}\right)\right)\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt\,.$$

По теореме о среднем и условий теоремы 6,

$$F^{-1}\left(\frac{k}{n}\exp\left(-\frac{x}{\sqrt{k}}\right)\right) = \left[b_n + \frac{k\left(\exp(x/\sqrt{k}) - 1\right)}{n\, f(F^{-1}(k/n))}\right]\ (1+o(1)).$$

Таким образом,

$$\lim_{n\to\infty} P\left(\frac{\xi_n^{(k(n))} - b_n}{a_n} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt\,.$$

ТЕОРЕМА 23. *Пусть $\xi_1\,,\ \xi_2\,,\ \ldots\,,\ \xi_n$ – нор св с абсолютно непрерывной фр $F(x)$ как в теореме 6. Предположим, $k = k(n)$ – последовательность натуральных чисел такая, что*

$$\lim_{n\to\infty} \frac{k(n)}{n} = \lambda\,,\ \ 0 < \lambda < 1\,,\ \ s_1 = s_1(n)\,,\ s_2 = s_2(n)\,,\ s_1 + s_2 \to \infty$$

при $n \to \infty$, *а* $(s_1+s_2)/n \to 0$. *Тогда*

$$\lim_{n\to\infty} P\left(\frac{n\, f_\lambda\,(\xi_n^{(k+s_2)} - \xi_n^{(k+s_1)}) - (s_1+s_2)}{\sqrt{s_1+s_2}} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt\,, \quad f_\lambda = f(F^{-1}(\lambda))\,.$$

Д о к а з а т е л ь с т в о. Так как

$$\varsigma_n^{(m+s_2)} - \varsigma_n^{(m+s_1)} = \sum_{j\,=\,n+1-m-s_2}^{n+1-m+s_1} \frac{\delta_j}{j}\,,$$

и

$$M_n = E\,(\varsigma_n^{(m+s_2)} - \varsigma_n^{(m+s_1)}) = \sum_{j\,=\,n+1-m-s_2}^{n+1-m+s_1} \frac{1}{j} = \ln\left(\frac{n+1-m+s_1}{n+1-m-s_2}\right) + O\left(\frac{1}{n}\right) =$$

$$= \frac{s_1+s_2}{n} + O\left(\left(\frac{s_1+s_2}{n}\right)^2\right),$$

$$B_n^2 = D\,(\varsigma_n^{(m+s_2)} - \varsigma_n^{(m+s_1)}) = \sum_{j\,=\,n+1-m-s_2}^{n+1-m+s_1} \frac{1}{j^2} = \frac{s_1+s_2}{n^2} + O\left(\frac{1}{n}\right) = O\left(\frac{s_1+s_2}{n^2}\right)$$

$$K_n^3 = \sum_{j\,=\,n+1-m-s_2}^{n+1-m+s_1} \frac{E|\delta_j - 1\,|^3}{j^3} = O\left(\frac{s_1+s_2}{n^3}\right)$$

Таким образом,
$\dfrac{K_n}{B_n} = O\left((s_1+s_2)^{-1/2}\right) = o(1)$ при $n\to\infty$ и по ц.п.т. Ляпунова

$$\lim_{n\to\infty} P\left(\frac{n\,(\varsigma_n^{(m+s_2)} - \varsigma_n^{(m+s_1)}) - (s_1+s_2)}{\sqrt{s_1+s_2}} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt\ .$$

Возьмем $m = n+1-k(n)$. Тогда
$\xi_n^{(k+s_2)} - \xi_n^{(k+s_1)} = \dfrac{\varsigma_n^{(m+s_2)} - \varsigma_n^{(m+s_1)}}{f_\lambda} + \alpha_n$, где $\dfrac{n\alpha_n}{\sqrt{s_1+s_2}} \xrightarrow[n\to\infty]{P} 0.$
Из последнего следует результат теоремы.

Аналогично предыдущему можно показать, что с.в.

$$w = \frac{\sqrt{n}\,(\xi_n^{(k)} - F^{-1}(\lambda))}{\sqrt{\lambda\,(1-\lambda)}}\,, \quad w_1 = \frac{(\xi_n^{(k)} - \xi_n^{(k-s_1)})}{\sqrt{s_1}}\,, \quad w_2 = \frac{(\xi_n^{(k+s_2)} - \xi_n^{(k)})}{\sqrt{s_2}}\,,$$

асимптотически (при $n\to\infty$) независимы. Это последнее утверждение верно и тогда, когда $s_1 = const$, $s_2 = const$. Более того, в последнем случае величины $2s_1 f_\lambda w_1$ и $2s_2 f_\lambda w_2$ распределены как $\chi^2_{2s_1}$ и $\chi^2_{2s_2}$ соответственно с $2s_1$ и $2s_2$ степенями свободы. Таким образом, $2n\, f_\lambda\,(\xi_n^{(k+s_2)} - \xi_n^{(k+s_1)})$ асимптотически имеет распределение $\chi^2_{2(s_1+s_2)}$, не зависящее от $\xi_n^{(k)}$.

Замечание. В теореме 5 мы предполагали, что $k(n)/n = \lambda + o(1/\sqrt{n})$. Теорема 7 ослабляет это предположение. Действительно, возьмем $b_n = F^{-1}(\lambda + s(n)/n)$, где $s(n)\to\infty$, $s(n)/n \to 0$, $a_n = \sqrt{\lambda\,(1-\lambda)}/(f_\lambda\sqrt{n})$, и рассмотрим предельное поведение разности

$$(\xi_n^{(k+s)} - b_n)/a_n = (\xi_n^{(k)} - F^{-1}(\lambda))/a_n + (\xi_n^{(k+s)} - \xi_n^{(k)} - s)\cdot(n\, f_\lambda/\sqrt{s})\cdot\sqrt{s/n},$$

и в силу того, что $s/n \to 0$, мы получаем

$$\lim_{n\to\infty} P\left(\frac{(\xi_n^{(k+s)} - b_n)}{a_n} < x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt\ .$$

6.4. Оценивание по цензурированным выборкам.

Рассмотрим последовательность статистических экспериментов, порожденных наблюдениями $X^n = (X_1, ..., X_n)$, где $X_1, X_2, ...$ – последовательность одинаково распределенных случайных величин с плотностью $f(x;\theta), \theta \in \Theta \subset \mathbf{R}$. Пусть $X_n^{(1)} < ... < X_n^{(n)}$ есть вариационный ряд и предположим, что наблюдению доступны следующие порядковые статистики: $X_n^{(\cdot)} = (X_n^{(k+1)}, ..., X_n^{(n)})$, плотность распределения которого имеет вид:

$$f_n(x_n^{(k+1)}, ..., x_n^{(n)}; \theta) = \frac{n!}{k!\,(n-k)!} \prod_{j=k+1}^{n} f(x_n^j; \theta) F^k(x_n^{(k+1)}), \quad \text{для} \quad x_n^{(k+1)} < ... < x_n^{(n)},$$

и нуль в остальных случаях, $F(x;\theta) = \int_{-\infty}^{x} f(u;\theta)\, du$.

Зададимся неотрицательной измеримой функцией $\pi(\theta)$ и рассмотрим оценку t_n неслучайного параметра θ, определенную следующим равенством:

$$t_n = \int_{\Theta} \theta f_n(\theta | x_n^{(\cdot)})\, d\theta,$$

где

$$f_n(\theta | x_n^{(\cdot)}) = \pi(\theta) f_n(x_n^{(k+1)}, ..., x_n^{(n)}; \theta) (\int \pi(\theta) f_n(x_n^{(k+1)}, ..., x_n^{(n)}; \theta)\, d\theta)^{-1}.$$

Рассмотрим случайную функцию

$$\mathcal{Z}_{n,\theta}(y) = \sum_{j=k+1}^{n} \ln \frac{f(X_n^{(j)}; \theta + y/\sqrt{n})}{f(X_n^{(j)}; \theta)} + k \ln \frac{F(X_n^{(k+1)}; \theta + y/\sqrt{n})}{F(X_n^{(k+1)}; \theta)}.$$

Согласно подходу, изложенному в [15] нас будет интересовать поведение $\mathcal{Z}_{n,\theta}(y)$ и $\sqrt{n}\,(t_n - \theta)$ при $n \to \infty$.

Мы будем предполагать, что выполнены следующие условия.

$(P.1)$ $f(x;\theta)$ для почти всех x дифференцируемая функция x, а $F(x;\theta)$ дифференцируема в точке $\lambda_p(\theta) = F^{-1}(p;\theta), 0 < p < 1$.

$(P.2)$ Функция

$$J^2(\theta) = \left[\frac{\partial}{\partial\theta} F(x;\theta)\right]^2 / F(x;\theta)\,|_{x=\lambda_p(\theta)} + \int\limits_{\lambda_p(\theta)}^{\infty} \left(\frac{\partial}{\partial\theta} f(x;\theta)\right)^2 f(x;\theta)\, dx$$

является непрерывной функцией $\theta \in \Theta$ и $0 < J^2(\theta) < \infty$.

$(P.3)$ При $n \to \infty, 1 \leqslant k(n) \leqslant n$,

$$k(n+1) - k(n) = o\,(\min\{\sqrt{k(n)}, \sqrt{k(n+1)}\}).$$

Заметим, что функция $k(n) = [np]$, где $[a]$ есть целая часть числа a, удовлетворяет этому условию.

Положим

$$G_{k,n}(x) = \mathbf{P}(X_n^{(k)} < x).$$

Пусть константы $a_n > 0$ и b_n выбраны таким образом, что

$$\mathbf{P}((X_n^{(k)} - b_n)/a_n < x) = G_{k,n}(a_n x + b_n) \underset{n\to\infty}{\longrightarrow} G(x) \tag{31}$$

в каждой точке непрерывности невырожденной функции распределения $G(x)$.

ТЕОРЕМА 24. [17] *Пусть выполнены условия* $(P.1), (P.2), (P.3)$. *Для того, чтобы при* $n \to \infty$ *и* $k/n \to p$ $(0 < p < 1)$ *для данной последовательности* a_n *и* b_n *имело место (25), необходимо и достаточно выполнения следующего условия*

$$u_n(x) = \frac{F(a_n x + b_n) - p_{k,n}}{\sqrt{\tau_{k,n}}} \underset{n\to\infty}{\longrightarrow} u(x), \tag{32}$$

где $p_{k,n} = k/(n+1)$, $\quad \sigma_{k,n} = (n-k+1)/(n+1)$, $\quad \tau_{k,n} = \sqrt{p_{k,n}\sigma_{k,n}}/\sqrt{n+1}$, *для неубывающей функции* $u(x)$, *однозначно определяемой уравнением*

$$G(x) = (2\pi)^{-1/2} \int\limits_{-\infty}^{u(x)} \exp(-x^2/2)\, dx = N[u(x)].$$

Заметим, что если в точке λ_p функция $f(\lambda_p;\theta) > 0$, то $u(x) = x$.

ТЕОРЕМА 25. *Пусть выполнены условия* $(P.1), (P.2), (P.3)$, $k/n \underset{n\to\infty}{\longrightarrow} p, 0 < p < 1$.

Тогда при $n \to \infty$ *конечномерные распределения функций* $\mathcal{Z}_{n,\theta}(y)$ *сходятся по мере* $\mathbf{P}_\theta^n$ *к конечномерным распределениям функции*

$$\mathcal{Z}_\theta(y) = yJ(\theta)\xi - y^2 J^2(\theta)/2,$$

где $\xi \in N(0,1)$.

ТЕОРЕМА 26. *Пусть выполнены условия теоремы 3. Тогда при* $n \to \infty$ *предельное распределение разностей* $\sqrt{n}\,(t_n - \theta)$ *асимптотически нормально* $N(0, J^2(\theta))$, *при этом*

$$J^2(\theta) \leqslant I^2(\theta) = \mathbf{E}_\theta[(\partial/\partial\theta)\ln f(X;\theta)]^2.$$

Д о к а з а т е л ь с т в о. Покажем сначала, что при фиксированном y последовательность $\mathcal{Z}_{n,\theta}(y)$ сходится по распределению к нормальной случайной величине $N(-y^2J^2(\theta)/2, y^2J^2(\theta))$. Для этого рассмотрим характеристическую функцию величины $\mathcal{Z}_{n,\theta}(y)$. Имеем

$$\varphi_n(t) = \mathbf{E}_\theta(\exp(it\mathcal{Z}_{n,\theta}(y))) = \int\limits_{x_n^{(k+1)} < \ldots < x_n^{(n)}} \!\!\cdots\!\! \int (n-k-1)! \prod_{j=k+2}^{n} \exp\left(it \ln \frac{f(x_n^{(j)}; \theta + y/\sqrt{n})}{f(x_n^{(j)};\theta)}\right)\cdot$$

$$\cdot \frac{f(x_n^{(j)};\theta)\, dx_n^{(j)}}{1 - F(x_n^{(k+1)};\theta)} \cdot S_{k+1,n}(x_n^{(k+1)};\theta) \exp\left(it \ln \frac{f(x_n^{(k+1)}; \theta + y/\sqrt{n})}{f(x_n^{(k+1)};\theta)}\right)\cdot$$

$$\cdot \exp\left(it \ln \frac{f(x_n^{(k+1)}; \theta + y/\sqrt{n})}{f(x_n^{(k+1)};\theta)}\right) dx_n^{(k+1)}, \tag{33}$$

$$S_{k+1,n}(x;\theta) = \frac{n!}{k!\,(n-k)!} F^k(x;\theta)(1 - F(x;\theta))^{n-k-1} f(x;\theta) \quad -$$

– плотность распределения $(k+1)$-й порядковой статистики.

Положим

$$U_{n,\theta}(t, x_n^{(k+1)}, y) = (1 - F(x_n^{(k+1)};\theta))^{-1} \int\limits_{x_n^{(k+1)}}^{\infty} \exp\left(it \ln \frac{f(x;\theta + y/\sqrt{n})}{f(x;\theta)}\right) dx,$$

$$V_{n,\theta}(t, x_n^{(k+1)}, y) = \exp\left(it \ln \frac{f(x_n^{(k+1)}; \theta + y/\sqrt{n})}{f(x_n^{(k+1)};\theta)}\right),$$

$$W_{n,\theta}(t, x_n^{(k+1)}, y) = \exp\left(it \ln \frac{F(x_n^{(k+1)}; \theta + y/\sqrt{n})}{F(x_n^{(k+1)}; \theta)}\right).$$

Тогда с учетом введенных обозначений равенство (27) перепишется в виде

$$\varphi_n(t) = \int\limits_{-\infty}^{\infty} (U_{n,\theta}(t, x_n^{(k+1)}, y))^{n-k-1} \cdot V_{n,\theta}(t, x_n^{(k+1)}, y) \cdot (W_{n,\theta}(t, x_n^{(k+1)}, y))^k S_{k+1,n}(x_n^{(k+1)}; \theta)\, dx_n^{(k+1)}.$$

Теперь подинтегральное выражение преобразуем с помощью замены переменной интегрирования $z = (x_n^{(k+1)} - b_n)/a_n$.

Лемма 4. *В условиях теоремы 2 существуют и конечны следующие интегралы, и при* $n \to \infty$

$$\int\limits_{a_n z + b_n}^{\infty} \ln \frac{f(x; \theta + y/\sqrt{n})}{f(x; \theta)} f(x; \theta)\, dx = F(a_n z + b_n; \theta) - F(a_n z + b_n; \theta + y/\sqrt{n}) -$$

$$-\frac{y^2}{2n} \int\limits_{\lambda_p(\theta)}^{\infty} ((\partial/\partial\theta) f(x; \theta))^2 (f(x; \theta))^{-1}\, dx + o(n^{-1}), \tag{34}$$

$$\int\limits_{a_n z + b_n}^{\infty} \ln^2 \frac{f(x; \theta + y/\sqrt{n})}{f(x; \theta)} f(x; \theta)\, dx = n^{-1} \int\limits_{\lambda_p(\theta)}^{\infty} ((\partial/\partial\theta) f(x; \theta))^2 (f(x; \theta))^{-1}\, dx + o(n^{-1}), \tag{35}$$

$$\ln \frac{F(a_n z + b_n; \theta + y/\sqrt{n})}{F(a_n z + b_n; \theta)} = \omega_{n,\theta}(a_n z + b_n) - \frac{1}{2}\omega_{n,\theta}^2(a_n z + b_n) + o(n^{-1}), \tag{36}$$

$$\omega_{n,\theta}(x) = F(x; \theta + y/\sqrt{n})(F(x; \theta))^{-1} - 1.$$

Д о к а з а т е л ь с т в о. Обозначим

$$\rho_{n,\theta}(x) = f(x; \theta + y/\sqrt{n})(f(x; \theta))^{-1} - 1.$$

Разложим

$$\ln \frac{f(x; \theta + y/\sqrt{n})}{f(x; \theta)} = \ln(1 + \rho_{n,\theta}(x))$$

в ряд по степеням $\rho_{n,\theta}(x)$ в области $A_{\varepsilon\delta}^c \bigcap (a_n z + b_n, \infty)$, где

$$A_{\varepsilon\delta} = \{x : |\ \ln(f(x; \theta + \varepsilon)/f(x; \theta))\ | > \delta\}.$$

Имеем

$$\ln(1 + \rho_{n,\theta}(x)) = \rho_{n,\theta}(x) - \frac{1}{2}\rho_{n,\theta}^2(x) + o(n^{-1}), \quad \ln^2(1 + \rho_{n,\theta}(x)) = \rho_{n,\theta}^2(x) + o(n^{-1}).$$

В области $A_{\varepsilon\delta}$ интегралы по плотности $f(x; \theta)$ есть $o(\varepsilon^2)$, отсюда получаем (28) и (29).

Аналогично, разлагая $\ln(1 + \omega_{n,\theta}(a_n z + b_n))$ в области
$B_{\varepsilon\delta}^c = \{x : |\ \ln(F(x; \theta + \varepsilon)/F(x; \theta))\ | \leqslant \delta\}$
получим соотношение (30). Лемма доказана.

Из леммы 1 следует, что существует момент порядка 2 случайной величины

$\ln \dfrac{f(x;\theta+y/\sqrt{n})}{f(x;\theta)\chi_{\{(a_nx+b_n,infty)\}}(X)}$ при каждом y, поэтому справедливо разложение

$$(37)\qquad [U_{n,\theta}(t,a_nz+b_n,y)=1+\frac{it}{1-F(a_n+z+b_n;\theta)}\int\limits_{a_n+z+b_n}^{\infty}\ln(1+\rho_{n,\theta}(x))f(x;\theta)\,dx-$$

$$-\frac{t^2}{2(1-F(a_n+z+b_n;\theta))}\int\limits_{a_n+z+b_n}^{\infty}\ln^2(1+\rho_{n,\theta}(x))f(x;\theta)\,dx+o(n^{-1}),$$

когда $n\to\infty$.

Обозначим $Q_n(t)=U_{n,\theta}^{n-k-1}V_{n,\theta}W_{n,\theta}^k$. С учетом (28) – (30) получим

$$\ln Q_{n,\theta}(t)=\ln(U_{n,\theta}^{n-k-1}V_{n,\theta}W_{n,\theta}^k)=itk(\omega_{n,\theta}(a_nz+b_n)-\frac{1}{2\omega_{n,\theta}^2(a_nz+b_n)})+$$

$$+it(n-k)(F(a_nz+b_n;\theta)-F(a_nz+b_n;\theta+y/\sqrt{n}))(1-F(a_nz+b_n;\theta))^{-1}-$$

$$-\frac{1}{2}t(n-k)(i+t)y^2(1-F(a_nz+b_n;\theta))^{-1}\int\limits_{\lambda_p(\theta)}^{\infty}((\partial/\partial\theta)f(x;\theta))^2(f(x;\theta))^{-1}\,dx+o(1)=$$

$$=it(F(a_nz+b_n;\theta)-F(a_nz+b_n;\theta+y/\sqrt{n}))F(a_nz+b_n;\theta)(1-F(a_nz+b_n;\theta))^{-1}-\frac{1}{2}t(i+t)J^2(\theta)+o(1).$$

Далее,

$$F(a_nz+b_n;\theta)-F(a_nz+b_n;\theta+y/\sqrt{n})=\int\limits_{\theta}^{\theta+y/\sqrt{n}}F'_\theta(a_nz+b_n;v)\,dv=\frac{\partial F(a_nz+b_n;v)}{\partial\theta}|_{v=\theta}\cdot\frac{y}{\sqrt{n}}(1+o(1)).$$

По теореме 1 при $n\to\infty$

$$u_n(z;\theta)=\frac{F(a_nz+b_n;\theta)-\rho_{k,n}}{\sqrt{\tau_{k,n}}}\sim\frac{nF(a_nz+b_n;\theta)-k}{\sqrt{np(1-p)}}\to u(z;\theta),$$

а предельный закон представляет собой $N[u(z;\theta)]$. Из результатов работы [17] следует, что

$$u_n(z;\theta)=\sim(nF(a_nz+b_n;\theta)-k)(F(a_nz+b_n;\theta)(1-F(a_nz+b_n;\theta)))^{-1}.$$

Суммируя вышесказанное мы получаем следующее представление для $\ln Q_{n,\theta}(t)$

$$\ln Q_{n,\theta}(t)=ity\frac{\partial F(a_nz+b_n;v)}{\partial v}|_{v=\theta}\cdot u(z;\theta)-\frac{itJ^2(\theta)}{2}-\frac{1}{2}\int\limits_{\lambda_p(\theta)}^{\infty}((\partial/\partial\theta)f(x;\theta))^2(f(x;\theta))^{-1}\,dx+o(1).$$

В [17] показано, что последовательность нормированных функций $a_nS_{k+1,n}(a_nz+b_n;\theta)$ сходится равномерно в любом конечном интервале, далее показывается, что интеграл в оставшейся части есть $o(1)$, откуда следует теорема 2 для фиксированного y. Чтобы доказать сходимость конечномерных распределений функции $\mathcal{Z}_\theta(y)$ остается воспользоваться стандартной техникой, применив те же рассуждения к произвольной линейной комбинации $\sum_{j=1}^{l}\alpha_j\mathcal{Z}_\theta(y_j)$. Теорема доказана.

