

# Математическая статистика

Маттиас Феттер  
(пер. Александр Самарин)

2016



# Оглавление

1	Условное математическое ожидание	1
2	Основы точечного оценивания	11
3	Байесовское и минимаксное оценивания параметров	25
4	Достаточность и полнота	33
5	Асимптотические свойства оценок	47
6	Основы тестирования	57
7	Асимптотические свойства критериев	69
8	Линейная модель	77
	Список литературы	87



# Глава 1

## Условное математическое ожидание

В этой главе мы рассмотрим понятие условного математического ожидания, которое представляет собой важную основу не только для анализа статистических методов, но и в общей сложности для стохастической теории.

**Замечание 1.1.** Рассмотрим интегрируемую случайную величину:  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$ . После проведения эксперимента и наблюдения результата информация о случайной величине полностью известна:

$$\omega \mapsto X(\omega).$$

Перед проведением эксперимента неизвестно ничего о его итоге. В такой ситуации лучший способ предсказать его – это использование математического ожидания:

$$\omega \mapsto \mathbb{E}[X](\omega)$$

Условное математическое ожидание – лучшая аппроксимация случайной величины  $X$ , при условии что часть информации о ней известна.

**Определение 1.2.** Пусть  $(\Omega, \mathcal{A})$  – измеримое пространство с заданными мерами  $\mu$  и  $\nu$ . Мера  $\nu$  называется **абсолютно непрерывной** относительно  $\mu$  (или же  $\mu$  доминирует  $\nu$ ), если

$$\mu(A) = 0 \implies \nu(A) = 0 \quad \forall A \in \mathcal{A}.$$

Обозначение:  $\nu \ll \mu$ .

**Теорема 1.3 (Теорема Радона-Никодима).** Пусть  $(\Omega, \mathcal{A})$  – измеримое пространство с заданными мерами  $\mu$  и  $\nu$ . Если  $\mu$   $\sigma$ -конечная, то следующие утверждения эквивалентны:

(i)  $\nu \ll \mu$ .

(ii) Существует  $\mathcal{A}$ -измеримая неотрицательная функция  $f$ , такая что:

$$\nu(A) = \int_A f d\mu \quad \forall A \in \mathcal{A}.$$

Иначе говоря,  $\nu$  обладает плотностью  $f$  по  $\mu$ . В качестве обозначения плотности часто используется т.н. **производная Радона-Никодима**:

$$f = \frac{d\nu}{d\mu}.$$

ДОКАЗАТЕЛЬСТВО: Теорема 17.10 в [1].

**Определение 1.4.** Пусть  $\mathcal{F} \subset \mathcal{A}$  –  $\sigma$ -алгебра. Случайная величина  $Y : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B})$  называется **условным математическим ожиданием**  $X$  относительно  $\mathcal{F}$ , если:

- (i)  $Y$   $\mathcal{F}$ - $\mathcal{B}$ -измеримая<sup>1</sup>,
- (ii)  $\mathbb{E}[X1_A] = \mathbb{E}[Y1_A] \quad \forall A \in \mathcal{F}$ .

**Теорема 1.5.** *Условное математическое ожидание  $X$  относительно  $\mathcal{F}$  существует и единственно почти всюду.*

**ДОКАЗАТЕЛЬСТВО:** *Единственность.* Пусть  $Y$  и  $\bar{Y}$  – два условных математических ожидания. Зададим множество  $A = \{\omega \mid Y(\omega) > \bar{Y}(\omega)\}$ . Тогда согласно Определению 1.4 (ii):

$$\mathbb{E}[(Y - \bar{Y})1_A] = \mathbb{E}[Y1_A] - \mathbb{E}[\bar{Y}1_A] = 0.$$

Так как  $(Y - \bar{Y})1_A \geq 0$ , то  $\mathbb{P}(A) = 0$ . Таким же образом доказывается, что  $\mathbb{P}(Y < \bar{Y}) = 0$ .

*Существование.* Представим  $X = X^+ - X^-$  в виде разницы положительной и отрицательной случайных величин. Зададим две меры на  $(\Omega, \mathcal{F})$ :

$$\mathbb{Q}^\pm(A) = \mathbb{E}[X^\pm 1_A], \quad A \in \mathcal{F}.$$

Они обе абсолютно непрерывны относительно меры  $\mathbb{P}$ . По Теореме 1.3 существуют такие  $\mathcal{F}$ -измеримые плотности  $Y^\pm$ , что

$$\mathbb{Q}^\pm(A) = \int_A Y^\pm d\mathbb{P} = \mathbb{E}[Y^\pm 1_A].$$

В таком случае,  $Y = Y^+ - Y^-$  – условное математическое ожидание.

**Замечание 1.6.** Для условного математического ожидания  $X$  относительно  $\mathcal{F}$  мы будем использовать обозначение  $Y = \mathbb{E}[X|\mathcal{F}]$ , что в соответствии с предыдущей теоремой должно пониматься как равенство  $\mathbb{P}$ -почти наверное. Если  $X = 1_C$  для  $C \in \mathcal{A}$ , то условная вероятность  $X$  при условии  $\mathcal{F}$ :

$$\mathbb{P}(C|\mathcal{F}) = \mathbb{E}[1_C|\mathcal{F}].$$

**Определение 1.7.** Пусть  $Y$  – случайная величина (не обязательно интегрируемая) в пространстве  $(\Omega, \mathcal{A}, \mathbb{P})$ . Тогда **условное ожидание  $X$  относительно  $Y$** :

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)],$$

где  $\sigma(Y)$  обозначает  $\sigma$ -алгебру, порожденную  $Y$ .

**Лемма 1.8.** Пусть  $(\Omega, \mathcal{A}, \mathbb{P})$  – вероятностное пространство,  $\mathcal{F} \subset \mathcal{A}$  –  $\sigma$ -алгебра и  $X, Y$  – интегрируемые случайные величины, принимающие действительные значения. Тогда выполняются следующие свойства:

- (i) **Линейность.** Для любых  $\alpha, \beta \in \mathbb{R}$ :

$$\mathbb{E}[\alpha X + \beta Y|\mathcal{F}] = \alpha \mathbb{E}[X|\mathcal{F}] + \beta \mathbb{E}[Y|\mathcal{F}].$$

- (ii) **Монотонность.** Если  $X \leq Y$   $\mathbb{P}$ -н.н., то  $\mathbb{E}[X|\mathcal{F}] \leq \mathbb{E}[Y|\mathcal{F}]$   $\mathbb{P}$ -н.н.

**ДОКАЗАТЕЛЬСТВО:**

- (i) Правая часть равенства  $\mathcal{F}$ -измерима. Следовательно, для любого множества  $A \in \mathcal{F}$ :

$$\mathbb{E}[(\alpha X + \beta Y)1_A] = \alpha \mathbb{E}[1_A X] + \beta \mathbb{E}[1_A Y] = \mathbb{E}[1_A(\alpha \mathbb{E}[X|\mathcal{F}] + \beta \mathbb{E}[Y|\mathcal{F}])].$$

- (ii) Зададим множество  $A = \{\mathbb{E}[X|\mathcal{F}] - \mathbb{E}[Y|\mathcal{F}] < 0\} \in \mathcal{F}$ . Тогда:

$$0 \geq \int_A (\mathbb{E}[X|\mathcal{F}] - \mathbb{E}[Y|\mathcal{F}]) d\mathbb{P} = \int_A \mathbb{E}[X - Y|\mathcal{F}] d\mathbb{P} = \int_A (X - Y) d\mathbb{P} \geq 0.$$

Следовательно,  $\mathbb{P}(A) = 0$ .

---

<sup>1</sup> $Y^{-1}(B) \in \mathcal{F} \quad \forall B \in \mathcal{B}$

**Пример 1.9.**

- (i) Если  $\mathcal{F} = \{\emptyset, \Omega\}$ , то  $\mathbb{E}[X|\mathcal{F}]$  по определению постоянная и с выбором  $A = \Omega$  в (ii) следует:

$$\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X].$$

Это соответствует ситуации, когда никакой дополнительной информации об итоге эксперимента не известно.

- (ii) Если  $\sigma(X) \subset \mathcal{F}$ , то  $\mathbb{E}[X|\mathcal{F}] = X$ .
- (iii) Пусть множество  $A \in \mathcal{A}$ ,  $\mathbb{P}(A) \in (0, 1)$  и  $\mathcal{F} = \sigma(A) = \{\Omega, A, A^c, \emptyset\}$ . Вследствие  $\mathcal{F}$ -измеримости условное ожидание  $\mathbb{E}[X|A]$  является постоянной:

$$\mathbb{E}[X|A] = \left( \mathbb{P}(A)^{-1} \int_A X d\mathbb{P} \right) 1_A + \left( \mathbb{P}(A^c)^{-1} \int_{A^c} X d\mathbb{P} \right) 1_{A^c}.$$

В случае, если семейство элементарных событий составляют непересекающиеся множества  $(\Omega = A_1 \cup \dots \cup A_n)$  с положительными вероятностями, то:

$$\mathbb{E}[X|A_1, \dots, A_n] = \sum_{i=1}^n \left( \mathbb{P}(A_i)^{-1} \int_{A_i} X d\mathbb{P} \right) 1_{A_i}.$$

**Теорема 1.10 (Монотонная сходимостъ).** Пусть  $X_n \geq 0$  – последовательность интегрируемых случайных величин, такая что  $X_n \nearrow X$ . Тогда существует  $\mathcal{F}$ -измеримая случайная величина  $Y$ , такая что  $\mathbb{E}[X_n|\mathcal{F}] \nearrow Y$ . В частности,

$$\mathbb{E}[X 1_A] = \mathbb{E}[Y 1_A] \quad \forall A \in \mathcal{F},$$

и это выполняется, если  $X$  интегрируемая в понимании обычного условного математического ожидания.

**ДОКАЗАТЕЛЬСТВО:** По свойству монотонности математического ожидания последовательность  $\mathbb{E}[X_n|\mathcal{F}]$  возрастает. Следовательно,  $Y$  – поточечный предел последовательности  $\mathcal{F}$ -измеримых функций, который также  $\mathcal{F}$ -измеримый. В дополнение, так как  $X_n \nearrow X$ , то  $X_n 1_A \nearrow X 1_A$ , и мы получаем классическую теорему о монотонной сходимости:

$$\mathbb{E}[X 1_A] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n 1_A] = \lim_{n \rightarrow \infty} \mathbb{E}[\mathbb{E}[X_n|\mathcal{F}] 1_A] = \mathbb{E}[Y 1_A].$$

**Замечание 1.11.**

- (i) Теорема 1.10 также показывает, что условное математическое ожидание существует для  $X$ , если задать последовательность  $X_n = \min(X, n)$ .
- (ii) Лемма 1.8 и Теорема 1.10 доказывают лишь некоторые свойства ожидаемых значений, которые могут быть обобщены до условных ожиданий. Последующие примеры включают неравенство Коши-Шварца, неравенство Йенсена и вариант теоремы о мажорируемой сходимости.

**Теорема 1.12.** Пусть  $X$  и  $Y$  – случайные величины, такие что  $\mathbb{E}[|X|] < \infty$ ,  $\mathbb{E}[|XY|] < \infty$  и  $Y$   $\mathcal{F}$ -измеримая, тогда

$$\mathbb{E}[XY|\mathcal{F}] = Y \mathbb{E}[X|\mathcal{F}].$$

Другими словами, мы вытаскиваем из-под условного математического ожидания все, что известно.

ДОКАЗАТЕЛЬСТВО: Случайная величина  $Y\mathbb{E}[X|\mathcal{F}]$   $\mathcal{F}$ -измерима. Остается доказать:

$$\mathbb{E}[XY1_A] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{F}]1_A] \quad \forall A \in \mathcal{F}.$$

Используем индукцию из теории меры:

(i) Пусть  $Y = 1_B$ ,  $B \in \mathcal{F}$ . Тогда

$$\mathbb{E}[XY1_A] = \mathbb{E}[X1_{A \cap B}] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]1_{A \cap B}] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{F}]1_A].$$

(ii) Обобщаем равенство до ступенчатых функций и используем свойство линейности (Лемма 1.8).

(iii) Для  $X \geq 0$  и  $Y \geq 0$  используем (ii) и монотонную сходимость (Теорема 1.10).

(iv) Раскладываем случайные величины  $X$  и  $Y$  на положительную и отрицательную части:  $X = X^+ - X^-$  и  $Y = Y^+ - Y^-$ .

**Теорема 1.13 (Башенное свойство).** Пусть  $X$  – интегрируемая случайная величина на  $(\Omega, \mathcal{A}, \mathbb{P})$  и  $\mathcal{F}_1, \mathcal{F}_2$  –  $\sigma$ -алгебры, такие что  $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{A}$ . Тогда

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}_2]|\mathcal{F}_1] = \mathbb{E}[X|\mathcal{F}_1] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_1]\mathcal{F}_2].$$

ДОКАЗАТЕЛЬСТВО: Первое равенство: для любого множества  $A \in \mathcal{F}_1$

$$\int_A \mathbb{E}[\mathbb{E}[X|\mathcal{F}_2]|\mathcal{F}_1] d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{F}_2] d\mathbb{P} = \int_A X d\mathbb{P} = \int_A \mathbb{E}[X|\mathcal{F}_1] d\mathbb{P}.$$

Второе равенство следует из Теоремы 1.12, так как случайная величина  $\mathbb{E}[X|\mathcal{F}_1]$   $\mathcal{F}_2$ -измерима.

**Теорема 1.14 (Независимость).** Пусть  $X$  – интегрируемая случайная величина и  $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$  –  $\sigma$ -алгебры, такие что  $\mathcal{G}$  и  $\sigma(\sigma(X), \mathcal{F})$  независимы. Тогда:

$$\mathbb{E}[X|\sigma(\mathcal{F}, \mathcal{G})] = \mathbb{E}[X|\mathcal{F}].$$

ДОКАЗАТЕЛЬСТВО: Условное ожидание  $\mathbb{E}[X|\mathcal{F}]$   $\sigma(\mathcal{F}, \mathcal{G})$ -измеримо. Остается показать, что

$$\mathbb{E}[X1_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]1_A] \quad \forall A \in \sigma(\mathcal{F}, \mathcal{G}).$$

Все множества, удовлетворяющие этому равенству составляют систему Динкина<sup>2</sup>. Нужно показать, что семейство этих множеств замкнуто относительно операции пересечения. Пусть  $F \in \mathcal{F}$  и  $D \in \mathcal{D}$ , тогда:

$$\mathbb{E}[X1_{F \cap G}] = \mathbb{E}[1_G(1_F X)] = \mathbb{E}[1_G]\mathbb{E}[1_F X] = \mathbb{E}[1_G]\mathbb{E}[\mathbb{E}[X|\mathcal{F}]1_F] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]1_{F \cap G}].$$

**Следствие 1.15.** Пусть  $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$  и  $\mathcal{H} \subset \mathcal{A}$  –  $\sigma$ -алгебра. Тогда:

(i)  $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]] = \mathbb{E}[X]$  (итерированное ожидание).

(ii) Если  $X$  не зависит от  $\mathcal{H}$ , то  $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$ .

ДОКАЗАТЕЛЬСТВО:

---

<sup>2</sup>Семейство  $\mathcal{D}$  подмножеств множества  $\Omega$  называется *системой Динкина* (или  *$\lambda$ -системой*), если оно удовлетворяет следующим свойствам:

- (i)  $\Omega \in \mathcal{D}$ ,
- (ii)  $A \in \mathcal{D} \implies A^c \in \mathcal{D}$ ,
- (iii) Если  $A_1, A_2, \dots \in \mathcal{D}$  – непересекающиеся множества, то  $\cup_{n=1}^{\infty} A_n \in \mathcal{D}$ .

Иными словами, система Динкина, замкнутая относительно конечного числа пересечений, образует  $\sigma$ -алгебру.



- (i) Теорема 1.13 для  $\mathcal{F}_1 = \{\Omega, \emptyset\}$ .  
(ii) Теорема 1.14 для  $\mathcal{F} = \{\Omega, \emptyset\}$  и  $\mathcal{G} = \mathcal{H}$ .

**Пример 1.16.**

- (i) Предположим, что  $X$  и  $Y$  дискретные случайные величины, т.е. существуют счетные подмножества  $I_X, I_Y \subset \mathbb{R}$ , такие что  $\mathbb{P}(X \in I_X) = \mathbb{P}(Y \in I_Y) = 1$ . Зададим для  $x \in I_X$  и  $y \in I_Y$  условную вероятность (при  $\mathbb{P}(Y = y) > 0$ ):

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} =: \frac{p_{xy}}{p_y}.$$

Тогда  $\mathbb{E}[X|Y] = g(Y)$  (если  $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ ), где

$$g(y) : \begin{cases} \sum_{x \in I_X} x \frac{p_{xy}}{p_y}, & p_y > 0, \\ 0, & p_y = 0. \end{cases}$$

- (ii) Предположим, что  $X$  и  $Y$  имеют плотность  $f_{X,Y}(x, y)$  относительно меры Лебега. Функции

$$f_X(x) = \int f_{X,Y}(x, y) dy \quad \text{и} \quad f_Y(y) = \int f_{X,Y}(x, y) dx$$

называются **маргинальными плотностями**. Если мы зададим

$$g(y) : \begin{cases} \int x \frac{f_{X,Y}(x, y)}{f_Y(y)} dx, & f_Y(y) > 0, \\ 0, & f_Y(y) = 0, \end{cases}$$

то  $\mathbb{E}[X|Y] = g(Y)$ .

ДОКАЗАТЕЛЬСТВО:

- (i) Функция  $g(Y) = \sigma(Y)$ -измеримая, так как она является измеримой по Борелю. Также пусть  $B \in \mathcal{B}$  и  $F = \{Y \in B\} \in \sigma(Y)$ . Тогда

$$\mathbb{E}[g(Y)1_F] = \sum_{y \in B \cap I_Y} \sum_{x \in I_X} x \frac{p_{xy}}{p_y} 1_{\{p_y > 0\}} p_y = \sum_{x \in I_X} \sum_{y \in B \cap I_Y} x p_{xy} = \mathbb{E}[X1_F].$$

- (ii) По теореме Фубини<sup>3</sup> имеет место измеримость  $y \mapsto \int x f(x, y) dx$  и  $y \mapsto \int f(x, y) dx$ . Следовательно,  $g$  измерима. В заключение, пусть  $F = \{Y \in B\} \in \sigma(Y)$ . Тогда,

$$\begin{aligned} \mathbb{E}[g(Y)1_F] &= \int_B g(y) f_Y(y) 1_{\{f_Y(y) > 0\}} dy \\ &= \int_B \frac{\int x f_{X,Y}(x, y) dx}{f_Y(y)} f_Y(y) 1_{\{f_Y(y) > 0\}} dy \\ \text{т. Фубини} \rightarrow &= \int x \int_B 1_{\{f_Y(y) > 0\}} f_{X,Y}(x, y) dy dx = \mathbb{E}[X1_F]. \end{aligned}$$

**Замечание 1.17.** В предыдущем примере мы увидели, что  $\mathbb{E}[X|Y]$  принимает форму функции  $g(Y)$ , где  $g$  – измеримая функция. Это необходимое свойство.

<sup>3</sup>Пусть даны два пространства с  $\sigma$ -конечными мерами  $(X_1, \mathcal{F}_1, \mu_1)$  и  $(X_2, \mathcal{F}_2, \mu_2)$ . Обозначим через  $(X_1 \times X_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$  их произведение. Пусть функция  $f: X_1 \times X_2 \rightarrow \mathbb{R}$  интегрируема относительно меры  $\mu_1 \otimes \mu_2$ . Тогда

(a) Функция  $x_1 \mapsto \int_{X_2} f(x_1, x_2) \mu_2(dx_2)$  определена и интегрируема относительно  $\mu_1$ .

(b) Функция  $x_2 \mapsto \int_{X_1} f(x_1, x_2) \mu_1(dx_1)$  определена и интегрируема относительно  $\mu_2$ .

(c)  $\iint_{X_1 \times X_2} f(x_1, x_2) \mu_1 \otimes \mu_2(dx_1, dx_2) = \int_{X_1} \int_{X_2} f(x_1, x_2) \mu_2(dx_2) \mu_1(dx_1) = \int_{X_2} \int_{X_1} f(x_1, x_2) \mu_1(dx_1) \mu_2(dx_2).$

**Теорема 1.18 (Лемма о факторизации).** Пусть  $Y : \Omega \rightarrow \Omega'$   $\mathcal{A}$ - $\mathcal{A}'$ -измерима и  $Z : \Omega \rightarrow \bar{\mathbb{R}}$   $\mathcal{A}$ - $\bar{\mathcal{B}}$ -измерима. Тогда  $Z$   $\sigma(Y)$ - $\bar{\mathcal{B}}$ -измерима тогда и только тогда, когда существует  $\mathcal{A}'$ - $\bar{\mathcal{B}}$ -измеримая функция  $g : \Omega' \rightarrow \bar{\mathbb{R}}$ , такая что  $Z = g \circ Y$ .

ДОКАЗАТЕЛЬСТВО:

" $\Leftarrow$ "  $Y$   $\sigma(Y)$ - $\mathcal{A}'$ -измерима и  $g$   $\mathcal{A}'$ - $\bar{\mathcal{B}}$ -измерима. Следовательно,  $g \circ Y$   $\sigma(Y)$ - $\bar{\mathcal{B}}$ -измерима.

" $\Rightarrow$ " Пусть  $Z = \sum_{i=1}^n \alpha_i 1_{A_i}$ , где  $A_i \in \sigma(Y)$ . Тогда, существуют  $A'_i \in \mathcal{A}'$ , такие что  $A_i = Y^{-1}(A'_i)$ . Тогда, пусть  $g = \sum_{i=1}^n \alpha_i 1_{A'_i}$ . В общем случае, любая  $\sigma(Y)$ - $\bar{\mathcal{B}}$ -измеримая функция  $Z$  является пределом таких элементарных функций.

**Теорема 1.19.** Пусть  $X : \Omega \rightarrow \mathbb{R}$  и  $Y : \Omega \rightarrow \Omega'$  – случайные величины и  $\mathbb{E}[X] < \infty$ . Тогда любая  $\mathcal{A}$ - $\mathcal{B}$ -измеримая функция  $g : \Omega' \rightarrow \mathbb{R}$ , такая что  $\mathbb{E}[X|Y] = g(Y)$   $\mathbb{P}^Y$ -интегрируема, удовлетворяет

$$\int_{A'} g d\mathbb{P}^Y = \int_{\{Y \in A'\}} X d\mathbb{P} \quad \forall A' \in \mathcal{A}' \quad (1.1)$$

и  $\mathbb{P}^Y$ -п.н. единственная. И наоборот, если  $g : \Omega' \rightarrow \mathbb{R}$   $\mathcal{A}$ - $\mathcal{B}$ -измерима и удовлетворяет (1.1), то  $g(Y)$  – версия<sup>4</sup> математического ожидания.

