ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ. ЧАСТЬ 1

В. Е. БЕНИНГ

МГУ, 2001 г.

Содержание

Введение

Предисловие

Данное учебное пособие предназначено для студентов и аспирантов математических специальностей университетов (математика, прикладная математика), знакомых с базовым университетским курсом теории вероятностей. Однако мы старались избегать слишком “продвинутых” в математическом отношении формулировок и доказательств, чтобы круг возможных читателей включал и специалистов в области прикладной статистики, желающих глубже ознакомиться с математическими аспектами математической статистики. Для удобства читателей в список литературы включены не только непосредственные источники приводимых результатов, но также и другие статьи и книги, которые, по мнению автора, могут оказаться полезными читателям, которые пожелают продолжить изучение математической статистики самостоятельно. Данное учебное пособие, состоящее из трёх частей, представляет собой курс лекций, который в течение многих лет автор читал студентам 3-го и 4-го курсов кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета им. М. В. Ломоносова в рамках обязательного курса “Дополнительные Главы Математической Статистики” и специального курса “Асимптотическая Статистика”, а также студентам отделения прикладной математики Вологодского государственного педагогического университета.

Со временем курс лекций неоднократно менялся в поисках варианта, который был бы по возможности более стройным и цельным, доступным и в то же время соответствовал современному состоянию предмета. Помимо традиционных разделов данный курс лекций содержит и такие разделы, мало освещённые в отечественной литературе, как эмпирический байесовский подход, асимптотическое разложение Эджворта и их применения в задачах проверки статистических гипотез, асимптотическое разложение Корниша - Фишера, "поправленное" или седловое асимптотическое разложение, симметричные статистики и разложение Хёффдинга, бутстрэп.

Преподаватели ВУЗов, уже знакомые хотя бы частично с математической статистикой, могут выбирать из книги совокупность лекций, используя которые (не обязательно полностью) можно составить семестровый курс математической статистики.

Автор признателен редактору д. ф.-м. н. профессору В.Ю. Королёву за тщательное редактирование и плодотворные обсуждения, рецензентам книги д. ф.-м. н. профессору А. И. Зейфману и д. ф.-м. н. С. Я. Шоргину, а также заведующему кафедрой математической статистики и эконометрики Тверского государственного университета д. ф.-м. н. профессору Ю. С. Хохлову за советы, которые, несомненно, способствовали улучшению изложения. Работа над книгой поддерживалась грантами Российского фонда фундаментальных исследований, проекты 99-01-00847, 00-01-00360 и 00-01-00657, Российского гуманитарного научного фонда, проект 00-02-00152а, а также the Committee on Knowledge Extension Research (CKER) Северо-Американского Общества Актуариев.

Обозначения

$\mathsf{P}(A)$	–	вероятность события A;
$\mathsf{E}X$	–	математическое ожидание случайной величины X;
$\mathsf{D}X$	–	дисперсия случайной величины X;
$\mathsf{Cov}(X,Y)$	–	ковариация случайных величин X и Y;
mX	–	медиана случайной величины X;
$\Longrightarrow$	–	слабая сходимость (сходимость по распределению);
$\xrightarrow{P_{n\theta}}$	–	сходимость по вероятности;
P-п.н.	–	почти наверное относительно вероятностной меры P (с вероятностью единица);
$\stackrel{d}{=}$	–	совпадение распределений;
$\mathbf{R}^1$	–	множество действительных чисел;
$\mathcal{B}^1$	–	борелевская σ – алгебра на прямой;
$\mathbf{R}^k$	–	прямое произведение k множеств $\mathbf{R}^1$;
$\mathcal{B}^k$	–	борелевская σ – алгебра множеств из $\mathbf{R}^k$;
$\mathbf{1}_A(\cdot)$	–	индикатор множества A;
$\Upsilon(\Omega)$	–	множество всех подмножеств множества Ω;
$\mathbf{N}$	–	множество натуральных чисел;
$\nu(\cdot)$	–	доминирующая мера;
$\mathcal{B}(n,p)$	–	биномиальное распределение с параметрами (n,p);
$\mathcal{P}(\lambda)$	–	распределение Пуассона с параметром λ;
$\mathcal{N}(\mu,\sigma)$	–	нормальное распределение с параметрами (μ,σ);
$\mathcal{R}(a,b)$	–	равномерное распределение на отрезке $[a,b]$;
$\Phi(x), \quad \varphi(x)$	–	функция распределения и плотность
$\overline{X}$	–	$\frac{1}{n}\sum\limits_{i=1}^{n} X_i$;
$\mathcal{L}(\theta;x)$	–	функция правдоподобия;
$l(\theta;x)$	–	логарифм функции правдоподобия $\mathcal{L}(\theta;x)$;
$L(\theta,\delta)$	–	функция потерь;
$R(\theta,\delta)$	–	риск оценки $\delta(X)$;

$r(\delta, Q)$	–	байесовский риск оценки $\delta(X)$, соответствующий априорному распределению Q;
$\delta_Q(X)$	–	байесовская оценка, соответствующая априорному распределению Q;
$r(Q)$	–	байесовский риск байесовской оценки $\delta_Q(X)$, соответствующий априорному распределению Q;
$\hat{\delta}(X)$	–	оценка максимального правдоподобия функции $g(\theta)$;
$\delta_*(X)$	–	минимаксная оценка функции $g(\theta)$;
$\delta^*(X)$	–	оптимальная оценка функции $g(\theta)$;
$\bar{\delta}(X)$	–	эффективная оценка функции $g(\theta)$;
$\mathcal{X}$	–	выборочное пространство;
$\mathcal{F}$	–	исходная σ – алгебра;
Θ	–	параметрическое пространство;
$(\Theta,\ \mathcal{V})$	–	измеримое параметрическое пространство;
$Q(\cdot)$	–	априорное распределение на $(\Theta,\ \mathcal{V})$;
$Q_x(\cdot)$	–	апостериорное распределение на $(\Theta,\ \mathcal{V})$;
$\{\mathsf{P}_\theta(\cdot), \theta \in \Theta\}$	–	исходное семейство вероятностных мер;
$g(\theta)$	–	оцениваемая функция;
u_α	–	решение уравнения $\Phi(x) = \alpha$;
a^+	–	$\max\{0,\ a\}, \quad a \in \mathbf{R}^1$;
a^-	–	$\max\{0,\ -a\}, \quad a \in \mathbf{R}^1$;
$\square$	–	конец доказательства;

В книге используется стандартная система нумерации формул и утверждений (определений, теорем, лемм, следствий, примеров и замечаний). Каждое из упомянутых утверждений снабжено тройным индексом: первое число – номер лекции, второе – номер раздела и третье число – непосредственный номер утверждения в этом разделе. Аналогичная нумерация применена и к формулам. Например, ссылка на формулу (4.1.1) означает ссылку на первую формулу первого раздела четвертой лекции.

Лекция 1

В Лекции приводятся основные определения из теории вероятностей, касающиеся измеримости, случайных величин и функций распределения. Подробные доказательства имеются в стандартных учебниках по теории вероятностей, приведенных во втором разделе.

1.1 СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ФУНКЦИИ РАСПРЕДЕЛЕНИЯ

1) Пусть Ω – некоторое непустое множество произвольной природы, состоящее из элементов ω. Эти элементы называются *элементарными событиями*, а множество Ω называется *пространством элементарных событий*.

   Пусть $\mathcal{A}$ – некоторое множество подмножеств пространства элементарных событий Ω, обладающее следующими свойствами:

   (a) $\Omega \in \mathcal{A}$,

   (b) если $A \in \mathcal{A}$, то $A^c \in \mathcal{A}$,

   (c) если $A_i \in \mathcal{A},\ i = 1, 2, \cdots$, то и

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}, \quad \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}.$$

   Множество $\mathcal{A}$ называется σ – *алгеброй событий*, или *борелевским полем событий*, а его элементы называются *измеримыми множествами* или *событиями*.

   Множество Ω вместе с σ – алгеброй его подмножеств называется *измеримым пространством* и обозначается $(\Omega, \mathcal{A})$.

Очевидным образом системы множеств

$$\underline{\mathcal{A}} = \{\emptyset, \Omega\}, \quad \overline{\mathcal{A}} = \{A : A \subseteq \Omega\} \equiv \Upsilon(\Omega)$$

являются σ – алгебрами. При этом $\underline{\mathcal{A}}$ – тривиальная, самая "бедная" σ – алгебра, а $\overline{\mathcal{A}}$ – самая "богатая" σ – алгебра, состоящая из всех подмножеств Ω.

Счётно-аддитивная мера P, определённая на $\mathcal{A}$ и нормированная условием $\mathsf{P}(\Omega) = 1$, называется *вероятностной мерой* или *вероятностью*. Значение $\mathsf{P}(A)$ называется *вероятностью события* A. Тройка $(\Omega, \mathcal{A}, \mathsf{P})$ называется *вероятностным пространством.*

2) $(\mathcal{X}, \mathcal{B})$ – *измеримое пространство*, где $\mathcal{X}$ – некоторое множество и $\mathcal{B}$ – σ – алгебра его подмножеств.

3) Отображение (функция) $X = X(\omega)$ вида

$$X : \Omega \longmapsto \mathcal{X}$$

называется *измеримым* ($\mathcal{A}$ – измеримым), если

$$X^{-1}(B) \in \mathcal{A} \quad \text{для всех} \quad B \in \mathcal{B},$$

где $X^{-1}(B)$ – *полный прообраз множества* B,

$$X^{-1}(B) = \{\omega : X(\omega) \in B\}.$$

Класс множеств вида

$$\mathcal{A}_X = \{X^{-1}(B) : \ B \in \mathcal{B}\} \subseteq \mathcal{A}$$

называется σ – алгеброй, порождённой X. Ясно, что свойство измеримости зависит от σ – алгебры $\mathcal{A}$.

4) Измеримая функция X называется *случайным элементом* (*случайным элементом со значениями в* $\mathcal{X}$). Действительная ($\mathcal{X} = \mathbf{R}^1$, $\mathcal{B}$ – *борелевская* σ – алгебра, то есть наименьшая σ – алгебра, содержащая все интервалы) конечная измеримая функция называется *случайной величиной*. Простейшим примером случайной величины является индикатор $\mathbf{1}_A(\omega)$ любого множества $A \in \mathcal{A}$

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}, \quad A \in \mathcal{A}.$$

Другим примером случайной величины служит дискретная случайная величина, принимающая не более чем счётное множество различных значений $\{x_1, x_2, \ldots\}$. Очевидно, что события $A_i = \{\omega : X(\omega) = x_i\}$ непересекаются и $\bigcup_i A_i = \Omega$ (всюду далее будем обозначать объединение *непересекающихся* множеств символом $\sum$, то есть

$$C = \sum_i A_i \Longleftrightarrow C = \bigcup_i A_i, \quad A_i \cap A_j = \emptyset, i \neq j).$$

Пусть

$$\mathsf{P}(A_i) = \mathsf{P}(X = x_i) = p_i.$$

Набор вероятностей $\{p_i\}$ и числа $\{x_i\}$ называются *распределением* дискретной случайной величины X. Оно полностью определяет вероятность попадания случайной величины X в любое борелевское множество $B \in \mathcal{B}$

$$\mathsf{P}(X \in B) = \sum_{i: x_i \in B} p_i.$$

Следует отметить, что для любого борелевского множества $B \in \mathcal{B}$ требование измеримости позволяет говорить о вероятностях событий вида $\{\omega : X(\omega) \in B\}$ состоящих в том, что значения случайной величины принадлежат некоторому борелевскому множеству B. Поэтому можно говорить о вероятностной мере P_X, определённой на множестве всех борелевских множеств $\mathcal{B}$ с помощью равенства

$$\mathsf{P}_X(B) = \mathsf{P}(\omega : X(\omega) \in B), \quad B \in \mathcal{B}.$$

Эта вероятностная мера называется *распределением* случайной величины X. В дальнейшем мы часто будем использовать более короткое обозначение $\mathsf{P}(X \in B)$ вместо $\mathsf{P}(\omega : X(\omega) \in B)$. Таким образом всякая случайная величина X порождает новое вероятностное пространство $(\mathbf{R}^1, \mathcal{B}, \mathsf{P}_X)$.

5) Случайная элемент X индуцирует на пространстве $(\mathcal{X},\ \mathcal{B})$ вероятностную меру P_X, называемую *распределением случайного элемента* X, вида

$$\mathsf{P}_X(B) = \mathsf{P}(X^{-1}(B)), \qquad B \in \mathcal{B}.$$

6) Пусть $\mathcal{X} = \mathbf{R}^1$, рассмотрим вероятности $\mathsf{P}(X \in B)$ в случае, когда множества B есть интервалы $(-\infty, x)$, то есть пусть $B = (-\infty, x)$. Положим в этом случае

$$F_X(x) \equiv F(x) = \mathsf{P}(X < x).$$

Функция $F(x)$ определена для любого действительного x и называется *функцией распределения* случайной величины X. Если X – дискретная случайная величина, для которой $\mathsf{P}(X = x_i) = p_i$, то

$$F_X(x) = \sum_{i:x_i<x} p_i.$$

Функция распределения $F(x)$ произвольной случайной величины обладает следующими свойствами:

(a) $F(x)$ не убывает и непрерывна слева,

(b) $\lim_{x\to-\infty} F(x) = 0$,

(c) $\lim_{x\to+\infty} F(x) = 1$.

Верно и обратное: любая функция $F(x)$, удовлетворяющая этим трём условиям, является функцией распределения некоторой случайной величины, определённой на некотором вероятностном пространстве.
Справедливо следующее

Утверждение 1.1.1. *Функция распределения $F(x)$ имеет самое большее конечное число точек, скачок в которых больше или равен $\delta > 0$ и, следовательно, самое большее счтное число точек разрыва. Производная $F'(x)$ функции $F(x)$ существует для почти всех значений x.*

$F(x)$ всегда может быть единственным образом представлена в виде суммы трх компонент:

$$F(x) = a_1 F_1(x) + a_2 F_2(x) + a_3 F_3(x),$$

где $a_1,\ a_2,\ a_3$ – некоторые неотрицательные числа, сумма которых равна единице, а $F_1(x),\ F_2(x),\ F_3(x)$ – функции рапределения, такие, что: $F_1(x)$ абсолютно непрерывна, то есть

$$F_1(x) = \int_{-\infty}^{x} F_1'(t)dt, \quad \textit{для всех} \ \ x;$$

(под интегралом понимается интеграл Лебега),
$F_2(x)$ – ступенчатая функция распределени, равная сумме скачков функции распределения $F(x)$ во всех точках разрыва меньших x;
$F_3(x)$ – сингулярная компонента, то есть непрерывная функция, производная которой почти всюду равна нулю.

Расмотрим, в частности, случаи, когда a_1 или a_2 равны единице, так что $F(x)$ совпадает с $F_1(x)$ или $F_2(x)$. Эти случаи чаще всего встречаются в приложениях. В первом случае распределение случайной величины X будем называть *абсолютно непрерывным* (можно показать, что распределение случайной величины X абсолютно непрерывно, если $\mathsf{P}(X \in B) = 0$ для любого борелевского множества $B \in \mathcal{B}$ нулевой лебеговой меры), производная $p_X(x) \equiv p(x) = F'(x)$ называется в этом случае *плотностью распределения* случайной величины X. Мы будем говорить о плотности распределения только в том случае, когда это распределение абсолютно непрерывно. Во втором случае, когда $a_2 = 1$ распределение X или сама случайная величина X называются *дискретными*. В этом случае существует конечное или счётное множество $\mathbf{X}$ точек действительной прямой, такое, что $\mathsf{P}(X \in \mathbf{X}) = 1$.

Если X – случайная величина с дискретным рапределением и $\mathsf{P}(X = x) > 0$, то число x называется *возможным значением* случайной величины X. Случайная величина X имеет *решетчатое распределение*, если с вероятностью единица она принимает значения вида $b+nh$, $n = 0, \pm 1, \pm 2, \ldots$, где b и $h > 0$ – фиксированные числа. Число h называтся *шагом рапределения*. Если ни при каких b_1 и $h_1 > h$ значения, принимаемые случайной величиной X с вероятностью единица, не могут быть представлены в виде $b_1 + nh_1$, $n = 0, \pm 1, \pm 2, \ldots$, то шаг h называется *максимальным*.

Особенно важную роль играют три дискретных рапределения – *вырожденное, биномиальное и пуассоновское*, и два абсолютно непрерывных распределения – *нормальное распределение* и *равномерное распределение*.

(a) Вырожденное рапределение. Случайная величина X имеет вырожденное рапределение, сосредеточенное в точке $a \in \mathbf{R}^1$, если

$$\mathsf{P}(X = a) = 1. \tag{1.1.1}$$

Функция распределения $F(x)$ равна нулю при $x \le a$ и равна единице при $x > a$. Вырожденное распределение описывает неслучайные величины.

(b) Биномиальное рапределение. Случайная величина X имеет биномиальное рапределение с параметрами (n, p), $0 < p < 1$, $n \ge$

1, если

$$\mathsf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{1.1.2}$$

Этот факт мы будем обозначать в виде

$$X \sim \mathcal{B}(n, p).$$

Функция распределения $F(x)$ равна нулю при $x \le 0$, равна единице при $x > n$ и равна

$$\sum_{k=1}^{l} \binom{n}{k} p^k (1-p)^{n-k},$$

при $l < x \le l+1$. Биномиальное распределение описывает случайный эксперимент состоящий из n независимых испытаний с вероятностью p наступления некоторого события в отдельном испытании. Тогда распределение общего числа наступлений этого события в эксперименте является биномиальным с параметрами n и p.

(c) Распределение Пуассона. Случайная величина X имеет рапределение Пуассона с параметрам $\lambda > 0$, если

$$\mathsf{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots \tag{1.1.3}$$

Обозначение:

$$X \sim \mathcal{P}(\lambda).$$

Распределение Пуассона даёт хорошую аппроксимацию биномиального распределения для больших n и малых значений p (случай редких событий).

(d) Нормальное рапределение. Случайная величина X имеет нормальное рапределение с параметрам (μ, σ), $\mu \in \mathbf{R}^1$, $\sigma > 0$, если она имеет плотность вида

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-(x-\mu)^2/2\sigma^2\}. \tag{1.1.4}$$

Обозначим это как

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

Нормальную функцию распределения и плотность с параметрами $(0, 1)$ всюду в дальнейшем будем обозначать через $\Phi(x)$ и $\varphi(x)$ и называть стандартными. Таким образом

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\{-x^2/2\}, \quad \Phi(x) = \int_{-\infty}^{x} \varphi(t)dt. \tag{1.1.5}$$

Фундаментальная роль, которое играет нормальное распределение, объясняется тем, что при широких предположениях суммы случайных величин с ростом числа слагаемых ведут себя асимптотически нормально.

(e) Равномерное распределение. Случайная величина X имеет равномерное рапределение на отрезке $[a, b]$, если она имеет плотность вида

$$p(x) = \begin{cases} (b-a)^{-1}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases} \tag{1.1.6}$$

Обозначим это как

$$X \sim \mathcal{R}(a, b).$$

Равномерное распределение естественно возникает в случаях полного отсутствия информации или при наличии симметрии.

7) Рассмотрим некоторое измеримое пространство ($\mathcal{F}$, $\mathcal{C}$) и $\mathcal{B}$ – измеримую функцию $T = T(x)$ со значениями в $\mathcal{F}$,

$$T : \mathcal{X} \longmapsto \mathcal{F}.$$

Тогда *суперпозицией* функций T и X называется функция

$$\Psi(\omega) = T(X(\omega)).$$

При этом эта функция является случайной величиной со значениями в $\mathcal{F}$. Обозначим её распределение через

$$\mathsf{P}_\psi(C) = \mathsf{P}(\Psi^{-1}(C)), \ C \in \mathcal{C}.$$

8)

$$\Psi^{-1}(C) = X^{-1}(T^{-1}(C)),$$

$$\mathsf{P}_\psi(C) = \mathsf{P}(\Psi^{-1}(C)) = \mathsf{P}(X^{-1}(T^{-1}(C))) = \mathsf{P}_X(T^{-1}(C)), \ C \in \mathcal{C},$$

то есть, если известно распределение случайной величины X и функция $T(x)$, то известно и распределение $T(X)$.

9) Справедливо следующее

УТВЕРЖДЕНИЕ 1.1.2 ([1], СТР. 60, ПРЕДЛОЖЕНИЕ 2.2.5). *Пусть* $\mathcal{A}_X \subset \mathcal{A}$ *есть* σ *– подалгебра* σ *– алгебры* $\mathcal{A}$*, порожденная случайной величиной* X*. Для того, чтобы случайная величина* Y *определенная на* $(\Omega, \mathcal{A})$ *была* $\mathcal{A}_X$ *– измеримой, необходимо и достаточно, чтобы* Y *можно было представить в виде*

$$Y = f(X),$$

где $f(x)$ *– измеримая функция отображающая* $\mathbf{R}^1$ *в себя.*

1.2 СПИСОК ЛИТЕРАТУРЫ

1) Ж. Невё, Математические Основы Теории Вероятностей,
Москва, Мир, 1969, Главы 1 – 2.

2) А.Н. Ширяев, Вероятность,
Москва, Наука, 1989, Главы 1 – 2.

3) М. Лоэв, Теория Вероятностей,
Москва, Иностранная Литература, 1962, Часть 1, Глава 1; Часть 2, Глава 4.

4) И.П. Натансон, Теория Функций Вещественной Переменной,
Москва, Наука, 1974, Главы 8 – 9.

Лекция 2

В Лекции приводится без доказательства основная схема построения интеграла Лебега. Определяются такие характеристики случайных величин как математическое ожидание, дисперсия, мода, медиана, асимметрия и эксцесс.

2.1 ИНТЕГРАЛ ЛЕБЕГА. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

1) Пусть $X(\omega)$ случайная величина, заданная на вероятностном пространстве $(\Omega, \mathcal{A}, \mathsf{P})$. Поскольку вероятностное пространство есть измеримое пространство с мерой, то можно ввести понятие интеграла. Опишем кратко схему построения интеграла Лебега.
Действительная случайная величина X $(\mathcal{X} = \mathbf{R^1})$ называется *ступенчатой*, если существует разбиение множества Ω

$$\Omega_1, \cdots, \Omega_N, \quad \bigcup_{i=1}^{N} \Omega_i = \Omega, \quad \Omega_i \cap \Omega_j = \emptyset, \quad i \neq j$$

такое, что

$$X(\omega) = x_i, \quad \omega \in \Omega_i, \quad i = 1, \cdots, N,$$

причём все x_i различны.

Обозначим через $\mathbf{S}$ – класс всех ступенчатых случайных величин.

2) *Индтикатором* (*индикаторной функцией*) назывется функция вида

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}, \quad A \in \mathcal{A}.$$

Используя индикаторы, ступенчатую случайную величину можно записать в виде

$$X \in \mathbf{S} \Longrightarrow X = \sum_{i=1}^{N} x_i \mathbf{1}_{\Omega_i}(\omega).$$

3) Если X *неотрицательная* случайная величина

$$X : \Omega \longmapsto \mathbf{R}^+ = [0, +\infty),$$

то существует неотрицательная неубывающая последовательность ступенчатых случайных величин

$$X_n \in \mathbf{S}, \quad X_n \geq 0, \quad X_n \leq X_{n+1},$$

монотонно сходящаяся к X

$$X_n \uparrow X, \quad n \to +\infty.$$

4) Для ступенчатой случайной величины X определим *математическое ожидание* по формуле

$$X \in \mathbf{S} \Rightarrow \mathsf{E}X \stackrel{def}{=} \sum_{i=1}^{N} x_i \mathsf{P}(\Omega_i).$$

5) Математическое ожидание для неотрицательных случайных величин теперь определим как

$$X \geq 0, \quad \mathsf{E}X \stackrel{def}{=} \operatorname*{Lim}_{n\to\infty} \mathsf{E}X_n, \quad \text{где} \quad X_n \in \mathbf{S}, \quad X_n \uparrow X.$$

6) Случайную величину X, принимающую произвольные знаки, можно всегда представить в виде *разности неотрицательных* случайных величин

$$X = X^+ - X^-, \quad \text{где} \quad X^+ = \max(X, 0),\ X^- = \max(-X, 0).$$

Теперь определим математическое ожидание произвольной случайной величины по формуле

$$\mathsf{E}X \stackrel{def}{=} \mathsf{E}X^+ - \mathsf{E}X^- \equiv \int_\Omega X(\omega) d\mathsf{P}(\omega) \equiv \int X(\omega) d\mathsf{P}(\omega),$$

$$\int_A X(\omega) d\mathsf{P}(\omega) \stackrel{def}{=} \mathsf{E}X\mathbf{1}_A.$$

7) Случайная величина X называется *интегрируемой*, если

$$\mathsf{E}X^{+} < +\infty, \qquad \mathsf{E}X^{-} < +\infty.$$

8) На множестве *интегрируемых* случайных величин математическое ожидание обладает следующими свойствами

(a) Линейность

$$\mathsf{E}\alpha X = \alpha\mathsf{E}X, \quad \mathsf{E}(\alpha X + \beta Y) = \alpha\mathsf{E}X + \beta\mathsf{E}Y, \quad \alpha, \beta \in \mathbf{R}^1.$$

(b) Сохранение порядка

$$X \le Y \Longrightarrow \mathsf{E}X \le \mathsf{E}Y.$$

(c) Непрерывность относительно монотонной сходимости

$$X_n \uparrow\downarrow X \Longrightarrow \mathsf{E}X_n \uparrow\downarrow \mathsf{E}X.$$

(d) Лемма Фату.

$$X_n \ge Z, \quad Z - \text{интегририруема} \Longrightarrow \mathsf{E}\liminf_{n\to+\infty} X_n \ \le\ \liminf_{n\to+\infty} \mathsf{E}X_n,$$

$$X_n \le Y, \quad Y - \text{интегририруема} \Longrightarrow \mathsf{E}\limsup_{n\to+\infty} X_n \ge \limsup_{n\to+\infty} \mathsf{E}X_n.$$

Отсюда следует Теорема Лебега о мажорируемой сходимости: если последовательность случайных величин X_n сходится и существует интегрируемая случайная величина $U \ge 0$ такая, что

$$|X_n| \le U, \quad \mathsf{E}|U| < +\infty,$$

то

$$\mathsf{E}\operatorname*{Lim}_{n\to\infty} X_n = \operatorname*{Lim}_{n\to\infty} \mathsf{E}X_n.$$

(e) σ – аддитивность неопределённого интеграла

$$X \ge 0 \Longrightarrow \int\limits_{\bigcup\limits_{i=1}^{\infty} A_i} X(\omega)d\mathsf{P}(\omega) = \sum_{i=1}^{\infty}\int\limits_{A_i} X(\omega)d\mathsf{P}(\omega),$$

где $\{A_i \in \mathcal{A},\ i = 1, 2, \cdots\}$ – семейство попарно непересекающихся измеримых множеств.

$$A_n \uparrow\downarrow A \Longrightarrow \int\limits_{A_n} X(\omega)d\mathsf{P}(\omega) \uparrow\downarrow \int\limits_{A} X(\omega)d\mathsf{P}(\omega), \quad n \to +\infty.$$

(f) Формула замены переменного: если

$$\Psi(\omega) = T(X(\omega)), \quad \mathcal{F} = \mathbf{R}^1$$

то

$$\mathsf{E}\Psi = \int\limits_{\Omega} T(X(\omega))d\mathsf{P}(\omega) = \int\limits_{\mathbf{R}^1} td\mathsf{P}_{\psi}(t) = \int\limits_{\mathcal{X}} T(x)d\mathsf{P}_X(x).$$

(g) Теорема Радона – Никодима. *Пусть на измеримом пространстве $(\mathcal{X}, \mathcal{B})$ имеется σ – конечная мера ν и σ – аддитивная функция множеств $\mathbf{G}$, абсолютно непрерывная относительно ν ($\mathbf{G} \ll \nu$). Тогда существует плотность $\mathbf{G}$ относительно ν, то есть существует $\mathcal{B}$ – измеримая функция $g(x)$ такая, что*

$$\mathbf{G}(B) = \int\limits_{B} g(x)d\nu(x), \quad \textit{для всех} \quad B \in \mathcal{B}.$$

Функция $g(x)$ называется производной *Радона – Никодима* и обозначается как

$$g(x) = \frac{d\mathbf{G}(x)}{d\nu} \equiv \frac{d\mathbf{G}}{d\nu}(x).$$

С точностью до ν – множеств меры нуль функция $g(x)$ единственна. Теорема Радона – Никодима даёт общие условия существования плотности.

Пример. Пусть на измеримом пространстве $(\mathcal{X}, \mathcal{B})$ заданы две вероятностные меры P_1, P_2, имеющие соответственно плотности $p_1(x)$, $p_2(x)$ относительно меры ν. Предположим, что выполнено условие

$$p_2(x) = 0 \implies p_1(x) = 0.$$

Тогда

$$\mathsf{P}_1(B) = \int\limits_{B} p_1(x)d\nu(x) = \int\limits_{B} \frac{p_1(x)}{p_2(x)} p_2(x)\, d\nu(x) = \int\limits_{B} \frac{p_1(x)}{p_2(x)}\, d\mathsf{P}_2(x), \ \ B \in \mathcal{B}.$$

То есть производна Радона – Никодима меры P_1 относительно меры P_2 имет вид

$$\frac{d\mathsf{P}_1}{d\mathsf{P}_2}(x) = \frac{p_1(x)}{p_2(x)}.$$

(h) Если через $F(x)$ обозначить функцию распределения случайной величины X, то справедливо равенство

$$\mathsf{E}X = \int\limits_{-\infty}^{+\infty} x dF(x),$$

здесь в правой части стоит интеграл Лебега – Стилтьеса. Если у случайной величины X существует плотность $p_X(x)$, то

$$\mathsf{E}X = \int\limits_{-\infty}^{+\infty} x p_X(x) dx.$$

Если $h(x)$ борелевская функция, то есть действительная функция, область определения которой является действительная прямая и при любом $c \in \mathbf{R}^1$ множество $\{x : h(x) < c\}$ является борелевским, то

$$\mathsf{E}h(X) = \int\limits_{-\infty}^{+\infty} h(x) dF(x),$$

при условиии, что существует хотя бы один из этих интегралов.

9) *Моментами* α_s и *абсолютными моментами* β_s порядка $s > 0$ случайной величины X называются математические ожидания случайных величин X^s и $|X|^s$

$$\alpha_s = \mathsf{E}X^s = \int\limits_{-\infty}^{+\infty} x^s dF(x), \tag{2.1.1}$$

$$\beta_s = \mathsf{E}|X|^s = \int\limits_{-\infty}^{+\infty} |x|^s dF(x). \tag{2.1.2}$$

Центральный момент μ_s и *абсолютный центральный моментами* ν_s порядка $s > 0$ случайной величины X определяются соответственно равенствами

$$\mu_s = \mathsf{E}(X - \mathsf{E}X)^s = \int\limits_{-\infty}^{+\infty} (x - \alpha_1)^s dF(x), \tag{2.1.3}$$

$$\nu_s = \mathsf{E}|X - \mathsf{E}X|^s = \int\limits_{-\infty}^{+\infty} |x - \alpha_1|^s dF(x). \tag{2.1.4}$$

Особую роль играет второй центральный момент μ_2, который называется *дисперсией* случайной величины X и обозначается символом $\mathsf{D}X$, то есть

$$\mathsf{D}X = \mathsf{E}(X - \mathsf{E}X)^2 = \mathsf{E}X^2 - (\mathsf{E}X)^2.$$

Заметим, что $\mathsf{D}X$ всегда определена, если определено $\mathsf{E}X$, но может принимать значения $+\infty$.

Величина $\sigma = \sqrt{\mathsf{D}X}$ называется *среднеквадратичным отклонением* случайной величины X.

Отметим важное свойство величины $\mathsf{D}X$: если $\mathsf{D}X = 0$, то

$$\mathsf{P}(X = \mathsf{E}X) = 1,$$

то есть в этом случае случайная величина X с вероятностью единица постоянна. Далее, если дисперсия конечна, то

$$\mathsf{D}(aX + b) = a^2\mathsf{D}X, \quad a, \ b \in \mathbf{R}^1.$$

В частности, нормированная случайная величина

$$X^* = \frac{X - \mathsf{E}X}{\sqrt{\mathsf{D}X}} \tag{2.1.5}$$

всегда имеет среднее 0 и дисперсию 1.

Если у случайной величины X существует момент α_k порядка k, то $\beta_m^{1/m} \le \beta_k^{1/k}$ и $\nu_m^{1/m} \le \nu_k^{1/k}$ для любого положительного $m \le k$. Отсюда следует, что $\beta_m\beta_l \le \beta_{m+l}$ и $\nu_m\nu_l \le \nu_{m+l}$ для любых l и m.

Рассмотрим теперь некоторые характеристики формы и расположения распределения случайной величины.

Если случайная величина X абсолютно непрерывна, то значения x, в которых плотность $p_X(x)$ достигает своего максимального значения, называются *модами*. Если мода единственна, то распределение случайной величины называют *унимодальным*, в противном случае – *мультимодальным*.

Если X – дискретная случайная величина и

$$p_k = \mathsf{P}(X = x_k),$$

то её модами называют те значения x_i, для которых

$$\mathsf{P}(X = x_i) = \max_k p_k.$$

Медианой случайной величины X называется любое число mX, для которого справедливы соотношения

$$\mathsf{P}(X \geq mX) \geq \frac{1}{2} \quad \text{и} \quad \mathsf{P}(X \leq mX) \geq \frac{1}{2}.$$

Для случайной величины X с абсолютно непрерывным распределением медиана определяется как значение mX, для которого

$$\int_{-\infty}^{mX} p_X(x)dx = \int_{mX}^{\infty} p_X(x)dx = \frac{1}{2}.$$

Квантиль порядка α, $\alpha \in (0,1)$, есть значение x_α, для которого

$$\mathsf{P}(X \geq x_\alpha) \geq \alpha \quad \text{и} \quad \mathsf{P}(X \leq x_\alpha) \geq \alpha.$$

Если X – случайная величина с абсолютно непрерывным распределением, то квантиль x_α порядка α определяется равенством

$$F_X(x_\alpha) = \alpha.$$

Медиана является квантилью порядка $1/2$.

Если случайная величина X имеет конечные моменты до четвёртого порядка включительно, то величина

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\mathsf{E}(X - \mathsf{E}X)^3}{\mathsf{D}^{3/2}X}$$

называется *коэффициентом асимметрии*, а

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{\mathsf{E}(X - \mathsf{E}X)^4}{\mathsf{D}^2X} - 3$$

коэффициентом эксцесса её распределения. Эти величины характеризуют степень отличия функции распределени $F_X(x)$ от функции распределения $\Phi(x)$ стандартного нормального распределения, для которого коэффициенты асимметрии и эксцесса равны нулю.

2.2 СПИСОК ЛИТЕРАТУРЫ

1) А.Н. Ширяев, Вероятность,
Москва, Наука, 1989, Глава 2.

2) П. Халмош, Теория Меры,
Москва, Иностранная Литература, 1953, Главы 4 – 5.

3) М.Дж. Кендалл, А. Стьюарт, Теория Распределений,
Москва, Наука, 1966, Главы 1 – 2.

Лекция 3

В Лекции определяются характеристические функции случайных величин, семиинварианты, приводятся основные свойства характеристических функций.

3.1 ХАРАКТЕРИСТИЧЕСКИЕ ФУНКЦИИ СЛУЧАЙНЫХ ВЕЛИЧИН

1) Пусть случайная величина X имеет функцию распределения $F(x)$, тогда *характеристической функцией* называется комплекснозначная функция вида

$$f_X(t) \equiv f(t) = \mathsf{E}e^{itX} = \int\limits_{-\infty}^{+\infty} e^{itx} dF(x) = \tag{3.1.1}$$

$$= \int\limits_{-\infty}^{+\infty} \cos(tx) dF(x) + i \int\limits_{-\infty}^{+\infty} \sin(tx) dF(x).$$

В частности, если у случайной величины X существует плотность $p(x) = F'(x)$, то характеристическая функция является преобразованием Фурье плотности $p(x)$

$$f(t) = \int\limits_{-\infty}^{+\infty} e^{itx} p(x) dx. \tag{3.1.2}$$

Для дискретной случайной величины X, принимающей значения x_k с вероятностями p_k, характеристическая функция $f(t)$ представима

рядом
$$f(t) = \sum_k e^{itx_k} p_k. \tag{3.1.3}$$
Несложно видеть, что если
$$X \sim \mathcal{N}(\mu, \sigma^2),$$
то
$$f(t) = \exp\{it\mu - t^2\sigma^2/2\}.$$
Характеристические функции определены при всех действительных t для любых случайных величин. Приведём основные свойства характеристических функций:

(a) справедливы соотношения
$$f(0) = 1, \quad |f(t)| \le 1, \quad t \in \mathbf{R}^1;$$

(b) характеристическая функция равномерно непрерывна на всей действительной оси;

(c) (положительная определённость характеристических функций) при каждом $n \in \mathbf{N}$ для любых комплексных чисел $z_1, \ldots, z_n$ и любых вещественных чисел $t_1, \ldots, t_n$
$$\sum_{l,m}^{n} f(t_l - t_m) z_l \bar{z}_m \ \ge\ 0;$$

(d) эрмитовость:
$$\bar{f}(-t) = f(t);$$

(e) при $Y = aX + b$, где a и b – действительные числа
$$f_Y(t) = e^{ibt} f_X(at).$$

Для решётчатого распределения
$$p_n = \mathsf{P}(X = b + nh), \quad n = 0, \pm 1, \pm 2, \ldots$$
характеристическая функция $f(t)$ представима в виде ряда Фурье
$$f(t) = e^{itb} \sum_{n=-\infty}^{+\infty} e^{itnh} p_n \tag{3.1.4}$$

так что
$$|f(2\pi/h)| = 1.$$
Обратно, если при некотором $t_0 \neq 0$ справедливо равенство
$$|f(t_0)| = 1,$$
то соответствующее распределение решётчатое.
Максимальный шаг распределения равен h тогда и только тогда, когда модуль характеристической функции меньше единицы при $0 < |t| < 2\pi/h$ и равен единице при $t = 2\pi/h$.

2) Согласно Утверждению 1.1.1, мы можем каждую функцию распределения $F(x)$ представить в виде суммы суммы трёх компонент. Используя этот факт, получаем соответствующее представление для характеристических функций
$$f(x) = a_1 f_1(x) + a_2 f_2(x) + a_3 f_3(x), \tag{3.1.5}$$
где каждый член содержит характеристическую функцию соответствующей компоненты $F(x)$. Рассмотрим теперь в отдельности поведение каждого из этих трёх членов.

(a) Так как $F_1(x)$ абсолютно непрерывна, то
$$f_1(t) = \int_{-\infty}^{+\infty} e^{itx} F_1'(x) dx$$
и, следовательно по Теореме Римана-Лебега
$$f_1(t) \to 0 \quad \text{при} \quad |t| \to \infty. \tag{3.1.6}$$
Если для всех x существует абсолютно интегрируемая n-я производная $F_1^{(n)}(x)$, то интегрирорванием по частям несложно показать, что поведение $f_1(t)$ на бесконечности описывается соотношением
$$f_1(t) = O\Big(\frac{1}{|t|^{n-1}}\Big) \quad \text{при} \quad t \to \infty. \tag{3.1.7}$$

(b) Если через x_k и p_k, $k = 1, 2, \ldots$ обозначить соответственно точки разрыва и величины скачков функции распределения $F(x)$ в этих точках, то
$$a_2 f_2(t) = \sum_{k=1}^{\infty} p_k e^{itx_k}.$$

Это выражение представляет собой сумму абсолютно сходящегося тригонометрического ряда и

$$\limsup_{|t|\to\infty} |f_2(t)| = 1. \tag{3.1.8}$$

(c) Характеристическая функция $f_3(t)$ является характеристической функцией непрерывной функции распределения $F_3(x)$, имеющей производную, почти всюду равную нулю. При этом $f_3(t)$ может не стремиться к нулю при $|t| \to \infty$.

Таким образом справедливо следующее

Утверждение 3.1.1. *Если в представлении функции распределения $F(x)$ в виде суммы трех компонент (см. Утверждение 1.1.1), $a_1 > 0$, то*

$$\limsup_{|t|\to\infty} |f(t)| < 1.$$

Отсюда следует, что для $|t| \geq \varepsilon > 0$

$$|f(t)| < q_\varepsilon < 1,$$

при любом сколь угодно малом $\varepsilon > 0$.
Если $a_1 = 1$, то

$$\lim_{|t|\to\infty} f(t) = 0.$$

Если $a_2 = 1$, то

$$\limsup_{|t|\to\infty} |f(t)| = 1.$$

Для любой характеристической функции $f(t)$ справедливо равенство

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |f(t)|^2 dt = \sum_{k=1}^{\infty} p_k^2,$$

где p_k величины скачков функции распределения $F(x)$ во всех ее точках разрыва x_k, $k = 1, 2, \ldots$

Справедливо также следующее утверждение.

Утверждение 3.1.2. *Если $f(t)$ – характеристическая функция такая, что*

$$|f(t)| \leq q < 1 \quad \text{при} \quad |t| \geq c, \quad c > 0,$$

то при $|t| < c$ *справедливо неравенство*

$$|f(t)| \le 1 - (1-q^2)\frac{t^2}{8c^2} \le \exp\Big\{-(1-q^2)\frac{t^2}{8c^2}\Big\}.$$

При малых t поведение $f(t)$ описывается следующим неравенством.

Утверждение 3.1.3. *Пусть* $f(t)$ *– характеристическая функция некоторого невырожденного рапределения. Тогда существуют такие* $\delta > 0$ *и* $\gamma > 0$, *что*

$$|f(t)| \le 1 - \delta t^2, \quad \textit{при} \quad |t| \le \gamma.$$

3) Случайная величина X и её распределение называются *симмметричными*, если функции распределения случайных величин X и $-X$ совпадают, то есть, если

$$X \stackrel{d}{=} -X.$$

Если X – симметричная случайная величина и $f(t)$ её характеристическая функция, то

$$f(t) = \mathsf{E}e^{itX} = \mathsf{E}e^{-itX} = f(-t) = \bar{f}(t).$$

Таким образом характеристическая функция симметричной случайной величины всегда действительна.

Если у случайной величины X существует момент $\alpha_k = \mathsf{E}X^k$ некоторого целого порядка $k \ge 1$, то характеристическая функция этой случайной величины дифференцируема k раз и, кроме того, справедливо соотношение

$$f_X^{(k)}(0) = i^k\alpha_k = i^k\mathsf{E}X^k. \tag{3.1.9}$$

Используя формулу Тейлора, можно показать, что если случайная величина X с характеристической функцией $f_X(t)$ имеет момент $\alpha_k = \mathsf{E}X^k$ некоторого целого порядка $k \ge 1$, то справедливо разложение

$$f_X(t) = 1 + \sum_{j=1}^{k}\frac{\alpha_j}{j!}(it)^j + o(|t|^k), \quad t \to 0. \tag{3.1.10}$$

Для достаточно малых значений t главная ветвь $\log f_X(t)$, которая стремится к нулю вместе с t, представима в виде

$$\log f_X(t) = \sum_{j=1}^{k}\frac{\kappa_j}{j!}(it)^j + o(|t|^k), \quad t \to 0. \tag{3.1.11}$$

при этом коэффициенты $\{\kappa_j(X) \equiv \kappa_j,\ j = 1, 2, \ldots\}$ называются *кумулянтами* или *семиинвариантами* случайной величины X. Семиинварианты определяются также по формуле

$$\kappa_j = \frac{1}{i^j} l^{(j)}(0), \text{ где } l(t) = \log f_X(t). \tag{3.1.12}$$

Для нормального распределения с произвольными параметрами семиинварианты всех порядков, начиная с третьего, равны нулю. Для распределения Пуассона с параметром λ семиинварианты всех порядков равны λ.

Из формального тождества

$$\log\left(1 + \sum_{j=1}^{\infty} \frac{\alpha_j}{j!}(it)^j\right) = \sum_{j=1}^{\infty} \frac{\kappa_j}{j!}(it)^j$$

можно получить следующую формулу, связывающую семиинвариант κ_s произвольного порядка s с моментами $\alpha_1, \ldots, \alpha_s$

$$\kappa_s = s! \sum (-1)^{m_1 + \ldots + m_s - 1}(m_1 + \ldots + m_s - 1)! \prod_{i=1}^{s} \frac{1}{m_i!}\left(\frac{\alpha_i}{i!}\right)^{m_i}. \tag{3.1.13}$$

Здесь суммирование производится по всем целым неотрицательным решениям уравнения

$$m_1 + 2m_2 + \ldots + s m_s = s.$$

Отсюда несложно получить следующие формулы

$$\kappa_1 = \mathsf{E}X = \alpha_1, \quad \kappa_2 = \mathsf{D}X = \mu_2, \tag{3.1.14}$$

$$\kappa_3 = \mu_3, \quad \kappa_4 = \mu_4 - 3\mu_2^2, \quad \kappa_5 = \mu_5 - 10\mu_2\mu_3,$$

$$\kappa_6 = \mu_6 - 15\mu_2\mu_4 - 10\mu_3^2 + 30\mu_2^3,$$

$$\kappa_7 = \mu_7 - 21\mu_2\mu_5 - 35\mu_3\mu_4 + 210\mu_2^2\mu_3,$$

$$\kappa_8 = \mu_8 - 28\mu_2\mu_6 - 56\mu_3\mu_5 - 35\mu_4^2 + 420\mu_2^2\mu_4 + 560\mu_2\mu_3^2 - 630\mu_2^4.$$

Можно показать, что для семиинвариантов справедливы неравенства

$$|\kappa_n| \ \le \ n^n \beta_n, \quad n = 1, 2, \ldots \tag{3.1.15}$$

Пусть $X = (X_1, \cdots, X_n)$ – случайный вектор со значениями в евклидовом пространстве $\mathbf{R}^n$ и функцией распределения

$$F(x) = \mathsf{P}(X_1 < x_1, \cdots, X_n < x_n), \quad x = (x_1, \cdots, x_n) \in \mathbf{R}^n.$$

Характеристическая функция случайного вектора $X = (X_1, \cdots, X_n)$ определяется равенством

$$f(t) = \mathsf{E}\exp\Big\{\sum_{i=1}^n t_i X_i\Big\} = \int\limits_{\mathbf{R}^n} \exp\Big\{\sum_{i=1}^n t_i x_i\Big\} dF(x), \quad t = (t_1, \cdots, t_n) \in \mathbf{R}^n.$$

Свойства характеристических функций многомерных распределений аналогичны свойствам характеристических функций случайных величин. *Моментами (смешанными моментами)* случайного вектора $X = (X_1, \cdots, X_n)$ называются числа вида

$$\alpha_{k_1, \cdots, k_n} = \mathsf{E}X_1^{k_1} \cdots X_n^{k_n},$$

при этом число

$$k = k_1 + \cdots + k_n$$

называется *порядком момента*. Моменты с натуральными индексами можно определить дифференцированием характеристической функции

$$\alpha_{k_1, \cdots, k_n} = \frac{(-i)^k \partial^k f(0)}{\partial t_1^{k_1} \cdots \partial t_n^{k_n}}, \quad 0 \in \mathbf{R}^n.$$

3.2 СПИСОК ЛИТЕРАТУРЫ

1) А.Н. Ширяев, Вероятность,
Москва, Наука, 1989, Глава 2.

2) М. Лоэв, Теория Вероятностей,
Москва, Иностранная Литература, 1962, Часть 2, Глава 4.

3) М.Дж. Кендалл, А. Стьюарт, Теория Распределений,
Москва, Наука, 1966, Глава 3 – 4.

4) Е. Лукач, Характеристические Функции,
Москва, Наука, 1979, Глава 1 – 4.

Лекция 4

В Лекции определяется понятия независимости и случайного процесса. Формулируются Центральная Предельная Теорема и Закон Больших Чисел.

4.1 НЕЗАВИСИМОСТЬ. ОСНОВНЫЕ ЗАКОНЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

1) Пусть $X_1 = X_1(\omega), \ldots, X_n = X_n(\omega)$ – случайные величины, определённые на одном и том же вероятностном пространстве $(\Omega, \mathcal{A}, \mathsf{P})$, тогда вектор $\mathbf{X}_n = (X_1, \cdots, X_n)$ называется случайным вектором, или n – мерной случайной величиной. Областью значений случайного вектора $\mathbf{X}_n$ является n – мерное евклидово пространство $\mathbf{R}^n$. Для каждого борелевского множества B пространства $\mathbf{R}^n$ определена вероятность

$$\mathsf{P}(\mathbf{X}_n \in B)$$

называемая *распределением случайного вектора* $\mathbf{X}_n$. В частности, для любых действительных чисел $x_1, \ldots, x_n$ определена функция

$$F(x_1, \ldots, x_n) = \mathsf{P}(X_1 < x_1, \ldots, X_n < x_n),$$

которая называется *функцией распределения случайного* вектора $\mathbf{X}_n$.

2) Пусть $(\Omega, \mathcal{A}, \mathsf{P})$ – вероятностное пространство, и пусть $A_i \in \mathcal{A}, i = 1, 2, \ldots, n$. События $A_1, \ldots, A_n$ называются *взаимно независимыми*, если

$$\mathsf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \prod_{l=1}^{k} \mathsf{P}(A_{i_l})$$

для любого целого числа $2 \le k \le n$ и любых целых $1 \le i_1 < \ldots < i_k \le n$.

Пусть $(X_1, \cdots, X_n)$ – случайные величины, определённые на $(\Omega, \mathcal{A}, \mathsf{P})$. Эти случайные величины называются *независимыми*, если взаимно независимы события вида

$$\{\omega : X(\omega) \in B_k\}, \quad k = 1, \ldots, n$$

для любых борелевских множеств $B_1, \ldots, B_n$ на действительной прямой. Случайные величины $(X_1, \cdots, X_n)$ независимы тогда и только тогда, когда

$$F(x_1, \ldots, x_n) = \prod_{k=1}^{n} F_k(x_k)$$

для любых действительных $x_1, \ldots, x_n$. Здесь

$$F(x_1, \ldots, x_n) = \mathsf{P}(X_1 < x_1, \ldots, X_n < x_n) \quad \text{и} \quad F_k(x) = \mathsf{P}(X_k < x).$$

Последовательность случайных величин $X_1, X_2, \ldots$, определённых на одном и том же вероятностном пространстве, называется *последовательностью независимых случайных величин*, если случайные величины $X_1, \ldots, X_n$ независимы при любом n.

Для любой последовательности функций распределения $F_1(x), F_2(x), \ldots$ существует вероятностное пространство $(\Omega, \mathcal{A}, \mathsf{P})$ и определённая на нем последовательность независимых случайных величин $X_1, X_2, \ldots$ такая, что для любого n функция распределения случайной величины X_n есть $F_n(x)$.

Если $X_1, \ldots, X_{n+m}$ – независимые случайные величины, h и g – борелевские функции со значениями в $\mathbf{R}^1$, определённые соответственно на $\mathbf{R}^n$ и $\mathbf{R}^m$, то случайные величины $h(X_1, \ldots, X_n)$ и $g(X_{n+1}, \ldots, X_{n+m})$ независимы. Если случайные величины $h(X_1, \ldots, X_n)$ и $g(X_{n+1}, \ldots, X_{n+m})$ имеют математические ожидания, то

$$\mathsf{E}h(X_1, \ldots, X_n)g(X_{n+1}, \ldots, X_{n+m}) = \mathsf{E}h(X_1, \ldots, X_n)\mathsf{E}g(X_{n+1}, \ldots, X_{n+m}).$$

В частности, если случайные величины X и Y независимы и у них существуют дисперсии, то

$$\mathsf{E}XY = \mathsf{E}X\mathsf{E}Y, \quad \mathsf{D}(X+Y) = \mathsf{D}X + \mathsf{D}Y.$$

Утверждение 4.1.1. *Если X и Y – независимые случайные величины и $F_X(x)$, $F_Y(x)$ – их функции распределения, а $f_X(t)$ и $f_Y(t)$ – характеристические функции, то сумма $X+Y$ имеет функцию распределения (называемую сверткой или композицией функций распределения $F_X(x)$, $F_Y(x)$) вида*

$$F_X * F_Y(x) \equiv F_{X+Y}(x) = \int\limits_{-\infty}^{\infty} F_X(x-y)dF_Y(y) =$$

$$= F_Y * F_X(x) \equiv \int\limits_{-\infty}^{\infty} F_Y(x-y)dF_X(y),$$

и характеристическую функцию

$$f_{X+Y}(t) = f_X(t)f_Y(t).$$

3) *Случайным процессом* называется семейство случайных величин $\xi(t) = \xi(\omega, t)$, заданных на одном вероятностном пространстве $(\Omega, \mathcal{A}, \mathsf{P})$ и зависящих от параметра t принимающего значения из некоторого множества T. Обозначать случайный процесс мы будем символами $\{\xi(t),\ t \in T\}$ или $\xi(t)$. Последовательности независимых случайных величин $X_1, X_2, \ldots$ рассмотренные выше, являются случайными процессами, для которых $T = \{1, 2, \ldots\}$. Такие процессы, у которых множество T можно отождествить со всей или с частью последовательности $\{\ldots, -1, 0, 1, \ldots\}$, обычно называют *процессами с дискретным временем* или *случайными последовательностями.*

   Если множество T совпадает с некоторым числовым интервалом $T = [a, b]$, то семейство случайных величин $\xi(t) = \xi(\omega, t)$ называется *случайным процессом с непрерывным временем.* Интерпретация параметра t как времени, конечно, не обязательна.

   Рассмотрим некоторый случайный процесс $\xi(t) = \xi(\omega, t)$. Если фиксировать $\omega_0 \in \Omega$, то мы получим функцию $\xi(t) = \xi(\omega_0, t),\ t \in T$, которую часто называют *выборочной функцией* или *траекторией случайного процесса.* Таким образом, здесь в роли случайных величин выступают функции. Как и раньше, мы могли бы рассматривать пространство элементарных событий Ω_ξ, предположив, что Ω_ξ есть функциональное пространство элементов $\xi = \xi(t)$ и что σ-алгебра $\mathcal{A}_\xi$ содержит все множества вида

$$\{\xi : \xi(t_0) \in C\}$$

для любых t_0 и борелевских множеств C. В этом случае меру P_ξ в тройке $(\Omega_\xi, \mathcal{A}_\xi, \mathsf{P}_\xi)$ мы будем называть *распределением случайного процесса* $\{\xi(t),\ t \in T\}$. Если же фиксировать значения $t_1, \dots, t_k$, то мы получим многомерную случайную величину $(\xi(\omega, t_1), \dots, \xi(\omega, t_k))$. Распределение таких случайных величин называется *конечномерными распределениями процесса* $\{\xi(t),\ t \in T\}$. Из Теоремы Колмогорова о согласованных распределениях следует, что задание согласованных конечномерных распределений однозначно определяет распределение процесса.

Наиболее простой природой обладают так называемые однородные случайные процессы с независимыми приращениями.

Случайный процесс $\{\xi(t),\ t \in [a, b]\}$, опредлённый на отрезке $[a, b]$ называется *случайным процессом с независимыми приращениями*, если для любых чисел $a \le t_0 < t_1 < \ldots < t_n \le b$ случайные величины

$$\xi(t_0),\ \xi(t_1) - \xi(t_0), \dots, \xi(t_n) - \xi(t_{n-1})$$

независимы.

Случайный процесс $\{\xi(t),\ t \in [a, b]\}$ назывется *однородным*, если распределения случайных величин $\xi(t_1) - \xi(t_0)$ зависит лишь от длины интервала $t_1 - t_0$ и не зависит от t_0.

4) Пусть на измеримом пространстве $(\mathcal{X}, \mathcal{B})$ имеется вероятностная мера $\mathbf{Q}$ $(= \mathsf{P}_X)$ и σ – конечная мера ν. Говорят, что мера $\mathbf{Q}$ имеет плотность относительно ν, если существует $\mathcal{B}$ – измеримая функция $p(x)$ такая, что
$$\mathbf{Q}(B) = \int\limits_B p(x) d\nu(x), \quad \text{для всех} \quad B \in \mathcal{B}.$$
При этом если P_X имеет плотность $p(x)$, то
$$\mathsf{E}T(X) = \int\limits_{\mathcal{X}} T(x) d\mathsf{P}_X(x) = \int\limits_{\mathcal{X}} T(x) p(x) d\nu(x).$$

5) Центральная предельная теорема. *Пусть $X_1, X_2, \cdots$ – последовательность независимых одинаково распределнных невырожденных случайных величин таких, что*
$$\mathsf{E}X_1^2 < +\infty,$$

тогда

$$\operatorname*{Lim}_{n\to\infty}\sup_x\left|\mathsf{P}\left(\frac{S_n-n\mu}{\sqrt{n}\sigma}<x\right)-\Phi(x)\right|=0,$$

где

$$S_n=X_1+\cdots+X_n,\quad \mu=\mathsf{E}X_1,\quad \sigma^2=\mathsf{D}X_1,$$

$$\Phi(x)=\int\limits_{-\infty}^{x}\phi(t)dt,\quad \phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

6) Закон больших чисел. *Пусть $X_1, X_2, \cdots$ – последовательность независимых одинаково распределнных случайных величин таких, что*

$$\mathsf{E}|X_1|<+\infty,$$

тогда

$$\mathsf{P}\left(\operatorname*{Lim}_{n\to\infty}n^{-1}\sum_{i=1}^{n}X_i=\mathsf{E}X_1\right)=1.$$

4.2 СПИСОК ЛИТЕРАТУРЫ

1) Ж. Невё, Математические Основы Теории Вероятностей, Москва, Мир, 1969, Глава 4.

2) А.Н. Ширяев, Вероятность, Москва, Наука, 1989, Главы 3 – 4.

3) М. Лоэв, Теория Вероятностей, Москва, Иностранная Литература, 1962, Часть 3.

Лекция 5

В Лекции дается общее определение условных математических ожиданий и условных вероятностей. Рассмотрены их основные свойства.

5.1 УСЛОВНЫЕ МАТЕМАТИЧЕСКИЕ ОЖИДАНИЯ И УСЛОВНЫЕ ВЕРОЯТНОСТИ

Пусть $(\Omega, \mathcal{A}, \mathsf{P})$ – вероятностное пространство и множества (события) A, B принадлежат σ – алгебре $\mathcal{A}$, причём

$$\mathsf{P}(B) > 0.$$

Определим *условную вероятность* события A при условии события B по формуле

$$\mathsf{P}(A \mid B) \stackrel{def}{=} \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}.$$

Рассмотрим сначала условную вероятность относительно дискретных случайных величин.