Следуя схеме рассуждений работы [2] можно показать, что

$$\mathbf{E}_\theta(\exp\mathcal{Z}_\theta(y_1)-\exp\mathcal{Z}_\theta(y_2))^2\leqslant H(y_1-y_2)^2$$

для $n>n_0$ и для любого $N>0$ существует постоянная C_N, такая, что при $n>n_1$ и всех y

$$\mathbf{E}_\theta(\exp\mathcal{Z}_\theta(y))\leqslant C_N/N.$$

Из последних соотношений следует теорема 3.

Приведем следующий **пример**, иллюстрирующий применение предыдущих результатов.

Пусть $x_1, ..., x_n$ – независимые одинаково распределенные по показательному закону случайные величины с плотностью $f(x) = \lambda \exp(-\lambda x), x > 0$. Наилучшей оценкой параметра λ в определенном смысле является статистика

$$t^(k)_n = \left(\sum_{j=1}^{k} x_n^{(j)+(n-k)x_n^{(k)}} \right) /k,$$

где $x_n^{(1)} < ... < x_n^{(n)}$ – вариационный ряд, построенный по $x_1, ..., x_n$. Она обладает следующими свойствами: $t^(k)_n$ есть несмещенная оценка параметра λ с дисперсией $\mathbf{D}_\lambda(t^(k)_n) = \lambda^2/k$, не зависящей от n. Последнее обстоятельство дает возможность сравнить ее с оценкой $t^(k)_k$ по полной выборке объема k, дисперсия которой равна $\mathbf{D}_\lambda(t^(k)_k) = \lambda^2/k$, но средняя длительность испытания для оценки $t^(k)_n$ равна

$$\tau_1 = \mathbf{E}_\lambda(X_n^k) = \lambda \left(\frac{1}{n} + \frac{1}{n-1} + ... + \frac{1}{n-k+1} \right),$$

в то время как для оценки $t^(k)_k$ среднее время испытания равно

$$\tau_2 = \mathbf{E}_\lambda(X_k^k) = \lambda \left(\frac{1}{k} + \frac{1}{k-1} + ... + \frac{1}{2} + 1 \right).$$

Следовательно, так как $n \geqslant k$, то $\tau_1 \leqslant \tau_2$. Так при $n = 10, k = 5$ имеем: $\tau_1/\tau_2 = 0.283$. Эти выводы легко следуют из представления Реньи.

Глава 7.

Асимптотическая теория оценивания.

7.1. Асимптотическая нормальность оценок максимального правдоподобия.

Условия регулярности.

Условие 0. Область, где $f = f(x;\theta) > 0$ не зависит от θ.

Условие 1. Существуют производные $\dfrac{\partial}{\partial\theta}\ln f(x;\theta)$, $\dfrac{\partial^2}{\partial\theta^2}\ln f(x;\theta)$, $\dfrac{\partial^3}{\partial\theta^3}\ln f(x;\theta)$ для почти всех x и при каждом θ, принадлежащем невырожденному интервалу A и включающем «истинное» значение параметра.

Условие 2. Для любого $\theta \in A$

$$\left|\frac{\partial f}{\partial\theta}\right| \le M_1(x), \left|\frac{\partial^2 f}{\partial\theta^2}\right| \le M_2(x), \left|\frac{\partial}{\partial\theta}\ln f(x;\theta)\right| \le H(x),$$

$$\int M_1(x)\,dx < \infty, \int M_2(x)\,dx < \infty, \mathbf{E}_\theta(H(X)) \le K.$$

Условие 3. Для каждого $\theta \in A$

$$0 < I^2(\theta) = \mathbf{E}_\theta\left[\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right] < \infty.$$

ТЕОРЕМА 27. *Пусть выполнены условия 0-3. Тогда $\tilde\theta$ — состоятельная оценка.*

Доказательство. Из разложения Тейлора функции $\dfrac{\partial \ln f}{\partial\theta}$ в точке θ_0 («истинное» значение параметра) имеем

$$\frac{\partial \ln f}{\partial\theta} = \left(\frac{\partial \ln f}{\partial\theta}\right)_{\theta_0} + (\theta-\theta_0)\left(\frac{\partial^2 \ln f}{\partial\theta^2}\right)_{\theta_0} + \frac{1}{2}\alpha(\theta-\theta_0)^2 H(x),\ \ |\,\alpha\,| < 1. \tag{38}$$

Поэтому для уравнения МП

$$\frac{1}{n}\frac{\partial \ln L}{\partial\theta} = B_0 + B_1(\theta-\theta_0) + \frac{1}{2}\alpha B_2(\theta-\theta_0)^2 = 0, \tag{39}$$

где

$$B_0 = \frac{1}{n}\sum_{i=1}^n\left(\frac{\partial \ln f(x_i;\theta)}{\partial\theta}\right)_{\theta_0}, B_1 = \frac{1}{n}\sum_{i=1}^n\left(\frac{\partial^2 \ln f(x_i;\theta)}{\partial\theta^2}\right)_{\theta_0}, B_2 = \frac{1}{n}\sum_{i=1}^n H(x_i).$$

По условиям 1 и 2 для любого $\theta \in A$

$$\int_{-\infty}^{\infty}\frac{\partial f}{\partial\theta}\,dx = \int_{-\infty}^{\infty}\frac{\partial^2 f}{\partial\theta^2}\,dx = 0,$$

поэтому

$$\mathbf{E}_\theta\left(\frac{\partial \ln f}{\partial\theta}\right) = \int_{-\infty}^{\infty}\frac{\partial f}{\partial\theta}\,dx = 0,$$

$$\mathbf{E}_\theta\left(\frac{\partial^2 \ln f}{\partial\theta^2}\right) = \int_{-\infty}^{\infty}\left(\frac{1}{f}\frac{\partial^2 f}{\partial\theta^2} - \left(\frac{1}{f}\frac{\partial f}{\partial\theta}\right)^2\right) f(x;\theta)\,dx = -\mathbf{E}_\theta\left(\left(\frac{\partial \ln f}{\partial\theta}\right)^2\right) = -I^2(\theta)$$

и по условию 3 $I^2(\theta) > 0$. По закону больших чисел при $n \to \infty$,

$$B_0 \xrightarrow{P_\theta} 0,\ B_1 \xrightarrow{P_\theta} -I^2(\theta),\ B_2 \xrightarrow{P_\theta} \mathbf{E}_\theta(H(X)) < K\ .$$

Следовательно, существует $n_0 = n_0(\varepsilon, \delta)$ такое, что для $n > n_0$,

$$p_1 = \mathbf{P}(|\,B_0|\, \geq \delta^2) < \frac{1}{3}\varepsilon, p_2 = \mathbf{P}(\,B_1\, \geq -\frac{I^2}{2}) < \frac{1}{3}\varepsilon, p_3 = \mathbf{P}(|\,B_2|\, \geq 2K) < \frac{1}{3}\varepsilon.$$

Рассмотрим событие $S = \{\,x = (x_1, ..., x_n)\, :\ |\,B_0|\, < \delta^2,\ B_1\, < -\frac{I^2}{2},\ |\,B_2|\, < 2K\}$. Для противоположного события $\bar{S}$ имеем: $\mathbf{P}\,(\bar{S}) \leq p_1 + p_2 + p_3 < \varepsilon$, откуда следует, что $\mathbf{P}\,(S) > 1 - \varepsilon$.

Если $\theta = \theta_0 \pm \delta$, то правая часть уравнения (*) принимает вид:

$$B_0 \pm B_1\delta + \frac{1}{2}\alpha B_2\delta^2.$$

Если $x \in S$, то $\left|\,B_0 + \frac{1}{2}\alpha B_2\delta^2\,\right|\; < \delta^2(1+K)\,,\ B_1\delta < -\frac{1}{2}I^2\delta$. Если $\delta < \frac{I^2}{2\,(1+K)}$, то знак всего выражения определяется при $\theta = \theta_0 \pm \delta$ вторым слагаемым, так что $\frac{\partial \ln L}{\partial \theta} > 0$ при $\theta = \theta_0 - \delta$ и $\frac{\partial \ln L}{\partial \theta} < 0$ при $\theta = \theta_0 + \delta$. Но $\frac{\partial \ln L}{\partial \theta}$ есть непрерывная функция θ. Поэтому с вероятностью $1 - \varepsilon$ существует корень между $\theta = \theta_0 - \delta$ и $\theta = \theta_0 + \delta$ для $n > n_0$. В силу произвольности δ отсюда следует состоятельность этого корня.

Другое доказательство этой теоремы.

Докажем сначала следующее неравенство.

(I) *Пусть f и g — плотности, а S —область, на которой $f > 0$. Тогда*

$$\int_S f\, \ln\frac{f}{g}\, d\mu \geq 0 \ \ \Leftrightarrow\ \ \int_S f\, \ln\frac{g}{f}\, d\mu \leq 0$$

со знаком равенства только при $f = g$.

Д о к а з а т е л ь с т в о. Заметим, что при $x > 0$,

$$\ln\, x = (x-1) + \frac{1}{2y^e}\,(x-1)^2,$$

где $y \in (1, x)$. Применяя это разложение, получим

$$\int\limits_S f\, \ln\frac{g}{f}\, d\mu = \int\limits_S f\, d\mu - \int_S g\, d\mu - \frac{1}{2}\int\limits_S \frac{1}{y^2}\left(\frac{g}{f} - 1\right)^2 d\mu = -\frac{1}{2}\int\limits_S \frac{1}{y^2}\left(\frac{g}{f} - 1\right)^2 d\mu \leq 0. \quad \text{(II)}$$

Если функция $\ln\, f(x\,;\ \theta)$ *дифференцируема в интервале, содержащем истинное значение параметра, то с вероятностью 1, когда $n \to \infty$, уравнение м.п. имеет корень, являющейся состоятельной оценкой.*

Д о к а з а т е л ь с т в о. Пусть θ_0 — истинное значение параметра. Возьмем точки $\theta = \theta_0 \pm \delta$. Согласно предыдущему неравенству имеем

$$E_{\theta_0}\left(\ln\frac{f(x\,;\ \theta_0 - \delta)}{f(x\,;\ \theta_0)}\right) < 0\,,\ \ E_{\theta_0}\left(\ln\frac{f(x\,;\ \theta_0 + \delta)}{f(x\,;\ \theta_0)}\right) < 0\,. \tag{40}$$

Согласно усиленному закону больших чисел с вероятностью 1 при $n \to \infty$

$$\frac{1}{n}\ [\,L_x(\theta_0 \pm \delta) - L_x(\theta_0)] < 0.$$

Другими словами, для почти всех выборок значение функции $L_x(\theta_0)$ в конце концов превзойдет ее значение $L_x(\theta_0 \pm \delta)$. Так как $L_x(\theta)$ непрерывна в интервале $(\theta_0 \pm \delta)$, то она имеет локальный максимум на этом интервале. Если $L_x(\theta)$ дифференцируема, то ее производная в точке $\tilde{\theta}$ локального максимума обращается в нуль. Поскольку δ произвольно, то $\tilde{\theta}$ – состоятельная оценка θ.

ТЕОРЕМА 28. *Пусть выполнены условия 0 – 3. Если $\tilde\theta$ – состоятельный корень уравнения правдоподобия, то с вероятностью 1*

$$\sqrt{n}\,\left[(\tilde\theta-\theta_0)\,I(\theta_0)-\frac{1}{n}\,\frac{dL}{d\theta_0}\right]\to 0.$$

Д о к а з а т е л ь с т в о. Из (9) и (10) получаем

$$I\,\sqrt{n}\,(\tilde\theta-\theta_0)=\frac{\dfrac{1}{I\sqrt{n}}\sum_1^n\left(\dfrac{\partial\,\ln\,f_i}{\partial\theta}\right)}{-\dfrac{B_1}{I^2}-\dfrac{1}{2}\xi\,B_2\dfrac{\tilde\theta-\theta_0}{I^2}}.$$

Из сказанного выше следует, что знаменатель дроби в правой части этого равенства сходится по вероятности к 1. По центральной предельной теореме (теорема Линдеберга-Леви) сумма $\dfrac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}\left(\dfrac{\partial\,\ln\,f_i}{\partial\theta}\right)$ асимптотически нормальна $N(0,I^2)$, значит, $\dfrac{1}{I\sqrt{n}}\sum\limits_{i=1}^{n}\left(\dfrac{\partial\,\ln\,f_i}{\partial\theta}\right)$ асимптотически нормальна $N(0,1)$.

Теорема доказана.

7.2. $M-$оценки. Эффективность метода максимального правдоподобия.

Метод максимального правдоподобия состоит в том, чтобы найти оценку параметра θ из условия

$$\sum_{i=1}^{n}\ln\,f(x_i\,;\,\theta)\ \ \to\ \ \max_{\theta}.$$

Необходимым условием этого является

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\,\ln\,f(x_i\,;\,\theta)=0.$$

Предположим, что $x_1\,,\ x_2\,,\ \dots\,,\ x_n$ независимы и одинаково с плотностью $f(x\,;\ \theta)$. Определим оценку параметра θ следующим образом:

$$T_n=\arg\min_{\theta}\sum_{i=1}^{n}\rho\,(\,x_i\,;\,\theta),$$

где $\rho(x\,;\ \theta)$ – заданная функция. Если $\rho(x\,;\ \theta)\ =-\ln\,f(x\,;\ \theta)$, то Tn оценка максимального правдоподобия для θ. Мы можем также определить Tn как решение уравнения

$$\sum_{i=1}^{n}\psi\,(\,x_i\,;\,T_n)=0,$$

где $\psi\,(x\,;\ \theta)=\frac{\partial}{\partial\theta}\,\rho\,(\,x\,;\ \theta)$. Такие оценки называются *оценками минимального контраста.*

Рассмотрим следующую модель. Пусть $x_1\,,\ x_2\,,\ \dots\,,\ x_n$ – случайные величины и

(41) $$x_i=\mu+\varepsilon_i\,,\quad i=1\,,\,2\,,\,\dots\,,\,n,$$

где ε_i независимые ошибки с общим распределением F_0 и конечной дисперсией σ^2. Нас интересует параметр μ. Функция распределения с.в. x_i имеет вид $F(x)=F_0((x-\mu)/\sigma)$. В этом случае $\theta=\mu$ и мы определим оценку как решение уравнения

$$L_n(t_n)=\sum_{i=1}^{n}\psi\,(x_i-t_n)=0.$$

Оценка t_n параметра θ называется *$M-$оценкой.*

Пусть
$$\mathbf{E}_\theta(\psi\,(X-\theta))=0.$$
Тогда (рассуждая формально)
$$L_n(t_n)=0=L_n(\theta_0)+(t_n-\theta)\,(L_n'(\theta_0)+\delta_n)\,,$$
откуда
$$\sqrt{n}\,(t_n-\theta_0)=-\frac{\frac{1}{\sqrt{n}}\sum\psi\,(x_i-\theta_0)}{\frac{1}{n}\sum\psi'\,(x_i-\theta_0)+\frac{\delta_n}{n}}\,.$$
По центральной предельной теореме
$$\frac{1}{\sqrt{n}}\sum\psi\,(x_i-\theta_0)\xrightarrow[n\to\infty]{d}N(0\,,\ \sigma^2)\,,$$
где
$$\sigma^2=\mathbf{E}_\theta(\psi^{\,2}(X-\theta))=\int_{-\infty}^{\infty}\psi^2(x)\,f(x)\,dx\,;$$
$$\frac{1}{n}\sum\psi\,(x_i-\theta_0)\xrightarrow[n\to\infty]{p}\mathbf{E}_\theta(\psi'\,(X-\theta))=\int_{-\infty}^{\infty}\psi'(x)\,f(x)\,dx\,.$$
Тогда
$$\sqrt{n}\,(t_n-\theta_0)\xrightarrow[n\to\infty]{d}N(0\,,\ \sigma^2(F,\psi))\,,$$
$$\sigma^2(F,\psi)=\frac{\int_{-\infty}^{\infty}\psi^2(x)\,f(x)\,dx}{\left(\int_{-\infty}^{\infty}\psi'(x)\,f(x)\,dx\right)^2}.$$
Пусть $(\psi f)_{-\infty}^{\infty}=0$. Тогда
$$(\psi f)'=\psi' f+\psi f'\ \ \Rightarrow\ \ 0=(\psi f)_{-\infty}^{\infty}=\int\psi' f+\int\psi f',$$
откуда $\int\psi' f\,dx=-\int\psi f'\,dx$.

Следовательно,
$$\frac{\int_{-\infty}^{\infty}\psi^2(x)\,f(x)\,dx}{\left(\int_{-\infty}^{\infty}\psi'(x)\,f(x)\,dx\right)^2}=\frac{\int_{-\infty}^{\infty}\psi^2(x)\,f(x)\,dx}{\left(\int_{-\infty}^{\infty}\psi(x)\,f'(x)\,dx\right)^2}.$$
Но
$$\left(\int\psi f'\,dx\right)^2=\left(\int\psi\sqrt{f}\cdot\frac{f'}{\sqrt{f}}\,dx\right)^2\le\int\psi^2 f\,dx\cdot\int\left(\frac{f'}{f}\right)^2 f\,dx\,,\ \ I(f)=\int\left(\frac{f'}{f}\right)^2 f\,dx\,.$$
Значит,
$$\frac{\int_{-\infty}^{\infty}\psi^2(x)\,f(x)\,dx}{\left(\int_{-\infty}^{\infty}\psi'(x)\,f(x)\,dx\right)^2}\ge\frac{1}{I(f)}$$
с равенством т. и т.т., когда $\frac{f'}{\sqrt{f}}=c\cdot\psi\sqrt{f}\quad\Rightarrow\quad\psi=c\frac{f'}{f}=c\,\frac{d}{dx}\ln\,f$.

Таким образом,
$$I(f)\ge\frac{\left(\int_{-\infty}^{\infty}\psi'(x)\,f(x)\,dx\right)^2}{\int_{-\infty}^{\infty}\psi^2(x)\,f(x)\,dx}.$$
Введем
$$I(F)=\sup_{\psi}\frac{\left(\int\psi'\,dF\right)^2}{\int\psi^2\,dF},$$

где супремум берется по множеству всех непрерывно дифференцируемых функций с компактным носителем, удовлетворяющих неравенству $\int \psi^2 dF > 0$.

ТЕОРЕМА 29. *Следующие условия эквивалентны:*
(A1) $I\,(F) < \infty$;
(A2) *распределение* F *имеет абсолютно непрерывную плотность* f,
причем $\int (f'/f)^2 f\,dx < \infty$.
В обоих случаях имеем: $I(F) = \int (f'/f)^2 f\,dx < \infty$.

Д о к а з а т е л ь с т в о. Пусть распределение имеет абсолютно непрерывную плотность, причем $\int (f'/f)^2 f\,dx < \infty$. Возьмем $\psi = f'/f = \dot f$. Тогда $\psi' = \dfrac{d^2}{dx^2}\ln f = \ddot f$ и

$$I(F) = \sup_{\psi} \frac{\left(\int \psi' f\,dx\right)^2}{\int \psi^2 f\,dx} \geq \frac{\left(\int \ddot f\, f\,dx\right)^2}{\int \dot f^2 f\,dx} = \frac{\left(-\int \dot f^2 f\,dx\right)^2}{\int \dot f^2 f\,dx} = \int \dot f^2 f\,dx.$$

С другой стороны,

$$\left(\int \psi' f\,dx\right)^2 = \left(\int \psi f'\,dx\right)^2 \leq \int \dot f^2 f\,dx \cdot \int \psi^2 f\,dx\,,$$

откуда $I(F) \leq \int \dot f^2 f\,dx$.

Из этих неравенств следует, что

$$I(F) = \int \dot f^2 f\,dx.$$

Докажем обратную импликацию. Пусть $I(F) < \infty$. Определим линейный функционал равенством

$$A\psi = -\int \psi' dF \tag{42}$$

на всюду плотном подмножестве гильбертова пространства $L_2(F)$ функций $F-$интегрируемых с квадратом. Норма функционала определяется как

$$\|\,A\,\|^2 = \sup_{\psi} \frac{|\,A\psi\,|^2}{\|\,\psi\,\|^2} = I(F) < \infty.$$

Следовательно, функционал A можно продолжить по непрерывности на все гильбертово пространство $L_2(F)$ и, более того, по теореме Рисса существует такая функция $g \in L_2(F)$, что для всех $\psi \in L_2(F)$,

$$A\psi = \int \psi g\,dF \quad \text{и} \quad \|\,A\,\|^2 = \|\,g\,\|^2 = \int g^2\,dF.$$

Заметим, что

$$Ag = \int g^2\,dF\,, \quad A\,1 = -\int 1' dF = \int g\,dF = 0\,.$$

В этом нетрудно убедиться, исходя из непрерывности A: нужно аппроксимировать 1 гладкими функциями с компактным носителем.

Пока мы еще не знаем, имеет ли распределение F абсолютно непрерывную плотность f, но если это так, то из определения A интегрированием по частям выводится равенство

$$A\psi = -\int \psi' f\,dx = \int \psi f'\,dx = \int \psi(f'/f)f\,dx\,.$$

Следовательно, $g = f'/f$. Поэтому *определим* функцию f равенством

$$f(x) = \int_{y<x} g(y)\, dF(y) \tag{43}$$

и проверим, что так определенная функция действительно есть плотность распределения F.