ДОКАЗАТЕЛЬСТВО:

" $\Rightarrow$ " Для любого  $A' \in \mathcal{A}'$  имеет место

$$\int_{\{Y \in A'\}} X d\mathbb{P} = \int_{\{Y \in A'\}} \mathbb{E}[X|Y] d\mathbb{P} = \int_{\{Y \in A'\}} g \circ Y d\mathbb{P} = \int_{A'} g d\mathbb{P}^Y.$$

$\mathbb{P}^Y$ -интегрируемость  $g$  следует из (1.1). Предположим, что  $\mathbb{E}[X|Y] = h(Y)$ , тогда

$$\int_{A'} g d\mathbb{P}^Y = \int_{A'} h d\mathbb{P}^Y \quad \forall A' \in \mathcal{A}',$$

вследствие (1.1). Разложив функции на положительные и отрицательные части:

$$g = g^+ - g^- \quad \text{и} \quad h = h^+ - h^-,$$

получаем

$$\int_{A'} (g^+ + h^-) d\mathbb{P}^Y = \int_{A'} (h^+ + g^-) d\mathbb{P}^Y \quad \forall A' \in \mathcal{A}'.$$

Как и в доказательстве Теоремы 1.5:

$$g^+ + h^- = h^+ + g^- \quad \mathbb{P}^Y - \text{п.н.} \quad \Rightarrow \quad g^+ - g^- = h^+ - h^- \quad \mathbb{P}^Y - \text{п.н.}$$

" $\Leftarrow$ "  $g \circ Y$   $\sigma(Y)$ - $\bar{\mathcal{B}}$ -измерима по определению. Используя замену переменных и (1.1), мы получаем:

$$\int_{\{Y \in A'\}} g \circ Y d\mathbb{P} = \int_{A'} g d\mathbb{P}^Y = \int_{\{Y \in A'\}} X d\mathbb{P} \quad \forall A' \in \mathcal{A}'.$$

Следовательно,

$$\int_C g \circ Y d\mathbb{P} = \int_C X d\mathbb{P} \quad \forall C \in \sigma(Y)$$

и также  $g \circ Y = \mathbb{E}[X|Y]$ .

<sup>4</sup>Версия от версии отличается не более чем на множестве нулевой меры Лебега.

**Пример 1.20.** Рассмотрим  $(\Omega', \mathcal{A}')$ , где  $\{y\} \in \mathcal{A}'$  для определенного события  $y \in \Omega'$ . В этом случае (1.1) следует читать следующим образом:

$$\int_{\{y\}} g d\mathbb{P}^Y = g(y)\mathbb{P}(Y = y) = \int_{\{Y=y\}} X d\mathbb{P}.$$

Если  $\mathbb{P}(Y = y) > 0$ , то

$$g(y) = \frac{1}{\mathbb{P}(Y = y)} \int_{\{\mathbb{P}(Y=y)\}} X d\mathbb{P} =: \mathbb{E}[X|Y = y].$$

В большинстве случаев  $\mathbb{P}(Y = y) = 0$ .

**Определение 1.21.** Пусть  $X : \Omega \rightarrow \mathbb{R}$  интегрируема и  $Y : \Omega \rightarrow \Omega'$ . Если  $g$  удовлетворяет равенству (1.1) и  $\mathcal{A}'$ - $\mathcal{B}$ -измерима и  $\mathbb{P}^Y$ -интегрируема, тогда

$$g(y) := \mathbb{E}[X|Y = y]$$

называется *математическим ожиданием  $X$  при условии  $Y = y$* .

**Замечание 1.22.** В общем случае,  $\mathbb{E}[X|Y]$  – случайная величина, которая принимает различные значения в соответствии с реализацией  $\omega \mapsto Y(\omega)$ . В свою очередь,  $\mathbb{E}[X|Y = y]$  – вещественное число, принадлежащее реализации  $Y(\omega) = y$ .

**Пример 1.23.** Мы уже посчитали  $\mathbb{E}[X|Y]$  в двух важных случаях (см. Пример 1.16):

(i) Для дискретных случайных величин  $X$  и  $Y$ , мы получаем:

$$\mathbb{E}[X|Y = y] = g(y) = \begin{cases} \sum_{x \in I_X} x \frac{p_{xy}}{p_y}, & p_y > 0, \\ 0, & p_y = 0. \end{cases}$$

(ii) Для непрерывных:

$$\mathbb{E}[X|Y = y] = g(y) = \begin{cases} \int x \frac{f_{X,Y}(x,y)}{f_Y(y)}, & f_Y(y) > 0, \\ 0, & f_Y(y) = 0. \end{cases}$$

Оба равенства схожи с формулами для стандартного математического ожидания, за исключением того, что мы замещаем  $p_x$  и  $f_X(x)$  их условными версиями:

$$\mathbb{P}(X = x|Y = y) = \frac{p_{xy}}{p_y} \quad \text{и} \quad f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

Если  $X$  и  $Y$  независимы, то мы получаем обыкновенное математическое ожидание (см. Следствие 1.15).

**Замечание 1.24.** В начале главы, мы заметили, что  $\mathbb{E}[X|\mathcal{F}]$  – лучшая аппроксимация  $X$  при условии  $\mathcal{F}$ . Давайте, опишем это строго.

**Теорема 1.25.** Пусть  $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ . Тогда для любой случайной величины  $Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$  соблюдается неравенство

$$\mathbb{E}[(X - Y)^2] \geq \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}])^2],$$

которое вырождается в равенство тогда и только тогда, когда  $Y = \mathbb{E}[X|\mathcal{F}]$  п.н.

**ДОКАЗАТЕЛЬСТВО:** Из условного неравенства Йенсена и Следствия 1.15 следует:

$$\mathbb{E}[(\mathbb{E}[X|\mathcal{F}])^2] \leq \mathbb{E}[\mathbb{E}[X^2|\mathcal{F}]] = \mathbb{E}[X^2] < \infty.$$

Далее, используя Следствие 1.15 и Теорему 1.12, получаем

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY|\mathcal{F}]] = \mathbb{E}[Y\mathbb{E}[X|\mathcal{F}]]$$

и

$$\mathbb{E}[X\mathbb{E}[X|\mathcal{F}]] = \mathbb{E}[\mathbb{E}[X\mathbb{E}[X|\mathcal{F}]]] = \mathbb{E}[\mathbb{E}[X^2|\mathcal{F}]].$$

Следовательно,

$$\begin{aligned}\mathbb{E}[(X - Y)^2] - \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}])^2] &= \mathbb{E}[X^2 - 2XY + Y^2] - \mathbb{E}[X^2 - 2X\mathbb{E}[X|\mathcal{F}] + (\mathbb{E}[X|\mathcal{F}])^2] \\ &= \mathbb{E}[Y^2 - 2\mathbb{E}[Y\mathbb{E}[X|\mathcal{F}]] + (\mathbb{E}[X|\mathcal{F}])^2] \\ &= \mathbb{E}[(Y - \mathbb{E}[X|\mathcal{F}])^2] \geq 0.\end{aligned}$$

## Упражнения

**1.1.** Пусть  $X$  – конечная интегрируемая вещественнозначная случайная величина, заданная на вероятностном пространстве  $(\Omega, \mathcal{A}, \mathbb{P})$  и  $\mathcal{F} \subset \mathcal{A}$  –  $\sigma$ -алгебра.

(i) Докажите **условное неравенство Йенсена**: для любой выпуклой функции  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$

$$\varphi(\mathbb{E}[X|\mathcal{F}]) \leq \mathbb{E}[\varphi(X)|\mathcal{F}] \quad \mathbb{P}\text{-п.н.}$$

*Подсказка:* для любой выпуклой функции  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  существует последовательность  $(a_n, b_n)_{n \in \mathbb{N}} \subset \mathbb{R}^2$ , такая что равенство

$$\varphi = \sup_{(a_n, b_n)} (a_n x + b_n)$$

соблюдается  $\forall x \in \mathbb{R}$ .

(ii) Докажите, что

$$\|\mathbb{E}[X|\mathcal{F}]\|_p \leq \|X\|_p \quad \forall p \geq 1,$$

где  $\|\cdot\|_p = (\mathbb{E}[|\cdot|^p])^{1/p}$ .

**1.2.** Пусть  $X$  – вещественнозначная случайная величина, заданная на вероятностном пространстве  $(\Omega, \mathcal{A}, \mathbb{P})$  и  $\mathcal{F} \subset \mathcal{A}$  –  $\sigma$ -алгебра. **Условная дисперсия** задается следующим образом:

$$\text{Var}(X|\mathcal{F}) := \mathbb{E}[(X - \mathbb{E}[X|\mathcal{F}])^2|\mathcal{F}].$$

Покажите, что

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|\mathcal{F})] + \text{Var}(\mathbb{E}[X|\mathcal{F}]).$$

**1.3.** Пусть  $X = (X_1, X_2)^T \in \mathbb{R}^2$  – двумерный нормально распределенный случайный вектор:  $X \sim \mathcal{N}(\mu, \Sigma)$ , с математическим ожиданием  $\mu = (0, 0)^T$  и ковариационной матрицей

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Покажите, что

$$\mathbb{E}[X_1|X_2] = \rho \frac{\sigma_1}{\sigma_2} X_2.$$

**1.4.** Пусть  $X, Y$  – две случайные величины с совместной плотностью

$$f_{X,Y}(x, y) = \lambda^2 e^{-\lambda x} 1_{(0, \infty)}(x - y) 1_{(0, \infty)}(y).$$

(i) Найдите маргинальные плотности  $f_X$  и  $f_Y$ . Что это за распределения?

(ii) Вычислите  $\mathbb{E}[Y|X]$ .

(iii) Покажите, что  $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$ .

**1.5.** Пусть  $X$  и  $Y$  – две вещественнозначные случайные величины, заданные на вероятностном пространстве  $(\Omega, \mathcal{A}, \mathbb{P})$  и  $\varphi : \Omega \times \Omega \rightarrow \mathbb{R}$  –  $\mathcal{A}$ -измеримая функция. Покажите с помощью теоремы Фубини, что для функции  $g : x \mapsto \mathbb{E}[\varphi(x, Y)]$  выполняется равенство

$$g(X) = \mathbb{E}[\varphi(X, Y)|X] \quad \mathbb{P}\text{-п.н.}$$



## Глава 2

# Основы точечного оценивания

**Пример 2.1.** Перед введением лекарства в производство проводятся опыты на животных для оценки его качества в зависимости от дозирования. Обследуется животное и проверяется, выздоровело ли оно или нет, приняв дозу  $X$ . Модель:

$$Y \sim \text{Bin}(1, p(X)),$$

где  $p(X)$  - вероятность выздоровления животного, принявшего дозу  $X$ . Как правило, обследуется несколько животных:  $Y_1, \dots, Y_n$ . Предположим, что эти случайные величины независимы. Подбираем различные дозы  $X_1, \dots, X_n$  таким образом, что:

$$Y_i \sim \text{Bin}(1, p(X_i)).$$

Цель: оценить функцию  $p : [0, \infty) \rightarrow [0, 1]$ . Упростим до параметрической модели, например:

$$p(x) = 1 - e^{-\beta x}, \quad \beta > 0.$$

Тогда оценка функции  $p(x)$  эквивалентна оценке параметра  $\beta$ .

**Предположение 2.2.** Пусть  $(\Omega, \mathcal{A}, \mathbb{P})$  – вероятностное пространство,  $(\mathcal{X}, \mathcal{B})$  – измеримое пространство и  $X : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathcal{X}, \mathcal{B})$  – случайная величина. Зададим

$$P(B) = \mathbb{P}^X(B) = \mathbb{P}(X^{-1}(B)), \quad B \in \mathcal{B}.$$

Тогда  $(\mathcal{X}, \mathcal{B}, P)$  является вероятностным пространством,  $\mathcal{X}$  называется **выборочным пространством**, а  $x = X(\omega)$  – **выборкой**.

**Определение 2.3.** Пусть  $X \neq \emptyset$ ,  $\mathcal{B}$  –  $\sigma$ -алгебра на  $\mathcal{X}$  и  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  – семейство вероятностных мер на  $(\mathcal{X}, \mathcal{B})$ , в котором  $|\Theta| \geq 2$  и  $P_\vartheta \neq P_{\vartheta'}$  для любых  $\vartheta \neq \vartheta'$ . Тогда  $\Theta$  называется **вероятностным пространством** и  $(\mathcal{X}, \mathcal{B}, \mathcal{P})$  называется **статистическим экспериментом**.

**Замечание 2.4.** Интерпретация: мы заинтересованы в реальном распределении  $P \in \mathcal{P}$  случайной величины  $X : \Omega \rightarrow \mathcal{X}$ . На основе выборки  $x = X(\omega)$  принимаем решение о неизвестном распределении  $P$ . Идентифицируя  $\mathcal{P}$  с параметрическим пространством  $\Theta$ , выбор  $P$  эквивалентен выбору  $\vartheta$ .

**Пример 2.5.** Пусть  $Y_1, \dots, Y_n$  независимы и имеют распределение:

$$Y_i \sim \text{Bin}(1, p(X_i)) = \text{Bin}(1, 1 - \exp(-\beta X_i)) = P_i^\beta.$$

Формально, составляющие статистического эксперимента:

$$\mathcal{X} = \{0, 1\}^n, \quad \mathcal{B} = \mathcal{P}(\mathcal{X}), \quad \mathcal{P} = \{\otimes_{i=1}^n P_i^\beta \mid \beta > 0\}, \quad \Theta = [0, \infty).$$

**Определение 2.6.** Пусть  $(\Gamma, \mathcal{A}_\Gamma)$  – измеримое пространство и  $\gamma : \Theta \rightarrow \Gamma$  – отображение. Измеримая функция

$$g : (\mathcal{X}, \mathcal{B}) \rightarrow \Gamma(\mathcal{A}_\Gamma)$$

называется *(точечной) оценкой*  $\gamma(\vartheta)$ .

**Пример 2.7.**

(i) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2) = P_{\mu, \sigma^2}$ ,  $X = (X_1, \dots, X_n)^T$  и

$$\mathcal{X} = \mathbb{R}^n, \quad \mathcal{B} = \mathcal{B}^n, \quad \mathcal{P} = \{\otimes_{i=1}^n P_{\mu, \sigma^2} \mid \mu \in \mathbb{R}, \sigma^2 > 0\}, \quad \Theta = \mathbb{R} \times [0, \infty).$$

Типичная оценка для параметра  $\gamma(\vartheta) = \vartheta = (\mu, \sigma^2)$ :

$$g : \begin{cases} \mathbb{R}^n & \rightarrow \Theta \\ x & \mapsto \begin{pmatrix} \bar{x}_n \\ \hat{s}_n^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_i \\ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \end{pmatrix} \end{cases}$$

(ii) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim F$ , где  $F(x) = \mathbb{P}(X_i \leq x)$  – неизвестная функция распределения. В этом случае  $\Theta$  – бесконечномерное семейство всех функций распределения. Если мы заинтересованы в значении функции только в одной точке:

$$\gamma : \begin{cases} \Theta & \rightarrow \Gamma = [0, 1] \\ F & \mapsto F(0) = \mathbb{P}(X_i \leq 0), \end{cases}$$

то точечная оценка будет:

$$g : \begin{cases} \mathbb{R}^n & \rightarrow \Gamma \\ x & \mapsto \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq 0\}}. \end{cases}$$

**Замечание 2.8.** Из предыдущего примера становится ясно, зачем нужно вводить отображение  $\gamma : \Theta \rightarrow \Gamma$ , поскольку мы не всегда заинтересованы в значении самого параметра  $\vartheta$ , но в подходящем функционале.

**Определение 2.9.** Измеримая функция  $L : \Gamma \times \Gamma \rightarrow [0, \infty)$  называется *функцией потерь*. Для точечной оценки  $g : \mathcal{X} \rightarrow \Gamma$  функция

$$R(\cdot, g) : \begin{cases} \Theta & \rightarrow [0, \infty) \\ \vartheta & \mapsto \mathbb{E}[L(\gamma(\vartheta), g(X))] = \int_{\mathcal{X}} L(\gamma(\vartheta), g(x)) P_\vartheta(dx) \end{cases}$$

называется *риском*  $g$  от  $L$ .

**Замечание 2.10.** Если  $\vartheta$  – истинный параметр, а  $g(x)$  – оценка, то функция  $L(\gamma(\vartheta), g(x))$  измеряет соответствующие потери. Если  $\Gamma$  – измеримое пространство, то, как правило, функции потерь зависят от дистанции между  $\gamma(\vartheta)$  и  $g(x)$ , например квадратичная функция потерь  $L(x, y) = (x - y)^2$  для  $\Gamma = \mathbb{R}$ . Тогда риск – это ожидаемые потери.

**Определение 2.11.** Пусть  $L$  – функция потерь и  $\gamma(\vartheta)$  – оцениваемый параметр. Если  $\mathcal{K}$  – множество всех точечных оценок  $\gamma(\vartheta)$ , то  $g^* \in \mathcal{K}$  называется *равномерно лучшей оценкой*, если

$$R(\vartheta, g^*) = \inf_{g \in \mathcal{K}} R(\vartheta, g) \quad \forall \vartheta \in \Theta.$$

**Пример 2.12.** Как правило, не существует равномерно лучших оценок, как и не существует оценки, которая была бы равномерно лучше другой. Например, пусть

$$\mathcal{X} = \mathbb{R}, \quad \mathcal{P} = \{P_\mu = \mathcal{N}(\mu, 1) \mid \mu \in \mathbb{R}\}, \quad \gamma(\mu) = \mu$$

и функция потерь квадратичная. Возьмем тривиальную оценку  $g_\nu(x) = \nu$ . Тогда риск:

$$R(\mu, g_\nu) = \mathbb{E}_\mu[(\mu - \nu)^2] = (\mu - \nu)^2.$$

В частности,  $R(\nu, g_\nu) = 0$ . Таким образом, никакая оценка  $g_\nu$  не является равномерно лучшей, чем некоторая  $g_\mu$ . Также, чтобы получить равномерно лучшую оценку равенство

$$\mathbb{E}_\mu[(g^*(X) - \mu)^2] = 0 \quad \forall \mu \in \mathbb{R} \iff g^*(x) = \mu \text{ } P_\mu\text{-п.н.} \quad \forall \mu \in \mathbb{R}$$

должно выполняться, что приводит к противоречию.



**Замечание 2.13.** Для того, чтобы всё же найти "оптимальную" оценку, можно выбрать две опции:

- (i) ограничиться подклассами  $\bar{\mathcal{K}} \subset \mathcal{K}$
- (ii) выбрать иной критерий, нежели равномерно меньший риск

**Определение 2.14.**

- (i) Оценка  $g^* \in \mathcal{K}$  называется *допустимой*, если не существует такой оценки  $g \in \mathcal{K}$ , что

$$R(\vartheta, g) \leq R(\vartheta, g^*) \quad \forall \vartheta \in \Theta$$

и хотя бы для одного значения  $\vartheta^* \in \Theta$

$$R(\vartheta^*, g) < R(\vartheta^*, g^*).$$

- (ii) Класс  $\tilde{\mathcal{K}} \subset \mathcal{K}$  называется *полным*, если для любой оценки  $g \in \mathcal{K} \setminus \tilde{\mathcal{K}}$  существует такая оценка  $\tilde{g} \in \tilde{\mathcal{K}}$ , что

$$R(\vartheta, \tilde{g}) \leq R(\vartheta, g) \quad \forall \vartheta \in \Theta.$$

Если класс  $\tilde{\mathcal{K}}$  не содержит полного подкласса, то  $\tilde{\mathcal{K}}$  называется *минимальным полным* классом.

**Замечание 2.15.** Оценки наподобие  $g_\nu(x) = \nu$  являются допустимыми, так что даже априори плохие оценки допустимые или принадлежат полным классам. Как следствие, мы нуждаемся в больших ограничениях.

**Определение 2.16.**

- (i) Пусть  $g$  – оценка функции  $\gamma : \Theta \rightarrow \Gamma$ . Тогда,

$$B_\vartheta(g) = \mathbb{E}_\vartheta[g(X)] - \gamma(\vartheta)$$

называется *смещением*  $g$ . Оценка  $g$  называется *несмещенной*, если

$$B_\vartheta(g) = 0 \quad \forall \vartheta \in \Theta.$$

- (ii) Оценка  $g^*$  называется *несмещенной с равномерно минимальной дисперсией (UMVU)*, если

$$g^* \in \mathcal{E}_\gamma = \{g \mid g \text{ несмещенная и } g \in L^2(P_\vartheta)\}$$

и

$$\text{Var}_\vartheta(g^*(X)) = \mathbb{E}[(g^*(X) - \gamma(\vartheta))^2] = \inf_{g \in \mathcal{E}_\gamma} \text{Var}_\vartheta(g(X)) \quad \forall \vartheta \in \Theta. \quad (2.1)$$

**Замечание 2.17.**

- (i) Оценки, удовлетворяющие равенству (2.1) лишь для одного значения  $\vartheta \in \Theta$  называются локально оптимальными. Так как истинное значение  $\vartheta$  неизвестно, то на практике это неприменимо.
- (ii) Если в качестве функции потерь мы выбираем  $L(x, y) = (x - y)^2$ , то для любой оценки  $g \in L^2(P_\vartheta)$  величина

$$MSE_\vartheta(g) = R(\vartheta, g) = \mathbb{E}_\vartheta[(g(X) - \gamma(\vartheta))^2] = \text{Var}_\vartheta(g(X)) + B_\vartheta^2(g)$$

называется *среднеквадратической ошибкой*. Если  $g$  несмещенная, то

$$MSE_\vartheta(g) = \text{Var}_\vartheta(g(X)).$$

(iii) Аналогичное определение для несмещенных оценок  $g$ , если  $\gamma : \Theta \rightarrow \mathbb{R}^d$ .