Пусть $\xi = \xi(\omega)$ – случайная величина, принимающая значения $\{x_1, x_2, \cdots\}$ с вероятностями

$$p_i = \mathsf{P}(\xi = x_i) > 0, \quad i \in \mathbf{N}, \quad \sum_i p_i = 1.$$

Рассмотрим множества

$$B_i = \{\omega : \xi(\omega) = x_i\}, \quad i \in \mathbf{N}.$$

Заметим, что σ – алгебра $\mathcal{A}_\xi$, порождённая случайной величиной ξ, состоит из множеств вида

$$\Lambda = \sum_{i \in I} B_i, \quad I \subset \mathbf{N}. \tag{5.1.1}$$

При этом существуют условные вероятности вида

$$\mathsf{P}(A \mid B_i) \equiv \mathsf{P}(A \mid \xi = x_i), \quad i \in \mathbf{N}.$$

Определим теперь *условную вероятность события A относительно случайной величины* ξ как функцию следующего вида:

$$\mathsf{P}(A \mid \xi(\omega)) \equiv \mathsf{P}(A \mid \xi = x_i), \quad \text{при} \quad \omega \in B_i, \quad i \in \mathbf{N}.$$

При этом нетрудно видеть, что справедливо следующее основное соотношение, дающее основание для определения условной вероятности в общем случае, вида

$$\mathsf{P}(A \cap \Lambda) = \sum_{i \in I} \mathsf{P}(A \cap B_i) = \sum_{i \in I} \mathsf{P}(A \mid \xi = x_i)\mathsf{P}(B_i),$$

или

$$\mathsf{P}(A \cap \Lambda) = \int\limits_{\Lambda} \mathsf{P}(A \mid \xi(\omega)) d\mathsf{P}(\omega), \quad \text{при всех} \quad \Lambda \in \mathcal{A}_{\xi}. \tag{5.1.2}$$

Определим теперь *условное математическое ожидание* дискретной случайной величины $\eta = \eta(\omega)$ относительно дискретной случайной величины ξ. Пусть η – дискретная случайная величина, принимающая значения $\{y_1, y_2, \cdots\}$ с вероятностями

$$q_k = \mathsf{P}(\eta = y_k), \quad k \in \mathbf{N}, \quad \sum_k q_k = 1.$$

Предположим, что случайная величина η интегрируема, то есть

$$\mathsf{E}|\eta| = \sum_k |y_k| q_k < \infty.$$

Определим теперь *условное математическое ожидание случайной величины η относительно случайной величины* ξ как функцию вида:

$$\mathsf{E}(\eta \mid \xi(\omega)) \equiv \mathsf{E}(\eta \mid \xi = x_i) \stackrel{def}{=} \sum_k y_k \mathsf{P}(\eta = y_k \mid \xi = x_i), \quad \text{при} \quad \omega \in B_i, \quad i \in \mathbf{N}.$$

Из этого определения непосредственно следует, что

$$\sum_k y_k \mathsf{P}(\eta = y_k, \xi = x_i) = \mathsf{E}(\eta \mid \xi = x_i)\mathsf{P}(\xi = x_i), \quad i \in \mathbf{N}.$$

Суммируя последние равенства по всем $i \in I$, получим

$$\sum_k y_k \mathsf{P}(\eta = y_k, \Lambda) = \sum_{i \in I} \mathsf{E}(\eta \mid \xi = x_i)\mathsf{P}(\xi = x_i), \quad I \subset \mathbf{N}.$$

То есть справедливо следующее основное интегральное соотношение, мотивирующее общее определение условного математического ожидания

$$\int\limits_{\Lambda} \eta(\omega) d\mathsf{P}(\omega) = \int\limits_{\Lambda} \mathsf{E}(\eta \mid \xi(\omega)) d\mathsf{P}(\omega), \quad \text{при всех} \quad \Lambda \in \mathcal{A}_\xi. \tag{5.1.3}$$

Рассмотрим теперь общий случай.

Пусть $\mathcal{D}$ – некоторая σ – подалгебра исходной σ – алгебры $\mathcal{A}$ и пусть $\eta = \eta(\omega)$ – интегрируемая случайная величина.

Определение 5.1.1. *Условным математическим ожиданием интегрируемой случайной величины η относительно σ – подалгебры $\mathcal{D}$ называется случайная величина $\mathsf{E}(\eta \mid \mathcal{D})(\omega) \equiv \mathsf{E}(\eta \mid \mathcal{D})$, удовлетворяющая следующим условиям*

1) $\mathsf{E}(\eta \mid \mathcal{D})$ является $\mathcal{D}$ – измеримой случайной величиной.

2) Справедливо следующее интегральное соотношение (ср. (5.1.3))

$$\int\limits_{D} \eta(\omega) d\mathsf{P}(\omega) = \int\limits_{D} \mathsf{E}(\eta \mid \mathcal{D}) d\mathsf{P}(\omega), \quad \textit{при всех} \quad D \in \mathcal{D}. \tag{5.1.4}$$

Докажем, что для любой интегрируемой случайной величины η существует условное математическое ожидание $\mathsf{E}(\eta \mid \mathcal{D})$. С этой целью рассмотрим на $\mathcal{D}$ счётно-аддитивную функцию множеств вида

$$\rho(D) = \int\limits_{D} \eta(\omega) d\mathsf{P}(\omega), \ D \in \mathcal{D}.$$

Она является абсолютно непрерывна относительно P, и поэтому по Теореме Радона – Никодима (см. Теорему из пункта 8(g), Лекции 2) существует $\mathcal{D}$ – измеримая функция $g(\omega)$ такая, что

$$\rho(D) = \int\limits_{D} g(\omega) d\mathsf{P}(\omega), \ D \in \mathcal{D},$$

то есть в качестве $\mathsf{E}(\eta \mid \mathcal{D})$ можно взять $g(\omega)$. Таким образом в соответствии с Теоремой Радона – Никодима условное математическое ожидание $\mathsf{E}(\eta \mid \mathcal{D})$

определяется однозначно с точностью до множеств P – меры нуль. Иначе говоря в качестве $\mathsf{E}(\eta \mid \mathcal{D})$ можно взять любую $\mathcal{D}$ – измеримую функцию $h(\omega)$, называемую *вариантом* условного математического ожидания, для которой

$$\rho(D) = \int\limits_D h(\omega) d\mathsf{P}(\omega), \ D \in \mathcal{D}.$$

Отметим также, что из Теоремы Радона – Никодима следует равенство

$$\mathsf{E}(\eta \mid \mathcal{D}) = \frac{d\rho}{d\mathsf{P}}(\omega),$$

то есть условное математическое ожидание есть не что иное, как производная Радона – Никодима меры ρ относительно меры P (рассматриваемых на $(\Omega, \mathcal{D})$).

В связи с соотношением (5.1.4) заметим, что мы не можем, вообще говоря, положить

$$\mathsf{E}(\eta \mid \mathcal{D}) = \eta,$$

поскольку случайная величина η не обязана быть $\mathcal{D}$ – измеримой.

Определение 5.1.2. *Условной вероятностью события $A \in \mathcal{A}$ относительно σ – подалгебры $\mathcal{D}$ называется случайная величина вида*

$$\mathsf{P}(A \mid \mathcal{D})(\omega) \equiv \mathsf{P}(A \mid \mathcal{D}) = \mathsf{E}(\mathbf{1}_A \mid \mathcal{D}).$$

Таким образом это функция удовлетворяющая следующим условиям

1) $\mathsf{P}(A \mid \mathcal{D})$ является $\mathcal{D}$ – измеримой случайной величиной.

2) Справедливо следующее интегральное соотношение (ср. (5.1.2))

$$\mathsf{P}(A \cap D) = \int\limits_D \mathsf{P}(A \mid \mathcal{D}) d\mathsf{P}(\omega), \quad \text{при всех} \quad D \in \mathcal{D}. \tag{5.1.5}$$

Пусть $\xi = \xi(\omega)$ – случайная величина со значениями в $\mathbf{R}^n$; Обозначим σ – алгебру, порождённую ξ через $\mathcal{A}_\xi$.

Определение 5.1.3. *Условным математическим ожиданием интегрируемой случайной величины η относительно случайной величины ξ называется случайная величина вида*

$$\mathsf{E}(\eta \mid \xi(\omega)) \equiv \mathsf{E}(\eta \mid \xi) \stackrel{def}{=} \mathsf{E}(\eta \mid \mathcal{A}_\xi).$$

Условной вероятностью события $A \in \mathcal{A}$ относительно случайной величины ξ называется случайная величина вида

$$\mathsf{P}(\eta \mid \xi(\omega)) \equiv \mathsf{P}(A \mid \xi) \stackrel{def}{=} \mathsf{P}(A \mid \mathcal{A}_\xi).$$

Рассмотрим теперь подробнее условные математические ожидания $\mathsf{E}(\eta \mid \xi)$ относительно случайной величины ξ. Поскольку по определению $\mathsf{E}(\eta \mid \xi)$ является $\mathcal{A}_\xi$ – измеримой функцией, то согласно Утверждению 1.1.2, найдётся такая измеримая функция $m(x)$, что

$$m(\xi(\omega)) = \mathsf{E}(\eta \mid \xi(\omega)).$$

Эту функцию $m(x)$ будем обозначать через

$$\mathsf{E}(\eta \mid \xi = x), \quad \text{то есть} \quad m(x) \stackrel{def}{=} \mathsf{E}(\eta \mid \xi = x)$$

и называть *условным матемтатическим ожиданием η при условии, что $\xi = x$.*

Из Определений 5.1.1 и 5.1.3 следует, что справедливы равенства

$$\int\limits_B \eta d\mathsf{P} = \int\limits_B \mathsf{E}(\eta \mid \xi) d\mathsf{P} = \int\limits_B m(\xi) d\mathsf{P}, \quad \text{при всех} \quad B \in \mathcal{A}_\xi.$$

Поэтому используя формулу замены переменного в интеграле Лебега (см. Лекцию 2, пункт 8(f)), последнюю формулу можно переписать в виде

$$\int\limits_{\{\omega:\xi\in C\}} \eta d\mathsf{P} = \int\limits_C m(x) d\mathsf{P}_\xi(x), \quad \text{при всех} \quad C \in \mathcal{B}^1, \tag{5.1.6}$$

где P_ξ – распределение вероятностей случайной величины ξ.

Условные математические ожидания почти наверное обладают свойствами обычного математического ожидания. Ниже предполагается существование и почти наверное конечность всех условных математических ожиданий. Приведём теперь без доказательств (которые, впрочем, непосредственно следуют из определений) основные свойства условных математических ожиданий. Приведённые ниже свойства выполняются *почти наверное.* Доказательства могут быть найдены, например, в [1] (стр. 231 –234).

1) Если $\eta \equiv C$, то

$$\mathsf{E}(\eta \mid \mathcal{D}) = C.$$

2) Для любых чисел $a, b \in \mathbf{R}^1$
$$\mathsf{E}(a\eta + b\zeta \mid \mathcal{D}) = a\mathsf{E}(\eta \mid \mathcal{D}) + b\mathsf{E}(\zeta \mid \mathcal{D}).$$

3) Если $\xi \leq \zeta$, то и
$$\mathsf{E}(\eta \mid \mathcal{D}) \leq \mathsf{E}(\zeta \mid \mathcal{D}).$$

4)
$$|\mathsf{E}(\eta \mid \mathcal{D})| \leq \mathsf{E}(|\eta| \mid \mathcal{D}).$$

5) Если
$$\mathcal{D} = \{\emptyset, \Omega\}, \quad \text{то} \quad \mathsf{E}(\eta \mid \mathcal{D}) = \mathsf{E}\eta.$$

Если
$$\mathcal{D} = \Upsilon(\Omega) = \{D : D \subseteq \Omega\}, \quad \text{то} \quad \mathsf{E}(\eta \mid \mathcal{D}) = \eta.$$

6)
$$\mathsf{E}(\mathsf{E}(\eta \mid \mathcal{D})) = \mathsf{E}\eta.$$

7) Если $\mathcal{D}_1 \subseteq \mathcal{D}_2$, то
$$\mathsf{E}[\mathsf{E}(\eta \mid \mathcal{D}_2) \mid \mathcal{D}_1] = \mathsf{E}(\eta \mid \mathcal{D}_1).$$

8) Если $\mathcal{D}_1 \supseteq \mathcal{D}_2$, то
$$\mathsf{E}[\mathsf{E}(\eta \mid \mathcal{D}_2) \mid \mathcal{D}_1] = \mathsf{E}(\eta \mid \mathcal{D}_2).$$

9) Пусть случайная величина η не зависит от σ – алгебры $\mathcal{D}$, то есть не зависит от $\mathbf{1}_D(\omega),\ D \in \mathcal{D}$. Тогда
$$\mathsf{E}(\eta \mid \mathcal{D}) = \mathsf{E}\eta.$$

10) Пусть ζ – $\mathcal{D}$ – измеримая случайная величина, тогда
$$\mathsf{E}(\zeta\eta \mid \mathcal{D}) = \zeta\mathsf{E}(\eta \mid \mathcal{D}).$$

11) Пусть $\xi = \xi(\omega)$ – случайная величина, принимающая значения $\{x_1, x_2, \cdots\}$ с вероятностями
$$p_i = \mathsf{P}(\xi = x_i) > 0, \quad i \in \mathbf{N}, \quad \sum_i p_i = 1.$$

Тогда
$$\mathsf{E}(\eta \mid \xi = x_i) = \frac{1}{\mathsf{P}(\xi = x_i)} \int\limits_{\{\omega : \xi = x_i\}} \eta d\mathsf{P}, \quad i \in \mathbf{N}.$$

12) Пусть $(\eta,\ \xi)$ – пара случайных величин, распределение которых обладает плотностью $p_{\eta\xi}(x,y)$:

$$\mathsf{P}((\eta,\ \xi)\in C)=\int\limits_C p_{\eta\xi}(x,y)dxdy,\qquad C\in\mathcal{B}^2.$$

Пусть $p_\eta(x),\ p_\xi(y)$ – соответственно плотности случайных величин η и ξ. Обозначим

$$p_{\eta|\xi}(x\mid y)=\frac{p_{\eta\xi}(x,y)}{p_\xi(y)},$$

полагая $p_{\eta|\xi}(x\mid y)=0$, если $p_\xi(y)=0$.

Тогда

$$\mathsf{P}(\eta\in B\mid \xi=y)=\int\limits_B p_{\eta|\xi}(x\mid y)dx,\qquad B\in\mathcal{B}^1$$

и

$$\mathsf{E}(\eta\mid\xi=y)=\int\limits_{-\infty}^{+\infty}xp_{\eta|\xi}(x\mid y)dx.$$

5.2 СПИСОК ЛИТЕРАТУРЫ

1) Ж. Невё, Математические Основы Теории Вероятностей,
Москва, Мир, 1969, Глава 4 ¡3.

2) А.Н. Ширяев, Вероятность,
Москва, Наука, 1989, Глава 1 ¡3 и ¡8, Глава 2 ¡7.

3) М. Лоэв, Теория Вероятностей,
Москва, Иностранная Литература, 1962, Часть 4 Глава 7.

Лекция 6

Понятие статистической структуры, вводимое в Лекции, играет в математической статистике такую же роль, что и вероятностное пространство в теории вероятностей. В частности, относительно выбора исходной статистической структуры справедливы те же замечания, что и относительно выбора того или иного вероятностного пространства в теории вероятностей.

В Лекции даются основные определения математичекой статистики, используемые в дальнейшем.

6.1 СТАТИСТИЧЕСКИЕ СТРУКТУРЫ

Определение 6.1.1. *Пусть $\mathcal{P}$ – семейство вероятностных мер (распределений) на измеримом пространстве $(\mathcal{X},\, \mathcal{F})$. Статистической структурой называется тройка $(\mathcal{X},\, \mathcal{F},\, \mathcal{P})$.*

Пространство $\mathcal{X}$ имеет смысл пространства наблюдений (в дальнейшем, как правило, мы будем считать $\mathcal{X} \subseteq \mathbf{R}^m, \;\; m \geq 1$), то есть имеется случайный элемент $X = X(\omega)$ со значениями в $\mathcal{X}$, заданный на вероятностном пространстве $(\Omega,\, \mathcal{A})$ и являющийся $\mathcal{A} - \mathcal{F}$ – измеримым. При этом предполагается, что распределение *неизвестно*, но принадлежит семейству $\mathcal{P}$. Основная задача математической статистики состоит в том, чтобы по наблюдению $=$ сделать выводы или "оценить" распределение P_X случайного элемента . Часто считают семейство $\mathcal{P}$ *параметризованным*, то есть имеющим вид

$$\mathcal{P} = \{\mathsf{P}_\theta,\, \theta \in \Theta\},\; \Theta \subseteq \mathbf{R}^k,$$

и предполагают, что отображение

$$\theta \longrightarrow \mathsf{P}_\theta$$

инъективно, то есть обладает свойством

$$\theta_1 \neq \theta_2 \Longrightarrow \exists\ A \in \mathcal{F}\ :\ \mathsf{P}_{\theta_1} \neq \mathsf{P}_{\theta_2}.$$

Такие статистические структуры называются *отделимыми* или *идентифицируемыми.* В этом случае задача статистики сводится к "оцениванию" параметра θ.

Примеры.

1) При статистическом эксперименте, состоящем в проведении n независимых наблюдений над случайной величиной, принимающей конечное число m значений, вероятности которых полностью не известны, но остаются постоянными в течение эксперимента, приходим к статистической структуре вида

$$\mathcal{X} = \{1, 2, \cdots, m\}^n, \qquad \mathcal{F} = \Upsilon(\mathcal{X}),$$

$$\Theta = \{\theta : \theta = (\theta_1, \cdots, \theta_m) \in \mathbf{R}^m,\ \sum_{i=1}^{m} \theta_i = 1,\ \theta_i \geq 0,\quad 1 \leq i \leq m\}, \quad \mathsf{P}_\theta = (\Pi_\theta)^n,$$

где Π_θ – распределение вероятностей на множестве $\{1, 2, \cdots, m\}$, задаваемое вероятностями $(\theta_1, \cdots, \theta_m)$.

2) Пусть наблюдается один раз биномиальная случайная величина (см. пункт 6(b) Лекции 1)

$$X \sim \mathcal{B}(n, \theta)$$

с неизвестной вероятностью успеха θ. В этом случай имеем статистическую структуру вида

$$\mathcal{X} = \{0, 1, 2, \cdots, n\}, \qquad \mathcal{F} = \Upsilon(\mathcal{X}), \qquad \Theta = (0, 1),$$

$$p_\theta(k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}, \quad \mathsf{P}_\theta(A) = \sum_{k \in A} p_\theta(k) = \int\limits_A p_\theta(x) d\nu(x), \quad A \in \mathcal{F},$$

где $\nu(\cdot)$ – *считающая мера*, то есть мера приписывающая каждому множеству из $\mathcal{F}$ число элементов в нём.

3) Пусть наблюдается нормальная случайная величина (см. пункт 6(d) Лекции 1)

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

с неизвестными параметрами (μ, σ^2). В этом случай статистическая структура имеет вид $\mathcal{X} = \mathbf{R}^1$, $\mathcal{F} = \mathcal{B}^1$, $\theta = (\mu, \sigma^2)$, $\Theta = \mathbf{R}^1 \times [0, \infty)$,

$$\mathsf{P}_\theta(A) = \int\limits_A \varphi_{\mu,\sigma}(x)dx, \quad A \in \mathcal{B}^1,$$

где

$$\varphi_{\mu,\sigma}(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right), \quad \varphi(x) = \frac{1}{\sqrt{2\pi}}\exp\{-x^2/2\}.$$

Определение 6.1.2. *Статистическая структура* $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ *(или семейство* $\mathcal{P}$*) называется доминируемой (доминируемым), если существует положительная* σ *– конечная мера* ν *на* $(\mathcal{X}, \mathcal{F})$ *такая, что выполнено одно из двух эквивалентных условий:*

1) каждое распределение из $\mathcal{P}$ *абсолютно непрерывно относительно меры* ν

$$\mathsf{P} \ll \nu, \qquad \text{для любого} \quad \mathsf{P} \in \mathcal{P}.$$

2) каждое распределение из $\mathcal{P}$ *имеет плотность относительно меры* ν

$$p_{\mathsf{P}}(x) \equiv p(x) = \frac{d\mathsf{P}}{d\nu}(x).$$

Эквивалентность этих двух условий следует из Теоремы Радона – Никодима. В дальнейшем будем рассматривать только доминируемые структуры. Если семейство параметризовано и доминируемо, то плотность относительно меры ν будем обозначать как

$$p_\theta(x) = \frac{d\mathsf{P}_\theta}{d\nu}(x),$$

то есть

$$\mathsf{P}_\theta(A) = \int\limits_A p_\theta(x)d\nu(x), \quad A \in \mathcal{F}.$$

Такую статистическую структуру можно записать в виде

$$(\mathcal{X}, \mathcal{F}, \{p_\theta(x), \theta \in \Theta\}).$$

Доминирующая мера не единственна, поскольку если мера ν доминирует структуру и является абсолютно непрерывной относительно меры μ, то μ также доминирует эту структуру и новые плотности имеют вид

$$\bar{p}(x) = \frac{d\mathsf{P}}{d\mu}(x) = p(x)\frac{d\nu}{d\mu}(x).$$

Определение 6.1.3.*Вещественная функция*

$$\mathcal{L}(\theta;\, x) = p_\theta(x),$$

определлнная на $(\Theta,\, \mathcal{X})$ *и рассматриваемая как функция* θ *при фиксированном называется функцией правдоподобия.*

Иногда будем рассматривать логарифм функции правдоподобия

$$l(\theta;\, x) = \log \mathcal{L}(\theta;\, x).$$

Полезный критерий доминируемости статистической структуры даётся следующей Теоремой.

Теорема 6.1.1.*Статистическая структура* $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ *является доминируемой тогда и только тогда, когда существует не более чем счетное подсемейство* $\mathcal{P}' \subseteq \mathcal{P}$ *семейства* $\mathcal{P}$ *эквивалентное* $\mathcal{P}$*, то есть такое, что*

$$\forall\ A \in \mathcal{F}\ :\ \mathsf{P}(A) = 0,\quad \forall\ \mathsf{P} \in \mathcal{P}' \Longleftrightarrow \mathsf{P}(A) = 0,\quad \forall\ \mathsf{P} \in \mathcal{P}.$$

При этом в качестве доминирующей меры ν *можно взять вероятностную меру*

$$\mathsf{P}^*(A) = \sum_{\mathsf{p} \in \mathcal{P}'} c_{\mathsf{p}} \mathsf{P}(A), \quad A \in \mathcal{F},$$

где числа c_{p} *такие, что*

$$c_{\mathsf{p}} > 0, \qquad \sum_{\mathsf{p} \in \mathcal{P}'} c_{\mathsf{p}} = 1.$$

Доказательство. Достаточность условия очевидна, если в качестве доминирующей меры ν выбрать указанную вероятность P^*.

Докажем необходимость. Предположим, что

$$\mathsf{P} \ll \nu, \quad \forall\ \mathsf{P} \in \mathcal{P}. \tag{6.1.1}$$

Меру ν можно считать вероятностью, так как если

$$\mathcal{X} = \sum_{i=1}^{\infty} B_i, \quad \text{и} \quad \nu(B_i) < \infty, \quad i \in \mathbf{N},$$

то эквивалентная ν мера

$$\nu'(A) = \sum_{i=1}^{\infty} \frac{\nu(A \cap B_i)}{2^i \nu(B_i)}, \quad A \in \mathcal{F}$$

является вероятностной. Пусть

$$p(x) = \frac{d\mathsf{P}}{d\nu}(x), \qquad A_{\mathsf{p}} = \{x : p(x) > 0\}, \quad \mathsf{P} \in \mathcal{P}$$

и $\mathcal{H}$ – класс *счетных объединений* множеств A_{p}. Докажем, что

$$\sup_{B \in \mathcal{H}} \nu(B) = C \leq 1$$

достижим. Пусть множества $A_i = A_{\mathsf{p}_i} \in \mathcal{H}$ такие, что

$$\nu(A_i) \ \geq \ C - 1/i.$$

Заметим, что

$$A = \bigcup_{i=1}^{\infty} A_i \in \mathcal{H}$$

и для любого i справедливы неравенства

$$C - 1/i \ \leq \ \nu(A_i) \leq \nu(A) \ \leq \ C,$$

поэтому

$$\nu(A) = \sup_{B \in \mathcal{H}} \nu(B).$$

Докажем, что в качестве не более чем счётного подсемейства $\mathcal{P}'$ можно взять семейство вида $\mathcal{P}' = \{\mathsf{P}_i\}$. При каждом P справедливы соотношения

$$A_{\mathsf{p}} \cup A = (A_{\mathsf{p}} - A) + A \in \mathcal{H},$$

поэтому

$$\nu(A_{\mathsf{p}} - A) + \nu(A) = \nu(A_{\mathsf{p}} \cup A) \ \leq \ C = \nu(A)$$

и значит

$$\nu(A_{\mathsf{p}} - A) = 0.$$

Теперь учитывая соотношение (6.1.1) отсюда следует, что при всех $\mathsf{P} \in \mathcal{P}$

$$\mathsf{P}(A_{\mathsf{p}} - A) = 0, \quad \mathsf{P} \in \mathcal{P}.$$

Но тогда для любого множества $F \in \mathcal{F}$ справедливы равенства

$$\mathsf{P}(F - A) = 0, \quad \mathsf{P} \in \mathcal{P}, \tag{6.1.2}$$

поскольку

$$\mathsf{P}(F \cap A_{\mathsf{p}} - A) \le \mathsf{P}(A_{\mathsf{p}} - A) = 0, \quad \mathsf{P}(F \cap A_{\mathsf{p}}^c - A) \le \mathsf{P}(A_{\mathsf{p}}^c) = 0$$

и

$$\mathsf{P}(F - A) = \mathsf{P}(F \cap A_{\mathsf{p}} - A) + \mathsf{P}(F \cap A_{\mathsf{p}}^c - A) = 0.$$

Пусть множество $D \in \mathcal{F}$ таково, что для всех i

$$\mathsf{P}_i(D) = 0.$$

Поскольку $p_i(x) > 0$ на A_i, и

$$\int\limits_{D \cap A_i} p_i(x) d\nu(x) \ \le \ \int\limits_{D} p_i(x) d\nu(x) = \mathsf{P}_i(D) = 0,$$

то при всех i

$$\nu(D \cap A_i) = 0$$

и значит

$$\nu(D \cap A) = \nu\left(D \cap \bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \nu(D \cap A_i) = 0,$$

поэтому из условия (6.1.1) следует, что для любого $\mathsf{P} \in \mathcal{P}$

$$\mathsf{P}(D \cap A) = 0, \quad \mathsf{P} \in \mathcal{P}.$$

Теперь соотношение (6.1.2) показывает, что вероятность события D

$$\mathsf{P}(D) = \mathsf{P}(D - A) + \mathsf{P}(D \cap A) = \mathsf{P}(D - A)$$

равна нулю. □

Из этой Теоремы непосредственно вытекает следующее

Следствие 6.1.1.*Если семейство $\mathcal{P}$ или пространство $\mathcal{X}$ не более чем счетны, то статистическая структура $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$ доминируема.*

В случае когда $\mathcal{X}$ счётно доминирующую σ – конечную меру, приписывающую каждому множеству из $\Upsilon(\mathcal{X})$ число элементов в нём, часто называют *считающей мерой.* В реальных задачах эта считающая мера в дискретном случае и мера Лебега в непрерывном случае используются наиболее часто.

Из Теоремы 6.1.1 также вытекает следующий основной результат, показывающий, что для доминируемой структуры всегда возможно в качестве доминирующей меры выбрать вероятность.

Теорема 6.1.2.*Статистическая структура $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$ доминируема тогда и только тогда, когда найдется вероятностное распределение P^* на $(\mathcal{X},\ \mathcal{F})$, называемое привилегированным, доминирующее $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$ и обладающее следующими свойствами:*

1) *распределение* P^* *абсолютно непрерывно относительно любой меры, доминирующей* $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$;

2) *распределение* P^* *является строго выпуклой линейной комбинацией вероятностей из некоторого не более чем счетного подсемейства* $\mathcal{P}' \subseteq \mathcal{P}$, *то есть*

$$\mathsf{P}^*(A) = \sum_{\mathsf{p}\in\mathcal{P}'} c_{\mathsf{p}}\mathsf{P}(A), \quad A \in \mathcal{F},$$

где числа c_{p} *такие, что*

$$c_{\mathsf{p}} > 0, \qquad \sum_{\mathsf{p}\in\mathcal{P}'} c_{\mathsf{p}} = 1;$$

3) *распределение* P^* *эквивалентно* $\mathcal{P}$, *то есть*

$$\forall\ A \in \mathcal{F}\ : \mathsf{P}(A) = 0,\ \forall\ P \in \mathcal{P} \Longleftrightarrow\ \mathsf{P}^*(A) = 0.$$

Определение 6.1.4. *Множество* A *из* $\mathcal{F}$ *называется* $\mathcal{P}$ *– пренебрежимым (*$\mathcal{P}$ *– нулевым), если*

$$\mathsf{P}(A) = 0, \quad \forall\ \mathsf{P} \in \mathcal{P}.$$

В соответствии с этим Определением часто говорят о выполнении $\mathcal{P}$ – почти всюду ($\mathcal{P}$ – п.в.) некоторого свойства.

Заметим, что если статистическая структура $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$ является доминируемой и если P^* – доминирующее привилегированное распределение, то событие является $\mathcal{P}$ – пренебрежимым тогда и только тогда, когда оно P^* – пренебрежимо. В самом деле, если событие A является P^* – пренебрежимым событием, то оно и P – пренебрежимо для всех P из $\mathcal{P}$, поскольку P^* доминирует $\mathcal{P}$. Обратно, если событие A является $\mathcal{P}$ – пренебрежимым, то в силу того обстоятельства, что P^* есть выпуклая линейная комбинация элементов из $\mathcal{P}$ событие A является и P^*– пренебрежимым.

Определение 6.1.5. *Статистикой на статистической структуре* $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$ *называется измеримое отображение* T *измеримого пространства* $(\mathcal{X},\ \mathcal{F})$ *в измеримое пространство* $(\mathcal{Y},\ \mathcal{H})$, *не зависящее от* $\mathsf{P} \in \mathcal{P}$.

Подчеркнём ещё раз то важное обстоятельство, что статистика не зависит от семейства $\mathcal{P}$ или от параметра θ в случае его наличия. Если $\mathcal{X} = \mathbf{R}^1$, то статистика T называется вещественной, в случае $\mathcal{X} = \mathbf{R}^k$ говорят о векторной статистике. В математической статистике понятие статистики отвечает понятию случайной величины из теории вероятностей. На практике использование статистик связано с извлечением нужной информации из исходных или необработанных данных, подлежащих анализу.

Любая статистика T порождает статистическую структуру вида $(\mathcal{Y}, \mathcal{H}, \mathcal{P}_T)$, где

$$\mathcal{P}_T = \{\mathsf{P}_T,\ \mathsf{P} \in \mathcal{P}\}, \quad \mathsf{P}_T(B) = \mathsf{P}(T^{-1}(B)), \quad B \in \mathcal{H}. \qquad (6.1.3)$$

Эта статистическая структура называется статистической структурой, *индуцируемой* статистикой T.

Определение 6.1.6.

1) *Пусть T_1 и T_2 – две статистики на $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$ со значениями соответственно в $(\mathcal{Y}_1, \mathcal{H}_1)$ и $(\mathcal{Y}_2, \mathcal{H}_2)$. Говорят, что T_1 и T_2 эквивалентны ($T_1 \sim T_2$), если*
$$T_1^{-1}(\mathcal{H}_1) = T_2^{-1}(\mathcal{H}_2).$$
(Отметим, что это понятие никак не связано с семейством $\mathcal{P}$, в частности, если T_1 и T_2 связаны взаимно однозначным и двусторонне измеримым преобразованием, то они эквивалентны.)

2) *Пусть T_1 и T_2 – две статистики, заданнаые на $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$ со значениями в $(\mathcal{Y},\ \mathcal{H})$. Говорят, что статистика T_1 $\mathcal{P}$ – эквивалентна статистике T_2 ($T_1 \overset{\mathcal{P}}{\sim} T_2$), если событие $\{T_1 \neq T_2\}$ является $\mathcal{P}$ – пренебрежимым. (Заметим, что $\mathcal{P}$ – эквивалентные статистики имеют одинаковые распределения для всех $\mathsf{P} \in \mathcal{P}$).*

3) *Говорят, что статистики T_1 и T_2, заданные на $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$ независимы, если независимы случайные величины T_1 и T_2 для всех $\mathsf{P} \in \mathcal{P}$.*

4) *Вещественная статистика T, заданная на $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$, называется интегрируемой, если для каждого распределения P из $\mathcal{P}$ случайная величина T интегрируема, то есть существует математическое ожидание $\mathsf{E}_\mathsf{P} T$.*

5) *Вещественная интегрируемая статистика T, заданная на $(\mathcal{X}, \mathcal{F}, \mathcal{P})$, называется подобной в среднем (центрированной), если $\mathsf{E}_\mathsf{P} T$ не зависит от P из $\mathcal{P}$ ($\mathsf{E}_\mathsf{P} T = 0$ для всех P из $\mathcal{P}$).*

6) *Образом интегрируемой статистики T, определенной на статистической структуре $(\mathcal{X},\ \mathcal{F},\ \{\mathsf{P}_\theta, \theta \in \Theta\})$, называется функция $\gamma(\theta) \equiv$*

$\gamma_T(\theta)$, *определенная на* Θ *по формуле*

$$\gamma(\theta) = \mathsf{E}_\theta T = \int\limits_\Omega T d\mathsf{P}_\theta, \quad \theta \in \Theta.$$

Определение 6.1.7.

1) *Пусть* $(\mathcal{X}_1,\, \mathcal{F}_1,\, \mathcal{P}_1)$ *и* $(\mathcal{X}_2,\, \mathcal{F}_2,\, \mathcal{P}_2)$ *– две статистические структуры. Их прямым произведением*

$$(\mathcal{X}_1,\, \mathcal{F}_1,\, \mathcal{P}_1) \otimes (\mathcal{X}_2,\, \mathcal{F}_2,\, \mathcal{P}_2)$$

называется статистическая структура вида

$$(\mathcal{X}_1 \times \mathcal{X}_2,\, \mathcal{F}_1 \times \mathcal{F}_2,\, \mathcal{P}_1 \times \mathcal{P}_2),$$

где

$$\mathcal{P}_1 \times \mathcal{P}_2 = \{\mathsf{P}_1 \times \mathsf{P}_2,\ \mathsf{P}_1 \in \mathcal{P}_1,\ \mathsf{P}_2 \in \mathcal{P}_2\}.$$

2) *Пусть* $(\mathcal{X}_1,\, \mathcal{F}_1,\, \{\mathsf{P}_{\theta 1}, \theta \in \Theta\})$ *и* $(\mathcal{X}_2,\, \mathcal{F}_2,\, \{\mathsf{P}_{\theta 2}, \theta \in \Theta\})$ *– две статистические структуры с одинаковым параметрическим пространством* Θ. *Их полупрямым произведением*

$$(\mathcal{X}_1,\, \mathcal{F}_1,\, \{\mathsf{P}_{\theta 1}, \theta \in \Theta\}) \times (\mathcal{X}_2,\, \mathcal{F}_2,\, \{\mathsf{P}_{\theta 2}, \theta \in \Theta\})$$

называется статистическая структура вида

$$(\mathcal{X}_1 \times \mathcal{X}_2,\, \mathcal{F}_1 \times \mathcal{F}_2,\, \{\mathsf{P}_{\theta 1} \times \mathsf{P}_{\theta 2}, \theta \in \Theta\}).$$

В частности, полупрямое произведение конечного числа одной и той же статистической структуры называется структурой повторной выборки

$$(\mathcal{X},\, \mathcal{F},\, \mathcal{P})^n = (\mathcal{X}^n,\, \mathcal{F}^n,\, \{\mathsf{P}^n,\ \mathsf{P} \in \mathcal{P}\}).$$

3) *Пусть* $(\mathcal{X}, \mathcal{F}, \mathcal{P})^n$ *– статистическая структура повторной выборки. Для всякой точки* $(x_1, \cdots, x_n)$ *из* $\mathcal{X}^n$ *выборочным (или эмпирическим) распределением называется распределение на* $(\mathcal{X},\, \mathcal{F})$, *определяемое по формуле*

$$\mathsf{P}_n(x_1, \cdots, x_n;\, A) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_A(x_i), \quad A \in \mathcal{F}.$$

Произведению статистических структур на практике соответствует система независимых наблюдений. В Определении 6.1.7 (2) предполагается, что значение параметра одинаково. В Определении 6.1.7 (1) этого предположения не делается. Понятие повторной выборки весьма важно, оно отвечает конечному числу независимых наблюдений над одной случайной величиной, проводимых в одинаковых условиях. Если структуры доминируемы, то нетрудно указать вид функции правдоподобия. В очевидных обозначениях:

$$\mathcal{L}(\theta_1, \theta_2;\ x_1, x_2) = \mathcal{L}(\theta_1;\ x_1)\mathcal{L}(\theta_2;\ x_2) \quad \text{в Определении 6.1.7 (1)}$$

$$\mathcal{L}(\theta;\ x_1, x_2) = \mathcal{L}(\theta;\ x_1)\mathcal{L}(\theta;\ x_2) \quad \text{в Определении 6.1.7 (2).}$$

В случае повторной выборки удобно положить

$$l(\theta;\ x) = \log \mathcal{L}(\theta;\ x),$$

так что

$$l(\theta;\ x_1, \cdots, x_n) = \sum_{i=1}^{n} l(\theta;\ x_i).$$

Итак, если $X = (X_1, \cdots, X_n)$, где X_i – независимые одинаково распределённые случайные величины со значениями в $\mathcal{X}$ и распределением $\mathsf{P} \in \mathcal{P}$, то в этом случае имеем статистическую структуру

$$(\mathcal{X},\ \mathcal{F},\ \mathcal{P})^n.$$

Закон Больших Чисел (см. Лекция 4, пункт 6) показывает, что при больших n $(n \to \infty)$, если $\mathsf{P} \in \mathcal{P}$ есть общее распределение случайных величин $(X_1, \cdots, X_n)$, то

$$\mathsf{P}_n(X_1, \cdots, X_n;\ A), \quad A \in \mathcal{F}$$

"близко" к $\mathsf{P}(A)$, точнее

$$\mathsf{P}\Big(\operatorname*{Lim}_{n\to\infty} \mathsf{P}_n(X_1, \cdots, X_n;\ A) = \mathsf{P}(A)\Big) = 1, \quad \mathsf{P} \in \mathcal{P}, \quad A \in \mathcal{F}.$$

Этот факт весьма часто используется в статистике. Так, для изучения характеристик статистической структуры $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$, таких как моменты, функции распределения и т.д., рассматривают статистику на $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})^n$, совпадающую с такой же характеристикой для $(\mathcal{X},\ \mathcal{F},\ \mathsf{P}_n(X_1, \cdots, X_n; \cdot))$.

Удобно называть статистику, получаемую таким образом, тем же термином, что и рассматриваемая характеристика, с добавлением прилагательного "эмпирический" или "выборочный". Таким образом справедлив следующий "выборочный" принцип в статистике: при оценивании некоторого достаточно гладкого функционала $\Psi(\mathsf{P})$ от неизвестного вероятностного распределения $\mathsf{P} \in \mathcal{P}$ разумно в качестве оценки взять $\Psi(\mathsf{P}_n)$.

Примеры.

1) Эмпирические (выборочные) моменты.

   Пусть
$$\Psi(\mathsf{P}) = \alpha_k = \mathsf{E}_\mathsf{P} X_1^k = \int x^k d\mathsf{P}(x) < \infty, \quad k \in \mathbf{N}.$$

   Тогда эмпирический момент имеет вид
$$\overline{\alpha}_{kn} = \Psi(\mathsf{P}_n) = \int x^k d\mathsf{P}_n(X_1, \cdots, X_n; x) = \frac{1}{n} \sum_{i=1}^{n} X_i^k. \qquad (6.1.4)$$

2) Эмпирическая функция распределения.

   Пусть $F(x)$ – функция распределения случайной величины X_1, то есть
$$\Psi(\mathsf{P}) = F(x) = \mathsf{P}(X_1 < x) = \int\limits_{-\infty}^{x} d\mathsf{P}(y) < \infty, \quad x \in \mathbf{R}^1.$$

   Тогда имеем эмпирическую функцию распределения
$$F_n(x) = \Psi(\mathsf{P}_n) = \int\limits_{-\infty}^{x} d\mathsf{P}_n(X_1, \cdots, X_n; y) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty,\, x)}(X_i). \qquad (6.1.5)$$

3) Эмпирическая характеристическая функция.

   Пусть $f(t)$ – характеристическая функция случайной величины X_1, то есть
$$\Psi(\mathsf{P}) = f(t) = \mathsf{E}_\mathsf{P} e^{itX_1} = \int e^{itx} d\mathsf{P}(x), \quad t \in \mathbf{R}^1.$$

   Тогда эмпирическая характеристическая функция имеет вид
$$f_n(t) = \Psi(\mathsf{P}_n) = \int e^{itx} d\mathsf{P}_n(X_1, \cdots, X_n; x) = \frac{1}{n} \sum_{j=1}^{n} e^{itX_j}. \qquad (6.1.6)$$

6.2 СПИСОК ЛИТЕРАТУРЫ

1) Ж. – Р. Барра, Основные Понятия Математической Статистики, Москва, Мир, 1974, Глава 1 ¡1, ¡2, ¡5.

2) А.А. Боровков, Математическая Статисика, Москва, Наука, 1984, Глава 1.

3) Ж.– Л. Соле, Основные Структуры Математической Статистики, Москва, Мир, 1972, Глава 1.

4) П.Л. Хеннекен, А. Тортра, Теория Вероятностей и Некоторые её Приложения, Москва, Наука, 1974, Глава 7.

Лекция 7

В противоположнось понятиям, введенным в предыдущей Лекции, которые обобщают основные определения теории вероятностей, в Лекции вводится фундаментальное понятие достаточности, являющееся собственно статистическим.

7.1 ДОСТАТОЧНЫЕ СТАТИСТИКИ

Обработка результатов наблюдений и представление их в виде, наиболее подходящем для принятия решений, является одной из важнейших задача на начальной стадии статистического исследования. При этой первичной обработке результатов наблюдений объём исходного множества выборочных значений уменьшается до относително небольшого числа статистик. При этом было бы желательно, что бы при этом не было потери информации, необходимой для принятия решения. Понятие достаточности и служит для математической формализации этой процедуры.

Определение 7.1.1. *Пусть $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ – статистическая структура. $\mathcal{D} \subseteq \mathcal{F}$ – σ – подалгебра называется достаточной, если для любого A из $\mathcal{F}$ существует вариант условной вероятности*

$$\mathsf{P}(A \mid \mathcal{D}) \equiv \mathsf{P}(X \in A \mid \mathcal{D}),$$

не зависящий от P *из* $\mathcal{P}$.

Это определение равносильно тому, что для каждой интегрируемой статистики Z существует вариант условного математического ожидания $\mathsf{E}(Z \mid \mathcal{D})$, не зависящий от P из $\mathcal{P}$. Действительно, если указанное свойство верно для индикаторов множеств из $\mathcal{F}$, то, образуя их линейные комбинации и переходя к пределам, убеждаемся в справедливости этого факта для всех интегрируемых статистик.

Отметим, что понятие достаточности σ – алгебры непосредственно связано с семейством $\mathcal{P}$. Ясно, что при расширении $\mathcal{P}$ σ – алгебра $\mathcal{D}$ не обязана оставаться достаточной.

Определение 7.1.2. *Статистика T, заданная на $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$ со значениями в $(\mathcal{Y},\ \mathcal{H})$ называется достаточной, если достаточна σ – алгебра $T^{-1}(\mathcal{H})$.*

Хотя в большинстве классических задач важны именно достаточные статистики, понятие достаточной σ – алгебры является, по крайней мере с теоретической точки зрения, более удобным, чем понятие достаточной статистики. Отметим, что существуют примеры σ – алгебр, которые не порождаются никакой достаточной статистикой со значениями в заданном измеримом пространстве.

Пример 7.1.1. Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые случайные величины, имеющие рапределение Пуассона с неизвестным параметром $\theta > 0$.

$$X_i \sim \mathcal{P}(\theta), \quad i = 1, \cdots, n.$$

Докажем, что статистика

$$T = \sum_{i=1}^{n} X_i$$

является достаточной. С этой целью заметим, что

$$T = \sum_{i=1}^{n} X_i \sim \mathcal{P}(n\theta),$$

поэтому, если $\sum\limits_{i=1}^{n} k_i = m, \ \ k_i,\ m \in \{0, 1, 2, \cdots\}$, то

$$\mathsf{P}_\theta(X_1 = k_1, \cdots, X_n = k_n \mid T = m) = \frac{e^{-n\theta}\theta^m m!}{e^{-n\theta}(n\theta)^m k_1! \cdots k_n!} = \frac{m!}{n^m k_1! \cdots k_n!}.$$

Если $\sum\limits_{i=1}^{n} k_i \neq m$, то эта вероятность равна нулю. Итак

$$\mathsf{P}(X \in A \mid T = m) = \sum_{k_1, \cdots, k_n \in A} \mathsf{P}_\theta(X_1 = k_1, \cdots, X_n = k_n \mid T = m), \quad A \in \mathcal{F}$$

не зависит от θ.

Покажем на эвристическом уровне, что достаточная статистика $= T(X)$ содержит о $\theta \in \Theta$ ($\mathsf{P} \in \mathcal{P}$) ту же информацию, что и , то есть при переходе от к информация о θ не теряется.

Поскольку распределение $\mathsf{P}_\theta(X \in A \mid T = t)$ по определению не зависит от θ, то мы можем в принципе при каждом t смоделировать статистику Z_t, имеющую это распределение

$$\mathsf{P}(Z_t \in A) = \mathsf{P}_\theta(X \in A \mid T = t), \quad A \in \mathcal{F},$$

причём эта статистика Z_t информацию о θ не содержит. Смоделируем теперь статистику $T = T(X)$ так, что бы она не зависила от Z_t, тогда составная статистика $Z_{T(X)}$ имеет такое же распределение как и , и, значит, вся информация о θ содержится в статистике .

Докажем, что у X и $Z_{T(X)}$ совпадают распределения. Используя свойства условных математических ожиданий, имеем

$$\mathsf{P}_\theta(Z_{T(X)} \in A) = \mathsf{E}_\theta \mathsf{P}_\theta(Z_{T(X)} \in A \mid T = t) = \mathsf{E}_\theta \mathsf{P}_\theta(Z_t \in A \mid T = t) =$$
$$= \mathsf{E}_\theta \mathsf{P}(Z_t \in A) = \mathsf{E}_\theta \mathsf{P}_\theta(X \in A \mid T = t) = \mathsf{P}_\theta(X \in A), \quad A \in \mathcal{F}.$$

Пример 7.1.2. Предположим, что наблюдение X распределено нормально со средним ноль и неизвестной дисперсией $\theta^2 > 0$

$$X \sim \mathcal{N}(0,\, \theta^2).$$

Тогда распределение X симметрично относительно нуля. При условии что $|X| = t$, единственные два возможных значения X есть $\pm t$, и из симметрии следует, что условная вероятность каждого из них равна $1/2$

$$\mathsf{P}\Big(X = t \mid |X| = t\Big) = \mathsf{P}\Big(X = -t \mid |X| = t\Big) = \frac{1}{2}.$$

Таким образом, условное распределение X при заданном $|X|$ не зависит от θ^2, и значит статистика

$$T(X) = |X|$$

достаточна. Теперь наблюдение X' с тем же распределением, что и X, можно получить из T, бросая правильную монету и полагая $X' = T$ или $X' = -T$, когда монета выпадет гербом или решёткой.

Выделение достаточных статистик с помощью Опредление 7.1.2 неудобно, поскольку оно требует, во-первых, угадывания достаточной статистики $T(X)$, которая могла бы быть достаточной, а затем проверки того, является ли условное распределение X при заданном $T(X)$ независящим от θ.

Однако для доминируемых семейств существует простой критерий факторизации (см. Теорему 7.1.3).

Теорема 7.1.1.*Пусть $\mathcal{D} \subseteq \mathcal{F}$ – достаточная σ – подалгебра для статистической структуры $(\mathcal{X},\, \mathcal{F},\, \mathcal{P})$, тогда*

1) *σ – подалгебра $\mathcal{D}$ является достаточной для любой статистической структуры вида*
$$(\mathcal{X},\, \mathcal{F},\, \mathcal{P}'), \qquad \textit{где} \quad \mathcal{P}' \subseteq \mathcal{P}$$
или $\mathcal{P}'$ – выпуклая оболочка $\mathcal{P}$.

2) *Всякая статистика, эквивалентная достаточной, сама достаточна.*

3) *Если $\mathcal{D}' \subseteq \mathcal{D}$ – σ – подалгебра $\mathcal{D}$, тогда $\mathcal{D}'$ достаточна для статистической структуры $(\mathcal{X},\, \mathcal{D},\, \mathcal{P})$ в том и только в том случае, когда $\mathcal{D}'$ – достаточна для исходной статистической структуры $(\mathcal{X},\, \mathcal{F},\, \mathcal{P})$.*

Доказательство. Пункты 1 и 2 непосредственно следуют из определения достаточности. Для доказательства пункта 3 заметим, что если $\mathcal{D}'$ – достататочна для $(\mathcal{X},\, \mathcal{F},\, \mathcal{P})$, то она очевидно достаточна и для $(\mathcal{X},\, \mathcal{D},\, \mathcal{P})$. Обратно, если $\mathcal{D}'$ – достататочна для $(\mathcal{X},\, \mathcal{D},\, \mathcal{P})$, то

$$\mathsf{P}(A \mid \mathcal{D}') = \mathsf{E}(\mathsf{P}(A \mid \mathcal{D}) \mid \mathcal{D}'), \qquad A \in \mathcal{F}.$$

Поскольку по условию $\mathcal{D}$ – достаточная σ – подалгебра для $(\mathcal{X},\, \mathcal{F},\, \mathcal{P})$, то

$$\mathsf{P}(A \mid \mathcal{D})$$

не зависит от P и является $\mathcal{D}$ – измеримой функцией, поэтому, в силу достаточности $\mathcal{D}'$, это условное математическое ожидание не зависит от P, а значит и $\mathsf{P}(A \mid \mathcal{D}')$ не зависит от P. □

В случае доминируемых структур справедлив следующий фундаментальный теоретический результат.

Теорема 7.1.2.*Пусть $(\mathcal{X},\, \mathcal{F},\, \{\mathsf{P}_\theta,\, \theta \in \Theta\})$ – доминируемая статистическая структура, а P^* – привилегированное доминирующее вероятностное распределение. Тогда необходимым и достаточным условием достаточности σ – подалгебры $\mathcal{D} \subseteq \mathcal{F}$ является существование $\mathcal{D}$ – измеримых (для всех $\theta \in \Theta$) плотностей*

$$p_\theta(x) = \frac{d\mathsf{P}_\theta}{d\mathsf{P}^*}(x).$$

При этом условии для всякого $A \in \mathcal{F}$ можно в качестве общего значения условных вероятностей $\mathsf{P}_\theta(A \mid \mathcal{D})$ положить

$$\mathsf{P}_\theta(A \mid \mathcal{D}) = \mathsf{P}^*(A \mid \mathcal{D}), \quad A \in \mathcal{F}.$$

Доказательство. Необходимость. Пусть $\mathcal{D} \subseteq \mathcal{F}$ – достаточная σ – подалгебра. Для каждого события $A \in \mathcal{F}$ обозначим через $\mathsf{P}(A \mid \mathcal{D})$ вариант $\mathsf{P}_\theta(A \mid \mathcal{D})$, не зависящий от $\theta \in \Theta$. Тогда

$$\mathsf{P}_\theta(A \cap B) = \int\limits_B \mathsf{P}_\theta(A \mid \mathcal{D}) d\mathsf{P}_\theta = \int\limits_B \mathsf{P}(A \mid \mathcal{D}) d\mathsf{P}_\theta, \quad \theta \in \Theta,\ B \in \mathcal{D},\ A \in \mathcal{F}. \tag{7.1.1}$$

Поскольку P^* является выпуклой комбинацией не более чем счётного числа $\mathsf{P}_\theta,\ \ \theta \in \Theta'$ (Теорема 6.1.2, пункт 2)

$$\mathsf{P}^*(A) = \sum_{\theta \in \Theta'} c_\theta \mathsf{P}_\theta(A), \quad A \in \mathcal{F},$$

то переходя к выпуклым комбинациям в равенстве (7.1.1), получим

$$\mathsf{P}^*(A \cap B) = \int\limits_B \mathsf{P}(A \mid \mathcal{D}) d\mathsf{P}^*, \quad B \in \mathcal{D},\ A \in \mathcal{F}, \tag{7.1.2}$$

то есть $\mathsf{P}(A \mid \mathcal{D})$ задаёт вариант условной вероятности $\mathsf{P}^*(A \mid \mathcal{D})$. Полагая в соотношении (7.1.1) $B = \mathcal{X}$, теперь получим

$$\mathsf{P}_\theta(A) = \int \mathsf{P}(A \mid \mathcal{D}) d\mathsf{P}_\theta = \int \mathsf{P}^*(A \mid \mathcal{D}) d\mathsf{P}_\theta = \int \mathsf{P}^*(A \mid \mathcal{D}) p_\theta d\mathsf{P}^*, \ \ \theta \in \Theta,\ A \in \mathcal{F}, \tag{7.1.3}$$

причём, поскольку P_θ абсолютно непрерывны относительно P^* на $\mathcal{D}$, то плотность $p_\theta(x)$ является $\mathcal{D}$ – измеримой. Последнее равенство можно записать в виде (см. Лекция 5, свойства 10 и 6 условного математического ожидания)

$$\int \mathsf{P}^*(A \mid \mathcal{D}) p_\theta d\mathsf{P}^* = \mathsf{E}^*\mathsf{P}^*(A \mid \mathcal{D}) p_\theta = \mathsf{E}^*\mathsf{E}^*(\mathbf{1}_A \mid \mathcal{D}) p_\theta =$$

$$= \mathsf{E}^*\mathsf{E}^*(\mathbf{1}_A p_\theta \mid \mathcal{D}) = \mathsf{E}^*\mathbf{1}_A p_\theta = \int\limits_A p_\theta d\mathsf{P}^*, \quad \theta \in \Theta,\ A \in \mathcal{F},$$

то есть

$$\mathsf{P}_\theta(A) = \int\limits_A p_\theta d\mathsf{P}^*, \quad \theta \in \Theta,\ A \in \mathcal{F}. \tag{7.1.5}$$

Это равенство и означает, что

$$p_\theta(x) = \frac{d\mathsf{P}_\theta}{d\mathsf{P}^*}(x),$$

причём плотность $p_\theta(x)$ является $\mathcal{D}$ – измеримой.

Достаточность. Предположим, что можно выбрать $\mathcal{D}$ – измеримый вариант плотности

$$p_\theta(x) = \frac{d\mathsf{P}_\theta}{d\mathsf{P}^*}(x).$$

Покажем, что тогда $\mathsf{P}^*(A \mid \mathcal{D})$ может служить условной вероятностью $\mathsf{P}_\theta(A \mid \mathcal{D})$ для всех $\theta \in \Theta$. Имеем

$$\mathsf{P}_\theta(A \cap B) = \int \mathbf{1}_A \mathbf{1}_B d\mathsf{P}_\theta = \int \mathbf{1}_A \mathbf{1}_B p_\theta d\mathsf{P}^*, \quad \theta \in \Theta,\ B \in \mathcal{D},\ A \in \mathcal{F}, \quad (7.1.6)$$

но функция $\mathbf{1}_B(x)p_\theta(x)$ $\mathcal{D}$ – измерима, поэтому по свойствам 6 и 10 из Лекции 5 условных математических ожиданий, равенство (7.1.6) можно переписать в виде

$$\mathsf{P}_\theta(A \cap B) = \int \mathsf{E}^*(\mathbf{1}_A \mathbf{1}_B p_\theta \mid \mathcal{D}) d\mathsf{P}^* =$$

$$= \int \mathsf{P}^*(A \mid \mathcal{D}) \mathbf{1}_B p_\theta d\mathsf{P}^* = \int\limits_B \mathsf{P}^*(A \mid \mathcal{D}) d\mathsf{P}_\theta, \quad \theta \in \Theta,\ B \in \mathcal{D},\ A \in \mathcal{F}. \quad (7.1.7)$$

Равенство (7.1.7) показывает, что $\mathsf{P}^*(A \mid \mathcal{D})$ может служить условной вероятностью $\mathsf{P}_\theta(A \mid \mathcal{D})$ для всех $\theta \in \Theta$. □

Следствие 7.1.1.*Пусть $\mathcal{D} \subseteq \mathcal{F}$ – достаточная σ – подалгебра для доминируемой статистической структуры $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$. Тогда любая σ – подалгебра $\mathcal{D}'$, содержащая $\mathcal{D}$, $\mathcal{D} \subseteq \mathcal{D}'$, также достаточна.*

Следствие 7.1.2.*Пусть $\mathcal{D} \subseteq \mathcal{F}$, $\mathcal{D}' \subseteq \mathcal{F}'$ – две достаточные σ – подалгебры для доминируемых статистических структур $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$ и $(\mathcal{X}',\ \mathcal{F}',\ \mathcal{P}')$ соответственно. Тогда σ – подалгебра $\mathcal{D} \times \mathcal{D}'$ достаточна для произведения статистических структур*

$$(\mathcal{X},\ \mathcal{F},\ \mathcal{P}) \otimes (\mathcal{X}',\ \mathcal{F}',\ \mathcal{P}').$$

Доказательство Следствия 7.1.1 непосредственно следует из определения измеримости и Теоремы 7.1.2.

Для доказательства Следствия 7.1.2 заметим, что плотность $p_\theta(x)p'_\theta(x')$ на произведении $\mathcal{X} \times \mathcal{X}'$ является $\mathcal{D} \times \mathcal{D}'$ – измеримой тогда и только тогда, когда $p_\theta(x)$ и $p'_\theta(x)$ являются соответсвенно $\mathcal{D}$ и $\mathcal{D}'$ измеримыми.

В частности, если σ – алгебры $\mathcal{D}$ и $\mathcal{D}'$ порождаются статистиками T и T', то пара $(T,\, T')$ достаточна для произведения статистических структур.

Ясно, что эти свойства также верны и для полупрямых произведений статистических структур.

Теорема 7.1.2 позволяет установить следующий, часто применяемый на практике, критерий достаточности, позволяющий находить достаточные статистики.