Применив неравенство Шварца, получим

$$|\, f(x)\,|^2 \le F(x) \int g^2 dF,$$

которая стремится к нулю при $x \to -\infty$. Чтобы показать, что $f(x)$ стремится к нулю и при $x \to +\infty$ нужно использовать и определение и $\int g\, dF = 0$. По теореме Фубини имеем

$$-\int \psi'(x) f(x)\, dx = -\int_{y<x} \int \psi'(x)\, g(y)\, dF(y)\, dx = \int \psi(y) g(y)\, dF(y) = A\,\psi.$$

Сравнивая полученный результат с определением функционала A, заключаем, что $f(x)\, dx$ и $dF(x)$ определяют один и тот же линейный функционал. Поэтому $f(x)\, dx$ и $dF(x)$ определяют одну и ту же меру и, значит, f есть плотность распределения F.

Наконец, имеем:

$$I(F) = \|A\|^2 = \int g^2 dF = \int \left(\frac{f'}{f}\right)^2 f\, dx.$$

7.3. Наилучшие асимптотически нормальные оценки.

Метод максимального правдоподобия может приводить к вычислительным трудностям, которые бывает трудно преодолеть. Численные же методы могут давать дополнительную погрешность вычислений, которая может быть сравнима со случайной погрешностью. Предположим, что мы имеем какую-либо состоятельную оценку (быть может, не самую лучшую) $\tilde\theta_n$. Оказывается, что в таком случае можно построить и асимптотически наилучшую и асимптотически нормальную оценку.

ТЕОРЕМА 30. *Пусть выполнены условия для существования оценки максимального правдоподобия (условия регулярности) и* $n^{1/4}(\tilde\theta_n - \theta) \xrightarrow[n\to\infty]{d} 0$. *Тогда при* $n \to \infty$ *оценка*

$$t_n = \tilde\theta_n - \frac{a_n(\tilde\theta_n)}{B_n(\tilde\theta_n)}, \tag{44}$$

где

$$a_n(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln\ f(X_i\,;\,\theta)\,, \quad B_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2} \ln\ f(X_i\,;\,\theta)\,,$$

является асимптотически наилучшей и асимптотически нормальной оценкой.

Доказательство. Из (15) следует, что

$$\sqrt{n}\,(\,t_n - \theta_0) = \sqrt{n}\,(\,\tilde\theta_n - \theta_0) - \left[\frac{1}{B_n(\theta_0)} + O((\tilde\theta_n - \theta_0))\right] \cdot$$
$$\left[a_n(\theta_0) + \sqrt{n}\, B_n(\theta_0)(\tilde\theta_n - \theta_0) + \sqrt{n}\ O_p((\tilde\theta_n - \theta_0)^2)\right] =$$
$$= -\frac{a_n(\theta_0)}{B_n(\theta_0)} + o(1) + O(\sqrt{n}\,(\tilde\theta_n - \theta_0)^2) = -\frac{a_n(\theta_0)}{B_n(\theta_0)} + o(1),$$

а $\dfrac{a_n(\theta_0)}{B_n(\theta_0)} \xrightarrow[n\,\to\,\infty]{d} \varsigma \in N\left(0\,,\ \dfrac{1}{I(\theta_0)}\right)$.

Пример 55. Пусть X_1, X_2, ... , X_n – повторная выборка из распределения Коши с плотностью

$$f\,(x\,;\,\theta) = f(x-\theta) = \frac{1}{\pi\,(1+(x-\theta)^2)}\,,\ -\infty < x < \infty\,.$$

Требуется оценить параметр $\theta \in R^1$.

Рассмотрим сначала в качестве оценки параметра θ выборочную медиану $m_n = X_n^{([\frac{n}{2}]+1)}$. Здесь $f(x_{0.5}) = f(0) = \dfrac{1}{\pi}\,,\ \dfrac{1}{f^2(0)} = \pi^2$,

$$\sqrt{n}\,(\,m_n-\theta) \xrightarrow{d} N\left(0\,,\ \frac{\pi^2}{4}\right)\,. \tag{45}$$

Вычислим информационное количество $I(\theta)$ для этого распределения. Имеем:

$$I(\theta) = E_\theta\left(\left(\frac{\partial}{\partial\theta}\ \ln\ f(X\,;\,\theta)\right)^2\right) = \frac{4}{\pi}\int_{-\infty}^{+\infty}\frac{(x-\theta)^2\,dx}{(1+(x-\theta)^2)^3} = \frac{4}{\pi}\int_{-\infty}^{+\infty}\frac{x^2\,dx}{(1+x^2)^3} = \frac{4}{\pi}\cdot\frac{\pi}{8} = \frac{1}{2}\,.$$

Заметим, что

$$\frac{\pi^2}{4} \approx 2.4674011\,,\ \ \frac{\pi^2}{4} > 2\ \ \Leftrightarrow\ \ \pi^2 > 2\,.$$

Из (16) видно, что $\sqrt[4]{n}\,(m_n-\theta) \xrightarrow{p} 0$. Отсюда

$$t_n = m_n - \frac{\sum\limits_{i=1}^{n}\frac{x_i-m_n}{1+(x_i-m_n)^2}}{\sum\limits_{i=1}^{n}\frac{(x_i-m_n)^2-1}{(1+(x_i-m_n)^2)^2}}\qquad \text{является н.а.н. оценкой параметра } \theta.$$

Действительно,

$$\ln\ f = -\ln\,\pi - \ln\,\left((1+(x-\theta)^2)\right)\,,\quad \frac{\partial\ln\ f}{\partial\theta} = \frac{2\,(x-\theta)}{1+(x-\theta)^2}\,,$$

$$\frac{\partial^2\ln\ f}{\partial\theta^2} = \frac{2\,((x-\theta)^2-1)}{(1+(x-\theta)^2)^2}\,.$$

7.4. Информационное количество для параметра сдвига и масштаба. Распределения, на которых достигает минимума информационное количество.

В следующей теореме мы покажем, что для параметра сдвига, если известны два момента, то нормальное распределение в определенном смысле является наихудшим, т.е. для нормального распределения количество информации Фишера минимально (рассматривается параметр сдвига θ).

ТЕОРЕМА 31. *Пусть*

1) $f(x)$ – *непрерывно дифференцируемая плотность распределения*;
2) $\int x\,f(x)\,dx = 0$, $\int x^2\,f(x)\,dx = \sigma^2 < \infty$;
3) $|\,x\,|\,f(x) \xrightarrow[x\,\to\,\pm\infty]{} 0$.

Тогда

$$\min_f\ I(f) = \frac{1}{\sigma^2}\ \textit{и достигается на нормальном распределении.}$$

Д о к а з а т е л ь с т в о. Пусть

$$\int x\,f(x)\,dx = 0\,.$$

Поскольку $(xf)' = f + x \cdot f'$, то интегрирование по частям (в силу условия 3) дает

$$\int x\, f'(x)\, dx = -\int f(x)\, dx = -1.$$

Имеем

$$1 = \left(\int x\, f'\, dx\right)^2 = \left(\int x\, \frac{f'}{\sqrt{f}} \cdot \sqrt{f}\, dx\right)^2 \le \int \frac{(f')^2}{f}\, dx \cdot \int x^2 f\, dx\,,$$

откуда $I(f) \ge \dfrac{1}{\sigma^2}$, причем равенство $I(f) = \dfrac{1}{\sigma^2}$ т. и т.т., когда

$$\frac{f'}{f} = cx \quad \Rightarrow \quad f = A \cdot e^{\frac{cx^2}{2}} \quad \Rightarrow \quad c = -\frac{1}{\sigma^2}\,.$$

ТЕОРЕМА 32. *Пусть*
1) $f(x) \ = 0$ *для* $x \le 0$;
2) $\int_0^\infty x\, f(x)\, dx \ = \alpha_1,\ \int_0^\infty x^2 f(x)\, dx \ = \alpha_2$
3) $x^2 f(x) \xrightarrow[x \to \infty]{} 0\,,\ x\, f(x) \xrightarrow[x \to 0]{} 0.$
Тогда

$\min\limits_f\ I(f)$ *и достигается на распределении с плотностью*

$$f(x) = C\, e^{-\lambda x} x^{\nu - 1}\,,\ x > 0.$$

При этом

$$\lambda = \frac{\alpha_1}{\alpha_2 - \alpha_1^2} = \frac{E(X)}{D(X)}\,,\ \nu = \frac{\alpha_1^2}{\alpha_2 - \alpha_1^2} = \frac{E^2(X)}{D(X)}.$$

Д о к а з а т е л ь с т в о (без доказательства).

Пусть теперь $F_t = (1-t)\, F_0 + t\, F_1$, $F_t \in P$ – выпуклое множество. В силу выпуклости распределение $F_0 \in P$ минимизирует информацию Фишера тогда и только тогда, когда $(d/dt) I(F_t) \ \ge 0$ в $t = 0$ для любого распределения $F_1 \in P_1$, где P_1 – множество всех мер, таких, для которых $I(F) < \infty$.

Дифференцирование под знаком интеграла в равенстве

$$I(F_t) = \int [(f')^2/f]\, dx,$$

допустимое в силу теоремы о монотонной сходимости, приводит к неравенству

$$\left[\frac{d}{dt} I(F_t)\right]_{t=0} = \int \left[2\frac{f_0'}{f_0}\,(f_1' - f_0') - \left(\frac{f_0'}{f_0}\right)^2 (f_1 - f_0)\right] dx \ge 0\,. \tag{46}$$

Введем функцию $\psi(x) \ = \ -f_0'(x)/f_0(x)$.Если функция ψ имеет производную ψ' и, значит, возможно интегрирование по частям, то неравенство (46) можно переписать в виде

$$\int (2\psi' - \psi^2)(f_1 - f_0)\, dx \ge 0 \tag{47}$$

или в виде

$$-4 \int [(\sqrt{f_0})''/\sqrt{f_0}\,]\,(f_1 - f_0)\, dx \ge 0 \tag{48}$$

для всех распределений $F_1 \in P_1$.

Пример 56. Пусть $\mathcal{P}$ – множество всех вероятностных распределений F, для которых с некоторой заданной функцией V выполняется неравенство
$\int V(x)\, dF(x) \le 0.$

Для распределения F_0, минимизирующего информацию Фишера на P, в (18) и (16) имеют место равенства. Комбинируя (18), (16) и равенство

$$\int dF(x) = 1,$$

при помощи метода множителей Лагранжа (здесь это множители α и β) получаем дифференциальное уравнение

$$4\,(\sqrt{f_0})''/\sqrt{f_0} - \alpha V + \beta = 0\,, \tag{49}$$

которое при обозначении $u = \sqrt{f_0}$ принимает вид

$$4\,u'' - (\alpha V - \beta)\,u = 0\,. \tag{50}$$

Последнее уравнение есть уравнение Шредингера движения электрона в поле потенциала V.

В частном случае $V(x) = x^2 - 1$ получается известное нам решение. Это решение согласуется с тем фактом, что среди всех распределений с дисперсией, не превосходящей 1, наименьшее значение информации Фишера дает стандартное нормальное распределение.

Другой частный случай:

$$V(x) = \begin{cases} -a < 0 & \text{для } \ |\,x\,| \ \le 1\,, \\ b > 0 & \text{для } \ |\,x\,| \ > 1\,. \end{cases}$$

В этом случае уравнение (12) имеет общее решение

$$f_0(x) = \begin{cases} \dfrac{C}{\cos^2(\omega/2)}\cos^2\left(\dfrac{\omega x}{2}\right) & \text{для } \ |\,x\,| \ \le 1\,, \\ Ce^{\lambda}e^{-\lambda\,|\,x\,|} & \text{для } \ |\,x\,| \ > 1\,, \end{cases}$$

где ω и λ – некоторые постоянные. Для того, чтобы решение было строго положительным следует взять $0 < \omega < \pi$. Функция f_0 является непрерывной. Если требуется, чтобы была непрерывной и функция $\psi = -(\ln\, f_0)'$, то следует положить

$$\lambda = \omega\,\mathrm{tg}\,(\omega/2)$$

и определить C так, чтобы $\int f_0\,dx = 1$:

$$C = \frac{\cos^2(\omega/2)}{1 + 2/[\omega\,\mathrm{tg}\,(\omega/2)]}\,.$$

Заметим, что в этом случае

$$-\,4\frac{(\sqrt{f_0})''}{\sqrt{f_0}} = \begin{cases} \omega^2\,, & |\,x\,| \ \le 1, \\ -\lambda^2\,, & |\,x\,| \ > 1 \end{cases}$$

и, следовательно,

$$I(F_0) = -\,4\int \frac{(\sqrt{f_0})''}{\sqrt{f_0}} f_0\,dx = \frac{\omega^2}{1 + 2/[\omega\,\mathrm{tg}\,(\omega/2)]}.$$

Нетрудно убедиться, что самостоятельно, что неравенство (48) выполнено.

Пример 57. Пусть G – фиксированное вероятностное распределение с такой дважды дифференцируемой плотностью g, что функция $-\ln g(x)$ выпукла на выпуклом носителе распределения G. Зададим $\varepsilon > 0$. Пусть P есть множество всех вероятностных мер на действительной прямой.

Пусть x_0 и x_1 ($x_0 < x_1$) суть концы интервала, на котором выполняется неравенство $|\,g'/g| \ \le k$, а величина k связана со значением ε соотношением

$$\int_{x_0}^{x_1} g(x)\,dx + \frac{g(x_0)+g(x_1)}{k} = \frac{1}{1-\varepsilon}\,, \tag{51}$$

Теперь положим

$$f_0(x) = \begin{cases} (1-\varepsilon)\,g(x_0)\,e^{k(x-x_0)}, & x \le x_0\,, \\ (1-\varepsilon)\,g(x), & x_0 < x < x_1\,, \\ (1-\varepsilon)\,g(x_0)\,e^{-k(x-x_1)}, & x \ge x_1\,. \end{cases}$$

Из условия (51) следует, что интеграл от $f_0(x)$ равен 1. Неотрицательность $f_0(x)$ следует, из того что выпуклая функция $-\ln g(x)$ лежит над своими касательными в точках x_0 и x_1, т.е. $g(x) \le g(x_0)\,e^{k(x-x_0)}$ и $g(x) \le g(x_1)\,e^{-k(x-x_1)}$. Как сама функция $f_0(x)$, так и ее производные непрерывны. Имеем:

$$\psi(x) = \begin{cases} -[\ln\, f_0(x)]' = -k\,, & x \le x_0\,, \\ -g'(x)/g(x), & x_0 < x < x_1\,, \\ k\,, & x \ge x_1. \end{cases}$$

Теперь проверим, что неравенство (10) выполняется в данном случае. Поскольку $\psi'(x) \ge 0$ и $k^2+2\psi'-\psi^2 \ge 0$ при $x_0 \le x \le x_1$, причем в остальных случаях $k^2+2\psi'-\psi^2 = 0$, имеем:

$$\int (2\psi'-\psi^2)(f_1-f_0)\,dx = \int_{x_0}^{x_1} (k^2+2\psi'-\psi^2)(f_1-f_0)\,dx - k^2 \int_{x_0}^{x_1} (f_1-f_0)\,dx \ge 0$$

в силу того, что $f_1 \ge f_0$ на интервале $x_0 < x < x_1$, и того, что $\int (f_1-f_0)\,dx \le 0$.

В случае стандартного нормального распределения

$$f_0(x) = \begin{cases} \dfrac{1-\varepsilon}{\sqrt{2\pi}} e^{-x^2/2}, & |\,x\,| \ \le k\,, \\ \dfrac{1-\varepsilon}{\sqrt{2\pi}} e^{-x^2/2-k\,|\,x\,|}, & |\,x\,| \ > k\,, \end{cases}$$

где величины k и ε связаны соотношением

$$2\varphi(k)/k - 2\Phi(-k) = \varepsilon/(1-\varepsilon),$$

$$\varphi(x) = \Phi'(x)\,,\ \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right)\,dt.$$

Глава 8.

Современные методы прикладной теории оценивания.

8.1. Оценки как непрерывные функционалы отношения правдоподобия.

Пусть $\theta \in \Theta \subseteq \mathbf{R}^m$, а $f(x;\theta)$ – плотность распределения. Рассмотрим байесовскую оценку $\tilde{\theta}_n$. Возьмем неотрицательную функцию $\pi(\theta)$ и определим апостериорную плотность

$$p_n(y) = \frac{\pi(y)\prod\limits_{i=1}^{n} f(x_i;y)}{\int_\Theta \pi(z)\prod\limits_{i=1}^{n} f(x_i;z)\,dz}.$$

Определим *байесовскую* оценку равенством

$$\tilde{\theta}_n = \int_\Theta y p_n(y)\,dy. \tag{52}$$

Сделаем в (1) замену переменной вида $y = \theta_0 + \dfrac{\theta}{\mu(n)}$.

Назовем функцию $\mu(n)$ *правильным нормирующим множителем*, если распределения $\dfrac{1}{\mu(n)}p_n\left(\theta_0 + \dfrac{\theta}{\mu(n)}\right)$ и $\mu(n)(\tilde{\theta}_n - \theta_0)$ сходятся к собственным предельным.

Рассмотрим нормированную функцию правдоподобия

$$\mathcal{Z}_n(\theta) = \frac{\prod\limits_{i=1}^{n} f(x_i;\theta_0+\theta/\mu(n))}{\prod\limits_{i=1}^{n} f(x_i;\theta_0)} = \exp(\mathcal{Y}_n(\theta))$$

и будем трактовать её как случайную функцию θ.

Тогда

$$\mu(n)(\tilde{\theta}_n - \theta_0) = \frac{\int_{\mathbf{R}^m} z\exp(\mathcal{Y}_n(z))\pi\left(\theta_0+\dfrac{\theta}{\mu(n)}\right)\,dz}{\int_{\mathbf{R}^m} \exp(\mathcal{Y}_n(z))\pi\left(\theta_0+\dfrac{\theta}{\mu(n)}\right)\,dz}. \tag{53}$$

При определенных условиях конечномерные распределения $\exp(\mathcal{Y}_n(\theta))$ сходятся к конечномерным распределениям некоторой функции $\exp(\mathcal{Y}(\theta))$.

Если к тому же функция $\pi(\theta)$ непрерывна в окрестности θ_0, то предельные распределения $\mu(n)(\tilde{\theta}_n - \theta_0)$ не зависят от π и совпадают с распределением величины

$$\frac{\int_{\mathbf{R}^m} z\exp(\mathcal{Y}(z))\,dz}{\int_{\mathbf{R}^m} \exp(\mathcal{Y}(z))\,dz}.$$

Предположим, что функция $f(x;\theta)$ абсолютно непрерывна по θ и существует информационное количество Фишера

$$I(\theta) = \int\limits_{-\infty}^{\infty} \frac{|\,f'(x;\theta)\,|^2}{f(x;\theta)}\,dx < \infty.$$

В этом случае $\mu(n) = \sqrt{n}$, а предельный процесс выглядит особенно просто:

$$\mathcal{Y}(\theta) = \sqrt{I(\theta_0)}\xi\cdot\theta - \frac{\theta^2}{2}I(\theta_0),$$

где $\xi \in N(0,1)$.

Формально, переходя к пределу в (2), получаем

$$\frac{\int_{\mathbf{R}^m} z \exp(\mathcal{Y}(z))\, dz}{\int_{\mathbf{R}^m} \exp(\mathcal{Y}(z))\, dz} = \frac{\xi}{\sqrt{I(\theta_0)}} \in N\left(0, \frac{1}{I(\theta_0)}\right).$$

Основным объектом исследования является случайная функция – логарифм отношения правдоподобия:

$$\mathcal{Y}_n(\theta) = \sum_{i=1}^{n} \ln \frac{f\left(x_i; \theta_0 + \frac{\theta}{\sqrt{n}}\right)}{f(x_i;\theta_0)} = 2\sum_{i=1}^{n} \ln \left(1 + \frac{\sqrt{f\left(x_i; \theta_0 + \frac{\theta}{\sqrt{n}}\right)} - \sqrt{f(x_i;\theta_0)}}{\sqrt{f(x_i;\theta_0)}}\right) \sim$$

$$\sim 2\sum_{i=1}^{n} \frac{\sqrt{f\left(x_i; \theta_0 + \frac{\theta}{\sqrt{n}}\right)} - \sqrt{f(x_i;\theta_0)}}{\sqrt{f(x_i;\theta_0)}} - \sum_{i=1}^{n} \left(\frac{\sqrt{f\left(x_i; \theta_0 + \frac{\theta}{\sqrt{n}}\right)} - \sqrt{f(x_i;\theta_0)}}{\sqrt{f(x_i;\theta_0)}}\right)^2.$$

Таким образом, исследование случайной функции $\mathcal{Y}_n(\theta)$ сводится к исследованию двух последних слагаемых.

8.2. Оценки байесовского типа.

Пусть $X = X_1, X_2, \ldots, X_n$ – выборка из генеральной совокупности, $\mathcal{P} = \{f_n(x;\theta), \theta \in \Theta \subset R^1\}$ – семейство плотностей распределения X по σ-конечной мере ν и задано некоторое семейство $\mathcal{G} = \{g_n(x;\theta), \theta \in \Theta\}$, в общем случае не совпадающее с $\mathcal{P}$. Введем *оценку байесовского типа* в виде

$$\hat{\theta}_n(X) = \int_{\Theta} \theta\pi(\theta) g_n(X;\theta)\, d\theta \left(\int_{\Theta} \pi(\theta) g_n(X;\theta)\, d\theta\right)^{-1},$$

где $\pi(\theta)$ – неотрицательная измеримая функция θ.