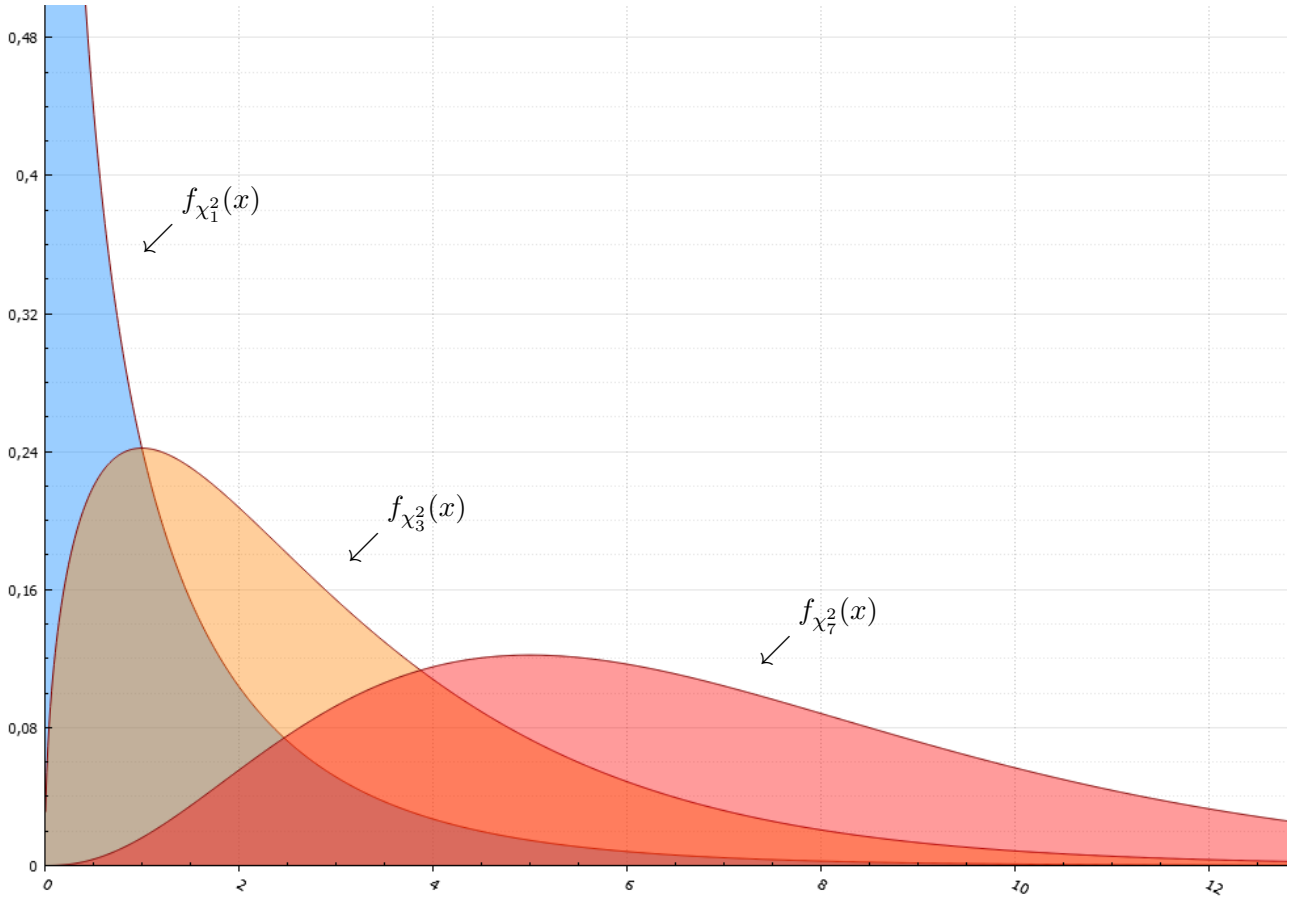
**Определение 2.18.**

(i) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(0, 1)$ . Тогда случайная величина  $Z = \sum_{i=1}^n X_i^2$  имеет **распределение хи-квадрат** с  $n$  степенями свободы. Обозначение:  $Z \sim \chi_n^2$ . Плотность распределения  $Z$ :

$$f_{\chi_n^2}(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} 1_{(0, \infty)}(x),$$

где  $\Gamma(\cdot)$  в знаменателе обозначает гамма-функцию:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0.$$



(ii) Пусть  $Y \sim \mathcal{N}(0, 1)$  и  $Z \sim \chi_n^2$  независимы. Тогда, случайная величина

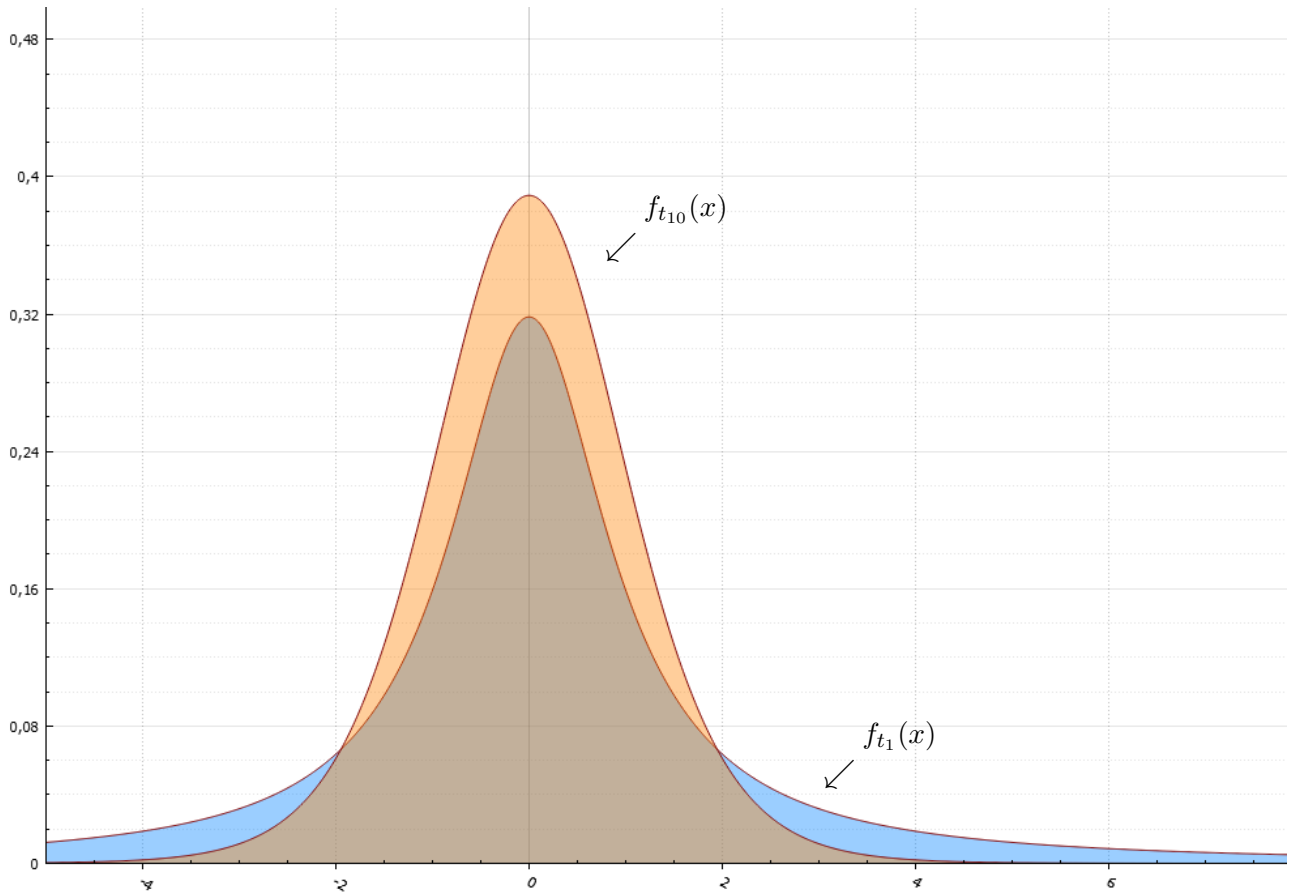
$$T = \frac{Y}{\sqrt{Z/n}}$$

имеет **t-распределение** с  $n$  степенями свободы. Обозначение:  $T \sim t_n$ . Плотность распределения:

$$f_{t_n}(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

**Замечание 2.19.** Если  $Z \sim \chi_n^2$ , то

$$\mathbb{E}[Z] = \sum_{i=1}^n \mathbb{E}[X_i^2] = n$$



и

$$\text{Var}(Z) = \sum_{i=1}^n \text{Var}(X_i^2) = n \left( \mathbb{E}[X_1^4] - (\mathbb{E}[X_1^2])^2 \right) = 2n.$$

**Лемма 2.20.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$ . Тогда

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

и

$$\hat{s}_n^2(X) = \frac{1}{n} \sum (X_i - \bar{X}_n)^2 \sim \frac{\sigma^2}{n} \chi_{n-1}^2,$$

и обе оценки независимы.

**ДОКАЗАТЕЛЬСТВО:** Распределение  $\bar{X}_n$  следует из свойства нормального распределения. Зададим

$$Y_i = \frac{X_i - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

и  $Y = (Y_1, \dots, Y_n)^T$ . Выберем ортогональную матрицу  $A$ , такую что её последняя строчка:

$$\left( \frac{1}{\sqrt{n}} \dots \frac{1}{\sqrt{n}} \right) = v^T.$$

Тогда для  $Z = AY$  имеет место равенство:

$$\|Z\|_2^2 = \sum_{i=1}^n Z_i^2 = Z^T Z = Y^T A^T A Y = Y^T Y = \sum_{i=1}^n Y_i^2 = \|Y\|_2^2.$$

Так как  $\text{Cov}(Z) = A^T A = \mathbb{I}_n$ , то  $Z \sim \mathcal{N}(0, \mathbb{I}_n)$ . Также:

$$\sqrt{n}\bar{X}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\sigma Y_i + \mu) = \sigma v^T Y + \sqrt{n}\mu = \sigma Z_n + \sqrt{n}\mu$$

и

$$\begin{aligned}
n\hat{s}_n^2(X) &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sigma^2 \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \\
&= \sigma^2 \left( \sum_{i=1}^n Y_i^2 - n\bar{Y}_n^2 \right) = \sigma^2 \left( \sum_{i=1}^n Y_i^2 - \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \right)^2 \right) \\
&= \sigma^2 \left( \sum_{i=1}^n Z_i^2 - Z_n^2 \right) = \sigma^2 \sum_{i=1}^{n-1} Z_i^2 \sim \sigma^2 \chi_{n-1}^2.
\end{aligned}$$

Обе оценки независимы как функции от  $Z_n$  и  $Z_1, \dots, Z_{n-1}$  соответственно.

**Следствие 2.21.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$ . Тогда

$$T = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{s}_n(X)} \sim t_{n-1}.$$

ДОКАЗАТЕЛЬСТВО: Представим  $T$  в виде:

$$T = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sqrt{\frac{\sigma^2 n}{n\hat{s}_n(X)}}$$

и используем предыдущую лемму.

**Пример 2.22.** Проверим, какие оценки из Примера 2.7 являются несмещенными. Мы знаем, что  $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ , значит оценка  $\bar{X}_n$  несмещенная. С другой стороны, используя Лемму 2.20 мы получаем:

$$\mathbb{E}_\vartheta[\hat{s}_n^2(X)] = \frac{\sigma^2}{n}(n-1) \neq \sigma^2.$$

Стоит также заметить, что

$$MSE_\vartheta(\bar{X}_n) = \text{Var}_\vartheta(\bar{X}_n) = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$$

и

$$MSE_\vartheta(\hat{s}_n^2(X)) = \text{Var}_\vartheta(\hat{s}_n^2(X)) + B_\vartheta^2(\hat{s}_n^2(X)) = \frac{\sigma^4}{n^2} 2(n-1) + \frac{\sigma^4}{n^2} = \frac{2n-1}{n^2} \sigma^4 \xrightarrow{n \rightarrow \infty} 0$$

**Теорема 2.23 (Неравенство Рао-Крамера).** Пусть  $\mu$  –  $\sigma$ -конечная мера и  $P_\vartheta \ll \mu$  с  $\mu$ -плотностью  $f(\cdot, \vartheta)$ . Пусть также  $\Theta \subset \mathbb{R}$  – открытое пространство и  $g : \mathcal{X} \rightarrow \mathbb{R}$  – оценка. Дальнейшие предположения (условия регулярности):

(i) Множество  $M_f = \{x \in \mathcal{X} \mid f(x, \vartheta) > 0\}$  не зависит от параметра  $\vartheta$ .

(ii) Частная производная  $\frac{\partial}{\partial \vartheta} f(x, \vartheta)$  существует  $\forall x \in \mathcal{X}$ .

(iii) (a)  $\mathbb{E}_\vartheta\left[\frac{\partial}{\partial \vartheta} \log f(X, \vartheta)\right] = 0$ ,

(b)  $\mathbb{E}_\vartheta[g(X) \frac{\partial}{\partial \vartheta} \log f(X, \vartheta)] = \frac{\partial}{\partial \vartheta} \mathbb{E}_\vartheta[g(X)]$ .

(iv)  $0 < I(f(\cdot, \vartheta)) = \mathbb{E}_\vartheta\left[\left(\frac{\partial}{\partial \vartheta} \log f(X, \vartheta)\right)^2\right] < \infty$

Тогда:

$$\text{Var}_\vartheta(g(X)) \geq \frac{\left(\frac{\partial}{\partial \vartheta} \mathbb{E}_\vartheta[g(X)]\right)^2}{I(f(\cdot, \vartheta))} \quad \forall \vartheta \in \Theta \quad (2.2)$$

ДОКАЗАТЕЛЬСТВО: Зададим функцию:

$$U_\vartheta(x) = \begin{cases} 0, & \text{если } x \notin M_f, \\ \frac{\partial}{\partial \vartheta} \log f(x, \vartheta), & \text{иначе.} \end{cases}$$

Из условия (iii)(a) мы знаем, что  $\mathbb{E}[U_{\vartheta}(X)] = 0$  и

$$\text{Var}_{\vartheta}(U_{\vartheta}(X)) = \mathbb{E}[(U_{\vartheta}(X))^2] = I(f(\cdot, \vartheta)).$$

Тогда используя неравенство Коши-Шварца мы получаем:

$$\begin{aligned} \left( \frac{\partial}{\partial \vartheta} \mathbb{E}_{\vartheta}[g(X)] \right)^2 &= \left( \mathbb{E}_{\vartheta}[g(X) \cdot U_{\vartheta}(X)] \right)^2 = \left( \text{Cov}_{\vartheta}(g(X), U_{\vartheta}(X)) \right)^2 \\ &\leq \text{Var}_{\vartheta}(g(X)) \cdot \text{Var}_{\vartheta}(U_{\vartheta}(X)) = I(f(\cdot, \vartheta)) \cdot \text{Var}_{\vartheta}(g(X)). \end{aligned}$$

### Определение 2.24.

- (i)  $I(f(\cdot, \vartheta))$  называется **информацией Фишера** семейства  $\mathcal{P} = \{P_{\vartheta} \mid \vartheta \in \Theta\}$ .
- (ii) Если оценка  $g$  такая, что неравенство (2.2) превращается в равенство, то такая оценка называется **эффективной** для  $\gamma(\vartheta) = \mathbb{E}_{\vartheta}[g(X)]$ .

### Замечание 2.25.

- (i) Запишем эквивалентные версии равенств в условии (iii):

$$(a) \int_{\mathcal{X}} \frac{\partial}{\partial \vartheta} \log f(x, \vartheta) f(x, \vartheta) d\mu(x) = \int_{\mathcal{X}} \frac{\partial}{\partial \vartheta} f(x, \vartheta) d\mu(x) = \frac{\partial}{\partial \vartheta} \int_{\mathcal{X}} f(x, \vartheta) d\mu(x) = 0,$$

$$(b) \int_{\mathcal{X}} g(x) \frac{\partial}{\partial \vartheta} f(x, \vartheta) d\mu(x) = \frac{\partial}{\partial \vartheta} \int_{\mathcal{X}} g(x) f(x, \vartheta) d\mu(x).$$

В обоих случаях условие (iii) означает, что можно менять местами интегралы и производные.

- (ii) Теорема 2.23 дает нижнюю границу для дисперсии оценки  $\gamma(\vartheta) = \mathbb{E}[g(X)]$  и может быть использована для получения UMVU-оценок. Если условия регулярности соблюдены, то любая эффективная и несмещенная оценка является UMVU.
- (iii) Если  $X_1, \dots, X_n$  i.i.d.  $\sim P_{\vartheta}^1 \ll \mu^1$  и  $X = (X_1, \dots, X_n)^T$ , то  $P_{\vartheta} = \otimes_{i=1}^n P_{\vartheta}^1 \ll \otimes_{i=1}^n \mu^1$  и  $f(x, \vartheta) = \prod_{i=1}^n f^1(x_i, \vartheta)$ , где  $f^1(\cdot, \vartheta) - \mu^1$ -плотность  $P_{\vartheta}^1$ . Несложно увидеть, что в данном случае:

$$I(f(\cdot, \vartheta)) = nI(f^1(\cdot, \vartheta)).$$

**Пример 2.26.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, 1)$  с плотностью

$$f^1(x, \vartheta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2} \right\}.$$

Тогда

$$I(f^1(\cdot, \mu)) = \mathbb{E} \left[ \left( \frac{\partial}{\partial \mu} \log f^1(X_1, \mu) \right)^2 \right] = \mathbb{E}[(X - \mu)^2] = 1.$$

В частности, информация Фишера для  $X = (X_1, \dots, X_n)^T$ :  $I(f(\cdot, \vartheta), \mu) = n$  и тогда неравенство Рао-Крамера для несмещенных оценок:

$$\text{Var}_{\mu}(g(X)) \geq \frac{1}{n} \left( \frac{\partial}{\partial \mu} \mathbb{E}_{\mu}[g(X)] \right)^2 = \frac{1}{n}.$$

Как следствие,  $g(x) = \bar{x}_n$  — UMVU-оценка.

**Определение 2.27.** Пусть

$$SYM(k) = \{A \in \mathbb{R}^{k \times k} \mid A \text{ симметричная}\}.$$

Для  $A, B \in SYM(k)$  будем использовать обозначения:

$$\begin{aligned} A \geq 0 &\iff A \text{ положительно полуопределенная,} \\ A \geq B &\iff A - B \geq 0. \end{aligned}$$

Частичное упорядочение определенное знаком  $\geq$  называется **упорядочением Левнера** на  $SYM(k)$ . Также зададим множества:

$$NND(k) = \{A \in SYM(k) \mid A \geq 0\},$$

$$PD(k) = \{A \in NND(k) \mid \det(A) \neq 0\}.$$

**Теорема 2.28 (Многомерное неравенство Рао-Крамера).** Пусть  $\mu$  –  $\sigma$ -конечная мера и  $P_\vartheta \ll \mu \forall \vartheta \in \Theta \subset \mathbb{R}$ . Предположим, что  $\Theta$  – открытое множество и  $g : \mathcal{X} \rightarrow \mathbb{R}^k$  – оценка. Зададим функцию

$$G(\vartheta) = \left( \frac{\partial}{\partial \vartheta_j} \mathbb{E}_\vartheta[g_i(X)] \right)_{i,j} \in \mathbb{R}^{k \times d}.$$

Тогда при условиях регулярности, схожими с условиями в Теореме 2.23 имеет место неравенство:

$$\text{Cov}_\vartheta(g(X)) \geq G(\vartheta) I^{-1}(f(\cdot, \vartheta)) G^T(\vartheta) \in \mathbb{R}^{k \times k}$$

в понимании упорядочения Левнера. Информация Фишера в данном случае:

$$I(f(\cdot, \vartheta)) = \left( \mathbb{E} \left[ \frac{\partial}{\partial \vartheta_i} \log f(X, \vartheta) \cdot \frac{\partial}{\partial \vartheta_j} \log f(X, \vartheta) \right] \right)_{i,j=1}^d \in \mathbb{R}^{d \times d}.$$

**ДОКАЗАТЕЛЬСТВО:** Для доказательства можно использовать многомерное неравенство Коши-Шварца (см. [2]) для  $Y \in \mathbb{R}^k$  и  $Z \in \mathbb{R}^d$ :

$$\mathbb{E}[YY^T] \geq \mathbb{E}[YZ^T](\mathbb{E}[ZZ^T])^{-1}\mathbb{E}[ZY^T],$$

и подставить:

$$Y = g(X) - \mathbb{E}_\vartheta[g(X)]$$

и

$$Z = \nabla_\vartheta \log f(X, \vartheta) = \begin{pmatrix} \frac{\partial}{\partial \vartheta_1} \log f(X, \vartheta) \\ \vdots \\ \frac{\partial}{\partial \vartheta_d} \log f(X, \vartheta) \end{pmatrix}.$$

**Пример 2.29.** В Примере 2.22:

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

Тогда:

$$U_\vartheta = \left( \frac{\partial}{\partial \mu} \log f^1(X_1, \vartheta), \frac{\partial}{\partial \sigma^2} \log f^1(X_1, \vartheta) \right)^T = \left( -\frac{(X_1 - \mu)/\sigma^2}{\frac{1}{2\sigma^2} + \frac{1}{\sigma^4}(X_1 - \mu)^2} \right)$$

и информация Фишера:

$$I(f^1(\cdot, \vartheta)) = \mathbb{E}[U_\vartheta, U_\vartheta^T] = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & \frac{1}{2}\sigma^{-4} \end{pmatrix} = \frac{1}{n} I(f(\cdot, \vartheta)).$$

Если оценка  $g(X)$  несмещенная, то  $G(\vartheta)$  – едичная матрица и граница Рао-Крамера:

$$\text{Cov}_\vartheta(g(X)) \geq G(\vartheta) I^{-1}(f(\cdot, \vartheta)) G^T(\vartheta) = I^{-1}(f(\cdot, \vartheta)) = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^4}{n} \end{pmatrix}.$$

В частности для оценки

$$\tilde{g}(X) = \left( \bar{X}_n, \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right)^T$$

имеет место

$$\text{Cov}_\vartheta(\tilde{g}(X)) = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{\sigma^4}{n-1} \end{pmatrix} \geq I(f(\cdot, \vartheta)),$$

то есть оценка  $\tilde{g}$  не является эффективной.

**Замечание 2.30.** В предыдущих примерах мы бездоказательно считали, что все условия регулярности для неравенства Рао-Крамера соблюдены. В дальнейшем мы обсудим семейство распределений, для которых неравенство Рао-Крамера превращается в равенство.