Теорема 7.1.3.(Критерий факторизации)*Пусть* $(\mathcal{X},\, \mathcal{F},\, \{p_\theta,\, \theta \in \Theta\})$ – *доминируемая статистическая структура. Статистика* T *со значениями в измеримом пространстве* $(\mathcal{Y},\, \mathcal{H})$ *является достаточной тогда и только тогда, когда существуют*

1) неотрицательная $\mathcal{F}$ *– измеримая функция* $h(x)$ *на* $\mathcal{X}$,

2) $\mathcal{H}$ *– измеримая для всех* $\theta \in \Theta$ *функция* $g_\theta(t)$ *на* $\mathcal{Y}$ *такие, что*

$$p_\theta(x) = g_\theta(T(x))h(x) \qquad \textit{п. в.} \quad \theta \in \Theta, \quad x \in \mathcal{X}.$$

Доказательство. Пусть P^* – привилегированное вероятностное распределение, доминирующее статистическую структуру $(\mathcal{X},\, \mathcal{F},\, \{p_\theta,\, \theta \in \Theta\})$. Тогда по Теореме 7.1.2 необходимым и достаточным условием достаточности статистики T является существование $T^{-1}(\mathcal{H})$ – измеримых (для всех $\theta \in \Theta$) плотностей

$$p_\theta(x) = \frac{d\mathsf{P}_\theta}{d\mathsf{P}^*}(x).$$

Учитывая Утверждение 1.1.2 это условие эквивалентно существованию $\mathcal{H}$ – измеримой при всех $\theta \in \Theta$ функции $g_\theta(t)$ на $(\mathcal{Y},\, \mathcal{H})$ такой, что

$$p_\theta(x) = \frac{d\mathsf{P}_\theta}{d\mathsf{P}^*}(x) = g_\theta(T(x)).$$

Если ν – мера, доминирующая исходную статистическую структуру $(\mathcal{X}, \mathcal{F}, \{p_\theta, \theta \in \Theta\})$, то по Теореме 6.1.2 P^* абсолютно непрерывна относительно ν и при

$$\frac{d\mathsf{P}^*}{d\nu}(x) = h(x)$$

имеем

$$p_\theta(x) = g_\theta(T(x))h(x) \qquad \text{п. в.} \quad \theta \in \Theta, \quad x \in \mathcal{X}.$$

□

Замечания.

1) Заметим, что функция $h(x)$ может равняться нулю только на $\mathcal{P}$ – пренебрежимых множествах. Поскольку, пусть $N \in \mathcal{F}$ таково, что
$$\mathsf{P}(N) = 0, \quad \text{для всех} \quad \mathsf{P} \in \mathcal{P}.$$
Но тогда и
$$\mathsf{P}^*(N) = 0.$$
Следовательно
$$\mathsf{P}^*(N) = \int\limits_N h(x) d\nu(x) = 0. \tag{7.1.8}$$
Если $\nu(N) = 0$, то положим $h(x) = 0$ при $x \in N$. С другой стороны, если $\nu(N) > 0$, то из (7.1.8) следует, что $h(x) = 0$ при $x \in N$.

2) Пусть P^*_T и $\mathsf{P}_{\theta T}$ – распределения, индуцируемые на $(\mathcal{Y}, \mathcal{H})$, исходя из P^* и P_θ соответственно, то есть
$$\mathsf{P}^*_T(B) = \mathsf{P}^*(T^{-1}(B)), \quad \mathsf{P}_{\theta T}(B) = \mathsf{P}_\theta(T^{-1}(B)), \quad B \in \mathcal{H}.$$
Тогда
$$g_\theta(t) = \frac{d\mathsf{P}_{\theta T}}{d\mathsf{P}^*_T}(t).$$
Поскольку, используя формулу замены переменного (см. Лекция 2, пункт 8(f)), имеем
$$\mathsf{P}_{\theta T}(B) = \mathsf{P}_\theta(T^{-1}(B)) = \int\limits_{T^{-1}(B)} d\mathsf{P}_\theta(x) =$$
$$= \int\limits_{T^{-1}(B)} g_\theta(T(x)) d\mathsf{P}^*(x) = \int\limits_B g_\theta(t)) d\mathsf{P}^*_T(t), \quad B \in \mathcal{H}.$$

Пример 7.1.3. Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково нормально распределённые наблюдения
$$X_i \sim \mathcal{N}(0, \theta^2), \ \theta^2 \in \Theta = (0, +\infty) \quad i \in \mathbf{N}.$$
Найдём достаточную статистику в этом случае. Применим Теорему 7.1.3. С этой целью заметим, что совместная плотность X имеет вид
$$p_\theta(x) = \sigma^{-n} \prod_{i=1}^n \varphi(x_i/\sigma) =$$

$$= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\Big\{-1/2\sigma^2 \sum_{i=1}^{n} x_i^2\Big\}, \quad x = (x_1, \cdots, x_n).$$

Таким образом, по Теореме 7.1.3 вместо n – мерного вектора наблюдений $X = (X_1, \cdots, X_n)$, имеем *одномерную* достаточную статистику вида

$$T = \sum_{i=1}^{n} X_i^2.$$

7.2 СПИСОК ЛИТЕРАТУРЫ

1) Ж. – Р. Барра, Основные Понятия Математической Статистики,
Москва, Мир, 1974, Глава 2 ¡1, ¡2.

2) Ш. Закс, Теория Статистических Выводов,
Москва, Мир, 1975, Глава 2, ¡2.1 – 2.3.

3) Ж.– Л. Соле, Основные Структуры Математической Статистики,
Москва, Мир, 1972, Глава 2, ¡2.

4) П.Л. Хеннекен, А. Тортра, Теория Вероятностей и Некоторые её Приложения,
Москва, Наука, 1974, Глава 7, ¡27.

Лекция 8

Статистическому эксперименту, в котором проводятся определенные наблюдения, отвечает некоторая статистическая структура
$(\mathcal{X},\ \mathcal{F},\ \{\mathsf{P}_\theta,\ \theta \in \Theta\})$. *В предыдущих Лекциях расматривались некоторые свойства этих структур. Выясним теперь, как обрабатывать полученные данные с помощью статистических методов.*

8.1 РЕШЕНИЯ И СТРАТЕГИИ

Определение 8.1.1.*Задача статистического решения определяется заданием статистической структуры* $(\mathcal{X},\ \mathcal{F},\ \{\mathsf{P}_\theta,\ \theta \in \Theta\})$ *и измеримого пространства* $(\Delta,\mathcal{U})$. *Стратегией* S *называется переходная вероятность* $S(x,D)$, *заданная на* $\mathcal{X} \times \mathcal{U}$ *такая, что:*

1) для всех $x \in \mathcal{X}$ $S(x,D)$ *– вероятность на* $\mathcal{U}$*;*

2) для всех $D \in \mathcal{U}$ *функция* $S(x,D)$ $\mathcal{F}$ *– измерима по .*

Пространство $(\Delta,\mathcal{U})$ интерпретируется как *пространство решений*. Требуемая информация о θ, нужная для принятия решения, содержится в указанном заранее множестве решений Δ. Если $X = x$ – наблюдения, то решение $\delta \in \Delta$ принимается согласно вероятностному распределению $S(x,\cdot)$ на измеримом пространстве $(\Delta,\ \mathcal{U})$. В частности, если $S(x,\cdot)$ – *вырожденное* распределение в точке $\delta(x) \in \Delta$, то есть

$$S(x,D) = \mathbf{1}_D(\delta(x)),\ D \in \mathcal{U},$$

то стратегия называется *нерандомизированной (детерминированной)*, и такая стратегия состоит просто в принятии решения $\delta(x) \in \Delta$ на основе наблюдения $X = x$.

С математической точки зрения легко понять, почему не следует ограничиваться рассмотрением нерандомизированных стратегий: множество всех стратегий является *выпуклым*, в то время как подмножество нерандомизированных стратегий таковым не является.

Примеры.

1) Пусть $\Delta = \Theta$ и стратегия $S(x,\cdot)$ вырожденна в точке $\delta(x) \in \Delta$, то есть

$$S(x, D) = \mathbf{1}_D(\delta(x)),\ D \in \mathcal{U}.$$

В этом случае имеем задачу *оценивания* – по каждому наблюдению X рассматриваем *оценку* $\delta(X)$ параметра $\theta \in \Theta$.

2) Пусть Δ – класс подмножеств Θ и решение – подмножество множества Θ. Это так называемая задача *доверительного оценивания*.

3) Пусть

$$\Theta = \sum_{j=1}^{k} \Theta_j.$$

То есть множество Θ разбито на k непересекающихся подмножества $\Theta_j,\ j = 1, \cdots, k$ и пусть

$$\Delta = \{\delta_1, \cdots, \delta_k\},$$

где решения $\delta_j,\ j = 1, \cdots, k$ интерпретируется как

$$\theta \in \Theta_j,\ j = 1, \cdots, k.$$

Это – *проверка статистических гипотез*.

4) Пусть из партии однородных деталей берутся n изделий для контроля. Как на основе случайного числа $X, X \leq n$, дефектных изделий принять решение о том, можно ли считать всю продукцию годной или нет?

Здесь $\mathcal{X} = \{0, 1, \cdots, n\},\ \mathcal{F} = \Upsilon(\mathcal{X})$,

$$\mathcal{P} = \{\mathcal{B}(n, \theta),\ \theta \in \Theta = [0, 1]\},\ \Delta = \{\delta_0, \delta_1\}, \quad \mathcal{U} = \Upsilon(\Delta),$$

где δ_0 – решение, состоящее в том, что партия принимается, а δ_1 – партия отвергается. Обычно рассматривают стратегию следующего типа: назначается пороговое значение m и считается, что если $X > m$, то партия отвергается, в противном случае она принимается. Таким образом стратегию $S(x,\cdot)$ можно задать только на двух точках

$$S(x, \delta_0) = \mathbf{1}_{(m,\,n]}(x), \qquad S(x, \delta_1) = \mathbf{1}_{[0,\,m]}(x).$$

8.2 ВЫБОР СТРАТЕГИИ

Основной задачей математической статистики является выбор стратегии в задаче статистического решения, которая была бы оптимальна относительно некоторой конкретной меры качества. Такой выбор с математической точки зрения естественно проводить, вводя отношение *частичного порядка* в классе всех стратегий. Даваемые ниже определения перефразируют понятия максимального элемента, максимума и кофинального множества.

Определение 8.2.1.

1) *Если пространство стратегий $S = S(x, \cdot)$ частично упорядочено с помощью данного упорядочения $\preceq$, то стратегия S^* называется оптимальной, если $S \preceq S^*$ для любой другой стратегии S. Стратегия называется допустимой, если не существует другой стратегии, строго превосходящей ее в смысле частичного упорядочения $\preceq$.*

2) *Семейство стратегий $\mathcal{S}$ называется полным относительно заданного упорядочения $\preceq$, если для любой стратегии S существует стратеги $S' \in \mathcal{S}$ такая, что $S \preceq S'$.*

3) *Функцией потерь называется измеримая функция вида*

$$L(\theta, \delta): \ \Theta \times \Delta \longrightarrow [0, +\infty).$$

При этом обычно $L(\theta, \delta)$ имеет смысл ущерба от принятия решения $\delta \in \Delta$ при истинном значении параметра $\theta \in \Theta$.

4) *Если $S(x, \cdot)$ – стратегия, а $L(\theta, \delta)$ – функция потерь, то величина вида*

$$W_S(\theta, x) = \int\limits_{\Delta} L(\theta, \delta) dS(x, \delta)$$

называется средним ущербом.

5) *Функция вида*

$$R_S(\theta) = \mathsf{E}_\theta W_S(\theta, X) = \int\limits_{\mathcal{X}} W_S(\theta, x) d\mathsf{P}_\theta(x), \ \theta \in \Theta,$$

называется функцией риска или риском.

6) *Если в параметрическом множестве* Θ *выделена* σ – *алгебра* $\mathcal{V}$, *то априорным распределением на статистической структуре* $(\mathcal{X}, \mathcal{F}, \{\mathsf{P}_\theta, \theta \in (\Theta,\ \mathcal{V}))$ *называется любая вероятностная мера* $Q(\cdot)$ *на измеримом пространстве* $(\Theta,\ \mathcal{V})$.

7) *Если* $Q(\cdot)$ – *априорное распределение на* $(\Theta,\ \mathcal{V})$, *то число*

$$R_S^Q = \int\limits_{\Theta} R_S(\theta) dQ(\theta)$$

называется байесовским риском.

8) *Функция риска* $R_S(\theta)$ *задает частичное упорядочение* $\preceq$ *в пространстве стратегий вида*

$$S' \preceq S \Longleftrightarrow R_S(\theta) \leq R_{S'}(\theta) \quad \text{для всех} \;\; \theta \in \Theta$$

и

$$S' \prec S \Longleftrightarrow R_S(\theta) \leq R_{S'}(\theta) \quad \text{для всех} \quad \theta \in \Theta;$$

$$\text{существует} \;\; \theta_0 \in \Theta \quad \text{такое, что} \;\; R_S(\theta_0) < R_{S'}(\theta_0).$$

9) *Функция потерь* $L(\theta, \delta)$ *и априорное распределение* $Q(\cdot)$ *задают линейное упорядочение на множестве всех стратегий с помощью байесовского риска*

$$S' \preceq S \Longleftrightarrow R_S^Q \leq R_{S'}^Q.$$

Введение априорного распределения $Q(\cdot)$ в статистической задаче является основным при так называемом *байесовском подходе* в статистике, в котором предполагается, что параметр θ является *случайной величиной* (хотя и ненаблюдаемой) с *известным* распределением $Q(\cdot)$. Это априорное распределение (априорное относительно наличия данных), цель которого – задание информации перед началом эксперимента или предварительных данных о неизвестном параметре θ, в некоторых задачах можно обосновать. Предположим, например, что мы хотим оценить вероятность выпадения "орла" при бросании монеты. До сих пор мы расматривали n бросаний монеты как множество n биномиальных испытани с неизвестной вероятностью выпадения "орла" θ. Предположим, однако, что мы имеем значительный опыт бросания монет, опыт, который, возможно, дал нам приближённое значение θ для большого числа подобных монет. Если мы считаем, что этот опыт имеет отношение и к даной монете, то было бы, быть может, разумно представить это прошлое знание в виде вероятностного распределение для θ, приближённая форма которого подсказана более ранними данными. Выбор априорного распределения $Q(\cdot)$ проводится обычно, как и выбор распределений $\mathsf{P}_\theta(\cdot)$, путём комбинирования опыта и удобства. Когда мы делаем допущение о том, что количество атмосферных осадков имеет гамма – распределение, мы делаем это не потому, что действительно верим, что это именно так, а потому, что семейство гамма – распределений есть двухпараметрическое семейство, которое, по-видимому, довольно хорошо соответствует данным и которое с математической точки зрения весьма удобно. Аналогично, мы можем получить априорное распределение, отправляясь от достаточно богатого семейства, с которым в математическом отношении легко обращаться, и выбирая из этого семейства распределение, которое аппроксимирует наш прошлый опыт. Такой подход, при котором модель включает в себя априорное распределение для θ с тем, чтобы отразить прошлый опыт, является полезным в тех областях, где имеется большой такой предшествующий опыт.

Примеры.

1) Пусть $\Delta = \Theta \subseteq \mathbf{R}^1$. Рассмотрим нерандомизированные стратегии $S(x, \cdot)$, вырожденные в точках $\delta(x) \in \Delta$. Пусть функция потерь имеет вид (квадратичная ошибка)

$$L_1(\theta, \delta) = c\,(\delta - \theta)^2, \qquad c > 0,$$

тогда

$$W_S(\theta, x) = \int\limits_{\Delta} L_1(\theta, \delta) dS(x, \delta) = L_1(\theta, \delta(x)).$$

Функция риска имеет вид

$$\begin{aligned} R_S(\theta) = \mathsf{E}_\theta W_S(\theta, X) = \mathsf{E}_\theta L_1(\theta, \delta(X)) = \\ = c\,\mathsf{E}_\theta(\delta(X) - \theta)^2. \end{aligned} \tag{8.2.1}$$

Теперь рассмотрим другую функцию потерь

$$L_2(\theta, \delta) = \begin{cases} 0, & |\theta - \delta| < \varepsilon \\ c, & |\theta - \delta| \geq \varepsilon \end{cases} \quad c > 0,\ \varepsilon > 0.$$

Тогда

$$W_S(\theta, x) = \int\limits_{\Delta} L_2(\theta, \delta) dS(x, \delta) = L_2(\theta, \delta(x))$$

и

$$R_S(\theta) = \mathsf{E}_\theta L_2(\theta, \delta(X)) = c\,\mathsf{P}_\theta(|\delta(X) - \theta| \geq \varepsilon). \tag{8.2.2}$$

2) Рассмотрим теперь задачу проверки двух гипотез

$$\Theta = \{\theta_0, \theta_1\}, \quad \Delta = \{\delta_0, \delta_1\},$$

где δ_i – решение, состоящее в том, что $\theta = \theta_i,\ i = 0, 1$. В этом случае любая стратегия $S(x, \cdot)$ задаётся измеримой функцией $\alpha(x) \in [0, 1]$

$$S(x, \delta_0) = \alpha(x), \qquad S(x, \delta_1) = 1 - \alpha(x).$$

Пусть, например, функция потерь имеет вид

$$L(\theta, \delta) = \begin{cases} 0, & \theta = \theta_0, \quad \delta = \delta_0 \\ c_1 > 0, & \theta = \theta_0, \quad \delta = \delta_1 \\ 0, & \theta = \theta_1, \quad \delta = \delta_0 \\ c_2 > 0, & \theta = \theta_1, \quad \delta = \delta_0 \end{cases}$$

В этом случае

$$W_S(\theta, x) = \int\limits_{\Delta} L(\theta, \delta) dS(x, \delta) = L(\theta, \delta_0)\alpha(x) + L(\theta, \delta_1)(1 - \alpha(x)) =$$

$$= \begin{cases} c_1(1 - \alpha(x)), & \theta = \theta_0 \\ c_2\alpha(x), & \theta = \theta_1. \end{cases}$$

Для функции риска имеем выражение

$$R_S(\theta) = \mathsf{E}_\theta W_S(\theta, X) = \begin{cases} c_1\, \mathsf{E}_{\theta_0}(1 - \alpha(X)), & \theta = \theta_0 \\ c_2\, \mathsf{E}_{\theta_1}\alpha(X), & \theta = \theta_1. \end{cases} \tag{8.2.3}$$

Теорема 8.2.1. *Пусть функция потерь $L(\theta, \delta)$ при всех $\theta \in \Theta$ непрерывна и выпукла вниз по $\delta \in \Delta$, и $\Delta \subseteq \mathbf{R}^1$ – выпуклое ограниченное множество, тогда семейство нерандомизированных стратегий полно относительно частичного упорядочения, порождаемого функцией потерь $L(\theta, \delta)$.*

Доказательство. Заметим, что если $g(x)$ – выпуклая вниз измеримая функция, то для любой интегрируемой случайной величины ξ справедливо неравенство Иенсена

$$\mathsf{E}g(\xi) \geq g(\mathsf{E}\xi). \tag{8.2.4}$$

(Доказательство можно найти, например, в [4], стр. 207.) Поскольку по условию множество Δ ограничено, то существует интеграл

$$\delta(x) = \int\limits_{\Delta} \delta\, dS(x, \delta),$$

из неравенства Иенсена следует оценка

$$\int_{\Delta} L(\theta, \delta)\, dS(x, \delta) \geq L(\theta, \delta(x)).$$

Интегрируя теперь по P_θ, получим

$$R_S(\theta) \geq R_{S^*}(\theta), \quad \text{для всех} \quad \theta \in \Theta,$$

где $S^*(x, \cdot)$ – вырожденная стратегия в точке $\delta(x)$, то есть для любой стратегии S существует *нерандомизированная* стратегия S^*, которая не хуже, чем S. Поэтому класс всех нерандомизированных стратегий полон. □

Задачи.

1) Какие условия накладываются на множество Θ?
2) Где в доказательстве Теоремы 8.2.1 использовалась выпуклость множества Δ?
3) Обобщить Теорему 8.2.1 на случай $\Delta \subseteq \mathbf{R}^k, \quad k > 1$.

8.3 СПИСОК ЛИТЕРАТУРЫ

1) Ж. – Р. Барра, Основные Понятия Математической Статистики, Москва, Мир, 1974, Глава 4, ¡1 – ¡5.

2) Н.Н. Ченцов, Статистические Решающие Правила и Оптимальные Выводы,
Москва, Наука, 1972, Введение, ¡1, ¡2, Глава 1, ¡5.

3) Г. Чернов, Л. Мозес, Элементарная Теория Статистических Решений, Москва, Советское Радио, 1962, Главы 1 – 6.

4) А.Н. Ширяев, Вероятность,
Москва, Наука, 1989.

Лекция 9

В Лекции приводятся основные понятия и факты теории оценивания, которая рассматривается как частный случай общей проблемы статистических решений.

9.1 ТЕОРИЯ ОЦЕНИВАНИЯ

Пусть $(\mathcal{X},\ \mathcal{F},\ \{\mathsf{P}_\theta, \theta\in\Theta\})$ – доминируемая статистическая структура и $(\Delta,\ \mathcal{U})$ – пространство решений. Пусть $g(\theta)$ некоторая измеримая функция, заданная на Θ и действующая в некоторое измеримое пространство $(\Gamma,\ \mathcal{W})$. Предположим, что по результатам наблюдений $X = x$ мы хотим "оценить" значение $g(\theta)$. В этом случае естественно положить $\Gamma = \Delta$. Предположим для простоты, что $\Gamma \subseteq \mathbf{R}^1$. Функция потерь

$$L(\theta, \delta):\ \Theta\times\Delta \longrightarrow [0,\infty)$$

характеризует "близость" $g(\theta)$ и δ. Обычно рассматривают функции потерь вида

$$L(\theta,\delta) = (g(\theta)-\delta)^2, \qquad L(\theta,\delta) = c(\theta)(g(\theta)-\delta)^2, \quad c(\theta)\ \geq\ 0,$$

$$L(\theta,\delta) = |g(\theta)-\delta|,$$

$$L(\theta,\delta) = \begin{cases} (g(\theta)-\delta)^2, & |g(\theta)-\delta|\ \leq\ k \\ 2k|g(\theta)-\delta| - k^2, & |g(\theta)-\delta|\ \geq\ k. \end{cases}$$

Все эти функции при фиксированном θ выпуклы вниз по δ, поэтому естественно в этом разделе рассматривать только выпуклые вниз функции потерь. Теперь, предполагая, что условия регулярности Теоремы 8.2.1 выполнены, получаем, что для любой стратегии $S(x,\cdot)$ существует *нерандомизированная* стратегия $S^*(x,\cdot)$ вида

$$S^*(x,D) = \mathbf{1}_D(\delta(x)), \quad D \in \mathcal{U},$$

которая не хуже чем $S(x,\cdot)$, то есть

$$R_S(\theta) \geq R_{S^*}(\theta), \quad \text{для всех} \quad \theta \in \Theta.$$

Таким образом будем рассматривать только нерандомизированные стратегии $S^*(x,\cdot)$ и отождествлять их с измеримой функцией $\delta(x) \in \Delta$.

Определение 9.1.1.

1) *Оценкой параметрической функции $g(\theta)$ называется измеримая функция*

$$\delta = \delta(x) : \mathcal{X} \longrightarrow \Gamma,$$

которая не зависит от θ.

2) *Риском оценки $\delta(X)$ называется функция (см. (8.2.1))*

$$R(\theta,\delta) = \mathsf{E}_\theta L(\theta,\delta(X)).$$

Покажем, как естественным образом возникают квадратичные функции потерь. Предполагая функцию потерь $L(\theta,\delta)$ достаточно гладкой и используя формулу Тейлора, имеем

$$L(\theta,\delta) = L(\theta,g(\theta)) + {L_\delta}'(\theta,g(\theta))(g(\theta)-\delta) + \\ + \frac{{L_\delta}''(\theta,g(\theta))}{2}(g(\theta)-\delta)^2 + R. \tag{9.1.1}$$

Из естественных предположений на функцию потерь $L(\theta,\delta)$ следует, что

$$L(\theta,g(\theta)) \equiv 0, \quad \theta \in \Theta.$$

Далее условие неотрицательности $L(\theta,\delta) \geq 0$ означает, что

$$L'_\delta(\theta,g(\theta)) \equiv 0, \quad \theta \in \Theta.$$

Таким образом, отбрасывая остаточный член R в формуле (9.1.1), получим аппроксимацию

$$L(\theta,\delta) \approx \frac{{L_\delta}''(\theta,g(\theta))}{2}(g(\theta)-\delta)^2 \equiv c(\theta)(g(\theta)-\delta)^2.$$

Поэтому обычно рассматривают квадратичные функции потерь.

Определение 9.1.2.

1) Если
$$L(\theta, \delta) = (g(\theta) - \delta)^2,$$
то величина
$$R(\theta, \delta) = \mathsf{E}_\theta(\delta(X) - g(\theta))^2$$
называется среднеквадратичной ошибкой.

2) Величина
$$b(\theta) = \mathsf{E}_\theta \delta(X) - g(\theta)$$
называется смещением оценки $\delta(X)$.

3) Оценка $\delta(X)$ называется несмещенной, если ее смещение равно нулю, то есть, если
$$\mathsf{E}_\theta \delta(X) \equiv g(\theta),\ \theta \in \Theta.$$

Если функция потерь квадратична, то
$$R(\theta, \delta) = \mathsf{E}_\theta(\delta(X) - g(\theta))^2 = \mathsf{D}_\theta \delta(X) + b^2(\theta).$$
И если $\delta(X)$ – *несмещенная оценка*, то её риск совпадает с дисперсией, то есть
$$R(\theta, \delta) = \mathsf{D}_\theta \delta(X).$$

Пример 9.1.1. Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения и
$$\mathsf{E}_\theta X_1 = \theta.$$
Тогда оценки вида
$$\delta(X) = \sum_{i=1}^n \alpha_i X_i, \quad \text{где} \quad \sum_{i=1}^n \alpha_i = 1$$
являются несмещёнными оценками параметра θ. Этот Пример показывает, что в общем случае несмещённых оценок много.

Итак, точность оценки $\delta(X)$ функции $g(\theta)$ измеряется функцией риска
$$R(\theta, \delta) = \mathsf{E}_\theta L(\theta, \delta(X)),$$
то есть средними потерями в результате использования оценки $\delta(X)$ в течение длительного промежутка времени. Хотелось бы найти такую оценку $\delta(X)$ которая бы минимизировала риск $R(\theta, \delta)$ при всех значениях параметра

$\theta \in \Theta$. В сформулированном виде эта задача решений не имеет. Поскольку, если

$$L(\theta, g(\theta)) \equiv 0, \quad \theta \in \Theta.$$

то риск $R(\theta, \delta)$ для каждой заданной точки $\theta_0 \in \Theta$ можно свести к нулю, выбирая $\delta(x)$ равным $g(\theta_0)$ при всех $x \in \mathcal{X}$. Поэтому *равномерно наилучшей* оценки $\delta(X)$ не существует, то есть нет такой оценки, которая одновременно минимизировала бы риск $R(\theta, \delta)$ *для всех* значений $\theta \in \Theta$, исключая тривиальный случай, когда $g(\theta)$ постоянна.

Один из способов избежания этой трудности состоит в *сужении* класса рассматриваемых оценок путём исключения тех оценок, которые оказывают слишком сильное предпочтение одному или нескольким значениям $\theta \in \Theta$ ценой пренебрежения остальными возможными значениями. Этого можно достигнуть, потребовав, чтобы оценки удовлетворяли некоторому условию, обеспечивающему определённую степень беспристрастности. Одним из таких условий является условие несмещенности оценки

$$\mathsf{E}_\theta \delta(X) \equiv g(\theta), \; \theta \in \Theta.$$

Это условие гарантирует, что в конце концов те количества, на которые оценка $\delta(X)$ пере – или недооценивает $g(\theta)$ сбалансируют друг друга, так что получаемые значения оцениваемой функции будут в среднем правильными. Заметим, однако, что требование несмещённости может приводить к проблемам. Например, несмещённые оценки могут просто не существовать.

Пример 9.1.2. Пусть наблюдение X имеет биномиальное распределение

$$X \sim \mathcal{B}(n, \, \theta), \quad \theta \in \Theta = (0, 1).$$

Предположим, что мы хотим оценить функцию вида

$$g(\theta) = \frac{1}{\theta}.$$

Тогда требование несмещённости равносильно выполнению условия

$$\sum_{k=0}^{n} \delta(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} \equiv \frac{1}{\theta}, \quad \text{для всех} \quad \theta \in (0, 1). \tag{9.1.2}$$

То, что такая оценка не существует, вытекает, например, из того, что при $\theta \to 0$ левая часть тождества (9.1.2) стремится к $\delta(0)$, а правая часть – к бесконечности.

Теорема 9.1.1.(Рао – Блекуэлл – Колмогоров)*Пусть функция потерь $L(\theta,\delta)$ непрерывна и выпукла вниз по $\delta \in \Delta$ для любого фиксированного значения парметра $\theta \in \Theta$ и $T = T(X)$ – достаточная статистика для статистической структуры $(\mathcal{X}, \mathcal{F}, \{\mathsf{P}_\theta, \theta \in \Theta\})$. Предположим, что $\delta(X)$ – некоторая интегрируемая оценка функции $g(\theta)$. Положим*

$$h(t) = \mathsf{E}_\theta(\delta(X) \mid T = t). \tag{9.1.3}$$

Тогда

1) Статистика $h(T)$ является оценкой параметрической функции $g(\theta)$.

2) Для всех $\theta \in \Theta$ риск оценки $h(T)$ не превосходит риска оценки $\delta(X)$

$$R(\theta,\delta) \geq R(\theta,h), \qquad \text{для всех} \quad \theta \in \Theta.$$

3) Если $\delta(X)$ – несмещенная оценка $g(\theta)$, то и $h(T)$ также несмещенная оценка $g(\theta)$.

Доказательство.

1) Измеримость и независимость $h(T)$ от θ следуют из Определения 5.1.3 условного математического ожидания и Определения 7.1.2 достаточной статистики.

2) Применим неравенство Иенсена (8.2.4) к условным математическим ожиданиям (см. также [5], стр. 250, задача 5), получим (используя свойство 6 условных математических ожиданий из Лекции 5)

$$R(\theta,\delta) = \mathsf{E}_\theta L(\theta,\delta(X)) = \mathsf{E}_\theta \mathsf{E}_\theta[L(\theta,\delta(X)) \mid T] \geq$$

$$\geq \mathsf{E}_\theta L(\theta,\, \mathsf{E}_\theta[\delta(X) \mid T]) = R(\theta,h) \quad \text{для всех} \quad \theta \in \Theta.$$

3) Поскольку

$$\mathsf{E}_\theta \delta(X) = g(\theta),$$

то используя свойство 6 условных математических ожиданий из Лекции 5

$$\mathsf{E}_\theta h(T) = \mathsf{E}_\theta \mathsf{E}_\theta[\delta(X) \mid T] = \mathsf{E}_\theta \delta(X) = g(\theta).$$

□

Эта Теорема, в частности, показывает, что при наличии *достаточной статистики* для любой оценки существует оценка, зависящая от наблюдений только через достаточную статистику, и которая не хуже её. (То есть такие оценки образуют *полный класс*.) Поэтому можно ограничиться рассмотрением оценок, зависящих от достаточных статистик. Операция нахождения оценки $h(T)$ по формуле (9.1.3) называется *проектированием* оценки $\delta(X)$ на достаточную статистику T, а сама оценка $h(T)$ называется *проекцией* оценки $\delta(X)$ на достаточную статистику T.

Пример 9.1.3. Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения и

$$\mathsf{E}_\theta X_1 = g(\theta). \tag{9.1.4}$$

Предположим, что достаточная статистика T имеет вид

$$T = \sum_{i=1}^{n} X_i.$$

Очевидно, что в силу (9.1.4) оценка $\delta(X) = X_1$ является несмещенной оценкой $g(\theta)$.

Найдём проекцию $h(T)$ этой оценки на достаточную статистику T.

$$h(T) = \mathsf{E}_\theta(\delta(X) \mid T) = \mathsf{E}_\theta(X_1 \mid T) = \mathsf{E}_\theta(X_2 \mid T) = \cdots = \mathsf{E}_\theta(X_n \mid T) =$$

$$= \frac{1}{n}\mathsf{E}_\theta\left(\sum_1^n X_i \mid T\right) = \frac{1}{n}\mathsf{E}_\theta(T \mid T) = \frac{T}{n} \equiv \overline{X}.$$

9.2 ОПТИМАЛЬНЫЕ ОЦЕНКИ

Рассмотрим в этом разделе более подробно случай квадратичных функций потерь

$$L(\theta, \delta) = c(\theta)(g(\theta) - \delta)^2, \quad c(\theta) \geq 0.$$

Как показано выше, в этом случае риск любой *несмещенной* оценки $\delta(X)$ параметрической функции $g(\theta)$ пропорционален её *дисперсии*, то есть

$$R(\theta, \delta) = c(\theta)\, \mathsf{D}_\theta \delta(X).$$

Таким образом проблема минимизации риска по $\delta(X)$ в этом случае сводится к проблеме минимизации дисперсии. Поэтому вполне естественно следующее Определение.

Определение 9.2.1. *Несмещенная оценка $\delta^*(X)$ функции $g(\theta)$ называется несмещенной оценкой с минимальной дисперсией (или оптимальной), если для любой несмещенной оценки $\delta(X)$ справедливо неравенство*

$$\mathsf{D}_\theta \delta^*(X) \leq \mathsf{D}_\theta \delta(X), \quad \text{для всех} \quad \theta \in \Theta.$$

Всюду в дальнейшем мы молчаливо предполагаем, что рассматриваемые оценки $\delta(X)$ квадратично интегрируемы

$$\mathsf{E}_\theta \delta(X) < \infty, \quad \text{для всех} \quad \theta \in \Theta.$$

Следующая Теорема показывает, что оптимальные оценки действительно существуют.

Теорема 9.2.1. *Относительная частота произвольного события A в n независимых бернуллиевских испытаниях является оптимальной оценкой вероятности этого события.*

Доказательство. Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения вида

$$X_i \sim \mathcal{B}(1, \theta),\ \theta \in \Theta = (0, 1),\ i = 1, \cdots, n.$$

Нам необходимо доказать, что оценка вида

$$\delta^*(X) = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

является оптимальной оценкой $g(\theta) = \theta$.

Найдём дисперсию этой оценки

$$\mathsf{D}_\theta \delta^*(X) = \frac{\mathsf{D}_\theta X_1}{n} = \frac{(1-\theta)\theta}{n}.$$

Теперь для доказательства Теоремы длстаточно показать, что если оценка $\delta(X)$ является несмещённой

$$\mathsf{E}_\theta \delta(X) \equiv \theta, \quad \text{для всех} \quad \theta \in \Theta, \tag{9.2.1}$$

то справедливо неравенство

$$\mathsf{D}_\theta \delta(X) \geq \frac{(1-\theta)\theta}{n}, \quad \text{для всех} \quad \theta \in \Theta. \tag{9.2.2}$$

Из условия несмещенности (9.2.1) имеем

$$\theta \equiv \mathsf{E}_\theta \delta(X) = \sum_x \delta(x)\mathcal{L}(\theta, x), \tag{9.2.3}$$

где

$$x = (x_1, \cdots, x_n), \quad x_i \in \{0,\ 1\}, \quad i = 1, \cdots, n;$$

и функция правдоподобия $\mathcal{L}(\theta, x)$ имеет вид

$$\mathcal{L}(\theta, x) = \theta^{\bar{x}}(1-\theta)^{n-\bar{x}}, \qquad \bar{x} = \sum_{i=1}^n x_i.$$

Справедливо также тождество

$$\sum_x \mathcal{L}(\theta, x) \equiv 1. \tag{9.2.4}$$

Дифференцируя тождества (9.2.3) и (9.2.4) по θ, получим

$$1 \equiv \sum_x \delta(x)\mathcal{L}'_\theta(\theta, x) = \sum_x \delta(x)\frac{\partial}{\partial\theta}\log\mathcal{L}(\theta, x)\cdot\mathcal{L}(\theta, x) =$$

$$= \mathsf{E}_\theta \delta(X)\frac{\partial}{\partial\theta}\log\mathcal{L}(\theta, X), \tag{9.2.5}$$

$$0 \equiv \sum_x \mathcal{L}'_\theta(\theta, x) = \sum_x \frac{\partial}{\partial\theta}\log\mathcal{L}(\theta, x)\cdot\mathcal{L}(\theta, x) = \mathsf{E}_\theta\frac{\partial}{\partial\theta}\log\mathcal{L}(\theta, X). \tag{9.2.6}$$

Учитывая соотношения (9.2.5), (9.2.6) и неравенство Коши – Буняковского, мы можем записать

$$1 = \mathsf{E}_\theta\Big(\delta(X) - \theta\Big)\frac{\partial}{\partial\theta}\log\mathcal{L}(\theta, X) \ \le\ \sqrt{\mathsf{D}_\theta\delta(X)}\sqrt{\mathsf{E}_\theta\left(\frac{\partial}{\partial\theta}\log\mathcal{L}(\theta, X)\right)^2}. \tag{9.2.7}$$

Из неравенства (9.2.7) получаем оценку снизу для дисперсии

$$\mathsf{D}_\theta\delta(X) \ \ge\ \mathsf{E}_\theta^{-1}\left(\frac{\partial}{\partial\theta}\log\mathcal{L}(\theta, X)\right)^2, \tag{9.2.8}$$

но

$$\mathsf{E}_\theta\left(\frac{\partial}{\partial\theta}\log\mathcal{L}(\theta, X)\right)^2 = \mathsf{E}_\theta\left(\frac{\sum\limits_{i=1}^n X_i}{\theta} - \frac{n - \sum\limits_{i=1}^n X_i}{1-\theta}\right)^2 = \frac{\mathsf{E}_\theta\left(\sum\limits_{i=1}^n X_i - n\theta\right)^2}{(\theta(1-\theta))^2} =$$

$$= \frac{\mathsf{D}_\theta \sum\limits_{i=1}^{n} X_i}{(\theta(1-\theta))^2} = \frac{n\mathsf{D}_\theta X_1}{(\theta(1-\theta))^2} = \frac{n}{\theta(1-\theta)}. \tag{9.2.9}$$

Теперь доказываемое соотношение (9.2.2) следует из (9.2.9) и (9.2.8). □

Рассмотрим теперь некоторые свойства *оптимальных* оценок.

Теорема 9.2.2. (Единственность оптимальной оценки)*Пусть $\delta_1^*(X)$ и $\delta_2^*(X)$ – оптимальные оценки функции $g(\theta)$, тогда они совпадают почти всюду, то есть*

$$\mathsf{P}_\theta\Big(\delta_1^*(X) \neq \delta_2^*(X)\Big) = 0, \quad \textit{для всех} \quad \theta \in \Theta.$$

Доказательство. Поскольку $\delta_1^*(X)$ и $\delta_2^*(X)$ – оптимальные оценки, то у них тождественно совпадают дисперсии. Обозначим

$$v = \mathsf{D}_\theta \delta_1^*(X) = \mathsf{D}_\theta \delta_2^*(X)$$

и рассмотрим оценку

$$\delta(X) = \frac{\delta_1^*(X) + \delta_2^*(X)}{2}.$$

Тогда эта оценка также является несмещённой оценкой функции $g(\theta)$ и поэтому для её дисперсии справедливо неравенство

$$v \ \leq \ \mathsf{D}_\theta \delta(X) = \frac{1}{4}\Big(\mathsf{D}_\theta \delta_1(X) +$$

$$+\mathsf{D}_\theta \delta_2(X) + 2\mathsf{Cov}_\theta\Big(\delta_1(X), \delta_2(X)\Big)\Big) \ \leq \ \frac{1}{4}(2v + 2v) = v. \tag{9.2.10}$$

Из этих неравенств следует, что

$$\mathsf{D}_\theta \delta(X) = v.$$

Но тогда из неравенств (9.2.10) получаем также соотношение для ковариации

$$4v = 2v + 2\mathsf{Cov}_\theta\Big(\delta_1(X), \delta_2(X)\Big),$$

то есть

$$\mathsf{Cov}_\theta\Big(\delta_1(X), \delta_2(X)\Big) = v. \tag{9.2.11}$$

Теперь, учитывая соотношение (9.2.11), найдём дисперсию *разности* оценок $\delta_1^*(X)$ и $\delta_2^*(X)$

$$\mathsf{D}_\theta\Big(\delta_1(X) - \delta_2(X)\Big) = \mathsf{D}_\theta \delta_1(X) + \mathsf{D}_\theta \delta_2(X) - 2\mathsf{Cov}_\theta\Big(\delta_1(X), \delta_2(X)\Big) = 2v - 2v = 0.$$

Отсюда следует, что

$$\mathsf{P}_\theta\Big(\delta_1^*(X) \neq \delta_2^*(X)\Big) = 0, \quad \text{для всех} \quad \theta \in \Theta.$$

(Здесь мы использовали следующее утверждение

$$\mathsf{D}Y = 0 \ \Rightarrow \mathsf{E}(Y - \mathsf{E}Y)^2 = 0 \ \Rightarrow Y = \mathsf{E}Y \quad \text{п.в.})$$

□

Теорема 9.2.3.

1) *Пусть $\delta^*(X)$ – оптимальная оценка для функции $g(\theta)$, тогда для любой оценки $\delta_0 = \delta_0(X)$ (несмещенная оценка нуля) такой, что*

$$\mathsf{E}_\theta\delta_0(X) \equiv 0, \quad \mathsf{E}_\theta\delta_0^2(X) < \infty, \qquad \textit{для всех} \quad \theta \in \Theta,$$

справедливо условие "ортогональности"

$$\mathsf{Cov}_\theta\Big(\delta^*(X), \delta_0(X)\Big) \equiv 0, \qquad \textit{для всех} \quad \theta \in \Theta.$$

2) *Пусть оценка $\bar{\delta}(X)$ является несмещенной оценкой своего математического ожидания $g(\theta)$ и для любой оценки нуля $\delta_0(X)$ такой, что*

$$\mathsf{E}_\theta\delta_0(X) \equiv 0, \quad \mathsf{E}_\theta\delta_0^2(X) < \infty, \qquad \textit{для всех} \quad \theta \in \Theta,$$

справедливо тождество

$$\mathsf{Cov}_\theta\Big(\bar{\delta}(X), \delta_0(X)\Big) \equiv 0, \qquad \textit{для всех} \quad \theta \in \Theta.$$

Тогда оценка $\bar{\delta}(X)$ является оптимальной оценкой своего математического ожидания $g(\theta)$.

Доказательство.

1) Для доказательства рассмотрим вспомогательную оценку вида

$$\delta(X) = \delta^*(X) + \lambda\delta_0(X), \quad \lambda \in \mathbf{R}^1.$$

Тогда для всех $\lambda \in \mathbf{R}^1$ эта оценка также является несмещённой оценкой функции $g(\theta)$, поэтому в силу оптимальности оценки $\delta^*(X)$

$$\mathsf{D}_\theta\delta(X) = \mathsf{D}_\theta\delta^*(X) + 2\lambda\mathsf{Cov}_\theta\Big(\delta^*(X), \delta_0(X)\Big) + \lambda^2\mathsf{D}_\theta\delta_0(X) \ \geq$$

$$\geq \mathsf{D}_\theta \delta^*(X) \quad \text{для всех} \quad \theta \in \Theta, \;\; \lambda \in \mathbf{R}^1$$

или

$$2\lambda \mathsf{Cov}_\theta\Big(\delta^*(X), \delta_0(X)\Big) +$$

$$+\lambda^2 \mathsf{D}_\theta \delta_0(X) \;\geq\; 0, \quad \text{для всех} \quad \theta \in \Theta, \;\; \lambda \in \mathbf{R}^1.$$

Этот квадратный многочлен от λ имеет два действительных корня $\lambda = 0$ и

$$\lambda = -\frac{2\mathsf{Cov}_\theta\Big(\delta^*(X), \delta_0(X)\Big)}{\mathsf{D}_\theta \delta_0(X)}$$

и, следовательно, принимает отрицательные значения, если только не выполнено условие

$$\mathsf{Cov}_\theta\Big(\delta^*(X), \delta_0(X)\Big) \equiv 0, \quad \text{для всех} \quad \theta \in \Theta.$$

2) Пусть $\delta(X)$ – произвольная несмещённая оценка $g(\theta)$. Поскольку при

$$\mathsf{D}_\theta \delta(X) = \infty$$

доказывать нечего, то предположим, что

$$\mathsf{D}_\theta \delta(X) < \infty.$$

Тогда ясно, что $\bar{\delta}(X) - \delta(X)$ является несмещённой оценкой нуля и поэтому

$$\mathsf{E}_\theta\Big(\bar{\delta}(X)(\bar{\delta}(X) - \delta(X))\Big) \equiv 0.$$

Отсюда следуют равенства

$$\mathsf{E}_\theta \bar{\delta}^2(X) = \mathsf{E}_\theta \bar{\delta}(X)\delta(X), \quad \mathsf{D}_\theta \bar{\delta}(X) = \mathsf{Cov}_\theta\Big(\bar{\delta}(X), \delta(X)\Big).$$

Поэтому применяя неравенство Коши – Буняковского, имеем

$$\mathsf{D}_\theta \bar{\delta}(X) \;\leq\; \mathsf{D}_\theta \delta(X) \quad \text{для всех} \quad \theta \in \Theta.$$

□

Теорема 9.2.4. *Пусть $\delta_1^*(X)$ и $\delta_2^*(X)$ – оптимальные оценки для функций $g_1(\theta)$ и $g_2(\theta)$ соответственно. Тогда для любых чисел a и b оценка вида*

$$\delta^*(X) = a\delta_1^*(X) + b\delta_2^*(X)$$

является оптимальной оценкой функции

$$g(\theta) = ag_1(\theta) + bg_2(\theta).$$

Доказательство. Доказательство Теоремы непосредственно следует из второй части Теоремы 9.2.3. Однако, для полноты мы дадим и прямое доказательство этой Теоремы.

Пусть $\delta(X)$ – произвольная несмещённая оценка функции $g(\theta) = ag_1(\theta) + bg_2(\theta)$. Тогда оценка

$$\delta_0(X) = \delta^*(X) - \delta(X)$$

является несмещённой оценкой нуля и поэтому по Теореме 9.2.3 справедливо тождество

$$0 = a\mathsf{Cov}_\theta\Big(\delta_1^*(X), \delta_0(X)\Big) + b\mathsf{Cov}_\theta\Big(\delta_2^*(X), \delta_0(X)\Big) =$$

$$= \mathsf{Cov}_\theta\Big(\delta^*(X), \delta_0(X)\Big) = \mathsf{D}_\theta\delta^*(X) - \mathsf{Cov}_\theta\Big(\delta^*(X), \delta(X)\Big),$$

то есть

$$\mathsf{D}_\theta\delta^*(X) = \mathsf{Cov}_\theta\Big(\delta^*(X), \delta(X)\Big) \ \leq\ \sqrt{\mathsf{D}_\theta\delta^*(X)\mathsf{D}_\theta\delta(X)},$$

или

$$\mathsf{D}_\theta\delta^*(X) \ \leq\ \mathsf{D}_\theta\delta(X), \qquad \text{для всех} \quad \theta \in \Theta.$$

□

9.3 БАЙЕСОВСКОЕ ОЦЕНИВАНИЕ

Рассмотрим доминируемую статистическую структуру $(\mathcal{X}, \mathcal{F}, \{\mathsf{P}_\theta, \theta \in \Theta\})$, и задачу оценки параметрической функции $g(\theta)$ по наблюдению $X = x$, но предположим, что θ является значением случайной величины Ξ с *известным* распределением (априорное распределение) $Q(\cdot)$ на $(\Theta, \mathcal{V})$. Случай *неизвестного* априорного распределения будет рассмотрен в Лекциях 18 – 21. В подобной ситуации задачу оценивания называют *задачей оценивания в байесовской постановке.*

Определение 9.3.1. *Байесовской оценкой параметрической функции $g(\theta)$, соответствующей априорному распределению Q называется измеримая функция*

$$\delta_Q = \delta_Q(x) :\ \mathcal{X} \longrightarrow \Gamma,$$

которая минимизирует байесовский риск (см. Определние 8.2.1)

$$r(\delta_Q, Q) = \inf_\delta r(\delta, Q), \qquad r(\delta, Q) = \int\limits_\Theta R(\theta, \delta) dQ(\theta),$$

$$R(\theta, \delta) = \mathsf{E}_\theta L(\theta, \delta(X)) = \int\limits_{\mathcal{X}} L(\theta, \delta(x)) p_\theta(x) d\nu(x).$$

При байесовском подходе плотность $p_\theta(x)$ интерпретируется как *условная плотность* вида $p_\theta(x) = p(x \mid \Xi = \theta)$ а риск $R(\theta, \delta)$ – как условный риск $R(\theta, \delta) = R(\delta \mid \Xi = \theta)$, тогда байесовский риск получается как усреднение условного риска

$$r(\delta, Q) = \mathsf{E} R(\delta \mid \Xi) = \int\limits_\Theta R(\theta, \delta) dQ(\theta).$$

Теорема 9.3.1.*Пусть случайная величина Ξ имеет распределение Q и при данном $\Xi = \theta$ наблюдение X имеет распределение P_θ. Предположим, кроме того, что в задаче оценивания параметрической функции $g(\theta)$ с неотрицательной функцией потерь $L(\theta, \delta)$ выполнены следующие условия*

1) *Существует оценка $\delta_0(X)$ с конечным байесовским риском $r(\delta_0, Q) < \infty$.*

2) *Для почти всех x существует значение $\delta_Q(x)$, минимизирующее по $\delta \in \Delta$*
$$\mathsf{E}(L(\Xi, \delta) \mid X = x). \tag{9.3.1}$$

Здесь X имеет распределение

$$\mathsf{P}(A) = \int\limits_\Theta \mathsf{P}_\theta(A) dQ(\theta), \quad A \in \mathcal{F},$$

а условное распределение $Q_x(\cdot)$ случайной величины Ξ при условии $X = x$ называется апостериорным распределением *(в отличие от априорного распределения $Q(\cdot)$) и имеет вид*

$$Q_x(B) = \frac{\int\limits_B p_\theta(x)\, dQ(\theta)}{\int\limits_\Theta p_\theta(x)\, dQ(\theta)}, \quad B \in \mathcal{V}.$$

Тогда $\delta_Q(X)$ есть байесовская оценка.

Доказательство. Пусть $\delta(X)$ – любая оценка с конечным риском. Тогда выражение (9.3.1) почти всюду конечно, поскольку функция потерь $L(\theta,\delta)$ неотрицательна. Поэтому почти всюду справедливо неравенство

$$\mathsf{E}(L(\Xi,\delta(x)) \mid X=x) \ \geq \ \mathsf{E}(L(\Xi,\delta_Q(x)) \mid X=x)$$

и результат следует после взятия математических ожиданий от обеих частей этого неравенства. □

Следствие 9.3.1. *Пусть выполнены условия Теоремы 9.3.1. Тогда*

1) *Если*
$$L(\theta,\delta) = (\delta - g(\theta))^2,$$
то
$$\delta_Q(x) = \mathsf{E}(g(\Xi) \mid X=x),$$
и, более общим образом, если
$$L(\theta,\delta) = c(\theta)(\delta - g(\theta))^2,$$
то
$$\delta_Q(x) = \frac{\mathsf{E}(c(\Xi)g(\Xi) \mid X=x)}{\mathsf{E}(c(\Xi) \mid X=x)}.$$

2) *Если*
$$L(\theta,\delta) = |\delta - g(\theta)|,$$
то $\delta_Q(x)$ есть любая медиана условного распределения Ξ при данном $X=x$.

3) *Если*
$$L(\theta,\delta) = \begin{cases} 0, & \text{если} \quad |\theta-\delta| \leq c, \\ 1, & \text{если} \quad |\theta-\delta| > c, \end{cases}$$
то $\delta_Q(x)$ есть середина интервала J длины $2c$, который максимизирует вероятность вида
$$\mathsf{P}(\Xi \in J \mid X=x).$$

Доказанная Теорема означает, что при нахождении байесовских оценок можно поступить следующим образом: сначала до проведения наблюдений, когда Ξ имеет распределение Q, найти байесовскую оценку δ_Q для функции $g(\theta)$, которая минимизирует по $\delta \in \Delta$ выражение $\mathsf{E}L(\Xi,\delta)$. Далее, после наблюдения $X = x$, априорное распределение Q случайной величины Ξ заменить на *апостериорное распределение* Q_x, то есть на условное распределение Ξ при данном $X = x$. Теперь байесовская оценка имеет вид $\delta_Q(x) = \delta_{Q_x}$.

Примеры.

1) Пусть

$$\mathcal{X} = \{0,1\}, \quad \Delta = \Theta = \{1/2, 1/3\}, \quad p_\theta(x) = \theta^x(1-\theta)^{1-x}, \quad x \in \mathcal{X}, \ \theta \in \Theta,$$

то есть наблюдение X принимает только два значения 0 и 1 соответственно с вероятностями $1-\theta$ и θ. Постороим байесовскую оценку $\delta_Q(x)$ параметра $\theta \in \Theta$, соответствующую априорному распределению Q вида

$$Q = \{\alpha, \ 1-\alpha\}, \quad \alpha \in (0,1),$$

и функции потерь

$$L(\theta, \delta) = \begin{cases} 0, & \text{если} \quad \theta = \delta, \\ 1, & \text{если} \quad \theta \neq \delta. \end{cases}$$

Тогда

$$\mathsf{E}L(\Xi, \delta) = L(1/2, \delta)\alpha + L(1/3, \delta)(1-\alpha) =$$
$$= \begin{cases} 1-\alpha, & \text{если} \quad \delta = 1/2, \\ \alpha, & \text{если} \quad \delta = 1/3. \end{cases}$$

Отсюда следует, что

$$\delta_Q = \begin{cases} 1/2, & \text{если} \quad \alpha > 1/2, \\ \forall, & \text{если} \quad \alpha = 1/2, \\ 1/3, & \text{если} \quad \alpha < 1/2. \end{cases}$$

Апостериорное распределение имеет вид

$$Q_x = \left\{ \frac{\alpha/2}{\alpha/2 + 2^{1-x}(1-\alpha)/3}, \quad \frac{2^{1-x}(1-\alpha)/3}{\alpha/2 + 2^{1-x}(1-\alpha)/3} \right\}.$$

Таким образом байесовская оценка есть

$$\delta_Q(x) = \delta_{Q_x} = \begin{cases} 1/2, & \text{если} \quad x > \log_2 \frac{4(1-\alpha)}{3\alpha}, \\ \forall, & \text{если} \quad x = \log_2 \frac{4(1-\alpha)}{3\alpha}, \\ 1/3, & \text{если} \quad x < \log_2 \frac{4(1-\alpha)}{3\alpha}. \end{cases}$$

2) Пусть $X = (X_1, \cdots, X_n)$, где X_i – независимые одинаково нормально распределённые наблюдения
$$X_i \sim \mathcal{N}(\theta,\, \sigma^2), \quad i = 1, \cdots, n,$$
с известной дисперсией σ^2. Построим байесовскую оценку $\delta_Q(X)$ параметра θ, соответствующую нормальному априорному распределению Q
$$\Xi \sim \mathcal{N}(\mu,\, \gamma^2).$$
Совместная плотность $X = (X_1, \cdots, X_n)$ и Ξ пропорциональна выражению
$$p(\theta, x) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \theta)^2\right\}\exp\Big\{-\frac{1}{2\gamma^2}(\theta - \mu)^2\Big\}.$$
Чтобы получить апостериорное распределение $\Xi \mid X = x$, необходимо совместную плотность Ξ, X разделить на маргинальную плотность X, поэтому апостериорное распределение имеет вид $C(x)p(\theta, x)$, что можно записать в виде
$$\bar{C}(x)\exp\left\{-\frac{\theta^2}{2}\Big(\frac{n}{\sigma^2} + \frac{1}{\gamma^2}\Big) + \theta\Big(\frac{n\bar{x}}{\sigma^2} + \frac{\mu}{\gamma^2}\Big) - \frac{\mu^2}{2\gamma^2}\right\} =$$
$$= \tilde{C}(x)\exp\left\{-\frac{1}{2}\Big(\frac{n}{\sigma^2} + \frac{1}{\gamma^2}\Big)\Big(\theta - \frac{n\bar{x}/\sigma^2 + \mu/\gamma^2}{n/\sigma^2 + 1/\gamma^2}\Big)^2\right\}, \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$
Это выражение представляет собой *нормальную плотность* с параметрами
$$\mathsf{E}(\Xi \mid X = x) = \frac{n\bar{x}/\sigma^2 + \mu/\gamma^2}{n/\sigma^2 + 1/\gamma^2}, \quad \mathsf{D}(\Xi \mid X = x) = \frac{1}{n/\sigma^2 + 1/\gamma^2}.$$
Таким образом, если функция потерь есть квадратичная ошибка, то байесовская оценка для θ есть
$$\delta_Q(X) = \frac{n/\sigma^2}{n/\sigma^2 + 1/\gamma^2}\overline{X} + \frac{1/\gamma^2}{n/\sigma^2 + 1/\gamma^2}\mu, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{9.3.2}$$

Как и в случае биномиального рапределния, возникает вопрос, является ли $\overline{X}$ байесовской оценкой для *некоторого* априорного распределения ? Ответ даётся следующей Теоремой.

Теорема 9.3.2. *Пусть Ξ имеет распределение Q и пусть P_θ обозначает условное распределение X при данном $\Xi = \theta$. Тогда ни одна несмещенная оценка $\delta(X)$ параметрической функции $g(\theta)$, при квадратичной функции потерь,* НЕ МОЖЕТ *быть байесовской, за исключением случая, когда*

$$\mathsf{E}(\delta(X) - g(\Xi))^2 = 0,$$

здесь математическое ожидание берется относительно совместного распределения X и Ξ.

Доказательство. Пусть $\delta(X)$ есть байесовская оценка и предположим, что она несмещённа для $g(\theta)$. Тогда в силу Теоремы 9.3.1 она почти всюду имеет вид

$$\delta(X) = \mathsf{E}(g(\Xi) \mid X), \quad \mathsf{E}(\delta(X) \mid \Xi = \theta) = \mathsf{E}_\theta \delta(X) = g(\theta).$$

Используя свойства условного математического ожидания отсюда следует, что

$$\mathsf{E}(g(\Xi)\delta(X)) = \mathsf{E}[\delta(X)\mathsf{E}(g(\Xi) \mid X)] = \mathsf{E}\delta^2(X)$$

и

$$\mathsf{E}(g(\Xi)\delta(X)) = \mathsf{E}[g(\Xi)\mathsf{E}(\delta(X) \mid \Xi)] = \mathsf{E}g^2(\Xi).$$

Поэтому

$$\mathsf{E}(\delta(X) - g(\Xi))^2 = \mathsf{E}\delta^2(X) + \mathsf{E}g^2(\Xi) - 2\mathsf{E}(\delta(X)g(\Xi)) = 0.$$

□

Применим теперь этот результат к нормальному и биномиальному случаям.

1) *Нормальный случай.* Пусть $X = (X_1, \cdots, X_n)$, где X_i – независимые одинаково нормально распределённые наблюдения

$$X_i \sim \mathcal{N}(\theta, \sigma^2), \quad i = 1, \cdots, n,$$

с известной дисперсией σ^2. Тогда для рассматриваемой оценки $\delta(X) = \overline{X}$ при $\Xi = \theta \in \Theta$ справедливо тождество

$$\mathsf{E}_\theta(\overline{X} - \theta)^2 \equiv \frac{\sigma^2}{n},$$

поэтому для любого априорного рапсределния Q

$$\mathsf{E}(\overline{X} - \Xi)^2 = \frac{\sigma^2}{n} \neq 0.$$

Таким образом $\delta(X) = \overline{X}$ *не является* байесовской оценкой.