Обозначим

$$\Phi_n(X;\theta) = g_n(X;\theta)/f_n(X;\theta).$$

Пусть $\pi(\theta)$ – плотность априорного распределения на Θ. Если $g_n(X;\theta)$ – плотность распределения, то оценка $\hat{\theta}_n(X)$ минимизирует следующий функционал

$$\int_{-\infty}^{\infty} \int_{\Theta} (\,\hat{\theta}_n(x) - \theta\,)^2 g_n(X;\theta)\pi(\theta)\, d\nu(x) d\theta = \mathbf{E}_{f_n}(\,[\,\hat{\theta}_n(X) - \theta\,]^2 \Phi_n(X;\theta)\,).$$

ТЕОРЕМА 33. *В классе оценок, для которых*

$$\rho(\pi, \delta_n) = \mathbf{E}(\,[\,\delta_n(X) - \theta\,]^2 \Phi_n(X;\theta) \mathbf{E}(\Phi_n(X;\theta)|X)\,) < \infty,$$

имеет место соотношение

$$\rho(\pi, \hat{\theta}_n) = \min_{\delta_n} \rho(\pi, \delta_n).$$

Д о к а з а т е л ь с т в о. Отметим прежде всего, что

$$\hat{\theta}_n(X) = \mathbf{E}(\theta\Phi_n(X;\theta)|X)/\mathbf{E}(\Phi_n(X;\theta)|X),$$

где мы полагаем отношение $g_n(x;\theta)/f_n(x;\theta)$ равным нулю, если $f_n(x;\theta)=0$. Правая часть этого равенства определена в силу условия предложения. Используя свойства условного математического ожидания, получим

$$\mathbf{E}(\,[\,\delta_n(X)-\theta\,]^2\Phi_n(X;\theta)\mathbf{E}(\Phi_n(X;\theta)|X)\,)=\mathbf{E}(\,[\,\delta_n(X)-\hat\theta_n(X)+\hat\theta_n(X)-\theta\,]^2\Phi_n(X;\theta)\mathbf{E}(\Phi_n(X;\theta)|X)\,)=$$

$$=\mathbf{E}(\,[\,\delta_n(X)-\hat\theta_n(X)\,]^2\Phi_n(X;\theta)\mathbf{E}(\Phi_n(X;\theta)|X)\,)+\mathbf{E}(\,[\hat\theta_n(X)-\theta\,]^2\Phi_n(X;\theta)\mathbf{E}(\Phi_n(X;\theta)|X)\,)+$$
$$+2\mathbf{E}(\,[\,(\delta_n(X)-\hat\theta_n(X))(\hat\theta_n(X)-\theta)\,]\Phi_n(X;\theta)\mathbf{E}(\Phi_n(X;\theta)|X)\,).$$

Рассмотрим первое слагаемое в правой части последнего равенства. Имеем

$$\mathbf{E}(\,[\,\delta_n(X)-\hat\theta_n(X)\,]^2\Phi_n(X;\theta)\mathbf{E}(\Phi_n(X;\theta)|X)\,)=\mathbf{E}\{\mathbf{E}(\,[\,\delta_n(X)-\hat\theta_n(X)\,]^2\Phi_n(X;\theta)\mathbf{E}(\Phi_n(X;\theta)|X)\,)|X\}=$$
$$=\mathbf{E}(\,[\,\delta_n(X)-\hat\theta_n(X)\,]^2\mathbf{E}^2(\Phi_n(X;\theta)|X)\,)\geqslant 0.$$

Теперь рассмотрим третье слагаемое в той же сумме. По свойству условных ожиданий

$$\mathbf{E}(\,[\,(\delta_n(X)-\hat\theta_n(X))(\hat\theta_n(X)-\theta)\,]^2\Phi_n(X;\theta)\mathbf{E}(\Phi_n(X;\theta)|X)\,)=$$
$$=\mathbf{E}\{\mathbf{E}(\,[\,(\delta_n(X)-\hat\theta_n(X))(\hat\theta_n(X)-\theta)\,]^2\Phi_n(X;\theta)\mathbf{E}(\Phi_n(X;\theta)|X)\,)|X\}.$$

Поскольку

$$\mathbf{E}\,[\,(\theta\Phi_n(X;\theta)-\hat\theta_n(X)\mathbf{E}(\Phi_n(X;\theta)|X)\,)|X]=0,$$

то

$$\mathbf{E}\{\mathbf{E}(\,[\,(\delta_n(X)-\hat\theta_n(X))(\hat\theta_n(X)-\theta)\,]\Phi_n(X;\theta)\mathbf{E}(\Phi_n(X;\theta)|X)\,)|X\}=0.$$

Окончательно получаем

$$\rho(\pi,\delta_n)=\rho(\pi,\hat\theta_n)+\mathbf{E}[(\hat\theta_n(X)-\delta_n(X))^2\mathbf{E}^2(\Phi_n(X;\theta)|X)]\geqslant\rho(\pi,\hat\theta_n),$$

что доказывает теорему 33.

Суть доказательства теоремы 33 состоит в том, что оценка, минимизирующая апостериорный риск, минимизирует и полный риск. Множитель $\mathbf{E}(\Phi_n(X;\theta)|X)$ введен для того, чтобы в случае общих функций $g_n(x;\theta)$ (а не только неотрицательных) риск $\rho(\pi,\delta_n(X))$ был бы неотрицательным (это очевидно, если $g_n(x;\theta)\geqslant 0$). Неотрицательность риска $\rho(\pi,\hat\theta_n(X))$ доказывается применением неравенства Коши-Буняковского.

Пусть заданы: Θ – открытый интервал в $\mathbf{R}^1$ значений параметра θ; статистическая структура
$\{\mathcal{X},\mathcal{A},\mathbf{P}_\theta,\theta\in\Theta\}$, где $\mathcal{X}$ – выборочное пространство, $\mathcal{A}$ – σ-алгебра подмножеств $\mathcal{X}$, $\mathbf{P}_\theta$ – вероятностное распределение на $\{\mathcal{X},\mathcal{A}\}$. Обозначим через $\mathcal{A}_n$ – σ-алгебру, порожденную набором случайных величин $X_0,X_1,...,X_n$, а через $\mathbf{P}_{n,\theta}$ – сужение $\mathbf{P}_\theta$ на $\mathcal{A}_n$ – σ-подалгебре $\mathcal{A}$. Предположим, что при каждом $n\geqslant 0$ семейство $\{\mathbf{P}_{n,\theta},\theta\in\Theta\}$ доминировано σ-конечной мерой ν и $d\mathbf{P}_{n,\theta}/d\nu(x)=f_n(x;\theta)$.

Пусть существуют такие функции $g(x_o;\theta),g_j(x_j|x_{j-1};\theta),j\geqslant 1$, что

$$g_n(x;\theta)=g(x_o;\theta)\prod_{j=1}^{n}g_j(x_j|x_{j-1};\theta).$$

Определим

$$\varphi_0(\theta,\theta^*)=\frac{g(X_0;\theta^*)}{g(X_0;\theta)},\quad \varphi_j(\theta,\theta^*)=\frac{g_j(X_j|X_{j-1};\theta^*)}{g_j(X_j|X_{j-1};\theta)},\quad j\geqslant 1.$$

УСЛОВИЯ /8.1/
/8.1.0/ Для любых $x,\theta,n\geqslant 0$

$$\int\limits_{-\infty}^{\infty}...\int\limits_{-\infty}^{\infty}g_n(x;\theta)\,d\nu(x)<\infty.$$

/8.1.1/ Равномерно по θ из компактных подмножеств Θ

$$\inf_{t:|\theta-t|>\delta>0} \sup_{A\in\mathcal{A}_n} |\mathbf{P}_{n,\theta}(A)-\mathbf{P}_{n,t}(A)|>0, \quad \forall n\geqslant 0.$$

/8.1.2/ Для каждого $\theta\in\Theta$ определен марковский процесс $\{X_n\}, n\geqslant 0$.

/8.1.3/ (i) Для каждого $\theta\in\Theta$ имеет место разложение

$$\varphi_j(\theta,\theta+h)=1+h\dot\varphi_j(\theta)+(1/2)h^2\ddot\varphi_j(\theta)+h^2\delta_{h,j},$$

где $\dot\varphi_j(\theta)$ – производная в ср. кв. от $\varphi_j(\theta,\theta^*)$ по θ^* в точке (θ,θ) относительно вероятностной меры $\mathbf{P}_\theta$, $\ddot\varphi_j(\theta)$ – вторая производная в ср. кв. от $\varphi_j(\theta,\theta^*)$ по θ^*,

$$\delta_{h,j}\xrightarrow[h\to 0]{L_2[\mathbf{P}_\theta]}0 \quad \text{для каждого} \quad j;$$

(ii) Функции $\varphi_j(\theta,\theta^*),\dot\varphi_j(\theta),\ddot\varphi_j(\theta)$ $\mathcal{A}_1\times\mathcal{B}$-измеримы, где $\mathcal{B}$ – σ-алгебра борелевских множеств на $\mathbf{R}^1$, $j\geqslant 1$, а $\varphi_0(\theta,\theta^*)$ $\mathcal{A}_0\times\mathcal{B}$-измерима.

/8.1.4/ Функции

$$\Gamma_j^2(\theta)=\mathbf{E}\,[\,\dot\varphi_j(\theta\,]^2<\infty \quad \text{и} \quad m_j(\theta)=\mathbf{E}\,[\,\ddot\varphi_j(\theta\,]^2<\infty$$

таковы, что для всех θ

$$\text{(i)}\quad \frac{1}{n}\sum_{j=1}^{n}\Gamma_j^2(\theta)\underset{n\to\infty}{\longrightarrow}\Gamma_0^2(\theta);\qquad \text{(ii)}\quad \frac{1}{n}\sum_{j=1}^{n}m_j(\theta)\underset{n\to\infty}{\longrightarrow}m_0^2(\theta).$$

/8.1.5/ Для любых $\theta\in\Theta$ и $\varepsilon\in(0,\infty)$ при $n\to\infty$,

$$\text{(i)}\quad \frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta\{\dot\varphi_j^2(\theta)I(|\dot\varphi_j(\theta)|>\varepsilon n^{1/2})\}=o(1),$$

$$\text{(ii)}\quad \frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta\{\ddot\varphi_j(\theta)I(|\ddot\varphi_j(\theta)|>\varepsilon n^{1/2})\}=o(1),$$

$$\text{(iii)}\quad \frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta\{\ddot\varphi_j(\theta)I(|\ddot\varphi_j(\theta)|\leqslant\varepsilon n^{1/2})\}\leqslant M(\theta).$$

/8.1.6/ Для каждого $\theta\in\Theta$ ($\ln x$ обозначает главное значение и по определению $0\cdot\ln 0=0$)

$$\ln\varphi_0(\theta,\theta+h)=h\xi_n(\theta)-(1/2)\tau_n^2(\theta)h^2+o(h^2), \qquad (h\to 0),$$

где

$$\frac{\xi_n(\theta)}{\sqrt{n}}\xrightarrow[n\to\infty]{d}N(0,\gamma^2(\theta)), \qquad \frac{1}{n}\tau_n^2(\theta)\underset{n\to\infty}{\longrightarrow}\tau^2(\theta)$$

причем

$$m(\theta)=\tau^2(\theta)+m_0(\theta)>0, \qquad \Gamma^2(\theta)=\gamma^2(\theta)+\Gamma_0^2(\theta)>0.$$

/8.1.7/ $\qquad \mathbf{E}_\theta(\,\dot\varphi_j(\theta)|\mathcal{A}_{j-1}\,)=0 \qquad \mathbf{P}_\theta$ – п.н. $\quad j=1,2,...,n.$

Положим

$$\theta_n=\theta+\frac{h}{\sqrt{n}}, \quad \Lambda_n(\theta,\theta_n)=\ln\varphi_0(\theta,\theta_n)+\sum_{j=1}^{n}\ln\varphi_j(\theta,\theta_n), \quad \Delta_n(\theta)=hn^{-1/2}\sum_{j=1}^{n}\dot\varphi_j(\theta).$$

ТЕОРЕМА 34. *Пусть выполнены условия* /8.1.1./ – /8.1.7/, $h\in\mathbf{R}^1$. *Тогда при* $n\to\infty$

(i) $\qquad \Lambda_n(\theta,\theta_n)-h\Delta_n(\theta)+(1/2)h^2m(\theta)-h\xi_n(\theta))-(1/2)\tau^2(\theta)h^2\xrightarrow{\mathbf{P}_\theta}0.$

(ii) $\qquad \Delta_n \xrightarrow[n\to\infty]{d} N(0, \Gamma_0^2(\theta)).$

Д о к а з а т е л ь с т в о теоремы 34 длинно, поэтому оно разбито на промежуточные леммы, в которых мы будем предполагать, что выполнены условия /8.1.1./ – /8.1.7/.

Лемма 5. *Пусть* $\varphi_{nj} = \varphi_j(\theta, \theta_n)$. *Тогда для любого* $\varepsilon > 0$,

$$\mathbf{P}_\theta \left\{ \max_{1\leqslant j\leqslant n} |\varphi_{nj} - 1| > \varepsilon \right\} \underset{n\to\infty}{\longrightarrow} 0.$$

Д о к а з а т е л ь с т в о. Из определения максимума и полуаддитивности вероятности имеем

$$\mathbf{P}_\theta \left\{ \max_{1\leqslant j\leqslant n} |\varphi_{nj} - 1| > \varepsilon \right\} \leqslant \sum_{j=1}^{n} \mathbf{P}_\theta \left\{ |\varphi_{nj} - 1| > \varepsilon \right\}.$$

Кроме того, из условия /8.1.3/ (i) следует, что

(54) $$\varphi_{nj} - 1 = hn^{-1/2}\dot{\varphi}_j + n^{-1/2}R_{nj},$$

где

$$R_{nj} = R_{nj}(\theta, h) = n^{1/2}(\varphi_{nj} - 1) - h\dot{\varphi}_j, \quad \dot{\varphi}_j = \dot{\varphi}_j(\theta).$$

По этому условию

(55) $$\mathbf{E}_\theta |R_{nj}|^2 = \mathbf{E}_\theta |n^{1/2}(\varphi_{nj} - 1) - h\dot{\varphi}_j|^2 \underset{n\to\infty}{\longrightarrow} 0.$$

Ввиду соотношения (54), получаем

$$\mathbf{P}_\theta\{|\varphi_{nj}-1| > \varepsilon\} = \mathbf{P}_\theta\{|n^{-1/2}h\dot{\varphi}_j + n^{-1/2}R_{nj}| > \varepsilon\} \leqslant \mathbf{P}_\theta\{|h\dot{\varphi}_j| > \frac{\varepsilon}{2}n^{1/2}\} + \mathbf{P}_\theta\{|R_{nj}|\frac{\varepsilon}{2}n^{1/2}\}.$$

По неравенству Чебышева имеем

$$\sum_{j=1}^{n} \mathbf{P}_\theta\{|R_{nj}| > \frac{\varepsilon}{2}n^{1/2}\} \leqslant \frac{4}{\varepsilon^2 n} \sum_{j=1}^{n} \mathbf{E}_\theta |R_{nj}|^2.$$

Из (55) и леммы Теплица следует, что

$$\frac{1}{n}\sum_{j=1}^{n} \mathbf{E}_\theta |R_{nj}|^2 \underset{n\to\infty}{\longrightarrow} 0.$$

Далее,

$$\sum_{j=1}^{n} \mathbf{P}_\theta\{|h\dot{\varphi}_j| > \frac{\varepsilon}{2}n^{1/2}\} = \sum_{j=1}^{n} \mathbf{E}_\theta\{I(|h\dot{\varphi}_j| > \frac{\varepsilon}{2}n^{1/2}\} =$$

$$= \sum_{j=1}^{n} \frac{4}{n\varepsilon^2} \mathbf{E}_\theta\{(\frac{\varepsilon}{2}n^{1/2})^2 I(|h\dot{\varphi}_j| > \frac{\varepsilon}{2}n^{1/2}\} \leqslant \frac{4h^2}{n\varepsilon^2} \sum_{j=1}^{n} \mathbf{E}_\theta\{(\dot{\varphi}_j I(|h\dot{\varphi}_j| > \frac{\varepsilon}{2}n^{1/2}/2\}$$

и из условия /8.1.5/ (i) получаем, что правая часть последнего соотношения стремится к нулю при $n \to \infty$. Это завершает доказательство леммы 5.

Лемма 6.

$$\sum_{j=1}^{n} (\varphi_{nj} - 1)^2 \xrightarrow[n\to\infty]{\mathbf{P}_\theta} h^2\Gamma_0^2(\theta).$$

Д о к а з а т е л ь с т в о. Так как $\varphi_j(\theta, \theta) = 1$, то условие /8.1.3/ (i) означает, что

(56) $$n^{1/2}(\varphi_j(\theta, \theta_n) - 1) \xrightarrow[n\to\infty]{L_2[\mathbf{P}_\theta]} h\dot{\varphi}_j(\theta).$$

Обозначим $\zeta_n = n^{1/2}(\varphi_j(\theta, \theta_n) - 1)$ и $\zeta = h\dot{\varphi}_j(\theta)$. По неравенству Чебышева

$$\mathbf{P}_\theta(|\zeta_n - \zeta| > \varepsilon) \leqslant \varepsilon^{-2}\mathbf{E}_\theta(\zeta_n - \zeta)^2,$$

поэтому соответствие (56) эквивалентно сходимости

$$\mathbf{E}_\theta(\zeta_n-\zeta)^2 \underset{n\to\infty}{\longrightarrow} 0,$$

что влечет сходимость ζ_n к ζ по $\mathbf{P}_\theta$-вероятности, т.е.

$$\zeta_n^2 \xrightarrow[n\to\infty]{\mathbf{P}_\theta} \zeta^2. \tag{57}$$

Неравенство

$$\zeta_n^2=(\zeta_n-\zeta+\zeta)^2\leqslant 2(\zeta_n-\zeta)^2+2\zeta^2$$

влечет соответствующее неравенство для математических ожиданий:

$$\mathbf{E}_\theta(\zeta_n^2)\leqslant 2\,\mathbf{E}_\theta((\zeta_n-\zeta)^2)+2\,\mathbf{E}_\theta(\zeta^2).$$

Пусть $\varepsilon>0$. Тогда найдется $n(\varepsilon)$, что для $n>n(\varepsilon)$, будет выполнено неравенство $\mathbf{E}_\theta((\zeta_n-\zeta)^2)<\varepsilon/2$. Поэтому для $n>n(\varepsilon)$ имеем

$$\mathbf{E}_\theta(\zeta_n^2)\leqslant \varepsilon+2\,\mathbf{E}_\theta(\zeta^2).$$

Соотношение (57) позволяют сделать вывод о том, что

$$\mathbf{E}_\theta(\zeta_n^2) \underset{n\to\infty}{\longrightarrow} \mathbf{E}_\theta(\zeta^2). \tag{58}$$

Рассмотрим $(\zeta^2-\zeta_n^2)$. Из (1.1.4) получаем, что $(\zeta^2-\zeta_n^2)_+ \xrightarrow[n\to\infty]{\mathbf{P}_\theta} 0$ и так как $(\zeta^2-\zeta_n^2)_+\leqslant \zeta^2$, то

$$\mathbf{E}_\theta[(\zeta^2-\zeta_n^2)_+] \underset{n\to\infty}{\longrightarrow} 0. \tag{59}$$

Из (58) и (59) вытекает, что

$$\mathbf{E}_\theta[(\zeta^2-\zeta_n^2)_-] \underset{n\to\infty}{\longrightarrow} 0. \tag{60}$$

Соотношения (59) и (60) эквивалентны тому, что

$$fE_\theta|\zeta_n^2-\zeta^2| \underset{n\to\infty}{\longrightarrow} 0. \tag{61}$$

Используя неравенство Чебышева, получаем, что для любого $\varepsilon>0$

$$\mathbf{P}_\theta\{|\sum_{j=1}^n(\varphi_j(\theta,\theta_n)-1)^2-\frac{1}{n}\sum_{j=1}^n(h\dot\varphi_j(\theta))^2|>\varepsilon\}\leqslant\frac{1}{\varepsilon n}\sum_{j=1}^n\mathbf{E}_\theta\{|n(\varphi_j(\theta,\theta_n)-1)^2-(h\dot\varphi_j(\theta))^2|\}.$$

Из условия /8.1.3/ (i) и леммы Теплица следует, что последнее выражение стремится к нулю. Таким образом,

$$\sum_{j=1}^n(\varphi_j(\theta,\theta_n)-1)^2-\frac{1}{n}\sum_{j=1}^n(h\dot\varphi_j(\theta))^2 \xrightarrow[n\to\infty]{\mathbf{P}_\theta} 0.$$

Покажем теперь, что

$$\lim_{n\to\infty}\mathbf{P}_\theta\{|\frac{1}{n}\sum_{j=1}^n(h^2\dot\varphi_j^2-h^2\mathbf{E}_\theta(\dot\varphi_j^2))|>\varepsilon\}=0. \tag{62}$$

Пусть $b_1,b_2,...,b_n$ – действительные числа. Тогда

$$|\exp(i\sum_{j=1}^n b_j)-1|=|\sum_{k=1}^n[\exp(i\sum_{j=k}^n b_j)-\exp(i\sum_{j=k+1}^n b_j)]|\leqslant\sum_{k=1}^n|\exp(ib_k)-1|, i=\sqrt{-1}, \tag{63}$$

где мы полагаем по определению $\sum_{j=n+1}^n b_j=0$.

Положим $\mu_j = \mathbf{E}_\theta(\dot\varphi_j^2)$. Если $\mathcal{F}_{j\theta}(x)$ – функция распределения величины $(\dot\varphi_j^2 - \mu_j)$, то применение неравенства (63) приводит к следующим оценкам:

$$|\,\mathbf{E}_\theta \exp\{ith^2n^{-1}\sum_{j=1}^{n}(\dot\varphi_j^2-\mu_j)\}-1| \leqslant \sum_{j=1}^{n}|\int\limits_{-\infty}^{\infty}(e^{ith^2n^{-1}x}-1)\,d\mathcal{F}_{j\theta}(x)| =$$

$$=\sum_{j=1}^{n}|\int\limits_{|x|\leqslant\varepsilon n}(e^{ith^2n^{-1}x}-1-ith^2xn^{-1})\,d\mathcal{F}_{j\theta}(x)|+\sum_{j=1}^{n}|\int\limits_{|x|>\varepsilon n}(e^{ith^2n^{-1}x}-1-ith^2xn^{-1})\,d\mathcal{F}_{j\theta}(x)| \leqslant$$

$$\leqslant \varepsilon h^4t^2n^{-1}\sum_{j=1}^{n}\mathbf{E}_\theta(\dot\varphi_j^2)+2|t|h^2n^{-1}\sum_{j=1}^{n}\int\limits_{|x|>\varepsilon n}|x|\,d\mathcal{F}_{j\theta}(x).$$

В силу условий /8.1.4/ (i) и /8.1.5/ (i) правую часть последнего неравенства можно сделать меньше заданного $\delta > 0$. Из соотношений (61) и (62) следует результат леммы.