**Предложение 2.31.** Пусть  $\mu$  –  $\sigma$ -конечная мера и  $P_\vartheta \ll \mu$  с  $\mu$ -плотностью:

$$f(x, \vartheta) = c(\vartheta)h(x) \exp\{\vartheta T(x)\} \quad \forall \vartheta \in \Theta.$$

Тогда равенство в (2.2) достигается при оценке  $g(x) = T(x)$ .

**ДОКАЗАТЕЛЬСТВО:** В одной из дальнейших теорем мы докажем, что функция  $c$  бесконечно дифференцируема ( $c \in C^\infty(\Theta)$ ) и что производные и интегралы могут быть поменяны местами. Сначала, заметим, что так как  $\int_{\mathcal{X}} f(x) \mu(dx) = 1 \quad \forall \vartheta \in \Theta$ , то

$$c(\vartheta) = \left( \int_{\mathcal{X}} h(x) \exp\{\vartheta T(x)\} \mu(dx) \right)^{-1}.$$

и

$$\begin{aligned} 0 &= \frac{\partial}{\partial \vartheta} \int_{\mathcal{X}} c(\vartheta) h(x) \exp\{\vartheta T(x)\} \mu(dx) \\ &= \int_{\mathcal{X}} (c'(\vartheta) + c(\vartheta) T(x)) h(x) \exp\{\vartheta T(x)\} \mu(dx). \end{aligned}$$

Используя эти два равенства получаем:

$$\begin{aligned} \mathbb{E}_\vartheta[T(X)] &= c(\vartheta) \int_{\mathcal{X}} h(x) T(x) \exp\{\vartheta T(x)\} \mu(dx) \\ &= -c'(\vartheta) \int_{\mathcal{X}} h(x) \exp\{\vartheta T(x)\} \mu(dx) \\ &= -\frac{c'(\vartheta)}{c(\vartheta)} = (-\log c(\vartheta))'. \end{aligned}$$

Информация Фишера:

$$I(f(\cdot, \vartheta)) = \mathbb{E}_\vartheta \left[ \left( \frac{\partial}{\partial \vartheta} \log f(X, \vartheta) \right)^2 \right] = \mathbb{E}_\vartheta [(T(X) + (\log c(\vartheta))')^2] = \text{Var}_\vartheta(T(X)).$$

Также:

$$\begin{aligned} \frac{\partial}{\partial \vartheta} \mathbb{E}_\vartheta[T(X)] &= \int_{\mathcal{X}} c'(\vartheta) h(x) T(x) \exp\{\vartheta T(x)\} \mu(dx) + \int_{\mathcal{X}} c(\vartheta) h(x) T^2(x) \exp\{\vartheta T(x)\} \mu(dx) \\ &= \frac{c'(\vartheta)}{c(\vartheta)} \int_{\mathcal{X}} c(\vartheta) h(x) T(x) \exp\{\vartheta T(x)\} \mu(dx) + \mathbb{E}_\vartheta[(T(X))^2] \\ &= \mathbb{E}_\vartheta[(T(X))^2] - (\mathbb{E}_\vartheta[T(X)])^2. \end{aligned}$$

Из этого следует:

$$\frac{\left( \frac{\partial}{\partial \vartheta} \mathbb{E}_\vartheta[T(X)] \right)^2}{I(f(\cdot, \vartheta))} = \text{Var}_\vartheta(T(X)).$$

**Определение 2.32.**

- (i) Семейство  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  называется **экспоненциальным семейством**, если существуют такие вещественнозначные  $c, Q_1, \dots, Q_k : \Theta \rightarrow \mathbb{R}$  и измеримые функции  $h, T_1, \dots, T_k : \mathcal{X} \rightarrow \mathbb{R}$ , что  $\mu$ -плотность  $P_\vartheta$ :

$$f(x, \vartheta) = c(\vartheta)h(x) \exp \left\{ \sum_{j=1}^k Q_j(\vartheta) T_j(x) \right\}$$

для  $\sigma$ -конечной меры  $\mu$ .

- (ii)  $\mathcal{P}$  называется  *$k$ -параметрическим экспоненциальным семейством*, если функции  $1, Q_1, \dots, Q_k$  и  $1, T_1, \dots, T_k$  линейно независимы ( $\mu$ -почти наверное).

**Замечание 2.33.**

- (i) В экспоненциальных семействах равенство в Теореме Рао-Крамера достигается при оценке  $(T_1, \dots, T_k)^T$ . С другой стороны, это свойство и характеризует экспоненциальные семейства (Теорема 3.42 в [3]).
- (ii) Без ограничения общности можно считать, что  $h(x) = 1$ , в другом случае можно использовать меру:

$$\mu'(dx) = h(x)\mu(dx).$$

- (iii) С этого момента мы обсуждаем только  $k$ -параметрические экспоненциальные семейства.

**Пример 2.34.**

- (i) Если  $X \sim \text{Bin}(n, \vartheta)$ , то плотность по счетной мере:

$$f(k, \vartheta) = \binom{n}{k} \vartheta^k (1 - \vartheta)^{n-k} = (1 - \vartheta)^n \binom{n}{k} \exp \left\{ k \log \left( \frac{\vartheta}{1 - \vartheta} \right) \right\},$$

где  $c(\vartheta) = (1 - \vartheta)^n$ ,  $h(x) = \binom{n}{x}$ ,  $T_1(x) = x$  и  $Q_1(\vartheta) = \log \left( \frac{\vartheta}{1 - \vartheta} \right)$ .

- (ii) Если  $X \sim \mathcal{N}(\mu, \sigma^2)$ , то  $\vartheta = (\mu, \sigma^2)^T$  и плотность:

$$\begin{aligned} f(x, \vartheta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{\mu^2}{2\sigma^2} \right\} \exp \left\{ -\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2} \right\} \\ &= c(\vartheta) \exp \left\{ Q_1(\vartheta)T_1(x) + Q_2(\vartheta)T_2(x) \right\}. \end{aligned}$$

- (iii) Если  $X \sim \text{Po}(\lambda)$ , то

$$f(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \frac{1}{x!} \exp \{x \log \lambda\}.$$

**Замечание 2.35.**

- (i) Если  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  – экспоненциальное семейство, то  $\mathcal{P}^{(n)} = \{\otimes_{j=1}^n P_\vartheta \mid \vartheta \in \Theta\}$  – также экспоненциальное семейство.
- (ii) Используя обозначение  $Q(\vartheta) = (Q_1(\vartheta), \dots, Q_k(\vartheta))^T$ , мы получаем новое параметрическое пространство  $Q(\Theta)$ , которое ныне обозначается  $\Theta$ . В таком случае  $\mu$ -плотность:

$$f(x, \vartheta) = c(\vartheta) \exp \left\{ \sum_{j=1}^k \vartheta_j T_j(x) \right\}, \quad \vartheta \in \Theta,$$

и множество

$$\Theta^* = \{\vartheta \in \mathbb{R}^k \mid f(\cdot, \vartheta) \in L^1(\vartheta)\}$$

называется *естественным параметрическим пространством*.

**Пример 2.36.** В предыдущем примере естественное параметрическое пространство:

- (i)  $X \sim \text{Bin}(n, \vartheta) : \Theta^* = \{\log \left( \frac{\vartheta}{1 - \vartheta} \right) \mid \vartheta \in (0, 1)\} = \mathbb{R}$ .
- (ii)  $X \sim \mathcal{N}(\mu, \sigma^2) : \Theta^* = \left\{ \left( \frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \right) \mid \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+ \right\} = \mathbb{R} \times \mathbb{R}^+.$



(iii)  $X \sim Po(\lambda) : \Theta^* = \{\log \lambda \mid \lambda \in \mathbb{R}^+\} = \mathbb{R}$ .

**Теорема 2.37.** *Естественное параметрическое пространство  $\Theta^*$   $k$ -параметрического экспоненциального семейства выпуклое и с непустой внутренней частью.*

ДОКАЗАТЕЛЬСТВО: Непустая внутренняя часть следует из линейной независимости. Выпуклость: пусть  $\vartheta = (\vartheta_1, \dots, \vartheta_k)^T$  и  $\vartheta' = (\vartheta'_1, \dots, \vartheta'_k)^T$  – элементы  $\Theta^*$ . Тогда для  $\alpha \in (0, 1)$  имеет место

$$\int \exp \left\{ \sum_{j=1}^k (\alpha \vartheta_j + (1 - \alpha) \vartheta'_j) T_j(x) \right\} \mu'(dx) = \int \left( \exp \left\{ \sum_{j=1}^k \vartheta_j T_j(x) \right\} \right)^\alpha \left( \exp \left\{ \sum_{j=1}^k \vartheta'_j T_j(x) \right\} \right)^{1-\alpha} \mu'(dx)$$

неравенство Гельдера

$$\left( p = \frac{1}{\alpha}, q = \frac{1}{1-\alpha} \right) \longrightarrow \leq \left( \int \exp \left\{ \sum_{j=1}^k \vartheta_j T_j(x) \right\} \mu'(dx) \right)^\alpha \left( \int \exp \left\{ \sum_{j=1}^k \vartheta'_j T_j(x) \right\} \mu'(dx) \right)^{1-\alpha} < \infty.$$

Следовательно,  $\alpha \vartheta + (1 - \alpha) \vartheta' \in \Theta^*$ .

**Теорема 2.38.** *Пусть  $\mathcal{P}$  –  $k$ -параметрическое экспоненциальное семейство с плотностями*

$$f(x, \vartheta) = c(\vartheta) \exp \left\{ \sum_{j=1}^k \vartheta_j T_j(x) \right\}.$$

Пусть также  $\Theta^{**} \subset \Theta^*$  открытое множество и  $\varphi \in L^1(P_\vartheta) \forall \vartheta \in \Theta^{**}$ . Тогда функция

$$\beta : \begin{cases} \Theta^{**} & \rightarrow \mathbb{R} \\ \vartheta & \mapsto \beta(\vartheta) := \int \varphi(x) \exp \left\{ \sum_{j=1}^k \vartheta_j T_j(x) \right\} \mu(dx) \end{cases}$$

бесконечно дифференцируемая, и её производные:

$$\left( \frac{\partial}{\partial \vartheta_1} \right)^{l_1} \dots \left( \frac{\partial}{\partial \vartheta_k} \right)^{l_k} \beta(\vartheta) = \int \varphi(x) T_1^{l_1}(x) \dots T_k^{l_k}(x) \exp \left\{ \sum_{j=1}^k \vartheta_j T_j(x) \right\} \mu(dx).$$

ДОКАЗАТЕЛЬСТВО: Теорема 2.71 в [4].

**Замечание 2.39.** Прежде чем задаваться вопросом о качестве, необходимы методы получения хорошей или хотя бы какой-нибудь оценки. Если  $\mathcal{P}$  – экспоненциальное семейство, то оценка  $T(X) = (T_1(X), \dots, T_k(X))$  является UMVU для  $\mathbb{E}_\vartheta[T(X)]$ .

**Пример 2.40.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$ , тогда совместная плотность:

$$f(x, \vartheta) = c(\vartheta) \exp \left\{ -\frac{n}{2\sigma^2} \left( \frac{1}{n} \sum_{i=1}^n x_i^2 \right) + \frac{n\mu}{\sigma^2} \left( \frac{1}{n} \sum_{i=1}^n x_i \right) \right\}.$$

Тогда оценка

$$T(X) = \left( \frac{1}{n} \sum_{i=1}^n X_i, \frac{1}{n} \sum_{i=1}^n X_i^2 \right)^T$$

является эффективной для  $(\mu, \mu^2 + \sigma^2)^T$ .

**Замечание 2.41.** Если распределение не принадлежит параметрическому семейству, то для такого случая существуют два классических метода оценивания:

(i) **Метод моментов:**

Пусть,  $X_1, \dots, X_n$  i.i.d.  $\sim P_\vartheta$ . Пусть также

$$\gamma(\vartheta) = f(m_1, \dots, m_k),$$

где  $m_j = \mathbb{E}[X_1^j] = \int x^j P_\vartheta(dx)$ . Тогда **оценка, полученная методом моментов** будет

$$\hat{\gamma}(X) = f(\hat{m}_1, \dots, \hat{m}_k),$$

где  $\hat{m}_j = \frac{1}{n} \sum_{i=1}^n X_i^j$ . Вследствие Закона Больших Чисел, при дополнительных условиях имеет место сходимость  $\hat{m}_j \xrightarrow{\mathbb{P}} m_j$ .

(ii) **Метод максимального правдоподобия:**

Пусть  $X \sim P_\vartheta \ll \mu$ ,  $\vartheta \in \Theta \subset \mathbb{R}^k$  и  $\gamma(\vartheta) = \vartheta$ . Тогда, оценка  $\hat{\theta} = \hat{\theta}(x)$  называется **оценкой максимального правдоподобия**, если

$$f(x, \hat{\theta}) = \sup_{\vartheta \in \Theta} f(x, \vartheta).$$

**Пример 2.42.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$  и мы оцениваем параметр  $\vartheta = (\mu, \sigma^2)^T = (m_1, m_2 - m_1^2)^T$ . Тогда по методу моментов:

$$\hat{\gamma}(\vartheta) = (\hat{m}_1, \hat{m}_2 - \hat{m}_1^2)^T = (\bar{x}_n, \hat{s}_n^2)^T.$$

Несложно доказать, что эта оценка совпадает с оценкой, полученной по методу максимального правдоподобия.

## Упражнения

**2.1.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \text{Exp}(\lambda)$ :

$$f(x) = \lambda e^{-\lambda x}$$

и  $X = (X_1, \dots, X_n)^T$ . Покажите, что  $g(X) = \bar{X}_n$  является UMVU-оценкой для  $\lambda^{-1}$ .

**2.2.** Пусть  $X = (X_1, \dots, X_n)^T$  – вектор i.i.d. нормально распределенных случайных величин:  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  с известным параметром  $\sigma$ .

(i) Покажите, что оценка

$$g(X) = (\bar{X}_n)^2 - \frac{\sigma^2}{n}$$

является несмещенной для  $\mu^2$ .

(ii) Вычислите квадратичный риск для  $g(X)$ .

(iii) Найдите границу Рао-Крамера для  $g(X)$ . Какой можно сделать вывод?

**2.3.** Выясните, какие из следующих распределений образуют экспоненциальное семейство. Найдите параметр  $k$ , статистики  $T_1, \dots, T_k$  и естественное параметрическое пространство в соответствующих случаях.

(i)  $f_\vartheta(x) = \frac{\vartheta^{\vartheta_1}}{\Gamma(\vartheta_1)} x^{\vartheta_1-1} e^{-\vartheta_2 x}$ ,  $x \geq 0$ ,  $\vartheta_1, \vartheta_2 > 0$ ,

(ii)  $f_\vartheta(x) = 1_{(0,\vartheta)}(x) \exp(-2 \log \vartheta + \log(2x))$ ,  $x \in \mathbb{R}$ ,  $\vartheta > 0$ ,

(iii)  $f_\vartheta(x) = \vartheta^{x-1}(1 - \vartheta)$ ,  $x \in \mathbb{N}_0$ ,  $\vartheta \in (0, 1)$ .

**2.4.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$  и  $X = (X_1, \dots, X_n)^T$ . Покажите, что оценка

$$g(X) = (\bar{X}_n, \hat{s}_n^2(X))^T$$

является оценкой максимального правдоподобия для  $\vartheta = (\mu, \sigma^2)^T$ .

**2.5.** Пусть  $X = (X_1, \dots, X_n)^T$  – вектор i.i.d. равномерно распределенных случайных величин:  $X_i \sim \mathcal{U}(0, \vartheta)$ , где оцениваемый параметр  $\vartheta > 0$ . Рассмотрим оценку

$$g(X) = \max\{X_1, \dots, X_n\}.$$

(i) Покажите, что  $g(X)$  – оценка максимального правдоподобия

(ii) Найдите математическое ожидание и дисперсию  $g(X)$ . Покажите, что

$$\text{Var}_\vartheta(g(X)) = O\left(\frac{1}{n^2}\right).$$

(iii) Найдите оценку  $\hat{g}$  методом моментов. Найдите её математическое ожидание и дисперсию.

(iv) Найдите границу Рао-Крамера для несмещенной оценки  $\vartheta$ . Применима ли здесь теорема Рао-Крамера?

(v) Сравните дисперсии  $\hat{g}$  и  $\frac{n+1}{n}g$  с границей Рао-Крамера.



## Глава 3

# Байесовское и минимаксное оценивания параметров

В предыдущей главе мы заметили, что крайне маловероятно получить равномерно лучшую оценку. Альтернативный вариант для сравнения функций риска - это интегрирование или вычисление максимума.

**Определение 3.1.** Пусть  $(\mathcal{X}, \mathcal{B}, \mathcal{P})$  - статистический эксперимент,  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  и  $(\Theta, \mathcal{A}_\Theta)$  - измеримое пространство. Вероятностная мера  $\pi$  на  $\mathcal{A}_\Theta$  называется **априорным распределением** для  $\vartheta$ . Для оценки  $g \in \mathcal{K}$  и её риска  $R(\cdot, g)$  функция

$$R(\pi, g) = \int_{\Theta} R(\vartheta, g) \pi(d\vartheta)$$

называется **Байесовским риском**  $g$  относительно  $\pi$ . Оценка  $g^* \in \mathcal{K}$  называется **Байесовской оценкой**, если она минимизирует Байесовский риск по всем возможным оценочным функциям:

$$R(\pi, g^*) = \inf_{g \in \mathcal{K}} R(\pi, g).$$

**Замечание 3.2.**

- (i) В байесовской интерпретации параметр  $\vartheta$  является случайным, а точнее - реализацией случайной величины  $\theta : (\Omega, \mathcal{A}) \rightarrow (\Theta, \mathcal{A}_\Theta)$  с распределением  $\pi$ .

$$\begin{array}{c} (\Omega, \mathcal{A}, \mathbb{P}) \xrightarrow{X} (\mathcal{X}, \mathcal{B}) \\ \theta \downarrow \\ (\Theta, \mathcal{A}_\Theta) \end{array}$$

- (ii) Функция  $R(\pi, g)$  играет роль среднего значения по всем функциям риска, где возможные значения  $\theta$  имеют вес соответственно их вероятностям. Распределение  $\pi$  может быть интерпретировано как априорное знание статистика о неизвестном параметре.

**Предположение 3.3.** В дальнейшем пусть  $(\Theta, \mathcal{A}_\Theta) = (\mathbb{R}^l, \mathcal{B}^l)$  и  $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}^n)$ . Также мы обозначаем  $Q^{X, \theta}$  в качестве распределения  $(X, \theta)$  на  $(\mathcal{X} \times \Theta, \mathcal{B} \otimes \mathcal{A}_\Theta)$  и  $P_\vartheta$  в качестве условного распределения  $X$  (при условии  $\theta = \vartheta$ ):

$$P_\vartheta = Q^{X|\theta=\vartheta}.$$

Мера  $\pi$  является маргинальным распределением  $\theta$  под  $Q^{X, \theta}$ . Для совместного распределения  $(X, \theta)$  правило итерированного ожидания (Следствие 1.15) дает:

$$Q^{X, \theta}(A) = \int_{\Theta} \int_{\mathcal{X}} 1_A(x, \vartheta) P_\vartheta(dx) \pi(d\vartheta).$$

Следует также заметить, что существует **апостериорное распределение**  $Q^{\theta|X=x}$  случайной величины  $\theta$  при условии  $X = x$ .

**Замечание 3.4.** В Байесовской статистике значения  $\pi$  и  $P_\vartheta$  интерпретируются следующим образом: перед экспериментом  $\pi = Q^\theta$  – предполагаемое статистиком распределение параметра  $\vartheta$ . После наблюдения  $X(\omega) = x$  информация о  $\theta$  изменяется с  $\pi$  на  $Q^{\theta|X=x}$ . Для функции риска от  $\gamma(\vartheta)$  используются следующие представления:

$$\begin{aligned} R(\pi, g) &= \int_{\Theta} R(\vartheta, g) \pi(d\vartheta) = \int_{\Theta} \int_{\mathcal{X}} L(\gamma(\vartheta), g(x)) P_\vartheta(dx) \pi(d\vartheta) \\ &= \int_{\Theta \times \mathcal{X}} L(\gamma(\vartheta), g(x)) Q^{X, \theta}(dx, d\vartheta) \\ &= \int_{\mathcal{X}} \int_{\Theta} L(\gamma(\vartheta), g(x)) Q^{\theta|X=x}(d\vartheta) Q^X(dx) \\ &= \int_{\mathcal{X}} R_\pi^x(g) Q^X(dx). \end{aligned}$$

Величина

$$R_\pi^x(g) := \int_{\Theta} L(\gamma(\vartheta), g(x)) Q^{\theta|X=x}(d\vartheta)$$

называется **апостериорным риском**  $g$  при данном  $X=x$ .

**Теорема 3.5.**

(i) Оценка  $g^*$  является Байесовской оценкой для  $\vartheta$  тогда и только тогда, когда  $g^*$  доставляет минимум апостериорному риску  $R_\pi^x(g)$   $Q^X$ -п.н.

$$R_\pi^x(g^*) = \inf_{g \in \mathcal{K}} R_\pi^x(g) = \inf_{a \in \Theta} \int L(\vartheta, a) Q^{\theta|X=x}(d\vartheta).$$

(ii) Пусть  $\Theta \subset \mathbb{R}$ ,  $L(\vartheta, a) = (\vartheta - a)^2$  и  $\int \vartheta^2 Q^{\theta|X=x}(d\vartheta) < \infty$   $Q^X$ -п.н. Тогда Байесовская оценка для  $\vartheta$ :

$$g^*(x) = \mathbb{E}[\theta|X=x] = \int_{\Theta} \vartheta Q^{\theta|X=x}(d\vartheta)$$

ДОКАЗАТЕЛЬСТВО:

(i) В соответствии с Замечанием 3.4,  $R(\pi, g)$  достигает минимума по  $g$  тогда и только тогда, когда  $R_\pi^x(g)$  минимальна относительно  $g$   $Q^X$ -п.н. В частности, мы минимизируем по всем возможным оценкам. Если существует  $g^*$ , то эта оценка должна минимизировать  $R_\pi^x(g)$   $Q^X$ -п.н., в противном случае возможно построить оценку, доставляющую меньшее значение риска.

(ii) В соответствии с Теоремой 1.25, величина  $\int_{\Theta} (\vartheta - a)^2 Q^{\theta|X=x}(d\vartheta)$  минимальна, если  $a = \mathbb{E}[\theta|X=x]$ .