2) *Биномиальный случай.* Пусть наблюдения X имеют вид

$$X \sim \mathcal{B}(n, \theta), \quad \theta \in (0,1).$$

Рассмотрим оценку $\delta(X) = \overline{X} = X/n$ для параметра θ. При фиксированном $\Xi = \theta \in (0,1)$ её функция риска равна

$$\mathsf{E}_\theta(\overline{X} - \theta)^2 = \frac{\theta(1-\theta)}{n},$$

поэтому

$$\mathsf{E}(X/n - \Xi)^2 = \frac{1}{n}\int\limits_0^1 \theta(1-\theta)dQ(\theta).$$

Этот интеграл равен нулю тогда и только тогда, когда распределение Q приписывает вероятность единица множеству $\{0,1\}$. Но байесовская оценка имеет вид

$$\delta_Q(X) = \mathsf{E}(\Xi \mid X),$$

поэтому для такого распределения Q

$$\delta_Q(0) = \delta_Q(n) = 1$$

и любая оценка, удовлетворяющая этому условию является байесовской для такого Q. Значит, в частности, X/n есть байесовская оценка. Конечно, если Q – истинное распределение, то значения $1, 2, \cdots, n-1$ никогда не наблюдаются. Таким образом, оценка X/n будет байесовской только в довольно тривиальном случае.

9.4 МИНИМАКСНОЕ ОЦЕНИВАНИЕ

Определение 9.4.1.*Минимаксной оценкой параметрической функции $g(\theta)$ называется оценка $\delta_*(x)$ такая, что*

$$\sup_{\theta\in\Theta} R(\theta, \delta_*) = \inf_\delta \sup_{\theta\in\Theta} R(\theta,\delta),$$

$$R(\theta,\delta) = \mathsf{E}_\theta L(\theta, \delta(X)) = \int\limits_{\mathcal{X}} L(\theta, \delta(x)) p_\theta(x) d\nu(x).$$

Таким образом, минимаксный подход заключается в выборе такой оценки, которая *минимизирует максимальный риск.* Можно также сказать, что минимаксная оценка является байесовской оценкой при априорном распределении, являющемся наименее благоприятным. Чтобы сделать это понятие точным, обозначим байесовский риск байесовской оценки $\delta_Q(X)$ через

$$r(Q) = r(\delta_Q, Q) = \int\limits_{\Theta} R(\theta, \delta_Q) dQ(\theta).$$

Априорное распределение Q_* называется *наименее благоприятным*, если

$$r(Q) \ \leq \ r(Q_*)$$

для всех априорных распределений Q на Θ. С байесовской точки зрения это есть априорное распределение, которое причиняет статистику наибольшие средние потери. Следующая Теорема даёт условия, при которых байесовская оценка $\delta_Q(X)$ является минимаксной.

Теорема 9.4.1.*Пусть существует априорное распределение Q на Θ такое, что*

$$\int\limits_{\Theta} R(\theta, \delta_Q) dQ(\theta) = \sup_{\theta \in \Theta} R(\theta, \delta_Q).$$

Тогда

1) *Оценка $\delta_Q(X)$ является минимаксной.*

2) *Если оценка $\delta_Q(X)$ является единственной байесовской оценкой, то оценка $\delta_Q(X)$ – единственная минимаксная оценка.*

3) *Априорное распределение Q является наименее благоприятным распределением.*

Доказательство.

1) Пусть $\delta(X)$ – любая другая оценка. Тогда

$$\sup_{\theta \in \Theta} R(\theta, \delta) \ \geq \ \int\limits_{\Theta} R(\theta, \delta) dQ(\theta) \ \geq \ \int\limits_{\Theta} R(\theta, \delta_Q) dQ(\theta) = \sup_{\theta \in \Theta} R(\theta, \delta_Q).$$

Таким образом

$$\inf_{\delta} \sup_{\theta \in \Theta} R(\theta, \delta) = \sup_{\theta \in \Theta} R(\theta, \delta_Q).$$

2) Пусть оценка $\delta_*(X) \neq \delta_Q(X)$ является минимаксной оценкой, тогда

$$\sup_{\theta\in\Theta} R(\theta,\delta_*) \;\geq\; \int\limits_{\Theta} R(\theta,\delta_*)dQ(\theta) > \int\limits_{\Theta} R(\theta,\delta_Q)dQ(\theta) = \sup_{\theta\in\Theta} R(\theta,\delta_Q).$$

Что противоречит минимаксности оценки $\delta_*(X)$.

3) Пусть $\bar{Q}$ – любое другое априорное распределение на Θ. Тогда

$$r(\bar{Q}) = \int\limits_{\Theta} R(\theta,\delta_{\bar{Q}})d\bar{Q}(\theta) \;\leq\; \int\limits_{\Theta} R(\theta,\delta_Q)d\bar{Q}(\theta) \;\leq\; \sup_{\theta\in\Theta} R(\theta,\delta_Q) = r(Q).$$

□

Условие Теоремы утверждает, что усреднённый риск $R(\theta,\delta_Q)$ равняется его *максимуму.* Это выполняется в том случае, когда функция риска *постоянна* или, более общим образом, когда априорное распределение Q приписывает вероятность 1 множеству, на котором функция риска достигает своего максимального значения. Более формальные утверждения содержит следующее Следствие.

Следствие 9.4.1.*Пусть существует априорное распределение Q на Θ такое, что*

1) байесовская оценка $\delta_Q(X)$ имеет постоянный риск. Тогда она является минимаксной.

2) для байесовской оценки $\delta_Q(X)$ справедливо соотношение

$$Q(\Xi \in A_Q) = 1, \qquad A_Q = \{\theta :\; R(\theta,\delta_Q) = \sup_{\bar{\theta}\in\Theta} R(\bar{\theta},\delta_Q)\}.$$

Тогда оценка $\delta_Q(X)$ является минимаксной оценкой.

Предположение Теоремы 9.4.1 влечёт существование наименее благоприятного распределения Q_*. Когда такого распределения не существует, Теорема 9.4.1 неприменима. Рассмотрим, например, задачу оценивания среднего θ нормального распределения с известной дисперсией. Поскольку все возможные значения θ играют полностью симметричную роль в том смысле, что ни одно из них не оценивается легче, чем любое другое, естественно предположить, что наименее благоприятное распределение есть "равномерное" распределение на действительной прямой, то есть мера Лебега. В этом случае оно является *несобственным.* Можно попытаться аппроксимировать несобственное распределение *последовательностью собственных* *распределений*, например, меру Лебега *равномерными* распределениями на $(-n, n)$, $n = 1, 2, \cdots$, и обобщить понятие наименее благоприятного распределения до понятия *наименее благоприятной последовательности распределений*. Рассмотрим более подробно этот подход.

Пусть $\{Q_n\}$ – последовательность априорных распределений на Θ и $\delta_{Q_n}(X)$ – байесовская оценка, соответствующая Q_n. Пусть её байесовский риск равен

$$r_n(Q_n) = r_n(\delta_{Q_n}, Q_n) = \int\limits_{\Theta} R(\theta, \delta_{Q_n}) dQ_n(\theta)$$

и предположим, что существует предел

$$\operatorname*{Lim}_{n\to\infty} r_n(Q_n) = r. \tag{9.4.1}$$

Тогда последовательность априорных распределений $\{Q_n\}$ называется *наименее благоприятной*, если для любого априорного распределения Q справедливо неравенство

$$r(Q) \;\leq\; r.$$

Теорема 9.4.2. *Пусть существует последовательность априорных распределений $\{Q_n\}$ на Θ такая, что выполняется соотношение (9.4.1), и предположим, что $\delta(X)$ есть оценка такая, что*

$$\sup_{\theta\in\Theta} R(\theta, \delta) = r.$$

Тогда

1) *Оценка $\delta(X)$ является минимаксной.*

2) *Последовательность априорных распределений $\{Q_n\}$ является наименее благоприятной.*

Доказательство.

1) Пусть $\tilde{\delta}(X)$ – любая другая оценка. Тогда

$$\sup_{\theta\in\Theta} R(\theta, \tilde{\delta}) \;\geq\; \int\limits_{\Theta} R(\theta, \tilde{\delta}) dQ_n(\theta) \;\geq\; r_n(Q_n),$$

и это выполняется при каждом n. Следовательно

$$\sup_{\theta\in\Theta} R(\theta, \tilde{\delta}) \;\geq\; \sup_{\theta\in\Theta} R(\theta, \delta)$$

и значит $\delta(X)$ есть минимаксная оценка.

2) Пусть $\bar{Q}$ – любое другое априорное распределение на Θ. Тогда

$$r(\bar{Q}) = \int\limits_{\Theta} R(\theta, \delta_{\bar{Q}}) d\bar{Q}(\theta) \ \leq \ \int\limits_{\Theta} R(\theta, \delta) d\bar{Q}(\theta) \ \leq \ \sup_{\theta \in \Theta} R(\theta, \delta) = r.$$

□

Эта Теорема менее удовлетворительна, чем Теорема 9.4.1, в двух отношениях. Во-первых, если даже байесовские оценки $\delta_{Q_n}(X)$ единственны, то отсюда невозможно заключить, что $\delta(X)$ есть единственная минимаксная оценка. Причина этого в том, что при переходе к пределу строгое неравенство заменяется нестрогим. Другая сложность состоит в том, что для того, чтобы проверить условие Теоремы 9.4.2, необходимо вычислить r и, следовательно, байесовские риски $r_n(Q_n)$. Для этого часто бывает полезна следующая Теорема.

Теорема 9.4.3.*Если $\delta_Q(X)$ – байесовская оценка для функции $g(\theta)$, соответствующая априорному распределению Q и если ее байесовский риск есть*

$$r(Q) = \mathsf{E}(\delta_Q(X) - g(\Xi))^2,$$

здесь Ξ и X имеют соответственно распределения $Q(\cdot)$ и

$$\mathsf{P}(A) = \int\limits_{\Theta} \mathsf{P}_\theta(A) dQ(\theta), \quad A \in \mathcal{F},$$

то

$$r(Q) = \int\limits_{\mathcal{X}} \mathsf{D}(g(\Xi) \mid X = x) d\mathsf{P}(x).$$

В частности, если апостериорная дисперсия $\mathsf{D}(g(\Xi) \mid X = x)$ не зависит от x, то

$$r(Q) = \mathsf{D}(g(\Xi) \mid X = x).$$

Доказательство. Доказательство следует из соотношения

$$r(Q) = \mathsf{E}(\delta_Q(X) - g(\Xi))^2 = \mathsf{E}[\mathsf{E}((g(\Xi) - \delta_Q(x))^2 \mid X = x)],$$

и Следствия 9.3.1, согласно которому

$$\delta_Q(x) = \mathsf{E}(g(\Xi) \mid X = x).$$

□

Пример 9.4.1. Пусть $X = (X_1, \cdots, X_n)$, где X_i – независимые одинаково нормально распределённые наблюдения

$$X_i \sim \mathcal{N}(\theta, \sigma^2), \quad i = 1, \cdots, n,$$

с известной дисперсией σ^2. Докажем, что $\overline{X}$ – минимаксная оценка. Рассмотрим в качестве априорных распределений нормальные распределения Q_γ вида

$$\Xi \sim \mathcal{N}(\mu, \gamma^2).$$

Тогда соотношение (9.3.2) показывает, что байесовская оценка имеет вид

$$\delta_{Q_\gamma}(X) = \frac{n/\sigma^2}{n/\sigma^2 + 1/\gamma^2}\overline{X} + \frac{1/\gamma^2}{n/\sigma^2 + 1/\gamma^2}\mu, \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Тамже найдена и апостериорная дисперсия

$$\mathsf{D}(\Xi \mid X = x) = \frac{1}{n/\sigma^2 + 1/\gamma^2},$$

которая не зависит от x. Поэтому из Теоремы 9.4.3 следует, что

$$r(Q_\gamma) = \frac{1}{n/\sigma^2 + 1/\gamma^2}.$$

Пусть $\gamma \to \infty$, тогда

$$r(Q_\gamma) \uparrow \sigma^2/n = r = \mathsf{D}_\theta \overline{X},$$

теперь используя Теорему 9.4.2, получаем минимаксность оценки $\overline{X}$.

9.5 СПИСОК ЛИТЕРАТУРЫ

1) Э. Леман, Теория Точечного Оценивания,
Москва, Наука, 1991, Глава 1, ¡1, ¡6; Глава 2, ¡1; Глава 4, ¡1, ¡2.

2) А. Вальд, Статистические решающие функции,
Позиционные Игры, Москва, Наука, 1967, стр. 300–522.

3) Г.И. Ивченко, Ю.И. Медведев, Математическая Статистика,
Москва, Высшая Школа, 1992, Глава 2, ¡2.1, ¡2.3.

4) Ш. Закс, Теория Статистических Выводов,
Москва, Мир, 1975, Глава 3, ¡3.1, ¡3.2; Глава 6, ¡6.1 – 6.5.

5) И.А. Ибрагимов, Р.З. Хасьминиский, Асимптотическая Теория оценивания,
Москва, Наука, 1979, Глава 1, ¡3.

6) А.Н. Ширяев, Вероятность,
Москва, Наука, 1989.

Лекция 10

В Лекции рассматриваются методы построения оптимальных оценок.

10.1 ПОЛНЫЕ ДОСТАТОЧНЫЕ СТАТИСТИКИ. МЕТОДЫ НАХОЖДЕНИЯ ОПТИМАЛЬНЫХ ОЦЕНОК

Теорема Рао – Блекуэлла – Колмогорова (Теорема 9.1.1) показывает, что оптимальные оценки нужно искать среди функций от достаточной статистики. При отыскании явного вида оптимальных оценок важную роль играет свойство *полноты* достаточных статистик.

Определение 10.1.1. *Достаточная статистика T называется полной, если для любой измеримой функции $\phi(t)$ выполнение тождества*

$$\mathsf{E}_\theta \phi(T) \equiv 0, \qquad \text{для всех} \qquad \theta \in \Theta$$

влечет равенство функции $\phi(t)$ нулю почти всюду, то есть

$$\mathsf{P}_\theta\Big(\phi(T) \neq 0\Big) = 0, \qquad \text{для всех} \qquad \theta \in \Theta.$$

Пример 10.1.1. Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения, причём

$$X_i \sim \mathcal{B}(1, \theta), \quad i = 1, \cdots, n, \quad \theta \in \Theta = (0, 1).$$

Совместная плотность наблюдений $X = (X_1, \cdots, X_n)$ имеет вид

$$p_\theta(x) = \theta^{\bar{x}}(1-\theta)^{n-\bar{x}},$$

где

$$x = (x_1, \cdots, x_n), \quad x_i \in \{0, 1\}, \quad i = 1, ..., n; \quad \bar{x} = \sum_{i=1}^{n} x_i.$$

Теперь из критерия факторизации (Теорема 7.1.3) следует, что статистика вида

$$T = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n, \theta)$$

является достаточной статистикой. Докажем, что эта достаточная статистика является *полной*. С этой целью предположим, что для функции $\phi(t)$ выполнено тождество

$$\mathsf{E}_\theta \phi(T) \equiv 0, \qquad \text{для всех} \qquad \theta \in (0, 1),$$

то есть

$$\sum_{k=0}^{n} \phi(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} \equiv 0, \qquad \text{для всех} \qquad \theta \in (0, 1).$$

В левой части этого тождества произведём замену переменной

$$u = \frac{\theta}{1-\theta} \in (0, +\infty),$$

тогда, получим

$$\sum_{k=0}^{n} \phi(k) \binom{n}{k} u^k \equiv 0, \qquad \text{для всех} \qquad u \in (0, +\infty).$$

В левой части здесь стоит *полином* по u степени n, который имеет бесконечно много корней, поэтому применяя основную Теорему алгебры, получаем, что он тождественно равен нулю. Таким образом

$$\phi(k) = 0, \quad k = 0, \cdots, n$$

или

$$\mathsf{P}_\theta\Big(\phi(T) \neq 0\Big) = 0, \qquad \text{для всех} \qquad \theta \in \Theta.$$

Приведём пример *не полной* достаточной статистики.

Пример 10.1.2. Покажем, что типичным образом вся выборка $X = (X_1, \cdots, X_n)$ *не является* полной достаточной статистикой. Пусть $X =$

$(X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения, имеющие плотность, зависящую от параметра $\theta \in \Theta$, причём

$$\mathsf{E}_\theta X_1 < \infty, \qquad \text{для всех} \quad \theta \in \Theta.$$

Из критерия факторизации (Теорема 7.1.3) непосредственно следует, что статистика вида

$$T(X) = X = (X_1, \cdots, X_n)$$

является достаточной статистикой. Но она не полна, поскольку, например, для функции не равной почти всюду нулю

$$\phi(x) = \frac{1}{n}\sum_{i=1}^{n} x_i - x_1, \qquad x = (x_1, \cdots, x_n)$$

справедливо тождество

$$\mathsf{E}_\theta \phi(T) \equiv 0, \qquad \text{для всех} \quad \theta \in \Theta.$$

Замечание 10.1.1. *Если $\mathcal{P}_0$ и $\mathcal{P}_1$ – два семейства распределений, таких что каждое $\mathcal{P}_0$ – нулевое множество (см. Определение 6.1.4) является и $\mathcal{P}_1$ – нулевым, тогда достаточная статистика $T(X)$, полная для семейства $\mathcal{P}_0$ будет также полной и для семейства $\mathcal{P}_1$. Заметим, также что если $\mathcal{P}_0$ есть семейство биномиальных распределений*

$$\mathcal{P}_0 = \{\mathcal{B}(n,\, \theta),\, \theta \in (0,1)\},$$

n фиксировано и, если

$$\mathcal{P}_1 = \mathcal{P}_0 \cup \mathcal{P}(1),$$

где $\mathcal{P}(1)$ есть распределение Пуассона с параметром 1, то по доказанному выше семейство $\mathcal{P}_0$ является полным (Пример 10.1.1), в то время как семейство $\mathcal{P}_1$ не полно.

Наличие полной достаточной статистики обеспечивает *единственность и оптимальность* несмещённой оценки, зависящей от такой статистики.

Теорема 10.1.1.*Пусть $= T(X)$ – полная достаточная статистика, тогда*

1) *Если $\delta_1(T)$ и $\delta_2(T)$ – две несмещенные оценки для функции $g(\theta)$, то они совпадают почти всюду, то есть*

$$\mathsf{P}_\theta\Big(\delta_1(T) \neq \delta_2(T)\Big) = 0, \qquad \textit{для всех} \quad \theta \in \Theta.$$

Если $\delta_1(X)$ и $\delta_2(X)$ – две несмещенные оценки для функции $g(\theta)$, то почти всюду справедливо равенство

$$\mathsf{E}(\delta_1(X) \mid T) = \mathsf{E}(\delta_2(X) \mid T).$$

2) *Если $\delta^*(T)$ – несмещенная оценка функции $g(\theta)$, то $\delta^*(T)$ оптимальна, то есть для любой несмещенной оценки $\delta(X)$ функции $g(\theta)$ справедливо неравенство*

$$\mathsf{D}_\theta \delta^*(T) \leq \mathsf{D}_\theta \delta(X), \qquad \text{для всех} \qquad \theta \in \Theta.$$

Доказательство.

1) Если $\delta_1(T)$ и $\delta_2(T)$ – две несмещённые оценки функции $g(\theta)$, то их разность
$$\phi(T) = \delta_1(T) - \delta_2(T)$$
удовлетворяет тождеству
$$\mathsf{E}_\theta \phi(T) \equiv 0, \qquad \text{для всех} \qquad \theta \in \Theta,$$
которое в силу полноты статистики T влечёт
$$\mathsf{P}_\theta\Big(\delta_1(T) \neq \delta_2(T)\Big) = 0, \qquad \text{для всех} \qquad \theta \in \Theta.$$
Второе утверждение непосредственно следует из условия несмещённости
$$\mathsf{E}_\theta \delta_i(X) \equiv g(\theta), \quad \theta \in \Theta, \quad i = 1, 2,$$
определения полноты и тождества
$$0 \equiv \mathsf{E}_\theta(\delta_1(X) - \delta_2(X)) \equiv \mathsf{E}_\theta \mathsf{E}\Big((\delta_1(X) - \delta_2(X)) \mid T\Big).$$

2) Пусть $\delta(X)$ – любая несмещённая оценка функции $g(\theta)$. Тогда по Теореме Рао – Блекуэлла – Колмогорова (Теорема 9.1.1) проекция оценки $\delta(X)$ на достаточную статистику T
$$h(T) = \mathsf{E}_\theta(\delta(X) \mid T)$$
является несмещённой оценкой функции $g(\theta)$ и выполнено неравенство
$$\mathsf{D}_\theta h(T) \leq \mathsf{D}_\theta \delta(X), \qquad \text{для всех} \qquad \theta \in \Theta.$$

Однако, эта оценка $h(T)$, будучи несмещённой, является единственной в силу пункта 1, то есть не зависит от оценки $\delta(X)$ и, поэтому последнее неравенство справедливо для любой несмещённой оценки $\delta(X)$. Следовательно оценка $h(T)$ оптимальна и

$$\mathsf{P}_\theta\Big(\delta^*(T) = h(T)\Big) = 1, \qquad \text{для всех} \qquad \theta \in \Theta.$$

□

Следствия.

1) Если существует полная достаточная статистика, то любая измеримая функция от неё является оптимальной оценкой своего математического ожидания.

2) Существует единственная несмещённая оценка функции $g(\theta)$, зависящая от полной достаточной статистики, и она оптимальна. Если $T = T(X)$ – полная достаточная статистика, то оптимальная оценка $\delta^*(T)$ любой параметрической функции $g(\theta)$, допускающей несмещённую оценку, однозначно определяется совокупностью уравнений

$$\mathsf{E}_\theta \delta^*(T) = g(\theta), \qquad \text{для всех} \qquad \theta \in \Theta.$$

3) Алгоритм получения оптимальных оценок.

   Для нахождения оптимальной оценки $\delta^*(T)$ функции $g(\theta)$ достаточно поступить следующим образом

   (a) найти *какую-нибудь несмещенную* оценку $\delta(X)$ функции $g(\theta)$;

   (b) спроектировать её на полную достаточную статистику T, то есть найти
   $$h(T) = \mathsf{E}_\theta(\delta(X) \mid T),$$
   тогда это и будет оптимальной оценкой, то есть
   $$\mathsf{P}_\theta\Big(\delta^*(T) = h(T)\Big) = 1, \qquad \text{для всех} \qquad \theta \in \Theta.$$

Пример 10.1.3. Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения и

$$X_i \sim \mathcal{B}(1, \theta), \quad i = 1, \cdots, n, \quad \theta \in \Theta = (0, 1).$$

Тогда

$$T = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n,\, \theta)$$

является *полной достаточной статистикой* (см. Пример 10.1.1). Найдём двумя способами оптимальную оценку, например, для функции $g(\theta) = \theta^2$.

1) Попытаемся найти оптимальную оценку $\delta^*(T)$ из условия несмещённости

$$\mathsf{E}_\theta \delta^*(T) = \theta^2, \qquad \text{для всех} \qquad \theta \in \Theta.$$

С этой целью заметим, что

$$\mathsf{E}_\theta T = n\theta, \quad \mathsf{D}_\theta T = n\theta(1-\theta), \quad \mathsf{E}_\theta T^2 = \mathsf{D}_\theta T + (\mathsf{E}_\theta T)^2 = n\theta(1-\theta) + n^2\theta^2.$$

Будем искать оптимальную оценку $\delta^*(T)$ в виде полинома второй степени

$$\delta^*(t) = a + bt + ct^2,$$

тогда из условия несмещённости имеем

$$a + bn\theta + c\Big(n\theta(1-\theta) + n^2\theta^2\Big) \equiv \theta^2.$$

Приравнивая коэффициенты при одинаковых степенях θ, получим

$$c = \frac{1}{n(n-1)}, \quad b = -\frac{1}{n(n-1)}, \quad a = 0,$$

то есть

$$\delta^*(T) = -\frac{T}{n(n-1)} + \frac{T^2}{n(n-1)} = \frac{T(T-1)}{n(n-1)}.$$

2) Найдём $\delta^*(T)$ методом проекций. С этой целью возьмём произвольную несмещённую оценку θ^2, например, $\delta(X) = X_1X_2$. Тогда

$$\delta^*(t) = \mathsf{E}_\theta(\delta(X) \mid T = t) = \mathsf{E}_\theta(X_1X_2 \mid T = t) =$$

$$= \sum_{x_1,x_2 \in \{0,1\}\times\{0,1\}} x_1x_2 \mathsf{P}_\theta\Big(X_1 = x_1, X_2 = x_2 \mid T = t\Big) = \mathsf{P}_\theta\Big(X_1 = 1, X_2 = 1 \mid T = t\Big) =$$

$$= \frac{\mathsf{P}_\theta\Big(X_1 = 1,\ X_2 = 1,\ T = t\Big)}{\mathsf{P}_\theta(T = t)} = \begin{cases} 0, & t < 2 \\ \frac{\theta^2\binom{n-2}{t-2}\theta^{t-2}(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}}, & t \geq 2 \end{cases} =$$

$$= \begin{cases} 0, & t < 2 \\ \frac{t!(n-t)!(n-2)!}{n!(t-2)!(n-t)!}, & t \geq 2 \end{cases} = \begin{cases} 0, & t < 2 \\ \frac{t(t-1)}{n(n-1)}, & t \geq 2. \end{cases}$$

Таким образом опять получаем

$$\delta^*(T) = \frac{T(T-1)}{n(n-1)}.$$

10.2 СВОБОДНЫЕ СТАТИСТИКИ

Выше отмечалось, что достаточные статистики, используются для сокращения данных без потери информации, в этой связи уместно рассмотреть случай, когда статистики вообще не несут в себе никакой информации о параметре θ.

Определение 10.2.1. *Статистика $U = U(X)$*

$$U : (\mathcal{X}, \mathcal{F}) \longrightarrow (\mathcal{Y}, \mathcal{H})$$

называется свободной (подчиненной), если ее распределение не зависит от $\theta \in \Theta$, то есть, если

$$\mathsf{P}_\theta(U(X) \in B)$$

для любого $B \in \mathcal{H}$ не зависит от $\theta \in \Theta$.

Ясно, что свободная статистика $U(X)$ не содержит информации о θ.

Примеры.

1) Пусть $X = (X_1, \cdots, X_n)$ – независимые нормально распределённые наблюдения

$$X_i \sim \mathcal{N}(0,\, \theta^2), \quad i = 1, \cdots, n.$$

Тогда статистики вида

$$U_1 = \frac{X_1}{\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2}}, \qquad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

и

$$U_2 = \frac{X_1}{\sqrt{\sum\limits_{i=1}^{n} {X_i}^2}}$$

являются свободными.

2) Пусть $X = (X_1, \cdots, X_n)$ – независимые нормально распределённые наблюдения
$$X_i \sim \mathcal{N}(\theta,\, 1), \qquad i = 1, \cdots, n.$$
Тогда статистики вида
$$U_3 = X_1 - \overline{X}$$
и
$$U_4 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$
являются также свободными.

Поскольку свободная статистика $U(X)$ не содержит информации о θ, а достаточная статистика (X) содержит всю информацию о θ, то, по-видимому, $U(X)$ и $T(X)$ должны быть независимыми. При наличии свойства *полноты* это действительно так.

Теорема 10.2.1. (Басу) *Пусть (X) – полная достаточная статистика, а $U(X)$ – свободная статистика. Тогда статистики $T(X)$ и $U(X)$ – независимы.*

Доказательство. По условию выражение
$$\mathsf{P}_\theta(U \in B \mid T) - \mathsf{P}_\theta(U \in B)$$
не зависит от θ для любого фикированного множества $B \in \mathcal{H}$. Обозначим его через
$$\phi(T) = \mathsf{P}_\theta(U \in B \mid T) - \mathsf{P}_\theta(U \in B).$$
Тогда
$$\mathsf{E}_\theta \phi(T) \equiv 0 \quad \text{для всех} \quad \theta \in \Theta,$$
то есть
$$\mathsf{P}_\theta(U \in B \mid T) = \mathsf{P}_\theta(U \in B) \qquad \text{п. в.}$$
и, поэтому они независимы, поскольку
$$\mathsf{P}_\theta(U \in B,\, T \in A) = \mathsf{E}_\theta \mathbf{1}_B(U)\mathbf{1}_A(T) =$$
$$= \mathsf{E}_\theta \mathsf{E}_\theta[\mathbf{1}_B(U)\mathbf{1}_A(T) \mid T] = \mathsf{E}_\theta \mathbf{1}_A(T)\mathsf{E}_\theta[\mathbf{1}_B(U) \mid T] =$$
$$= \mathsf{E}_\theta \mathbf{1}_A(T)\mathsf{P}_\theta(U \in B \mid T) = \mathsf{P}_\theta(T \in A)\mathsf{P}_\theta(U \in B).$$
□

Пример 10.2.1. Пусть $X = (X_1, \cdots, X_n)$ – независимые равномерно распределённые наблюдения

$$X_i \sim \mathcal{R}(0,\, \theta), \quad i = 1, \cdots, n; \quad \theta > 0.$$

Докажем, что статистика

$$T(X) = X_{(n)} \equiv \max_{1 \le i \le n} X_i$$

является полнай достаточной статистикой. Совместная плотность наблюдений $(X_1, \cdots, X_n)$ равна

$$p_\theta(x) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbf{1}_{(0,\theta)}(x_i) = \frac{1}{\theta^n} \mathbf{1}_{(0,\theta)}(x_{(1)}) \mathbf{1}_{(0,\theta)}(x_{(n)}) =$$

$$= \frac{1}{\theta^n} \mathbf{1}_{(0,+\infty)}(x_{(1)}) \mathbf{1}_{(0,\theta)}(x_{(n)}), \quad \text{где} \quad x = (x_1, \cdots, x_n), \quad x_{(1)} = \min_{1 \le i \le n} x_i.$$

Теперь из критерия факторизации (Теорема 7.1.3) следует, что статистика вида

$$T(X) = X_{(n)}$$

является достаточной статистикой. Докажем теперь её полноту. Поскольку

$$\mathsf{P}_\theta(T(X) < t) = \mathsf{P}_\theta(X_1 < t, \cdots, X_n < t) = \frac{t^n}{\theta^n}, \quad 0 < t < \theta,$$

то статистика $T(X)$ имеет плотность

$$p_\theta(t) = \frac{n t^{n-1}}{\theta^n}, \quad 0 < t < \theta.$$

Пусть теперь

$$\mathsf{E}_\theta \phi(T) \equiv 0, \quad \text{для всех} \quad \theta > 0.$$

Обозначим через $\phi^+(t)$ и $\phi^-(t)$ соответственно положительную и отрицательную части функции $\phi(t)$. Тогда

$$\int_0^\theta t^{n-1} \phi^+(t)\, dt = \int_0^\theta t^{n-1} \phi^-(t)\, dt, \quad \text{для всех} \quad \theta > 0.$$

Отсюда по Теореме о продолжении меры следует, что для *всех борелевских множеств* $B \subseteq \mathbf{R}^1$ справедливо равенство

$$\int_B t^{n-1} \phi^+(t)\, dt = \int_B t^{n-1} \phi^-(t)\, dt.$$

Поэтому

$$\phi(t) = 0, \quad \text{п. в.}$$

Ясно, что статистика

$$U(X) = \frac{X_1}{X_{(n)}}$$

является свободной. Из Теоремы Басу следует, что статистики $T(X)$ и $U(X)$ *независимы.*

Пример 10.2.2. Пусть $X = (X_1, \cdots, X_n)$ – независимые нормально распределённые наблюдения

$$X_i \sim \mathcal{N}(\mu,\, \sigma^2), \quad i = 1, \cdots, n; \quad \theta = (\mu,\, \sigma^2).$$

Можно доказать, что статистика

$$T(X) = (\overline{X}, S^2), \quad \text{где} \quad \overline{X} = \frac{1}{n}\sum X_i, \quad S^2 = \frac{1}{n-1}\sum (X_i - \overline{X})^2$$

является полной достаточной статистикой, а статистика

$$U(X) = \left(\frac{X_1 - \overline{X}}{S},\ \cdots,\ \frac{X_n - \overline{X}}{S}\right)$$

является свободной статистикой. Из Теоремы Басу следует, что статистики $T(X)$ и $U(X)$ *независимы.*

10.3 СПИСОК ЛИТЕРАТУРЫ

1) Э. Леман, Теория Точечного Оценивания,
Москва, Наука, 1991, Глава 1, ¡5; Глава 2, ¡1.

2) А. Вальд, Статистические решающие функции,
Позиционные Игры, Москва, Наука, 1967, стр. 300–522.

3) Г.И. Ивченко, Ю.И. Медведев, Математическая Статистика,
Москва, Высшая Школа, 1992, Глава 2, ¡2.3.

4) Ш. Закс, Теория Статистических Выводов,
Москва, Мир, 1975, Глава 2, ¡2.6.

5) Ю.В. Линник, Лекции о Задачах Аналитической Статистики,
Москва, Наука, 1994, Лекция 2.

Лекция 11

В Лекции рассматриваются нижние оценки для дисперсий оценок, вводятся так называемые эффективные оценки.

11.1 ИНФОРМАЦИОННОЕ НЕРАВЕНСТВО

В общем случае оптимальная оценка $\delta^*(X)$ параметрической функции $g(\theta)$ может не существовать (Теорема 10.1.1 даёт лишь *достаточные* условия существования $\delta^*(X)$). Однако при выполнении естественных условий регулярности можно получить оценку *снизу* для дисперсии *любой* оценки функции $g(\theta)$ и указать условия, при которых эта граница достигается.

Любые две случайные величины Y и Z с конечными вторыми моментами удовлетворяют *ковариационному неравенству*

$$\mathsf{Cov}(Y,\,Z)\;\leq\;\sqrt{\mathsf{D}Y\mathsf{D}Z}. \tag{11.1.1}$$

Доказательство непосредственно следует из определения ковариации и неравенства Коши – Буняковского.

Применим неравенство (11.1.1) к любой оценке $\delta(X)$ функции $g(\theta)$ и любой функции $S(X,\theta)$ с конечным вторым моментом и положительной дисперсией

$$\mathsf{D}_\theta\delta(X)\;\geq\;\frac{\mathsf{Cov}^2_\theta\Big(\delta(X),\;S(X,\theta)\Big)}{\mathsf{D}_\theta S(X,\theta)}. \tag{11.1.2}$$

В общем случае неравенство (11.1.2) бесполезно, поскольку его левая часть также содержит оценку $\delta(X)$. Но если

$$\mathsf{Cov}_\theta\Big(\delta(X),\;S(X,\theta)\Big)$$

зависит от оценки $\delta(X)$ только через её математическое ожидание

$$\mathsf{E}_\theta \delta(X) = g(\theta),$$

то неравенство (11.1.2) действительно даёт нижнюю границу для дисперсии *всех несмещенных* оценок функции $g(\theta)$.

Теорема 11.1.1. *Для того чтобы*

$$\mathsf{Cov}_\theta\Big(\delta(X),\ S(X,\theta)\Big)$$

зависела от оценки $\delta(X)$ только через ее математическое ожидание

$$\mathsf{E}_\theta \delta(X) = g(\theta),$$

необходимо и достаточно, чтобы для любой несмещенной оценки нуля $\delta_0(X)$ выполнялось тождество

$$\mathsf{Cov}_\theta\Big(\delta_0(X),\ S(X,\theta)\Big) \equiv 0, \qquad \textit{для всех} \qquad \theta \in \Theta. \tag{11.1.3}$$

Доказательство. *Необходимость.* Предположим, что

$$\mathsf{Cov}_\theta\Big(\delta(X),\ S(X,\theta)\Big)$$

зависит от оценки $\delta(X)$ только через $g(\theta)$. Тогда для любой несмещённой оценки нуля $\delta_0(X)$ справедливо тождество

$$\mathsf{Cov}_\theta\Big(\delta(X)+\delta_0(X),\ S(X,\theta)\Big) \equiv \mathsf{Cov}_\theta\Big(\delta(X),\ S(X,\theta)\Big), \qquad \text{для всех} \qquad \theta \in \Theta$$

и, следовательно,

$$\mathsf{Cov}_\theta\Big(\delta_0(X),\ S(X,\theta)\Big) \equiv 0, \qquad \text{для всех} \qquad \theta \in \Theta.$$

Достаточность. Пусть выполнено тождество (11.1.3) для всех несмещённых оценок нуля. Пусть $\delta_1(X)$ и $\delta_2(X)$ – две несмещённые оценки функции $g(\theta)$

$$\mathsf{E}_\theta \delta_1(X) \equiv \mathsf{E}_\theta \delta_2(X) \equiv g(\theta), \qquad \text{для всех} \qquad \theta \in \Theta.$$

Тогда оценка

$$\delta_1(X) - \delta_2(X)$$

является несмещённой оценкой нуля, и поэтому

$$\mathsf{Cov}_\theta\Big(\delta_1(X)-\delta_2(X),\ S(X,\theta)\Big) \equiv 0, \qquad \text{для всех} \qquad \theta \in \Theta,$$

так что

$$\mathsf{Cov}_\theta\Big(\delta_1(X),\ S(X,\theta)\Big) \equiv \mathsf{Cov}_\theta\Big(\delta_2(X),\ S(X,\theta)\Big), \qquad \text{для всех} \qquad \theta \in \Theta.$$

□

Применим Теорему 11.1.1 и неравенство (11.1.2) для получения оценки снизу для дисперсии произвольной несмещённой оценки функции $g(\theta)$.

Теорема 11.1.2. (Неравенство Хаммерсли – Чепмена – Роббинса) *Предположим, что плотность $p_\theta(x)$ удовлетворяет условию*

$$p_\theta(x) > 0, \qquad \textit{для всех} \qquad \theta \in \Theta \quad \textit{и} \qquad x \in \mathcal{X}.$$

Тогда для любой несмещнной оценки $\delta(X)$ функции $g(\theta)$ справедливо неравенство

$$\mathsf{D}_\theta\delta(X) \ \geq\ \frac{\Big(g(\theta+\Delta) - g(\theta)\Big)^2}{\mathsf{E}_\theta\Big(\frac{p_{\theta+\Delta}(X)}{p_\theta(X)} - 1\Big)^2}, \qquad \textit{для всех} \qquad \theta,\ \ \theta + \Delta \in \Theta. \qquad (11.1.4)$$

Замечание 11.1.1. Заметим, что

$$\mathsf{E}_\theta\left(\frac{p_{\theta+\Delta}(X)}{p_\theta(X)} - 1\right)^2 > 0, \qquad \text{для всех} \qquad \theta,\ \ \theta+\Delta \in \Theta,$$

поскольку в противном случае если существуют $\theta \in \Theta$ и $\theta + \Delta \in \Theta$ такие, что

$$\mathsf{P}_\theta\Big(p_\theta(X) = p_{\theta+\Delta}(X)\Big) = 1,$$

то

$$\mathsf{P}_\theta\Big(p_\theta(X) = p_{\theta+\Delta}(X)\Big) = \int\limits_{\{x:p_\theta(x)=p_{\theta+\Delta}(x)\}} p_\theta(x)d\nu(x) =$$

$$= \int\limits_{\{x:p_\theta(x)=p_{\theta+\Delta}(x)\}} p_{\theta+\Delta}(x)d\nu(x) = \mathsf{P}_{\theta+\delta}\Big(p_\theta(X) = p_{\theta+\Delta}(X)\Big) = 1,$$

$$\mathsf{P}_\theta(X \in A) = \mathsf{P}_\theta\Big(X \in A, p_\theta(X) = p_{\theta+\Delta}(X)\Big) = \int\limits_{A\cap\{x:p_\theta(x)=p_{\theta+\Delta}(x)\}} p_\theta(x)d\nu(x) =$$

$$= \int\limits_{A\cap\{x:p_\theta(x)=p_{\theta+\Delta}(x)\}} p_{\theta+\Delta}(x)d\nu(x) =$$

$$= \mathsf{P}_{\theta+\Delta}\Big(X \in A,\ p_\theta(X) = p_{\theta+\Delta}(X)\Big) = \mathsf{P}_{\theta+\Delta}(X \in A), \qquad A \in \mathcal{F}$$

и нарушается условие *идентифицируемости* (см. Лекцию 6) семейства $\mathcal{P}$.

Доказательство. Рассмотрим функцию

$$S(x,\theta) = \frac{p_{\theta+\Delta}(x)}{p_\theta(x)} - 1 \tag{11.1.5}$$

тогда она удовлетворяет условиям Теоремы 11.1.1, поскольку

$$\mathsf{E}_\theta S(X,\theta) \equiv 0, \qquad \text{для всех} \qquad \theta \in \Theta$$

и, следовательно

$$\mathsf{Cov}_\theta\Big(\delta_0(X),\ S(X,\theta)\Big) = \mathsf{E}_\theta \delta_0(X) S(X,\theta) =$$

$$= \mathsf{E}_{\theta+\Delta}\delta_0(X) - \mathsf{E}_\theta\delta_0(X) \equiv 0, \qquad \text{для всех} \qquad \theta, \theta+\Delta \in \Theta.$$

При этом

$$\mathsf{Cov}_\theta\Big(\delta(X),\ S(X,\theta)\Big) = \mathsf{E}_\theta \delta(X) S(X,\theta) = g(\theta+\Delta) - g(\theta).$$

Поэтому неравенство (11.1.2) принимает вид

$$\mathsf{D}_\theta \delta(X) \ \ge\ \frac{\Big(g(\theta+\Delta) - g(\theta)\Big)^2}{\mathsf{E}_\theta\Big(\frac{p_{\theta+\Delta}(X)}{p_\theta(X)} - 1\Big)^2}.$$

□

Замечания.

1) Заметим, что неравенство (11.1.4) имеет смысл, если $\theta \in \Theta$ и $\theta+\Delta \in \Theta$ таковы, что
$$g(\theta) \neq g(\theta+\Delta).$$

2) Поскольку неравенство (11.1.4) выполнено при *всех* Δ таких, что $\theta + \Delta \in \Theta$, то его можно записать в виде

$$\mathsf{D}_\theta \delta(X) \ \ge\ \sup_{\Delta:\theta+\Delta\in\Theta} \frac{\Big(g(\theta+\Delta) - g(\theta)\Big)^2}{\mathsf{E}_\theta\Big(\frac{p_{\theta+\Delta}(X)}{p_\theta(X)} - 1\Big)^2}, \qquad \text{для всех} \qquad \theta \in \Theta.$$

3) Условие

$$p_\theta(x) > 0, \qquad \text{для всех} \qquad \theta \in \Theta; \qquad x \in \mathcal{X}$$

можно несколько ослабить. Обозначим

$$A(\theta) = \{x \in \mathcal{X} : \, p_\theta(x) > 0\}.$$

Тогда неравенство (11.1.4) выполнено, если это условие заменить на

$$A(\theta + \Delta) \subseteq A(\theta).$$

При этом справедливо неравенство

$$\mathsf{D}_\theta \delta(X) \; \geq \; \sup_{\Delta \in B(\theta)} \frac{\Big(g(\theta+\Delta) - g(\theta)\Big)^2}{\mathsf{E}_\theta \Big(\frac{p_{\theta+\Delta}(X)}{p_\theta(X)} - 1\Big)^2},$$

где $\quad B(\theta) = \{\Delta : \theta + \Delta \in \Theta, \; A(\theta+\Delta) \subseteq A(\theta)\}; \quad$ для всех $\quad \theta \in \Theta.$

При выполнении некоторых условий реглярности (см. ниже Условие R), классическое *информационное неравенство* получается, если в неравенстве (11.1.4) устремить Δ к нулю. Неравенство (11.1.4) не изменится если функцию $S(x, \theta)$ заменить на выражение

$$\frac{p_{\theta+\Delta}(x) - p_\theta(x)}{\Delta} \cdot \frac{1}{p_\theta(x)},$$

которое стремится к

$$\frac{\partial p_\theta(x)}{\partial \theta} \cdot \frac{1}{p_\theta(x)}$$

при $\Delta \to 0$, если плотность $p_\theta(x)$ дифференцируема по $\theta \in \Theta$, а выражение

$$g(\theta) - g(\theta + \Delta)$$

заменить на отношение

$$\frac{g(\theta) - g(\theta+\Delta)}{\Delta},$$

которое стремится к $g'(\theta)$ при $\Delta \to 0$, если $g(\theta)$ дифференцируемая по $\theta \in \Theta$ функция.

Таким образом, кажется правдоподобным, что в качестве функции $S(x, \theta)$ в (11.1.5) можно рассмотреть выражение

$$S(x, \theta) = \frac{\partial p_\theta(x)}{\partial \theta} \cdot \frac{1}{p_\theta(x)}. \tag{11.1.6}$$

Поскольку для любой несмещённой оценки нуля $\delta_0(X)$ выполняется тождество (при условии дифференцируемости $\mathsf{E}_\theta\delta_0(X)$ по θ)

$$\frac{d\,\mathsf{E}_\theta\delta_0(X)}{d\,\theta} \equiv 0, \quad \text{для всех} \quad \theta \in \Theta,$$

то функция $S(x,\theta)$ будет удовлетворять соотношению (11.1.3) при условии, что выражение

$$\mathsf{E}_\theta\delta_0(X) = \int \delta_0(x)p_\theta(x)d\nu(x)$$

можно дифференцировать по $\theta \in \Theta$ под знаком интеграла при всех $\delta_0(X)$. Чтобы получить окончательную нижнюю границу дисперсии, положим

$$p'_\theta(x) = \frac{\partial p_\theta(x)}{\partial\theta},$$

тогда

$$\mathsf{Cov}_\theta\Big(\delta(X),\ S(X,\theta)\Big) = \int \delta(x)p'_\theta(x)d\nu(x).$$

Если в тождестве

$$\int \delta(x)p_\theta(x)d\nu(x) \equiv g(\theta)$$

допускается дифференцирование по $\theta \in \Theta$ под знаком интеграла, то отсюда следует, что

$$\mathsf{Cov}_\theta\Big(\delta(X),\ S(X,\theta)\Big) = g'(\theta),$$

и следовательно справедливо *информационное неравенство* (см. (11.1.2))

$$\mathsf{D}_\theta\delta(X)\ \geq\ \frac{\Big(g'(\theta)\Big)^2}{\mathsf{D}_\theta\Big(\frac{\partial\log p_\theta(X)}{\partial\theta}\Big)}, \quad \text{для всех} \quad \theta \in \Theta. \tag{11.1.7}$$

Предположения, при которых выполняется это неравенство, будут приведены в более формальном виде в Теореме 11.1.4.

Функция $S(x,\theta)$, определённая равенством (11.1.6), представляет собой относительную скорость изменения плотности $p_\theta(x)$ в точке $x \in \mathcal{X}$. Среднее значение квадрата этой скорости обозначим через $I(\theta)$.

Определение 11.1.1. *Величина*

$$I_X(\theta) \equiv I(\theta) = \mathsf{E}_\theta\left(\frac{\partial\log p_\theta(X)}{\partial\theta}\right)^2 = \int\left(\frac{p'_\theta(x)}{p_\theta(x)}\right)^2 p_\theta(x)d\nu(x)$$

называется информацией по Фишеру (фишеровской информацией), которая содержится в наблюдении X о параметре $\theta \in \Theta$.

Сформулируем теперь условия регулярности.

Условие R.

1) Параметрическое множество Θ является открытым множеством из $\mathbf{R}^1$.

2) Множество (*носитель* распределения P_θ)

$$A = \{x \in \mathcal{X} : \ p_\theta(x) > 0\}$$

не зависит от $\theta \in \Theta$.

3) Для всех x из A и всех θ из Θ функция $p_\theta(x)$ дифференцируема по θ и

$$\frac{d}{d\theta}\int p_\theta(x)d\nu(x) = \int\limits_A p'_\theta(x)d\nu(x) < \infty.$$

4) Для всех x из A и всех θ из Θ функция $p_\theta(x)$ дважды дифференцируема по θ и

$$\frac{d}{d\theta}\int\limits_A p'_\theta(x)d\nu(x) = \int\limits_A p''_\theta(x)d\nu(x) < \infty, \qquad p''_\theta(x) = \frac{\partial^2 p_\theta(x)}{\partial\theta^2}.$$

5) Для всех оценок $\delta(X)$, всех x из A и всех θ из Θ функция $p_\theta(x)$ дифференцируема по θ и

$$\frac{d}{d\theta}\int \delta(x)p_\theta(x)d\nu(x) = \int\limits_A \delta(x)p'_\theta(x)d\nu(x) < \infty.$$

6) Функция $g(\theta)$ дифференцируема по $\theta \in \Theta$.

Некоторые свойства фишеровской информации $I(\theta)$, например её аддитивность относительно независимых наблюдений, описываются следующей Теоремой.

Теорема 11.1.3.

1) Если выполнены Условия R(1) – R(3), то справедливы равенства

$$\mathsf{E}_\theta\left(\frac{\partial \log p_\theta(X)}{\partial\theta}\right) \equiv 0, \qquad \text{для всех} \qquad \theta \in \Theta;$$

$$I(\theta) = \mathsf{D}_\theta\left(\frac{\partial \log p_\theta(X)}{\partial\theta}\right).$$

2) Если выполнены Условия R(1), R(2) и R(4), то справедливо равенство

$$I(\theta) = -\mathsf{E}_\theta\left(\frac{\partial^2 \log p_\theta(X)}{\partial\theta^2}\right).$$

3) Пусть X и Z независимые наблюдения, имеющие плотности $p_\theta(x)$ и $q_\theta(x)$ относительно мер $\nu(x)$ и $\mu(x)$. Пусть $I_X(\theta)$, $I_Z(\theta)$ и $I_{X,Z}(\theta)$ соответственно информации о θ, содержащиеся соответственно в X, Z и (X,Z). Тогда если плотности $p_\theta(x)$ и $q_\theta(x)$ удовлетворяют Условиям R(1) – R(3), то справедливо соотношение

$$I_{X,Z}(\theta) = I_X(\theta) + I_Z(\theta).$$

Если $X = (X_1, \cdots, X_n)$ – независимые одинаково распределенные наблюдения, для плотностей которых выполнены Условия R(1) – R(3), то

$$I_X(\theta) = nI_{X_1}(\theta).$$

Доказательство.

1) Доказываемое утверждение следует из Условий R(1) – R(3), тождества

$$\int p_\theta(x)d\nu(x) \equiv 1$$

и Определения 11.1.1 фишеровской информации $I(X)$.

2) Требуемый результат следует из тождества

$$\frac{\partial^2 \log p_\theta(x)}{\partial\theta^2} \equiv \frac{1}{p_\theta(x)} \cdot \frac{\partial^2 p_\theta(x)}{\partial\theta^2} - \left(\frac{\partial \log p_\theta(x)}{\partial\theta}\right)^2$$

после взятия математического ожидания E_θ от обеих частей.

3) По определению

$$I_{X,Z}(\theta) = \mathsf{E}_\theta\left(\frac{\partial \log p_\theta(X)}{\partial\theta} + \frac{\partial \log q_\theta(Z)}{\partial\theta}\right)^2,$$

поэтому требуемый результат следует из соотношения

$$\mathsf{E}_\theta\left(\frac{\partial \log p_\theta(X)}{\partial\theta} \cdot \frac{\partial \log q_\theta(Z)}{\partial\theta}\right) = \mathsf{E}_\theta\frac{\partial \log p_\theta(X)}{\partial\theta} \cdot \mathsf{E}_\theta\frac{\partial \log q_\theta(Z)}{\partial\theta} \equiv 0.$$

□

Вернёмся теперь к неравенству (11.1.7). В силу первого утверждения Теоремы 11.1.3 знаменатель в правой части этого неравенста можно заменить на фишеровскую игформацию $I(\theta)$. В результате получается следующая версия *информационного неравенства.*

Теорема 11.1.4. (Неравенство Крамéра – Рао) *Пусть выполнены Условия R(1) – R(3), R(5) и* $I(\theta) > 0$. *Пусть* $\delta(X)$ *– любая оценка, для которой выполнено Условие R(4). Тогда*

$$\mathsf{D}_\theta \delta(X) \ \geq \left(\frac{\partial \mathsf{E}_\theta \delta(X)}{\partial \theta} \right)^2 \cdot \frac{1}{I(\theta)}$$

Следствия.

1) *Если* $\delta(X)$ *есть оценка функции* $g(\theta)$ *и*

$$\mathsf{E}_\theta \delta(\theta) = g(\theta) + b(\theta),$$

где $b(\theta)$ *есть смещение оценки* $\delta(X)$*, то при выполнении условий Теоремы 11.1.4, Условия R(6) и дифференцируемости смещения* $b(\theta)$ *справедливо неравенство*

$$\mathsf{D}_\theta \delta(X) \geq \frac{\Big(g'(\theta) + b'(\theta)\Big)^2}{I(\theta)}, \quad \mathsf{E}_\theta (\delta(X) - g(\theta))^2 \geq b^2(\theta) + \frac{\Big(g'(\theta) + b'(\theta)\Big)^2}{I(\theta)}.$$

2) *Если* $\delta = \delta(X)$*, где* $X = (X_1, \cdots, X_n)$ *и наблюдения* $(X_1, \cdots, X_n)$ *независимы и одинаково распределены, тогда, если для отдельного наблюдения* X_i *выполнены условия регулярности из предыдущего Следствия, то*

$$\mathsf{D}_\theta \delta(X) \ \geq \ \frac{\Big(g'(\theta) + b'(\theta)\Big)^2}{n I_{X_1}(\theta)}.$$

Доказательство. Этот результат непосредственно следует из неравенства (11.1.7) и первого утверждения Теоремы 11.1.3. Однако мы дадим его полное доказательство (оно фактически повторяет доказательство Теоремы 9.2.1), поскольку из него можно получить условия, при которых информационное неравенство обращается в равенство. Дифференцируя тождества по θ

$$\int p_\theta(x) d\nu(x) \equiv 1, \quad \int \delta(x) p_\theta(x) d\nu(x) \equiv g(\theta) + b(\theta),$$

с использованием Условий R(1) – R(3), R(5), получим

$$0 \equiv \int\limits_A p'_\theta(x)d\nu(x) = \mathsf{E}_\theta \frac{\partial \log p_\theta(X)}{\partial\theta}, \quad \text{для всех} \quad \theta \in \Theta,$$

$$g'(\theta) + b'(\theta) \equiv \int\limits_A \delta(x)p'_\theta(x)d\nu(x) = \mathsf{E}_\theta \delta(X)\frac{\partial \log p_\theta(X)}{\partial\theta}, \quad \text{для всех} \quad \theta \in \Theta.$$

Поэтому

$$g'(\theta) + b'(\theta) = \mathsf{E}_\theta\Big(\delta(X) - g(\theta) - b(\theta)\Big)\frac{\partial \log p_\theta(X)}{\partial\theta}.$$

Таким образом, используя неравенство Коши – Буняковского, можно записать

$$\Big(g'(\theta) + b'(\theta)\Big)^2 \ \le\ \mathsf{D}_\theta\delta(X)\mathsf{E}_\theta\left(\frac{\partial \log p_\theta(X)}{\partial\theta}\right)^2 = \mathsf{D}_\theta\delta(X)I_X(\theta).$$

□

Задача 11.1.1. Пусть наблюдения $X = (X_1, \cdots, X_n)$ независимы и одинаково распределены с общей плотностью $p_\theta(x)$, которая положительна при всех $x \in \mathcal{X}$ и всех $\theta \in \Theta$. Тогда дисперсия любой несмещённой оценки $\delta(X)$ параметра θ удовлетворяет неравенству

$$\mathsf{D}_{\theta_0}\delta(X) \ \ge\ \frac{(\theta-\theta_0)^2}{\Big(\int \frac{p^2_\theta(x)}{p_{\theta_0}(x)}d\nu(x)\Big)^n - 1}, \quad \text{при всех} \quad \theta \in \Theta, \ \ \theta \ne \theta_0.$$

11.2 ЭФФЕКТИВНЫЕ ОЦЕНКИ

Рассмотрим теперь вопрос о том, когда в неравенстве Крамéра – Рао достигается *равенство*.

Определение 11.2.1. *Несмещенная оценка $\bar\delta(X)$ дифференцируемой функции $g(\theta)$ называется эффективной, если для ее дисперсии справедливо тождество*

$$\mathsf{D}_\theta\bar\delta(X) \equiv \frac{\Big(g'(\theta)\Big)^2}{I(\theta)}, \qquad \textit{при всех} \quad \theta \in \Theta.$$

Пусть выполнены условия регулярности из Теоремы 11.1.4, тогда из этой Теоремы следует, что если существует эффективная оценка, то она является *оптимальной* и значит Теорема 9.2.2 показывает, что она *единственна.* Простой критерий эффективности даётся следующей Теоремой.

Теорема 11.2.1. *Пусть выполнены условия регулярности из Теоремы 11.1.4, тогда для того, чтобы несмещенная оценка $\bar{\delta}(X)$ функции $g(\theta)$ была эффективной необходимо и достаточно, чтобы выполнялось следующее представление*

$$\frac{\partial \log p_\theta(x)}{\partial \theta} \equiv A(\theta)\Big(\bar{\delta}(x) - g(\theta)\Big), \qquad \text{для всех} \quad \theta \in \Theta,$$

где $A(\theta)$ – некоторая функция θ. При этом

$$\mathsf{D}_\theta \bar{\delta}(X) = \left|\frac{g'(\theta)}{A(\theta)}\right|.$$

Доказательство. Из доказательства Теоремы 11.1.4 следует, что в информационном неравенстве достигается равенство тогда и только тогда, когда достигается равенство в неравенстве Коши – Буняковского

$$\mathsf{E}_\theta\Big(\delta(X) - g(\theta)\Big)\frac{\partial \log p_\theta(X)}{\partial \theta} \ \le\ \mathsf{D}_\theta \delta(X) \mathsf{E}_\theta \left(\frac{\partial \log p_\theta(X)}{\partial \theta}\right)^2.$$

Хорошо известно, что это равенство достигается в случае линейной зависимости функций

$$\frac{\partial \log p_\theta(x)}{\partial \theta} \quad \text{и} \quad \bar{\delta}(x) - g(\theta).$$

Далее для эффективной оценки $\bar{\delta}(X)$ имеем

$$\mathsf{D}_\theta \bar{\delta}(X) = \frac{\Big(g'(\theta)\Big)^2}{\mathsf{E}_\theta\Big(\frac{\partial \log p_\theta(X)}{\partial \theta}\Big)^2} = \frac{\Big(g'(\theta)\Big)^2}{A^2(\theta)\mathsf{E}_\theta\Big(\bar{\delta}(X) - g(\theta)\Big)^2} = \frac{\Big(g'(\theta)\Big)^2}{A^2(\theta)\mathsf{D}_\theta \bar{\delta}(X)}.$$

□

Пример 11.2.1. Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые пуассоновские наблюдения

$$X_i \sim \mathcal{P}(\theta), \quad \theta > 0, \quad i = 1, \cdots, n.$$

Тогда условия регулярности Теоремы 11.2.1 выполнены и

$$p_\theta(x) = \mathsf{P}_\theta(X_1 = x_1, \cdots, X_n = x_n) = e^{-n\theta}\frac{\theta^{x_1 + \cdots + x_n}}{x_1! \cdots x_n!}, \qquad x = (x_1, \cdots, x_n),$$

поэтому

$$\frac{\partial \log p_\theta(X)}{\partial \theta} \equiv -n + \frac{\sum\limits_{i=1}^{n} X_i}{\theta} \equiv \frac{n}{\theta}\Big(\overline{X} - \theta\Big), \qquad \text{для всех} \quad \theta > 0,$$

и значит $A(\theta) = n/\theta$ и эффективная оценка для параметра θ имеет вид

$$\bar{\delta}(X) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

11.3 СПИСОК ЛИТЕРАТУРЫ

1) Э. Леман, Теория Точечного Оценивания,
Москва, Наука, 1991, Глава 1, ¡6.