Лемма 7. *При* $n \to \infty$

$$\Lambda_n(\theta,\theta_n)-\sum_{j=1}^{n}(\varphi_{nj}-1)+\frac{1}{2}\sum_{j=1}^{n}(\varphi_{nj}-1)^2-\frac{h\xi_n(\theta)}{\sqrt{n}}-\frac{h^2}{2n}\tau_n(\theta)\xrightarrow{\mathbf{P}_\theta}0.$$

Д о к а з а т е л ь с т в о. Обозначим для данного $\varepsilon > 0$ событие

$$A_n = \{\max_{1\leqslant j\leqslant n}|\varphi_j(\theta,\theta_n)-1|>\varepsilon\}.$$

По лемме 5 для $n > n_0$ – достаточно большого, вероятность $\mathbf{P}_\theta(A_n^c) > 1-\varepsilon$. Тогда на A_n^c имеем:

(64)
$$\sum_{j=1}^{n}\ln\varphi_j(\theta,\theta_n)=\sum_{j=1}^{n}\ln(1+\varphi_j(\theta,\theta_n)-1)=\sum_{j=1}^{n}(\varphi_j(\theta,\theta_n)-1)-\frac{1}{2}\sum_{j=1}^{n}(\varphi_j(\theta,\theta_n)-1)^2+B_n,$$

где

$$B_n=\sum_{j=1}^{n}\sum_{s=1}^{n}\frac{(-1)^{s-1}}{s}(\varphi_j(\theta,\theta_n)-1)^s. \tag{65}$$

Можно считать, что $|\,\varphi_j(\theta,\theta_n)-1| \leqslant 2/3$. В таком случае из равенства (65) следует, что

$$|B_n|\leqslant\sum_{j=1}^{n}\sum_{s=1}^{n}\frac{1}{3}(\varphi_j(\theta,\theta_n)-1)^s=\frac{1}{3}\sum_{j=1}^{n}\frac{|\,\varphi_j(\theta,\theta_n)-1|^3}{1-|\,\varphi_j(\theta,\theta_n)-1|}\leqslant\sum_{j=1}^{n}|\,\varphi_j(\theta,\theta_n)-1|^3\leqslant$$

$$\leqslant\max_{1\leqslant j\leqslant n}|\,\varphi_j(\theta,\theta_n)-1|\sum_{j=1}^{n}|\,\varphi_j(\theta,\theta_n)-1|^2,$$

поэтому в силу лемм 5 и 6,

$$|B_n|\xrightarrow[n\to\infty]{\mathbf{P}_\theta}0. \tag{66}$$

Из соотношений (64) и (66) получаем

$$\sum_{j=1}^{n}\ln\varphi_j(\theta,\theta_n)-\sum_{j=1}^{n}(\varphi_{nj}-1)+\frac{1}{2}\sum_{j=1}^{n}(\varphi_{nj}-1)^2\xrightarrow[n\to\infty]{\mathbf{P}_\theta}0. \tag{67}$$

Кроме того, из условия /8.1.6/ следует, что

$$\Lambda_n(\theta,\theta_n)-\sum_{j=1}^{n}\ln(\varphi_j(\theta,\theta_n)-\frac{h\xi_n(\theta)}{\sqrt{n}}-\frac{h^2\tau^2(\theta)}{2}=o(1). \tag{68}$$

Утверждение леммы 7 следует теперь из соотношений (67) и (68).

Положим

$$\psi_j(\theta,\theta^*)=\mathbf{E}_\theta(\,\varphi_j(\theta,\theta^*)|\mathcal{A}_{j-1}),\quad j=1,2,...,n.$$

Лемма 7а. *При* $n\to\infty$

(i) $$\sum_{j=1}^{n}[\,\psi_j(\theta,\theta_n)-1]\xrightarrow{\mathbf{P}_\theta}h^2m(\theta)/2,$$

(ii) $$\sum_{j=1}^{n}[\,\varphi_j(\theta,\theta_n)-1]-n^{-1/2}h\dot\varphi_j(\theta)-[\,\psi_j(\theta,\theta_n)-1]\xrightarrow{\mathbf{P}_\theta}0.$$

Д о к а з а т е л ь с т в о. (i) По условию /8.1.3/ имеет место разложение

$$\varphi_j(\theta,\theta_n)-1=\frac{h\dot\varphi_j(\theta)}{\sqrt{n}}+\frac{h^2\ddot\varphi_j(\theta)}{2n}+\frac{r_{nj}}{n}6 \tag{69}$$

где $r_{nj}\overset{L_2[\mathbf{P}_\theta]}{\longrightarrow}0$ при $n\to\infty$. Возьмем от обеих частей условное математическое ожидание при условии $\mathcal{A}_{j-1}$. Получаем:

$$\psi_j(\theta,\theta_n)-1=\frac{h^2}{2n}\mathbf{E}_\theta(\ddot\varphi_j(\theta)|\mathcal{A}_{j-1})+\frac{1}{n}\mathbf{E}_\theta(r_{nj}(\theta)|\mathcal{A}_{j-1}).$$

В последнем соотношении мы использовали условие /8.1.7/. Далее, аналогично доказательству соотношения (62) можно показать, что

$$\frac{1}{n}\sum_{j=1}^{n}(\ddot\varphi_j(\theta)-\mathbf{E}_\theta\ddot\varphi_j(\theta))\xrightarrow[n\to\infty]{\mathbf{P}_\theta}0.$$

Так как $\mathbf{E}_\theta\ddot\varphi_j(\theta)=-\,m(\theta)$ и выполнено условие /8.1.4/ (ii), то

$$\frac{1}{n}\sum_{j=1}^{n}(-\ddot\varphi_j(\theta)-m_0(\theta))\xrightarrow[n\to\infty]{\mathbf{P}_\theta}0.$$

Из неравенства

$$|\,\frac{1}{n}\sum_{j=1}^{n}-\mathbf{E}_\theta(\ddot\varphi_j(\theta)I(|\ddot\varphi_j(\theta)\,|>\varepsilon n))\,|\leqslant\frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta(|\ddot\varphi_j(\theta)\,|I(|\ddot\varphi_j(\theta)|>\varepsilon n))$$

и условия /8.1.5/ (ii) следует, что

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta(\ddot\varphi_j(\theta)I(|\ddot\varphi_j(\theta)|>\varepsilon n))\xrightarrow[n\to\infty]{\mathbf{P}_\theta}0,$$

поэтому

$$-\frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta(\ddot\varphi_j(\theta)I(|\ddot\varphi_j(\theta)\,|\leqslant\varepsilon n))\xrightarrow[n\to\infty]{\mathbf{P}_\theta}m(\theta).$$

Обозначим $V_j=-\mathbf{E}_\theta(\ddot\varphi_j(\theta)|\mathcal{A}_{j-1})$. Тогда $V_j=\tilde V_j+\tilde W_j$, где

$$\tilde V_j=-\mathbf{E}_\theta(\ddot\varphi_j(\theta)I(|\ddot\varphi_j(\theta)|\leqslant\varepsilon n)|\mathcal{A}_{j-1}),\qquad\tilde W_j=-\mathbf{E}_\theta(\ddot\varphi_j(\theta)I(|\ddot\varphi_j(\theta)|>\varepsilon n)|\mathcal{A}_{j-1}).$$

Так как

$$\mathbf{P}_\theta(\,|\,\frac{1}{n}\sum_{j=1}^{n}\tilde{W}_j\,|>\delta)\leqslant\frac{1}{n\delta}\sum_{j=1}^{n}\mathbf{E}_\theta(\,|\,\ddot{\varphi}_j(\theta)\,|\,I(|\ddot{\varphi}_j(\theta)|>\varepsilon n)|\mathcal{A}_{j-1})$$

и

$$\frac{1}{n\delta}\sum_{j=1}^{n}\mathbf{E}_\theta(\,|\,\ddot{\varphi}_j(\theta)\,|\,I(|\ddot{\varphi}_j(\theta)|>\varepsilon n)|\mathcal{A}_{j-1})\xrightarrow[n\to\infty]{\mathbf{P}_\theta}0$$

(условие /8.1.5/ (ii)), то

$$\frac{1}{n}\sum_{j=1}^{n}\tilde{W}_j\xrightarrow[n\to\infty]{\mathbf{P}_\theta}0.$$

Далее,

$$\mathbf{E}_\theta(\tilde{V}_j+\ddot{\varphi}_j(\theta)I(\,|\,\ddot{\varphi}_j(\theta)\,|\,\leqslant\varepsilon n))=0,$$

поэтому ($\ddot{\varphi}_j=\ddot{\varphi}_j(\theta)$)

$$\mathbf{D}_\theta(\tilde{V}_j+\ddot{\varphi}_j(\theta)I(\,|\,\ddot{\varphi}_j(\theta)\,|\,\leqslant\varepsilon n))=\mathbf{E}_\theta(\tilde{V}_j+\ddot{\varphi}_j(\theta)I(\,|\,\ddot{\varphi}_j(\theta)\,|\,\leqslant\varepsilon n))^2\leqslant$$

$$\leqslant 2\mathbf{E}_\theta(\mathbf{E}_\theta(\ddot{\varphi}_j^2I(\,|\,\ddot{\varphi}_j\,|\,\leqslant\varepsilon n)|\mathcal{A}_{j-1}))+2\mathbf{E}_\theta(\ddot{\varphi}_j^2I(\,|\,\ddot{\varphi}_j\,|\,\leqslant\varepsilon n)|\mathcal{A}_{j-1})\leqslant 4\varepsilon n\mathbf{E}_\theta(\,|\,\ddot{\varphi}_j\,|\,I(\,|\,\ddot{\varphi}_j\,|\,\leqslant\varepsilon n)|\mathcal{A}_{j-1}).$$

Кроме того, если $i<j$, то

$$\mathbf{COV}(\tilde{V}_i+\ddot{\varphi}_i(\theta)I(\,|\,\ddot{\varphi}_i(\theta)\,|\,\leqslant\varepsilon n),\tilde{V}_j+\ddot{\varphi}_j(\theta)I(\,|\,\ddot{\varphi}_j(\theta)\,|\,\leqslant\varepsilon n))=$$

$$=\mathbf{E}_\theta(\tilde{V}_i+\ddot{\varphi}_i(\theta)I(\,|\,\ddot{\varphi}_i(\theta)\,|\,\leqslant\varepsilon n))\mathbf{E}_\theta(\tilde{V}_i+\ddot{\varphi}_i(\theta)I(\,|\,\ddot{\varphi}_i(\theta)\,|\,\leqslant\varepsilon n))=0,$$

так как

$$\mathbf{E}_\theta(\tilde{V}_i+\ddot{\varphi}_i(\theta)I(\,|\,\ddot{\varphi}_i(\theta)\,|\,\leqslant\varepsilon n))=0,$$

и

$$\mathbf{P}_\theta\left(\left|\frac{1}{n}\sum_{j=1}^{n}(\tilde{V}_j+\ddot{\varphi}_j(\theta)I(\,|\,\ddot{\varphi}_j(\theta)\,|\,\leqslant\varepsilon n))\right|>\delta\right)\leqslant\tag{70}$$

$$\leqslant\frac{1}{n^2\delta^2}\frac{1}{n}\sum_{j=1}^{n}\mathbf{D}_\theta(\tilde{V}_j+\ddot{\varphi}_j(\theta)I(\,|\,\ddot{\varphi}_j(\theta)\,|\,\leqslant\varepsilon n))\leqslant\frac{4}{n^2\delta^2}\frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta(\,|\,\ddot{\varphi}_j(\theta)\,|\,I(\,|\,\ddot{\varphi}_j(\theta)\,|\,\leqslant\varepsilon n)).$$

Из неравенства Чебышева, свойств условных математических ожиданий и равенства (69) следуют соотношения:

$$\mathbf{P}_\theta\{\,|\,\sum_{j=1}^{n}(\psi_j(\theta,\theta_n)-1)-\frac{h^2}{2n}\sum_{j=1}^{n}\mathbf{E}_\theta(\ddot{\varphi}_j(\theta)|\mathcal{A}_{j-1})\,|\,>\varepsilon\}\leqslant$$

$$\leqslant\frac{1}{\varepsilon n}\sum_{j=1}^{n}\mathbf{E}_\theta\,|\,r_{nj}\,|\,\leqslant\frac{1}{\varepsilon n}\sum_{j=1}^{n}\mathbf{E}_\theta^{1/2}(r_{nj}^2)\xrightarrow[n\to\infty]{}0.$$

Из последнего и соотношения (70) получаем утверждение (i) леммы 7.

Перейдем к доказательству части (ii) леммы 7. Рассмотрим величины

$$Y_j=(\varphi_j(\theta,\theta_n)-1)-n^{-1/2}h\dot{\varphi}_j(\theta)-(\psi_j(\theta,\theta_n)-1),\qquad j=1,2,...,n.$$

В силу свойств условных математических ожиданий, определения величин $\psi_j(\theta,\theta_n)$ и условия /8.1.7/,

$$\mathbf{E}_\theta(Y_j|Y_1,...,Y_{j-1})=\mathbf{E}_\theta(\mathbf{E}_\theta(Y_j|\mathcal{A}_{j-1})|Y_1,...,Y_{j-1})=0.$$

Обозначим $S_N=\sum_{j=1}^{N}Y_j, 1\leqslant N\leqslant n$. Тогда

$$\mathbf{E}_\theta(S_N^2)=\mathbf{E}_\theta(Y_N^2+2Y_nS_{N-1}+S_{N-1}^2)=\mathbf{E}_\theta(Y_N^2)+2\mathbf{E}_\theta(S_{N-1}\mathbf{E}_\theta(Y_N|Y_1...Y_{N-1}))+\mathbf{E}_\theta(S_{N-1}^2)=$$

$$= \mathbf{E}_\theta(Y_N^2) + \mathbf{E}_\theta(S_{N-1}^2).$$

Значит,

$$\mathbf{E}_\theta(S_N^2) = \sum_{j=1}^{N} \mathbf{E}_\theta(Y_j^2) \leqslant \sum_{j=1}^{n} \mathbf{E}_\theta(Y_j^2).$$

Из элементарного неравенства $(a+b)^2 \leqslant 2a^2 + 2b^2$ получаем, что

$$\sum_{j=1}^{n} \mathbf{E}_\theta(Y_j^2) \leqslant \frac{2}{n}\sum_{j=1}^{n} \mathbf{E}_\theta[\, n^{1/2}(\varphi_j(\theta,\theta_n)-1) - h\dot\varphi_j(\theta)]^2 + \frac{2}{n}\sum_{j=1}^{n} \mathbf{E}_\theta[\, n^{1/2}(\psi_j(\theta,\theta_n)-1)]^2 \underset{n\to\infty}{\longrightarrow} 0.$$

Таким образом, $(S_N, \mathcal{F}_N), 1 \leqslant N \leqslant n$, где $\mathcal{F}_N = \sigma(Y_1, ..., Y_N)$, есть квадратично интегрируемый мартингал. По неравенству Колмогорова

$$\mathbf{P}_\theta(\max_{N\leqslant n} |\, S_N \,| \geqslant \varepsilon) \leqslant \varepsilon^{-2}\mathbf{E}_\theta(S_n^2) = \varepsilon^{-2}\sum_{j=1}^{n} \mathbf{E}_\theta(Y_j^2),$$

откуда

$$\max_{N\leqslant n} |\, S_N \,| \xrightarrow[n\to\infty]{\mathbf{P}_\theta} 0,$$

и, значит,

$$\sum_{j=1}^{n} Y_j^2 \xrightarrow[n\to\infty]{\mathbf{P}_\theta} 0.$$

Д о к а з а т е л ь с т в о теоремы 34. По лемме 6

$$\Lambda_n(\theta,\theta_n) - \sum_{j=1}^{n}(\varphi_{nj}-1) - \frac{1}{2}\sum_{j=1}^{n}(\varphi_{nj}-1)^2 - \ln\varphi_0(\theta,\theta_n) \xrightarrow[n\to\infty]{\mathbf{P}_\theta} 0,$$

а по лемме 5

$$\sum_{j=1}^{n}(\varphi_{nj}-1)^2 \xrightarrow[n\to\infty]{\mathbf{P}_\theta} h^2\Gamma_0^2(\theta).$$

По лемме 7

$$\sum_{j=1}^{n}(\varphi_{nj}-1) - n^{-1/2}h\sum_{j=1}^{n}\dot\varphi_j(\theta) \xrightarrow[n\to\infty]{\mathbf{P}_\theta} h^2 m_0(\theta)/2.$$

Результат (i) теоремы 34 следует из приведенных соотношений.

Перейдем к доказательству части (ii) теоремы 34. Для этого рассмотрим сумму $U_n = \sum_{j=1}^{n}\dot\varphi_j/\sqrt{n}$.

Поскольку $\mathbf{E}_\theta(\dot\varphi_j^2) < \infty$ и $\mathbf{E}_\theta(\dot\varphi_j|\mathcal{A}_{j-1}) = 0$, то $(\dot\varphi_j, \mathcal{A}_{j-1}, j \geqslant 1$ есть мартингал-разность, причем (здесь $\dot\varphi_j = \dot\varphi_j(\theta)$)

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta(\dot\varphi_j^2 I(|\,\dot\varphi_j\,| > \varepsilon)|\mathcal{A}_{j-1}) \xrightarrow[n\to\infty]{\mathbf{P}_\theta} 0 \tag{71}$$

(можно считать, что $\varepsilon \in (0,1]$) и

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta(\dot\varphi_j^2|\mathcal{A}_{j-1}) \xrightarrow[n\to\infty]{\mathbf{P}_\theta} \Gamma_0^2(\theta). \tag{72}$$

Соотношения (71) и (72) доказываются аналогично предыдущим соотношениям (см., например, соотношение (70)).

Тогда,

$$\mathbf{P}_\theta(U_n < x) \underset{n\to\infty}{\longrightarrow} \Phi(x/\Gamma_0(\theta)),$$

поскольку выполнены условия (l_2) и (c_2) этой теоремы. В нашем случае это есть соотношения (71) и (72). Последнее доказывает часть (ii) теоремы 34.

УСЛОВИЯ /8.1/.
/8.1.8/ Функции $\Gamma^2(\theta)$ и $m(\theta)$ растут по $|\,\theta\,|$ не быстрее некоторого полинома, а

$$\mathbf{E}_\theta(\exp((1/2)\Lambda_n(\theta,\theta_n)) \leqslant a_N|\,h\,|^{-N}$$

для $n \geqslant n_0, N > 0$ – любое число, a_N – константа, зависящая от N и θ.

Теорема 35. *Пусть выполнены условия /8.1.1/ – /8.1.8/. Тогда оценка $\hat{\theta}_n$ состоятельна, а $\sqrt{n}(\hat{\theta}_n - \theta)$ асимптотически нормальна с ожиданием нуль и дисперсией* $\Gamma^2(\theta)m^{-2}(\theta)$.

Для д о к а з а т е л ь с т в а теоремы 35 достаточно показать:

(i) конечномерные распределения функций $\Lambda_n(\theta,\theta_n)$ сходятся к конечномерным распределениям функции $\Lambda(h) = \xi\Gamma(\theta)h - h^2m(\theta)/2$,

(ii) $\mathbf{E}_\theta[\exp((1/2)\Lambda_n(\theta,\theta_{1n}) - \exp((1/2)\Lambda_n(\theta,\theta_{2n})]^2 \leqslant (H_2-h_1)^2M(1+|\,\theta\,|+(|h_1|+|h_2|)/\sqrt{n})^l$,

$\theta_i = \theta + h_in^{-1/2}$, $i = 1,2, n \geqslant n_0, M$ – некоторая константа, l – степень полинома, определенная условием /8.1.7/, и применить результаты предыдущие рассуждения.

Для д о к а з а т е л ь с т в а утверждения (i) теоремы 35 надо воспользоваться соотношением ($n \to \infty$)

$$\mathbf{COV}(\varphi_j(\theta,\theta_{1n}), \varphi_j(\theta,\theta_{1n})) = \Gamma^2(\theta)h_1h_2n^{-1}(1+o(1))$$

для фиксированных h_1, h_2 и повторить предыдущие рассуждения.

Утверждение (ii) теоремы 35 следует из условия /8.1.7/. Действительно, рассмотрим $\mathbf{E}_\theta(\mathcal{T}_n(h_2) - \mathcal{T}_n(h_1)$, где $\mathcal{T}_n(h) = \exp(\Lambda_n(\theta,\theta_n))$. В силу условия /8.1.7/ для $n \geqslant n_0$, используя также неравенство Коши-Буняковского, получим ($h_2 > h_1$)

$$\leqslant \mathbf{E}_\theta[(g_n(X;\theta + h_2n^{-1/2}) - g_n(X;\theta + h_1n^{-1/2})/g_n(X;\theta)]^2 \leqslant$$

$$C_1(\theta)\mathbf{E}_\theta\{g_n^{-2}(X;\theta)[\int_{h_1n^{-1/2}}^{h_1n^{-1/2}} \frac{\partial}{\partial t}g_n(X;\theta+t)\,dt]^2\} \leqslant$$

$$\leqslant C_1(\theta)(h_2-h_1)n^{-1/2}\mathbf{E}_\theta\{\} \int_{h_1n^{-1/2}}^{h_1n^{-1/2}} \frac{\partial}{\partial t}g_n(X;\theta+t)/g_n(X;\theta\,dt]^2\} \leqslant$$

$$\leqslant C_2(\theta)(h_2-h_1)^2 \sup_{h_1n^{-1/2}\leqslant u-\theta\leqslant h_2n^{-1/2}} \Gamma^2(u) \leqslant M(h_2-h_1)^2(1+|\theta|+(|h_1|+|h_2|)n^{-1/2})^l.$$

Последнее завершает доказательство теоремы 35.