**Замечание 3.6.** Если  $P_\vartheta \ll \mu$  с  $\mu$ -плотностью  $f(x|\vartheta)$  и также  $\pi \ll \nu$  с  $\nu$ -плотностью  $h(\vartheta)$ , то совместное распределение  $(X, \theta)$  удовлетворяет:

$$Q^{X, \theta} \ll \mu \otimes \nu$$

с плотностью  $f(x|\vartheta)h(\vartheta)$ . Также, апостериорное распределение  $Q^{\theta|X=x}$  обладает  $\nu$ -плотностью:

$$f(\vartheta|x) = \frac{f(x|\vartheta)h(\vartheta)}{\int_{\Theta} f(x|\vartheta)h(\vartheta)\nu(d\vartheta)}$$

если знаменатель положительный. Апостериорный и Байесовский риски соответственно:

$$R_\pi^x(g) = \frac{\int_{\Theta} L(\vartheta, g(x)) f(x|\vartheta) h(\vartheta) \nu(d\vartheta)}{\int_{\Theta} f(x|\vartheta) h(\vartheta) \nu(d\vartheta)},$$

$$R(\pi, g) = \int_{\mathcal{X}} \int_{\Theta} L(\vartheta, g(x)) f(x|\vartheta) h(\vartheta) \nu(d\vartheta) \mu(dx).$$

**Пример 3.7.**

Рассмотрим пример с оцениванием параметра биномиального распределения. Пусть  $\Theta = (0, 1)$ ,  $\mathcal{X} = \{0, \dots, n\}$ :

$$P_{\vartheta}(X = x) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x}.$$

Функция потерь:  $L(x, y) = (x - y)^2$

- UMVU-оценка для  $\vartheta$ :  $g(x) = \frac{x}{n}$  (экспоненциальное семейство)

$$\text{Var}_{\vartheta}(g(X)) = \frac{\vartheta(1 - \vartheta)}{n}.$$

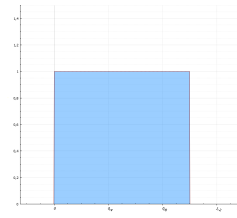
- Пусть  $\pi \sim \mathcal{U}(0, 1)$ .

Тогда

$$f(x | \vartheta) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x} 1_{\{0, \dots, n\}}(x)$$

и априорная функция плотности распределения параметра:

$$h(\vartheta) = 1_{(0,1)}(\vartheta).$$



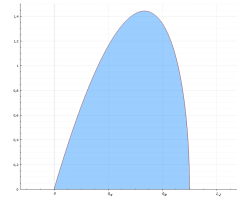
$h(\vartheta)$

Апостериорная плотность вероятности:

$$f(\vartheta | x) = \frac{\vartheta^x (1 - \vartheta)^{n-x} 1_{(0,1)}(\vartheta)}{B(x+1, n-x+1)},$$

где в знаменателе бета-функция:

$$B(a, b) = \int_0^1 \vartheta^{a-1} (1 - \vartheta)^{b-1} d\vartheta.$$



$f(\vartheta | x)$

Тогда Байесовская оценка:

$$g^*(x) = \mathbb{E}[\vartheta | X = x] = \int_0^1 \frac{\vartheta^{x+1} (1 - \vartheta)^{n-x}}{B(x+1, n-x+1)} d\vartheta = \frac{B(x+2, n-x+1)}{B(x+1, n-x+1)} = \frac{x+1}{n+2},$$

и Байесовский риск:

$$\begin{aligned} R(\pi, g^*) &= \int_0^1 R(\vartheta, g^*) d\vartheta = \int_0^1 \mathbb{E}_{\vartheta} \left[ \left( \frac{X+1}{n+2} - \vartheta \right)^2 \right] d\vartheta \\ &= \frac{1}{(n+2)^2} \int_0^1 (n\vartheta - n\vartheta^2 + 1 - 4\vartheta + 4\vartheta^2) d\vartheta = \frac{1}{6(n+2)}. \end{aligned}$$

**Пример 3.8.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim P_{\mu}^1 = \mathcal{N}(\mu, \sigma^2)$  с заранее известным параметром  $\sigma^2$ . Априорное распределение  $\mu$ :

$$h(\mu) = \frac{1}{\sqrt{2\pi\tau^2}} \exp \left\{ -\frac{(\mu - \mu_0)^2}{2\tau^2} \right\}.$$

Используя плотность распределения  $X$

$$f(x | \mu) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2 \right\},$$

получаем апостериорное распределение:

$$Q^{\mu | X=x} \sim \mathcal{N} \left( g_{\mu_0, \tau^2}(x), \left( \frac{n}{\sigma^2} + \frac{1}{\tau^2} \right)^{-1} \right),$$

где

$$g_{\mu_0, \tau^2}(x) = \left(1 + \frac{\sigma^2}{n\tau^2}\right)^{-1} \bar{x}_n + \left(\frac{n\tau^2}{\sigma^2} + 1\right)^{-1} \mu_0.$$

Для квадратичного риска функция  $g_{\mu_0, \tau^2}(x)$  – Байесовская оценка. Интерпретация: при большом значении  $\tau^2$  (мало априорной информации) оценка  $g_{\mu_0, \tau^2}(x) \approx \bar{x}_n$ , иначе  $g_{\mu_0, \tau^2}(x) \approx \mu_0$ .

**Определение 3.9.** Пусть  $g$  – оценка  $\gamma(\vartheta)$ . Тогда:

$$R^*(g) = \sup_{\vartheta \in \Theta} R(\vartheta, g)$$

называется **максимальным риском**  $g$  и

$$R^*(g^*) = \inf_{g \in \mathcal{K}} R^*(g)$$

называется **минимаксным риском**, а соответствующая оценка  $g^*$  – **минимаксной**.

**Замечание 3.10.**

- (i) Использование минимаксной оценки нацелено на защиту от больших потерь.
- (ii) Пусть  $\mathcal{M} = \{\pi \mid \pi - \text{вероятностная мера на } \mathcal{A}_\Theta\}$ . Тогда несложно заметить, что:

$$R^*(g) = \sup_{\pi \in \mathcal{M}} R(\pi, g).$$

**Определение 3.11.** Априорное распределение  $\pi^*$  на  $\mathcal{A}_\Theta$  называется **наименее благоприятным априорным**, если

$$\inf_{g \in \mathcal{K}} R(\pi^*, g) \geq \inf_{g \in \mathcal{K}} R(\pi, g) \quad \forall \pi \in \mathcal{M}.$$

**Теорема 3.12.**

- (i) Если  $g_\pi$  – Байесовская оценка относительно  $\pi$  и

$$R(\pi, g_\pi) = \sup_{\vartheta \in \Theta} R(\vartheta, g_\pi), \tag{3.1}$$

то  $g_\pi$  – минимаксная оценка.

- (ii) Если  $g_\pi$  – единственная Байесовская оценка относительно  $\pi$ , удовлетворяющая равенству (3.1), то  $g_\pi$  – единственная минимаксная оценка.
- (iii) Если равенство (3.1) выполняется, то  $\pi$  – наименее благоприятное априорное распределение.

ДОКАЗАТЕЛЬСТВО:

- (i) Для любой оценки  $g \in \mathcal{K}$ :

$$\sup_{\vartheta \in \Theta} R(\vartheta, g) \geq \int_{\Theta} R(\vartheta, g) \pi(d\vartheta) \geq \int_{\Theta} R(\vartheta, g_\pi) \pi(d\vartheta) = R(\pi, g_\pi) = \sup_{\vartheta \in \Theta} R(\vartheta, g_\pi).$$

- (ii) Если  $g_\pi$  – единственная Байесовская оценка для  $\pi$ , то

$$\int_{\Theta} R(\vartheta, g) \pi(d\vartheta) > \int_{\Theta} R(\vartheta, g_\pi) \pi(d\vartheta) \quad \forall g \neq g_\pi.$$

Тогда  $\sup_{\vartheta \in \Theta} R(\vartheta, g) > \sup_{\vartheta \in \Theta} R(\vartheta, g_\pi)$  и  $g_\pi$  единственная минимаксная оценка.

- (iii) Для любого распределения  $\mu \in \mathcal{M}$ :

$$\inf_{g \in \mathcal{K}} \int_{\Theta} R(\vartheta, g) \mu(d\vartheta) \leq \int_{\Theta} R(\vartheta, g_\pi) \mu(d\vartheta) \leq \sup_{\vartheta \in \Theta} R(\vartheta, g_\pi) = R(\pi, g_\pi) = \inf_{g \in \mathcal{K}} \int_{\Theta} R(\vartheta, g) \pi(d\vartheta).$$



**Замечание 3.13.** Иногда функция риска Байесовской оценки  $g_\pi$  является постоянной:

$$R(\vartheta, g_\pi) = c \quad \forall \vartheta \in \Theta.$$

Тогда

$$\sup_{\vartheta \in \Theta} R(\vartheta, g_\pi) = c = \int_{\Theta} R(\vartheta, g_\pi) \pi(d\vartheta) = R(\pi, g_\pi),$$

равенство (3.1) выполняется,  $g_\pi$  минимаксная оценка и  $\pi$  – наименее благоприятное априорное распределение.

**Пример 3.14.** Пусть  $\Theta = (0, 1)$ ,  $\mathcal{X} = \{0, \dots, n\}$  и

$$P_\vartheta(X = x) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x}.$$

Мы снова используем квадратичный риск и выбираем бета-распределение в качестве априорного:

$$h(\vartheta) = \frac{\vartheta^{a-1} (1 - \vartheta)^{b-1} 1_{[0,1]}(\vartheta)}{B(a, b)}.$$

Апостериорное распределение  $Q^{\vartheta|X=x} \sim B(x + a, n - x + b)$  с плотностью:

$$f(\vartheta|x) = \frac{\vartheta^{x+a-1} (1 - \vartheta)^{n-x+b-1} 1_{[0,1]}(\vartheta)}{B(x + a, n - x + b)}.$$

Несложно доказать, что для случайной величины  $Z$  с бета-распределением  $B(p, q)$

$$\mathbb{E}[Z] = \frac{p}{p+q} \quad \text{и} \quad \text{Var}(Z) = \frac{pq}{(p+q)^2(p+q+1)}.$$

Используя Теорему 3.5, получаем Байесовскую оценку для  $\vartheta$ :

$$g_{a,b}(x) = \frac{x + a}{n + a + b}.$$

Соответствующий ей риск:

$$R(\vartheta, g_{a,b}) = \mathbb{E}[(g_{a,b}(X) - \vartheta)^2] = \frac{\vartheta^2(-n + (a+b)^2 + \vartheta(n - 2a(a+b))) + a^2}{(n + a + b)^2}.$$

Если выбрать постоянные  $a^* = b^* = \sqrt{n}/2$ , то риск будет равен:

$$R(\vartheta, g_{a^*,b^*}) = \frac{1}{4(\sqrt{n} + 1)^2}.$$

Такой риск не зависит от  $\vartheta$ , а значит оценка  $g_{a^*,b^*}(x) = \frac{x + \sqrt{n}/2}{n + \sqrt{n}}$  является минимаксной и  $B(a^*, b^*)$  – наименее благоприятное распределение.

**Определение 3.15.** Пусть

$$r_\pi = \inf_{g \in \mathcal{K}} R(\pi, g), \quad \pi \in \mathcal{M}.$$

Последовательность  $(\pi_m)_{m \in \mathbb{N}}$  в  $\mathcal{M}$  называется **наименее благоприятной последовательностью априорных распределений**, если

$$(i) \quad \lim_{m \rightarrow \infty} r_{\pi_m} = r,$$

$$(ii) \quad \forall \pi \in \mathcal{M} \quad r_\pi \leq r.$$

**Теорема 3.16.** Пусть  $(\pi_m)$  в  $\mathcal{M}$  последовательность, такая что  $r_{\pi_m} \rightarrow r \in \mathbb{R}$ . Также пусть существует такая оценка  $g^* \in \mathcal{K}$ , что:

$$\sup_{\vartheta \in \Theta} R(\vartheta, g^*) = r.$$

Тогда:

- (i)  $g^*$  – минимаксная оценка,
- (ii)  $(\pi_m)$  – наименее благоприятная последовательность априорных распределений.

ДОКАЗАТЕЛЬСТВО:

- (i) Для любой оценки  $g \in \mathcal{K}$

$$\sup_{\vartheta \in \Theta} R(\vartheta, g) \geq \int_{\Theta} R(\vartheta, g) \pi_m(d\vartheta) \geq r_{\pi_m} \rightarrow r = \sup_{\vartheta \in \Theta} R(\vartheta, g^*).$$

- (ii) Для любого распределения  $\pi \in \mathcal{M}$

$$r_{\pi} \leq R(\pi, g^*) = \int_{\Theta} R(\vartheta, g^*) \pi(d\vartheta) \leq \sup_{\vartheta \in \Theta} R(\vartheta, g^*) = r.$$

**Пример 3.17.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$  с известным параметром  $\sigma^2$ . В качестве априорного распределения выбираем:

$$h_m(\mu) = \frac{1}{\sqrt{2\pi m}} \exp \left\{ -\frac{(\mu - \mu_0)^2}{2m} \right\}.$$

Байесовская оценка:

$$g_m(x) = \left(1 + \frac{\sigma^2}{nm}\right)^{-1} \bar{x}_n + \left(\frac{nm}{\sigma^2} + 1\right)^{-1} \mu_0.$$

Для любого значения  $\mu \in \mathbb{R}$ :

$$\begin{aligned} R(\mu, g_m) &= \mathbb{E}_{\mu}[(g_m(X) - \mu)^2] \\ &= \mathbb{E}_{\mu} \left[ \left( \left(1 + \frac{\sigma^2}{nm}\right)^{-1} (\bar{x}_n - \mu) + \left(\frac{nm}{\sigma^2} + 1\right)^{-1} (\mu_0 - \mu) \right)^2 \right] \\ &= \left(1 + \frac{\sigma^2}{nm}\right)^{-2} \frac{\sigma^2}{n} + \left(1 + \frac{nm}{\sigma^2}\right)^{-2} (\mu_0 - \mu)^2 \xrightarrow{m \rightarrow \infty} \frac{\sigma^2}{n} \end{aligned}$$

Так как риск ограничен сверху:

$$R(\mu, g_m) \leq \frac{\sigma^2}{n} + (\mu - \mu_0)^2,$$

то по теореме Лебега о мажорируемой сходимости<sup>1</sup>:

$$r_{\pi_m} = R(\pi_m, g_m) = \int_{\mathbb{R}} R(\mu, g_m) \pi_m(d\mu) \rightarrow \frac{\sigma^2}{n}.$$

Очевидно,  $g^*(x) = \bar{x}_n$  удовлетворяет равенству

$$R(\mu, g^*) = \mathbb{E}_{\mu}[(\bar{X}_n - \mu)^2] = \frac{\sigma^2}{n},$$

следовательно по Теореме 3.16  $g^*$  – минимаксная оценка и  $\pi_m$  – наименее благоприятная последовательность априорных распределений.

<sup>1</sup>Пусть фиксировано измеримое пространство  $(X, \mathcal{F}, \mu)$ . Предположим, что  $\{f_n\}_{n=1}^{\infty}$  и  $f$  – измеримые функции на  $X$ , причем  $f_n(x) \rightarrow f(x)$  почти всюду. Тогда если существует определённая на том же пространстве интегрируемая функция  $g$ , такая что

$$|f_n(x)| \leq g(x) \quad \forall n \in \mathbb{N}$$

почти всюду, то  $f_n$  и  $f$  интегрируемы и

$$\lim_{n \rightarrow \infty} \int_X f_n(x) \mu(dx) = \int_X f(x) \mu(dx).$$

## Упражнения

**3.1.** Пусть  $X \sim Po(\lambda)$  (распределение Пуассона). Функция потерь:

$$L(\lambda, a) = (\lambda - a)^2 / \lambda.$$

Докажите, что оценка  $g(X) = X$  является минимаксной для параметра  $\lambda$ . Используйте следующие шаги:

- (i) Выберите гамма-распределение  $\pi_{\alpha, \beta} = \Gamma(\alpha, \beta)$  в качестве априорного. Найдите апостериорное распределение.
- (ii) Рассчитайте апостериорный риск:

$$R_{\pi_{\alpha, \beta}}^x(a) := \int_{\Theta} L(\lambda, a) Q^{\lambda|X=x}(d\lambda).$$

для  $a \in \mathbb{R}$  и  $\alpha > 1$ .

*Подсказка:* если  $\Lambda \sim \Gamma(\alpha, \beta)$ , то

$$\mathbb{E}[\Lambda] = \frac{\alpha}{\beta}, \quad \mathbb{E}[1/\Lambda] = \frac{\beta}{\alpha - 1}.$$

- (iii) Найдите значение  $a^*$ , доставляющее минимум апостериорному риску  $R_{\pi_{\alpha, \beta}}^x(a)$ . Рассчитайте соответствующий минимум  $R_{\pi_{\alpha, \beta}}^x(a^*)$ .
- (iv) Рассчитайте  $R(\lambda, X)$  и тем самым завершите доказательство.

**3.2.** Докажите следующие утверждения:

- (i) Если  $g^*$  – допустимая оценка с постоянным риском, то  $g^*$  – минимаксная.
- (ii) Если  $g^*$  – Байесовская оценка для априорного распределения  $\pi$  и единственная в том смысле, что для любой другой Байесовской оценки  $\tilde{g}$

$$R(\vartheta, g^*) = R(\vartheta, \tilde{g}) \quad \forall \vartheta \in \Theta,$$

то  $g^*$  – допустимая.

**3.3.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$ ,  $X = (X_1, \dots, X_n)^T$  и функция потерь:

$$L(\sigma^2, d) = (d/\sigma^2 - 1)^2.$$

Предположим, что  $\mu = 0$ . Докажите, что

$$g(X) = \frac{1}{n+2} \sum_{i=1}^n X_i^2$$

является минимаксной оценкой  $\sigma^2$ .

*Подсказка:* произведите замену  $\lambda = 1/\sigma^2$  и возьмите для этого параметра априорное распределение  $\pi_{\alpha, \beta} = \Gamma(p, b)$ . Заметьте, что для апостериорного риска  $g$ :

$$R_{\pi_{\alpha, \beta}}^x(\sigma^2, g) = \int_{\Theta} L(\sigma^2, g(x)) Q^{\theta|X=x} d(1/\sigma^2) = \int_{\Theta} (g(x)\lambda - 1)^2 Q^{\theta|X=x} d(\lambda).$$



## Глава 4

# Достаточность и полнота

Теперь, пусть  $(\mathcal{X}, \mathcal{B}, P)$ , где  $P \in \mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$ , – вероятностное пространство и  $(\mathcal{T}, \mathcal{D})$  – измеримое пространство.

**Определение 4.1.**  $\mathcal{B}$ - $\mathcal{D}$ -измеримая функция  $T : \mathcal{X} \rightarrow \mathcal{T}$  называется *статистикой*.

**Пример 4.2.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \text{Bin}(1, \vartheta)$  с совместным распределением:

$$f(x, \vartheta) = \vartheta^{\sum_{i=1}^n x_i} (1 - \vartheta)^{n - \sum_{i=1}^n x_i} = \vartheta^{T(x)} (1 - \vartheta)^{n - T(x)},$$

где  $T(x) = \sum_{i=1}^n x_i$  – статистика. Если мы выбираем  $u_1, \dots, u_n \in \{0, 1\}$ , то

$$P_\vartheta(X_i = u_i \ \forall i = 1, \dots, n \mid T(X) = k) = \begin{cases} \frac{1}{\binom{n}{k}}, & \text{если } \sum_{i=1}^n u_i = k \\ 0, & \text{иначе} \end{cases}$$

Мы видим, что в данном примере, зная значение  $T(X)$ , никакой дополнительной информации о  $\vartheta$  не может быть получено из информации о векторе  $X = X_1, \dots, X_n$ .

**Определение 4.3.**

(i)  $\sigma$ -алгебра  $\mathcal{C} \subset \mathcal{B}$  называется *достаточной* для  $\vartheta$ , если

$$k_B = P_\vartheta(B \mid \mathcal{C}) \quad \forall \vartheta \in \Theta, \ \forall B \in \mathcal{B}.$$

Это означает, что  $\mathcal{C}$ -условные вероятности не зависят от  $\vartheta$ .

(ii) Статистика  $T : \mathcal{X} \rightarrow \mathcal{T}$  называется *достаточной* для  $\vartheta$ , если  $\sigma(T)$  достаточна для  $\vartheta$ .

**Замечание 4.4.**

(i) Из Леммы о факторизации (Теорема 1.18) мы знаем, что  $T$  достаточная для  $\vartheta$  тогда и только тогда, когда  $\forall B \in \mathcal{B}$  существует функция  $h_B : \mathcal{T} \rightarrow \mathbb{R}$ , такая что:

$$h_B(t) = P_\vartheta(B \mid T = t).$$

(ii) В Примере 4.2 статистика  $T(x) = \sum_{i=1}^n x_i$  достаточная для  $\vartheta$ .

(iii) Пусть  $g : \mathcal{X} \rightarrow \mathbb{R}$  из  $L^1(\mathcal{P}) = \bigcap_{\vartheta \in \Theta} L^1(P_\vartheta)$ .

(a) Если  $\mathcal{C}$  достаточная для  $\vartheta$ , то существует функция

$$k = \mathbb{E}_\vartheta[g \mid \mathcal{C}],$$

не зависящая от  $\vartheta$ .

(b) Если  $T : \mathcal{X} \rightarrow \mathcal{T}$  достаточная для  $\vartheta$ , то существует функция

$$h(t) = \mathbb{E}_\vartheta[g \mid T = t],$$

не зависящая от  $\vartheta$ .

**Пример 4.5.** Пусть  $Q$  – конечная группа отображений  $\pi : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{X}, \mathcal{B})$ . Система

$$\mathcal{C} = \mathcal{C}(Q) = \{B \in \mathcal{B} \mid \pi(B) = B \quad \forall \pi \in Q\}$$

называется  $\sigma$ -алгеброй  $Q$ -инвариантных множеств. Если  $\mathcal{P}$  инвариантна относительно  $Q$ , то есть:

$$P_{\vartheta}^{\pi}(B) = P_{\vartheta}(\pi^{-1}(B)) = P_{\vartheta}(B) \quad \forall B \in \mathcal{B}, \quad \forall \pi \in Q, \quad \forall \vartheta \in \Theta$$

и если  $g \in L^1(\mathcal{P})$  и  $q = |Q|$ , то

$$k(x) = \frac{1}{q} \sum_{\pi \in Q} g(\pi(x))$$

является версией  $\mathbb{E}_{\vartheta}[g|\mathcal{C}]$ , независимой от  $\vartheta$ . Как следствие,  $\mathcal{C}$  – достаточная.