2) Г.И. Ивченко, Ю.И. Медведев, Математическая Статистика,
Москва, Высшая Школа, 1992, Глава 2, ¡2.2.

3) Л.Н. Большев, Уточнение неравенства Краме́ра – Рао,
Теория вероятностей и её применения, 1961, т. 6, н. 3, стр. 319 – 326.

4) Ш. Закс, Теория Статистических Выводов,
Москва, Мир, 1975, Глава 4, ¡4.1.

5) Э. Питмен, Основы Теории Статистических Выводов,
Москва, Мир, 1986, Глава 5.

Лекция 12

В Лекции рассматриваются некоторые методы построения оценок, приводящие к разумным результатам. Рассмотрены асимптотические свойства получаемых оценок.

12.1 СОСТОЯТЕЛЬНЫЕ ОЦЕНКИ

Рассмотрим доминируемую статистическую структуру $(\mathcal{X}_n,\, \mathcal{F}_n,\, \{\mathsf{P}_{n\theta},\, \theta \in \Theta\})$, зависящую от параметра $n \in \mathbf{N}$, который может интерпретироваться как *размер выборки* (например, если наблюдение $\mathbf{X}_n$, имеет вид $\mathbf{X}_n = (X_1, \cdots, X_n)$, где X_i – независимые наблюдения и исходная статистическая структура является прямым произведением n структур). До сих пор мы считали размер выборки n *фиксированным*. Предположим теперь, что параметр n "большой", то есть пусть $n \to \infty$. Будем обозначать наблюдения и оценки параметрической функции $g(\theta)$ соответственно через $\mathbf{X}_n \in \mathcal{X}_n$ и $\delta_n = \delta_n(\mathbf{X}_n)$ и рассмотрим *последовательности* статистических структур $(\mathcal{X}_n,\, \mathcal{F}_n,\, \{\mathsf{P}_{n\theta},\, \theta \in \Theta\})$ и оценок $\delta_n = \delta_n(\mathbf{X}_n)$.

Предположим, что $\mathbf{X}_n = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения, имеющие одинаковое распределение $\mathsf{P}_\theta,\ \theta \in \Theta$, и что мы хотим оценить функцию $g(\theta)$. С ростом n информации о $\theta \in \Theta$ становится всё больше и бльше, и хотелось бы ожидать, что при достаточно больших значениях n можно было бы оценить $g(\theta)$ достаточно точно. Если $\delta_n(\mathbf{X}_n)$ – некоторая разумная оценка функции $g(\theta)$, то, конечно, нельзя ожидать, что её значения близки к $g(\theta)$ для всех конкретных значений наблюдений $X_1 = x_1, \cdots, X_n = x_n$. Но можно надеяться, что $\delta_n(\mathbf{X}_n)$ будет близка к $g(\theta)$ с *большой вероятностью*.

Эта идея формализуется в следующем Определении, при этом наблюдения $\mathbf{X}_n$ не обязаны иметь вид $\mathbf{X}_n = (X_1, \cdots, X_n)$.

Определение 12.1.1. *Последовательность оценок $\delta_n = \delta_n(\mathbf{X}_n)$ функции $g(\theta)$ называется состоятельной, если для любого $\varepsilon > 0$ справедливо соотношение*

$$\mathsf{P}_{n\theta}\Big(|\delta_n(\mathbf{X}_n) - g(\theta)| \ \geq \ \varepsilon\Big) \to 0, \quad n \to \infty, \qquad \textit{для всех} \quad \theta \in \Theta.$$

Символически это обозначается как

$$\delta_n(\mathbf{X}_n) \xrightarrow{P_{n\theta}} g(\theta), \quad n \to \infty, \qquad \textit{для всех} \quad \theta \in \Theta.$$

Пример 12.1.1. Пусть $\mathbf{X}_n = (X_1, \cdots, X_n)$, где $X_i, \quad i = 1, \cdots, n$ – независимые одинаково распределённые наблюдения

$$X_i \sim \mathcal{N}(\theta,\, 1), \quad i = 1, \cdots, n.$$

Тогда в силу Закона Больших Чисел

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P_{n\theta}} \theta, \quad n \to \infty, \quad \text{для всех} \quad \theta \in \mathbf{R}^1.$$

Если же

$$X_i \sim \mathcal{N}(0,\, \theta^2), \quad i = 1, \cdots, n,$$

то аналогично

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

$$= \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right] \xrightarrow{P_{n\theta}} \theta^2, \quad n \to \infty, \quad \text{для всех} \quad \theta^2 \ \geq \ 0.$$

Следующая Теорема часто является полезной при доказательстве состоятельности.

Теорема 12.1.1.

1) *Пусть $\delta_n(\mathbf{X}_n)$ – последовательность оценок параметрической функции $g(\theta)$ с функцией риска*

$$R_n(\theta, \delta_n) = c(\theta)\mathsf{E}_{n\theta}(\delta_n(\mathbf{X}_n) - g(\theta))^2, \quad c(\theta) > 0.$$

Тогда, если

$$R_n(\theta, \delta_n) \to 0, \qquad \textit{для всех} \quad \theta \in \Theta,$$

то $\delta_n(\mathbf{X}_n)$ – состоятельная оценка функции $g(\theta)$.

2) Пусть

$$\mathsf{E}_{n\theta}\delta_n(\mathbf{X}_n) = g(\theta) + b_n(\theta), \quad \mathsf{D}_{n\theta}\delta_n(\mathbf{X}_n) = \eta_n(\theta).$$

Тогда, если

$$b_n(\theta) \to 0, \quad \text{и} \quad \eta_n(\theta) \to 0, \qquad \text{для всех} \quad \theta \in \Theta,$$

то $\delta_n(\mathbf{X}_n)$ *– состоятельная оценка* $g(\theta)$.

3) В частности, $\delta_n(\mathbf{X}_n)$ *состоятельна, если она несмещенная при каждом* n *и*

$$\mathsf{D}_{n\theta}\delta_n(\mathbf{X}_n) \to 0, \quad n \to \infty, \qquad \text{для всех} \quad \theta \in \Theta.$$

Доказательство.

1) Доказательство следует из неравенства Чебышева

$$\varepsilon^2 \mathsf{P}_{n\theta}\Big(|\delta_n(\mathbf{X}_n) - g(\theta)| \geq \varepsilon\Big) \leq \mathsf{E}_{n\theta}(\delta_n(\mathbf{X}_n) - g(\theta))^2 \to 0, \quad n \to \infty, \qquad \text{для всех} \quad \theta \in \Theta.$$

2) Доказательство также следует из неравенства Чебышева

$$\mathsf{P}_{n\theta}\Big(|\delta_n(\mathbf{X}_n) - g(\theta)| \geq \varepsilon\Big) = \mathsf{P}_{n\theta}\Big(|\delta_n(\mathbf{X}_n) - \mathsf{E}_{n\theta}\delta_n(\mathbf{X}_n) + \mathsf{E}_{n\theta}\delta_n(\mathbf{X}_n) - g(\theta)| \geq \varepsilon\Big) \leq$$

$$\leq \mathsf{P}_{n\theta}\Big(|\delta_n(\mathbf{X}_n) - \mathsf{E}_{n\theta}\delta_n(\mathbf{X}_n)| + |b_n(\theta)| \geq \varepsilon\Big) = \mathsf{P}_{n\theta}\Big(|\delta_n(\mathbf{X}_n) - \mathsf{E}_{n\theta}\delta_n(\mathbf{X}_n)| > \varepsilon - |b_n(\theta)|\Big) \leq$$

$$\leq \frac{\mathsf{D}_{n\theta}\delta_n(\mathbf{X}_n)}{(\varepsilon - |b_n(\theta)|)^2} = \frac{\eta_n(\theta)}{(\varepsilon - |b_n(\theta)|)^2} \to 0, \quad n \to \infty, \quad \text{для всех} \quad \theta \in \Theta.$$

□

Пример 12.1.2. Пусть $\mathbf{X}_n = (X_1, \cdots, X_n)$ – независимые одинаково равномерно распределённые наблюдения

$$X_i \sim \mathcal{R}(0, \theta), \quad i = 1, \cdots, n.$$

Тогда

$$\mathsf{E}_{n\theta}X_{(n)} = \frac{n\theta}{n+1}, \qquad \mathsf{D}_{n\theta}X_{(n)} = \frac{n\theta^2}{(n+1)^2(n+2)}, \qquad X_{(n)} = \max_{1 \leq i \leq n} X_i.$$

И значит оценка

$$\delta_n(\mathbf{X}_n) = \frac{n+1}{n} X_{(n)}$$

является состоятельной для параметра $\theta > 0$.

Теорема 12.1.2. *Пусть $\delta_n(\mathbf{X}_n)$ состоятельная оценка функции $g(\theta)$ и функция $h(t)$ непрерывна в точке $g(\theta)$ для каждого $\theta \in \Theta$, тогда оценка $h(\delta_n(\mathbf{X}_n))$ является состоятельной для функции $h(g(\theta))$.*

Доказательство. Фиксируем θ и обозначим $a = g(\theta)$. Из непрерывности функции $h(t)$ в точке a следует, что для любого $\varepsilon > 0$ существует $\gamma = \gamma(\varepsilon) > 0$ такое, что если $|t - a| < \gamma$, то

$$|h(t) - h(a)| < \varepsilon.$$

Отсюда следует, что

$$\mathsf{P}_{n\theta}\Big(|h(\delta_n(\mathbf{X}_n)) - h(g(\theta))| < \varepsilon\Big) \ \geq \ \mathsf{P}_{n\theta}\Big(|\delta_n(\mathbf{X}_n) - g(\theta)| < \gamma\Big) =$$

$$= 1 - \mathsf{P}_{n\theta}\Big(|\delta_n(\mathbf{X}_n) - g(\theta)| \ \geq \ \gamma\Big),$$

поэтому

$$\mathsf{P}_{n\theta}\Big(|h(\delta_n(\mathbf{X}_n)) - h(g(\theta))| \ \geq \ \varepsilon\Big) \ \leq$$

$$\leq \ \mathsf{P}_{n\theta}\Big(|\delta_n(\mathbf{X}_n) - g(\theta)| \ \geq \ \gamma\Big) \to 0, \quad n \to \infty, \qquad \text{для всех} \quad \theta \in \Theta.$$

$\square$

Пример 12.1.3. Пусть $\mathbf{X}_n = (X_1, \cdots, X_n)$ – независимые одинаково равномерно распределённые наблюдения

$$X_i \sim \mathcal{R}(0, \theta), \quad i = 1, \cdots, n.$$

Тогда из этой Теоремы и Примера 12.1.2 следует, что оценка

$$\delta_n(\mathbf{X}_n) = \arctan\Big(X_{(n)}(n+1)/n\Big)$$

является состоятельной оценкой функции $g(\theta) = \arctan\theta$.

12.2 МЕТОД МОМЕНТОВ

Пусть $\mathbf{X}_n = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения, причём существуют моменты вида

$$\mathsf{E}_\theta X_1^j = \alpha_j(\theta), \quad j = 1, \cdots, r, \quad \text{для всех} \quad \theta \in \Theta,$$

зависящие от параметра $\theta = (\theta_1, \cdots, \theta_r) \in \Theta \subseteq \mathbf{R}^r$. Определим *эмпирические моменты* (см. Лекцию 6) по формуле

$$\overline{\alpha}_{jn} = \frac{1}{n}\sum_{i=1}^{n} X_i^j, \qquad j = 1, \cdots, r.$$

Предположим, что оцениваемая функция $g(\theta)$ представима в виде *непрерывной* функции от

$$\alpha_1(\theta), \cdots, \alpha_r(\theta),$$

то есть

$$g(\theta) = h(\alpha_1(\theta), \cdots, \alpha_r(\theta)).$$

Тогда оценкой $g(\theta)$ по *методу моментов* называется оценка вида

$$\delta_n(\mathbf{X}_n) = h(\overline{\alpha}_{1n}, \cdots, \overline{\alpha}_{rn}).$$

Непрерывность функции h и многомерный вариант Теоремы 12.1.2 обеспечивают *состоятельность* оценки $\delta_n(\mathbf{X}_n)$, поскольку в силу Закона Больших Чисел (см. Лекцию 4, п.6)

$$\overline{\alpha}_{jn} \xrightarrow{P_{n\theta}} \alpha_j(\theta), \quad j = 1, \cdots, r, \qquad \text{для всех} \quad \theta \in \Theta$$

и поэтому

$$\delta_n(\mathbf{X}_n) \xrightarrow{P_{n\theta}} h(\alpha_1(\theta), \cdots, \alpha_r(\theta)) = g(\theta) \qquad \text{для всех} \quad \theta \in \Theta.$$

Если $\Theta \subseteq \mathbf{R}^r$ и $g(\theta) = \theta$, то оценка по *методу моментов* находится как решение системы уравнений

$$\alpha_j(\theta) = \overline{\alpha}_{jn}, \quad j = 1, \cdots, r,$$

принадлежащее Θ. Если эта система допускае *однозначное и непрерывное* решение

$$\theta = H(\overline{\alpha}_{1n}, \cdots, \overline{\alpha}_{rn}),$$

то в качестве оценки берём оценку вида

$$\delta_n(\mathbf{X}_n) = H(\overline{\alpha}_{1n}, \cdots, \overline{\alpha}_{rn}).$$

Поскольку

$$\theta \equiv H(\alpha_1(\theta), \cdots, \alpha_r(\theta)),$$

то опять в силу непрерывности функции H, имеем

$$\delta_n(\mathbf{X}_n) = H(\overline{\alpha}_{1n}, \cdots, \overline{\alpha}_{rn}) \xrightarrow{P_{n\theta}} H(\alpha_1(\theta), \cdots, \alpha_r(\theta)) = \theta, \quad \text{для всех} \quad \theta \in \Theta.$$

Таким образом $\delta_n(\mathbf{X}_n)$ – состоятельная оценка.

Пример 12.2.1. Пусть $\mathbf{X}_n = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения, имеющие плотность

$$p_\theta(x) = \theta e^{-\theta x}, \quad x > 0; \quad \theta > 0.$$

Найдём оценку по методу моментов для θ, используя только второй момент.

$$\alpha_2(\theta) = \int_0^\infty \theta e^{-\theta x} x^2 dx = -\int_0^\infty x^2 de^{-\theta x} = 2\int_0^\infty x e^{-\theta x} dx = \frac{2}{\theta^2}.$$

Поэтому имеем уравнение

$$\frac{2}{\theta^2} = \overline{\alpha}_{2n} = \frac{1}{n}\sum_{i=1}^n X_i^2$$

и значит оценка по методу моментов имеет вид

$$\delta_n(\mathbf{X}_n) = \frac{1}{\sqrt{\frac{1}{2n}\sum\limits_{i=1}^n X_i^2}} \xrightarrow{P_{n\theta}} \theta, \quad \text{для всех} \quad \theta > 0.$$

12.3 МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Рассмотрим ещё один метод построения оценок, приводящий к разумным результатам. Поскольку мы рассматриваем только доминируемые статистические структуры $(\mathcal{X}_n, \mathcal{F}_n, \{\mathsf{P}_{n\theta}, \theta \in \Theta\})$, то обозначим через $p_{n\theta}(x), \ \ x \in \mathcal{X}_n, \ \ \theta \in \Theta$ – плотность относительно доминирующей меры ν_n.

Определение 12.3.1 *Оценка* $\hat{\delta}_n = \hat{\delta}_n(\mathbf{X}_n)$ *называается оценкой максимального правдоподобия, если*

$$\mathcal{L}_n(\hat{\delta}_n;\, x_n) = \sup_{\theta \in \Theta} \mathcal{L}_n(\theta;\, x_n),$$

где через

$$\mathcal{L}_n(\theta;\, x_n) = p_{n\theta}(x_n)$$

обозначена функция правдоподобия.

Рассмотрим теперь некоторые свойства оценок максимального правдоподобия, причём будем считать n, сначала, *фиксированным* и поэтому будем опускать индекс n в обозначениях.

Заметим, что оценка максимального правдоподобия может не существовать, может определяться неоднозначно или может не быть оптимальной. Однако покажем на эвристическом уровне, что типичным образом метод максимального правдоподобия приводит к разумным результатам.

Пусть наблюдение имеет дискретное распределение и

$$\mathcal{L}(\theta;\, x) = \mathsf{P}_\theta(X = x), \quad x \in \mathcal{X}, \quad \theta \in \Theta.$$

Предположим, что мы наблюдаем конкретное значение $X = x$, тогда поскольку обычно происходят события, имеющие *наибольшую вероятность*, то этому значению x соответствует θ *максимизирующее* функцию правдоподобия $\mathcal{L}(\theta;\, x)$ при фиксированном $x \in \mathcal{X}$.

Справедливы следующие свойства оценок максимального правдоподобия.

1) В регулярном случае оценка максимального правдоподобия удовлетворяет уравнению
$$\left.\frac{\partial \mathcal{L}(\theta;\, x)}{\partial \theta}\right|_{\theta = \hat{\delta}(x)} = 0.$$

2) Если в регулярном случае существует *эффективная* оценка $\bar{\delta}(X)$ параметра θ, то
$$\bar{\delta}(X) = \hat{\delta}(X),$$
поскольку (см. Теорему 11.2.1) в этом случае
$$\frac{\partial \log \mathcal{L}(\theta;\, x)}{\partial \theta} = A(\theta)(\bar{\delta}(x) - \theta) = 0$$
и значит
$$\bar{\delta}(X) = \hat{\delta}(X).$$

3) Если сущестует достаточная статистика $T = T(X)$ и существует оценка максимального правдоподобия $\hat{\delta}(X)$, то она зависит от X только через достаточную статистику $T(X)$
$$\hat{\delta}(X) = \delta(T(X)),$$

поскольку по критерию факторизации (Теорема 7.1.3)

$$\mathcal{L}(\theta;\, x) = h(x) g_\theta(T(x))$$

и максимизация $\mathcal{L}(\theta;\, x)$ по θ сводится к максимизации $g_\theta(T(x))$.

Теорема 12.3.1. (Принцип инвариантности оценок максимального правдоподобия) *Пусть оцениваемая функция $g(\theta)$ измерима и*

$$g : \Theta \longrightarrow \Gamma.$$

Тогда, если $\hat{\delta}(X)$ – оценка максимального правдоподобия для параметра θ, то $g(\hat{\delta}(X))$ – оценка максимального правдоподобия для $g(\theta)$.

Доказательство. Обычно считают, что $g(\theta)$ – взаимнооднозначная функция. Здесь мы этого не предполагаем. Для каждого $g \in \Gamma$ определим множества

$$Z(g) = \{\theta \in \Theta :\ g(\theta) = g\} = g^{-1}(g).$$

Тогда оценкой максимального правдоподобия для функции $g(\theta)$ называется оценка $\hat{g}(x)$ такая, что

$$\sup_\theta \mathcal{L}(\theta;\, x) = \sup_{g \in \Gamma} \sup_{\theta \in Z(g)} \mathcal{L}(\theta;\, x) = \sup_{\theta \in Z(\hat{g}(x))} \mathcal{L}(\theta;\, x).$$

Пусть

$$M(g, x) = \sup_{\theta \in Z(g)} \mathcal{L}(\theta;\, x), \qquad g \in \Gamma$$

и $\hat{\delta}(x)$ – оценка максимального правдоподобия для θ.

Тогда существует элемент $\tilde{g}(x) \in \Gamma$ такой, что

$$\hat{\delta}(x) \in Z(\tilde{g}(x)),$$

поэтому

$$\mathcal{L}(\hat{\delta}(x);\, x)) \ \le \sup_{\theta \in Z(\tilde{g}(x))} \mathcal{L}(\theta;\, x) = M(\tilde{g}(x), x) \ \le$$

$$\le \sup_{g \in \Gamma} M(g, x) = \sup_{g \in \Gamma} \sup_{\theta \in Z(g)} \mathcal{L}(\theta;\, x) = \sup_\theta \mathcal{L}(\theta;\, x) = \mathcal{L}(\hat{\delta}(x);\, x).$$

Отсюда следует, что

$$M(\tilde{g}(x), x) = \sup_{g \in \Gamma} M(g, x)$$

и значит $\tilde{g}(x)$ – оценка максимального правдоподобия для $g(\theta)$, то есть $\tilde{g}(x) = \hat{g}(x)$ и

$$\hat{\delta}(x) \in Z(\tilde{g}(x)) = Z(\hat{g}(x)),$$

поэтому

$$\hat{g}(x) = g(\hat{\delta}(x)).$$

□

Пример 12.3.1. Пусть $\mathbf{X}_n = (X_1, \cdots, X_n)$ – независимые нормально распределённые наблюдения

$$X_i \sim \mathcal{N}(\theta,\ 1), \quad i = 1, \cdots, n; \quad \theta \in \mathbf{R}^1.$$

Тогда, поскольку, оценка вида

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

является оценкой максимального правдоподобия для параметра θ, то оценкой максимального правдоподобия для функции

$$g(\theta) = \Phi(x - \theta) = \mathsf{P}_\theta(X_1 < x) \quad (x \text{ фиксировано})$$

будет $\Phi(x - \overline{X})$.

12.4 СПИСОК ЛИТЕРАТУРЫ

1) Э. Леман, Теория Точечного Оценивания,
   Москва, Наука, 1991, Глава 5, ¡1.

2) Г.И. Ивченко, Ю.И. Медведев, Математическая Статистика,
   Москва, Высшая Школа, 1992, Глава 2, ¡2.4, ¡2.5.

3) Ш. Закс, Теория Статистических Выводов,
   Москва, Мир, 1975, Глава 4, ¡4.5, Глава 5, ¡5.1.

4) А.А. Боровков, Математическая Статистика,
   Москва, Наука, 1984, Глава 2, ¡4, ¡6.

Лекция 13

В Лекции доказывается состоятельность и асимптотическая нормальность оценок максимального правдоподобия.

13.1 АСИМПТОТИЧЕСКИЕ СВОЙСТВА ОЦЕНОК МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Пусть теперь наблюдения имеют вид $\mathbf{X}_n = (X_1, \cdots, X_n), \quad n \in \mathbf{N}$, где X_i – независимы и одинаково распределены с общей плотностью $p_\theta(x), \; \theta \in \Theta$. Докажем, состоятельность оценок максимального правдоподобия $\hat{\delta}_n(\mathbf{X}_n)$.

Теорема 13.1.1. (Состоятельность оценок максимального правдоподобия) *Пусть $\Theta \subseteq \mathbf{R}^k$ – открытое ограниченное множество и $\theta_0 \in \Theta$. Пусть выполнены следующие условия регулярности*

1) Множество

$$\{x \in \mathcal{X} : \; p_\theta(x) > 0\}$$

не зависит от $\theta \in \Theta$.

2) Для любого $\theta \in \Theta, \quad \theta \neq \theta_0$ справедливо неравенство

$$\int |p_\theta(x) - p_{\theta_0}(x)| d\nu(x) > 0.$$

3) Для любого $x \in \mathcal{X}$ плотность $p_\theta(x)$ непрерывна по $\theta \in \overline{\Theta}$, где $\overline{\Theta}$ – замыкание множества Θ.

4) Для любого $\theta \in \overline{\Theta}$

$$\mathsf{E}_{\theta_0}|l_\theta(X_1)| = \int |l_\theta(x)| p_{\theta_0}(x) d\nu(x) < \infty,$$

где

$$l_\theta(x) = \log p_\theta(x).$$

5) Для любого $\theta \in \Theta$ существует окрестность $U_\theta \subseteq \Theta,\ \theta \in U_\theta$ такая, что

$$\mathsf{E}_{\theta_0}|\sup_{\theta_1 \in U_\theta} l_{\theta_1}(X_1)| < \infty.$$

Определим оценку максимального правдоподобия $\hat{\delta}_n(\mathbf{X}_n)$ как

$$\sum_{i=1}^{n} \log p_{\hat{\delta}_n(\mathbf{X}_n)}(X_i) = \sup_{\theta \in \overline{\Theta}} \sum_{i=1}^{n} \log p_\theta(X_i).$$

Тогда оценка максимального правдоподобия $\hat{\delta}_n(\mathbf{X}_n)$ состоятельна, то есть для любого $\varepsilon > 0$

$$\mathsf{P}_{n\theta_0}\Big(\|\hat{\delta}_n(\mathbf{X}_n) - \theta_0\| \ \geq\ \varepsilon\Big) \to 0, \qquad n \to \infty.$$

Доказательство. Докажем сначала вспомогательную Лемму.

Лемма 13.1.1. *Из Условий 1 и 2 следует что*

$$\mathsf{E}_{\theta_0} l_{\theta_0}(X_1) > \mathsf{E}_{\theta_0} l_\theta(X_1), \quad \text{для любого} \ \ \theta \in \Theta, \quad \theta \neq \theta_0.$$

и

$$\mathsf{P}_{n\theta_0}\Big(\prod_{i=1}^{n} p_{\theta_0}(X_i) > \prod_{i=1}^{n} p_\theta(X_i)\Big) \to 1, \quad n \to \infty, \quad \text{для любого} \ \ \theta \in \Theta, \quad \theta \neq \theta_0.$$

Доказательство Леммы. Первое утверждение Леммы эквивалентно неравенству

$$\mathsf{E}_{\theta_0} \log \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} < 0.$$

Поскольку логарифмическая функция выпукла вверх, то из неравенства Йенсена (см. (8.2.4)) следует, что

$$\mathsf{E}_{\theta_0} \log \frac{p_\theta(X_1}{p_{\theta_0}(X_1)} \ \leq\ \log \mathsf{E}_{\theta_0} \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} = \log 1 = 0.$$

Причём из Условия 2 вытекает, что равенство здесь возможно только при $\theta = \theta_0$. Заметим, что Лемму можно доказать без использования неравенства Йенсена. Для этого используем неравенство

$$\log(1+x) \le x, \quad x \ge -1$$

причём равенство здесь возможно только при $x = 0$. Имеем

$$\mathsf{E}_{\theta_0} \log \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} = \int \log \frac{p_\theta(x)}{p_{\theta_0}(x)} p_{\theta_0}(x) d\nu(x) \le \int \left(\frac{p_\theta(x)}{p_{\theta_0}(x)} - 1 \right) p_{\theta_0}(x) d\nu(x) = 0.$$

Причём, если здесь равенство, то

$$1 = \mathsf{P}_{\theta_0} \left(\log \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} = \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} - 1 \right) = \mathsf{P}_{\theta_0} \left(\frac{p_\theta(X_1)}{p_{\theta_0}(X_1)} = 1 \right).$$

И это соотношение противоречит Условию 2. Второе утверждение Леммы следует из соотношения

$$\mathsf{P}_{n\theta_0} \left(\prod_{i=1}^n p_{\theta_0}(X_i) \le \prod_{i=1}^n p_\theta(X_i) \right) = \mathsf{P}_{n\theta_0} \left(\frac{1}{n} \sum_{i=1}^n \log \frac{p_{\theta_0}(X_i)}{p_\theta(X_i)} \le 0 \right) =$$

$$= \mathsf{P}_{n\theta_0} \left(\frac{1}{n} \sum_{i=1}^n \Big(\log \frac{p_{\theta_0}(X_i)}{p_\theta(X_i)} - \mathsf{E}_{\theta_0} \log \frac{p_{\theta_0}(X_1)}{p_\theta(X_1)} \Big) \le -\mathsf{E}_{\theta_0} \log \frac{p_{\theta_0}(X_1)}{p_\theta(X_1)} \right) \to 0,$$

$$\text{для любого} \quad \theta \in \Theta, \quad \theta \ne \theta_0,$$

которое следует из Закона Больших Чисел и доказанного неравенства

$$\mathsf{E}_{\theta_0} \log \frac{p_{\theta_0}(X_1)}{p_\theta(X_1)} > 0.$$

□

Заметим, что оценка максимального правдоподобия $\hat{\delta}_n(\mathbf{X}_n)$ действительно существует, поскольку по условию $\overline{\Theta}$ – компакт и плотность $p_\theta(x)$ непрерывна по θ на $\overline{\Theta}$ при любом фиксированном $x \in \mathcal{X}$. Обозначим

$$U_{\theta,k} = \{\theta_1 \in \Theta\colon \|\theta_1 - \theta\| < 1/k\}, \quad k = 1, 2, \cdots; \quad g_{\theta,k}(x) = \sup_{\theta_1 \in U_{\theta,k}} l_{\theta_1}(x).$$

Из Условия 5 следует, что для любого $\theta \in \Theta$ существует $k_0(\theta)$ такое, для любого $k > k_0(\theta)$ справедливо включение

$$U_{\theta,k} \subseteq U_\theta, \quad \mathsf{E}_{\theta_0} g_{\theta,k}(X_1) < \infty.$$

В силу непрерывности плотности $p_\theta(x)$ по $\theta \in \overline{\Theta}$ (Условие 3) для любого $x \in \mathcal{X}$

$$g_{\theta,k}(x) \downarrow l_\theta(x), \quad k \to \infty.$$

Теперь из Теоремы о монотонной сходимости (см. Лекция 2, п. 8) следует сходимость

$$\mathsf{E}_{\theta_0} g_{\theta,k}(X_1) \downarrow \mathsf{E}_{\theta_0} l_\theta(X_1), \quad k \to \infty. \tag{13.1.1}$$

В силу доказанной Леммы 13.1.1

$$\mathsf{E}_{\theta_0} l_{\theta_0}(X_1) > \mathsf{E}_{\theta_0} l_\theta(X_1), \quad \theta \neq \theta_0,$$

поэтому из (13.1.1) следует, что для любого $\theta \neq \theta_0$ существует натуральное $k(\theta)$ такое, что

$$\mathsf{E}_{\theta_0} l_{\theta_0}(X_1) > \mathsf{E}_{\theta_0} g_{\theta,k(\theta)}(X_1). \tag{13.1.2}$$

Обозначим

$$C = \{\theta \in \overline{\Theta} : \|\theta - \theta_0\| \ \geq \ \varepsilon\},$$

тогда множество C ограниченно и замкнуто и поэтому является компактом. Окрестности вида $U_{\theta,k(\theta)}$ образуют открытое покрытие множества C, поэтому существует *конечное* подпокрытие вида

$$U_{\theta_j,k(\theta_j)}, \quad j = 1, \cdots, m; \quad C \subseteq \bigcup_{j=1}^{m} U_{\theta_j,k(\theta_j)}.$$

Обозначим

$$U_j = U_{\theta_j,k(\theta_j)}, \quad g_j(x) = g_{\theta_j,k(\theta_j)}(x).$$

Тогда из неравенства (13.1.2) непосредственно следует, что справедливы неравенства

$$\mathsf{E}_{\theta_0} l_{\theta_0}(X_1) > \mathsf{E}_{\theta_0} g_j(X_1), \quad j = 1, \cdots, m. \tag{13.1.3}$$

Далее имеем

$$\{\mathbf{X}_n : \|\hat{\delta}_n(\mathbf{X}_n) - \theta_0\| \ \geq \ \varepsilon\} = \{\mathbf{X}_n : \hat{\delta}_n(\mathbf{X}_n) \in C\} \subseteq \bigcup_{j=1}^{m} \{\mathbf{X}_n : \hat{\delta}_n(\mathbf{X}_n) \in U_j\}$$

и из этого соотношения следует неравенство

$$\mathsf{P}_{n\theta_0}(\|\hat{\delta}_n(\mathbf{X}_n) - \theta_0\| \ \geq \ \varepsilon) \ \leq \ \sum_{j=1}^{m} \mathsf{P}_{n\theta_0}(\hat{\delta}_n(\mathbf{X}_n) \in U_j).$$

Теперь для доказательства Теоремы достаточно доказать, что каждое слагаемое в этой сумме стремится к нулю при $n \to \infty$.

Рассмотрим j - ое слагаемое. Пусть $\hat{\delta}_n(\mathbf{X}_n) \in U_j$, тогда в силу Закона Больших Чисел (см. Лекция 4, п.6)

$$\frac{1}{n}\sum_{i=1}^{n} l_{\theta_0}(X_i) \xrightarrow{\mathsf{P}_{n\theta_0}} \mathsf{E}_{\theta_0} l_{\theta_0}(X_1),$$

но

$$\frac{1}{n}\sum_{i=1}^{n} l_{\theta_0}(X_i) \;\le\; \frac{1}{n}\sum_{i=1}^{n} l_{\hat{\delta}_n(\mathbf{X}_n)}(X_i) \;\le\; \frac{1}{n}\sum_{i=1}^{n} g_j(X_i) \xrightarrow{\mathsf{P}_{n\theta_0}} \mathsf{E}_{\theta_0} g_j(X_1),$$

что противоречит неравенствам (13.1.3). Формализуем теперь эту идею. Для любого $a \in \mathbf{R}^1$ справедливо соотношение

$$\{\mathbf{X}_n : \hat{\delta}_n(\mathbf{X}_n) \in U_j\} \subseteq \left\{\mathbf{X}_n : \frac{1}{n}\sum_{i=1}^{n} l_{\theta_0}(X_i) \;\le\; a\right\} \bigcup$$

$$\bigcup \left\{\mathbf{X}_n : \frac{1}{n}\sum_{i=1}^{n} g_j(X_i) > a\right\} \equiv A_n \cup B_n. \qquad (13.1.4)$$

Выберем теперь a так, чтобы получить доказательство. Возьмём

$$a = \frac{1}{2}\Big(\mathsf{E}_{\theta_0} l_{\theta_0}(X_1) + \mathsf{E}_{\theta_0} g_j(X_1)\Big),$$

тогда

$$A_n = \left\{\mathbf{X}_n : \frac{1}{n}\sum_{i=1}^{n}(l_{\theta_0}(X_i) - \mathsf{E}_{\theta_0} l_{\theta_0}(X_1)) \;\le\; \frac{1}{2}(\mathsf{E}_{\theta_0} g_j(X_1) - \mathsf{E}_{\theta_0} l_{\theta_0}(X_1))\right\}.$$

теперь заметим, что в силу (13.1.3) правая часть этого неравенства строго меньше нуля, и поэтому в силу Закона Больших Чисел

$$\mathsf{P}_{n\theta_0}(A_n) \to 0, \qquad n \to \infty.$$

Аналогично получаем

$$\mathsf{P}_{n\theta_0}(B_n) \to 0, \qquad n \to \infty.$$

Теперь утверждение Теоремы следует из соотношения (13.1.4) и неравенства

$$\mathsf{P}_{n\theta_0}(A_n \cup B_n) \;\le\; \mathsf{P}_{n\theta_0}(A_n) + \mathsf{P}_{n\theta_0}(B_n).$$

□

Замечание 13.1.1. Заметим, что из Леммы 13.1.1 непосредственно следует, что если $\Theta \subseteq \mathbf{R}^k$ и *конечно*, то оценка максимального правдоподобия $\hat{\delta}_n(\mathbf{X}_n)$ существует, единственна с вероятностью стремящейся к единице, и состоятельна.

Доказательство следует из того факта, что в случае конечности Θ оценка $\delta_n(\mathbf{X}_n)$ состоятельна тогда и только тогда, когда

$$\mathsf{P}_{n\theta}(\delta_n(\mathbf{X}_n) = \theta) \to 1, \quad n \to \infty, \quad \text{для всех} \quad \theta \in \Theta.$$

И значит

$$\mathsf{P}_{n\theta_0}(\hat{\delta}_n(\mathbf{X}_n) = \theta_0) =$$

$$= \mathsf{P}_{n\theta_0}\Big(\prod_{i=1}^n p_{\theta_0}(X_i) > \prod_{i=1}^n p_\theta(X_i)\Big) \to 1, \quad n \to \infty, \quad \text{для всех} \quad \theta_0 \in \Theta.$$

□

Заметим, что Теорема 13.1.1 доказана при весьма слабых условиях регулярности. В частности не требуется дифференцируемость по θ плотности $p_\theta(x)$. При наличии указанной дифференцируемости доказательство было бы короче и проще. Предполагая дифференцируемость плотности $p_\theta(x)$, для полноты, докажем состоятельность оценок максимального правдоподобия.

Заметим, что в случае $\mathbf{X}_n = (X_1, \cdots, X_n)$, совместная плотность $\mathbf{X}_n$ имеет вид

$$p_{n\theta}(\mathbf{x}_n) = \prod_{i=1}^n p_\theta(x_i), \quad \mathbf{x}_n = (x_1, \cdots, x_n)$$

и значит функция правдоподобия и её логарифм есть

$$\mathcal{L}_n(\theta;\, \mathbf{x}_n) = p_{n\theta}(\mathbf{x}_n) = \prod_{i=1}^n p_\theta(x_i), \quad l_n(\theta;\, \mathbf{x}_n) = \log \mathcal{L}_n(\theta;\, \mathbf{x}_n) = \sum_{i=1}^n \log p_\theta(x_i).$$

Пусть для простоты Θ – открытое множество из $\mathbf{R}^1$. Напомним, что в регулярном случае оценка максимального правдоподобия $\hat{\delta}_n(\mathbf{X}_n)$ удовлетворяет *уравнению правдоподобия*

$$\frac{\partial l_n(\theta;\, \mathbf{x}_n)}{\partial \theta} \equiv l_n'(\theta;\, \mathbf{x}_n) = \sum_{i=1}^n \frac{p_\theta'(x_i)}{p_\theta(x_i)} = 0, \quad \mathbf{x}_n = (x_1, \cdots, x_n). \tag{13.1.5}$$

Теорема 13.1.2. *Пусть $\Theta \subseteq \mathbf{R}^1$ – открытое множество и $\theta_0 \in \Theta$. Пусть выполнены следующие условия регулярности*

1) Множество

$$\{x \in \mathcal{X} : \, p_\theta(x) > 0\}$$

не зависит от $\theta \in \Theta$.

2) Для любого $\theta \in \Theta, \quad \theta \neq \theta_0$ *справедливо неравенство*

$$\int |p_\theta(x) - p_{\theta_0}(x)| d\nu(x) > 0.$$

3) Для ν *– почти всех* $x \in \mathcal{X}$ *плотность* $p_\theta(x)$ *дифференцируема по* $\theta \in \Theta$ *и ее производная есть* $p'_\theta(x)$.

Тогда с вероятностью, стремящейся к единице при $n \to \infty$, *уравнение правдоподобия (13.1.5) имеет корень* $\tilde{\delta}_n(\mathbf{X}_n)$, *такой что для любого* $\varepsilon > 0$

$$\mathsf{P}_{n\theta_0}(|\tilde{\delta}_n(\mathbf{X}_n) - \theta_0| > \varepsilon) \to 0, \quad n \to \infty.$$

Доказательство. Пусть $\varepsilon > 0$ столь мало, что $(\theta_0 - \varepsilon, \theta_0 + \varepsilon) \subset \Theta$ и пусть

$$D_n = \{\mathbf{x}_n : \, l_n(\theta_0;\, \mathbf{x}_n) > l_n(\theta_0 - \varepsilon;\, \mathbf{x}_n), \; l_n(\theta_0;\, \mathbf{x}_n) > l_n(\theta_0 + \varepsilon;\, \mathbf{x}_n)\}.$$

Тогда из Леммы 13.1.1 следует, что

$$\mathsf{P}_{n\theta_0}(D_n) \to 1, \quad n \to \infty.$$

Поэтому для любого $\mathbf{X}_n \in D_n$ существует точка $\tilde{\delta}_n(\mathbf{X}_n, \varepsilon)$, такая что $\tilde{\delta}_n(\mathbf{X}_n, \varepsilon) \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ и в которой функция правдоподобия $l_n(\theta;\, \mathbf{X}_n)$ имеет *локальный максимум*, так что

$$l'_n(\tilde{\delta}_n(\mathbf{X}_n, \varepsilon);\, \mathbf{X}_n) = 0.$$

Следовательно для любого достаточно малого $\varepsilon > 0$ существует последовательность корней $\tilde{\delta}_n(\mathbf{X}_n, \varepsilon)$ этого уравнения, такая что

$$\mathsf{P}_{n\theta_0}(|\tilde{\delta}_n(\mathbf{X}_n, \varepsilon) - \theta_0| > \varepsilon) \to 0, \quad n \to \infty.$$

Остаётся показать, что мы можем определить такую последовательность, которая от ε не зависит. Пусть $\tilde{\delta}_n(\mathbf{X}_n)$ – корень, *ближайший* к θ_0. Он существует, потому что предел последовательности корней вновь есть корень в силу непрерывности функция правдоподобия $l_n(\theta;\, \mathbf{X}_n)$. Тогда ясно, что

$$\mathsf{P}_{n\theta_0}(|\tilde{\delta}_n(\mathbf{X}_n) - \theta_0| > \varepsilon) \to 0, \quad n \to \infty.$$

□

Замечания.

1) Доказанная Теорема *не устанавливает существование состоятельной последовательности оценок*, поскольку когда истинное значение θ_0 неизвестно, данные не указывают нам, какой корень выбирать, чтобы получить состоятельную последовательность. Исключением, конечно, является случай, когда корень *единственный*.

2) Следует подчеркнуть также, что существование корня $\tilde{\delta}_n(\mathbf{x}_n)$ при всех $\mathbf{x}_n$ (или для любого n при заданном $\mathbf{x}_n$) *не утверждается*. Это не влияет на состоятельность, для которой требуется лишь, чтобы оценка $\tilde{\delta}_n(\mathbf{x}_n)$ была определена на множестве, вероятность которого стремится к единице при $n \to \infty$.

3) Если в предположених Теоремы 13.1.2 уравнение правдоподобия (13.1.5) имеет *единственный* корень $\tilde{\delta}_n(\mathbf{x}_n)$ для каждого n и всех $\mathbf{x}_n$, то $\tilde{\delta}_n(\mathbf{X}_n)$ есть состоятельная последовательность оценок для θ. Если, кроме того, параметрическое пространство Θ есть открытый интервал (c, d) (не обязательно конечный), то с вероятностью, стремящейся к единице, $\tilde{\delta}_n(\mathbf{X}_n)$ максимизирует функцию правдоподобия, то есть является оценкой максимального правдоподобия, которая является состоятельной.

   Первое утверждение очевидно. Чтобы доказать второе, предположим, то вероятность того, что $\tilde{\delta}_n(\mathbf{X}_n)$ есть оценка максимального правдоподобия, не стремится к единице. Тогда при достаточно больших n функция правдоподобия должна стремиться с положительной вероятностью к своему супремуму, когда θ стремится к c или к d. Но с вероятностью, стремящейся к единице, $\tilde{\delta}_n(\mathbf{X}_n)$ есть точка локального максимума функции правдоподобия, которая должна тогда обладать также и локальным минимумом. Это противоречит предположению о единственности корня.

Теорема 13.1.2 устанавливает существование состоятельного корня уравнения правдоподобия (13.1.5). Следующая Теорема утверждает, что любая такая последовательность асимптотически нормальна.

Теорема 13.1.3. *Пусть $\theta_0 \in \Theta$ и выполнены следующие условия регулярности.*

1) *Параметрическое пространство Θ есть открытый интервал (не обязательно конечный).*

2) *Множество*
$$A = \{x \in \mathcal{X} :\, p_\theta(x) > 0\}$$

не зависит от $\theta \in \Theta$.

3) *Плотность* $p_\theta(x)$ *трижды дифференцируема по* $\theta \in \Theta$ *в некоторой окрестности точки* θ_0 *при каждом* $x \in A$ *и третья ее производная непрерывна по* $\theta \in \Theta$ *в этой окрестности.*

4) *Интеграл вида*

$$\int p_\theta(x) d\nu(x)$$

можно дважды дифференцировать по $\theta \in \Theta$ *под знаком интеграла в точке* θ_0.

5) *Фишеровская информация (см. Определение 11.1.1)* $I(\theta_0)$, *где*

$$I(\theta) = \mathsf{E}_\theta \left(\frac{\partial \log p_\theta(X_1)}{\partial \theta} \right)^2$$

такова, что $0 < I(\theta_0) < \infty$.

6) *Существуют окрестность точки* θ_0 *и функция* $M(x) \geq 0$ *такие, что в этой окрестности*

$$\left| \frac{\partial^3 \log p_\theta(x)}{\partial \theta^3} \right| \leq M(x), \quad \text{для всех} \quad x \in A$$

и

$$\mathsf{E}_{\theta_0} M(X_1) < \infty.$$

Тогда любая состоятельная последовательность $\tilde{\delta}_n(\mathbf{X}_n)$ *корней уравнения правдоподобия (13.1.5) асимптотически нормальна*

$$\mathsf{P}_{n\theta_0}\Big(\sqrt{n}(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) < x\Big) \to \Phi\Big(\sqrt{I(\theta_0)}x\Big), \quad n \to \infty.$$

Доказательство. Выберем $\varepsilon > 0$ так, чтобы интервал $(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ принадлежал окрестности из формулировки Теоремы. Тогда при фиксированном x и достаточно большом n

$$\mathsf{P}_{n\theta_0}\Big(\sqrt{n}(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) < x\Big) = \mathsf{P}_{n\theta_0}\Big(\sqrt{n}(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) < x, |\tilde{\delta}_n(\mathbf{X}_n) - \theta_0| < \varepsilon\Big) +$$

$$+ \mathsf{P}_{n\theta_0}\Big(\sqrt{n}(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) < x, \ |\tilde{\delta}_n(\mathbf{X}_n) - \theta_0| \geq \varepsilon\Big) =$$

$$= \mathsf{P}_{n\theta_0}\Big(-\sqrt{n}\varepsilon < \sqrt{n}(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) < x\Big) +$$

$$+\mathsf{P}_{n\theta_0}\Big(\sqrt{n}(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) < x,\ |\tilde{\delta}_n(\mathbf{X}_n) - \theta_0| \geq \varepsilon\Big). \qquad (13.1.6)$$

Причём, для второго слагаемого справедлива оценка

$$\mathsf{P}_{n\theta_0}\Big(\sqrt{n}(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) < x,\ |\tilde{\delta}_n(\mathbf{X}_n) - \theta_0| \geq \varepsilon\Big) \leq \mathsf{P}_{n\theta_0}\Big(|\tilde{\delta}_n(\mathbf{X}_n) - \theta_0| \geq \varepsilon\Big)$$

и правая часть этого неравенства стремится к нулю в силу Теоремы 13.1.2. Таким образом из равенства (13.1.6) следует, что достаточно показать, что

$$\mathsf{P}_{n\theta_0}\Big(-\sqrt{n}\varepsilon < \sqrt{n}(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) < x\Big) \to \Phi\Big(\sqrt{I(\theta_0)}x\Big), \qquad n \to \infty. \qquad (13.1.7)$$

Пусть $\tilde{\delta}_n(\mathbf{X}_n) \in (\theta_0 - \varepsilon, \theta_0 + \varepsilon)$, тогда по формуле Тейлора можно записать

$$l'_n(\tilde{\delta}_n(\mathbf{X}_n); \mathbf{X}_n) = l'_n(\theta_0; \mathbf{X}_n) + (\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) l''_n(\theta_0; \mathbf{X}_n) + \frac{1}{2}(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0)^2 l'''_n(\theta_n^*; \mathbf{X}_n),$$

где θ_n^* лежит между θ_0 и $\tilde{\delta}_n(\mathbf{X}_n)$. По предположению левая часть этого равенства равна нулю, поэтому

$$\sqrt{n}\Big(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0\Big) = \frac{-\frac{1}{\sqrt{n}} l'_n(\theta_0;\ \mathbf{X}_n)}{\frac{1}{n} l''_n(\theta_0;\ \mathbf{X}_n) + \frac{1}{2}(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0)\frac{1}{n} l'''_n(\theta_n^*;\ \mathbf{X}_n)}. \qquad (13.1.8)$$

Рассмотрим числитель дроби (13.1.8). Он имеет вид суммы независимых одинаково распределённых случайных величин

$$-\frac{1}{\sqrt{n}} l'_n(\theta_0;\ \mathbf{X}_n) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{p'_{\theta_0}(X_i)}{p_{\theta_0}(X_i)},$$

поэтому из Центральной Предельной Теоремы (см. Лекция 4, п. 5) следует слабая сходимость

$$-\frac{1}{\sqrt{n}} l'_n(\theta_0;\ \mathbf{X}_n) \Longrightarrow \mathcal{N}(0,\ I(\theta_0)), \qquad n \to \infty. \qquad (13.1.9)$$

Рассмотрим теперь знаменатель выражения (13.1.8). Имеем

$$\frac{1}{n} l''_n(\theta_0;\ \mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log p_{\theta_0}(X_i)}{\partial \theta^2},$$

поэтому из Закона Больших Чисел (см. Лекция 4, п. 6) и второго утверждения Теоремы 11.1.3 следует сходимость

$$\frac{1}{n} l_n''(\theta_0;\, \mathbf{X}_n) \xrightarrow{\mathsf{P}_{n\theta_0}} \mathsf{E}_{\theta_0} \frac{\partial^2 \log p_{\theta_0}(X_1)}{\partial \theta^2} = -I(\theta_0), \qquad n \to \infty. \tag{13.1.10}$$

Рассмотрим теперь последнее слагаемое в знаменателе выражения (13.1.8). Справедлива оценка

$$\frac{1}{2n} |(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) l_n'''(\theta_n^*;\, \mathbf{X}_n)| \ \le \ \frac{1}{n} |\tilde{\delta}_n(\mathbf{X}_n) - \theta_0| \sum_{i=1}^{n} \left| \frac{\partial^3 \log p_{\theta_n^*}(X_i)}{\partial \theta^3} \right| \ \le$$

$$\le \ |\tilde{\delta}_n(\mathbf{X}_n) - \theta_0| \, \frac{1}{n} \sum_{i=1}^{n} M(X_i).$$

Поскольку по условию Теоремы $\tilde{\delta}_n(\mathbf{X}_n)$ состоятельна, то с учётом Закона Больших Чисел имеем

$$|\tilde{\delta}_n(\mathbf{X}_n) - \theta_0| \, \frac{1}{n} \sum_{i=1}^{n} M(X_i) \xrightarrow{\mathsf{P}_{n\theta_0}} 0 \cdot \mathsf{E}_{\theta_0} M(X_1) = 0,$$

поэтому

$$\frac{1}{2n} |(\tilde{\delta}_n(\mathbf{X}_n) - \theta_0) l_n'''(\theta_n^*;\, \mathbf{X}_n)| \xrightarrow{\mathsf{P}_{n\theta_0}} 0. \tag{13.1.11}$$

Теперь из соотношений (13.1.8) – (13.1.11) следует утверждение Теоремы. □

Следствие 13.1.1. Если в предположениях Теоремы 13.1.3 уравнение правдоподобия (13.1.5) имеет единственный корень при всех n и $\mathbf{x}_n$, или, более общо, если вероятность наличия нескольких корней стремится к нулю при $n \to \infty$, то оценка максимального правдоподобия $\hat{\delta}_n(\mathbf{X}_n)$ асимптотически нормальна

$$\mathsf{P}_{n\theta_0}\Big(\sqrt{n}(\hat{\delta}_n(\mathbf{X}_n) - \theta_0) < x\Big) \to \Phi\Big(\sqrt{I(\theta_0)}x\Big), \qquad n \to \infty.$$

13.2 СПИСОК ЛИТЕРАТУРЫ

1) Э. Леман, Теория Точечного Оценивания,
Москва, Наука, 1991, Глава 6, ¡ 2.

2) И.А. Ибрагимов, Р.З. Хасьминский, Асимптотическая Теория Оценивания,
Москва, Наука, 1979, Глава 1, ¡ 4.

Лекция 14

В Лекции рассматривается случай, когда в качестве оцениваемого параметра выступает неизвестная плотность наблюдений.

14.1 ОЦЕНКА ПЛОТНОСТИ

Пусть наблюдения имеют вид $\mathbf{X}_n = (X_1, \cdots, X_n), \quad n \in \mathbf{N}$, где X_i – независимы и одинаково распределены с общей неизвестной плотностью $p(x)$. Если плотность $p(x)$ зависит от *конечного* числа параметров и является известной функцией x и этих параметров, то мы снова приходим к задаче *параметрического оценивания*. Если, однако, известно лишь, что плотность $p(x)$ принадлежит некоторому достаточно обширному множеству функций, то задача оценивания $p(x)$ становится *бесконечномерной* или *непараметрической*.

Будем исходить из естественной оценки функции распределения $F(x)$ – *эмпирической функции распределения* $F_n(x)$ (см. Лекция 6, пример 2)

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty,\, x)}(X_i), \qquad F(x) = \int\limits_{-\infty}^{x} p(y)dy. \tag{14.1.1}$$

При большом n эмпирическая функция распределения $F_n(x)$, в силу Закона Больших Чисел (см. Лекция 4, п. 6), близка к истинной функции распределения $F(x)$, поэтому можно было бы ожидать, что её производная $F_n'(x)$ близка к $p(x) = F'(x)$. Однако

$$F_n'(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - X_i),$$

где $\delta(x)$ – δ - функция Дирака – не является даже функцией в смысле классического анализа. Естественно "сгладить" эмпирическую функцию распределения $F_n(x)$ и использовать в качестве оценки для плотности $p(x)$ производную от такой сглаженной функции. Таким образом, обычно рассматривают оценки вида

$$p_n(x,\, \mathbf{X}_n) = \frac{1}{nh_n}\sum_{i=1}^{n} V\left(\frac{x - X_i}{h_n}\right), \tag{14.1.2}$$

где функция $V(x)$ интегрируема и удовлетворяет условию

$$\int\limits_{-\infty}^{\infty} V(x)dx = 1, \tag{14.1.3}$$

а последовательность h_n такова, что

$$h_n \to 0, \quad nh_n \to \infty, \quad n \to \infty. \tag{14.1.4}$$

Поясним на эвристическом уровне эти условия. Заметим, что естественным условием на оценку $p_n(x,\, \mathbf{X}_n)$ было бы требование

$$\mathsf{E}_p p_n(x,\, \mathbf{X}_n) \to p(x), \quad n \to \infty. \tag{14.1.5}$$

Поэтому, если выполнены условия (14.1.3) и (14.1.4), то при больших n

$$\mathsf{E}_p p_n(x,\, \mathbf{X}_n) = \frac{1}{h_n}\mathsf{E}_p V\left(\frac{x - X_1}{h_n}\right) = \frac{1}{h_n}\int V\left(\frac{x-y}{h_n}\right)p(y)dy =$$

$$= \int V(z)p(x - h_n z)dz \approx p(x)\int V(z)dz = p(x).$$

Заметим, что при этом мы не использовали второе условие из (14.1.4).

Конечно, сходимость

$$p_n(x,\, \mathbf{X}_n) \to p(x), \qquad n \to \infty$$

в том или ином смысле имеет место лишь при некоторых ограничениях на плотность $p(x)$. Если, например, $p(x)$ имеет точки разрыва, то сходимость не может быть равномерой ни при каком выборе h_n и $V(x)$. Если заранее известно, что $p(x)$ принадлежит тому или иному классу непрерывных функций, то в классе оценок (14.1.2) можно найти оценки, сходящиеся к $p(x)$ с той или иной скоростью.

Пусть, например, заранее известно, что плотность $p(x)$ принадлежит множеству функций, удовлетворяющих условию Липшица с постоянной $D > 0$

$$p(x) \in \mathbf{L}(1,\ D), \tag{14.1.6}$$

где

$$\mathbf{L}(1,\ D) = \Big\{ p(x)\colon |p(x_1) - p(x_2)| \ \leq \ D|x_1 - x_2|,$$
$$\text{для любых} \quad x_1,\ x_2 \quad \text{и} \quad p(x) - \text{плотность} \Big\}.$$

Тогда можно доказать следующую Теорему.

Теорема 14.1.1. *Если $h_n = n^{-1/3}$, функции $xV(x)$, $V^2(x)$ интегрируемы и выполнено условие (14.1.3), то для оценки (14.1.2) справедливо неравенство*

$$\sup_{p(x)\in\mathbf{L}(1,\ D)} \sup_{x\in\mathbf{R}^1} \mathsf{E}_p\Big(p_n(x,\ \mathbf{X}_n) - p(x)\Big)^2 \ \leq \ Cn^{-2/3}, \quad C > 0.$$

Доказательство. Пусть h_n таково, что выполнены условия (14.1.4). Тогда заметим, что

$$\mathsf{E}_p\Big(p_n(x,\ \mathbf{X}_n) - p(x)\Big)^2 = \Big(\mathsf{E}_p p_n(x,\ \mathbf{X}_n) - p(x)\Big)^2 + \mathsf{D}_p p_n(x,\ \mathbf{X}_n). \tag{14.1.7}$$

Рассмотрим первое слагаемое в этом равенстве, имеем

$$\Big|\mathsf{E}_p p_n(x,\ \mathbf{X}_n) - p(x)\Big| = \left|\int V(z)(p(x - h_n z) - p(x))dz\right| \ \leq$$
$$\leq \ Dh_n \int |V(z)z|dz = O(h_n). \tag{14.1.8}$$

Рассмотрим второе слагаемое в (14.1.7)

$$\mathsf{D}_p p_n(x,\ \mathbf{X}_n) = \frac{1}{nh_n^2}\mathsf{D}_p V\left(\frac{x - X_1}{h_n}\right) \ \leq \ \frac{1}{nh_n^2}\mathsf{E}_p V^2\left(\frac{x - X_1}{h_n}\right) =$$
$$= \frac{1}{nh_n}\int V^2(z)p(x - h_n z)dz \leq \frac{1}{nh_n}\sup_{x\in\mathbf{R}^1} p(x)\int V^2(z)dz = O\left(\frac{1}{nh_n}\right) \tag{14.1.9}$$

(здесь используется условие $nh_n \to \infty$). Из соотношений (14.1.7) – (14.1.9) следует, что равномерно по $x \in \mathbf{R}^1$

$$\mathsf{E}_p\Big(p_n(x,\ \mathbf{X}_n) - p(x)\Big)^2 = O(h_n^2) + O\left(\frac{1}{nh_n}\right).$$

Минимизируя правую часть этого выражения по h_n, получим, что $h_n = O(n^{-1/3})$ и выполнено неравенство из формулировки Теоремы. □

Рассмотрим теперь обобщение Теоремы 14.1.1 на другие семейства плотностей $p(x)$. Мы увидим, в частности, что для семейств $p(x)$, удовлетворяющих более жёстким условиям гладкости, среди оценок (14.1.2) найдутся такие, которые сходятся к $p(x)$ быстрее, причём скорость сходимости существенно зависит от степени гладкости.

Обозначим через $\mathbf{L}(\beta,\, D)$, $\beta = k + \alpha,\ k \in \{0, 1, \cdots\},\ 0 < \alpha \le 1$ множество k раз дифференцируемых плотностей таких, что их k – я производная удовлеторяет условию Гёльдера с показателем α и константой $D > 0$

$$\mathbf{L}(\beta,\, D) = \Big\{ p(x)\colon |p^{(k)}(x_1) - p^{(k)}(x_2)| \;\le\; D|x_1 - x_2|^{\alpha},$$

$$\text{для любых} \quad x_1,\ x_2 \quad \text{и} \quad p(x) \text{ – плотность} \Big\}.$$

Чтобы не оговаривать каждый раз условия сходимости соответсвующих интегралов, ограничимся изучением оценок вида (14.1.2) с *финитными* функциями $V(x)$.