Так как прямая проверка условий /8.1.5/ бывает затруднительной, то докажем следующий результат.

ТЕОРЕМА 36. (i) *Если для некоторого $\delta > 0$*

$$\limsup_n \frac{1}{n}\sum_{j=1}^n \mathbf{E}_\theta(|\,\dot{\varphi}_j\,|^{2+\delta}) < \infty,$$

то условие /8.1.5/ (i) *будет выполнено.*

(ii) *Если для некоторого $\delta > 0$*

$$\limsup_n \frac{1}{n}\sum_{j=1}^n \mathbf{E}_\theta(|\,\ddot{\varphi}_j\,|^{1+\delta}) < \infty,$$

то условия /8.1.5/ (ii) *и* /8.1.5/ (iii) *будут выполнены.*

Д о к а з а т е л ь с т в о. Начнем с части (i). Запишем следующую цепочку неравенств ($\varepsilon > 0$)

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta\{\dot\varphi_j^2 I(|\,\dot\varphi_j\,| > \varepsilon n^{1/2})\} \leqslant \varepsilon^{-\delta} n^{-(1+\delta/2)}\sum_{j=1}^{n}\mathbf{E}_\theta\{|\,\dot\varphi_j\,|^{2+\delta} I(|\,\dot\varphi_j\,| > \varepsilon n^{1/2})\} \leqslant$$

$$\leqslant \varepsilon^{-\delta} n^{-(1+\delta/2)}\sum_{j=1}^{n}\mathbf{E}_\theta\{|\,\dot\varphi_j\,|^{2+\delta}\},$$

поэтому, если

$$\limsup_n \frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta(|\,\dot\varphi_j\,|^{2+\delta}) < \infty, \qquad \text{то} \qquad \frac{1}{\varepsilon^\delta n^{1+\delta/2}}\sum_{j=1}^{n}\mathbf{E}_\theta\{|\,\dot\varphi_j\,|^{2+\delta}\} \underset{n\to\infty}{\longrightarrow} 0,$$

и далее – результат теоремы 34 (i).

Точно так же доказывается часть (ii) теоремы 34.

В следующем примере рассмотрим задачу оценивания неизвестного параметра распределения (байесовское оценивание) по цензурированным типа II выборкам. Мы покажем, что она есть частный случай рассмотренной оценки байесовского типа и отсюда выведем её асимптотическое поведение при $n \to \infty$.

Пример 58. Пусть $x_1, x_2, ..., x_n$ – повторная выборка из распределения с общей плотностью распределения $f(x;\theta)$. Плотность совместного распределения отрезка вариационного ряда $X_n^{(1,n)}$ равна

$$f_n^{(n)}(y_1, ..., y_n;\theta) = n!\prod_{j=1}^{n} f(y_j;\theta), \qquad -\infty < y_1 < ... < y_n < \infty.$$

В качестве модельной плотности $g_n(x;\theta)$ возьмем плотность распределения порядковых статистик $(X_n^{(1)}, ..., X_n^{(n)})$, равную

$$g_n(y_1, ..., y_k;\theta) = f_n^{(k)}(y_1, ..., y_k;\theta) = \frac{n!}{(n-k)!}\prod_{j=1}^{k} f(y_j;\theta) G^{n-k}(y_k;\theta), \quad -\infty < y_1 < ... < y_k < \infty,$$

где $G(x;\theta) = 1 - F(x;\theta)$.

Обозначим через $f_n(y_j|y_{j-1};\theta)$ условную плотность распределения j-й порядковой статистики, когда фиксировано значение $X_n^{(j-1)} = y_{j-1}$. Очевидно, что

$$f_n(y_j|y_{j-1};\theta) = (n-j+1) f(y_j;\theta) G^{n-j}(y_j;\theta)/G^{n-j+1}(y_{j-1};\theta), \quad 2 \leqslant j \leqslant n,$$

и

$$f_n(y_1|y_0;\theta) = f_n^{(1)}(y_1;\theta) = n f(y_1;\theta) G^{n-1}(y_1;\theta).$$

Тогда

$$f_n^{(k)}(y_1, ..., y_k;\theta) = \prod_{j=1}^{n} f(y_j|y_{j-1};\theta), \quad y_1 < ... < y_k.$$

Пусть $k/n \to \lambda$ при $n \to \infty$, где $\lambda \in (0,1)$. Предположим, что Θ – открытый интервал в $\mathbf{R}^1$, $f(x;\theta) > 0$, для каждых $x \in \mathbf{R}^1, \theta \in \Theta$, $f(x;\theta)$ – дважды непрерывно дифференцируемая функция θ и для каждого $\theta \in \Theta$

/8.1.9/ $\qquad |\,(\partial^i/\partial\theta^i) f(x;\theta)\,| \leqslant U_i(x), \qquad \int\limits_{-\infty}^{\infty} U_i(x)\,dx < \infty, \quad i = 1, 2\,;$

/8.1.10/ $\quad 0 < \int\limits_{-\infty}^{\infty} \dot f^2(x;\theta)\,dF(x;\theta) < \infty,$

где

$$\dot f(x;\theta) = (\partial/\partial\theta)\ln f(x;\theta), \qquad \ddot f(x;\theta) = (\partial/\partial\theta)\dot f(x;\theta).$$

Пусть

$$r(x;\theta) = f(x;\theta)/G(x;\theta), \quad x \in \mathbf{R}^1,$$

и предположим, что для каждого $\theta \in \Theta$

/8.1.11/ $\quad |\,(\partial^i/\partial\theta^i) r(x;\theta)\,| \leqslant W_i(x), \quad \int\limits_{-\infty}^{\infty} W_i(x)\,dx < \infty, \quad i = 1, 2\,; \quad \dot r(x;\theta) = \dfrac{\partial}{\partial\theta}\ln r(x;\theta).$

Пусть существуют $\delta, \varepsilon > 0$, что для любого $\theta \in \Theta$

/8.1.12/ $\quad \int\limits_{-\infty}^{\infty} |\,\dot r(x;\theta)\,|^{2+\delta}\,dF(x;\theta) < \infty, \quad \int\limits_{-\infty}^{\infty} |\,\ddot r(x;\theta)\,|^{1+\delta}\,dF(x;\theta) < \infty$

Заметим, что

$$\dot G(x;\theta) = \dot f(x;\theta) - \dot r(x;\theta) \quad \text{и} \quad \ddot G(x;\theta) = \ddot f(x;\theta) - \ddot r(x;\theta).$$

Из условий /8.1.9/ – /8.1.12/ следует, что

$$\mathbf{E}_\theta(|\,\dot G(x;\theta)\,| < \infty \quad \text{для любого} \quad \theta \in \Theta. \tag{73}$$

Пусть

$$\dot\varphi_j(\theta) = (\partial/\partial\theta)\ln f_n(y_j\,|\,y_{j-1};\theta), \quad 1 \leqslant j \leqslant n; \quad \ddot\varphi_j(\theta) = \partial/\partial\theta)\dot\varphi_j(\theta).$$

Далее, требования /8.1.9/ – /8.1.12/ гарантируют интегрируемость $\dot\varphi_j(\theta)$ и дифференцируемость её под знаком интеграла, т.е. для каждого $1 \leqslant j \leqslant n$ $(y_0 = -\infty)$,

$$\int\limits_{y_{j-1}}^{\infty} [(\partial/\partial\theta)\ln f_n(x\,|\,y_{j-1};\theta)] f_n(x\,|\,y_{j-1};\theta)\,dx = \frac{\partial}{\partial\theta}\int\limits_{y_{j-1}}^{\infty} f_n(x\,|\,y_{j-1};\theta)\,dx = 0,$$

что означает выполнение условия /8.1.7/.

Покажем, что для $\delta, \varepsilon > 0$, определенных требованиями /8.1.12/, будут выполнены неравенства

$$\limsup_n \frac{1}{n}\sum_{j=1}^{n} \mathbf{E}_\theta|\,\dot\varphi_j(\theta)\,|^{2+\delta} < \infty \tag{74}$$

и

$$\limsup_n \frac{1}{n}\sum_{j=1}^{n} \mathbf{E}_\theta|\,\ddot\varphi_j(\theta)\,|^{1+\delta} < \infty, \tag{75}$$

откуда, согласно теореме 36, будет следовать условие /8.1.5/.

Докажем соотношение (74), соотношение (75) доказывается аналогичными рассуждениями.

Из определения $\dot\varphi_j(\theta)$ и неравенства Гёльдера

$$|\,\dot\varphi_j(\theta)\,|^{2+\delta} \leqslant 2^{2+\delta}(|\,\dot r(X_n^{(j)};\theta)\,|^{2+\delta} + (n-j+1)|\,\dot G(X_n^{(j)};\theta) - \dot G(X_n^{(j-1)};\theta)\,|^{2+\delta}),$$

и для $1 \leqslant k \leqslant n$,

$$\frac{1}{n}\sum_{j=1}^{k} \mathbf{E}_\theta|\,\dot r(X_n^{(j)};\theta)\,|^{2+\delta} \leqslant \frac{1}{n}\sum_{j=1}^{n} \mathbf{E}_\theta|\,\dot r(X_n^{(j)};\theta)\,|^{2+\delta} \xrightarrow[n\to\infty]{} \int\limits_{-\infty}^{\infty} |\,\dot r(x;\theta)\,|^{2+\delta}\,dF(x;\theta) < \infty.$$

Аналогично,

$$\frac{1}{n}\sum_{j=1}^{k} \mathbf{E}_\theta(n-j+1)|\,\dot G(X_n^{(j)};\theta) - \dot G(X_n^{(j-1)};\theta)\,|^{2+\delta} \leqslant$$

$$\leqslant \frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta(n-j+1)|\,\dot G(X_n^{(j)};\theta)-\dot G(X_n^{(j-1)};\theta)\,|^{2+\delta}\xrightarrow[n\to\infty]{}\Gamma(3+\delta)\int\limits_{-\infty}^{\infty}|\,\dot r(x;\theta)\,|^{2+\delta}\,dF(x;\theta)<\infty.$$

Далее получаем

$$\frac{1}{n}\sum_{j=1}^{k}\mathbf{E}_\theta(\dot\varphi_j^2(\theta))\xrightarrow[n\to\infty]{}\int\limits_{-\infty}^{x_\lambda(\theta)}(\dot f(x;\theta))^2\,dF(x;\theta)+(1-\lambda)(\dot F(x;\theta))^2\Big|_{x=x_\lambda(\theta)}\equiv J_\lambda(\theta)<\infty,$$

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{E}_\theta[\,-\ddot\varphi_j(\theta)]\xrightarrow[n\to\infty]{}J_\lambda(\theta)<\infty.$$

Таким образом, выполнено и условие /8.1.4/. Пусть выполнено условие /8.1.8/. Тогда из теоремы 35 получаем результат: если $t_n=t_n(X_n^{(1)},...,X_n^{(k)})$ – байесовская оценка, построенная по цензурированной выборке $X_n^{(1,k)}$ из потенциальной повторной выборки $X_1,...,X_n$, то

$$\mathbf{P}_\theta(\sqrt{n}(t_n-\theta)<x)\xrightarrow[n\to\infty]{}\Phi(x/\sqrt{J_\lambda(\theta)}).$$

Что же касается условия /8.1.7/, то как нетрудно заметить, его можно ослабить, предполагая, что

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\mathbf{E}_\theta(\dot\varphi_j(\theta)\,|\,\mathcal{A}_{j-1})\xrightarrow[n\to\infty]{\mathbf{P}_\theta}0.$$

Далее, если условие /1.1.7/ будет выполнено для некоторых $g_{1n}(x;\theta)$ и $g_{2n}(x;\theta)$, то $g_{1n}^{\alpha}(x;\theta)g_{2n}^{\beta}(x;\theta)$ также ему будет удовлетворять. Так, если θ – параметр сдвига и

$$\int_{-\infty}^{\infty}xf(x)\,dx=0,\qquad\int_{-\infty}^{\infty}x^2f(x)\,dx<\infty,$$

то класс функций $g(x)$(, для которых выполнено условие /8.1.7/ имеет вид: $g(x)=f^{\alpha}(x)\exp(\beta x^2)$.

Если же

$$\int_{0}^{\infty}xf(x)\,dx=\int_{-\infty}^{0}xf(x)\,dx,$$

то $g(x)=f^{\alpha}(x)\exp(\beta|\,x\,|)$.

Для параметра масштаба с условием

$$\mathbf{P}_\theta(X>0)=1,\quad\gamma=\int_{0}^{\infty}xf(x)\,dx<\infty,$$

будем иметь $g(x)=f^{\alpha}(x)e^{-\mu\beta x}x^{\beta(\nu-1)}$, где $\nu=\gamma\mu$.

Пример 59. Пусть наблюдается пара (Z,W), где $Z=\min(X,Y)$, а W – индикатор события $(X<Y)$, X и Y независимы, выборка $(z,w)=((z_1,w_1),(z_2,w_2),...,(z_n,w_n))$ – повторная, с.в. X имеет неизвестную плотность распределения $f(x;\theta)$ и функцию распределения $F(x;\theta)$, с.в. Y имеет плотность распределения $g(x)$ и функцию распределения $G(x)$, которые неизвестны. Возможны два случая: 1) $g(x)$ не зависит от θ; 2) $g(x)$ зависит от параметра θ. В обоих случаях возьмем $g_n(z,w;\theta)$, равную

$$g_n(z,w;\theta)=\prod_{j=1}^{n}f^{w_j}(z_j;\theta)(1-F(x;\theta))^{1-w_j}.$$

Рассмотрим первый случай. Здесь

$$\dot\varphi(\theta) = w\,l_1(z;\theta) + (1-w)l_2(z;\theta) \quad \text{где} \quad l_1(z;\theta) = \dot f(z;\theta), \quad l_2(z;\theta) = \frac{\partial}{\partial\theta}\ln(1-F(x;\theta)).$$

В таком случае

$$\mathbf{E}_\theta(\dot\varphi(\theta)) = \mathbf{E}_\theta[\mathbf{E}_\theta(W\,|\,Z)\,l_1(W;\theta) + \mathbf{E}_\theta((1-W)\,|\,Z)l_2(Z;\theta)].$$

Найдем $\mathbf{E}_\theta(W\,|\,Z)$. Имеем

$$\mathbf{E}_\theta(W\,|\,Z) = \mathbf{P}_\theta(W=1|Z) = \frac{(1-G(Z))f(Z;\theta)}{h(Z;\theta)},$$

где

$$h(Z;\theta) = f(Z;\theta)(1-G(Z)) + g(Z)1-F(Z;\theta))$$

– плотность распределения Z. Поэтому

$$\mathbf{E}_\theta(\dot\varphi(\theta)) = \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(z;\theta)(1-G(z))\,dz - \int_{-\infty}^{\infty}g(z)\frac{\partial}{\partial\theta}F(z;\theta)\,dz.$$

Дифференцируя по θ обе части равенства

$$\int_{-\infty}^{\infty}h(z;\theta)\,dz = 1$$

и меняя знаки интегрирования и дифференцирования (предполагая, что это можно сделать) получим, что

$$\mathbf{E}_\theta(\dot\varphi(\theta)) = \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(z;\theta)(1-G(z))\,dz - \int_{-\infty}^{\infty}g(z)\frac{\partial}{\partial\theta}F(z;\theta)\,dz = 0.$$

Обратимся ко второму случаю и рассмотрим модель пропорциональных рисков, т.е. случай, когда

$$(1-G(x;\theta)) = (1-F(x;\theta))^\beta, \quad \beta > 0,$$

β неизвестно. Здесь

$$h(z;\theta) = (1+\beta)f(z;\theta)(1-F(z;\theta))^\beta, \quad \mathbf{E}_\theta(W|Z) = \mathbf{E}_\theta(W) = 1/(1+\beta)$$

и

$$\mathbf{E}_\theta(\dot\varphi(\theta)) = (1+\beta)^{-1}\int\limits_{-\infty}^{\infty}(\dot f(z;\theta) - \beta\frac{\frac{\partial}{\partial\theta}F(z;\theta)}{1-F(z;\theta)})\,h(z;\theta)\,dz = 0,$$

что следует из возможности дифференцирования под знаком интеграла равенства

$$\int\limits_{-\infty}^{\infty}h(z;\theta)\,dz = (1+\beta)\int\limits_{-\infty}^{\infty}f(z;\theta)(1-F(z;\theta))^\beta\,dz = 1.$$

Пример 60. Пусть исходное семейство есть $\mathcal{P} = \{f_\varepsilon(x;\theta), \theta\in\Theta\subset\mathbf{R}^1\}$, где

$$f_\varepsilon(x;\theta) = (1-\varepsilon)f(x;\theta) + \varepsilon h(x;\theta), \quad 0\le\varepsilon\le 1,$$

$f(x;\theta)$ известна, а $h(x;\theta)$ может быть как известной, так и нет. Предположим, что θ – параметр сдвига. Тогда условие /8.1.7/ состоит здесь в том, чтобы

Пример 61. Пусть η_1 и η_2 – случайные величины такие, что $\mathbf{P}_\theta(\eta_1\le\eta_2) = 1$. Для случайной величины η_0, не зависящей от η_1 и η_2, назовем наблюдением следующий набор случайных величин (Благовещенский Ю.Н., Агеев В.В.)

$$\bar w = (w_0, w_1, w_2), \quad \zeta = w_0\eta_0 + w_1\eta_1 + w_2\eta_2,$$

где $w_1 = I_{(-\infty,\eta_1]}(\eta_0)$, $w_0 = I_{(\eta_1,\eta_2]}(\eta_0)$, $w_1 = I_{(\eta_2,\infty]}(\eta_0)$, $I_A(x)$ – индикатор множества $A \subset R^1$. Пусть далее функция распределения с.в. η_0 равна

$$F(x;\theta) = \mathbf{P}_\theta(\eta_0 < x),$$

$f(x;\theta)$ – плотность распределения. Введем

$$\dot{\varphi}(\theta) = w_0 l_0(x;\theta) + w_1 l_1(x;\theta) + w_2 l_2(x;\theta),$$

где $l_0(x;\theta) = \dot{f}(x;\theta)$, $l_1(x;\theta) = \dot{F}(x;\theta)$,$l_2(x;\theta) = (\partial/\partial\theta)\ln(1 - F(x;\theta))$.

Тогда

$$\mathbf{E}_\theta(\dot{\varphi}(\theta)) = 0,$$

что говорит о выполнении условия /8.1.7/.

Глава 9.

Последовательное оценивание

9.1. Последовательные планы оценивания параметра сдвига

Пусть на вероятностном пространстве ($\Omega,\ \mathcal{F},\ \mathbf{P}$) задана последовательность

$$x_i = \theta + \varepsilon_i, \quad i = 1, 2, \ldots, k, \quad \theta \in R^1,$$

где $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ – независимые случайные величины, $x_k^{(1)} < \ldots < x_k^{(r+1)} < \ldots < x_k^{(k-s)} < \ldots < x_k^{(k)}$ – вариационный ряд, построенный по повторной выборке $x_1,\ x_2, ...,\ x_k$. Введем семейство σ-алгебр $\mathcal{F}_k = \sigma$ ($x_k^{(r+2)} - x_k^{(r+1)}, \ldots, x_k^{(k-s)} - x_k^{(r+1)}$). Пусть $\tau(\omega) \geqslant r + s + 2$ – целочисленная случайная величина, $\mathfrak{M} = \{\ \tau(\omega)\ \}$ – класс марковских моментов относительно неубывающего семейства σ-алгебр $\mathcal{F}_k,\ k \geqslant r{+}s{+}2$ ($\tau = \tau(\omega)$ определяет момент прекращения наблюдений), $\{\delta_{\tau+m}(\omega)\}$ – класс $\mathcal{F}_{\tau+m}$-измеримых функций, функция $\delta_{\tau+m} = \delta_{\tau+m}(\omega)$ есть оценка параметра θ по наблюдениям $x_1, x_2, \ldots, x_{\tau+m}$.

Определение 15. *Пару функций $\Delta = (\tau, \delta_{\tau+m})$ назовем последовательным планом оценивания.*

Отметим, что в нашем случае решение об остановке принимается по наблюдениям $x_1, x_2, \ldots, x_\tau$ до момента τ, а оценка параметра θ производится после того, как мы получим наблюдения $x_1, x_2, \ldots, x_{\tau+m}$ до момента $\tau + m$.

Средние потери от применения последовательного плана оценивания $\Delta = (\tau, \delta_{\tau+m})$ будем измерять величиной

$$R(\theta, \Delta) = \mathbf{E}_\theta(\,|\,\delta_{\tau+m} - \theta\,|^\alpha\,), \quad \alpha > 1.$$

Определим класс $\mathcal{D}_n$ последовательных планов $\Delta = (\tau, \delta_{\tau+m})$ следующим образом:

$$\mathcal{D}_n = \{(\tau, \delta_{\tau+m}) : \mathbf{E}_\theta(\,\tau\,) \leqslant n,\ R(\theta, \Delta) < \infty\}.$$

Мы хотим найти такой план Δ_n^0, чтобы

$$R(\theta, \Delta_n^0) = \min_{\Delta \in \mathcal{D}_n} R(\theta, \Delta).$$

Перейдем к решению поставленной задачи.