**ДОКАЗАТЕЛЬСТВО:** Так как  $Q$  является группой, то  $k(x) = k(\pi(x)) \quad \forall \pi \in Q$ . Следовательно,

$$k^{-1}(B) = (\pi)^{-1}(k^{-1}(B)) \quad \forall B \in \mathcal{B},$$

иными словами,  $k$  –  $\mathcal{C}$ -измерима. Пусть  $C \in \mathcal{C}$ , тогда:

$$\begin{aligned} \int_C k(x) P_{\vartheta}(dx) &= \frac{1}{q} \sum_{\pi \in Q} \int_C g(\pi(x)) P_{\vartheta}(dx) = \frac{1}{q} \sum_{\pi \in Q} \int_{\pi(C)} g(x) P_{\vartheta}^{\pi}(dx) \\ &= \frac{1}{q} \sum_{\pi \in Q} \int_{\pi(C)} g(x) P_{\vartheta}(dx) = \frac{1}{q} \sum_{\pi \in Q} \int_C g(x) P_{\vartheta}(dx) \\ &= \int_C g(x) P_{\vartheta}(dx). \end{aligned}$$

**Пример 4.6.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim F$  и

$$\mathcal{P} = \{P \sim F^n \mid F - \text{функция распределения}\}, \quad F^n(x) = \prod_{i=1}^n F(x_i).$$

Также  $\mathcal{X} = \mathbb{R}^n$  и  $\mathcal{B} = \mathcal{B}^n$ . Пусть

$$Q = S_n = \{\pi : \mathbb{R}^n \rightarrow \mathbb{R}^n \mid \pi - \text{перестановка}\},$$

тогда  $P^{\pi} = P \quad \forall \pi \in S_n$  и таким образом  $\mathcal{P}$  – инвариантна относительно  $S_n$ . Следовательно,  $\mathcal{C} = \mathcal{C}(S_n)$  – достаточная.

**Определение 4.7.** Пусть  $\mathbb{R}_{\leq}^n = \{X \in \mathbb{R}^n \mid X_1 \leq \dots \leq X_n\}$  и  $X_{(j)}$  –  $j$ -ое наименьшее число среди  $X_1, \dots, X_n$ . Тогда статистика

$$T : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{R}_{\leq}^n \\ X & \mapsto X_{(\cdot)} \end{cases}$$

называется **порядковой статистикой**  $(X_1, \dots, X_n)^T$ .

**Замечание 4.8.**

(i) По определению  $\sigma(T) = \mathcal{C}(S_n)$ , поэтому  $T$  – достаточная статистика. Другими словами, всегда достаточно хранить упорядоченный вектор наблюдений.

(ii) Статистика

$$\tilde{T} : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{R}^n \\ X & \mapsto (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2, \dots, \sum_{i=1}^n X_i^n)^T \end{cases}$$

также достаточная, потому что можно показать, что  $\sigma(T) = \sigma(\tilde{T})$ .

**Определение 4.9.** Пусть  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  – семейство вероятностных мер. Тогда мера  $\nu$  называется *эквивалентной*  $\mathcal{P}$ , если

$$\nu(N) = 0 \iff P_\vartheta(N) = 0 \quad \forall \vartheta \in \Theta.$$

**Теорема 4.10 (Теорема Халмоса-Савая).** Пусть  $\mu$  –  $\sigma$ -конечная мера,  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  и  $P_\vartheta \ll \mu \quad \forall \vartheta \in \Theta$ .

(i) Существует мера  $\nu$ , эквивалентная  $\mathcal{P}$ , вида

$$\nu = \sum_{i=1}^{\infty} c_i P_{\vartheta_i}, \text{ где } c_i \geq 0 \quad \forall i \text{ и } \sum_{i=1}^{\infty} c_i = 1. \quad (4.1)$$

(ii)  $\sigma$ -алгебра  $\mathcal{C} \subset \mathcal{B}$  достаточная для  $\vartheta$  тогда и только тогда, когда существует  $\mathcal{C}$ -измеримая функция  $\frac{dP_\vartheta}{d\nu}$  для любого параметра  $\vartheta$ .

(iii) Статистика  $T$  является достаточной для  $\vartheta$  тогда и только тогда, когда для любого параметра  $\vartheta$  существует функция  $g_\vartheta$ , такая что:

$$\frac{dP_\vartheta}{d\nu}(x) = g_\vartheta(T(x)).$$

ДОКАЗАТЕЛЬСТВО:

(i) Мы докажем утверждение только для конечной меры  $\mu$ . Пусть

$$\mathcal{J} = \{\nu \mid \nu - \text{мера вида (4.1)}\}$$

и

$$q = \frac{dQ}{d\mu}, \quad Q \in \mathcal{J}.$$

Достаточно показать, что существует мера  $Q_0 \in \mathcal{J}$ , такая что:

$$Q_0(A) = 0 \implies Q(A) = 0 \quad \forall Q \in \mathcal{J}. \quad (4.2)$$

Тогда  $P_\vartheta(A) = 0 \quad \forall \vartheta$ , так как  $P_\vartheta \in \mathcal{J}$ . Верно и обратное, если  $P_\vartheta(A) = 0 \quad \forall \vartheta$ , то  $Q_0(A) = 0$ , так как  $Q_0 \in \mathcal{J}$ . Рассмотрим множество

$$\mathcal{C} := \{B \in \mathcal{B} \mid \exists Q \in \mathcal{J}, \text{ такая что } Q(B) > 0 \text{ и } q(x) > 0 \text{ для } \mu\text{-почти всех } x \in B\}$$

и зададим  $\sup_{B \in \mathcal{C}} \mu(B) = r$ . Тогда существуют  $B_i \in \mathcal{C}$ , такие что  $\mu(B_i) \rightarrow r$ . Пусть

$$B_0 = \bigcup_{i \in \mathbb{N}} B_i,$$

тогда вследствие непрерывности  $\mu$  мы получаем  $\mu(B_0) = r$ . Пусть также  $Q_i$  – меры  $Q \in \mathcal{J}$ , соответствующие  $B_i$ , и

$$Q_0 = \sum_{i=1}^n c_i Q_i \in \mathcal{J},$$

где  $c_i > 0$  и  $\sum_{i=1}^n c_i = 1$ . Тогда  $\mu$ -плотность  $Q_0$ :

$$\frac{dQ_0}{d\mu}(x) = q_0(x) = \sum_{i=1}^n c_i q_i(x).$$

Очевидно, что  $q_0(x) > 0$  для  $\mu$ -почти всех  $x \in B_0$ . Следовательно  $Q_0(B_0) > 0$  и  $B_0 \in \mathcal{C}$ . Докажем (4.2). Допустим,  $Q_0(A) = 0$  и пусть  $Q \in \mathcal{J}$  – произвольная мера. Пусть  $q$  – плотность  $Q$  и

$$B = \{x \mid q(x) > 0\}.$$

Тогда

$$0 = Q_0(A \cap B_0) = \int_{A \cap B_0} q_0(x) \mu(dx).$$

Так как на  $q_0 > 0$  на  $B_0$ , мы получаем

$$\mu(A \cap B_0) = 0$$

и, как следствие,

$$Q(A \cap B_0) = 0 \quad \forall Q \in \mathcal{J}.$$

Кроме того, в соответствии с определением  $B$ :

$$Q(A \cap B_0^c \cap B^c) = 0.$$

Если  $Q(A \cap B_0^c \cap B) > 0$ , то  $B_0$  и  $A \cap B_0^c \cap B$  являются элементами  $\mathcal{C}$ . Тогда легко увидеть, что объединение  $B_0$  и  $A \cap B_0^c \cap B$  также принадлежит  $\mathcal{C}$ . Тогда

$$\mu(B_0 \cup (A \cap B_0^c \cap B)) = \mu(B_0) + \mu(A \cap B_0^c \cap B) > \mu(B_0),$$

что противоречит максимальнойности  $B_0$  в  $\mathcal{C}$ . Следовательно,

$$Q(A \cap B_0^c \cap B) = 0 \implies Q(A) = 0.$$

- (ii) " $\implies$ " Допустим, что  $\mathcal{C}$  – достаточная. Тогда  $\forall B \in \mathcal{B}$  существует  $\mathcal{C}$ -измеримая функция  $k_B$ , независимая от  $\vartheta$ , такая что

$$\int_C k_B dP_\vartheta = \int_C 1_B dP_\vartheta \quad \forall C \in \mathcal{C} \quad \forall \vartheta \in \Theta.$$

Из (i) следует, что

$$\int_C k_B d\nu = \int_C 1_B d\nu \quad \forall C \in \mathcal{C}.$$

Следовательно,  $k_B = \mathbb{E}_\nu[1_B | \mathcal{C}]$  и

$$P_\vartheta(B) = \int_{\mathcal{X}} 1_B dP_\vartheta = \int_{\mathcal{X}} k_B dP_\vartheta = \int_{\mathcal{X}} \mathbb{E}_\nu[1_B | \mathcal{C}] dP_\vartheta.$$

Пусть теперь

$$f_\vartheta^{\mathcal{C}} = \frac{dP_\vartheta^{\mathcal{C}}}{d\nu}$$

–  $\nu$ -плотность  $P_\vartheta$ , ограниченная  $\sigma$ -алгеброй  $\mathcal{C}$ . Тогда вследствие  $\mathcal{C}$ -измеримости  $f_\vartheta^{\mathcal{C}}$  (Теорема 1.3) мы получаем по Теореме 1.12:

$$P_\vartheta(B) = \int_{\mathcal{X}} \mathbb{E}_\nu[1_B | \mathcal{C}] f_\vartheta^{\mathcal{C}} d\nu = \int_{\mathcal{X}} \mathbb{E}_\nu[1_B f_\vartheta^{\mathcal{C}} | \mathcal{C}] d\nu = \int_B f_\vartheta^{\mathcal{C}} d\nu.$$

Следовательно,  $\frac{dP_\vartheta}{d\nu} = f_\vartheta^{\mathcal{C}}$  и  $f_\vartheta^{\mathcal{C}}$   $\mathcal{C}$ -измерима.

" $\Leftarrow$ " Пусть  $B \in \mathcal{B}$  и  $k_B$  – версия  $\mathbb{E}_\nu[1_B | \mathcal{C}]$ , где  $\nu$  – мера из (i). Тогда для  $f_\vartheta = \frac{dP_\vartheta}{d\nu}$  имеет место

$$\int_C 1_B dP_\vartheta = \int_{B \cap C} f_\vartheta d\nu = \int_C 1_B f_\vartheta d\nu = \int_C \mathbb{E}_\nu[1_B f_\vartheta | \mathcal{C}] d\nu \quad \forall C \in \mathcal{C}.$$

$f_\vartheta$   $\mathcal{C}$ -измерима по предположению. Таким образом,

$$\int_C 1_B dP_\vartheta = \int_C f_\vartheta \mathbb{E}_\nu[1_B | \mathcal{C}] d\nu = \int_C f_\vartheta k_B d\nu = \int_C k_B dP_\vartheta$$

и  $\mathcal{C}$  – достаточная по определению.



(iii) Следует из (ii) и Леммы о факторизации (Теорема 1.18).

**Теорема 4.11 (Критерий Неймана).** Пусть  $\mu$  –  $\sigma$ -конечная мера и  $P \ll \mu$ . Тогда:

(i)  $\sigma$ -алгебра  $\mathcal{C}$  является достаточной для  $\vartheta \in \Theta$  тогда и только тогда, когда существуют  $\mathcal{C}$ -измеримые функции  $f_\vartheta : \mathcal{X} \rightarrow \mathbb{R}$  и  $\mathcal{B}$ -измеримая функция  $r : \mathcal{X} \rightarrow \mathbb{R}$ , такие что

$$\frac{dP_\vartheta}{d\mu} = r(x)f_\vartheta(x).$$

(ii) Статистика  $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{T}, \mathcal{D})$  является достаточной тогда и только тогда, когда существуют  $\mathcal{D}$ -измеримые функции  $g_\vartheta : \mathcal{T} \rightarrow \mathbb{R}$  и  $\mathcal{B}$ -измеримая функция  $r : \mathcal{X} \rightarrow \mathbb{R}$ , такие что

$$\frac{dP_\vartheta}{d\mu} = r(x)g_\vartheta(T(x)).$$

ДОКАЗАТЕЛЬСТВО: Мы покажем только (i), так как (ii) следует из 1.18.

" $\Rightarrow$ " Пусть  $r$  – мера из Теоремы 4.10,  $P_\vartheta \ll r \ll \mu$ . Из Теоремы 1.3 следует:

$$\frac{dP_\vartheta}{d\mu}(x) = \underbrace{\frac{dP_\vartheta}{d\nu}(x)}_{f_\vartheta(x)} \underbrace{\frac{d\nu}{d\mu}(x)}_{r(x)}$$

$f_\vartheta(x)$   $\mathcal{C}$ -измерима по Теореме 4.10(ii).

" $\Leftarrow$ "  $f_\vartheta(x)$   $\mathcal{C}$ -измерима по предположению. Возьмем  $c_i$  и  $f_{\vartheta_i}$  из  $r$  из Теоремы 4.10 и определим:

$$\tilde{f}_\vartheta(x) = \begin{cases} f_\vartheta(x) / \sum_{i=1}^{\infty} c_i f_{\vartheta_i}(x), & \exists i : f_{\vartheta_i}(x) > 0, \\ 0, & \text{иначе.} \end{cases}$$

Следовательно,  $\tilde{f}_\vartheta$   $\mathcal{C}$ -измерима и

$$\int_B \tilde{f}_\vartheta d\nu = \int_B \tilde{f}_\vartheta \sum_{i=1}^{\infty} c_i dP_{\vartheta_i} = \int_B \tilde{f}_\vartheta \sum_{i=1}^{\infty} c_i r f_{\vartheta_i} d\mu = \int_B f_\vartheta r d\mu \quad \forall B \in \mathcal{B},$$

что не что иное, как  $P_\vartheta(B)$ . Следовательно,  $\tilde{f}_\vartheta$   $\mathcal{C}$ -измеримая версия  $\frac{dP_\vartheta}{dr}$ . Следовательно, по Теореме 4.10 (ii)  $\mathcal{C}$  достаточная для  $\vartheta$ .

**Пример 4.12.**

(i) Пусть  $P$  –  $k$ -параметрическое экспоненциальное семейство с плотностями

$$\frac{dP_\vartheta}{d\mu}(x) = c(\vartheta)h(x) \exp \left\{ \sum_{j=1}^k Q_j(\vartheta)T_j(x) \right\}.$$

По Теореме 4.11  $T = (T_1, \dots, T_n)^T$  достаточная для  $\vartheta$ .

(ii) Если  $X_i \sim \mathcal{N}(\mu, \sigma^2)$ , то

$$T(X) = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$$

достаточная для  $(\mu, \sigma^2)^T$ .

(iii) Если  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{U}[0, \vartheta]$  ( $\vartheta > 0$ ), то для  $X$  плотность вероятности будет:

$$f_\vartheta(X) = \left( \frac{1}{\vartheta} \right)^n \prod_{i=1}^n 1_{[0, \vartheta]}(X_i) = \left( \frac{1}{\vartheta} \right)^n 1_{[0, \vartheta]} \left( \max_{1 \leq i \leq n} X_i \right).$$

Следовательно,  $T(X) = \max_{1 \leq i \leq n} X_i$  достаточная для  $\vartheta$  по Теореме 4.11.

**Замечание 4.13.** Пусть  $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{T}, \mathcal{D})$  и  $\tilde{T} : (\mathcal{X}, \mathcal{B}) \rightarrow (\tilde{\mathcal{T}}, \tilde{\mathcal{D}})$  – статистики и, без ограничения общности,  $\mathcal{T} = T(\mathcal{X})$  и  $\tilde{\mathcal{T}} = \tilde{T}(\mathcal{X})$ . Если  $T$  – достаточная для  $\vartheta$  и существует биекция  $b : \mathcal{T} \rightarrow \tilde{\mathcal{T}}$ , такая что  $\tilde{T} = b \circ T$  и  $b(\mathcal{D}) = \tilde{\mathcal{D}}$ , то

$$\tilde{T}^{-1}(\tilde{D}) = T^{-1}(b^{-1}(b(D))) = T^{-1}(D), \quad \tilde{D} \in \tilde{\mathcal{D}} \quad \text{и} \quad D = b^{-1}(\tilde{D}) \in \mathcal{D}.$$

Следовательно,  $\sigma(\tilde{T}) = \sigma(T)$  и  $\tilde{T}$  – достаточная.

Вкратце: биективное отображение сохраняет достаточность.

**Пример 4.14.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$ . Мы знаем из Примера 4.12, что

$$T(X) = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2 \right)$$

достаточная для  $(\mu, \sigma^2)^T$ . Из Замечания 4.13 следует, что  $(\bar{X}_n, \hat{s}_n^2)$  достаточная, если взять

$$b(x, y) = \left( \frac{x}{n}, \frac{y}{n} - \left( \frac{x}{n} \right)^2 \right), \quad \mathcal{T} = \left\{ (x, y) \mid x \in \mathbb{R}, y \geq \frac{x^2}{n} \right\} \quad \text{и} \quad \tilde{\mathcal{T}} = \mathbb{R} \times \mathbb{R}^+.$$

Множество  $\mathcal{T}$  может быть проверено с помощью неравенства Коши-Шварца.

**Теорема 4.15 (Теорема Рао-Блэквелла).** Пусть  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  – семейство распределений на  $(\mathcal{X}, \mathcal{B})$ ,  $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{T}, \mathcal{D})$  – достаточная статистика для  $\vartheta$ ,  $\gamma : \Theta \rightarrow \Gamma \subset \mathbb{R}^l$ ,  $L : \Gamma \times \Gamma \rightarrow \mathbb{R}^+$  – функция потерь, такая что  $y \mapsto L(\gamma(\vartheta), y)$  – выпуклая  $\forall \vartheta \in \Theta$ . Если  $g$  – несмещенная оценка  $\gamma(\vartheta)$  и  $\mathbb{E}[L(\gamma(\vartheta), g)] < \infty \forall \vartheta \in \Theta$ , то:

(i) Существует  $\sigma(T)$ -измеримая несмещенная оценка  $k$ , такая что

$$R(\vartheta, k) \leq R(\vartheta, g) \quad \forall \vartheta \in \Theta,$$

точнее  $k = \mathbb{E}[g|T]$ .

(ii) Если  $y \mapsto L(\gamma(\vartheta), y)$  строго выпуклая  $\forall \vartheta \in \Theta$ , то

$$R(\vartheta, k) = R(\vartheta, g) \quad \forall \vartheta \in \Theta,$$

тогда и только тогда, когда  $g = h \circ T$ , где  $h(t) = \mathbb{E}[g|T = t]$ .

**ДОКАЗАТЕЛЬСТВО:** Используем неравенство Йенсена для условного математического ожидания:

$$f(\mathbb{E}[g(X)|\mathcal{V}]) \leq \mathbb{E}[f(g(X))|\mathcal{V}] \quad \mathbb{P}^\mathcal{V}\text{-п.н.}$$

Для строго выпуклой функции  $f$  имеет место равенство в случае  $g = \mathbb{E}[g|\mathcal{V}]$ .

(i) По свойству итерированного ожидания:

$$\mathbb{E}_\vartheta[k] = \mathbb{E}_\vartheta[\mathbb{E}_\vartheta[g|T]] = \mathbb{E}_\vartheta[g] = \gamma(\vartheta),$$

то есть оценка  $k$  несмещенная  $\forall \vartheta \in \Theta$ .

$$L(\gamma(\vartheta), k) = L(\gamma(\vartheta), \mathbb{E}_\vartheta[g|T]) \leq \mathbb{E}_\vartheta[L(\gamma(\vartheta), g)|T] \quad \forall \vartheta \in \Theta.$$

Интегрируя по  $P_\vartheta$ , получаем:

$$R(\vartheta, k) = \mathbb{E}_\vartheta[L(\gamma(\vartheta), k)] \leq \mathbb{E}_\vartheta[\mathbb{E}_\vartheta[L(\gamma(\vartheta), g)|T]] = \mathbb{E}_\vartheta[L(\gamma(\vartheta), g)] = R(\vartheta, g).$$

(ii) В случае строгой выпуклости равенство имеет место тогда и только тогда, когда  $g = \mathbb{E}[g|T]$ .

**Следствие 4.16.** Пусть  $\Theta \subset \mathbb{R}$  и  $L(x, y) = (x - y)^2$ . Если  $g$  – несмещенная оценка, а  $T$  – достаточная статистика, то для оценки  $k = \mathbb{E}[g|T]$  неравенство

$$\text{Var}_{\vartheta}(k) \leq \text{Var}_{\vartheta}(g) \quad \forall \vartheta \in \Theta$$

превращается в равенство тогда и только тогда, когда  $g = \mathbb{E}[g|T]$ .

**Пример 4.17.**

(i) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{U}(0, \vartheta)$ ,  $\vartheta > 0$ . Мы знаем, что статистика

$$T(X) = X_{(n)} = \max_{1 \leq i \leq n} X_i$$

является достаточной для  $\vartheta$ . Оценка  $g(X) = \frac{2}{n} \sum_{i=1}^n X_i$  является несмещенной для  $\vartheta$ . Следовательно, оценка

$$k(X) = \mathbb{E}[g(X) | X_{(n)}] = \frac{2}{n} \sum_{i=1}^n \mathbb{E}[X_{(i)} | X_{(n)}] = \frac{2}{n} \sum_{i=1}^n \frac{i}{n} X_{(n)} = \frac{n+1}{n} X_{(n)}$$

также является несмещенной и имеет меньшую (или, по крайней мере, такую же) дисперсию. Проверим это, используя распределение  $i$ -го элемента в порядковой статистике из стандартного равномерного распределения  $\mathcal{U}(0, 1)$ :

$$\mathbb{P}(X_{(i)} \leq x) = \sum_{j=i}^n \binom{n}{j} x^j (1-x)^{n-j} = \frac{\int_0^x t^{i-1} (1-t)^{n-i} dt}{B(i, n-i+1)}.$$

Другими словами,  $X_{(i)} \sim B(i, n-i+1)$ . Зная дисперсию бета-распределения (Пример 3.14) и масштабируя на  $\vartheta$ , мы получаем

$$\text{Var}_{\vartheta}(g) = \frac{\vartheta^2}{3n} \quad \text{и} \quad \text{Var}_{\vartheta}(k) = \frac{\vartheta^2}{n(n+2)}.$$

Таким образом, для  $n \geq 2$  оценка  $k$  имеет меньшую дисперсию.