Теорема 14.1.2. *Если $h_n = n^{-1/(2\beta+1)}$, $\beta = k + \alpha$ и ограниченная финитная функция $V(x)$ удовлетворяет условиям (14.1.3) и*

$$\int x^j V(x)dx = 0, \qquad j = 1, \cdots, k,$$

то для оценки (14.1.2) при любом $D > 0$ справедливо неравенство

$$\sup_{p(x)\in \mathbf{L}(\beta,\, D)} \sup_{x \in \mathbf{R}^1} \mathsf{E}_p \Big(p_n(x,\, \mathbf{X}_n) - p(x) \Big)^2 \;\le\; C n^{-2\beta/(2\beta+1)}.$$

Доказательство. Как и при доказательстве Теоремы 14.1.1, оценим отдельно смещение и дисперсию оценки $p_n(x,\, \mathbf{X}_n)$. Используя условия Теоремы и формулу Тейлора, имеем

$$\mathsf{E}_p p_n(x,\, \mathbf{X}_n) - p(x) = \int V(z)(p(x - h_n z) - p(x))dz =$$

$$= \frac{h_n^k}{k!} \int z^k V(z)(p^{(k)}(\xi) - p^{(k)}(x))dz,$$

где ξ – точка интервала $(x, x - h_n z)$. Поэтому из определения множества $\mathbf{L}(\beta,\, D)$ следует, что для некоторой постоянной $C_1 > 0$ справедливо неравенство

$$\Big| \mathsf{E}_p p_n(x,\, \mathbf{X}_n) - p(x) \Big| \;\le\; C_1 h_n^{\beta}.$$

При этом справедливо соотношение (14.1.9), поэтому

$$\mathsf{E}_p\Big(p_n(x,\ \mathbf{X}_n)-p(x)\Big)^2=O(h_n^{2\beta})+O\left(\frac{1}{nh_n}\right).$$

Минимизируя правую часть этого выражения по h_n, получим, что $h_n = O\left(n^{-1/(2\beta+1)}\right)$ и выполнено утверждение Теоремы. □

14.2 ПРОЕКЦИОННЫЕ ОЦЕНКИ

Пусть наблюдения имеют вид $\mathbf{X}_n=(X_1,\cdots,X_n)$, $n\in\mathbf{N}$, где X_i – независимы и одинаково распределены с общей неизвестной плотностью $p(x)$. Рассмотрим другой метод оценки неизвестной плотности $p(x)$. Для простоты будем считать плотность заданной на отрезке $[-\pi,\pi]$. Тогда плотности $p(x)$ можно сопоставить *ряд Фурье* по тригонометрической системе

$$p(x)\sim\frac{1}{2\pi}+\sum_{m=1}^{\infty}(a_m\cos mx+b_m\sin mx), \tag{14.2.1}$$

где коэффициенты Фурье a_m и b_m имеют вид

$$a_m=\frac{1}{\pi}\int_{-\pi}^{\pi}\cos(mx)p(x)dx=\frac{1}{\pi}\mathsf{E}_p\cos(mX_1),$$

$$b_m=\frac{1}{\pi}\int_{-\pi}^{\pi}\sin(mx)p(x)dx=\frac{1}{\pi}\mathsf{E}_p\sin(mX_1),\quad m=1,2,\cdots \tag{14.2.2}$$

Из этих формул видно, что *несмещенными* оценками коэффициентов Фурье являются, например, оценки

$$a_{mn}(\mathbf{X}_n)=\frac{1}{n\pi}\sum_{i=1}^{n}\cos(mX_i),\qquad b_{mn}(\mathbf{X}_n)=\frac{1}{n\pi}\sum_{i=1}^{n}\sin(mX_i),$$

$$\mathsf{E}_p a_{mn}(\mathbf{X}_n)=a_m,\qquad \mathsf{E}_p b_{mn}(\mathbf{X}_n)=b_m,\qquad m=1,2,\cdots \tag{14.2.3}$$

Если ряд Фурье (14.2.1) *сходится* к плотности $p(x)$, то при больших n справедлива аппроксимация

$$S_n(x)\approx p(x),\quad n\gg 1,$$

где

$$S_n(x) = \frac{1}{2\pi} + \sum_{m=1}^{n} (a_m \cos mx + b_m \sin mx) \qquad (14.2.4)$$

частичная сумма ряда Фурье. Таким образом выбирая некоторую последовательность натуральных чисел $k_n \to \infty$ и заменяя коэффициенты Фурье в формуле (14.2.4) их оценками из (14.2.3), получаем *проекционную* оценку $\bar{p}_n(x,\, \mathbf{X}_n)$ плотности $p(x)$

$$\bar{p}_n(x,\, \mathbf{X}_n) = \frac{1}{2\pi} + \sum_{m=1}^{k_n} \Big(a_{mn}(\mathbf{X}_n) \cos mx + b_{mn}(\mathbf{X}_n) \sin mx \Big). \qquad (14.2.5)$$

Заметим, что

$$\mathsf{E}_p \bar{p}_n(x,\, \mathbf{X}_n) = S_{k_n}(x) \approx p(x), \qquad k_n \gg 1.$$

Теорема 14.2.1. *Пусть плотность $p(x)$ задана на отрезке $[-\pi, \pi]$,*

$$p(\pi) = p(-\pi), \qquad \int_{-\pi}^{\pi} p^2(x) dx < \infty$$

и $k_n \to \infty$, тогда справедливо неравенство

$$\mathsf{E}_p \int_{-\pi}^{\pi} \Big(\bar{p}_n(x,\, \mathbf{X}_n) - p(x) \Big)^2 dx \ \le\ \frac{k_n^2}{n\pi^2} + \sum_{m=k_n+1}^{\infty} (a_m^2 + b_m^2).$$

Доказательство. Заметим сначала, что

$$\mathsf{E}_p \int_{-\pi}^{\pi} \Big(\bar{p}_n(x,\, \mathbf{X}_n) - p(x) \Big)^2 dx = \mathsf{E}_p \int_{-\pi}^{\pi} \Big(\bar{p}_n(x,\, \mathbf{X}_n) - \mathsf{E}_p \bar{p}_n(x,\, \mathbf{X}_n) \Big)^2 dx +$$

$$+ \int_{-\pi}^{\pi} \Big(S_{k_n}(x) - p(x) \Big)^2 dx. \qquad (14.2.6)$$

Второе слагаемое в правой части (14.2.6) в силу равенства Бесселя равно

$$\int_{-\pi}^{\pi} \Big(S_{k_n}(x) - p(x) \Big)^2 dx = \sum_{m=k_n+1}^{\infty} (a_m^2 + b_m^2). \qquad (14.2.7)$$

Рассмотрим первое слагаемое в правой части (14.2.6)

$$\mathsf{E}_p \int\limits_{-\pi}^{\pi} \Big(\bar{p}_n(x,\, \mathbf{X}_n) - \mathsf{E}_p \bar{p}_n(x,\, \mathbf{X}_n)\Big)^2 dx = \int\limits_{-\pi}^{\pi} \mathsf{D}_p \bar{p}_n(x,\, \mathbf{X}_n) dx. \tag{14.2.8}$$

Но

$$\mathsf{D}_p \bar{p}_n(x,\, \mathbf{X}_n) = \frac{1}{n^2\pi^2} \mathsf{D}_p \sum_{i=1}^{n} \sum_{m=1}^{k_n} \Big(\cos(mX_i)\cos(mx) + \sin(mX_i)\sin(mx)\Big) =$$

$$= \frac{1}{n\pi^2} \mathsf{D}_p \sum_{m=1}^{k_n} \Big(\cos(mX_1)\cos(mx) + \sin(mX_1)\sin(mx)\Big) \ \le$$

$$\le \ \frac{1}{n\pi^2} \mathsf{E}_p \left(\sum_{m=1}^{k_n} \cos(m(X_1 - x)) \right)^2 \ \le \ \frac{k_n^2}{n\pi^2}. \tag{14.2.9}$$

Из соотношений (14.2.6) – (14.2.9) следует утверждение Теоремы. □

Замечание 14.2.1. Заметим, что из доказанной Теоремы следует, что для получения разумных оценок $\bar{p}_n(x,\, \mathbf{X}_n)$ плотности $p(x)$, последовательность k_n следует выбирать растущей не быстрее чем $\sqrt{n}$. Слагаемое вида

$$\sum_{m=k_n+1}^{\infty} (a_m^2 + b_m^2)$$

характеризует степень *гладкости* плотности $p(x)$, например, если у плотнотси $p(x)$ существует $r \ge 1$ непрерывных производных и $r-1$ производных принимают на концах отрезка $[-\pi, \pi]$ равные значения, то этот член имеет вид $O\left(\frac{1}{k_n^{2r-1}}\right)$.

14.3 СПИСОК ЛИТЕРАТУРЫ

1) И.А. Ибрагимов, Р.З. Хасьминский, Асимптотическая Теория Оценивания,
Москва, Наука, 1979, Глава 4, ¡4; Глава 7, ¡4.

2) Л. Деврой, Л. Дьёрфи, Непараметрическое Оценивание Плотности,
Москва, Мир, 1988, Глава 12, ¡1 – ¡4.

3) Д. Джексон, Ряды Фурье и Ортогональные Полиномы,
Москва, Иностранная Литература, 1948, Главы 1 и 2.

Лекция 15

В Лекции рассматриваются достаточные статистики, которые редуцируют данные в максимальной степени.

15.1 МИНИМАЛЬНЫЕ ДОСТАТОЧНЫЕ СТАТИСТИКИ

Напомним, что достаточные статистики (см. Определение 7.1.2) сокращают наблюдения без потери информации. Ясно, что эквивалентные формы достаточной статистики редуцируют данные в одной и той же степени. Могут, однако, существовать также достаточные статистики которые дают различные степени редукции. Пусть, например, $X = (X_1, \cdots, X_n)$ – независимые нормально распределённые наблюдения с неизвестной дисперсией

$$X_i \sim \mathcal{N}(0,\, \theta^2), \quad i = 1, \cdots, n.$$

Тогда из критерия факторизации (см. Теорему 7.1.3) следует, что следующие статистики являются достаточными

$$T_1(X) = (X_1, \cdots, X_n), \;\; T_2(X) = (X_1^2, \cdots, X_n^2),$$

$$T_3(X) = (X_1^2 + \cdots + X_m^2, X_{m+1}^2 + \cdots + X_n^2), \quad T_4(X) = X_1^2 + \cdots + X_n^2.$$

Причём степень редукции данных статистиками $T_i(X)$ возрастает с ростом i.

Из критерия факторизации следует также, что если $T(X)$ достаточная статистика и $T(X) = H(S(X))$, где $S(X)$ некоторая статистика, а $H(\cdot)$ измеримая функция, то статистика $S(X)$ также достаточна. Знание $S(X)$ влечёт знание $T(X)$ и, следовательно, позволяет "восстанавливать" исходные данные. Кроме того, статистика $T(X)$ обеспечивает бо́льшую редукцию данных, чем статистика $S(X)$, если только функция $H(\cdot)$ *не является взаимно однозначной*, в противном случае статистики $T(X)$ и $S(X)$ являются *эквивалентными.* Достаточная статистика $T^*(X)$ называется *минимальной*, если она даёт наибольшую возможную редукцию данных среди всех достаточных статистик, то есть если для любой достаточной статистики $S(X)$ существует измеримая функция $H(\cdot)$ такая, что $T^*(X) = H(S(X))$ ($\mathcal{P}$ – п.в.). Дадим теперь формальное определение. Пусть дана доминируемая статистическая структура $(\mathcal{X},\ \mathcal{F},\ \mathcal{P})$.

Определение 15.1.1.

1) *Достаточная σ – подалгебра $\mathcal{D}^* \subseteq \mathcal{F}$ называется минимальной, если она содержится в любой другой достаточной σ – подалгебре.*

2) *Достаточная статистика $T^*(X)$ называется минимальной достаточной статистикой, если она индуцирует минимальную достаточную σ – подалгебру.*

Минимальная достаточная σ – подалгебра, в случае её существования, *единственна*; она является пересечением всех достаточных σ – подалгебр. Из второго определения следует, что если $T^*(X)$ вещественная минимальная достаточная статистика, а $S(X)$ другая вещественная достаточная статистика, то из Утверждения 1.1.2 следует, что существует измеримая функция $H(\cdot)$ такая, что $T^*(X) = H(S(X))$.

Теорема 15.1.1. *Пусть $\mathcal{P} = \{\mathsf{P}_0, \cdots, \mathsf{P}_k\}$ – конечное семейство распределений с плотностями $\{p_0(x), \cdots, p_k(x)\}$, имеющими общий носитель. Тогда статистика*

$$T^*(X) = \left(\frac{p_1(X)}{p_0(X)}, \cdots, \frac{p_k(X)}{p_0(X)}\right)$$

является минимальной достаточной статистикой.

Доказательство. Напомним, что из критерия факторизации (Теорема 7.1.3) следует, что статистика $S(X)$ достаточна тогда и только тогда, когда

$$p_\theta(x) = g_\theta(S(x))h(x) \quad \mathcal{P} - \text{п.в.}, \quad \theta \in \Theta, \quad x \in \mathcal{X}.$$

Поскольку по условию все плотности $\{p_0(x), \cdots, p_k(x)\}$ имеют общий носитель, то последнее равенство эквивалентно

$$\frac{p_i(x)}{p_0(x)} = g_i(S(x)) \quad i = 1, \cdots, k, \quad x \in A,$$

где A – общий носитель плотностей $\{p_0(x), \cdots, p_k(x)\}$ $(A = \{x \in \mathcal{X} : p_i(x) > 0, \quad i = 0, 1, \cdots, k.\})$. Отсюда следует, что статистика $T^*(X)$ является достаточной и статистика $T^*(X)$ есть функция от $S(X)$. □

Заметим, что из доказательства Теоремы 15.1.1 следует, что справедливо более общее

Утверждение 15.1.1. *Пусть $\mathcal{P} = \{\mathsf{P}_0, \cdots, \mathsf{P}_k\}$ – конечное семейство распределений с плотностями $\{p_0(x), \cdots, p_k(x)\}$, и пусть для каждого $x \in \mathcal{X}$ множество $A(x)$ имеет вид*

$$A(x) = \{(i, j) : \, p_i(x) + p_j(x) > 0, \quad (i, j) \in \{0, 1, \cdots, k\}^2\}.$$

Тогда статистика

$$T^*(X) = \left(\frac{p_j(X)}{p_i(X)}, \quad i < j, \, (i, j) \in A(X)\right)$$

является минимальной достаточной статистикой. Здесь $p_j(x)/p_i(x) = \infty$, если $p_i(x) = 0$ и $p_j(x) > 0$.

Следующая Теорема часто является полезной при практическом нахождении минимальных достаточных статистик.

Теорема 15.1.2. *Пусть $\mathcal{P}$ – семейство распределений с общим носителем, $\mathcal{P}_0 \subseteq \mathcal{P}$ и $T^*(X)$ – минимальная достаточная статистика для подсемейства $\mathcal{P}_0$ и является достаточной статистикой для семейства $\mathcal{P}$. Тогда $T^*(X)$ – минимальная достаточная статистика для семейства $\mathcal{P}$.*

Доказательство. Пусть $S(X)$ есть достаточная статистика для семейства $\mathcal{P}$, тогда она достаточна и для $\mathcal{P}_0$ и, следовательно σ – подалгебра, порождённая статистикой $T^*(X)$ содержится в σ – подалгебре, порождённой статистикой $S(X)$ (достаточная статистика $T^*(X)$ есть функция от $S(X)$). □

Замечание 15.1.1. *Заметим, что в Теореме 15.1.2 предположение об общем носителе можно заменить на более слабое предположение о том, что каждое $\mathcal{P}_0$ – нулевое множество (см. Определение 6.1.4) является также и $\mathcal{P}$ – нулевым, так что $\mathcal{P}_0$ – п.в. эквивалентно $\mathcal{P}$ – п.в.*

Примеры.

1) Пусть имеется n независимых нормально распределённых наблюдения $X = (X_1, \cdots, X_n)$ и семейство $\mathcal{P}_0^n$ – n - кратное произведение семейств $\mathcal{P}_0$, состоит из двух нормальных распределений

$$\mathcal{P}_0^n = \{\mathcal{N}^n(\theta_0, \, 1), \quad \mathcal{N}^n(\theta_1, \, 1), \quad \theta_0 \neq \theta_1\}.$$

Тогда по Теореме 15.1.1 минимальная достаточная статистика $T^*(X)$ для семейства $\mathcal{P}_0^n$ имеет вид

$$T^*(X) = \frac{p_{\theta_1}(X)}{p_{\theta_0}(X)},$$

что эквивалентно

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

Однако, по критерию факторизации статистика $\overline{X}$ является достаточной статистикой для семейства

$$\mathcal{P}^n = \{\mathcal{N}^n(\theta,\, 1), \theta \in \mathbf{R}^1\},$$

поэтому по Теореме 15.1.2 минимальная достаточная статистика для семейства $\mathcal{P}^n$ есть

$$T^*(X) = \overline{X}.$$

2) Пусть теперь $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения, каждое из которых имеет распределение Пуассона с параметром $\theta > 0$. Рассмотрим семейство $\mathcal{P}_0^n$, состоящее из двух пуассоновских распределений

$$\mathcal{P}_0^n = \{\mathcal{P}^n(\theta_0),\ \mathcal{P}^n(\theta_1), \quad \theta_0 \neq \theta_1\}.$$

Тогда по Теореме 15.1.1 минимальная достаточная статистика $T^*(X)$ для семейства $\mathcal{P}_0^n$ имеет вид (здесь выбрана считающая доминирующая мера ν)

$$T^*(X) = \frac{p_{\theta_1}(X)}{p_{\theta_0}(X)} = \frac{e^{-n\theta_1}\frac{\theta_1^{X_1+\cdots+X_n}}{X_1!\cdots X_n!}}{e^{-n\theta_0}\frac{\theta_0^{X_1+\cdots+X_n}}{X_1!\cdots X_n!}} = e^{n(\theta_0-\theta_1)}\left(\frac{\theta_1}{\theta_0}\right)^{n\overline{X}},$$

что эквивалентно

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

Аналогично предыдущему минимальная достаточная статистика для семейства

$$\mathcal{P}^n = \{\mathcal{P}^n(\theta),\ \theta > 0\}$$

есть

$$T^*(X) = \overline{X}.$$

Теорема 15.1.3 (Существование минимальной достаточной σ – подалгебры) *Пусть* $(\mathcal{X},\, \mathcal{F},\, \mathcal{P} = \{\mathsf{P}_\theta,\, \theta \in \Theta\})$ *– доминируемая статистическая структура,* P^* *– привилегированное вероятностное распределение и* ν *– доминирующая мера. Тогда* σ *– подалгебра* $\mathcal{D}^* \subseteq \mathcal{F}$*, порожденная (то есть минимальная) функциями вида*

$$r(x,\, \theta) = \frac{p_\theta(x)}{p^*(x)} \quad \text{\textit{при всех}} \quad \theta \in \Theta,$$

где

$$p_\theta(x) = \frac{d\mathsf{P}_\theta}{d\nu}(x), \quad p^*(x) = \frac{d\mathsf{P}^*}{d\nu}(x), \quad \theta \in \Theta,$$

является минимальной достаточной σ *– подалгеброй.*

Доказательство. Докажем сначала, что σ – подалгебра $\mathcal{D}^*$ достаточна. Это следует из Теоремы 7.1.2, поскольку

$$\frac{d\mathsf{P}_\theta}{d\mathsf{P}^*}(x) = \frac{\frac{d\mathsf{P}_\theta}{d\nu}(x)}{\frac{d\mathsf{P}^*}{d\nu}(x)} = \frac{p_\theta(x)}{p^*(x)} = r(x,\, \theta)$$

и по определению σ – подалгебры $\mathcal{D}^*$ эта плотность $\mathcal{D}^*$ – измерима при всех $\theta \in \Theta$.

Докажем, что σ – подалгебра $\mathcal{D}^*$ – минимальна, с этой целью рассмотрим любую другую достаточную σ – подалгебру $\mathcal{B} \subseteq \mathcal{F}$, тогда опять по Теореме 7.1.2 плотность

$$\frac{d\mathsf{P}_\theta}{d\mathsf{P}^*}(x) = \frac{p_\theta(x)}{p^*(x)} = r(x,\, \theta)$$

$\mathcal{B}$ – измерима, и поэтому по определению σ –подалгебры $\mathcal{D}^*$, справедливо включение $\mathcal{D}^* \subseteq \mathcal{B}$. $\square$

Рассмотрим теперь связь между минимальными и полными (см. Определение 10.1.1) достаточными статистиками.

Теорема 15.1.4 (Связь минимальности и полноты) *Любая полная вещественная достаточная статистика* $T(X)$*, заданная на доминируемой статистической структуре* $(\mathcal{X},\, \mathcal{F},\, \mathcal{P} = \{\mathsf{P}_\theta,\, \theta \in \Theta\})$ *является также и минимальной достаточной статистикой.*

Доказательство. Пусть $\mathcal{D}^*$ – минимальная достаточная σ – подалгебра. По предыдущей Теореме она существует. Предположим, что у статистики $T(X)$ существует математическое ожидание

$$\mathsf{E}_\theta T(X) < \infty, \quad \theta \in \Theta$$

и рассмотрим функцию

$$h(X) = T(X) - \mathsf{E}_\theta(T(X) \mid \mathcal{D}^*).$$

Тогда в силу достаточности σ – подалгебры $\mathcal{D}^*$ эта функция не зависит от θ, и с учётом минимальности $\mathcal{D}^*$ эта функция является измеримой относительно σ – алгебры, порождённой статистикой $T(X)$, поэтому в силу Утверждения 1.1.2 она имеет вид $h(x) = \bar{h}(T(x))$. Но из определения $h(x)$ следует, что

$$\mathsf{E}_\theta \bar{h}(T) \equiv 0, \quad \theta \in \Theta,$$

поэтому с учётом полноты статистики $T(X)$, имеем

$$\bar{h}(T(X)) \equiv 0, \quad \mathcal{P} - \text{п.в.}$$

Итак

$$T(X) \equiv \mathsf{E}_\theta(T(X) \mid \mathcal{D}^*), \quad \mathcal{P} - \text{п.в.},$$

поэтому статистика $T(X)$ является $\mathcal{D}^*$ – измеримой и значит σ – алгебра, порождённая статистикой $T(X)$ совпадает с $\mathcal{D}^*$ (здесь под $\mathcal{D}^*$ следует понимать σ – алгебру, пополненную множествами N, для которых $\mathsf{P}_\theta(N) = 0$ при всех $\theta \in \Theta$).

Если $\mathsf{E}_\theta T(X)$ не существует, то надо вместо статистики $T(X)$ рассмотреть, например, статистику $\arctan T(X)$, которая, очевидно, эквивалентна $T(X)$ относительно свойств достаточности, полноты и минимальности. □

Пример 15.1.1. Пусть $X = (X_1, \cdots, X_n)$ – независимые равномерно распределённые наблюдения

$$X_i \sim \mathcal{R}(0,\, \theta), \quad i = 1, \cdots, n; \quad \theta > 0.$$

Тогда в Примере 10.2.2 показано, что статистика

$$T(X) = X_{(n)} \equiv \max_{1 \le i \le n} X_i$$

является полнай достаточной статистикой. По Теореме 15.1.4 эта статистика является *минимальной* достаточной статистикой.

Заметим, однако, что в общем случае утверждение, обратное Теореме 15.1.4 *не верно*, то есть из минимальности достаточной статистики *не следует* её полнота.

Задача 15.1.1. Пусть $X = (X_1, \cdots, X_n)$ – независимые равномерно распределённые наблюдения

$$X_i \sim \mathcal{R}(\theta - 1/2,\, \theta + 1/2), \quad i = 1, \cdots, n; \quad \theta \in \Theta = \mathbf{R}^1.$$

Доказать, что статистика

$$T(X) = (X_{(1)},\, X_{(n)}) = (\min_{1\le i\le n} X_i,\ \max_{1\le i\le n} X_i)$$

является минимальной достаточной статистикой, но она не полна. (Последнее утверждение следует, например, из тождества

$$\mathsf{E}_\theta\Big(X_{(n)} - X_{(1)} - (n-1)/(n+1)\Big) \equiv 0, \qquad \theta \in \Theta.)$$

15.2 СПИСОК ЛИТЕРАТУРЫ

1) Э. Леман, Теория Точечного Оценивания, Москва, Наука, 1991, Глава 1, ¡5.

2) А.А. Боровков, Математическая Статистика, Москва, Наука, 1984, Глава 2, ¡13.

3) Ж.– Р. Барра, Основные Понятия Математической Статистики, Москва, Мир, 1974, Глава 2, ¡5.

Лекция 16

В Лекции рассматриваются экспоненциальные структуры, для которых многие общие конструкции математической статистики можно реализовать в явном виде.

16.1 ЭКСПОНЕНЦИАЛЬНЫЕ СТРУКТУРЫ

Определение 16.1.1.

1) *Статистическая доминируемая структура*

$$\left(\mathcal{X} = \mathbf{R}^m,\ \mathcal{F} = \mathcal{B}^m,\ \{p_\theta(x),\quad \theta \in \Theta \subseteq \mathbf{R}^k\}\right)$$

называется экспоненциальной, если носитель плотностей $p_\theta(x)$

$$A = \{x \in \mathcal{X}:\ p_\theta(x) > 0\}$$

не зависит от $\theta \in \Theta$ *и плотности* $p_\theta(x)$ *имеют вид*

$$p_\theta(x) = C(\theta)\exp\left\{\sum_{j=1}^{k} Q_j(\theta)U_j(x)\right\}h(x),\ \ \theta \in \Theta \subseteq \mathbf{R}^k,\ \ x \in \mathbf{R}^m, \quad (16.1.1)$$

где все функции, входящие в правую часть, конечны и измеримы, а функции $Q_0(\theta) \equiv 1, Q_1(\theta), \cdots, Q_k(\theta)$ *линейно независимы на* Θ.

2) *Семейство распределений* $\{\mathsf{P}_\theta, \theta \in \Theta \subseteq \mathbf{R}^k\}$ *с плотностями* $p_\theta(x)$ *вида (16.1.1) называется экспоненциальным семейством.*

Пример 16.1.1. Пусть X имеет гамма – распределение с параметром $\theta = (\alpha,\, \lambda)$, $\alpha > 0$, $\lambda > 0$, то есть

$$p_\theta(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-x\alpha} = \frac{\alpha^\lambda}{\Gamma(\lambda) x} e^{\lambda \log x - \alpha x}, \qquad x > 0,$$

$$\Gamma(\lambda) = \int\limits_0^\infty x^{\lambda-1} e^{-x} dx,$$

то есть здесь имеем

$$h(x) = \begin{cases} x^{-1}, & x > 0, \\ 0, & x < 0, \end{cases}$$

$$U_1(x) = \log x, \quad U_2(x) = x, \quad C(\theta) = \frac{\alpha^\lambda}{\Gamma(\lambda)}, \quad Q_1(\theta) = \lambda, \quad Q_2(\theta) = -\alpha.$$

Заметим, что биномиальное распределение, распределение Пуассона, отрицательное биномиальное распределение, нормальное распределение, бета распределение, хи – квадрат распределение (с параметром масштаба), гамма распределение (с параметром масштаба) являются экспоненциальными семействами. В то время как, например, равномерное распределение $\mathcal{R}(0,\, \theta)$, $\theta > 0$ и распределение Коши не являются экспоненциальными семействами. Заметим также, что полупрямое произведение экспоненциальных структур также является экспоненциальной структурой. То есть, если $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения, каждое из которых имеет плотность вида (16.1.1), то совместная плотность X также является экспоненциальной.

Иногда используют более естественную параметризацию включая множитель $h(x)$ в доминирующую меру ν, то есть относительно меры $d\tilde{\nu}(x) = h(x) d\nu(x)$ рассматривают семейства вида (каноническая форма экспоненциального семейства)

$$p_\theta(x) = C(\theta) \exp\left\{ \sum_{j=1}^k \theta_j U_j(x) \right\}, \qquad \theta \in \Theta \subseteq \mathbf{R}^k. \tag{16.1.2}$$

Правая часть (16.1.2), если её интеграл конечен, может быть надлежащим выбором $C(\theta)$ превращена в плотность вероятности. Множество Ω параметрических точек $\theta \in \Theta$, для которых это имеет место, называется *естественным параметрическим пространством* экспоненциального семейства

(16.1.2), то есть

$$\Omega = \Big\{\theta:\ \int \exp\Big(\sum_{j=1}^{k}\theta_j U_j(x)\Big) h(x)d\nu(x) < \infty\Big\}. \tag{16.1.3}$$

Теорема 16.1.1.

1) *Естественное параметрическое пространство Ω экспоненциального семейства (16.1.2) выпукло.*

2) *Для любой ограниченной измеримой функции $f(x)$ интеграл*

$$\int f(x)\exp\Big(\sum_{j=1}^{k}\theta_j U_j(x)\Big) h(x)d\nu(x),$$

рассматриваемый как функция комплексных переменных $\theta_j = \xi_j + i\eta_j$, $j = 1,\cdots,k$ является аналитической функцией по каждой из этих переменных в области $\overline{\Omega}$ параметрических точек, для которых $(\xi_1,\cdots,\xi_k)$ является внутренней точкой естественного параметрического пространства Ω. Производные любого порядка от этого интеграла по θ могут вычисляться дифференцированием под знаком интеграла.

3) *Если плотность $p_\theta(x)$ имеет вид (16.1.2) и функция $C(\theta)$ дважды дифференцируема, то справедливы равенства*

$$\mathsf{E}_\theta U_j(X) = -\frac{\partial \log C(\theta)}{\partial\theta_j}, \qquad \mathsf{Cov}_\theta(U_i(X),\, U_j(X)) = -\frac{\partial^2 \log C(\theta)}{\partial\theta_i\partial\theta_j}.$$

4) *Если плотность $p_\theta(x)$ имеет вид (16.1.2), то для любого θ из внутренности естественного параметрического пространства Ω в некоторой окрестности нуля существуют производящие функции моментов и семиинвариантов $M_\theta(s)$ и $K_\theta(s)$, $s \in \mathbf{R}^k$ статистики*

$$T(X) = \Big(U_1(X),\cdots,U_k(X)\Big)$$

и они имеют вид

$$M_\theta(s) = \mathsf{E}_\theta \exp\{s_1U_1(X)+\cdots+s_kU_k(X)\} = \frac{C(\theta)}{C(\theta+s)},$$

$$K_\theta(s) = \log M_\theta(s) = \log C(\theta) - \log C(\theta+s), \quad s = (s_1,\cdots,s_k).$$

5) *Пусть X имеет распределение с плотностью $p_\theta(x)$ вида (16.1.2), тогда существует мера $\bar\nu$ (зависящая от T) такая, что распределение статистики*

$$T(X) = \Big(U_1(X), \cdots, U_k(X)\Big)$$

также принадлежит экспоненциальному семейству вида

$$\bar p_\theta(t) = C(\theta)\exp\left\{\sum_{j=1}^k \theta_j t_j\right\}, \qquad \theta \in \Theta \subseteq \mathbf{R}^k, \qquad t = (t_1, \cdots, t_k) \in \mathbf{R}^k.$$

6) *Пусть X имеет распределение с плотностью $p_\theta(x)$ вида (16.1.2), тогда условное распределение статистики*

$$T_r(X) = \Big(U_1(X), \cdots, U_r(X)\Big), \quad r = 1, \cdots, k-1$$

при данном $\bar T_r(X) = t \in \mathbf{R}^{k-r}$, где

$$\bar T_r(X) = \Big(U_{r+1}(X), \cdots, U_k(X)\Big), \quad r = 1, \cdots, k-1$$

также принадлежит экспоненциальному семейству вида (относительно меры ν_t)

$$p_{\theta t}(u) = C_t(\theta)\exp\left\{\sum_{j=1}^r \theta_j u_j\right\}, \qquad \theta \in \Theta \subseteq \mathbf{R}^k, \qquad u = (u_1, \cdots, u_r) \in \mathbf{R}^r.$$

Доказательство.

1) Пусть $(\bar\theta_1, \cdots, \bar\theta_k)$ и $(\tilde\theta_1, \cdots, \tilde\theta_k)$ – две параметрические точки, принадлежащие множеству Ω, тогда из неравенства Гёльдера для любого $0 < \gamma < 1$ следует, что

$$\int \exp\Big(\sum_{j=1}^k (\gamma\bar\theta_j + (1-\gamma)\tilde\theta_j)U_j(x)\Big)h(x)d\nu(x) \ \le$$

$$\le \left(\int \exp\Big(\sum_{j=1}^k \bar\theta_j U_j(x)\Big)h(x)d\nu(x)\right)^\gamma \left(\int \exp\Big(\sum_{j=1}^k \tilde\theta_j U_j(x)\Big)h(x)d\nu(x)\right)^{1-\gamma} < \infty,$$

то есть точка $(\gamma\bar\theta_1 + (1-\gamma)\tilde\theta_1, \cdots, \gamma\bar\theta_k + (1-\gamma)\tilde\theta_k)$ также принадлежит Ω.

2) Рассматриваемый интеграл действительно существует при всех $(\xi_1, \cdots, \xi_k) \in \Omega$, поскольку, если $|f(x)| \le D$, $x \in \mathcal{X}$, то

$$\left|\int f(x) \exp\Big(\sum_{j=1}^{k} \theta_j U_j(x)\Big) h(x) d\nu(x)\right| \le D \int \exp\Big(\sum_{j=1}^{k} \xi_j U_j(x)\Big) h(x) d\nu(x) < \infty.$$

Докажем теперь аналитичность рассматриваемого интеграла, например, по θ_1. Выделяя в подинтегральном выражении действительную и мнимую части и разлагая каждую из последних на положительную и отрицательную части, включая затем надлежащие множители в меру ν получим, что требуемый результат достаточно доказать для интегралов вида

$$g(\theta_1) = \int \exp\Big(\theta_1 U_1(x)\Big) d\bar{\nu}(x).$$

Пусть $(\xi_1^*, \cdots, \xi_k^*) \in \Omega$ – некоторая *внутренняя* точка множества Ω, тогда существует $\delta > 0$ такое, что $g(\theta_1)$ конечна при всех θ_1 таких, что $|\theta_1 - \theta_1^*| < \delta$, $\theta_1^* = \xi_1^* + i\eta_1^*$. Рассмотрим разностное отношение

$$\frac{g(\theta_1) - g(\theta_1^*)}{\theta_1 - \theta_1^*} = \int \frac{\exp\{\theta_1 U_1(x)\} - \exp\{\theta_1^* U_1(x)\}}{\theta_1 - \theta_1^*} d\bar{\nu}(x) =$$

$$= \int \exp\{\theta_1^* U_1(x)\} \left(\frac{\exp\{(\theta_1 - \theta_1^*) U_1(x)\} - 1}{\theta_1 - \theta_1^*}\right) d\bar{\nu}(x).$$

Применим к подинтегральному выражению неравенство

$$\left|\frac{\exp\{az\} - 1}{z}\right| = \left|\int\limits_0^a \exp\{tz\} dt\right| \le$$

$$\le \int\limits_0^{|a|} \exp\{t\delta\} dt \le \frac{1}{\delta} e^{\delta|a|}, \qquad \text{при} \quad |z| \le \delta,$$

получим, что оно не превосходит

$$\frac{|\exp\{\theta_1^* U_1(x) + \delta |U_1(x)|\}|}{\delta} \le \frac{|\exp\{(\theta_1^* + \delta) U_1(x)\}| + |\exp\{(\theta_1^* - \delta) U_1(x)\}|}{\delta},$$

при $|\theta_1 - \theta_1^*| < \delta$. Поскольку последнее выражение интегрируемо относительно $\bar{\nu}$, то из Теоремы Лебега о мажорируемой сходимости (см.

Лекция 10, п.8) следует, что для любой последовательности точек $\theta_{1n} \in \bar{\Omega}$, сходящейся к θ_1^*, разностное отношение $g(\theta_{1n})$ стремится к

$$\int U_1(x)\exp\Big(\theta_1^* U_1(x)\Big)d\bar{\nu}(x).$$

Этим завершается доказательство первого утверждения и доказывается второе утверждение для случая первой производной. Доказательство для высших производных проводится аналогичным образом по индукции.

3) Доказательство непосредственно следует из возможности дифференцирования под знаком интеграла тождества

$$C(\theta)\int \exp\Big(\sum_{j=1}^k \theta_j U_j(x)\Big)h(x)d\nu(x) \equiv 1,$$

поскольку, например, имеем

$$\frac{\partial C(\theta)}{\partial \theta_j}\cdot\int \exp\Big(\sum_{i=1}^k \theta_i U_i(x)\Big)h(x)d\nu(x)+$$

$$+C(\theta)\int U_j(x)\exp\Big(\sum_{i=1}^k \theta_i U_i(x)\Big)h(x)d\nu(x) \equiv \frac{\partial \log C(\theta)}{\partial \theta_j} + \mathsf{E}_\theta U_j(X) \equiv 0.$$

4) Доказательство следует из соотношения

$$M_\theta(s) = \mathsf{E}_\theta \exp\{s_1 U_1(X) + \cdots + s_k U_k(X)\} =$$

$$= C(\theta)\int \exp\Big(\sum_{j=1}^k (s_j+\theta_j)U_j(x)\Big)h(x)d\nu(x) = \frac{C(\theta)}{C(\theta+s)},$$

справедливого в любой *внутренней* точке θ множества Ω и при изменении s в некоторой окрестности нуля.

5) Поскольку плотность $p_\theta(x)$ зависит от $T(x)$, то доказательство непосредственно следует из формулы замены переменного в интеграле Лебега (см. Лекция 2, п. 8)

$$\mathsf{P}_\theta(T(X)\in B) = \mathsf{P}_\theta(X \in T^{-1}(B)) =$$

$$= C(\theta)\int\limits_{T^{-1}(B)} \exp\left\{\sum_{j=1}^k \theta_j U_j(x)\right\}h(x)d\nu(x) = \int\limits_B \bar{p}_\theta(t)d\bar{\nu}(t).$$

6) Доказательство следует из Свойства 12 условных вероятностей Лекции 5 и доказанного пункта 5.

□

Утверждения 3 и 4 Теоремы 16.1.1 позволяют достаточно легко находить моменты и семиинварианты статистики

$$T(X) = \Big(U_1(X), \cdots, U_k(X)\Big).$$

Пример 16.1.2. Пусть X имеет распределение Пуассона с параметром $\lambda > 0$, то есть

$$X \sim \mathcal{P}(\lambda).$$

Тогда плотность $p_\lambda(x)$ относительно считающей меры равна

$$p_\lambda(x) = \frac{\lambda^x}{x!}e^{-\lambda} = e^{-\lambda}\exp\{x\log\lambda\}\frac{1}{x!}, \quad x = 0, 1, \cdots, \quad \lambda > 0.$$

Вводя новый параметр $\theta = \log\lambda$, получаем каноническое экспоненциальное семейство типа (16.1.2) с

$$k = 1, \ \ U_1(x) = x, \ \ C(\theta) = \exp\{-e^\theta\}$$

и, следовательно, производящие функции моментов и семиинварианто соответственно равны

$$M_\theta(s) = \mathsf{E}_\theta e^{sX} = \frac{C(\theta)}{C(\theta+s)} = \exp\{e^\theta(e^s-1)\} = \exp\{\lambda(e^s-1)\},$$

$$K_\theta(s) = \log M_\theta(s) = e^\theta(e^s-1) = \lambda(e^s-1),$$

так что, в частности, все семиинварианты (см. (3.1.11)) $\kappa_j(X)$ равны λ для всех $j \in \mathbf{N}$ и, например, для первых двух моментов справедливы равенства

$$\mathsf{E}X = \frac{\partial M_\theta(0)}{\partial s} = \lambda, \ \ \mathsf{E}X^2 = \frac{\partial^2 M_\theta(0)}{\partial s^2} = \lambda(1+\lambda).$$

Теорема 16.1.2.

1) *Если $X = (X_1, \cdots, X_n)$ – независимые одинаково распределенные наблюдения, каждое из которых имеет плотность вида (16.1.1), то статистика*

$$T^*(X) = \Big(\sum_{i=1}^n U_1(X_i), \cdots, \sum_{i=1}^n U_k(X_i)\Big) \tag{16.1.4}$$

является достаточной. Таким образом, сколь бы велик ни был объем выборки $n \geq 1$, для $(X_1, \cdots, X_n)$ всегда существует k – мерная достаточная статистика.

2) *Статистика (16.1.4) является минимальной достаточной статистикой.*

3) *Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково распределенные наблюдения, каждое из которых имеет плотность вида (16.1.1) и пусть существует подмножество $\Theta_0 \in \Theta$ такое, что образ отображения*

$$\theta \in \Theta_0 \longrightarrow Q(\theta) = \{Q_1(\theta), \cdots, Q_k(\theta)\} \in \mathbf{R}^k$$

содержит хотя бы одну точку вместе с некоторой окрестностью и $C(\theta) \neq 0$ в прообразе этой окрестности. Тогда достаточная статистика

$$T^*(X) = \Big(\sum_{i=1}^n U_1(X_i), \cdots, \sum_{i=1}^n U_k(X_i)\Big)$$

является полной.

Доказательство.

1) Доказательство непосредственно следует из Критерия Факторизации (Теорема 7.1.3).

2) Для доказательства применим Теорему 15.1.3. Поскольку функции $Q_i(\theta)$, $U_i(x)$, $C(\theta)$ конечны, а экспонента в (16.1.1) строго положительна, то в качестве привилигерованного распределения P^* можно взять распределение с плотностью из экспоненциального семейства

$$C^n(\theta) \exp\left\{\sum_{j=1}^k Q_j(\theta) \sum_{i=1}^n U_j(x_i)\right\} \prod_{i=1}^n h(x_i),$$

$$\theta \in \Theta \subseteq \mathbf{R}^k, \qquad x_i \in \mathbf{R}^m, \quad i = 1, \cdots, n$$

в фиксированной точке θ_0. Поэтому минимальная достаточная σ – алгебра порождена функциями вида

$$r(x,\, \theta) = \frac{p_\theta(x)}{p_{\theta_0}(x)} = \frac{C^n(\theta)}{C^n(\theta_0)} \exp\left\{\sum_{j=1}^k (Q_j(\theta) - Q_j(\theta_0)) \sum_{i=1}^n U_j(x_i)\right\},$$

$x = (x_1, \cdots, x_n) \in \mathbf{R}^{mn}$, при всех $\theta \in \Theta$. Из линейной независимости функций $1, Q_1(\theta), \cdots, Q_k(\theta)$ следует линейная независимость функций

$Q_1(\theta) - Q_1(\theta_0), \cdots, Q_k(\theta) - Q_k(\theta_0)$, а это означает, что существуют точки $\theta_1, \cdots, \theta_k \in \Theta$ такие, что определитель матрицы с элементами

$$Q_{ij} = Q_i(\theta_j) - Q_i(\theta_0)$$

не равен 0, и поэтому уравнения

$$\sum_{j=1}^{k}(Q_j(\theta_l) - Q_j(\theta_0))\sum_{i=1}^{n}U_j(x_i) = \log r(x,\, \theta_l) - n(\log C(\theta_l) - \log C(\theta_0)),$$

$l = 1, \cdots, k$ однозначно разрешимы относительно $T^*(x)$. Таким образом σ – подалгебра, порождённая статистикой $T^*(x)$ содержится в σ – подалгебре, порождённой функциями

$$r(x,\, \theta_l), \quad l = 1, \cdots, k,$$

а эта σ – подалгебра, очевидно, содержится в минимальной достаточной σ – подалгебре. Итак $T^*(X)$ – минимальная достаточная статистика.

3) Очевидно достаточно рассмотреть случай $n = 1$. Будем считать (производя, если необходимо, сдвиг в пространстве параметров), что образ $Q(\Theta_0)$ множества Θ_0 содержит прямоугольник вида

$$\Pi = \{(q_1, \cdots, q_k) :\ -c < q_j < c;\ j = 1, \cdots k\}, \quad c > 0.$$

Пусть для некоторой измеримой функции $\phi(t)$ справедливо тождество

$$\mathsf{E}_\theta \phi(T^*(X)) \equiv 0, \quad \theta \in \Theta_0. \tag{16.1.5}$$

Необходимо доказать, что

$$\mathsf{P}_\theta(\phi(T^*(X)) = 0) \equiv 1, \quad \theta \in \Theta. \tag{16.1.6}$$

Обозначим через $\phi^+(t)$ и $\phi^-(t)$ положительную и отрицателную части функции $\phi(t)$, то есть

$$\phi^+(t) = \max\{\phi(t),\, 0\}, \quad \phi^-(t) = \max\{-\phi(t),\, 0\}, \quad \phi(t) = \phi^+(t) - \phi^-(t).$$

Тогда из соотношения (16.1.5) следует, что

$$\int \phi^+(T^*(x)) \exp\left\{\sum_{j=1}^{k} Q_j(\theta)U_j(x)\right\} h(x)d\nu(x) =$$

$$= \int \phi^-(T^*(x)) \exp\left\{\sum_{j=1}^k Q_j(\theta)U_j(x)\right\} h(x)d\nu(x), \quad \theta \in \Theta_0. \qquad (16.1.7)$$

Поизводя в тождестве (16.1.7) замену переменных $t = T^*(x)$, получим что для некоторой меры $\lambda(\cdot)$ справедливо тождество

$$\int \phi^+(t) \exp\Big\{\sum_{j=1}^k q_j t_j\Big\} d\lambda(t) = \int \phi^-(t) \exp\Big\{\sum_{j=1}^k q_j t_j\Big\}) d\lambda(t), \qquad (16.1.8)$$

$q = (q_1, \cdots, q_k) \in \Pi$ и в частности

$$\int \phi^+(t) d\lambda(t) = \int \phi^-(t) d\lambda(t),$$

при этом последние интегралы без ограничения общности можно считать равными единице. Полагая

$$\mathsf{P}^{\pm}(B) = \int\limits_B \phi^{\pm}(t) d\lambda(t), \quad B \in \mathcal{B}^k, \qquad (16.1.9)$$

имеем, что P^+ и P^- есть вероятноятные меры на $(\mathbf{R}^k,\, \mathcal{B}^k)$ и при этом из (16.1.8) следует, что

$$\int \exp\Big\{\sum_{j=1}^k q_j t_j\Big\} d\mathsf{P}^+(t) = \int \exp\Big\{\sum_{j=1}^k q_j t_j\Big\} d\mathsf{P}^-(t), \quad q = (q_1, \cdots, q_k) \in \Pi.$$

Рассмотрим теперь эти интегралы, как функции комплексных переменных $q_j = \xi_j + i\eta_j$, $j = 1, \cdots, k$. При любых фиксированных

$$q_1, \cdots, q_{j-1}, q_{j+1}, \cdots, q_k,$$

действительные части которых лежат строго внутри промежутка от $-c$ до $+c$, эти интегралы по Теореме 16.1.1, п.2 являются аналитическими функциями q_j в полосе

$$\Lambda_j = \{q_j :\ -c < \xi_j < +c, \ \ -\infty < \eta_j < +\infty\}, \quad j = 1, \cdots, k$$

комплексной плоскости. При фиксированных действительных $\xi_2, \cdots, \xi_k$, лежащих между $-c$ и $+c$, равенство интегралов имеет место в полосе Λ_1, в которой они аналитичны. По индукции равенство может быть распространено на многомерную комплексную область

$$\{(q_1, \cdots, q_k) :\ (\xi_j, \eta_j) \in \Lambda_j, \ \ j = 1, \cdots, k\}.$$

Отсюда, в частности, следует, что при всех действительных $\eta_1, \cdots, \eta_k$

$$\int \exp\Big\{i\sum_{j=1}^{k}\eta_j t_j\Big\} d\mathsf{P}^+(t) = \int \exp\Big\{i\sum_{j=1}^{k}\eta_j t_j\Big\}) d\mathsf{P}^-(t).$$

последние интегралы представляют собой *характеристические функции* (см. Лекцию 3) распределений P^+ и P^- соответственно, и по Теореме единственности для характеристических функций, распределения P^+ и P^- должны совпадать. Но из их определения (16.1.9) следует, что

$$\phi^+(t) \equiv \phi^-(t), \qquad \nu - \text{почти всюду},$$

и поэтому справедливо (16.1.6).

□

Пример 16.1.3. Пусть X имеет Гамма – распределение с параметром $\theta = (\alpha,\ \lambda)$, $\alpha > 0$, $\lambda > 0$, тогда в Примере 16.1.1 было показано, что это экспоненциальное семейство с

$$U_1(x) = \log x, \quad U_2(x) = x, \quad C(\theta) = \frac{\alpha^\lambda}{\Gamma(\lambda)}, \quad Q_1(\theta) = \lambda, \quad Q_2(\theta) = -\alpha,$$

поэтому по Теореме 16.1.2 статистика

$$T^*(X) = \Big(\sum_{i=1}^{n}\log X_i,\ \sum_{i=1}^{n} X_i\Big)$$

является полной минимальной достаточной статистикой. Заметим также, что и эквивалентная статистика

$$\bar{T}^*(X) = \Big(\prod_{i=1}^{n} X_i,\ \sum_{i=1}^{n} X_i\Big)$$

также является полной минимальной достаточной статистикой.

16.2 СПИСОК ЛИТЕРАТУРЫ

1) Э. Леман, Проверка Статистических Гипотез,
Москва, Наука, 1979, Глава 2, ¡7.

2) Э. Леман, Теория Точечного Оценивания,
Москва, Наука, 1991, Глава 1, ¡4.

3) А.А. Боровков, Математическая Статистика,
Москва, Наука, 1984, Глава 2, ¡15.

4) Ж.– Р. Барра, Основные Понятия Математической Статистики,
Москва, Мир, 1974, Глава 10.

5) Ю.В. Линник, Лекции о Задачах Аналитической Статистики,
Москва, Наука, 1991, Лекция 1.

Лекция 17

До сих пор изучались свойства и методы построения ТОЧЕЧНЫХ *оценок неизвестного параметра θ. В Лекции рассматривается другой подход к проблеме, когда неизвестное значение параметра θ оценивается с помощью* МНОЖЕСТВА.

17.1 ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ

Пусть $(\mathcal{X},\ \mathcal{F},\ \{\mathsf{P}_\theta, \theta \in \Theta\})$ – доминируемая статистическая структура и $(\Delta,\ \mathcal{U})$ – пространство решений. Пусть $g(\theta)$ некоторая оцениваемая измеримая функция, заданная на Θ и действующая в некоторое измеримое пространство $(\Gamma, \mathcal{W})$. Предположим, что по результатам наблюдений $X = x$ мы хотим "оценить" значение $g(\theta)$ с помощью некоторого *множества* $C(x) \in \mathcal{W}$. В этом случае естественно положить $\Delta = \mathcal{W}$. Будем рассмотрим только вырожденные стратегии

$$S(x, D) = \mathbf{1}_D(C(x)), \quad D \in \mathcal{U}$$

и отождестим их с множеством $C(x) \in \mathcal{W}$. Рассмотрим функцию потерь (см. Определение 8.2.1) $L(\theta, d)$ вида

$$L(\theta, d) = \begin{cases} 0,\ g(\theta) \in d \\ 1,\ g(\theta) \notin d, \end{cases} \qquad d \in \mathcal{W}.$$

Тогда средние потери (см. Определение 8.2.1) равны

$$W_S(\theta, x) = \int L(\theta, \delta) dS(x, \delta) = L(\theta, C(x))$$

и риск представляет собой вероятность "непокрытия" значения $g(\theta)$ случайным множеством $C(X)$

$$R(\theta, C) = \mathsf{E}_\theta W_S(\theta, X) = \mathsf{E}_\theta L(\theta, C(X)) = \mathsf{P}_\theta(g(\theta) \notin C(X)).$$

Определение 17.1.1.

1) *Семейство множеств* $C(x) \in \mathcal{W}, \quad x \in \mathcal{X}$ *называется семейством доверительных множеств, если*

$$\{x: \, g(\theta) \in C(x)\} \in \mathcal{F}, \qquad \textit{для всех} \ \ \theta \in \Theta.$$

2) *Величина*

$$\gamma_C(\theta) = \mathsf{P}_\theta(g(\theta) \in C(X))$$

называется доверительной вероятностью, а число

$$\gamma_C = \inf_{\theta \in \Theta} \gamma_C(\theta)$$

называется коэффициентом доверия семейства доверительных множеств $C(x) \in \mathcal{W}, \quad x \in \mathcal{X}$.

Заметим, что первая часть Определения 17.1.1 означает просто, что определены вероятности вида

$$\mathsf{P}_\theta(g(\theta) \in C(X)), \qquad \text{для всех} \ \ \theta \in \Theta.$$

Часто встречается ситуация, когда $\Gamma = \mathbf{R}^1, \ \mathcal{W} = \mathcal{B}^1$, а семейство доверительных множеств $C(x), x \in \mathcal{X}$ являются интервалами вида $(a(x), \, b(x))$, тогда в этом случае $C(x)$ называются доверительными интервалами, а $a(x)$ и $b(x)$ – доверительными границами. Обычно задаются числом (близким к единице) $1-\alpha$, $\alpha \in (0,1)$ и рассматривают только те доверительные множества $C(x)$, для которых доверительная вероятность ограничена снизу этим числом $\gamma_C(\theta) \ \geq \ 1-\alpha$, для всех $\theta \in \Theta$, то есть ограничивают риск *сверху*

$$R(\theta, C) = 1 - \mathsf{P}_\theta(g(\theta) \in C(X)) = 1 - \gamma_C(\theta) \ \leq \ 1 - (1-\alpha) = \alpha, \quad \theta \in \Theta.$$

Рассмотрим общий способ построения доверительных множеств. Пусть сначала $g(\theta) = \theta \in \Theta$ и для каждого θ построим множество $S_\theta \in \mathcal{F}$ такое, что

$$\mathsf{P}_\theta(X \in S_\theta) \ \geq \ 1-\alpha.$$

Положим
$$C(x) = \{\theta : \ x \in S_\theta\},$$
тогда $C(X)$ – доверительное множество с коэффициентом доверия не меньшим чем $1-\alpha$, поскольку
$$\mathsf{P}_\theta(\theta \in C(X)) = \mathsf{P}_\theta(X \in S_\theta) \ \geq \ 1-\alpha.$$
В общем случае для функции $g(\theta)$ рассмотрим множества $S_g, g \in \Gamma$ такие, что
$$\inf_{\theta:\, g(\theta)=g} \mathsf{P}_\theta(X \in S_g) \ \geq \ 1-\alpha, \qquad \text{для всех} \quad g \in \Gamma.$$
Положим
$$C(X) = \{g \in \Gamma : \ x \in S_g\},$$
тогда
$$\mathsf{P}_\theta(g(\theta) \in C(X)) = \mathsf{P}_\theta(X \in S_{g(\theta)}) \ \geq$$
$$\geq \inf_{\bar\theta:\, g(\bar\theta)=g(\theta)} \mathsf{P}_\theta(X \in S_{g(\theta)}) \ \geq \ 1-\alpha.$$
Этот метод основан на самих наблюдениях X. Отметим, что этот метод применим для случая *произвольного* множества Γ и , в частности, в случае векторного параметра θ. Отметим также, что получаемое доверительное множество $C(X)$ *неоднозначно*, поскольку при заданном $1-\alpha$ множества S_g можно выбрать различными способами и задача состоит в построении доверительного множества *минимальных* "размеров", обеспечивающего наиболее точную (при заданном $1-\alpha$) локализацию оцениваемой функции.

Рассмотрим теперь другие методы построения доверительных множеств.

17.2 МЕТОД ПОСТРОЕНИЯ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ, ОСНОВАННЫЙ НА ЦЕНТРАЛЬНЫХ СТАТИСТИКАХ

Рассмотрим для простоты случай скалярного параметра $\theta \in \Theta \subseteq \mathbf{R}^1$

Определение 17.2.1.*Вещественная функция $G(\theta, X)$, определенная на $\Theta \times \mathcal{X}$, называется центральной статистикой, если*

1) *При каждом $\theta \in \Theta$ функция рапределения случайных величин $G(\theta, X)$ непрерывна и не зависит от θ.*

2) Для всех $x \in \mathcal{X}$ функция $G(\theta, x)$ непрерывна и строго монотонна по $\theta \in \Theta$.

Построим доверительное множество с помощью центральной статистики. Поскольку её функция распределения непрерывна и не зависит от θ, то для любого $\alpha \in (0,\,1)$ существуют числа $\beta_1 < \beta_2$ (не зависящие от θ) такие, что

$$\mathsf{P}_\theta(\beta_1 < G(\theta, X) < \beta_2) = 1 - \alpha, \quad \alpha \in (0,\,1), \qquad \theta \in \Theta.$$

Поскольку функция $G(\theta, x)$ строго монотонна и непрерывна по θ, то при каждом $x \in \mathcal{X}$ существуют решения относительно θ уравнений

$$G(\theta, x) = \beta_1, \quad G(\theta, x) = \beta_2. \tag{17.2.1}$$

Обозначим эти решения через $u(x)$ и $v(x)$. Предположим, например, что функция $G(\theta, x)$ строго возрастает по θ, тогда $u(x) < v(x)$ и справедливы соотношения

$$\mathsf{P}_\theta(u(X) < \theta < v(X)) = \mathsf{P}_\theta(G(u(X), X) < G(\theta, X) < G(v(X), X)) =$$

$$= \mathsf{P}_\theta(\beta_1 < G(\theta, X) < \beta_2) = 1 - \alpha,$$

то есть множество $C(X) = (u(X),\, v(X))$ является доверительным интервалом с коэффициентом доверия $1 - \alpha$. Таким образом доказана следующая Теорема.

Теорема 17.2.1.*Пусть $G(\theta, X)$ – центральная статистика и $u(x)$, $v(x)$ – решения при каждом $x \in \mathcal{X}$ относительно θ уравнений (17.2.1), где $\beta_1 < \beta_2$ и такие, что*

$$\mathsf{P}_\theta(\beta_1 < G(\theta, X) < \beta_2) = 1 - \alpha, \quad \alpha \in (0,\,1).$$

Тогда, если функция $G(\theta, x)$ строго возрастает по θ, то $(u(X),\, v(X))$ – доверительный интервал с коэффициентом доверия $1 - \alpha$. Если же $G(\theta, x)$ строго убывает θ, то $(v(X),\, u(X))$ – доверительный интервал с коэффициентом доверия $1 - \alpha$.

В каждом конкретном случае при построении центральной статистики приходится учитывать специфику рассматриваемой модели, однако можно выделить класс моделей, для которых центральная статистика всегда существует и имеет достаточно простой вид.