Пусть $f_k(x_{r+1} - \theta, \ldots, x_{k-s} - \theta)$ – плотность распределения вектора $x_k^{(r+1,k-s)}$. Положим

$$p_k(\theta) = \frac{f_k(x_{r+1} - \theta, \ldots, x_{k-s} - \theta)}{\int\limits_{-\infty}^{\infty} f_k(x_{r+1} - \theta, \ldots, x_{k-s} - \theta)\, d\theta}.$$

Лемма 8. *Если $\delta_k = \delta_k(x_1, x_2, \ldots, x_k)$ какая-либо правильная оценка, т.е. оценка со свойством*

$$\delta_k(x_1 + c, x_2 + c, \ldots, x_k + c) = \delta_k(x_1, x_2, \ldots, x_k) + c, \quad \forall c \in R^1,$$

и $\mathbf{E}_\theta[\,W(\delta_k - \theta)\,] < \infty$, где $W(x)$ – выпуклая функция, то

$$\mathbf{E}_\theta\{\,W(\delta_k - \theta)\,|\,\mathcal{F}_k\,\} = \int\limits_{-\infty}^{\infty} W(\delta_k - u)\, p_k(u)\, du.$$

Правильная оценка $t_k^{(W)}(x_k^{(r+1,k-s)}) = t_k$ называется *оценкой Питмена* для параметра θ по отношению к функции потерь $W(\delta_k - \theta)$, если

$$\mathbf{E}_\theta(\,W(t_k - \theta)\,) = \inf_{\delta_k} \mathbf{E}_\theta(\,W(\delta_k - \theta)\,),$$

где inf берется по всем правильным оценкам δ_k.

Таким образом, оценка Питмена t_k есть наилучшая правильная оценка, построенная по выборке объема k. Из леммы 8 следует, что оценка Питмена существует и однозначно определяется равенством

$$\int\limits_{-\infty}^{\infty} W(t_k-\theta)p_k(\theta)\,d\theta = \inf_t \int\limits_{-\infty}^{\infty} W(t-\theta)p_k(\theta)\,d\theta,$$

так что оценку Питмена можно трактовать как байесовскую относительно равномерного априорного распределения на R^1 и функции потерь $W(x)$. В частности, если $W(x)=x^2$, то

$$t_k = \int\limits_{-\infty}^{\infty} up_k(u)\,du.$$

Определим теперь *последовательную* оценку Питмена, полагая $t_{\tau+m}=t_{k+m}$ на множестве $\{\tau=k\}$.

Лемма 9 [Факеев А.Г.]. *Пусть $\delta_{\tau+m}$ – произвольная правильная последовательная оценка и $\mathbf{E}_\theta(W(\delta_{\tau+m}-\theta))<\infty$, а $t_{\tau+m}$ – последовательная оценка Питмена по отношению к функции потерь $W(x)$. Тогда*

$$\mathbf{E}_\theta(W(\delta_{\tau+m}-\theta)) \geqslant \mathbf{E}_\theta(W(t_{\tau+m}-\theta)).$$

Из этой леммы следует, что для нахождения оптимального плана Δ_n^0 достаточно рассматривать планы вида $(\tau, t_{\tau+m})$.

Таким образом, если через $\mathfrak{M}_n$ обозначим множество моментов остановки таких, что

$$\mathbf{E}_\theta\tau \leqslant n, \quad \mathbf{E}_\theta(W(t_{\tau+m}-\theta))<\infty,$$

то

$$\inf_{\Delta\in\mathcal{D}_n} R(\theta,\Delta) = \inf_{\tau\in\mathfrak{M}_n} R(\theta,\tau,t_{\tau+m}) = R(\theta,\Delta_n^0).$$

Положим

$$d_k = \int\limits_{-\infty}^{\infty} W(t_k-u)p_k(u)\,du.$$

Лемма 10. *Если τ – марковский момент остановки относительно $\{\mathcal{F}_k\}$, $k\geqslant r+s+2$, то*

$$\mathbf{E}_\theta(W(t_{\tau+m}-\theta)) = \mathbf{E}_\theta(d_{\tau+m}) = \mathbf{E}_0(d_{\tau+m}) = \mathbf{E}(d_{\tau+m}).$$

Обозначим

$$\mathcal{B}_k = \{\mathbf{E}\{d_{k+m+1}|\,\mathcal{F}_k\} \geqslant d_{k+m}\}, \quad k\geqslant r+s+2.$$

Определение 16. [Чжоу, Роббинс]. *Будем говорить, что имеет место монотонный случай, если*

$$\mathcal{B}_{r+s+2}\subset\mathcal{B}_{r+s+3}\subset\ldots, \quad \bigcup_k \mathcal{B}_k=\Omega.$$

Лемма 11. *В монотонном случае оптимальное правило остановки имеет вид*

$$\tau_0=\min\{k\geqslant r+s+2 : \mathbf{E}\{d_{k+m+1}|\,\mathcal{F}_k\}\geqslant d_{k+m},$$

для него

$$\mathbf{E}(d_{\tau_0+m}) = \inf_\tau \mathbf{E}(d_{\tau+m}),$$

если

$$\liminf_k \int\limits_{(\tau>k)} d^+_{k+m}=0; \quad a^+=\max(0,a).$$

Лемма 12. *Если τ_0 – марковский момент остановки такой, что*

$$\mathbf{E}\,(\lambda_0(\tau_0)) = \min_\tau \mathbf{E}\,(\lambda_0(\tau)),$$

при условии

$$\mathbf{E}\,(\lambda_i(\tau)) \leqslant 0, \quad i = 1, 2, \ldots, M, \tag{76}$$

то существуют такие постоянные $\nu_0 \geqslant 0$, $\nu_1 \geqslant 0$, $\ldots$, $\nu_M \geqslant 0$, *не все равные нулю одновременно, что* τ_0 *минимизирует величину*

$$H(\tau) = \mathbf{E}\,(\nu_0\lambda_0(\tau)) + \sum_{i=1}^{M} \nu_i\lambda_i(\tau)) \tag{77}$$

среди всех марковских моментов τ, *причем*

$$\mathbf{E}\,(\sum_{i=1}^{M} \nu_i\lambda_i(\tau)) = 0. \tag{78}$$

Если $\nu_0 > 0$, *то условия* (77) *и* (78) *являются достаточными для оптимальности* τ_0.

Лемма 13. *Оценка* $\hat\theta(x)$ *является байесовской тогда и только тогда, когда она удовлетворяет почти всюду следующим неравенствам*

$$\int\limits_{\theta \geqslant \hat\theta(x)} W'(\theta - \hat\theta(x))\,\xi(d\theta|\,x) \geqslant \int\limits_{\theta < \hat\theta(x)} W'(\hat\theta(x) - \theta)\,\xi(d\theta|\,x) \tag{79}$$

и

$$\int\limits_{\theta > \hat\theta(x)} W'(\theta - \hat\theta(x))\,\xi(d\theta|\,x) < \int\limits_{\theta \leqslant \hat\theta(x)} W'(\hat\theta(x) - \theta)\,\xi(d\theta|\,x), \tag{80}$$

где $\xi(d\theta|\,x)$ *– апостериорное распределение* θ *при заданном* x. *Кроме того, если* $W(t)$ *строго выпукла, то для каждого* x *существует единственная байесовская оценка* $\hat\theta(x)$, *удовлетворяющая* (79) – (80). *Если* $W'(0) = 0$, *то неравенства сводятся к уравнению*

$$\int\limits_{\theta > \hat\theta(x)} W'(\theta - \hat\theta(x))\,\xi(d\theta|\,x) = \int\limits_{\theta < \hat\theta(x)} W'(\hat\theta(x) - \theta)\,\xi(d\theta|\,x).$$

Пусть $D_n = (\,[\,n\,], t_{[\,n\,]+m}\,)$ – план фиксированного объема.

Определение 16. *Оптимальный последовательный план оценивания* Δ_n^0 *назовем* эффективным, *если*

$$\gamma_n(r, s) = \mathbf{E}_\theta W(t_{\tau_0+m} - \theta)/\mathbf{E}_\theta W(t_{[\,n\,]+m} - \theta) < 1,$$

и асимптотически эффективным, *если*

$$\varlimsup_{n\to\infty} \gamma_n(r, s) = \gamma(r, s) < 1.$$

Рассмотрим задачу построения оптимального плана последовательного оценивания параметра сдвига равномерного на $(\theta - 1/2, \theta + 1/2)$ распределения по отрезку вариационного ряда $x_k^{(r+1)} < \ldots < x_k^{(k-s)}$. Пусть сначала $m = 0,\ r = s$. Докажем следующее утверждение.

ТЕОРЕМА 37. *Оптимальный последовательный план оценивания* Δ_n^0 *при симметричной и выпуклой функции потерь* $W(x)$, *такой, что* $W(0) = 0$, *в классе* $\mathcal{D}_n$ *имеет вид*

$$\tau_0 = \min\{\,k \geqslant 2r + 2 : x_k^{(k-r)} - x_k^{(r+1)} \geqslant \frac{2r+2}{n}\,\},$$

$$t_{\tau_0} = (1/2)(x_{\tau_0}^{(r+1)} + x_{\tau_0}^{(\tau_0 - r)}).$$

Д о к а з а т е л ь с т в о. Положим $b_k = \mathbf{E}_\theta(W(t_k-\theta)\,|\,\mathcal{F}_k)+ck$. В случае $r=s$ оценка $t_k=(1/2)(x_k^{(r+1)}+x_k^{(k-r)})$ не зависят от вида симметричной и выпуклой функции потерь. Имеем

$$b_k=\frac{2}{\beta\,\ell_k^{2r+1}}\int\limits_0^{\ell_k/2}W(x)(\,\ell_k^2-x^2)^r\,dx+ck$$

где

$$\ell_k=1-x_k^{(k-r)}+x_k^{(r+1)},\quad \beta=\int\limits_0^1 t^r(1-t)^r\,dt=\frac{(r!)^2}{(r+1)!}.$$

Далее,

$$\mathbf{E}_\theta(b_{k+1}\,|\,\mathcal{F}_k)=\frac{4}{\beta^2\,\ell_k^{2r+1}}\int\limits_0^{\ell_k}v^r\,dv\int\limits_v^{\ell_k}(r+1)z^{-2r-1}(z-v)^r\,dz\int\limits_0^{z/2}W(x)(z^2/4-x^2)^r\,dx+$$

$$+\frac{2}{\beta\,\ell_k^{2r+1}}\int\limits_0^{\ell_k/2}W(x)(\ell_k^2/2-x^2)^r\,dx-\frac{2}{\beta\,\ell_k^{2r}}\int\limits_0^{\ell_k/2}W(x)(\ell_k^2/2-x^2)^r\,dx+c(k+1).$$

В случае $r=0$,

$$\mathbf{E}_\theta(b_{k+1}|\,\mathcal{F}_k)=\frac{4}{\ell_k}\int\limits_0^{\ell_k}dv\int\limits_v^{\ell_k}z^{-1}\,dz\int\limits_0^{z/2}W(x)\,dx+\frac{2}{\ell_k}\int\limits_0^{\ell_k/2}W(x)\,dx-2\int\limits_0^{\ell_k/2}W(x)\,dx+c(k+1).$$

Обозначим $\psi(\ell_k)=\mathbf{E}_\theta(b_{k+1}|\,\mathcal{F}_k)-b_k$. Покажем, что $\psi(x)$ монотонно не возрастает по $x<1$. Для этого рассмотрим производную функции $\psi(x)$. Мы изучим отдельно случаи $r=0$ и $r\neq 0$.

В случае $r=0$

$$\psi'(\ell)=-4\ell^2\int\limits_0^{\ell}du\int\limits_u^{\ell}z^{-1}\,dz\int\limits_0^{z/2}W(x)\,dx+\frac{4}{\ell}\int\limits_0^{\ell/2}W(x)\,dx-W(\ell/2).$$

Из выпуклости функции $W(x)$ имеем $W(x)\leqslant 2x\,\ell^{-1}W(\ell/2)$ для $0<x<\ell/2$. Следовательно,

$$\frac{4}{\ell}\int\limits_0^{\ell/2}W(x)\,dx\leqslant\frac{8W(\ell/2)}{\ell^2}\int\limits_0^{\ell/2}x\,dx=W(\ell/2),$$

откуда следует, что $\psi'(\ell)<0$ (здесь $0<\ell<1$).

В случае $r\neq 0$

$$\psi'(\ell)=-\frac{4(\,2r+1)}{\beta\,\ell^{2r+2}}\int\limits_0^{\ell}v^r\,dv\int\limits_v^{\ell}(r+1)z^{-2r-1}(z-v)^r\,dz\int\limits_0^{z/2}W(x)(z^2/4-x^2)^r\,dx+$$

$$+\frac{4}{\beta\,\ell^{2r+1}}\int\limits_0^{\ell/2}W(x)(\ell^2/4-x^2)^r\,dx-\frac{r}{\beta\,\ell^{2r-1}}\int\limits_0^{\ell/2}W(x)(\ell^2/4-x^2)^{r-1}\,dx.$$

Из выпуклости функции $W(x)$ следует также, что

$$\int\limits_0^{z}W(x)\,dx\leqslant zW(z/2),$$

откуда

$$\int\limits_0^{\ell/2}W(x)(\ell^2/4-x^2)^r\,dx=r\int\limits_0^{\ell/2}2z(\ell^2/4-z^2)^{r-1}\,dz\int\limits_0^{z}W(x)\;\leqslant r\,\ell^2/4\int\limits_0^{\ell/2}W(x)(\ell^2/4-x^2)^{r-1}\,dx,$$

поэтому и здесь $\psi'(\ell) \leqslant 0$.

Итак, $\psi(\ell)$ монотонно не возрастает, причем $\psi(0) > 0$, $\psi(1) < 0$. Используя леммы 12 и 13 и выбирая c так, чтобы $\psi((2r+2)/n) = 0$, получаем результат теоремы 37.

Пусть $w_1(\tau, \delta_\tau, \theta)$ и $w_2(\tau, \delta_\tau, \theta)$ – неотрицательные измеримые функции потерь (необязательно выпуклые) от принятия последовательного плана Δ, когда истинное значение параметра равно θ. Будем рассматривать класс последовательных планов, удовлетворяющих условиям:

$$0 < \mathbf{E}_\theta(\, w_1(\Delta, \theta)\,) < \infty, \quad 0 < \mathbf{E}_\theta(\, w_2(\Delta, \theta)\,) < \infty.$$

Лемма 14. *Пусть существует при каждом θ такой план Δ_c, $c > 0$, что*

$$\mathbf{E}_\theta(\, w_1(\Delta_c, \theta) + c w_2(\Delta_c, \theta)) = \inf_{\Delta} \mathbf{E}_\theta(\, w_1(\Delta, \theta) + c w_2(\Delta, \theta)), \tag{81}$$

и пусть функции $q_\theta(c) = \mathbf{E}_\theta(\, w_2(\Delta_c, \theta))$, $Q_\theta(c) = \mathbf{E}_\theta(\, w_1(\Delta_c, \theta))$ – функции ограниченной вариации, причем существуют обратные функции $q_\theta^{-1}(c)$ и $Q_\theta^{-1}(c)$. Тогда

$$Q_\theta(c) = -\int\limits_0^c x\, dq_\theta(x) = \int\limits_{\mathbf{E}_\theta(\, w_2(\Delta_c, \theta))}^{\infty} q_\theta^{-1}(y)\, dy, \tag{82}$$

$$q_\theta(c) = -\int\limits_0^c x^{-1}\, dQ_\theta(x) = \int\limits_{\mathbf{E}_\theta(\, w_1(\Delta_c, \theta))}^{\infty} (Q_\theta^{-1}(y))^{-1}\, dy, \tag{83}$$

Д о к а з а т е л ь с т в о. Докажем соотношение (82), соотношение (83) доказывается аналогично.

Если $c_1 \neq c$, то из (81) следует, что

$$\begin{cases} Q_\theta(c) + c\, q_\theta(c) \leqslant Q_\theta(c_1) + c\, q_\theta(c_1), \\ Q_\theta(c_1) + c_1 q_\theta(c_1) \leqslant Q_\theta(c) + c_1\, q_\theta(c), \end{cases}$$

откуда

$$\begin{cases} Q_\theta(c) - Q_\theta(c_1) + c\,(\, q_\theta(c) - q_\theta(c_1)\,) \leqslant 0, \\ Q_\theta(c) - Q_\theta(c_1) + c_1\,(\, q_\theta(c) - q_\theta(c_1)\,) \geqslant 0. \end{cases}$$

Разобьем отрезок $[0, c]$ точками $0 \equiv c_0 < c_1 \ldots < c_k < c \equiv c_{k+1}$. Учитывая, что $Q_\theta(0) = 0$, складывая последовательно предыдущие неравенства, мы получаем

$$\begin{cases} Q_\theta(c) + \sum\limits_{i:\, 0 \leqslant c_i \leqslant c} c_i\,(\, q_\theta(c_i) - q_\theta(c_{i+1})\,) \leqslant 0, \\ Q_\theta(c) + \sum\limits_{i:\, 0 \leqslant c_i \leqslant c} c_i\,(\, q_\theta(c_i) - q_\theta(c_{i+1})\,) \geqslant 0. \end{cases}$$

Так как функция $f(x) = x$ – непрерывна, а $q_\theta(c)$ есть функция ограниченной вариации, то

$$\begin{cases} Q_\theta(c) + \int\limits_0^c x\, dq_\theta(x) \leqslant 0, \\ Q_\theta(c) + \int\limits_0^c x\, dq_\theta(x) \geqslant 0. \end{cases}$$

Последние неравенства совместимы, только если

$$Q_\theta(c) = -\int\limits_0^c x\, dq_\theta(x),$$

что составляет утверждение леммы 14.

ТЕОРЕМА 38. (i) *Оптимальный последовательный план оценивания Δ_n^0 параметра сдвига равномерного на $(\theta - 1/2, \theta + 1/2)$ распределения в классе $\mathcal{D}_n$ имеет вид*

$$\tau_0 = \min\{\, k \geqslant r+s+2 : x_k^{(k-r)} - x_k^{(r+1)} \geqslant \frac{r+s+2}{n}\,\},$$

$$t_{\tau_o} = x_{\tau_o+m}^{(\tau_o+m-s)} - 1/2 + h(\alpha, r, s)(1 + x_{\tau_o+m}^{(r+1)} - x_{\tau_o+m}^{(\tau_o+m-s)}),$$

где $h(\alpha, r, s)$ есть решение уравнения

$$\int_0^1 \operatorname{sign}(h-y)\,|\,h-y\,|^{\alpha-1} y^r (1-y)^s \, dy = 0.$$

(ii) *Для последовательного плана Δ_n^0 характеристическая функция с.в τ_0 равна*

$$\varphi(t) = \left(\frac{\varepsilon e^{it}}{1 - e^{it} + \varepsilon e^{it}} \right)^q,$$

где

$$\varepsilon = \frac{r+s+2}{n}, \quad q = r+s+1,$$

(84)
$$\mathbf{E}_\theta(\tau_0) = n, \quad \mathbf{D}_\theta(\tau_0) = \frac{n^2}{r+s+2} - n,$$

$$\mathbf{E}_\theta\,|\,t_{\tau_0+m}\,|^\alpha = \mu_\alpha \frac{(r+s+2)^{\alpha+1}}{(r+s+\alpha+2)n^\alpha} F(-m, \alpha; r+s+3; \tfrac{r+s+2}{n}),$$

$$\mu_\alpha = \frac{\Gamma(r+s+2)}{\Gamma(r+1)\Gamma(s+1)} \int_0^1 |\,h-y\,|^\alpha y^r (1-y)^s \, dy,$$

$F(a, b; c; x)$ *– гипергеометрическая функция Гаусса.*

(iii) *Для оптимальных последовательного Δ_n^0 и непоследовательного $\mathcal{D}_n$ планов имеет место соотношение*

$$\gamma_{n+m}(r,s) = \frac{(r+s+2)^\alpha\,\Gamma(r+s+3)\,\Gamma(n+m+\alpha+1)}{\Gamma(n+m+1)\,\Gamma(r+s+\alpha+3)\,n^\alpha}\, F\left(-m, \alpha; r+s+3; \frac{r+s+2}{n} \right).$$

Пусть

$$\gamma(r,s) = \lim_{n\to\infty} \gamma_n(r,s) \qquad (\ r+s = const\ ).$$

<u>Следствие 1.</u> *Если $r = s = 0,\ m = 0$, то*

$$\gamma(0,0) = 2^{\alpha+1} / \,\Gamma(\alpha+3).$$

<u>Замечание.</u> Как видно из представленной теоремы оптимальный план Δ_n^0 зависит от разности порядковых статистик $w_{rs} = x_k^{(k-s)} - x_k^{(r+1)}$. В книге Дэйвид Г. Порядковые Статистики, М.: Наука, 1979, с.21, найдена плотность распределения w_{rs}. Она имеет вид

$$f(w) = \frac{w^{k-s-r-2}(1-w)^{s+r+1}}{\beta(k-s-r-1, s+r+2)}, \qquad 0 \leqslant w \leqslant 1,$$

где $\beta(\alpha, \gamma)$ – бета-функция. Таким образом, эта плотность зависит от суммы $r+s$, а не от r и s по отдельности. Значит последовательный план тоже зависит только от $r+s$.

Следствие 2. *Если $m = const$, $r+s = const$, то относительная эффективность $\gamma_{n+m}(r,s)) < 1$, т.е. планы последовательного оценивания эффективны по сравнению с планами фиксированного объема. Кроме того, функция $\gamma(r,s)$ является монотонно возрастающей функцией $r+s$. При $r+s \to \infty$*

$$\gamma(r,s) = \sum_{k=0}^{N-1} \frac{(-1)^k B_k^{(1-\alpha)}(1)\,\Gamma(\alpha+k)}{\Gamma(\alpha)\,k!\,(r+s+2)^k} + O\left(\frac{1}{(r+s+2)^N}\right),$$

где $B_k^{(a)}(x)$ – обобщенный полином Бернулли.