(ii) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim F$  и

$$\mathcal{P} = \{F \mid F \text{ — функция распределения}\}.$$

Допустим, мы заинтересованы в оценке  $\gamma : F \mapsto F(z)$ ,  $z \in \mathbb{R}$ . Функция

$$g(X) = 1_{\{X_1 \leq z\}}$$

является несмещенной оценкой. Статистика

$$X_{(\cdot)} = (X_{(1)}, \dots, X_{(n)})^T$$

является достаточной для  $F$  (Замечание 4.8). Заметив, что

$$\mathbb{E}[1_{\{X_n \leq z\}} | X_{(\cdot)}] = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq z\}}, \quad (4.3)$$

мы увидим, что  $\hat{F}_n(z) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq z\}}$  – несмещенная оценка  $F(z)$ , с дисперсией, не большей чем дисперсия  $g$ . Функция  $\hat{F}_n(z)$  называется **эмпирической функцией распределения**.

Чтобы увидеть, откуда берется (4.3), вспомним из Замечания 4.8, что

$$\sigma(X_{(\cdot)}) = \mathcal{C}(S_n).$$

Так как  $\hat{F}_n(z)$  инвариантна по отношению к перестановкам, она  $\sigma(X_{(\cdot)})$ -измерима. Заметим далее, что

$$\int_B 1_{\{X_1 \leq z\}} d\mathbb{P} = \int_B 1_{\{X_j \leq z\}} d\mathbb{P} \quad \forall B \in \sigma(X_{(\cdot)}) = \mathcal{C}(S_n)$$

и (4.3) следует из определения условного математического ожидания.

**Замечание 4.18.** Хорошие оценки факторизуются в общем над достаточными статистиками. Таким способом, уменьшаются данные, как в Примере 4.12, где мы храним  $(\bar{X}_n, \hat{s}_n^2(X))$ , вместо целого вектора  $(X_1, \dots, X_n)^T$ . Оптимальным будет сокращение до минимальных достаточных статистик.

**Определение 4.19.** Достаточная статистика  $T^* : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{T}, \mathcal{D})$  называется **минимальной достаточной**, если для любой другой достаточной статистики  $T$  существует функция  $h$ , такая что

$$T^* = h \circ T.$$

**Пример 4.20.** Пусть  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  – семейство эквивалентных вероятностных мер, где  $\Theta = \{\vartheta_0, \dots, \vartheta_k\}$  и  $\mu$ -плотности  $f_{\vartheta_i}, i = 0, \dots, k$ . Тогда статистика

$$T^*(x) = \left( \frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)}, \dots, \frac{f_{\vartheta_k}(x)}{f_{\vartheta_0}(x)} \right)^T$$

минимальная достаточная для  $\vartheta$ .

**ДОКАЗАТЕЛЬСТВО:** Введем обозначения:  $P_i = P_{\vartheta_i}$  и  $f_i = f_{\vartheta_i}$ . Выберем  $P_0$  в качестве доминирующей меры в критерии Неймана (Теорема 4.11). Тогда

$$\frac{dP_i}{dP_0} = \frac{dP_i/d\mu}{dP_0/d\mu} = \frac{f_i}{f_0} = \pi_i \circ T^*, \quad i = 1, \dots, k$$

и

$$P_i(A) = \int_A \frac{dP_i}{dP_0} dP_0 = \int_A \frac{dP_i}{dP_0} \frac{dP_0}{d\mu} d\mu = \int_A \frac{dP_i}{d\mu} d\mu \quad i = 1, \dots, k,$$

где  $\pi_i$  обозначает проекцию  $i$ -й компоненты. Также, так как  $\frac{dP_0}{dP_0} = 1 = k \circ T^*$  для  $k \equiv 1$ , то  $T^*$  – достаточная по критерию Неймана. Допустим теперь, что  $T$  – другая достаточная статистика. Тогда

$$f_i(x) = h(x)g_i(T(x))$$

для определенных функций  $h$  и  $g$ . Тогда

$$\frac{f_i(x)}{f_0(x)} = \frac{g_i(T(x))}{g_0(T(x))}.$$

Следовательно,  $T^*(x)$  – функция от  $T(x)$ .

**Лемма 4.21.** Пусть  $\mathcal{P}$  – семейство эквивалентных мер и  $\mathcal{P}_0 \subset \mathcal{P}$  – конечное подсемейство. Тогда любая статистика  $T$ , достаточная для  $\mathcal{P}$  и минимальная достаточная для  $\mathcal{P}_0$  также минимальная достаточная для  $\mathcal{P}$ .

**ДОКАЗАТЕЛЬСТВО:** Пусть  $S$  – достаточная для  $\mathcal{P}$ . Тогда  $S$  также достаточная для  $\mathcal{P}_0$  и

$$T = h \circ S \quad \mathcal{P}_0\text{-п.н.}$$

Так как все меры эквивалентны,

$$T = h \circ S \quad \mathcal{P}\text{-п.н.}$$

**Теорема 4.22.** Пусть  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  –  $k$ -параметрическое экспоненциальное семейство с плотностями:

$$\frac{dP_\vartheta}{d\mu}(x) = c(\vartheta)h(x) \exp \left\{ \sum_{i=1}^k Q_i(\vartheta)T_i(x) \right\}.$$

Если  $Z = \{(Q_1(\vartheta), \dots, Q_k(\vartheta))^T \mid \vartheta \in \Theta\}$  имеет непустую внутреннюю часть, то  $(T_1(x), \dots, T_k(x))^T$  – минимальная достаточная для  $\vartheta$ .

ДОКАЗАТЕЛЬСТВО: Статистика  $(T_1(x), \dots, T_k(x))^T$  – достаточная по критерию Неймана. Пусть  $\mathcal{P}_0 = \{P_{\vartheta_i} \mid i = 0, \dots, k\}$  – конечное подсемейство. Из Примера 4.20 и Замечания 4.13 мы знаем, что

$$\tilde{T}(x) = \left( \sum_{i=1}^k (Q_i(\vartheta_1) - Q_i(\vartheta_0))T_i(x), \dots, \sum_{i=1}^k (Q_i(\vartheta_k) - Q_i(\vartheta_0))T_i(x) \right)^T$$

минимальная достаточная для  $\mathcal{P}_0$ . Пусть  $T(x) = (T_1(x), \dots, T_k(x))^T$ , тогда имеет место равенство

$$\tilde{T} = \Delta Q \cdot T = (Q_i(\vartheta_j) - Q_i(\vartheta_0))_{i,j=1}^k \cdot T.$$

Если мы выберем подсемейство так, что  $\Delta Q$  обратима (это возможно, благодаря непустой части  $Z$ ), то

$$T = (\Delta Q)^{-1} \cdot \tilde{T}$$

минимальная достаточная для  $\mathcal{P}_0$ . Тогда теорема следует из Леммы 4.21.

**Замечание 4.23.** Используя Теорему 4.15 все кандидаты на UMVU-оценку – достаточные статистики. Мы будем искать условия, при которых класс этих статистик будет относительно небольшим.

**Определение 4.24.**

- (i) Пусть  $\mathcal{P} = \{P_{\vartheta} \mid \vartheta \in \Theta\}$  – семейство вероятностных мер на  $(\mathcal{X}, \mathcal{B})$ . Тогда  $\sigma$ -алгебра  $\mathcal{B}_0 \subset \mathcal{B}$  называется **полной** для  $\mathcal{P}$ , если для любой  $\mathcal{B}_0$ -измеримой функции  $g : \mathcal{X} \rightarrow \mathbb{R}$ :

$$\mathbb{E}_{\vartheta}[g(X)] = 0 \quad \forall \vartheta \in \Theta \quad \Longleftrightarrow \quad g = 0 \text{ } P_{\vartheta}\text{-п.н.} \quad \forall \vartheta \in \Theta.$$

- (ii) Статистика  $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{T}, \mathcal{D})$  называется **полной** для  $\vartheta$ , если  $\sigma(T)$  полная для  $\vartheta$ :

$$\mathbb{E}_{\vartheta}[g \circ T] = 0 \quad \forall \vartheta \in \Theta \quad \Longleftrightarrow \quad g \circ T = 0 \text{ } P_{\vartheta}\text{-п.н.} \quad \forall \vartheta \in \Theta.$$

**Теорема 4.25.** Пусть  $\mathcal{P} = \{P_{\vartheta} \mid \vartheta \in \Theta\}$  –  $k$ -параметрическое экспоненциальное семейство с непустой внутренней частью и плотностями:

$$\frac{dP_{\vartheta}}{d\mu}(x) = c(\vartheta)h(x) \exp \left\{ \sum_{i=1}^k Q_i(\vartheta)T_i(x) \right\}.$$

Тогда  $T(x) = (T_1(x), \dots, T_k(x))^T$  – полная статистика для  $\vartheta$ .

ДОКАЗАТЕЛЬСТВО: Естественное параметрическое пространство:  $\vartheta_i = Q_i(\vartheta)$ . Предположим, что  $[-a, a]^k \subset \Theta^*$ . Пусть  $g$  измерима и

$$0 = \mathbb{E}_{\vartheta}[g(T(X))] = \int c(\vartheta)g(t) \exp \left\{ \sum_{i=1}^k \vartheta_i t_i \right\} \mu^T(dt) \quad \forall \vartheta \in [-a, a]^k.$$

Разложим  $g = g^+ - g^-$  и получим:

$$\int c(\vartheta)g^+(t) \exp \left\{ \sum_{i=1}^k \vartheta_i t_i \right\} \mu^T(dt) = \int c(\vartheta)g^-(t) \exp \left\{ \sum_{i=1}^k \vartheta_i t_i \right\} \mu^T(dt) \quad \forall \vartheta \in [-a, a]^k. \quad (4.4)$$

В частности мы получаем для  $\vartheta = 0$ :

$$A = \int g^+ \mu^T(dt) = \int g^- \mu^T(dt).$$

Нам нужно показать,  $g^+ = g^- = 0$   $\mu^T$ -п.н. Если  $A = 0$ , то утверждение очевидно. Пусть  $A \neq 0$ , тогда зададим

$$P^{\pm}(B) = \frac{1}{A} \int_B g^{\pm} \mu^T(dt).$$

Тогда равенство (4.4) эквивалентно

$$\int \exp \left\{ \sum_{i=1}^k \vartheta_i t_i \right\} P^+(dt) = \int \exp \left\{ \sum_{i=1}^k \vartheta_i t_i \right\} P^-(dt) \quad \forall \vartheta \in [-a, a]^k,$$

что несколько напоминает равенство характеристических функций, но без комплексной части в экспоненте. Зададим  $\vartheta_i = \xi_i + i\eta_i$ , где  $|\xi_i| \leq a \quad \forall i = 1, \dots, k$ . Рассмотрим функции

$$f_l^\pm : \begin{cases} \{z \in \mathbb{C} \mid |Re(z)| \leq a\} & \rightarrow \mathbb{C} \\ z & \mapsto \int \exp \left\{ \sum_{i \neq l} \vartheta_i t_i + z t_l \right\} P^\pm(dt). \end{cases}$$

Используя комплексную версию Теоремы 2.38 мы заключаем, что эти функции аналитические (и таким образом голоморфные). На множестве  $\mathbb{C} \cap [-a, a]$  имеет место равенство  $f_l^+(z) = f_l^-(z)$ . Используя тождественную теорему для голоморфных функций<sup>1</sup>, мы получаем равенство

$$f_l^+(z) = f_l^-(z) \quad \forall z \in \{z \in \mathbb{C} \mid |Re(z)| \leq a\}.$$

Так как  $l$  выбирается произвольно:

$$\int \exp \left\{ i \sum_{i=1}^k \eta_i t_i \right\} P^+(dt) = \int \exp \left\{ i \sum_{i=1}^k \eta_i t_i \right\} P^-(dt) \quad \forall \eta = (\eta_1, \dots, \eta_k)^T.$$

Следовательно, исходя из теоремы о единственности характеристических функций, имеет равенство мер  $P^+$  и  $P^-$  и

$$g^+ = g^- \quad \mu\text{-п.н.} \quad \implies \quad g = 0 \quad \mu\text{-п.н.}$$

#### Пример 4.26.

(i) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(0, \sigma^2)$  с плотностью

$$f_{\sigma^2}(x) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{\sum_{i=1}^n X_i^2}{2\sigma^2} \right\}.$$

Мы видим из Теоремы 4.11, что

$$\begin{aligned} T_1(X) &= (X_1, \dots, X_n)^T, \\ T_2(X) &= (X_1^2, \dots, X_n^2)^T, \\ T_3(X) &= \left( \sum_{i=1}^m X_i^2, \sum_{i=m+1}^n X_i^2 \right)^T, \\ T_4(X) &= \sum_{i=1}^n X_i^2 \end{aligned}$$

достаточные статистики в порядке убывания сложности.

(ii) Несложно увидеть, что можно подобрать набор функций  $h_{ij}$ , такой что

$$T_j(X) = h_{ij}(T_i), \quad i < j.$$

Следовательно,  $T_1, T_2, T_3$  не минимальные достаточные, в отличие от  $T_4$  (по Теореме 4.22).

---

<sup>1</sup>Пусть заданы функции  $f$  и  $g$  на связном открытом множестве  $D$ . Тогда, если  $f = g$  на некотором непустом открытом подмножестве  $D$ , то  $f = g$  на всем множестве  $D$ .

(iii) Статистики  $T_1, T_2, T_3$  не полные, так как можно подобрать следующие функции:

$$g_1(X) = X_1 \text{ (для } T_1),$$

$$g_2(X) = X_2 - X_1 \text{ (для } T_2),$$

$$g_3(X) = (n - m)X_1 - mX_2 \text{ (для } T_3).$$

Статистика  $T_4$  является полной вследствие Теоремы 4.25.

#### Пример 4.27.

(i) В Примере 4.17(i) мы показали, что

$$X_{(n)} = \max_{i=1}^n X_i$$

является достаточной для  $\vartheta \in (0, \infty)$ . Она также является полной, так как плотность  $X_{(n)}$

$$f_{\vartheta}(x) = \frac{n}{\vartheta^n} x^{n-1} 1_{[0, \vartheta]}(x).$$

То есть, если

$$\mathbb{E}_{\vartheta}[g(X_{(n)})] = \frac{n}{\vartheta^n} \int_0^{\vartheta} g(x) x^{n-1} dx = 0 \quad \forall \vartheta > 0,$$

то  $g(x) \equiv 0$   $\lambda$ -п.н.

(ii) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim F \ll \lambda$  и

$$\mathcal{P} = \{F \mid F - \text{функция распределения}\}.$$

Тогда порядковая статистика  $X_{(\cdot)}$  является полной (см. Пример 4.34 в [4]).

**Теорема 4.28 (Теорема Леманна-Шеффе).** Пусть  $\mathcal{P} = \{P_{\vartheta} \mid \vartheta \in \Theta\}$  – семейство вероятностных мер на  $(\mathcal{X}, \mathcal{B})$ ,  $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{T}, \mathcal{D})$  – достаточная и полная статистика для  $\vartheta$ . Пусть также  $\gamma : \Theta \rightarrow \Gamma$  – функционал, для которого может быть получена несмещенная оценка и  $L$  – функция потерь, такая что  $L(\gamma(\vartheta), \cdot)$  – выпуклая  $\forall \vartheta \in \Theta$ . Тогда существует (почти наверное) единственная несмещенная оценка  $\gamma(\vartheta)$  вида  $h \circ T$ . Она имеет равномерно наименьший риск среди всех несмещенных оценок.

ДОКАЗАТЕЛЬСТВО: Существование следует из Теоремы 4.15.

Единственность: пусть  $g \circ T$  – другая несмещенная оценка  $\gamma(\vartheta)$ . Тогда:

$$\mathbb{E}_{\vartheta}[h(T) - g(T)] = 0 \quad \forall \vartheta \in \Theta.$$

Так как  $T$  – полная статистика, то:

$$h \circ T = g \circ T \quad \forall \vartheta \in \Theta.$$

Наконец, пусть  $k$  – некоторая несмещенная оценка  $\gamma(\vartheta)$ . Тогда

$$k \circ T = \mathbb{E}[k|T]$$

также несмещенная и мы знаем из Теоремы Рао-Блэквелла, что

$$R(\vartheta, k) \geq R(\vartheta, \mathbb{E}[k|T]) = R(\vartheta, h \circ T).$$

**Следствие 4.29.** В случае, когда функция потерь  $L(x, y) = (x - y)^2$  и  $T$  – достаточная и полная, любая несмещенная оценка вида  $h \circ T$  единственная и UMVU.

**Пример 4.30.**

(i) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{U}[0, \vartheta]$ . Тогда

$$\frac{n+1}{n}X_{(n)}$$

является UMVU-оценкой.

(ii) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim F \ll \lambda$ . Тогда эмпирическая функция распределения  $\hat{F}(z)$  – UMVU-оценка для  $F(z)$ .

(iii) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \text{Bin}(1, \vartheta)$ ,  $\vartheta \in [0, 1]$ . Тогда

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

является достаточной (Пример 4.2). Она также является полной, что можно видеть из Примера 2.34 и Теоремы 4.25 или же из того, что, если

$$0 = \mathbb{E}_\vartheta[h \circ T] = \sum_{j=0}^n h(j) \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \quad \forall \vartheta \in [0, 1],$$

то  $h \equiv 0$ . Следовательно,  $\bar{X}_n$  – UMVU-оценка.

**Замечание 4.31.** В тех же условиях, что и в Замечании 4.13 для статистик  $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{T}, \mathcal{B})$  и  $\tilde{T} = b \circ T : (\mathcal{X}, \mathcal{B}) \rightarrow (\tilde{\mathcal{T}}, \tilde{\mathcal{D}})$  имеет место следующая импликация: если  $T$  полная и достаточная, то  $\tilde{T}$  также полная и достаточная. Например, мы видим из Примера 4.30 (iii), что

$$g(X) = \frac{n}{n-1} \bar{X}_n (1 - \bar{X}_n)$$

UMVU-оценка для  $\vartheta(1-\vartheta)$ .



## Упражнения

4.1. Пусть  $X = (X_1, \dots, X_n)^T$  – вектор i.i.d. случайных величин, имеющих

(i) распределение Вейбулла:

$$f_{\theta, \alpha}(x) = \theta \alpha (\theta x)^{\alpha-1} \exp\{-(\theta x)^\alpha\} 1_{(0, \infty)}(x), \quad \theta, \alpha > 0.$$

(ii) равномерное распределение:

$$f_{\theta_1, \theta_2}(x) = \frac{1_{(\theta_1, \theta_2)}(x)}{\theta_2 - \theta_1}, \quad (\theta_1, \theta_2) \in \{(x, y) \in \mathbb{R}^2 \mid x < y\}.$$

Найдите достаточную статистику для  $X$ , используя критерий Неймана.

4.2. Пусть  $X_1, \dots, X_n$  i.i.d. случайные величины, распределенные по Парето с параметром  $(\theta, \alpha)$ , т.е. имеющие плотность:

$$f_{\theta, \alpha}(x) = \frac{\theta \alpha^\theta}{x^{\theta+1}} 1_{(\alpha, \infty)}(x), \quad \theta, \alpha > 0.$$

(i) Найдите достаточную статистику для  $(\theta, \alpha)$ .

(ii) Найдите достаточную статистику для  $\theta$  и для  $\alpha$ , если другой параметр известен.

4.3. Пусть  $X$  – случайная величина, принимающая значения в  $\mathcal{X}$ , и, имеющая распределение  $P_\vartheta$ ,  $\vartheta \in \Theta$ . Статистика  $S : \mathcal{X} \rightarrow \mathcal{T}$  называется **свободной от распределения**, если её распределение не зависит от  $\vartheta$ . Статистика называется **свободной от распределения первого рода**, если её математическое ожидание  $\mathbb{E}_\vartheta[S(X)]$  не зависит от  $\vartheta$ . Докажите, что статистика  $T : \mathcal{X} \rightarrow \mathcal{T}$  является полной тогда и только тогда, когда существует неконстантная функция  $f : \mathcal{T} \rightarrow \tilde{\mathcal{T}}$ , такая что  $f(T)$  является свободной от распределения первого рода.

4.4. Пусть  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  – семейство распределений и  $T : \mathcal{X} \rightarrow \mathbb{R}^d$  – достаточная и полная статистика для  $\vartheta$ . Докажите, что  $T$  является минимальной достаточной статистикой, если таковая существует.

4.5. Докажите следующее утверждение. Пусть  $P_\vartheta$ ,  $\vartheta \in \Theta$  – семейство эквивалентных мер с  $\mu$ -плотностями  $f_\vartheta$ . Пусть  $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{T}, \mathcal{D})$  – статистика, обладающая следующим свойством:

$$\frac{f_\vartheta(x)}{f_\vartheta(y)}$$

не зависит от  $\vartheta$  тогда и только тогда, когда  $T(x) = T(y)$ . Тогда  $T$  – минимальная достаточная. Используйте следующие шаги:

(i) Покажите, что  $T$  – достаточная статистика. Для этого рассмотрите

$$f_\vartheta(x) = \frac{f_\vartheta(x)}{f_\vartheta(y)} f_\vartheta(y).$$

(ii) Докажите, что для любой достаточной статистики  $S$  имеет место импликация

$$S(x) = S(y) \implies T(x) = T(y),$$

и тем самым завершите доказательство.

4.6. Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \text{Cauchy}(\vartheta, 1)$ , т.е. плотность распределения  $X_i$ :

$$f_\vartheta(x) = \frac{1}{\pi(1 + (x - \vartheta)^2)}.$$

Покажите, используя утверждение из предыдущего упражнения, что порядковая статистика  $X_{(\cdot)}$  является минимальной достаточной для  $\vartheta$ .



## Глава 5

# Асимптотические свойства оценок

Пусть  $X^{(n)} = (X_1, \dots, X_n)^T$  – вектор случайных величин на пространстве  $\mathcal{X}_n = \mathcal{X}^n$  с распределением  $\mathcal{P}^n = \{P_\vartheta^n \mid \vartheta \in \Theta\}$ . Для любого  $n$  пусть функция

$$T_n : \begin{cases} \mathcal{X}_n & \rightarrow \Gamma \\ x^{(n)} & \mapsto T_n(x^{(n)}) \end{cases}$$

будет оценкой  $\gamma(\vartheta)$ . Минимальное условие для хорошей оценки – это стремление  $T_n$  к  $\gamma(\vartheta)$  при растущем значении  $n$ .