Теорема 17.2.2.*Пусть наблюдения X имеют вид $X = (X_1, \cdots, X_n)$, где X_i, $i = 1, \cdots, n$ независимы и одинаково распределены. Предположим,* *что функция распределения $F(x,\theta)$ наблюдения X_1 непрерывна и строго монотонна по $\theta \in \Theta$. Тогда статистика*

$$G(\theta, X) = -\sum_{i=1}^{n} \log F(X_i, \theta)$$

является центральной и доверительный интервал с коэффициентом доверия $1-\alpha$, $\alpha \in (0,1)$ имеет вид $(u(X), v(X))$, где $u(X) < v(X)$ – решения относительно θ уравнений

$$-\sum_{i=1}^{n} \log F(X_i, \theta) = \beta_1, \quad -\sum_{i=1}^{n} \log F(X_i, \theta) = \beta_2 \tag{17.2.2}$$

где β_1 и β_2 удовлетворяют равенству

$$\frac{1}{(n-1)!}\int\limits_{\beta_1}^{\beta_2} x^{n-1} e^{-x} dx = 1 - \alpha.$$

Доказательство. Поскольку слагаемые $F(X_i, \theta)$ независимы и имеют *равномерное* $\mathcal{R}(0,1)$ распределение (см. Теорему 17.3.1), то распределение статистики $G(\theta, X)$ совпадает с гамма-распределением с параметрами $(1, n)$. $\square$

Заметим, что наибольшая трудность в применении Теоремы 17.2.2, возникает при нахождении решений уравнений (17.2.2).

17.3 ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНЫХ МНОЖЕСТВ С ИСПОЛЬЗОВАНИЕМ ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ СТАТИСТИК

Рассмотрим другой метод построения доверительных множеств, основанный на следующей Теореме.

Теорема 17.3.1. *Пусть случайная величина X имеет функцию распределения $F(x)$, тогда для любого $y \in [0,1]$, справедливы неравенства*

$$\mathsf{P}(F(X+0) \le y) \le y \le \mathsf{P}(F(X) < y).$$

Доказательство. Докажем сначала правое неравенство из формулировки Теоремы. Напомним, что функция распределения $F(x) = \mathsf{P}(X < x)$ непрерывна *слева*, то есть

$$F(x) = \lim_{\varepsilon \downarrow 0} F(x-\varepsilon), \qquad \mathsf{P}(X \le x) = \lim_{\varepsilon \downarrow 0} F(x+\varepsilon) = F(x+0).$$

Для любого $y \in [0,1]$ определим число $z = \sup\{x : F(x) < y\}$, тогда, если для $y = 1$, $z = \infty$, то $\mathsf{P}(F(X) < 1) = 1$ и для $y = 1$ утверждение доказано. Поэтому будем предполагать, что $z < \infty$. По определению супремума это число обладает следующими свойствами

1) Для любого $\varepsilon > 0$ справедливо неравенство $F(z-\varepsilon) < y$.

2) $F(z+\varepsilon) \ge y$.

Поэтому устремляя $\varepsilon > 0$ к нулю в этих неравенствах, получим $F(z) \le y \le F(z+0)$. Рассмотрим два случая

1) Пусть $F(z) = y$. Тогда

$$\mathsf{P}(F(X) < y) = \mathsf{P}(X < z) = F(z) = y.$$

2) Пусть $F(z) < y$, тогда

$$\mathsf{P}(F(X) < y) = \mathsf{P}(X \le z) = F(z+0) \ge y.$$

То есть в любом случае $\mathsf{P}(F(x) < y\} \ge y$ и правое неравенство доказано. Докажем теперь левое неравенство. С этой целью рассмотрим случайную величину $Y = -X$. Тогда

$$G(x) = \mathsf{P}(Y < x) = \mathsf{P}(X > -x) = 1 - F(-x+0)$$

и по доказанному

$$x \le \mathsf{P}(G(Y) < x) = \mathsf{P}(F(X+0) > 1-x) = 1 - \mathsf{P}(F(X+0) \le 1-x),$$

полагая $y = 1-x$, получим $\mathsf{P}(F(X+0) \le y) \le y$. □

Следствие 17.3.1. *Если функция распределения $F(x)$ непрерывна, то*

$$\mathsf{P}(F(X) < y) = y = \mathsf{P}(F(X) \le y),$$

то есть случайная величина $F(X)$ имеет равномерное распределение $F(X) \sim \mathcal{R}(0,\,1)$.

Для доказательства достаточно заметить, что

$$\mathsf{P}(F(X+0) \le y) = \mathsf{P}(F(X) \le y) \ge \mathsf{P}(F(X) < y).$$

Применим эту Теорему к построению доверительных множеств. С этой целью предположим, что имеется статистика $T(X)$ с функцией распределения $F(t,\theta),\ \theta \in \Theta$.

Теорема 17.3.2. *Пусть статистика $T(X)$ имеет функцию распределения $F(t,\theta),\ \theta \in \Theta$ и числа $\alpha_1 \in (0,1),\ \alpha_2 \in (0,1)$ таковы, что $\alpha_1+\alpha_2 < 1$. Тогда множества*

$$C_1(X) = \{\theta:\ F(T(X),\theta) < 1-\alpha_1\}, \quad C_2(X) = \{\theta:\ F(T(X)+0,\theta) > \alpha_2\},$$

$$C(X) = C_1(X) \cap C_2(X)$$

являются доверительными множествами с коэффициентами доверия не меньшими соответственно $1-\alpha_1,\ 1-\alpha_2,\ 1-\alpha_1-\alpha_2$.

Доказательство. Применим Теорему 17.3.1 с $X = T(X),\ F(t) = F(t,\theta)$. Имеем

$$\mathsf{P}_\theta(\theta \in C_1(X)) = \mathsf{P}_\theta(F(T(X),\theta) < 1-\alpha_1) \ge 1-\alpha_1,$$

$$\mathsf{P}_\theta(\theta \in C_2(X)) = \mathsf{P}_\theta(F(T(X)+0,\theta) > \alpha_2) = 1-\mathsf{P}_\theta(F(T(X)+0,\theta) \le \alpha_2) \ge 1-\alpha_2,$$

$$\mathsf{P}_\theta(\theta \in C_1(X) \cap C_2(X)) = \mathsf{P}_\theta(F(T(X),\theta) < 1-\alpha_1,\ F(T(X)+0,\theta) > \alpha_2) =$$

$$= \mathsf{P}_\theta(F(T(X),\theta) < 1-\alpha_1) - \mathsf{P}_\theta(F(T(X),\theta) < 1-\alpha_1,\ F(T(X)+0,\theta) \le \alpha_2) \ge$$

$$\ge \mathsf{P}_\theta(F(T(X),\theta) < 1-\alpha_1) - \mathsf{P}_\theta(F(T(X)+0,\theta) \le \alpha_2) \ge 1-\alpha_1-\alpha_2.$$

□

Пример 17.3.1. Пусть наблюдения X имеют вид $X = (X_1,\cdots,X_n)$, где $X_i, i = 1,\cdots,n$ – независимы и одинаково нормально распределены

$$X_i \sim \mathcal{N}(\theta,\, 1), \quad i = 1,\cdots,n, \quad \theta \in \Theta = \mathbf{R}^1.$$

Наилучшей оценкой для параметра θ является

$$T(X) = \overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \sim \mathcal{N}(\theta,\, 1/n), \quad F(t,\theta) = \Phi(\sqrt{n}(t-\theta)).$$

Применим последнюю Теорему для постороения доверительного множества $C(X)$, используя статистику $T(X)$. Имеем при $\alpha_1+\alpha_2 < 1$

$$C_1(X) = \{\theta:\ \Phi(\sqrt{n}(\overline{X}-\theta)) < 1-\alpha_1\} = \{\theta:\ \sqrt{n}(\overline{X}-\theta) < u_{1-\alpha_1}\} =$$

$$= \left\{\theta:\ \theta > \overline{X} - \frac{u_{1-\alpha_1}}{\sqrt{n}}\right\} = \left(\overline{X} - \frac{u_{1-\alpha}}{\sqrt{n}},\ +\infty\right),$$

$$C_2(X) = \{\theta:\ \Phi(\sqrt{n}(\overline{X} - \theta)) > \alpha_2\} = \left(-\infty,\ \overline{X} - \frac{u_{\alpha_2}}{\sqrt{n}}\right),$$

$$C(X) = \left(\overline{X} - \frac{u_{1-\alpha_1}}{\sqrt{n}},\ \overline{X} - \frac{u_{\alpha_2}}{\sqrt{n}}\right), \quad \text{где} \ \ \Phi(u_\alpha) = \alpha.$$

Пусть теперь необходимо построить доверительное множество для значений функции $g(\theta)$, которая не обязана быть однозначной. Тогда вместо функции $F(t,\theta)$ достаточно рассмотреть функцию

$$H(t,g) = \inf_{\theta: g(\theta)=g} F(t,\theta), \quad g \in \Gamma$$

и определить множества

$$C_1(X) = \{g \in \Gamma : H(T(X),g) < 1-\alpha_1\}, \ \ C_2(X) = \{g \in \Gamma : H(T(X)+0,g) > \alpha_2\},$$

$$C(X) = C_1(X) \cap C_2(X).$$

Справедлива следующая Теорема.

Теорема 17.3.3 *Множества $C_i(X), i = 1,2,3$, определенные выше, являются доверительными множествами для $g(\theta)$ с коэффициентами доверия не меньше, соответственно, чем $1-\alpha_1,\ 1-\alpha_2,\ 1-\alpha_1-\alpha_2$.*

Доказательство. Доказательство Теоремы полностью аналогично доказательству Теоремы 17.3.2. Докажем Теорему, например, для множества $C_1(X)$

$$\mathsf{P}_\theta(g(\theta) \in C_1(X)) = \mathsf{P}_\theta(H(T(X), g(\theta)) < 1-\alpha_1) =$$

$$= \mathsf{P}_\theta\Big(\inf_{\bar\theta: g(\bar\theta)=g(\theta)} F(T(X),\theta) < 1-\alpha_1\Big) \ \geq\ \mathsf{P}_\theta(F(T(X),\theta) < 1-\alpha_1) \ \geq\ 1-\alpha_1.$$

$\square$

17.4 АСИМПТОТИЧЕСКИЕ ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

Предположим, что существует оценка $T_n = T_n(\mathbf{X}_n)$ параметра θ, которая при $n \to \infty$ состоятельна (см. Лекцию 12) и асимптотически нормальна с параметрами $(\theta, \sigma^2(\theta))$, то есть для любого $\varepsilon > 0$

$$\mathsf{P}_\theta(|T_n(\mathbf{X}_n) - \theta| > \varepsilon) \to 0, \quad n \to \infty, \tag{17.4.1}$$

и

$$\mathsf{P}_\theta(\sqrt{n}(T_n(\mathbf{X}_n)-\theta)/\sigma(\theta)<x)\to\Phi(x). \qquad (17.4.2)$$

Тогда справедлива следующая

Теорема 17.4.1. *Пусть выполнены соотношения (17.4.1) и (17.4.2) и функция $\sigma(\theta)$ непрерывна по $\theta\in\Theta$, тогда для любого $\alpha\in(0,1)$ интервал вида*

$$\Big(T_n(\mathbf{X}_n)-n^{-1/2}u_{1-\alpha/2}\sigma(T_n(\mathbf{X}_n)),\ T_n(\mathbf{X}_n)+n^{-1/2}u_{1-\alpha/2}\sigma(T_n(\mathbf{X}_n))\Big),$$

где $\Phi(u_\gamma)=\gamma\in(0,1)$, является асимптотическим доверительным интервалом с коэффициентом доверия $1-\alpha$, то есть

$$\mathsf{P}_\theta\Big(T_n(\mathbf{X}_n)-n^{-1/2}u_{1-\alpha/2}\sigma(T_n(\mathbf{X}_n))<\theta<T_n(\mathbf{X}_n)+n^{-1/2}u_{1-\alpha/2}\sigma(T_n(\mathbf{X}_n))\Big)\to$$

$$\to 1-\alpha,\quad n\to\infty,\quad \theta\in\Theta.$$

Доказательство. Из соотношения (17.4.2) следует, что

$$\mathsf{P}_\theta(\sqrt{n}|T_n(\mathbf{X}_n)-\theta|/\sigma(\theta)<u_{1-\alpha/2})\to\Phi(u_{1-\alpha/2})-\Phi(-u_{1-\alpha/2})=$$

$$=2\Phi(u_{1-\alpha/2})-1=2-\alpha-1=1-\alpha.$$

По условию функция $\sigma(\theta)$ непрерывна, поэтому из Теоремы 12.1.2 следует, что

$$\sigma(T_n(\mathbf{X}_n))\xrightarrow{P_{n\theta}}\sigma(\theta),\quad n\to\infty.$$

Таким образом, учитывая свойство слабой сходимости, имеем

$$\mathsf{P}_\theta(\sqrt{n}|T_n(\mathbf{X}_n)-\theta|/\sigma(T_n(\mathbf{X}_n))<u_{1-\alpha/2})\to 1-\alpha,\quad n\to\infty,\quad \theta\in\Theta.$$

Разрешая неравенства под знаком вероятности относительно θ, получим искомый доверительный интервал. □

Заметим, что асимптотический доверительный интервал будет тем короче, чем "меньше" функция $\sigma^2(\theta)$, поэтому при построении асимптотических доверительных интервалов, по-видимому, разумно использовать оценки $T_n(\mathbf{X}_n)$, имеющие минимальную дисперсию (см. Лекции 9 – 11).

Пример17.4.1 Пусть $\mathbf{X}_n=(X_1,\cdots,X_n)$ – независимые одинаково распределённые бернуллиевские наблюдения

$$X_i\sim\mathcal{B}(1,\,\theta),\quad \theta\in\Theta=(0,\,1),\quad i=1,\cdots,n.$$

Тогда оптимальная оценка для параметра θ имеет вид

$$T_n(\mathbf{X}_n) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Учитывая Закон Больших Чисел (Лекция 4, п.6), имеем

$$T_n(\mathbf{X}_n) \xrightarrow{P_{n\theta}} \theta, \quad n \to \infty$$

и в силу Центральной Предельной Теоремы (Лекция 4, п.5) справедливо соотношение

$$\operatorname*{Lim}_{n\to\infty} \mathsf{P}_\theta\Big(\sqrt{n}(\theta(1-\theta))^{-1/2}(T_n(\mathbf{X}_n)-\theta) < x\Big) = \Phi(x),$$

то есть в этом случае $\sigma^2(\theta) = \theta(1-\theta)$ и эта функция непрерывна. Итак, асимптотический доверительный интервал с коэффициентом доверия $1-\alpha$ есть

$$\left(\overline{X} - u_{1-\alpha/2} n^{-1/2}\sqrt{\overline{X}(1-\overline{X})},\ \overline{X} + u_{1-\alpha/2} n^{-1/2}\sqrt{\overline{X}(1-\overline{X})}\right).$$

17.5 СПИСОК ЛИТЕРАТУРЫ

1) Э. Леман, Проверка Статистических Гипотез,
Москва, Наука, 1979, Глава 3, ¡5.

2) Л.Н. Большев, О построении доверительных пределов,
Теория вероятностей и её применения, 1965, т. 10, н. 1, стр. 187 –192.

3) Л.Н. Большев, Э.А. Логинов, Интервальные оценки при наличии мешающих параметров,
Теория вероятностей и её применения, 1966, т. 11, н. 1, стр. 94 –107.

4) А.А. Боровков, Математическая Статистика,
Москва, Наука, 1984, Глава 2, ¡31.

5) Ж.– Р. Барра, Основные Понятия Математической Статистики,
Москва, Мир, 1974, Глава 6, ¡3, ¡4.

6) Г.И. Ивченко, Ю.И. Медведев, Математическая Статистика,
Москва, Высшая Школа, 1992, Глава 2, ¡2.6.

Лекция 18

В Лекции рассматривается введение в эмпирическую байесовскую проблему решений.

18.1 СТРУКТУРА БАЙЕСОВСКИХ РЕШЕНИЙ

Рассмотрим ситуацию, когда многократно возникает однотипная статистическая задача, например, задача проверки гипотез или оценивания параметров. При этом в прошлом многократно наблюдаются случайные величины с плотностью вида (смесь)

$$p_Q(x) = \int\limits_{\Theta} p_\theta(x) Q(\theta).$$

Как и при байесовском подходе, будем считать, что на Θ задано априорное распределение $Q(\cdot)$ и предпологается, что каждый раз (ненаблюдаемый) параметр θ порождается случайно в соответствии с этим распределением на Θ, то есть θ является значением случайной величены Ξ, имеющей распределение $Q(\cdot)$. После этого случайная величина X порождается в соответствии с распределением, заданным условной плотностью $p_\theta(x) = p(x \mid \Xi = \theta)$. При каждом $\theta \in \Theta$ функция $p_\theta(x)$ предполагается известной. Здесь рассматриваеся задача "оценивания" текущего значения θ.

В противоположность чисто байесовской схеме здесь нет необходимости точно задавать априорное распределение, в соответствии с которым порождается Ξ. При эмпирическом байесовском подходе необходимо только определить *семейство* априорных распределений, которому принадлежит $Q(\cdot)$. Задача состоит в том, чтобы построить такую последовательность решающих функций (см. Лекцию 8), которая при определённой функции потерь будет "близка" к оптимальной байесовской решающей функции, а соответсвующая последовательность априорных рисков будет "близка" к байесовскому риску. Впервые этот подход был предложен Г.Роббинсом (Herbert Robbins) в 1955 году и который назвал его *эмпирической байесовской процедурой* (см. [1], [2]).

Итак предположим, что нам заданы

1) Параметрическое пространство Θ с элементами θ, которые интерпретируется как "состояния природы" и которые неизвестны для нас.

2) Пространство решений Δ с элементами δ, которое характеризует решаемую задачу.

3) Функция потерь $L(\theta, \delta) \geq 0$, которая характеризует "потери", которые мы несём выбирая решение δ в случае, когда истинное значение параметра есть θ.

4) Априорное распределение $Q(\cdot)$ на Θ случайной величины Ξ, которое может быть как известным так и неизвестным для нас.

5) Наблюдаемая случайная величина X, значения которой принадлежат пространству $\mathcal{X}$, на котором определена σ – конечная мера $\nu(\cdot)$. Если случайная величина Ξ приняла значение θ, то случайная величина X имеет условную плотность $p_\theta(x)$ относительно этой меры $\nu(\cdot)$. Безусловная плотность X имеет вид (*смесь*)

$$p_Q(x) = \int\limits_{\Theta} p_\theta(x) dQ(\theta).$$

Проблема состоит в том, чтобы выбрать измеримую *решающую функцию* $\delta(x)$, определённую на $(\mathcal{X}, \mathcal{F})$ и принимающую значения в $(\Delta, \mathcal{U})$

$$\delta(x): \mathcal{X} \longrightarrow \Delta,$$

минимизирующую байесовский риск (см. Лекцию 8). А именно, если наблюдается значение $X = x$ и принимается решение $\delta(x)$, то имеем "случайные" потери вида

$$L(\theta, \delta(x)).$$

Далее для данной решающей функции $\delta(\cdot)$ определим условные средние потери по формуле

$$R(\theta, \delta) = R(\delta \mid \Xi = \theta) = \mathsf{E} L(\theta, \delta(X)) =$$

$$= \int_{\mathcal{X}} L(\theta, \delta(x)) p_\theta(x) d\nu(x). \qquad (18.1.1)$$

Усредняя по априорному распределению $Q(\cdot)$ определим *байесовский риск* (*априорный байесовский риск*) решающей функции $\delta(\cdot)$ соответствующий априорному распределению $Q(\cdot)$

$$r(\delta, Q) = \mathsf{E} R(\delta \mid \Xi) = \int_{\Theta} R(\theta, \delta) dQ(\theta). \qquad (18.1.2)$$

По теореме Фубини байесовский риск можно переписать в виде

$$r(\delta, Q) = \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) p_\theta(x) d\nu(x) dQ(\theta) =$$

$$= \int_{\mathcal{X}} \Big(\int_{\Theta} L(\theta, \delta(x)) p_\theta(x) dQ(\theta) \Big) d\nu(x),$$

поэтому, если определить функцию

$$h_Q(\delta, x) = \int_{\Theta} L(\theta, \delta) p_\theta(x) dQ(\theta), \qquad \delta \in \Delta, \qquad (18.1.3)$$

то байесовский риск $r(\delta, Q)$ можно переписать в виде

$$r(\delta, Q) = \int_{\mathcal{X}} h_Q(\delta(x), x) d\nu(x). \qquad (18.1.4)$$

Будем предпологать, что существует решающая функция $\delta_Q(x)$ такая, что

$$h_Q(\delta_Q(x), x) = \min_{\delta \in \Delta} h_Q(\delta, x). \qquad (18.1.5)$$

Тогда для любой решающей функции $\delta(x)$ справедливо неравенство

$$r(\delta_Q, Q) = \int_{\mathcal{X}} \min_{\delta \in \Delta} h_Q(\delta, x) d\nu(x) \leq$$

$$\leq \int_{\mathcal{X}} h_Q(\delta(x), x) d\nu(x) = r(\delta, Q). \qquad (18.1.6)$$

Поэтому, если определить функционал $r(Q)$ от $Q(\cdot)$ по формуле

$$r(Q) = r(\delta_Q, Q) = \int\limits_{\mathcal{X}} h_Q(\delta_Q(x), x) d\nu(x), \tag{18.1.7}$$

то

$$r(Q) = \min_{\delta \in \Delta} r(\delta, Q). \tag{18.1.8}$$

Таким образом решающая функция $\delta_Q(x)$, определённая в формуле (18.1.5), является *байесовской решающей функцией*.

18.2 ЭМПИРИЧЕСКИЙ БАЙЕСОВСКИЙ ПОДХОД

Формулы (18.1.3) – (18.1.8) можно интерпретировать следующим образом: предположим, что априорное распределение $Q(\cdot)$ имеет плотность $q(\theta)$ относительно некоторой σ – конечной меры $\mu(\cdot)$, тогда определим *апостериорную* плотность случайной величины Ξ по формуле

$$q(\theta \mid X = x) = q(\theta \mid x) = \frac{p_\theta(x) q(\theta)}{p_Q(x)},$$

тогда выражения (18.1.3) и (18.1.4) можно переписать в следующем виде

$$h_Q(\delta, x) = p_Q(x) \int\limits_{\Theta} L(\theta, \delta) q(\theta \mid x) d\mu(\theta) \equiv p_Q(x) h_Q(\delta \mid X = x),$$

$$r(\delta, Q) = \int\limits_{\mathcal{X}} p_Q(x) h_Q(\delta(x) \mid X = x) d\nu(x) =$$

$$= \int\limits_{\mathcal{X}} p_Q(x) \Big(\int\limits_{\Theta} L(\theta, \delta(x)) q(\theta \mid x) d\mu(\theta) \Big) d\nu(x). \tag{18.2.1}$$

При этом из определения (18.1.5) решающей функции $\delta_Q(x)$ и этих формул следует, что в формуле (18.1.5) можно $h_Q(\delta, x)$ заменить на $h_Q(\delta \mid X = x)$. Таким образом оптимальная решающая функция $\delta_Q(x)$ определяется на основе *апостериорного* распределения с плотностью $q(\theta \mid X = x)$ и байесовский риск $r(\delta, Q)$ может быть записан в виде

$$r(\delta, Q) = \mathsf{E} r(\delta, Q \mid X),$$

где $r(\delta, Q \mid X)$ – апостериорный риск

$$r(\delta, Q \mid X = x) = \int\limits_{\Theta} L(\theta, \delta(x)) q(\theta \mid x) d\mu(\theta)$$

и случайная величина X имеет плотность вида

$$p_Q(x) = \int\limits_{\Theta} p_\theta(x) dQ(\theta).$$

Итак, после наблюдения значения $X = x$ априорное распределение $Q(\cdot)$ с плотностью $q(\theta)$ "заменяется" апостериорным распределением с плотностью $q(\theta \mid X = x)$ и оптимальная решающая функция $\delta_Q(x)$ получается путём минимизации байесовского апостериорного риска $r(d, Q \mid X = x)$, а априорный риск $r(\delta, Q)$ получается усреднением апостериорного риска $r(\delta, Q \mid X = x)$ по распределению с плотностью $p_Q(x)$. Эти факты важны для интерпретации описанной ниже схемы.

Определение 18.2.1. *Любая решающая функция $\delta_Q(x)$, удовлетворяющая соотношению (18.1.5) и минимизирующая байесовский риск $r(\delta, Q)$ при априорном распределении $Q(\cdot)$, называется байесовской решающей функцией соответствующей априорному распределению $Q(\cdot)$. Функционал $r(Q)$ от $Q(\cdot)$, определенный равенством (18.1.7) называется байесовским огибающим функционалом.*

Если априорное распределение $Q(\cdot)$ *известно*, то используя $\delta_Q(x)$ мы получим минимальный байесовский риск $r(Q)$.

Предположим теперь, что априорное распределение $Q(\cdot)$ *не известно*, но рассматриваемая ситуация наблюдается многократно и независимо. Итак, пусть

$$(x_1, \theta_1),\ (x_2, \theta_2),\ (x_3, \theta_3),\ \cdots \tag{18.2.2}$$

последовательность значений независимых пар наблюдений над случайными величинами X и Ξ, причём значения θ_i *ненаблюдаемы*, то есть неизвестны нам, и являются значениями случайной величины Ξ с распределением $Q(\cdot)$. Значения x_i наблюдаемы и являются реализациями случайной величины X, условное распределение которой при условии $\Xi = \theta$ имеет плотность вида $p_\theta(x)$ и безусловную плотность относительно меры $\nu(\cdot)$ вида (смесь)

$$p_Q(x) = \int\limits_{\Theta} p_\theta(x) dQ(\theta).$$

Итак мы наблюдаем независимые случайные величины $X_1, \cdots, X_{n+1}$, причём i – ая случайная величина X_i имеет плотность вида $p_Q(x)$, $\ i = 1, \cdots, n$, а случайная величина X_{n+1} имеет плотность $p_{\theta_{n+1}}(x)$. Наблюдая $x_1, \cdots, x_{n+1}$ мы хотим "оценить" текущее значение θ_{n+1}. Поскольку значения $\theta_1, \cdots, \theta_n$ порождены тем же вероятностным распределением $Q(\cdot)$, что и θ_{n+1}, то значения $x_1, \cdots, x_n$ содержат также информацию о θ_{n+1} и поэтому будем рассматривать решающие функции, зависящие и от $x_1, \cdots, x_n$, то есть пусть

$$\delta_n(x) = \delta_n(x_1, \cdots, x_n;\ x) \in \Delta \tag{18.2.3}$$

и значит текущие потери имеют вид

$$L(\theta_{n+1}, \delta_n(x_{n+1})).$$

Основная проблема состоит в нахождении такой решающей функции $\delta_n(\cdot)$, чтобы асимптотически, то есть при больших n, она была "близка" к оптимальной, но неизвестной решающей функции $\delta_Q(\cdot)$, то есть чтобы в некотором смысле выполнялось соотношение

$$\delta_n(x) \approx \delta_Q(x), \quad n \to +\infty.$$

Определим теперь "эмпирическую" или "адаптивную" решающую процедуру, как некоторую последовательность $\delta = \{\delta_n\}$ решающих функций $\delta_n(x)$ вида (18.2.3). Для данной последовательности δ условный байесовский риск (см. (18.1.3) и (18.1.4)) при "оценивании" θ_{n+1} при *данных* $x_1, \cdots, x_n$ есть

$$r_n(\delta, Q \mid X_1 = x_1, \cdots, X_n = x_n) = \int\limits_{\mathcal{X}} h_Q(\delta_n(x), x) d\nu(x) =$$

$$= \int\limits_{\mathcal{X}} p_Q(x) \Big(\int\limits_{\Theta} L(\theta, \delta_n(x_1, \cdots, x_n;\ x), \theta) q(\theta \mid X_{n+1} = x) d\mu(\theta) \Big) d\nu(x). \tag{18.2.4}$$

Определим теперь *глобальный априорный байесовский риск* (последнее выражение случайно и зависит от значений $x_1, \cdots, x_n$ независимых одинаково распределённых случайных величин $X_1, \cdots, X_n$, имеющих такое же распределение, как и случайная величина X) по формуле

$$r_n(\delta, Q) = \int\limits_{\mathcal{X}} \mathsf{E} h_Q(\delta_n(x), x) d\nu(x), \tag{18.2.5}$$

где E обозначает математическое ожидание относительно независимых одинаково распределённых случайных величин $X_1, \cdots, X_n$, имеющих общую плотность относительно меры $\nu(\cdot)$ вида

$$p_Q(x) = \int\limits_{\Theta} p_\theta(x) dQ(\theta). \tag{18.2.6}$$

Выражение (18.2.5) можно переписать в виде

$$r_n(\delta, Q) = \int\limits_{\mathcal{X}} p_Q(x) \Big(\int\limits_{\Theta} q(\theta \mid x) \mathsf{E} L(\theta, \delta_n(x)) d\mu(\theta) \Big) d\nu(x) = \tag{18.2.7}$$

$$= \int\limits_{\mathcal{X}} p_Q(x) \Big(\int\limits_{\mathcal{X}^n} \Big(\int\limits_{\Theta} L(\theta, \delta_n(x_1, \cdots, x_n; x)) q(\theta \mid x) d\mu(\theta) \Big) \prod_{i=1}^{n} p_Q(x_i) d\nu(x_i) \Big) d\nu(x),$$

которое является непосредственным аналогом выражения (18.2.1) и поэтому интерпретируется в терминах апостериорных распределений.

При таком определении (18.2.7) байесовского риска $r_n(\delta, Q)$ всюду далее все вероятности и математические ожидания рассматривются относительно независимых одинаково распределённых случайных величин $(X_1, \cdots, X_n)$, имеющих общую плотность $p_Q(x)$ вида (18.2.6) с неизвестным априорным распределением $Q(\cdot)$. Таким образом имея значения таких случайных величин

$$X_1 = x_1, \cdots, X_n = x_n$$

мы наблюдаем случайную величину X_{n+1}, имеющую условную плотность $p_\theta(x)$ и хотим "оценить" текущее значение θ. С помощью случайных величин $(X_1, \cdots, X_n)$ мы "оцениваем" неизвестное априорное распределение $Q(\cdot)$.

Из соотношений (18.1.5) и (18.2.5) непосредственно следует, что всегда

$$r_n(\delta, Q) \geq r(Q). \tag{18.2.8}$$

Определение 18.2.2. *Если*

$$\operatorname*{Lim}_{n \to \infty} r_n(\delta, Q) = r(Q), \tag{18.2.9}$$

то последовательность решающих функций δ называется асимптотически оптимальной отностительно априорного распределения $Q(\cdot)$. Основная проблема состоит в нахождении последовательности решающих функции δ, которая была бы асимптотически оптимальна относительно некоторого *класса* $\Pi = \{Q(\cdot)\}$ априорных распределений, который содержит неизвестное истинное распределение $Q(\cdot)$. В частности может ли быть Π классом *всех* априорных распределений на Θ?

18.3 СПИСОК ЛИТЕРАТУРЫ

1) H. Robbins, The empirical Bayes approach to statistics,
Proc. Third Berkley Symp. Math. Statist. Probab., 1955, v.1, p. 157 – 164.

2) H. Robbins, The empirical Bayes approach to statistical decision problems,
Ann. Math. Statist., 1964, v.35, p. 1– 20.

3) Л.Н. Большев, Приложения эмпирического байесовского подхода,
Труды Международного Конгресса Математиков, Ницца 1970, Москва, Наука, 1972, стр. 48 – 55.

4) J.S. Maritz, Empirical Bayes Methods,
Methuen and Co LTD, London, 1970, Chapter 1.

5) Ш. Закс, Теория Статистических Выводов,
Москва, Наука, 1975, Глава 6, ¡6.9.

Лекция 19

В Лекции строятся асимптотически эффективные байесовские решающие функции.

19.1 АСИМПТОТИЧЕСКАЯ ОПТИМАЛЬНОСТЬ

Напомним, что (см. (18.1.7) и (18.2.5))

$$r(Q) = r(\delta_Q, Q) = \int\limits_{\mathcal{X}} h_Q(\delta_Q(x), x) d\nu(x),$$

$$r_n(\delta, Q) = \int\limits_{\mathcal{X}} \mathsf{E} h_Q(\delta_n(x), x) d\nu(x),$$

поэтому, учитывая Теорему Лебега о мажорируемой сходимости (см. Лекция 2, п.8), непосредственно из определения асимптотической оптимальности (Определение 18.2.2) следует, что для асимптотической оптимальности последовательности решающих функций δ относительно априорного распределения $Q(\cdot)$ достаточно, чтобы

(A) $\quad \mathrm{Lim}_{n\to\infty} \mathsf{E} h_Q(\delta_n(x), x) = h_Q(\delta_Q(x), x), \qquad \nu$ – почти всюду

и

(B) $\quad \mathsf{E} h_Q(\delta_n(x), x) \le H(x), \qquad$ причём $\quad \int\limits_{\mathcal{X}} H(x) d\nu(x) < +\infty.$

Основная проблема состоит в доказательстве соотношения (A). Для проверки неравенства (B) предположим, что

(C) $\quad \int\limits_{\Theta} L(\theta) dQ(\theta) < +\infty,$

где
$$0 \le L(\theta) = \sup_{\delta\in\Delta} L(\theta,\delta) \ \le\ +\infty. \tag{19.1.1}$$
Тогда полагая
$$H(x) = \int\limits_{\Theta} L(\theta) p_\theta(x) dQ(\theta) \ \ge\ 0, \tag{19.1.2}$$
и учитывая соотношение (18.1.3), получаем неравенство
$$h_Q(\delta,x) = \int\limits_{\Theta} L(\theta,\delta) p_\theta(x) dQ(\theta) \ \le\ H(x). \tag{19.1.3}$$
Тогда в силу предположения (C), имеем
$$\int\limits_{\mathcal{X}} H(x) d\nu(x) = \int\limits_{\Theta} L(\theta) \int\limits_{\mathcal{X}} p_\theta(x) d\nu(x) dQ(\theta) =$$
$$= \int\limits_{\Theta} L(\theta) dQ(\theta) < +\infty. \tag{19.1.4}$$
Из соотношений (18.2.3) и (18.2.4) теперь следует справедливость неравенства (B). Более того из (18.2.4) также следует, что
$$H(x) < +\infty, \qquad \nu - \text{почти всюду} \tag{19.1.5}$$
и значит для доказательства (A) достаточно показать, что (функция $h_Q(\delta,x)$ равномерно ограничена по δ (см.(19.1.3))

(D) $\quad \mathsf{P} - \mathrm{Lim}_{n\to\infty}\, h_Q(\delta_n(x),x) = h_Q(\delta_Q(x),x), \quad \nu$ – почти всюду,

где $\mathsf{P} - \lim$ означает предел по вероятности относительно распределения независимых случайных величин $(X_1,\cdots,X_n)$ с плотностью (18.2.6). Таким образом для доказательства асимптотической оптимальности δ относительно $Q(\cdot)$ достаточно проверить соотношения (C) и (D).

Рассмотрим соотношение (D). Пусть δ_0 – произвольный фиксированный элемент множества Δ и определим
$$\Delta_Q(\delta,x) = \int\limits_{\Theta} [L(\theta,\delta) - L(\theta,\delta_0)] p_\theta(x) dQ(\theta) \tag{19.1.6}$$
и
$$L_0(x) = \int\limits_{\Theta} L(\theta,\delta_0) p_\theta(x) dQ(\theta). \tag{19.1.7}$$

При выполнении условия (C) для ν – почти всех x мы можем записать

$$h_Q(\delta, x) = L_0(x) + \Delta_Q(\delta, x). \tag{19.1.8}$$

Предположим, что мы можем найти последовательность функций

$$\Delta_n(\delta, x) = \Delta_n(x_1, \cdots, x_n;\ \delta, x) \tag{19.1.9}$$

такую, что для ν – почти всех x справедливо соотношение

$$\mathsf{P} - \operatorname*{Lim}_{n\to\infty} \sup_{\delta\in\Delta} |\Delta_n(\delta, x) - \Delta_Q(\delta, x)| = 0. \tag{19.1.10}$$

Пусть ε_n произвольная последовательность чисел, стремящаяся к нулю и определим последовательность решающих функций

$\delta_n(x) = \delta_n(x_1, \cdots, x_n;\ x) =$ произвольному элементу $\delta^* \in \Delta$ такому, что

$$\Delta_n(\delta^*, x) \ \le\ \inf_{\delta\in\Delta} \Delta_n(\delta, x) + \varepsilon_n. \tag{19.1.11}$$

Очевидно, справедливо соотношение (см. (18.1.5) и (19.1.8))

$$0 \ \le\ \Delta_Q(\delta_n(x), x) - \Delta_Q(\delta_Q(x), x) = \tag{19.1.12}$$

$$= [\Delta_Q(\delta_n(x), x) - \Delta_n(\delta_n(x), x)] + [\Delta_n(\delta_n(x), x) - \Delta_n(\delta_Q(x), x)] +$$
$$+ [\Delta_n(\delta_Q(x), x) - \Delta_Q(\delta_Q(x), x)].$$

Теперь для любого $\varepsilon > 0$, используя (19.1.10) и (19.1.11), получим, что для достаточно больших n с вероятностью сколь угодно близкой к единице, правая часть (19.1.12) не превосходит $\varepsilon + \varepsilon_n$. Таким образом

$$\mathsf{P} - \operatorname*{Lim}_{n\to\infty} \Delta_Q(\delta_n(x), x) = \Delta_Q(\delta_Q(x), x), \qquad \nu - \text{почти всюду}, \tag{19.1.13}$$

поэтому из (19.1.8) следует соотношение (D). Таким образом доказана

Теорема 19.1.1 *Пусть априорное распределение $Q(\cdot)$ удовлетворяет условию (C)*

$$\int\limits_{\Theta} L(\theta) dQ(\theta) < +\infty, \quad 0 \ \le\ L(\theta) = \sup_{\delta\in\Delta} L(\theta, \delta) \ \le\ +\infty$$

и пусть $\Delta_n(\delta, x) = \Delta_n(x_1, \cdots, x_n; \delta, x)$ – последовательность функций вида (19.1.9), которая удовлетворяет соотношению (19.1.10)

$$\mathsf{P} - \operatorname*{Lim}_{n\to\infty} \sup_{\delta\in\Delta} |\Delta_n(\delta, x) - \Delta_Q(\delta, x)| = 0,$$

где

$$\Delta_Q(\delta, x) = \int\limits_{\Theta} [L(\theta, \delta) - L(\theta, \delta_0)] p_\theta(x) dQ(\theta).$$

Определим последовательность $\delta = \{\delta_n\}$ с помощью соотношения (19.1.11)

$$\delta_n(x) = \delta_n(x_1, \cdots, x_n;\, x) = \text{произвольному элементу } \delta^* \in \Delta \text{ такому, что}$$

$$\Delta_n(\delta^*, x) \ \le \ \inf_{\delta \in \Delta} \Delta_n(\delta, x) + \varepsilon_n, \quad 0 \ \le \ \varepsilon_n \to 0.$$

Тогда последовательность решающих функций $\delta = \{\delta_n\}$ асимптотически оптимальна относительно априорного распределения $Q(\cdot)$.

В случае, если множество Δ *конечно* справедливо следующее

Следствие 19.1.1. *Пусть $\Delta = \{\delta_0, \cdots, \delta_m\}$ – конечное множество и априорное распределение $Q(\cdot)$ таково, что*

$$\int\limits_{\Theta} L(\theta, d_j) dQ(\theta) < +\infty, \qquad j = 0, \cdots, m \tag{19.1.14}$$

и пусть $\Delta_{j,n}(x) = \Delta_{j,n}(x_1, \cdots, x_n;\, x)$, $\quad j = 1, \cdots, m; \quad n = 1, 2, \cdots$ – такая последовательность функций, что

$$\mathsf{P} - \operatorname*{Lim}_{n \to \infty} \Delta_{j,n}(x) = \int\limits_{\Theta} [L(\theta, d_j) - L(\theta, d_0)] p_\theta(x) dQ(\theta), \quad \nu - \text{почти всюду}. \tag{19.1.15}$$

Положим $\Delta_{0,n}(x) = 0$ и определим

$$\delta_n(x) = \delta_k, \qquad \text{где } 0 \ \le \ k \ \le \ m \text{ – произвольное число такое, что}$$

$$\Delta_{k,n}(x) = \min[0, \Delta_{1,n}(x), \cdots, \Delta_{m,n}(x)]. \tag{19.1.16}$$

Тогда последовательность $\delta = \{\delta_n\}$ асимптотически оптимальна относительно априорного распределения $Q(\cdot)$.

В важном частном случае $m = 1$ (*проверка гипотез*) это Следствие приобретает следующий вид.

Следствие 19.1.2. *Пусть $\Delta = \{\delta_0, \delta_1\}$ и априорное распределение $Q(\cdot)$ таково, что*

$$\int\limits_{\Theta} L(\theta, \delta_j) dQ(\theta) < +\infty, \qquad j = 0, 1 \tag{19.1.17}$$

и пусть функция $\Delta_n(x) = \Delta_n(x_1, \cdots, x_n;\, x)$ такая, что

$$\mathsf{P} - \operatorname*{Lim}_{n \to \infty} \Delta_n(x) = \Delta_Q(x) =$$

$$= \int_{\Theta} [L(\theta,\delta_1) - L(\theta,\delta_0)] p_\theta(x) dQ(\theta), \qquad \nu - \textit{почти всюду}. \tag{19.1.18}$$

Определим решающую функцию

$$\delta_n(x) = \begin{cases} \delta_0, & \textit{если} \;\; \Delta_n(x) \geq 0, \\ \delta_1, & \textit{если} \;\; \Delta_n(x) < 0. \end{cases} \tag{19.1.19}$$

Тогда последовательность $\delta = \{\delta_n\}$ асимптотически оптимальна относительно априорного распределения $Q(\cdot)$.

Ниже будет приведён пример последовательности $\Delta_n(x)$, удовлетворяющей соотношению (19.1.18).

19.2 СЛУЧАЙ РАСПРЕДЕЛЕНИЯ ПУАССОНА

Рассмотрим задачу проверки односторонней гипотезы вида

$$\mathbf{H}_0 : \theta \leq \theta^*$$

о параметре распределения Пуассона θ (значение θ^* известно). Пусть

$$\Theta = \{\theta : 0 < \theta < +\infty\}, \quad \Delta = \{\delta_0, \delta_1\},$$

где решения δ_0 и δ_1 интерпретируются как

$$\delta_0 - \text{"принять гипотезу } \mathbf{H}_0\text{"}, \qquad \delta_1 - \text{"отвергнуть гипотезу } \mathbf{H}_0\text{"}.$$

Далее

$$\mathcal{X} = \{0, 1, 2, \cdots\}, \qquad p_\theta(x) = \frac{e^{-\theta}\theta^x}{x!} \tag{19.2.1}$$

и $\nu(\cdot)$ – считающая мера на действительной прямой $\mathbf{R}^1$. Определим теперь функцию потерь

$$L(\theta,\delta_0) = \begin{cases} 0, & \text{если} \;\; \theta \leq \theta^*, \\ \theta - \theta^*, & \text{если} \;\; \theta \geq \theta^*, \end{cases} \tag{19.2.2}$$

$$L(\theta,\delta_1) = \begin{cases} \theta^* - \theta, & \text{если} \;\; \theta \leq \theta^*, \\ 0, & \text{если} \;\; \theta \geq \theta^*. \end{cases}$$

Из определения функции потерь непосредственно следует, что

$$L(\theta,\delta_1) - L(\theta,\delta_0) = \theta^* - \theta, \qquad (0 < \theta < +\infty) \tag{19.2.3}$$

и

$$\Delta_Q(x) = \int\limits_{\Theta} [L(\theta, \delta_1) - L(\theta, \delta_0)] p_\theta(x) dQ(\theta) = \frac{1}{x!} \int\limits_0^{+\infty} (\theta^* - \theta) e^{-\theta} \theta^x dQ(\theta). \quad (19.2.4)$$

Далее используя (18.2.6), получим

$$p_Q(x) = \mathsf{P}\{X_j = x\} = \frac{1}{x!} \int\limits_0^{+\infty} e^{-\theta} \theta^x dQ(\theta) \qquad (19.2.5)$$

и значит мы можем записать

$$\Delta_Q(x) = \theta^* p_Q(x) - (x+1) p_Q(x+1). \qquad (19.2.6)$$

Определим функцию

$$\alpha(x, y) = \begin{cases} 1, & \text{если} \quad x = y, \\ 0, & \text{если} \quad x \neq y \end{cases} \qquad (19.2.7)$$

и рассмотрим выражение

$$u_n(x) = u_n(x_1, \cdots, x_n; x) = \frac{1}{n} \sum_{i=1}^{n} \alpha(x, x_i). \qquad (19.2.8)$$

Заметим, что

$$u_n(x) = \frac{\text{число} \quad x_1, \cdots, x_n \quad \text{равных} \quad x}{n}$$

и

$$\mathsf{E}\alpha(x, X_j) = \mathsf{P}\{X_j = x\} = p_Q(x). \qquad (19.2.9)$$

Из Закона Больших Чисел следует (см. Лекция 4, п.6), что

$$\mathsf{P} - \lim_{n \to \infty} u_n(x) = \mathsf{E}\alpha(x, X_1) = p_Q(x), \qquad x = 0, 1, 2, \cdots. \qquad (19.2.10)$$

Поэтому если положить (см. (19.2.6))

$$\Delta_n(x) = \theta^* u_n(x) - (x+1) u_n(x+1), \qquad (19.2.11)$$

то непосредственно из соотношения (19.2.10) следует, что

$$\mathsf{P} - \lim_{n \to \infty} \Delta_n(x) =$$

$$= \theta^* p_Q(x) - (x+1)p_Q(x+1) = \Delta_Q(x), \quad x = 0, 1, 2, \cdots. \tag{19.2.12}$$

Теперь из Следствия 19.1.2 следует, что асимптотически оптимальная решающая функция имеет вид

$$\delta_n(x) = \begin{cases} \delta_0, & \text{если} \quad \theta^* u_n(x) - (x+1)u_n(x+1) \ \geq \ 0, \\ \delta_1, & \text{в противном случае} \end{cases} \tag{19.2.13}$$

для всех априорных распределений $Q(\cdot)$, обладающих свойством

$$\int_0^{+\infty} \theta dQ(\theta) < +\infty. \tag{19.2.14}$$

Заметим, что можно было бы определить $u_n(x)$ как (см. (19.2.8))

$$u_n(x) = \frac{1}{n+1} \sum_{j=1}^{n+1} \alpha(x, x_j), \tag{19.2.15}$$

тогда соотношение (19.2.10) выполняется для функции $u_n(x)$, определённой по формуле (19.2.15). Используя (19.2.15), соответствующая асимптотически оптимальная решающая функция δ имеет вид: мы должны принять гипотезу $\mathbf{H}_0$ (решение δ_0), касающуюся параметра θ_{n+1}, если

$$\theta^* \ \geq \ (x_{n+1}+1) \frac{\text{число наблюдений} \ \ x_1, \cdots, x_{n+1} \ \ \text{равных} \ \ x_{n+1}+1}{\text{число наблюдений} \ \ x_1, \cdots, x_{n+1} \ \ \text{равных} \ \ x_{n+1}}.$$

Соотношение (19.2.6) является основным для построения $\Delta_n(x)$, удовлетворяющего (19.1.18). Однако, соотношение (19.2.6) является специфическим свойством рапределения Пуассона (19.2.1) и функции потерь (19.2.2) и поэтому может показаться, что применение Следствия 19.1.2 к проверке гипотез весьма ограничительно, однако это не так. В следующей Лекции будет развита общая теория.

Задача. Рассмотрим также задачу проверки односторонней гипотезы вида

$$\mathbf{H}_0 : \theta \ \leq \ \theta^*$$

о параметре геометрического распределения θ (значение $\theta^* \in (0,1)$ известно). Пусть

$$\Theta = \{\theta : 0 < \theta < 1\}, \quad \Delta = \{\delta_0, \delta_1\},$$

где решения δ_0 и δ_1 интерпретируются как

$$\delta_0 - \text{"принять гипотезу } \mathbf{H}_0\text{"}, \quad \delta_1 - \text{"отвергнуть гипотезу } \mathbf{H}_0\text{"}.$$

Далее

$$\mathcal{X} = \{0, 1, 2, \cdots\}, \quad p_\theta(x) = (1-\theta)\theta^x$$

и $\nu(\cdot)$ – считающая мера на действительной прямой $\mathbf{R}^1$. Определим теперь функцию потерь

$$L(\theta, \delta_0) = \begin{cases} 0, & \text{если } \theta \le \theta^*, \\ \theta - \theta^*, & \text{если } \theta \ge \theta^*, \end{cases}$$

$$L(\theta, \delta_1) = \begin{cases} \theta^* - \theta, & \text{если } \theta \le \theta^*, \\ 0, & \text{если } \theta \ge \theta^*. \end{cases}$$

Найти асимптотически оптимальную решающую функцию δ относительно класса всех априорных распределений $Q(\cdot)$ таких, что

$$\int_0^1 \theta dQ(\theta) < +\infty.$$

19.3 СПИСОК ЛИТЕРАТУРЫ

1) H. Robbins, The empirical Bayes approach to statistical decision problems, Ann. Math. Statist., 1964, v.35, p. 1– 20.

2) J.S. Maritz, Empirical Bayes Methods, Methuen and Co LTD, London, 1970, Chapter 1.

3) Ш. Закс, Теория Статистических Выводов, Москва, Наука, 1975, Глава 6, ¡6.9.

Лекция 20

В Лекции рассмотрен один из возможных методов оценки априорного распределения.

20.1 ОЦЕНКА АПРИОРНОГО РАСПРЕДЕЛЕНИЯ: ОБЩИЙ СЛУЧАЙ

Рассмотрим более подробно общую схему эмпирического байесовского подхода, описанную в Следствиях 19.1.1 и 19.1.2. Для простоты предположим, что

$$\Delta = \{\delta_0,\, \delta_1\}, \quad \mathcal{X} = \Theta = \mathbf{R}^1.$$

Напомним, что последовательность $\delta = \{\delta_n\}$ решающих функций является асимптотически оптимальной относительно класса Π априорных распределений $Q(\cdot)$, определённого соотношением (19.1.17)

$$\Pi = \left\{ Q(\cdot) : \int\limits_{\Theta} L(\theta, \delta_j) dQ(\theta) < +\infty, \;\; j = 0,1 \right\},$$

если можно найти последовательность функций

$$\Delta_n(x) = \Delta_n(x_1, \cdots, x_n; x),$$

такую, что

$$\mathsf{P}-\operatorname*{Lim}_{n\to\infty} \Delta_n(x) = \Delta_Q(x) = \int\limits_{\Theta} [L(\theta,\delta_1) - L(\theta, d_0)] p_\theta(x) dQ(\theta), \quad \nu-\text{п.в.} \quad (20.1.1)$$

для всех $Q \in \Pi$.

Один из возможных путей построения такой последовательности (отличный от описанного в Лекции 19) состоит в нахождении последовательности случайных функций распределения

$$Q_n(\theta) = Q_n(x_1, \cdots, x_n; \theta),$$

такой, что

$$\mathsf{P}\Big(\operatorname*{Lim}_{n\to\infty} Q_n(\theta) = Q(\theta), \quad \text{в каждой точке непрерывности } \ Q(\theta)\Big) = 1, \quad (20.1.2)$$

то есть последовательность $Q_n(\theta)$ сходится почти всюду к $Q(\theta)$ *слабо.* Здесь мы обозначили функцию распределения, соответствующую априорному распределению $Q(\cdot)$, через $Q(\theta)$. Имея такую последовательность $Q_n(\theta)$, положим

$$\Delta_n(x) = \int\limits_{\Theta} [L(\theta, \delta_1) - L(\theta, \delta_0)] p_\theta(x) dQ_n(\theta). \quad (20.1.3)$$

Теперь, если предположить, что для ν – почти всех $x \in \mathcal{X}$ функция

$$[L(\theta, \delta_1) - L(\theta, \delta_0)] p_\theta(x) \quad (20.1.4)$$

ограничена и непрерывна по θ, то из определения слабой сходимости следует соотношение (20.1.1). Заметим, что здесь возникает *основная трудность* в построении такой последовательности $Q_n(\theta)$, поскольку мы наблюдаем случайные величины $(X_1, \cdots, X_n)$, имеющие функции распределения $F_Q(x)$, а не независимые копии случайной величины Ξ, имеющие функцию распределения $Q(\theta)$.

Рассмотрим так называемый метод "минимума расстояния", предложенный Волфовицем (J. Wolfowitz), построения последовательности оценок $Q_n(\theta)$ функции распределения $Q(\theta)$, удовлетворяющей равенству (20.1.2) для любого $Q \in \Pi$.

С этой целью ослабим условия регулярности на семейство плотностей $p_\theta(x)$ и будем предпологать лишь, что соответствующие функции распределения $F(x, \theta)$, $\theta \in \Theta$ являются при каждом фиксированном $x \in \mathcal{X} = \mathbf{R}^1$ борелевскими функциями. (Функция $F(x)$, определённая при $x \in \mathcal{X}$, называется функцией распределения, если она неубывает, непрерывна слева и $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to+\infty} F(x) = 1$.)

Для любой функции распределения $Q(\theta)$ определим *смесь* функций распределения $F(x, \theta)$ по формуле

$$F_Q(x) = \int\limits_{-\infty}^{+\infty} F(x, \theta) dQ(\theta), \quad (20.1.5)$$

тогда $F_Q(x)$ также является функцией распределения на $\mathcal{X}$.

Пусть $(X_1, \cdots, X_n)$ являются последовательностью независимых одинаково распределённых случайных величин с общей функцией распределения $F_Q(x)$. Определим эмпирическую функцию распределения $F_n(x)$ (см. Лекция 6, Пример)

$$F_n(x) = \frac{\text{число} \quad X_1, \cdots, X_n \quad \text{меньших} \quad x}{n} \tag{20.1.6}$$

и для любых двух функций распределения $F_1(x),\ F_2(x)$ определим расстояние

$$\rho(F_1, F_2) = \sup_x |F_1(x) - F_2(x)|. \tag{20.1.7}$$

Пусть $0 < \varepsilon_n$ – произвольная последовательность, стремящаяся к нулю. Для произвольного класса $\bar{\Pi}$ априорных распределений $\bar{Q}(\cdot)$, содержащего "истинное" априорное распределение $Q(\cdot)$, определим величину

$$\gamma_n = \inf_{\bar{Q} \in \bar{\Pi}} \rho(F_n, F_{\bar{Q}}). \tag{20.1.8}$$

Пусть

$$Q_n(\theta) = Q_n(x_1, \cdots, x_n; \theta),$$

любой элемент из множества $\bar{\Pi}$, удовлетворяющий неравенству

$$\rho(F_n, F_{Q_n}) \ \leq \ \gamma_n + \varepsilon_n. \tag{20.1.9}$$

Будем называть так определённую последовательность $Q_n(\theta)$ *эффективной* для класса $\bar{\Pi}$, если соотношение (20.1.2) выполнено для всех $Q \in \bar{\Pi}$.

Теорема 20.1.1. *Предположим, что*

1) *Для любого фиксированного $x \in \mathbf{R}^1$ функции распределения $F(x,\theta)$ непрерывны по $\theta \in \Theta$.*

2) *Пределы вида*

$$F_{-\infty}(x) = \lim_{\theta \to -\infty} F(x,\theta), \quad F_{+\infty}(x) = \lim_{\theta \to +\infty} F(x,\theta)$$

существуют для всех x.

3) *Функции $F_{-\infty}(x)$ и $F_{+\infty}(x)$ не являются функциями распределения.*

4) *Если $Q_1(\theta)$ и $Q_2(\theta)$ любые функции распределения такие, что*

$$F_{Q_1}(x) \equiv F_{Q_2}(x),$$

то и

$$Q_1(\theta) \equiv Q_2(\theta).$$

(Это так называемое свойство РАЗДЕЛИМОСТИ *или* ИДЕНТИФИЦИРУЕМОСТИ *смесей.)*

Тогда последовательность $Q_n(\theta)$, определенная соотношением (20.1.9), является эффективной для класса $\bar{\Pi}$ всех *априорных распределений.*

Доказательство. По Теореме Гливенко – Кантелли (см. [4], стр. 28), имеем

$$\mathsf{P}\Big(\operatorname*{Lim}_{n\to\infty} \rho(F_n, F_Q) = 0\Big) = 1. \tag{20.1.10}$$

Далее, поскольку

$$\rho(F_{Q_n}, F_Q) \ \le\ \rho(F_{Q_n}, F_n) + \rho(F_n, F_Q) \ \le \tag{20.1.11}$$

$$\le\ \gamma_n + \varepsilon_n + \rho(F_n, F_Q)\ \le\ \rho(F_n, F_Q) + \varepsilon_n + \rho(F_n, F_Q),$$

поэтому из соотношения (20.1.10) следует, что с вероятностью единица, равномерно по x, справедливо равенство

$$\operatorname*{Lim}_{n\to\infty} F_{Q_n}(x) = \operatorname*{Lim}_{n\to\infty} \int\limits_{-\infty}^{+\infty} F(x,\theta) dQ_n(\theta) = \tag{20.1.12}$$

$$= F_Q(x) = \int\limits_{-\infty}^{+\infty} F(x,\theta) dQ(\theta).$$

Теперь докажем, что выполняется соотношение (20.1.2). Для этого рассмотрим фиксированную последовательность $X_1 = x_1, \cdots, X_n = x_n$ такую, что выполняется равенство (20.1.12). Из Теоремы Хелли (см. [3], стр. 340) следует, что из любой подпоследовательности функций распределения $Q_n(\theta)$ можно выделить подпоследовательность $Q_{k_n}(\theta)$ такую, что

$$Q_{k_n}(\theta) \to Q^*(\theta)$$

в каждой точке непрерывности $Q^*(\theta)$, где $Q^*(\theta)$ – неубывающая, непрерывная слева функция такая, что

$$0\ \le\ Q^*(-\infty)\ \le\ Q^*(+\infty)\ \le\ 1.$$

Поэтому определяя функцию распределения

$$\tilde{Q}^*(\theta) = \begin{cases} 1, & \text{если} \quad x = +\infty, \\ Q^*(\theta), & \text{если} \quad -\infty < x < +\infty, \\ 0, & \text{если} \quad x = -\infty, \end{cases}$$

из определения слабой сходимости и Предположений (1) и (2) следует, что

$$\operatorname*{Lim}_{n\to\infty} \int_{-\infty}^{+\infty} F(x,\theta) dQ_{k_n}(\theta) = \int_{-\infty}^{+\infty} F(x,\theta) d\tilde{Q}^*(\theta) =$$

$$= \int_{-\infty}^{+\infty} F(x,\theta) dQ^*(\theta) + Q^*(-\infty) F_{-\infty}(x) + (1 - Q^*(+\infty)) F_{+\infty}(x) \qquad (20.1.13)$$

и поэтому из (20.1.12) следует, что

$$\int_{-\infty}^{+\infty} F(x,\theta) dQ(\theta) = \int_{-\infty}^{+\infty} F(x,\theta) dQ^*(\theta) + Q^*(-\infty) F_{-\infty}(x) + (1 - Q^*(+\infty)) F_{+\infty}(x). \qquad (20.1.14)$$

Если мы покажем, что

$$Q^*(-\infty) = 0 \quad \text{и} \quad Q^*(+\infty) = 1,$$

то из Предположения (4) будет следовать , что

$$Q(\theta) \equiv Q^*(\theta)$$

и значит $Q(\theta)$ является слабым пределом *любой* сходящейся подпоследовательности $Q_n(\theta)$ и поэтому справедливо (20.1.2). Итак для завершения доказательства Теоремы достаточно показать, что из Предположения (3) следует, что

$$Q^*(-\infty) = 0 \quad \text{и} \quad Q^*(+\infty) = 1.$$

Поскольку $F_{-\infty}(x)$ является пределом при $\theta \to -\infty$ функций $F(x,\theta)$, то $F_{-\infty}(x)$ является неубывающей функцией x такой, что

$$0 \le F_{-\infty}(-\infty) \le F_{-\infty}(+\infty) \le 1. \qquad (20.1.15)$$

Аналогичное неравенство справедливо и для $F_{+\infty}(x)$. Пусть теперь в равенстве (20.1.14) $x \to -\infty$. Тогда по Теореме о мажорируемой сходимости (Лекция 2, п. 8), имеем

$$0 = Q^*(-\infty) F_{-\infty}(-\infty) + (1 - Q^*(+\infty)) F_{+\infty}(-\infty). \qquad (20.1.16)$$

Поэтому, если
$$Q^*(-\infty) \neq 0,$$
то $F_{-\infty}(-\infty) = 0$ и если $Q^*(+\infty) \neq 1$, то $F_{+\infty}(-\infty) = 0$. Аналогично полагая $x \to +\infty$ в (20.1.14) мы видим, что если $Q^*(-\infty) \neq 0$, то $F_{-\infty}(+\infty) = 1$ и если $Q^*(+\infty) \neq 1$, то $F_{+\infty}(+\infty) = 1$.