При $N = 2$ имеем

$$\gamma(r,s) = 1 - \frac{\alpha(\alpha+1)}{2\,(r+s+2)} + \frac{\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{24\,(r+s+2)^2} + O(\frac{1}{(r+s+2)^3}).$$

Следствие 3. *Если $m = 0$, $r+s \to \infty$, то асимптотическая эффективность $\gamma(r,s) \to 1$.*

Следствие 4. *Если $r+s = const$, $n \to \infty$, то*

$$\gamma_n(r,s) \to \gamma(r,s) = \frac{(r+s+2)^\alpha\,\Gamma(r+s+3)}{\Gamma(r+s+\alpha+3)},$$

более того,

$$\gamma_{n+m}(r,s) = \frac{(r+s+2)^\alpha\,\Gamma(r+s+3)}{\Gamma(r+s+\alpha+3)} \cdot$$

$$\cdot \left(\sum_{k=0}^{N-1} \frac{(-1)^k B_k^{1-\alpha}(1)\,(\alpha)_k}{n^k} + O\left(\frac{1}{n^N}\right)\right) \left(\sum_{j=0}^{m} \frac{(-m)_j\,(\alpha)_j\,(r+s+2)^j}{(r+s+\alpha+3)\,j!\,n^j}\right),$$

где $(a)_k = \Gamma(a+k)/\Gamma(a)$.

Если $m \to \infty$, $n = const$, то

$$\gamma_{n+m}(r,s) \to 1.$$

Доказательство теоремы 38. Для случая $r \neq s$ оценка Питмена зависит от вида функции потерь и для $W(x) = |\,x\,|^\alpha$, $\alpha > 1$, имеет вид

$$t_k = x_k^{(k-s)} - 1/2 + h(\alpha,r,s)\,\ell_k = x_k^{(r+1)} + 1/2 - h(\alpha,s,r)\,\ell_k, \tag{85}$$

где $\ell_k = 1 + x_k^{(r+1)} - x_k^{(k-s)}$.

Тогда

$$W'(x) = \alpha\,|\,x\,|^{\,\alpha}\,\mathrm{sign}\,x, \quad W'(0) = 0,$$

и

$$p_k(y) = \frac{(x_k^{(r+1)} - y + 1/2)^r\,(1/2 - x_k^{(k-s)} + y)^s}{\int\limits_{x_k^{(k-s)}-1/2}^{x_k^{(r+1)}+1/2} (x_k^{(r+1)} - y + 1/2)^r\,(1/2 - x_k^{(k-s)} + y)^s\,dy}\ .$$

Тогда оценка t_k находится из уравнения

$$\int\limits_{x_k^{(k-s)}-1/2}^{x_k^{(r+1)}+1/2} |\,t_k - y\,|^{\,\alpha-1}\,\mathrm{sign}\,(t_k - y)\,(x_k^{(r+1)} - y + 1/2)^r\,(1/2 - x_k^{(k-s)} + y)^s\,dy = 0.$$

Производя в последнем интеграле замену переменных вида $y = 1/2 - x_k^{(k-s)} + u$, а затем $y = x_k^{(r+1)} - u + 1/2$, получим оценку в виде (85).

Очевидно, имеем равенство $h(\alpha,r,s) + h(\alpha,s,t) = 1$. В частности, $h(\alpha,r,r) = 1/2$ (не зависит от α), а при $\alpha = 2$ (в случае общих r и s) получаем $h(\alpha,r,s) = (r+1)/(r+s+2)$.

Теперь

$$b_k = \mathbf{E}_\theta(\,|\,t_k - \theta\,|^\alpha\,|\,\mathcal{F}_k) + ck = \mu_\alpha\,\ell_k^\alpha + ck.$$

Обозначим $\rho_k = \mathbf{E}_\theta(\,b_{k+1}\,|\,\mathcal{F}_k)$. Покажем, что

$$\rho_k = \mu_\alpha\,\ell_k^\alpha\,F(-m,\alpha;r+s+\alpha+2;\,\ell_k) + c\,(\,k+m\,). \tag{86}$$

Действительно,

$$\mathbf{E}_\theta(\,b_{k+1}\,|\,\mathcal{F}_k) = \mu_\alpha \left\{ \int\limits_{x_k^{(k-s)}-1/2}^{x_k^{(r+1)}+1/2} \int\limits_{\theta-1/2}^{x_k^{(r+1)}} \left[\int\limits_{\theta-1/2}^{z} |\,1+z-x_k^{(k-s)}\,|^\alpha\,dy + \right.\right.$$

$$\left. + \int\limits_{z}^{x_k^{(r+1)}} |\,1+y-x_k^{(k-s)}\,|^\alpha\,dy \right] r(z-\theta+1/2)^{r-1}(1/2-x_k^{(k-s)}+\theta)^s\,dzd\theta +$$

$$+ \int\limits_{x_k^{(k-s)}-1/2}^{x_k^{(r+1)}+1/2} \int\limits_{x_k^{(k-s)}}^{\theta+1/2} \left[\int\limits_{x_k^{(k-s)}}^{z} |\,1+z-x_k^{(r+1)}\,|^\alpha\,dy + + \int\limits_{z}^{\theta+1/2} |\,1+y-x_k^{(k-s)}\,|^\alpha\,dy \right] \times$$

$$\left. \times\ s(x_k^{(r+1)}-\theta+1/2)^r(1/2-z+\theta)^{s-1}\,dzd\theta \right\} \ell_k^{-r-s-1}\frac{\Gamma(r+s+2)}{\Gamma(r-1)\,\Gamma(s+1)} + \ell_k^\alpha(1-\ell_k) + c\,(k+1) =$$

$$= \mu_\alpha\,\ell_k^\alpha\left(1 - \frac{\alpha\,\ell_k}{r+s+2} + c\,(k+1)\right).$$

Используя свойства гипергеометрической функции, индукцией по m получим равенство (86). Отсюда

$$\mathbf{E}_\theta(\,\rho_{k+1}\,|\,\mathcal{F}_k) - \rho_k = \mu_\alpha\,\ell_k^\alpha\,(\,F(-m-1,\alpha;r+s+2;\,\ell_k) - F(-m,\alpha;r+s+2;\,\ell_k)\,) + c =$$

$$= -\,\frac{\alpha}{r+s+\alpha+2}\,\mu_\alpha\,\ell_k^\alpha\,F(-m,1+\alpha;r+s+2;\,\ell_k) + c.$$

Функция $x^\alpha\,F(-m,1+\alpha;r+s+2;x) = \varphi(x)$ монотонно возрастает по x, так как

$$\varphi'(x) = (\alpha+1)\,x^\alpha\,F(-m,1+\alpha;r+s+2;x) > 0, \quad 0 < x < 1$$

Выбирая

$$c = \frac{\alpha\,\mu_\alpha\,(r+s+2)^{\alpha+1}}{(r+s+\alpha+2)\,n^{\alpha+1}}\,F(-m,1+\alpha;r+s+\alpha+3;\tfrac{r+s+2}{n})$$

и используя леммы 9 и 10, получаем утверждение (i).

(ii) Найдем вероятность того, что $\tau_o = k$. Определим событие

$$A_k = \{\,x_k^{(k-s)} - x_k^{(r+1)} \leqslant 1-\varepsilon\}, \quad \varepsilon = (r+s+2)/n,$$

для фиксированных k, r и s. Тогда $(\tau_o = k) = A_{k-1}\bar{A}_k$ и $\mathbf{P}(\tau_o = k) = \mathbf{P}(\,A_{k-1}\,) - \mathbf{P}(\,\bar{A}_k\,)$, поскольку $A_k \subset A_{k-1}$.

Пусть $q = r+s+1$. Для равномерного на $(\theta-1/2,\theta+1/2)$ распределения получаем

$$\mathbf{P}(\,A_k\,) = \iint\limits_{0<y-z\leqslant 1-\varepsilon} \frac{\Gamma(k+1)}{\Gamma(r+1)\,\Gamma(k-q)\,\Gamma(s+1)}\,(z{+}1/2)^r\,(z{-}y)^{k-q}\,(1/2{-}y)^s\,dydz = \sum_{j=0}^{q} C_k^j\,\varepsilon^j\,(1{-}\varepsilon)^{k-j};$$

$$\mathbf{P}(\tau_o = k) = \sum_{j=0}^{q} C_{k-1}^j\,\varepsilon^j\,(1-\varepsilon)^{k-j-1} - \sum_{j=0}^{q} C_k^j\,\varepsilon^j\,(1-\varepsilon)^{k-j} = C_{k-1}^q\,\varepsilon^{q+1}\,(1-\varepsilon)^{k-q-1},$$

для $k = q+1, q+2, \ldots$.

Таким образом, с.в. τ_o имеет отрицательное биномиальное распределение. Её х.ф. равна

$$\varphi_{\tau_o}(t) = e^{itq}\sum_{k=q+1}^{\infty} e^{itk}C_{k-1}^q\,\varepsilon_{\,q+1}\,(1-\varepsilon)^{k-q-1} = \left(\frac{\varepsilon e^{it}}{1-e^{it}+\varepsilon e^{it}}\right)^{q+1},$$

отсюда находим, что

$$\mathbf{E}_\theta(\tau_o) = \frac{q+1}{\varepsilon} = n, \quad \mathbf{D}_\theta(\tau_o) = \frac{n^2}{r+s+2} - n.$$

Заметим также, что х.ф. с.в. τ_o/n равна

$$\psi_n(t) = \left(\frac{\varepsilon e^{it/n}}{1 - e^{it/n} + \varepsilon e^{it/n}}\right)^{q+1},$$

и для $r+s = const$

$$\psi_n(t) \underset{n\to\infty}{\longrightarrow} \left(1 - \frac{it}{r+s+2}\right)^{-(r+s+2)},$$

а если $r+s \to \infty$, то

$$\left(1 - \frac{it}{r+s+2}\right)^{-(r+s+2)} \to e^{it},$$

что является характеристической функцией вырожденного закона.

Из этих рассуждений следует, что предельное при $n \to \infty$ распределение τ_o/ε имеет плотность (для $r+s = const$) $\dfrac{x^{r+s+1}}{\Gamma(r+s+2)} e^{-x}$ при $x > 0$, с начальными моментами $\alpha_\ell = (r+s+2)\,(r+s+3)\cdot\ldots\cdot(r+s+\ell+1)$.

Докажем (iii). Не теряя в общности, предположим, что n – целое. Пусть $\tau(c)$ момент остановки, для которого

$$\mathbf{E}_\theta\,|\,t_{\tau(c)+m} - \theta\,|^{\,\alpha} + c\,\mathbf{E}_\theta(\tau(c)) = \min_\tau\{\mathbf{E}_\theta\,|\,t_{\tau+m} - \theta\,|^{\,\alpha} + c\,\mathbf{E}_\theta(\tau)\},$$

и пусть

$$Q(c) = \mathbf{E}_\theta\,|\,t_{\tau(c)+m} - \theta\,|^{\,\alpha}, \quad q(c) = \mathbf{E}_\theta(\tau(c)).$$

В нашем случае

$$q^{\,-1}(x) = \mu_\alpha\,\frac{(r+s+2)^\alpha}{(r+s+\alpha+2)\,n^\alpha}\,\alpha\,x^{\alpha-1}\,F(-m, 1+\alpha; r+s+\alpha+3; \tfrac{r+s+2}{n}),$$

поэтому

$$\mathbf{E}_\theta(\,|\,t_{\tau_o} - \theta\,|^{\,\alpha}\,) = \mu_\alpha\,\frac{(r+s+2)^{1+\alpha}}{(r+s+\alpha+2)\,n^\alpha}\int\limits_0^1 t^{\,\alpha-1} F(-m, 1+\alpha; r+s+\alpha+3; \varepsilon t)\,dt =$$

$$= \mu_\alpha\,\frac{(r+s+2)^{1+\alpha}}{(r+s+\alpha+2)\,n^\alpha}\,F(-m, 1+\alpha; r+s+\alpha+3; \tfrac{r+s+2}{n}).$$

Пользуясь свойствами условных математических ожиданий, свойствами гипергеометрических функций и формулой Эйлера (см. [Бейтмен, Эрдейи], с.89, 2.4.(2)) получим

$$\mathbf{E}_\theta(\,|\,t_{\tau_o} - \theta\,|^{\,\alpha}\,) = \mu_\alpha \mathbf{E}_\theta(\,\ell^\alpha_{n+m}\,) =$$

$$= \mu_\alpha\,\frac{\Gamma(n+m+1)}{\Gamma(r+1)\,\Gamma(s+1)\,\Gamma(n+m-r-s-1)}\iint\limits_{0\leqslant x<y\leqslant 1}(1+x-y)^{\,\alpha}\,x^r\,(y-x)^{n+m-r-s-2}\,(1-y)^s\,dydx =$$

$$= \mu_\alpha\,\frac{\Gamma(n+m+1)}{\Gamma(s+1)\,\Gamma(n+m-s)}\int\limits_0^1 y^{n+m-s-1}\,(1-y)^s\,F(-\alpha, n+m-r-s-1; n+m-s; y)\,dy =$$

$$= \mu_\alpha\,F(-\alpha, n+m-r-s; n+m+1; 1) = \mu_\alpha\,\frac{\Gamma(n+m+1)\,\Gamma(r+s+\alpha+2)}{\Gamma(r+s+2)\,\Gamma(n+m+\alpha+1)}.$$

Отсюда выводим (iii).

Пусть теперь $m = 0$. Обозначим

$$b_k^{(\alpha)} = \mathbf{E}_\theta(\,|\,t_k - \theta\,|^{\,\alpha}\,|\,\mathcal{F}_k\,), \quad \zeta_k = k^{\,\alpha} b_k^{(\alpha)},$$

и пусть плотность распределения $f(x-\theta)=0$, если $x\notin[\theta-1/2,\theta+1/2]$,

$$\lim_{x\downarrow\theta-1/2} f(x-\theta)=\lim_{x\uparrow\theta+1/2} f(x-\theta)=1.$$

Рассмотрим план $\Delta_\sigma=(\sigma,t_\sigma)$, где

$$\sigma=\min\,\{k:\, b_k^{(\alpha)}\leqslant\gamma^{\alpha}\,n^{-\alpha}\},\quad \gamma=\mathbf{E}_\theta(\zeta^{1/\alpha}),\quad \zeta_k=b_k^{(\alpha)}\,k^{\alpha}\xrightarrow[k\to\infty]{d}\zeta.$$

Теорема 39. (i) *При $n\to\infty$, $\alpha>1$*

$$\varlimsup_{n\to\infty} n^{-1}\mathbf{E}_\theta(\sigma)\leqslant 1,\quad \varlimsup_{n\to\infty}\frac{\mathbf{E}_\theta(\,|\,t_\sigma-\theta\,|^{\,\alpha}\,)}{\mathbf{E}_\theta(\,|\,t_n-\theta\,|^{\,\alpha}\,)}\leqslant\frac{\mathbf{E}^{\alpha}\zeta^{1/\alpha}}{\mathbf{E}\,\zeta}<1.$$

(ii) *Для равномерного на $(\theta-1/2,\theta+1/2)$ распределения план Δ_σ оптимален.*

Д о к а з а т е л ь с т в о.

(i) Из определения σ следует, что $b_\sigma^{(\alpha)}\leqslant\gamma^{\alpha}\,n^{-\alpha}$, так что $\mathbf{E}_0\,|\,t_\sigma\,|^{\,\alpha}=\mathbf{E}_0(b_\sigma^{(\alpha)})\leqslant\gamma^{\alpha}\,n^{-\alpha}$. С другой стороны

$$n^{\alpha}\mathbf{E}_0\,|\,t_\sigma\,|^{\,\alpha}=n^{\alpha}\mathbf{E}_0(b_\sigma^{(\alpha)})=\mathbf{E}_0(\zeta_n)\xrightarrow[n\to\infty]{}\mathbf{E}_0(\zeta).$$

По неравенству Ляпунова, если $\alpha>1$, то $\mathbf{E}^{\alpha}(\zeta^{1/\alpha})\leqslant\mathbf{E}(\zeta)$, причем равенство достигается, когда ζ – вырожденная случайная величина. Это доказывает первую часть теоремы.

Для доказательства (ii) заметим, что в случае равномерного распределения $b_k^{(\alpha)}=\mu_\alpha\,\ell_k^{\alpha}$, где $\ell_k=1-x_k^{(k-s)+x_k^{(r+1)}}$, а μ_α определено формулой (84). Отсюда следует, что

$$k\,\ell_k=k(1-x_k^{(k-s)}+x_k^{(r+1)})\xrightarrow[k\to\infty]{d}\xi_{r+s+2},$$

для $r+s=const$, где ξ_q имеет гамма-распределение с параметром q.

В таком случае

$$\gamma=\mathbf{E}_0(\zeta^{1/\alpha})=\mu_\alpha^{1/\alpha}\mathbf{E}_0(\xi_{r+s+2})=\mu_\alpha^{1/\alpha}\,(r+s+2)$$

и поэтому момент остановки имеет вид:

$$\sigma=\min_k\{x_k^{(k-s)}-x_k^{(r+1)}\geqslant 1-\tfrac{r+s+2}{n}\},$$

т.е. план $\Delta_\sigma=(\sigma,t_\sigma)$ оптимален.

Найдем математическое ожидание $\mathbf{E}_0(\zeta)$. Имеем

$$\mathbf{E}_0(\zeta)=\mu_\alpha\,\mathbf{E}_0(\xi_{r+s+2}^{\alpha})=\mu_\alpha\,\frac{\Gamma(r+s+\alpha+2)}{\Gamma(r+s+2)},$$

и поэтому

$$G(r,s)=\frac{\mathbf{E}_0^{\alpha}(\zeta^{1/\alpha})}{\mathbf{E}_0(\zeta)}=\frac{(r+s+2)^{\,\alpha}\,\Gamma(r+s+2)}{\Gamma(r+s+\alpha+2)},$$

в то время как точный расчет эффективности дает

$$\gamma(r,s)=\frac{\mathbf{E}_0^{\alpha}(\zeta^{1/\alpha})}{\mathbf{E}_0(\zeta)}=\frac{(r+s+2)^{\,\alpha}\,\Gamma(r+s+2)}{\Gamma(r+s+\alpha+3)}.$$

Ввиду того, что $\Gamma(k)/\Gamma(k+\alpha)\sim k^{-\alpha}$ при $k\to\infty$, получаем, что

$$G(r,s))\xrightarrow[(r+s)\,\to\infty]{}1.$$

Список литературы

[1] Крамер Г. Математические методы статистики. – М.: Наука, 1975. – 648 с.

[2] Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. – М.: Наука, 1979. – 528 с.

[3] Ibragimov I.A., Has'minskii R.Z. Statistical estimation: asymptotic theory. – NY.: Springer-Verlag, 1981. – 404 p.

[4] Hajek J., Sidak Z., Sen P. Theory of Rank Tests. – Academic Press, NY, 1999. – 435 p.

[5] Roussas G. Contiguity of Probability Measures. – Cambridge University Press, Cambridge, 2008. – 264 p.

[6] Тихов М.С. Оценивание по методу моментов и методу максимального правдоподобия: методические рекомендации – Горький, ГГУ, 1988. – 16 с.

[7] Тихов М.С. Контрольные задания по курсу «Теория вероятностей и математическая статистика» – Н. Новгород, ННГУ, 2003. – 28 с.

[8] Тихов М.С. Асимптотические методы в теории порядковых статистик: учебное пособие – Горький, ГГУ, 1985. – 80 с.

[9] Tikhov M.S. Bayes type estimates, *Journal of Soviet Mathematics*, 1991, v.56, no.3, p.2438-2442.

[10] Каган А.М. Фишеровская информация, содержащаяся в конечномерном линейном пространстве, и корректный вариант метода моментов, *Проблемы передачи информации*, 1976, т.12, в.2, с.20-42.

[11] Hansen L. Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 1982, v.50, no.4, p.1029-1054.

[12] Pearson K. Contribution to the mathematical theory of evolution, *PTRS*, 1894, v.185, p.71-110.

[13] Fisher R.A. On an absolute criterion for fitting frequency curves, *Messenger of Mathematics*, v.41, p.155-160.

[14] Lehmann E., Casella G. Theory of Point estimation. – NY, Springer, 1998. – 616 p.

[15] Хьюбер Дж. Робастность в статистике. – М.: Мир, 1984. – 304 с.

[16] У.Феллер. Введение в теорию вероятностей и ее приложения, т.1. – М.: Наука, 1984. – 528 с.

[17] Tikhov M.S. Asymptotic analysis of statistical estimates according to censored central and intermediate terms of the variational series of samples. – *Journal Soviet Mathem.*, 1984, v.25, no.3, p.1219-1230.

[18] Рао С.Р. Линейные статистические методы и их применения. – М.: Наука, 1968 – 547 с.

[19] Rao C.R. Linear models: least squares and alternatives. – NY, Springer, 1999 – 428 p.

[20] Айвазян С.А., Фантаццини Д. Эконометрика-2: продвинутый курс с приложениями в финансах. – М.: Магистр ИНФРА-М, 2014 – 944 с.

[21] Тихов М.С. Асимптотики Т-оценок.– *Теор.вер. и ее прим*, 1992, т.37, в.4, с.658-675.

[22] Tikhov M.S. Sequential estimation of the location parameter of a uniform distribution based on censored samples of type II. – *Journal Soviet Mathem.*, 1986, v.33, no.1, p.810-816.

Оглавление

.

СОВРЕМЕННЫЕ МЕТОДЫ
СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ПАРАМЕТРОВ

Составители:
Михаил Семенович **Тихов**
Мария Владимировна **Котельникова**
Учебно-методическое пособие

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное
учреждение высшего образования
«Национальный исследовательский Нижегородский государственный
университет им. Н.И. Лобачевского»

603950, Нижний Новгород, пр. Гагарина, 23.