Пусть теперь a_n любая последовательность чисел, сходящаяся к a слева. Тогда из (20.1.14), полагая $x = a_n$, $n \to +\infty$ и вычитая (20.1.14) с $x = a$, получим

$$Q^*(-\infty)(F_{-\infty}(a) - F_{-\infty}(a-0)) + (1 - Q^*(+\infty))(F_{+\infty}(a) - F_{+\infty}(a-0)) = 0. \tag{20.1.17}$$

Следовательно, если $Q^*(-\infty) \neq 0$, то $F_{-\infty}(a-0) = F_{-\infty}(a)$, и если $Q^*(+\infty) \neq 1$, то $F_{+\infty}(a-0) = F_{+\infty}(a)$. Таким образом, если $Q^*(-\infty) \neq 0$, то $F_{-\infty}(x)$ является функцией распределения и если $Q^*(+\infty) \neq 1$, то $F_{+\infty}(x)$ также функция распределения. Что противоречит Предположению (3). Итак получаем
$$Q^*(-\infty) = 0 \quad \text{и} \quad Q^*(+\infty) = 1.$$
□

20.2 ПРИМЕРЫ

Пример 20.2.1. (параметр сдвига) Пусть $F(x)$ – непрерывная функция распределения с характеристической функцией нигде не обращающейся в ноль

$$f_F(t) = \int_{-\infty}^{+\infty} e^{itx} dF(x) \neq 0, \qquad \text{для всех } t. \tag{20.2.1}$$

Положим
$$F(x,\theta) = F(x-\theta).$$

Тогда Предположения (1), (2), (3) Теоремы 20.1.1 выполняются. Проверим справедливость Предположения (4). Если $Q_1(\theta)$, $Q_2(\theta)$ две функции распределения такие, что $F_{Q_1}(x) \equiv F_{Q_2}(x)$, то есть

$$\int_{-\infty}^{+\infty} F(x-\theta)dQ_1(\theta) = \int_{-\infty}^{+\infty} F(x-\theta)dQ_2(\theta), \qquad \text{для всех } x, \tag{20.2.2}$$

тогда (поскольку эти интегралы являются свёртками)

$$f_F(t) f_{Q_1}(t) = f_F(t) f_{Q_2}(t), \qquad \text{для всех } t, \tag{20.2.3}$$

и значит

$$f_{Q_1}(t) = f_{Q_2}(t), \qquad \text{для всех } t, \tag{20.2.4}$$

то есть $Q_1(\theta) \equiv Q_2(\theta)$. Таким образом Предположение (4) также выполняется и значит согласно Теоремы 20.1.1 последовательность $Q_n(\theta)$, определённая соотношением (20.1.9) является эффективной относительно класса $\bar{\Pi}$ всех априорных распределениеи $Q(\cdot)$.

Заметим также, что характеристическая функция нормального закона также удовлетворяет соотношению (20.2.1).

В случае, если параметрическое пространство Θ не является всей действительной прямой $\mathbf{R}^1$ утверждение и доказательство Теоремы 20.1.1 нуждаются в модификации. Например, если $\Theta = [0, +\infty)$, то Теорема 20.1.1 приобретает следующий вид.

Теорема 20.2.1 *Предположим, что*

1) *Для любого фиксированного x функции распределения $F(x, \theta)$ непрерывны по $\theta \in \Theta$.*

2) *Предел вида*
$$F_{+\infty}(x) = \lim_{\theta \to +\infty} F(x, \theta)$$
существуют для всех x.

3) *Функция $F_{+\infty}(x)$ не является функцией распределения.*

4) *Если $Q_1(\theta)$ и $Q_2(\theta)$ любые функции распределения сосредоточенные на $\Theta = [0, +\infty)$ и такие, что*
$$F_{Q_1}(x) \equiv F_{Q_2}(x),$$
то и
$$Q_1(\theta) \equiv Q_2(\theta).$$

Тогда последовательность $Q_n(\theta)$, определенная соотношением (20.1.9), является эффективной для класса $\bar{\Pi}$ всех *априорных распределений, сосредоточенных на $\Theta = [0, +\infty)$.*

Приведём примеры использования этой Теоремы.

Пример 20.2.2. (распределение Пуассона) Пусть $\Theta = [0, +\infty)$ и

$$F(x,0) = \begin{cases} 0, & \text{если } \ x < 0, \\ 1, & \text{если } \ x \geq 0 \end{cases} \tag{20.2.5}$$

и для $0 < \theta < +\infty$ пусть

$$F(x,\theta) = \sum_{0 \leq i \leq x} \frac{e^{-\theta}\theta^i}{i!}. \tag{20.2.6}$$

Тогда Предположения (1), (2) и (3) выполняются.

Пусть $Q \in \bar{\Pi}$, тогда

$$F_Q(x) = \int\limits_{\Theta} F(x,\theta) dQ(\theta) = \sum_{0 \leq i \leq x} \int\limits_{\Theta} p_\theta(i) dQ(\theta), \tag{20.2.7}$$

где

$$p_0(i) = \begin{cases} 1, & \text{если } \ i = 0, \\ 0, & \text{если } \ i = 1, 2, \cdots \end{cases} \tag{20.2.8}$$

и

$$p_\theta(i) = \frac{e^{-\theta}\theta^i}{i!}, \quad \text{при } \ i = 0, 1, \cdots; \ \ 0 < \theta < +\infty.$$

Теперь

$$F_Q(0) = \int\limits_{\Theta} p_\theta(0) dQ(\theta), \quad F_Q(n) - F_Q(n-1) =$$

$$= \int\limits_{\Theta} p_\theta(n) dQ(\theta), \quad n = 1, 2, \cdots, \tag{20.2.9}$$

поэтому, если

$$F_{Q_1}(x) \equiv F_{Q_2}(x),$$

то

$$\int\limits_{\Theta} p_\theta(n) dQ_1(\theta) = \int\limits_{\Theta} p_\theta(n) dQ_2(\theta), \quad n = 0, 1, 2, \cdots. \tag{20.2.10}$$

Определим теперь функцию множеств

$$H_j(B) = \frac{\int\limits_B e^{-\theta} dQ_j(\theta)}{\int\limits_{\Theta} e^{-\theta} dQ_j(\theta)}, \quad j = 1, 2; \quad B \in \mathcal{B}^1, \tag{20.2.11}$$

тогда $H_j(\cdot)$ являются вероятностными мерами, определёнными на борелевских множествах. Поскольку из равенств (20.2.10) следует, что

$$C = \int\limits_{\Theta} e^{-\theta} dQ_1(\theta) = \int\limits_{\Theta} p_\theta(0) dQ_1(\theta) = \tag{20.2.12}$$

$$= \int\limits_{\Theta} p_\theta(0) dQ_2(\theta) = \int\limits_{\Theta} e^{-\theta} dQ_2(\theta),$$

то мы можем записать

$$H_j(B) = \frac{1}{C} \int\limits_{B} e^{-\theta} dQ_j(\theta), \quad j = 1, 2, \tag{20.2.13}$$

где $0 < C < +\infty$. Далее, поскольку

$$\frac{dH_j}{dQ_j}(\theta) = \frac{e^{-\theta}}{C}, \tag{20.2.14}$$

то для $n = 1, 2, \cdots$ и $j = 1, 2$ имеем

$$\int\limits_{\Theta} \theta^n dH_j(\theta) = \frac{1}{C} \int\limits_{\Theta} e^{-\theta} \theta^n dQ_j(\theta) = \frac{n!}{C} \int\limits_{\Theta} p_\theta(n) dQ_j(\theta), \tag{20.2.15}$$

поэтому из (20.2.10) следует, что

$$\alpha_n = \int\limits_{\Theta} \theta^n dH_1(\theta) = \int\limits_{\Theta} \theta^n dH_2(\theta), \quad n = 1, 2, \cdots. \tag{20.2.16}$$

Более того, поскольку

$$e^\theta n! = \left(1 + \theta + \frac{\theta^2}{2!} + \cdots + \frac{\theta^n}{n!} + \cdots\right) n! \ \geq \ \theta^n,$$

то

$$0 \ \leq \ e^{-\theta} \theta^n \ \leq \ n!, \qquad \text{при} \quad 0 \ \leq \ \theta < +\infty,$$

имеем

$$0 \ \leq \ \alpha_n = \frac{1}{C} \int\limits_{\Theta} e^{-\theta} \theta^n dQ_j(\theta) \ \leq \ \frac{n!}{C}, \quad n = 1, 2, \cdots, \tag{20.2.17}$$

поэтому имеем сходящийся ряд вида

$$\sum_{n=1}^{+\infty} \frac{\alpha_n}{n!2^n} < +\infty.$$

Отсюда по известной Теореме, касающейся проблемы моментов, получаем, что распределения $H_1(\cdot)$ и $H_2(\cdot)$ определяются *однозначно* своими моментами и значит тождественны (моменты α_n однозначно определяют распределение, если ряд

$$\sum_{n=1}^{+\infty} \frac{\alpha_n s^n}{n!}$$

сходится для некоторого значения $s \neq 0$ (см. [3], стр. 315)). Далее, поскольку

$$Q_j(B) = \int\limits_B \frac{dQ_j}{dH_j}(\theta) dH_j(\theta) = \int\limits_B Ce^{\theta} dH_j(\theta), \quad j = 1, 2, \tag{20.2.18}$$

то тождественны и $Q_1(\cdot)$, $Q_2(\cdot)$. Таким образом Предположение (4) также выполнено.

Приведём теперь пример, в котором $\Theta = (0, +\infty)$ и ноль играет роль $-\infty$ в Теореме 20.1.1.

Пример 20.2.3. (равномерное распределение) Для $\theta \in \Theta = (0, +\infty)$ определим функцию распределения равномерного распределения

$$F(x, \theta) = \begin{cases} 0, & \text{если} \quad x \le 0, \\ x/\theta, & \text{если} \quad 0 < x < \theta, \\ 1, & \text{если} \quad x \ge \theta. \end{cases} \tag{20.2.19}$$

Тогда

$$\lim_{\theta \to 0+} F(x, \theta) = \begin{cases} 0, & \text{если} \quad x \le 0, \\ 1, & \text{если} \quad x > 0 \end{cases} \tag{20.2.20}$$

и предел

$$\lim_{\theta \to +\infty} F(x, \theta) \equiv 0 \tag{20.2.21}$$

не являются функциями распределения. Итак Предположения (1), (2) и (3) выполняются.

Для любого априорного распределения $Q(\cdot)$, сосредоточенного на Θ, имеем при $x > 0$

$$F_Q(x) = \int\limits_{\Theta} F(x, \theta) dQ(\theta) = \int\limits_{\{0 < \theta \le x\}} 1 \cdot dQ(\theta) + x \int\limits_{\{\theta > x\}} \frac{dQ(\theta)}{\theta}. \tag{20.2.22}$$

Следовательно, если $F_{Q_1}(x) \equiv F_{Q_2}(x)$, то

$$Q_1(x) + x \int\limits_{\{\theta > x\}} \frac{dQ_1(\theta)}{\theta} = Q_2(x) + x \int\limits_{\{\theta > x\}} \frac{dQ_2(\theta)}{\theta}. \tag{20.2.23}$$

Если x является одновременно точкой непрерывности $Q_1(x)$ и $Q_2(x)$, то

$$\int\limits_{\{\theta > x\}} \frac{dQ_j(\theta)}{\theta} = \frac{Q_j(\theta)}{\theta}\Big|_x^{+\infty} + \int\limits_{\{\theta > x\}} \frac{Q_j(\theta)}{\theta^2} d\theta = \tag{20.2.24}$$

$$= -\frac{Q_j(x)}{x} + \int\limits_{\{\theta > x\}} \frac{Q_j(\theta)}{\theta^2} d\theta,$$

поэтому

$$Q_j(x) + x \int\limits_{\{\theta > x\}} \frac{dQ_j(\theta)}{\theta} = x \int\limits_{\{\theta > x\}} \frac{Q_j(\theta)}{\theta^2} d\theta, \tag{20.2.25}$$

и значит из соотношения (20.2.23) следует, что

$$\int\limits_{\{\theta > x\}} \frac{Q_1(\theta)}{\theta^2} d\theta = \int\limits_{\{\theta > x\}} \frac{Q_2(\theta)}{\theta^2} d\theta \tag{20.2.26}$$

в каждой точке непрерывности $x > 0$ функций распределения $Q_1(x)$ и $Q_2(x)$. Дифференцирование по x даёт $Q_1(x) = Q_2(x)$ для всех таких точек и значит $Q_1(x) \equiv Q_2(x)$.

Одним из недостатков метода, описанного в Теореме 20.1.1, построения оценки неизвестного априорного рапределения $Q(\cdot)$, является *неконструктивность* выбора оценки $Q_n(\theta)$, удовлетворяющей (20.1.9). Метод, описанный в Лекции 21 свободен от этого недостатка, но существенно зависит от конкретного параметрического семейства. В следующей Лекции будет описан иной метод оценки $Q(\cdot)$ в случае, если параметрическое пространство Θ является *конечным* множеством.

20.3 СПИСОК ЛИТЕРАТУРЫ

1) H. Robbins, The empirical Bayes approach to statistical decision problems, Ann. Math. Statist., 1964, v.35, p. 1– 20.

2) J.S. Maritz, Empirical Bayes Methods,
Methuen and Co LTD, London, 1970, Chapter 2.

3) А.Н. Ширяев,
Вероятность, Москва, Наука, 1989, Глава 3, ¡2.

4) М. Лоэв,
Теория Вероятности, Москва, Иностранная Литература, 1962, Вводная Часть, 2, ¡6.

Лекция 21

В Лекции рассмотрен метод оценки априорного распределения, сосредоточенного в конечном числе точек.

21.1 ОЦЕНКА АПРИОРНОГО РАСПРЕДЕЛЕНИЯ: КОНЕЧНЫЙ СЛУЧАЙ

Пусть теперь параметрическое пространство Θ *конечно*. Без ограничения общности можно считать, что оно имеет вид

$$\Theta = \{1, \cdots, r\}$$

и априорное распределение $Q(\cdot)$ на Θ задаётся вектором

$$(q_1, \cdots, q_r), \quad q_i \ \geq \ 0, \quad i = 1, \cdots, r, \quad \sum_{i=1}^{r} q_i = 1,$$

что мы будем обозначать в виде

$$Q = \{q_1, \cdots, q_r\}.$$

Таким образом

$$Q(\Xi = i) = q_i, \quad i = 1, \cdots, r.$$

Предположим также, что задано известное конечное семейство распределений на $\mathcal{X}$

$$\mathsf{P}_1(\cdot), \cdots, \mathsf{P}_r(\cdot)$$

и наблюдаемые независимые одинаково распределённые случайные величины $(X_1, \cdots, X_n)$ имеют распределение

$$\mathsf{P}_Q(X_1 \in B) = \int\limits_{\Theta} \mathsf{P}_\theta(B) dQ(\theta) = \sum_{i=1}^{r} q_i \mathsf{P}_i(B). \tag{21.1.1}$$

Нашей задачей является построение функций

$$q_{i,n} = q_{i,n}(x_1, \cdots, x_n) \tag{21.1.2}$$

таких, что

$$q_{i,n} \ \geq \ 0, \quad i = 1, \cdots, r, \quad \sum_{i=1}^{r} q_{i,n} = 1$$

и для любого априорного распределения $Q(\cdot)$ на Θ справедливо равенство

$$\mathsf{P}\Big(\operatorname*{Lim}_{n\to\infty} q_{i,n} = q_i, \quad i = 1, \cdots, r\Big) = 1. \tag{21.1.3}$$

Ясно, что *необходимое* условие для существования такой последовательности $q_{i,n}$ имеет вид

(A) Если

$$Q = \{q_1, \cdots, q_r\} \quad \text{и} \quad \bar{Q} = \{\bar{q}_1, \cdots, \bar{q}_r\}$$

два априорных распределения таких, что для любого борелевского множества $B \in \mathcal{B}^1$

$$\sum_{i=1}^{r} q_i \mathsf{P}_i(B) = \sum_{i=1}^{r} \bar{q}_i \mathsf{P}_i(B), \tag{21.1.4}$$

то $Q \equiv \bar{Q}$.

Докажем, что условие (A) является также и *достаточным* для сущестования такой последовательности.

Обозначим через $\lambda(\cdot)$ любую σ – конечную меру на $\mathcal{X}$, относительно которой все распределения $\mathsf{P}_i(\cdot)$ являются абсолютно непрерывными и такими, что их плотности

$$p_i(x) = \frac{d\mathsf{P}_i}{d\lambda}(x), \quad i = 1, \cdots, r$$

квадратично интегрируемы

$$\int\limits_{\mathcal{X}} p_i^2(x) d\lambda(x) < +\infty, \quad i = 1, \cdots, r. \tag{21.1.5}$$

Можно всегда, например, положить

$$\lambda(B) = \mathsf{P}_1(B) + \cdots + \mathsf{P}_r(B), \qquad B \in \mathcal{B}^1,$$

тогда

$$0 \ \leq \ p_i(x) \ \leq \ 1$$

и значит

$$\int\limits_{\mathcal{X}} p_i^2(x)d\lambda(x) \le \int\limits_{\mathcal{X}} p_i(x)d\lambda(x) = 1, \quad i = 1, \cdots, r. \tag{21.1.6}$$

Теперь функции $p_i(x), \quad i = 1, \cdots, r$ можно рассматривать как элементы гильбертова пространства H, порождённого измеримым пространством $(\mathcal{X}, \mathcal{B}^1, \lambda)$. Из условия (A) следует, что функции $p_i(x), \quad i = 1, \cdots, r$ *линейно независимы*. Поскольку, если

$$c_1 p_1(x) + \cdots + c_r p_r(x) \equiv 0$$

для некоторых констант $c_1, \cdots, c_r$, которые не все равны нулю, то изменяя обозначения, всегда можно записать

$$c_1 p_1(x) + \cdots + c_k p_k(x) \equiv c_{k+1} p_{k+1}(x) + \cdots + c_q p_q(x), \tag{21.1.7}$$

где $c_1, \cdots, c_q$ все *положительны* и $1 \le q \le r$. Интегрируя это тождество по всему пространству $\mathcal{X}$, получим

$$c_1 + \cdots + c_k = c_{k+1} + \cdots + c_q = c > 0 \tag{21.1.8}$$

и значит для *различных* априорных распределений

$$Q = \left\{ \frac{c_1}{c}, \cdots, \frac{c_k}{c}, 0, \cdots, 0 \right\} \ne \bar{Q} = \left\{ 0, \cdots, 0, \frac{c_{k+1}}{c}, \cdots, \frac{c_q}{c} \right\} \tag{21.1.9}$$

справедливо соотношение (21.1.4), что противоречит предположению (A).

Пусть теперь L_j обозначает линейное пространство, порождённое $r-1$ функцией $p_1(x), \cdots, p_{j-1}(x), p_{j+1}(x), \cdots, p_r(x)$. Тогда справедливо однозначное представление

$$p_j(x) = \bar{p}_j(x) + \tilde{p}_j(x), \quad j = 1, \cdots, r \tag{21.1.10}$$

с

$$\bar{p}_j(x) \in L_j, \quad \tilde{p}_j(x) \perp L_j, \quad \tilde{p}_j(x) \ne 0. \tag{21.1.11}$$

Полагая теперь

$$\psi_j(x) = \frac{\tilde{p}_j(x)}{\int\limits_{\mathcal{X}} (\tilde{p}_j(x))^2 d\lambda(x)}, \tag{21.1.12}$$

получим

$$\int\limits_{\mathcal{X}} \psi_j(x) p_k(x) d\lambda(x) = \begin{cases} 1, & \text{если } j = k, \\ 0, & \text{если } j \ne k. \end{cases} \tag{21.1.13}$$

Теперь определим

$$\bar q_{i,n} = \frac{1}{n}\sum_{l=1}^{n}\psi_i(x_l), \quad q_{i,n} = \frac{\bar q_{i,n}^{+}}{\sum\limits_{j=1}^{r}\bar q_{j,n}^{+}}, \tag{21.1.14}$$

где a^+ обозначает $\max(a,0)$. Если $(X_1,\cdots,X_n)$ – независимые одинаково распределённые случайные величины с общим распределением (21.1.1), то их общая плотность относительно меры $\lambda(\cdot)$ имеет вид

$$\sum_{j=1}^{r} q_j p_j(x), \tag{21.1.15}$$

поэтому принимая во внимание соотношение (21.1.13), получим

$$\mathsf{E}\psi_i(X_1) = \int\limits_{\mathcal{X}}\psi_i(x)\sum_{j=1}^{r}q_j p_j(x)d\lambda(x) =$$

$$= \sum_{j=1}^{r} q_j \int\limits_{\mathcal{X}}\psi_i(x)p_j(x)d\lambda(x) = q_j. \tag{21.1.16}$$

Теперь из Закона Больших Чисел непосредственно следует (21.1.3).

Применим теперь Теорему 19.1.1 с произвольным пространством решений Δ и функцией потерь $L(\theta,\delta)$, которая полностью определяется набором функций
$L(i,\delta),\ i=1,\cdots,r$. Предположим для простоты, что

$$0 \le L(i,\delta) \le L < +\infty, \quad \text{для всех } \delta\in\Delta \text{ и } i=1,\cdots,r. \tag{21.1.17}$$

Теперь соотношение (19.1.6) приобретает вид

$$\Delta_Q(\delta,x) = \sum_{i=1}^{r}[L(i,\delta)-L(i,\delta_0)]p_i(x)q_i \tag{21.1.18}$$

и если положить

$$\Delta_n(\delta,x) = \sum_{i=1}^{r}[L(i,\delta)-L(i,\delta_0)]p_i(x)q_{i,n}, \tag{21.1.19}$$

то нетрудно видеть, что справедливо неравенство

$$\sup_{\delta\in\Delta}|\Delta_n(\delta,x)-\Delta_Q(\delta,x)| \le L\sum_{i=1}^{r}p_i(x)|q_{i,n}-q_i|. \tag{21.1.20}$$

Поскольку $p_i(x) < +\infty$ для λ – почти всех x, то из соотношения (21.1.5) теперь следует, что с вероятностью единица выполнено равенство (19.1.10). Поэтому последовательность решающих функций $\delta = \{\delta_n\}$, определённая соотношнием (19.1.11), является асимптотически оптимальной для *любого* априорного распределения $Q = \{q_1, \cdots, q_r\}$.

Было бы интересно попытаться распространить описанный метод оценки априорного распределения на случай параметрического пространтсва вида $\Theta = \mathbf{R}^1$. Один из возможных путей состоит в следующем.

Предположим ради определённости, что θ является параметром сдвига нормального распределения с единичной дисперсией, то есть наблюдения $(X_1, \cdots, X_n)$ имеют общую плотность вида

$$p_Q(x) = \int_{-\infty}^{+\infty} \varphi(x-\theta) dQ(\theta), \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{21.1.21}$$

относительно меры Лебега на прямой $\mathcal{X} = \mathbf{R}^1$.

Для каждого $n \geq 1$ пусть

$$\theta_1^{(n)} < \cdots < \theta_{k_n}^{(n)} \tag{21.1.22}$$

являются константами и пусть $q_{i,n}, \quad i = 1, \cdots, k_n$ определяются соотношениями (21.1.14) с функциями $p_j(x)$ из (21.1.5), заменёнными на $\varphi(x - \theta_j^{(n)})$. Рассмотрим случайную функцию распределения

$$Q_n(\theta) = \sum q_{i,n}, \tag{21.1.23}$$

где суммирование распространяется на все i такие, что

$$\theta_i^{(n)} < \theta.$$

Можем ли мы выбрать величины k_n и (21.1.22) для каждого n так, чтобы при всех $Q(\cdot)$ выполнялось равенство

$$\mathsf{P}\Big(\operatorname*{Lim}_{n\to\infty} Q_n(\theta) \to Q(\theta), \quad \text{в каждой точке непрерывности} \quad Q(\theta)\Big) = 1?$$

21.2 СЛУЧАЙ ОТСУТСТВИЯ АСИМПТОТИЧЕСКИ ОПТИМАЛЬНОЙ РЕШАЮЩЕЙ ФУНКЦИИ

Приведём пример, когда не существует асимптотически оптимальной последовательности решающих функций, но эмпирический байесовский подход приводит к разумным результатам.

Пусть X – случайная величина, принимающая только два значения: ноль с вероятностью $1-\theta$ и единицу с вероятностью θ, где неизвестный параметр θ принадлежит множеству $\Theta = [0,1]$.

На основе одного наблюдения $X = x$ мы хотим оценить θ. Для оценки $\delta \in \Delta = \Theta = [0,1]$, пусть функция потерь имеет вид

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

Решающая функция $\delta(x)$ определяется двумя постоянными $\delta(0), \delta(1)$, которые принадлежат единичному интервалу $\Delta = [0,1]$. Условные средние потери, при использовании решающей функции $\delta(x)$ при данном θ, имеют вид (см. (18.1.1))

$$R(\theta,\delta) = (1-\theta)(\theta-\delta(0))^2 + \theta(\theta-\delta(1))^2 = \tag{21.2.1}$$

$$= \delta^2(0) + [\delta^2(1) - 2\delta(0) - \delta^2(0)]\theta + [1 - 2\delta(1) + 2\delta(0)]\theta^2.$$

Рассмотрим класс решающих функций $\delta_\alpha(x)$, зависящих от параметра $\alpha \in (0,1)$ и определённых как

$$\delta_\alpha(0) = \frac{\alpha}{2}, \qquad \delta_\alpha(1) = \frac{1+\alpha}{2}. \tag{21.2.2}$$

Из равенства (21.2.1) непосредственно следует, что

$$R(\theta,\delta_\alpha) = \frac{\alpha^2 + (1-2\alpha)\theta}{4}. \tag{21.2.3}$$

Обозначим $\delta_\alpha(x)$ при $\alpha = 1/2$ через $\delta^*(x)$, то есть

$$\delta^*(0) = \frac{1}{4}, \qquad \delta^*(1) = \frac{3}{4}, \qquad R(\theta,\delta^*) = \frac{1}{16}, \quad \text{для всех} \quad \theta. \tag{21.2.4}$$

Для *любого априорнорго распределения* $Q(\cdot)$ случайной величины Ξ, положим

$$\alpha_i = \int_0^1 \theta^i dQ(\theta), \qquad i = 1,2. \tag{21.2.5}$$

Тогда из (21.2.1) следует, что для любой решающей функции $\delta(x)$ справедливо равенство

$$r(\delta,Q) = \int_0^1 R(\theta,\delta)dQ(\theta) = \tag{21.2.6}$$

$$= \delta^*(0) + \alpha_1[\delta^2(1) - 2\delta(0) - \delta^2(0)] + \alpha_2[1 - 2\delta(1) + 2\delta(0)].$$

Исключая тривиальные случаи, когда $\alpha_1 = 0$ или $\alpha_1 = 1$, после некоторых вычислений, отсюда следует, что

$$r(\delta, Q) = \frac{(\alpha_1 - \alpha_2)(\alpha_2 - \alpha_1^2)}{\alpha_1(1 - \alpha_1)} + \tag{21.2.7}$$

$$+(1 - \alpha_1)\left[\delta(0) - \frac{\alpha_1 - \alpha_2}{1 - \alpha_1}\right]^2 + \alpha_1\left[\delta(1) - \frac{\alpha_2}{\alpha_1}\right]^2,$$

поэтому при данном априорном распределении $Q(\cdot)$ байесовский риск $r(\delta, Q)$ достигает единственного минимума на байесовской решающей функции $\delta_Q(x)$ вида

$$\delta_Q(0) = \frac{\alpha_1 - \alpha_2}{1 - \alpha_1}, \qquad \delta_Q(1) = \frac{\alpha_2}{\alpha_1}, \tag{21.2.8}$$

причём

$$r(Q) = r(\delta_Q, Q) = \frac{(\alpha_1 - \alpha_2)(\alpha_2 - \alpha_1^2)}{\alpha_1(1 - \alpha_1)}. \tag{21.2.9}$$

Каждая решающая функция $\delta(x)$ (в частности и $\delta^*(x)$) является байесовской решающей функцией, относительно априорного распределения $Q(\cdot)$ такого, что

$$\frac{\alpha_1 - \alpha_2}{1 - \alpha_1} = \frac{\alpha}{2}, \qquad \frac{\alpha_2}{\alpha_1} = \frac{1 + \alpha}{\alpha}. \tag{21.2.10}$$

Таким априорным распределением Q_α, например, является бета – распределение с плотностью

$$[B(\alpha, 1 - \alpha)]^{-1}\theta^{\alpha - 1}(1 - \theta)^{(1 - \alpha) - 1}, \tag{21.2.11}$$

для которого

$$\alpha_1 = \alpha, \quad \alpha_2 = \frac{\alpha(1 + \alpha)}{2}, \quad r(Q_\alpha) = \frac{\alpha(1 - \alpha)}{4}. \tag{21.2.12}$$

Тот факт, что $\delta^*(x)$ является байесовской решающей функцией при априорном распределении $Q_{1/2}$ и то, что для любого априорного распределения $Q(\cdot)$ справедливо тождество

$$r(\delta^*, Q) = \int_0^1 R(\theta, \delta^*) dQ(\theta) = \frac{1}{16} \tag{21.2.13}$$

имеет важное следствие вида

$$\sup_Q r(\delta, Q) > \frac{1}{16} \quad \text{для каждой решающей функции} \quad \delta(x) \neq \delta^*(x). \quad (21.2.14)$$

Таким образом, если для некоторой решающей функции $\delta'(x)$ справедливо неравенство

$$\sup_Q r(\delta', Q) \ \leq \ \frac{1}{16},$$

то, в частности,

$$\frac{1}{16} = r(\delta^*, Q_{1/2}) \ \leq \ r(\delta', Q_{1/2}) \ \leq \ \frac{1}{16} \qquad (21.2.15)$$

и значит

$$r(\delta^*, Q_{1/2}) = r(\delta', Q_{1/2}) = \frac{1}{16}, \qquad (21.2.16)$$

поэтому $\delta'(x) = \delta^*(x)$. Отсюда следует, что решающая функция $\delta^*(x)$ является единственной *минимаксной решающей функцией* в том смысле, что она минимизирует максимум, по всем априорным распределениям $Q(\cdot)$, байесовского риска. Когда ничего не известно об априорном распределении $Q(\cdot)$, то разумно использовать минимаксную решающую функцию $\delta^*(x)$, при этом байесовский риск всегда будет равняться $1/16$, независимо от априорного распределения $Q(\cdot)$. Если некоторая решающая функция $\delta(x) \neq \delta^*(x)$, то байесовский риск будет $> 1/16$ для некоторого априорного распределения $Q(\cdot)$ (в частности, для любого априорного распределения $Q(\cdot)$ с $\alpha_1 = 1/2$ и $\alpha_2 = 3/8$, например, для $Q_{1/2}(\cdot)$).

Для любого $0 < \alpha < 1$ обозначим через G_α класс всех априорных распределений $Q(\cdot)$ таких, что $\alpha_1(Q) = \alpha$. Для любого априорного распределения $Q(\cdot)$ из G_α (в частности для $Q_\alpha(\cdot)$) из соотношений (21.2.2) и (21.2.7), после некоторых преобразований, следует, что

$$r(\delta_\alpha, Q) = \frac{\alpha(1-\alpha)}{2}, \qquad \text{для любого} \quad Q \in G_\alpha, \qquad (21.2.17)$$

независимо от значения $\alpha_2(Q)$. Поэтому так же, как и выше

$$\sup_{Q \in G_\alpha} r(\delta, Q) > \frac{\alpha(1-\alpha)}{4} \qquad \text{для каждой решающей функции} \quad \delta(x) \neq \delta_\alpha(x), \qquad (21.2.18)$$

поэтому относительно класса G_α решающая функция $\delta_\alpha(x)$ является единственной минимаксной решающей функцией в том смысле, что она *минимизирует максимум байесовского риска, взятый по классу G_α*. Если ничего не известно об априорном распределении $Q(\cdot)$ кроме того, что $\alpha_1(Q) = \alpha$, то разумно использовать решающую функцию $\delta_\alpha(x)$, поскольку для неё байесовский риск будет равняться

$$\frac{\alpha(1-\alpha)}{4},$$

в то время, как для любой другой решающей функции байесовский риск будет $> \frac{\alpha(1-\alpha)}{4}$ для некоторого априорного распределения $Q(\cdot)$ (в частности, для любого априорного распределения $Q(\cdot)$ с

$$\alpha_1(Q) = \alpha, \quad \alpha_2(Q) = \frac{\alpha(1+\alpha)}{2},$$

то есть, например, для $Q_\alpha(\cdot)$).

Из предыдущего следует, что (впрочем, это может быть проверено и непосредственно)

$$\frac{(\alpha_1-\alpha_2)(\alpha_2-\alpha_1^2)}{\alpha_1(1-\alpha_1)} \le \frac{\alpha_1(1-\alpha_1)}{4} \le \frac{1}{16}. \tag{21.2.19}$$

Причём равенства достигается только, если соответственно

$$\alpha_2 = \frac{\alpha_1(1+\alpha_1)}{2} \quad \text{и} \quad \alpha_1 = \frac{1}{2}.$$

Предположим теперь, что априорное распределение $Q(\cdot)$ нам не известно, мы хотим "оценить" неизвестное значение θ и имеем независимые наблюдения $(X_1, \cdots, X_n)$ причём

$$\mathsf{P}(X_1 = 1) = \int_0^1 \theta dQ(\theta) = \alpha_1(Q),$$

$$\mathsf{P}(X_1 = 0) = \int_0^1 (1-\theta) dQ(\theta) = 1 - \alpha_1(Q). \tag{21.2.20}$$

Таким образом распределение X_i зависит только от $\alpha_1(Q)$. Поскольку байесовская решающая функция $\delta_Q(x)$, определённая соотношением (21.2.8), зависит так же и от $\alpha_2(Q)$, то отсюда следует, что *не существует* асимптотически отимальной, относительно любого априорного распределения $Q(\cdot)$ из класса G, решающей функции $\delta_n(x)$, если только α_2 не является функцией

α_1 в классе G. В практических приложениях α_2 редко является функцией α_1, поэтому типичным образом асимптотически оптимальной решающей функции не существует.

С другой стороны, пусть

$$u_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \tag{21.2.21}$$

и рассмотрим последовательность решающих функций $\tilde{\delta}(x) = \{\delta_n(x)\}$ вида

$$\delta_n(0) = \frac{u_n}{2}, \qquad \delta_n(1) = \frac{1+u_n}{2} \tag{21.2.22}$$

В силу Закона Больших Чисел, для любого априорного распределения $Q(\cdot)$ из класса G_α с вероятностью единица при $n \to \infty$

$$u_n \to \alpha$$

и значит

$$\delta_n(x) \to \delta_\alpha(x).$$

В действительности, поскольку

$$\mathsf{E}X_i = \alpha = \mathsf{E}X_i^2, \qquad \mathsf{D}X_i = \alpha(1-\alpha), \tag{21.2.23}$$

то

$$\mathsf{E}u_n = \alpha, \quad \mathsf{E}u_n^2 = \mathsf{D}u_n + \alpha^2 = \frac{\alpha(1-\alpha)}{n} + \alpha^2 \tag{21.2.24}$$

и значит из соотношения (21.2.6) следует, что

$$r_n(\tilde{\delta}, Q) = \mathsf{E}\left[\frac{u_n^2}{4} + \alpha\left\{\frac{1+2u_n+u_n^2}{4} - u_n - \frac{u_n^2}{4}\right\} + \alpha_2(1-1-u_n+u_n)\right] = \tag{21.2.25}$$

$$= \frac{1}{4}\mathsf{E}[u_n^2 - 2u_n + \alpha] = \frac{\alpha(1-\alpha)(n+1)}{4n} = r(\delta_\alpha, Q)\frac{n+1}{n}.$$

Таким образом при больших n используя решающую функцию $\tilde{\delta}$ мы бы получили почти такой же риск, как если бы знали $\alpha_1(Q) = \alpha$ и использовали решающую функцию $\delta_\alpha(x)$. Более того, для любого априорного распределения $Q \in G_\alpha$

$$r_n(\tilde{\delta}, Q) - r(\delta_\alpha, Q) = \frac{\alpha(1-\alpha)}{4n} \le \frac{1}{16n}, \tag{21.2.26}$$

в то время, как

$$r_n(\tilde{\delta}, Q) - r(\delta^*, Q) = \frac{\alpha(1-\alpha)(n+1)}{4n} - \frac{1}{16} = \tag{21.2.27}$$

$$= -\frac{(1-2\alpha)^2}{16} + \frac{\alpha(1-\alpha)}{4n}.$$

Этот пример иллюстрирует ситуацию, когда асимптотически оптимальная решающая функция δ не существует, либо существует, но риск $r(\delta, Q)$ слишком медленно сходится к $r(Q)$ и стоит использовать по крайней мере асимптолтически "субминимаксную" решающую функцию.

21.3 СПИСОК ЛИТЕРАТУРЫ

1) H. Robbins, The empirical Bayes approach to statistical decision problems, Ann. Math. Statist., 1964, v.35, p. 1– 20.

2) J.S. Maritz, Empirical Bayes Methods, Methuen and Co LTD, London, 1970, Chapter 2.

Лекция 22

Лекция содержит задачи, дополняющие теорию, изложенную в предыдущих Лекциях.

22.1 ЗАДАЧИ

1) Пусть $\mathcal{X} = \{x_1, x_2, \cdots\}$, $\nu(\cdot)$ – считающая мера на $\mathcal{X}$ а $f(x)$ интегрируемая функция. Доказать, что
$$\int\limits_{\mathcal{X}} f(x) d\nu(x) = \sum_{i \geq 1} f(x_i).$$

2) Пусть $(\mathcal{X},\ \mathcal{F}, \nu)$ есть пространство с мерой, и пусть $\mathcal{A}$ – класс всех множеств вида $F \cup C$, где $F \in \mathcal{F}$ и C есть подмножество множества $A \in \mathcal{F}$ с $\nu(A) = 0$. Доказать, что $\mathcal{A}$ есть σ – алгебра.

3) Пусть λ и μ – σ – конечные меры на измеримом пространстве $(\mathcal{X},\ \mathcal{F})$ и мера μ абсолютно непрерывна относительно λ, тогда
$$\int\limits_{\mathcal{X}} f(x) d\mu(x) = \int\limits_{\mathcal{X}} f(x) \frac{d\mu}{d\lambda}(x) d\lambda(x)$$
для любой μ – интегрируемой функции $f(x)$.

4) Пусть λ, μ и ν – σ – конечные меры на измеримом пространстве $(\mathcal{X}, \mathcal{F})$ такие, что мера λ абсолютно непрерывна относительно μ, а мера μ абсолютно непрерывна относительно ν, тогда
$$\frac{d\lambda}{d\nu}(x) = \frac{d\lambda}{d\mu}(x) \frac{d\mu}{d\nu}(x), \qquad \nu \text{ – почти всюду.}$$

5) Пусть λ и μ – σ – конечные меры на измеримом пространстве $(\mathcal{X},\ \mathcal{F})$, которые эквивалентны в том смысле, что каждая абсолютно непрерывна относительно другой, тогда

$$\frac{d\lambda}{d\mu}(x) = \left(\frac{d\mu}{d\lambda}(x)\right)^{-1}, \quad \lambda,\ \ \mu - \text{почти всюду.}$$

6) Пусть $\mu_k, k = 1, 2, \cdots$ и μ – σ – конечные меры на измеримом пространстве $(\mathcal{X},\ \mathcal{F})$ такие, что

$$\sum_{k=1}^{\infty} \mu_k(F) = \mu(F), \quad F \in \mathcal{F}.$$

Тогда, если μ_k абсолютно непрерывны относительно σ – конечной меры ν, то и μ абсолютно непрерывна относительно ν и

$$\frac{d\sum\limits_{i=1}^{n}\mu_i}{d\nu}(x) = \sum_{i=1}^{n}\frac{d\mu_i}{d\nu}(x), \quad \operatorname*{Lim}_{n\to\infty}\frac{d\sum\limits_{i=1}^{n}\mu_i}{d\nu}(x) = \frac{d\mu}{d\nu}(x), \quad \nu - \text{почти всюду.}$$

7) Напомним, что медианой случайной величины X называется любое значение mX такое, что

$$\mathsf{P}(X\ \le\ mX)\ \ge\ \frac{1}{2} \quad \text{и} \quad \mathsf{P}(X\ \ge\ mX)\ \ge\ \frac{1}{2}.$$

Доказать, что это определение равносильно следующему

$$\mathsf{P}(X < mX)\ \le\ \frac{1}{2} \quad \text{и} \quad \mathsf{P}(X > mX)\ \le\ \frac{1}{2}.$$

Доказать также, что множество медиан всегда есть замкнутый интервал $m_0\ \le\ mX\ \le\ m_1$.

8) Пусть

$$h(a) = \mathsf{E}|X - a| < \infty$$

при некотором $a \in \mathbf{R}^1$. Доказать, что $h(a)$ минимизируется на любой медиане случайной величины X.

9) Для любого набора различных вещественных чисел $x_1, \cdots, x_n$ медиана определяется как среднее из упорядоченных значений x – ов, когда n

нечётное, и как любое значение между двумя центральными среди упорядоченных значений x – ов, когда n чётное. Показать, что это есть также медиана случайной величины X, принимающей каждое из значений $x_1, \cdots, x_n$ с вероятностью $1/n$.

Для любого набора различных вещественных чисел $x_1, \cdots, x_n$ сумма абсолютных уклонений

$$\sum_{i=1}^{n} |x_i - a|$$

минимизируется любой медианой x – ов.

10) Рассмотрим статистическую структуру $(\mathcal{X},\ \mathcal{F},\ \{\mathsf{P}_\theta,\ \theta \in \Theta\})$, в которой $\mathcal{X} = \{-1, 0, 1, 2, \cdots\}$, $\mathcal{F} = \gamma(\mathcal{X})$ – множество всех подмножеств множества $\mathcal{X}$, $\Theta = (0, 1)$ и семейство распределений $\{\mathsf{P}_\theta,\ \theta \in \Theta\}$ имеет вид

$$\mathsf{P}_\theta(-1) = \theta,\ \mathsf{P}_\theta(k) = (1-\theta)^2\theta^k, \quad k = 0, 1, 2, \cdots, \quad \theta \in \Theta.$$

Доказать, что достаточная статистика $T(X) = X$ не полна.

11) Пусть $X = (X_1, \cdots, X_n)$, где наблюдения X_i имеют симметричное распределение, то есть для любого борелевского множества $B \in \mathcal{B}^n$ справедливо равенство

$$\mathsf{P}((X_1, \cdots, X_n) \in B) = \mathsf{P}((X_{i_1}, \cdots, X_{i_n}) \in B)$$

для любой перестановки $(i_1, \cdots, i_n)$ чисел $(1, \cdots, n)$. В частности, $(X_1, \cdots, X_n)$ могут быть независимыми и одинаково распределёнными. Пусть $h(x_1, \cdots, x_n)$ – любая измеримая и интегрируемая функция. Доказать, что

$$\mathsf{E}(h(X_1, \cdots, X_n) \mid T(X)) = \frac{1}{n!} \sum_{(i_1, \cdots, i_n)} h(X_{i_1}, \cdots, X_{i_n}),$$

где $T(X) = (X_{(1)}, \cdots, X_{(n)})$ и $X_{(1)} \le \cdots \le X_{(n)}$ – вариационный ряд. Эдесь суммирование производится по всем перестановкам $(i_1, \cdots, i_n)$ чисел $(1, \cdots, n)$. Этот результат означает, что статистика $T(X)$ является достаточной для случая, если неизвестным параметром θ является неизвестное симметричное распределение.

12) Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково равномерно распределённые наблюдения

$$X_i \sim \mathcal{R}(0, \theta), \qquad i = 1, 2, \cdots; \quad \theta > 0.$$

Доказать, что

$$F(x \mid t) = \mathsf{P}_\theta(X_1 < x \mid X_{(n)} = t) = \begin{cases} 0, & X < 0, \\ \frac{x}{t}\,\frac{n-1}{n}, & 0 \le x \le t, \\ 1, & x > t. \end{cases}$$

Вывести отсюда, что оптимальная оценка для параметра θ есть

$$\delta^*(X_{(n)}) = 2\mathsf{E}_\theta(X_1 \mid X_{(n)}) = \frac{n+1}{n}\, X_{(n)}$$

и

$$\mathsf{D}_\theta \delta^*(X_{(n)}) = \frac{\theta^2}{n(n+2)} < \mathsf{D}_\theta(2\overline{X}) = \frac{\theta^2}{3n}, \quad n > 1.$$

13) Пусть статистика $T(X)$ на статистической структуре $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ достаточна для семейства $\mathcal{P}$. Если она является минимальной достаточной статистикой для подсемества $\mathcal{P}' \subseteq \mathcal{P}$ и всякое $\mathcal{P}'$ – нулевое множество является и $\mathcal{P}$ – нулевым, то статистика $T(X)$ является минимальной достаточной статистикой и для семейства $\mathcal{P}$.

14) Если наблюдение X имеет биномиальное распределение

$$X \sim \mathcal{B}(n,\, \theta), \quad \theta \in (0,1),$$

то для риска справедливо равенство ($L(\theta, \delta) = |\theta - \delta|$)

$$\mathsf{E}_\theta |X/n - \theta| = 2\binom{n-1}{k-1}\theta^k (1-\theta)^{n-k+1} \quad \text{при} \quad \frac{k-1}{n} \le \theta \le \frac{k}{n}.$$

15) Пусть $X = (X_1, \cdots, X_n)$ – некоррелированные наблюдения с общим математическим ожиданием θ и дисперсией σ^2. Доказать, что среди всех линейных оценок для θ вида $\sum\limits_{i=1}^{n} \alpha_i X_i$, удовлетворяющих соотношению $\sum\limits_{i=1}^{n} \alpha_i = 1$, оценка $\overline{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$ имеет наименьшую дисперсию.

16) Доказать, что *единственная* байесовская оценка является допустимой.

17) Доказать, что *единственная* минимаксная оценка является допустимой.

18) Пусть
$$\min_{\delta} r(\delta, Q) < \infty.$$
Тогда, если
$$\Theta = \{\theta_1, \theta_2, \cdots\} \quad \text{и} \quad Q(\Xi = \theta_i) > 0, \quad i = 1, 2, \cdots$$
или $\Theta \subseteq \mathbf{R}^k$ и функция риска $R(\theta, \delta)$ непрерывна по $\theta \in \Theta$ для любой оценки $\delta(X)$ и априорное распределение Q имеет строго положительную плотность, то байесовская оценка $\delta_Q(X)$ допустима.

19) Пусть
$$X \sim \mathcal{B}(\theta, 1), \qquad \Theta = [1/3, 2/3].$$
Доказать, что оценка
$$\delta_*(X) = 4/9\, \mathbf{1}_{\{0\}}(X) + 5/9\, \mathbf{1}_{\{1\}}(X)$$
является минимаксной.

20) Пусть $X = (X_1, X_2)$, где $X_i,\ i = 1, 2$ – независимые одинаково распределённые наблюдения, имеющие плотность
$$p_\theta(x) = \frac{3x^2}{\theta^3}\, \mathbf{1}_{(0,\theta)}(x)$$
и
$$\Theta = \Delta = (0, \infty), \qquad L(\theta, \delta) = (\theta - \delta)^2.$$
Доказать, что оценки
$$\delta_1(X) = 2/3\,(X_1 + X_2), \quad \delta_2(X) = 7/6\, \max(X_1, X_2)$$
являются несмещёнными оценками параметра θ. Найти и сравнить риски этих оценок.

21) Пусть $\delta_Q(X)$ есть байесовская (соответственно оптимальная, минимаксная, допустимая) оценка для параметрической функции $g(\theta)$ при квадратичной функции потерь. Тогда оценка $a\delta_Q(X) + b,\quad a, b \in \mathbf{R}^1$ является байесовской (соответственно оптимальной, минимаксной, допустимой) для функции $ag(\theta) + b$.

22) Пусть $\delta(X)$ – оценка для параметра $\theta \in \Theta$ при квадратичной функции потерь. Тогда оценка $a\delta_Q(X) + b,\quad a, b \in \mathbf{R}^1$ является *недопустимой* оценкой для θ, если только $a > 1$ или $a < 0$ или $a = 1,\ b \neq 0$.

23) Если оценка имеет постоянный риск и допустима, то она минимаксна.

24) Пусть $\delta_*(X)$ есть минимаксная оценка для $g(\theta)$, когда $\theta \in \bar{\Theta} \subseteq \Theta$. Тогда, если

$$\sup_{\theta \in \bar{\Theta}} R(\theta, \delta_*) = \sup_{\theta \in \Theta} R(\theta, \delta_*),$$

то $\delta_*(X)$ минимаксна также для $g(\theta)$ и тогда, когда $\theta \in \Theta$.

25) Пусть существует последовательность априорных распределений $\{Q_n\}$ на Θ и оценка $\delta(X)$ такие, что

$$\sup_{\theta \in \Theta} R(\theta, \delta) \ \leq \ \limsup_{n \to \infty} \int\limits_{\Theta} R(\theta, \delta_{Q_n}) dQ_n(\theta).$$

Тогда $\delta(X)$ – минимаксная оценка.

26) Пусть наблюдение X имеет биномиальное распределение

$$X \sim \mathcal{B}(n, \theta), \quad \theta \in \Theta = (0, 1)$$

с неизвестным параметром θ и пусть функция потерь имеет вид

$$L(\theta, \delta) = \frac{(\delta - \theta)^2}{\theta(1 - \theta)}.$$

Рассмотрим равномерное априорное распределение Q на Θ, то есть пусть

$$\Xi \sim \mathcal{R}(0, 1).$$

Доказать, что единственная байесовская оценка $\delta_Q(X)$ для параметра θ есть

$$\delta_Q(X) = \frac{X}{n}$$

и её байесовский риск постоянен и равен $1/n$.

27) Пусть $X = (X_1, \cdots, X_n)$, где $X_i, \ i = 1, \cdots, n$ независимые одинаково нормально распредлённые наблюдения

$$X_i \sim \mathcal{N}(\theta, 2\theta), \quad i = 1, \cdots, n; \quad \theta > 0.$$

Доказать, что оценка максимального правдоподобия $\hat{\delta}(X)$ для параметра θ есть

$$\hat{\delta}(X) = \sqrt{1 + \frac{1}{n} \sum_{i=1}^{n} X_i^2} - 1.$$

Доказать состоятельность этой оценки.

28) Пусть $C(X)$ – доверительный интервал для параметра θ и $|C(X)|$ – его длина. Доказать, что
$$\mathsf{E}_{\theta_0}|C(X)| = \int\limits_{\Theta} \mathsf{P}_{\theta_0}(\theta \in C(X))\, d\theta.$$

29) Распределение наблюдения X вида
$$p_\theta(x) = \mathsf{P}_\theta(X = x) = \frac{a(x)\theta^x}{C(\theta)}, \quad x = 0, 1, 2, \cdots; \ \ a(x) \ \ge \ 0, \ \ \theta > 0,$$
называется *распределением степенного ряда*. Доказать, что биномиальное, отрицательно биномиальное и распределние Пуассона есть распределения степенного ряда.

30) Доказать, что распределение степенного ряда принадлежит экспоненциальному семейству и для производящей функции моментов справедливо равенство
$$M_\theta(s) = \mathsf{E}_\theta e^{sX} = \frac{C(\theta e^s)}{C(\theta)}.$$

31) Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения, причём распределние $X_i, \quad i = 1, \cdots, n$ является распределением степенного ряда. Доказать, что распределение статистики $T(X) = \sum\limits_{i=1}^{n} X_i$ также является распределением степенного ряда и
$$p_\theta(t) = \mathsf{P}_\theta(T(X) = t) = \frac{A(t,n)\theta^x}{C^n(\theta)}, \quad t = 0, 1, 2, \cdots,$$
где $A(t, n)$ – коэффициенты при θ^t в разложении $C^n(\theta)$ в степенной ряд. Доказать также, что $T(X)$ – полная достаточная статистика.

32) В условиях предыдущей задачи доказать, что оптимальные оценки для параметрических функций
$$g(\theta) = \theta^r, \ \ r \in \mathbf{N}; \quad g(\theta) = \mathsf{P}_\theta(X_1 = x)$$
соответственно имеют вид
$$\delta^*(T) = \begin{cases} 0, & T = 0, 1, \cdots, r-1, \\ \frac{A(T-r,n)}{A(T,n)}, & T \ \ge \ r, \end{cases}$$
$$\delta^*(T) = \frac{a(x)A(T-x, n-1)}{A(T,n)}.$$

33) Пусть наблюдение X имеет плотность (относительно меры ν) $p_\theta(x)$, которая *положительна* при всех x и пусть Q_1 и Q_2 – два распределения на действительной прямой с конечными первыми моментами. Доказать, что для дисперсии любой несмещённой оценки $\delta(X)$ параметра θ справедливо неравенство

$$\mathsf{D}_\theta \delta(X) \;\geq\; \frac{\left(\int x\, dQ_1(x) - \int x\, dQ_2(x)\right)^2}{\int \psi^2(x,\theta) p_\theta(x)\, d\nu(x)}, \quad \theta \in \Theta,$$

где

$$\psi(x,\theta) = \frac{1}{p_\theta(x)} \int\limits_{A_\theta} p_{\theta+y}(x)\, (dQ_1(y) - dQ_2(y)), \quad A_\theta = \{y :\, \theta + y \in \Theta\}.$$

34) Пусть $T_n(\mathbf{X}_n)$, $U_n(\mathbf{X}_n)$, $V_n(\mathbf{X}_n)$, $n = 1, 2, \cdots$ – последовательности статистик такие, что

$$U_n(\mathbf{X}_n) \xrightarrow{P_{n\theta}} 1, \quad V_n(\mathbf{X}_n) \xrightarrow{P_{n\theta}} 0, \quad \theta \in \Theta, \ \ n \to \infty$$

и

$$\mathsf{P}_\theta(T_n(\mathbf{X}_n) < x) \to G_\theta(x) \quad n \to \infty, \quad \theta \in \Theta$$

в каждой точке непрерывности x функции распределения $G_\theta(x)$. Доказать, что и

$$\mathsf{P}_\theta(T_n(\mathbf{X}_n) U_n(\mathbf{X}_n) + V_n(\mathbf{X}_n) < x) \to G_\theta(x) \quad n \to \infty, \quad \theta \in \Theta$$

в каждой точке непрерывности x функции распределения $G_\theta(x)$.

35) Пусть $\mathcal{D}$ класс всех оценок параметра $\theta \in \Theta$ при квадратичной функции потерь при условии, что выполнены условия регулярности Краме́ра – Рао (см. Теорему 11.1.4 и её Следствия). Предположим, что оценка $\delta_0(X) \in \mathcal{D}$ является эффективной оценкой, то есть

$$\mathsf{E}_\theta(\delta_0(X) - \theta)^2 = C_{\delta_0}(\theta), \quad \theta \in \Theta,$$

где

$$C_\delta(\theta) = b_\delta^2(\theta) + \frac{(1 + b'_\delta(\theta))^2}{I(\theta)}, \quad b_\delta(\theta) = \mathsf{E}_\theta \delta(X) - \theta.$$

Доказать, что если для любой оценки $\delta(X) \in \mathcal{D}$ из соотношения

$$C_\delta(\theta) \le C_{\delta_0}(\theta), \quad \theta \in \Theta$$

следует, что

$$b_\delta(\theta) \equiv b_{\delta_0}(\theta), \quad \theta \in \Theta,$$

то $\delta_0(X)$ – допустимая оценка.

36) Пусть $X = (X_1, \cdots, X_n)$ – независимые одинаково распределённые наблюдения с

$$\mathsf{E}_\theta X_i = \theta, \quad \mathsf{D}_\theta X_i = 1, \quad i = 1, \cdots, n$$

и

$$\Theta = \Delta = \mathbf{R}^1, \quad L(\theta, \delta) = (\theta - \delta)^2.$$

Доказать, что при этих условиях оценка

$$\delta(X) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

является допустимой и минимаксной.

37) Доказать, что для случая, когда наблюдение X имеет распределение Пуассона, геометрическое распределение или отрицательно биномиальное распределение с неизвестным параметром $\theta \in \Theta$, байесовская оценка $\delta_Q(X)$, соответствующая априорному распределению Q, есть (удобный вид для эмпирического байесовского подхода, см. Лекция 18)

$$\delta_Q(X) = C(X)\frac{p_Q(X+1)}{p_Q(X)}, \quad \theta \in \Theta,$$

где $C(x)$ – известная константа и

$$p_Q(x) = \int\limits_{\Theta} p_\theta(x)\, dQ(\theta).$$

Обобщить этот результат на случай, если наблюдение X имеет дискретное распределение вида

$$p_\theta(x) = \mathsf{P}_\theta(X = x) = \exp\{A(\theta) + B(\theta)h(x) + q(x)\}.$$

22.2 СПИСОК ЛИТЕРАТУРЫ

1) Э. Леман, Теория Точечного Оценивания,
Москва, Наука, 1991.

2) Э. Леман, Проверка Статистических Гипотез,
Москва, Наука, 1979.

3) Д. Дюге, Теоретическая и Прикладная Статистика,
Москва, Наука, 1972.

4) А.А. Боровков, Математическая Статистика,
Москва, Наука, 1984.

5) Ж.– Р. Барра, Основные Понятия Математической Статистики,
Москва, Мир, 1972.

6) Г.П. Климов, Теория Вероятностей и Математическая Статистика,
Москва, Издательство МГУ, 1983.

7) Г.И. Ивченко, Ю.И. Медведев, Математическая Статистика,
Москва, Высшая Школа, 1992.

8) Ш. Закс, Теория Статистических Выводов,
Москва, Мир, 1975.

9) И.А. Ибрагимов, Р.З. Хасьминиский, Асимптотическая Теория оценивания,
Москва, Наука, 1979.

10) А.Н. Ширяев, Вероятность,
Москва, Наука, 1989.