**Определение 5.1.** Пусть  $T_n : \mathcal{X}_n \rightarrow \Gamma$  – оценка  $\gamma(\vartheta)$ , принимающая значения в метрическом пространстве. Допустим, что все эксперименты определены на совместном вероятностном пространстве  $P_\vartheta^n \ll Q_\vartheta$  для любого  $n$ .

(i)  $T_n$  называется *(слабо) состоятельной* оценкой  $\gamma(\vartheta)$ , если

$$T_n \xrightarrow{Q_\vartheta} \gamma(\vartheta) \quad \forall \vartheta \in \Theta.$$

(ii)  $T_n$  называется *сильно состоятельной* оценкой  $\gamma(\vartheta)$ , если

$$T_n \rightarrow \gamma(\vartheta) \quad Q_\vartheta\text{-п.н.} \quad \forall \vartheta \in \Theta.$$

**Пример 5.2.** Вспомним метод моментов из Замечания 2.41:  $X_1, \dots, X_n$  i.i.d.  $\sim P_\vartheta$  вещественные случайные величины,  $\vartheta \in \Theta \subset \mathbb{R}^k$  и  $\gamma : \Theta \rightarrow \Gamma \subset \mathbb{R}^l$ . Также  $m_j = \mathbb{E}_\vartheta[X_1^j] = \int x^j P_\vartheta(dx)$ ,  $j = 1, \dots, k$  и

$$\gamma(\vartheta) = f(m_1, \dots, m_k).$$

Далее берем

$$\hat{\gamma}(X) = f(\hat{m}_1, \dots, \hat{m}_k),$$

где  $\hat{m}_j = \frac{1}{n} \sum_{i=1}^n X_i^j$ . Если  $\mathbb{E}_\vartheta[|X|^k] < \infty$ , то из Закона Больших чисел следует, что  $\hat{m}_j \rightarrow m_j$   $Q_\vartheta$ -п.н., где  $Q_\vartheta = \otimes_{i=1}^{\mathbb{N}} P_\vartheta$ . Так как  $f$  – непрерывная, мы получаем:

$$\hat{\gamma}(X) \rightarrow \gamma(\vartheta) \quad Q_\vartheta\text{-п.н.}$$

**Теорема 5.3 (Теорема Крамера-Вольда).** Пусть  $(X_n)$  – последовательность  $d$ -мерных случайных величин. Тогда  $X_n \xrightarrow{\mathcal{L}} X$  тогда и только тогда, когда

$$y^T X_n \xrightarrow{\mathcal{L}} y^T X \quad \forall y \in \mathbb{R}^d.$$

ДОКАЗАТЕЛЬСТВО: Согласно Теореме Леви о непрерывности:

$$X_n \xrightarrow{\mathcal{L}} X \iff \mathbb{E}[\exp\{iu^T X_n\}] \rightarrow \mathbb{E}[\exp\{iu^T X\}] \quad \forall u \in \mathbb{R}^d.$$

Остается только произвести замену  $u = ty$  для  $t \in \mathbb{R}$  и  $y \in \mathbb{R}^d$ .

**Теорема 5.4 (Центральная предельная теорема).** Пусть  $X_1, \dots, X_n$  i.i.d.  $d$ -размерные случайные величины,  $\mathbb{E}[X_j] = \mu \in \mathbb{R}^d$  и  $\text{Cov}(X_j) = \Sigma > 0 \in \mathbb{R}^{d \times d}$  (положительно-определенная). Тогда для случайного вектора

$$Z^{(n)} = \frac{1}{n} \sum_{j=1}^n \in \mathbb{R}^d$$

имеет место сходимость:

$$\sqrt{n}(Z^{(n)} - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma). \quad (5.1)$$

ДОКАЗАТЕЛЬСТВО: Из одномерной центральной предельной теоремы мы знаем, что

$$\sqrt{n}(y^T Z^{(n)} - y^T \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, y^T \Sigma y).$$

Применяем Теорему 5.3 и получаем сходимость (5.1).

**Определение 5.5.** Пусть  $T_n : \mathcal{X}_n \rightarrow \Gamma \subset \mathbb{R}^l$  – последовательность оценок.

- (i) Пусть  $\mu_n(\vartheta) = \mathbb{E}_\vartheta[T_n]$ , тогда  $T_n$  называется **асимптотически несмещенной** оценкой  $\gamma(\vartheta)$ , если

$$\mu_n(\vartheta) \rightarrow \gamma(\vartheta).$$

- (ii)  $T_n$  называется **асимптотически нормальной**, если существуют последовательности  $(\mu_n) \in \mathbb{R}^l$  и  $(\Sigma_n(\vartheta)) \in \mathbb{R}^{l \times l}$ , такие что  $\|\Sigma_n(\vartheta)\| \rightarrow 0$  и

$$\Sigma_n^{-\frac{1}{2}}(\vartheta)(T_n - \mu_n(\vartheta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbb{I}_l).$$

**Теорема 5.6 (Лемма Слуцкого).** Пусть  $(Y_n)$  и  $(Z_n)$  – последовательности  $d$ -мерных случайных величин, такие что

$$Z_n \xrightarrow{\mathcal{L}} Z \quad \text{и} \quad Y_n \xrightarrow{\mathbb{P}} y_0.$$

Тогда:

$$(i) \quad Z_n + Y_n \xrightarrow{\mathcal{L}} Z + y_0.$$

$$(ii) \quad Y_n^T Z_n \xrightarrow{\mathcal{L}} y_0^T Z.$$

ДОКАЗАТЕЛЬСТВО: Теорема 11.2.11 в [4].

**Пример 5.7.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \text{Bin}(1, p)$ ,  $p \in (0, 1)$ . Тогда оценка  $p$   $T_n = \bar{X}_n$  является несмещенной. Из центральной предельной теоремы мы знаем, что:

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Но так как  $X_n \xrightarrow{\mathbb{P}} p$ , то имеет место:

$$\sqrt{\bar{X}_n(1 - \bar{X}_n)} \xrightarrow{\mathbb{P}} \sqrt{p(1-p)}.$$

По Теореме 5.6:

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{\bar{X}_n(1 - \bar{X}_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Например, пусть  $n = 100$  и  $\bar{X}_n = 0.85$ . Тогда

$$P_p(|\bar{X}_n - p| < \varepsilon) \approx 2\Phi\left(\varepsilon \sqrt{\frac{n}{\bar{X}_n(1 - \bar{X}_n)}}\right) - 1 \quad \forall p \in (0, 1),$$

где  $\Phi$  – функция распределения  $\mathcal{N}(0, 1)$ .

$$P_p(|\bar{X}_n - p| < \varepsilon) \approx \begin{cases} 83.84\%, & \varepsilon = 0.05 \\ 99.46\%, & \varepsilon = 0.1 \end{cases}$$

**Теорема 5.8 (Дельта-метод).** Пусть  $(X_n)$  – последовательность  $k$ -мерных случайных векторов, такая что:

$$\frac{X_n - \mu}{c_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma),$$

где  $c_n \rightarrow 0$ ,  $\mu \in \mathbb{R}^k$  и  $\Sigma \geq 0 \in \mathbb{R}^{k \times k}$ . Пусть также  $g : \mathbb{R}^k \rightarrow \mathbb{R}^m$  – непрерывно дифференцируемая по  $\mu$  функция с матрицей Якоби  $D \in \mathbb{R}^{m \times k}$ . Тогда:

$$\frac{g(X_n) - g(\mu)}{c_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, D\Sigma D^T).$$

ДОКАЗАТЕЛЬСТВО: По Лемме 5.6:

$$X_n - \mu = \frac{X_n - \mu}{c_n} c_n \xrightarrow{\mathcal{L}} 0.$$

Из сходимости по распределению к постоянной следует сходимость по вероятности:

$$X_n \xrightarrow{\mathbb{P}} \mu.$$

Далее

$$\frac{g(X_n) - g(\mu)}{c_n} = g'(\mu) \frac{X_n - \mu}{c_n} + (g'(\xi_n) - g'(\mu)) \frac{X_n - \mu}{c_n},$$

где  $\xi_n$  – промежуточная точка:

$$\|\xi_n - \mu\| \leq \|X_n - \mu\|.$$

Следовательно,  $\xi_n \xrightarrow{\mathbb{P}} \mu$  и  $g'(\xi_n) \xrightarrow{\mathbb{P}} g'(\mu)$  (поскольку  $g$  – непрерывно дифференцируемая). Вновь по Лемме 5.6:

$$g'(\mu) \frac{X_n - \mu}{c_n} \xrightarrow{\mathcal{L}} g'(\mu) \mathcal{N}(0, \Sigma).$$

**Замечание 5.9.** Пусть в Примере 5.2  $\mathbb{E}_\vartheta[X_1^{2k}] < \infty$  для любого  $\vartheta \in \Theta$  и  $\gamma : \mathbb{R}^k \rightarrow \mathbb{R}^l$  – непрерывно дифференцируема по  $\mu = (m_1, \dots, m_k)^T$  с матрицей Якоби  $D$ . Мы знаем из Теоремы 5.4, что

$$\sqrt{n}((\hat{m}_1, \dots, \hat{m}_k)^T - (m_1, \dots, m_k)^T) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma),$$

где

$$\Sigma = (\Sigma)_{i,j=1}^k = (m_{i+j} - m_i m_j)_{i,j=1}^k.$$

Тогда

$$\sqrt{n}(\gamma(\hat{m}_1, \dots, \hat{m}_k) - \gamma(m_1, \dots, m_k)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, D\Sigma D^T).$$

**Пример 5.10.**

(i) Пусть  $X_1, \dots, X_n$  i.i.d.,  $\mathbb{E}_\vartheta[X_i] = \mu$  и  $\text{Var}_\vartheta(X_i) = \sigma^2$ . Из ЦПТ следует:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

В качестве оценки  $\mu^2$  выберем асимптотически несмещенную статистику  $\bar{X}_n^2$ . Используя Дельта-метод, получаем:

$$\sqrt{n}(\bar{X}_n^2 - \mu^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\mu^2\sigma^2).$$

(ii) Пусть

$$(X_i, Y_i)^T \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix}\right), \quad i = 1, \dots, n$$

i.i.d. с параметром  $\vartheta = (\mu_1, \mu_2, \sigma^2, \tau^2, \rho)^T$ . Оценка

$$\hat{\rho}_n = \frac{SQ_{xy}}{\sqrt{SQ_{xx}SQ_{yy}}},$$

где  $SQ_{xy} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)$ ,  $(SQ_{xx}, SQ_{yy})$  – аналогично), называется **коэффициентом корреляции Пирсона**. Без ограничения общности, пусть  $\mu_1 = \mu_2 = 0$ ,  $\sigma = \tau = 1$ , так как  $\hat{\rho}_n$  инвариантна по отношению к аффинным преобразованиям. Докажем сначала, что  $S_n = (SQ_{xx}, SQ_{yy}, SQ_{xy})^T$  удовлетворяет

$$\sqrt{n}(S_n - m) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V), \quad (5.2)$$

где  $m = (1, 1, \rho)^T$  и

$$V = 2 \begin{pmatrix} 1 & \rho^2 & \rho \\ \rho^2 & 1 & \rho \\ \rho & \rho & (1 + \rho^2)/2 \end{pmatrix}.$$

Чтобы доказать (5.2), используем Лемму 5.6 и ЦПТ и покажем, что

$$\sqrt{n}(\bar{X}_n, \bar{Y}_n) \xrightarrow{\mathbb{P}} 0, \quad \sqrt{n}(\bar{X}_n)^2 \xrightarrow{\mathbb{P}} 0, \quad \sqrt{n}(\bar{Y}_n)^2 \xrightarrow{\mathbb{P}} 0.$$

Далее, несложно увидеть, что

$$\sqrt{n}(S_n - m) - \sqrt{n}\left(\frac{1}{n} \sum_{i=1}^n Z_i - m\right) \xrightarrow{\mathbb{P}} 0,$$

где  $Z_i = (X_i^2, Y_i^2, X_i Y_i)^T$ . Затем, покажем, что

$$\text{Cov}(Z_i) = \mathbb{E}[Z_i Z_i^T] - \mathbb{E}[Z_i] \mathbb{E}[Z_i]^T = V.$$

Сходимость (5.2) следует из Теоремы 5.4. Наконец,  $g(S_n) = \hat{\rho}_n$ , где  $g(x_1, x_2, x_3) = \frac{x_3}{\sqrt{x_1 x_2}}$ . Тогда матрица Якоби функции  $g$  в точке  $m$ :

$$D = (-\rho/2, -\rho/2, 1).$$

Как результат,

$$\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, DVD^T) = \mathcal{N}(0, (1 - \rho^2)^2).$$

**Определение 5.11.** Пусть  $T_n : \mathcal{X} \rightarrow \mathbb{R}^l$  – асимптотически несмещенная и асимптотически нормальная последовательность оценок. При условиях регулярности из Теоремы 2.23 мы назовем  $T_n$  **асимптотически эффективной**, если

$$\lim_{n \rightarrow \infty} \Sigma_n(\vartheta) I(f_n(\cdot, \vartheta)) = \mathbb{I}_l \quad \forall \vartheta \in \Theta,$$

где  $\mathbb{I}_l$  – единичная матрица, а  $I(f_n(\cdot, \vartheta))$  – информация Фишера.

**Замечание 5.12.** Заданное выше определение можно интерпретировать следующим образом: если  $T_n$  несмещенная, то вследствие Теоремы Рао-Крамера

$$\text{Cov}_{\vartheta}(T_n) \geq I^{-1}(f_n(\cdot, \vartheta)).$$

Но, так как

$$\Sigma_n^{\frac{1}{2}}(\vartheta)(T_n - \mu_n(\vartheta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbb{I}_l),$$

то

$$\text{Cov}_{\vartheta}(T_n) \approx \Sigma_n(\vartheta) \approx I^{-1}(f_n(\cdot, \vartheta))$$

и  $T_n$  асимптотически несмещенная и асимптотически эффективная.

**Пример 5.13.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$ . Из Примера 2.29 следует, что для

$$g_n(X) = \left( \frac{1}{n-1} \sum_{i=1}^n \bar{X}_n (X_i - \bar{X}_n)^2 \right)$$

выполняется равенство

$$\text{Cov}_{\vartheta}(g_n) = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/(n-1) \end{pmatrix} = \Sigma_n(\vartheta).$$

Но информация Фишера:

$$I^{-1}(f_n(\cdot, \vartheta)) = \begin{pmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{pmatrix}$$

и  $g_n$  не эффективная, но эффективная асимптотически.

**Замечание 5.14.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim P_{\vartheta}$ ,  $\vartheta \in \Theta$  с  $\mu$ -плотностями  $f(\cdot, \vartheta)$ . Мы назовем

$$\ell(\cdot, \vartheta) = \log f(\cdot, \vartheta)$$

*логарифмической функцией правдоподобия* и зададим

$$\hat{\theta}_n(X) = \arg \sup_{\vartheta \in \Theta} f(X, \vartheta) = \arg \sup_{\vartheta \in \Theta} \ell(X, \vartheta) = \arg \sup_{\vartheta \in \Theta} \frac{1}{n} \sum_{i=1}^n \ell(X_i, \vartheta)$$

как оценку максимального правдоподобия для  $\vartheta$  (если таковая существует).

**Определение 5.15.** Пусть  $\mathbb{P}$  и  $\mathbb{Q}$  – вероятностные меры на  $(\mathcal{X}, \mathcal{B})$ , тогда

$$KL(\mathbb{P}|\mathbb{Q}) = \begin{cases} \int_{\mathcal{X}} \log \left( \frac{d\mathbb{P}}{d\mathbb{Q}}(x) d\mathbb{P}(x) \right), & \text{если } \mathbb{P} \ll \mathbb{Q}, \\ \infty, & \text{иначе} \end{cases}$$

называется *расстоянием Кульбака - Лейблера* для  $\mathbb{P}$  и  $\mathbb{Q}$ .

**Лемма 5.16.** Величина  $KL(\mathbb{P}|\mathbb{Q}) \geq 0$  и достигает нуля тогда и только тогда, когда  $\mathbb{P} = \mathbb{Q}$ .

ДОКАЗАТЕЛЬСТВО:

$$\begin{aligned} \int_{\mathcal{X}} \log \left( \frac{d\mathbb{P}}{d\mathbb{Q}} \right)(x) \mathbb{P}(dx) &= \int_{\mathcal{X}} -\log \left( \frac{d\mathbb{Q}}{d\mathbb{P}} \right)(x) \mathbb{P}(dx) \\ \text{неравенство Йенсена} &\rightarrow \geq -\log \int_{\mathcal{X}} \left( \frac{d\mathbb{Q}}{d\mathbb{P}} \right)(x) \mathbb{P}(dx) \\ &= -\log \int_{\mathcal{X}} \mathbb{Q}(dx) = 0. \end{aligned}$$

Неравенство вырождается в равенство, если  $\frac{d\mathbb{Q}}{d\mathbb{P}}(x) = 1$  п.н.

**Теорема 5.17.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim P_{\vartheta}$ ,  $\vartheta \in \Theta$ ,  $\ell(\cdot, \vartheta)$  – функция правдоподобия. Также:

- (i)  $\Theta \subset \mathbb{R}^k$  – компактное пространство.
- (ii)  $\eta \mapsto L(\eta, \vartheta) = \mathbb{E}[\ell(X_i, \eta)]$  непрерывная и  $\eta \mapsto L_n(\eta) = \frac{1}{n} \sum_{i=1}^n \ell(X_i, \eta)$   $\otimes_{i=1}^n P_{\vartheta}$ -п.н. непрерывная функции.
- (iii) Пусть  $Q_{\vartheta} = \otimes_{i=1}^{\infty} P_{\vartheta}$  и

$$\sup_{\eta \in \Theta} |L_n(\eta) - L(\eta, \vartheta)| \xrightarrow{Q_{\vartheta}} 0.$$

Тогда оценка максимального правдоподобия  $\hat{\theta}_n$  состоятельная.

ДОКАЗАТЕЛЬСТВО: Для любого  $\eta \in \Theta$ :

$$L_n(\eta) \rightarrow L(\eta, \vartheta) = \int \ell(x, \eta) f(x, \vartheta) \mu(dx) = \int \ell(x, \vartheta) f(x, \vartheta) \mu(dx) - KL(\vartheta|\eta)$$

$\mu$ -п.н. Используя Лемму 5.16, заключаем что  $\eta \mapsto L(\eta, \vartheta)$  достигает максимума при  $\eta = \vartheta$ . Функция

$$\arg \max : \begin{cases} C(\Theta, \mathbb{R}) & \rightarrow \Theta \\ f & \mapsto m_f = \arg \max_{\eta \in \Theta} f(\eta) \end{cases}$$

непрерывна для тех  $f$ , для которых  $m_f$  единственная. Тогда утверждение будет доказано, исходя из того, что

$$\vartheta = \arg \max L(\eta, \vartheta) \quad \text{и} \quad \hat{\theta}_n = \arg \max L_n(\eta)$$

и из условия (iii).

### Замечание 5.18.

- (i) Так как  $\eta \rightarrow L_n(\eta)$  (п.н.) непрерывна и  $\Theta$  – компакт, оценка максимального правдоподобия существует (п.н.) и измерима (Лемма 6.7 в [5])
- (ii) Наиболее сложным условием для проверки является (iii). Как правило, используется следующий вывод: Предположим, что  $T \subset \mathbb{R}$  компакт и случайные величины  $X_n(\gamma)$  и  $X(\gamma)$  такие, что

$$X_n(\gamma) \xrightarrow{\mathbb{P}} X(\gamma)$$

и

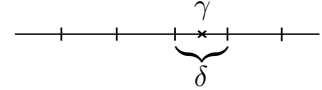
$$\gamma \mapsto X(\gamma) \quad \text{и} \quad \gamma \mapsto X_n(\gamma)$$

(п.н.) непрерывные. Тогда

$$\sup_{\gamma \in \Gamma} \|X_n(\gamma) - X(\gamma)\| \xrightarrow{\mathbb{P}} 0$$

тогда и только тогда, когда

$$\forall \varepsilon > 0 \quad \lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{|\gamma_1 - \gamma_2| < \delta} \|X_n(\gamma_1) - X_n(\gamma_2)\| \geq \varepsilon \right) = 0.$$



Идея заключается в интерполяции внутри интервалов  $\delta$ .

**Теорема 5.19 (Асимптотическая эффективность оценок МП).** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim P_\vartheta$ ,  $\vartheta \in \Theta$ ,  $\ell(\cdot, \vartheta)$  – функция правдоподобия. Также:

- (i)  $\Theta \subset \mathbb{R}^k$  – компактное пространство и  $\vartheta \subset \text{int}(\Theta)$ .
- (ii)  $\eta \mapsto \ell(x, \eta)$  непрерывна на  $\Theta$  и дважды непрерывно дифференцируема по  $\vartheta$  для почти всех  $x \in \mathcal{X}$ .
- (iii) Существуют функции  $H_0, H_2 \in L^1(P_\vartheta)$  и  $H_1 \in L^2(P_\vartheta)$ , такие что:

$$\sup_{\eta \in \Theta} |\ell(x, \eta)| \leq H_0(x), \quad \sup_{\eta \in \Theta} \|\dot{\ell}(x, \eta)\| \leq H_1(x), \quad \sup_{\eta \in \Theta} \|\ddot{\ell}(x, \eta)\| \leq H_2(x) \quad \forall x \in \mathcal{X}.$$

(iv) Информация Фишера

$$I(f(\cdot, \vartheta)) = \mathbb{E}_\vartheta[\dot{\ell}(X, \vartheta) \dot{\ell}(X, \vartheta)^T]$$

положительно определенная (обратима).

Тогда имеет место сходимость:

$$\sqrt{n}(\hat{\theta}_n - \vartheta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I(f(\cdot, \vartheta))^{-1}).$$



ДОКАЗАТЕЛЬСТВО:

*Шаг 1:* покажем состоятельность оценки  $\hat{\theta}_n$  (проверим, что условия из Теоремы 5.17 соблюдаются)

- (i) Соблюдается по предположению.
- (ii)  $\eta \mapsto L_n(\eta)$  п.н. непрерывная по условию (ii). Также, используя (ii), (iii) и мажорируемую сходимость, получаем

$$|L(\eta_1, \vartheta) - L(\eta_2, \vartheta)| \leq \int_{\mathcal{X}} |\ell(x, \eta_1) - \ell(x, \eta_2)| f(x, \vartheta) \mu(dx) \rightarrow 0,$$

при  $\eta_1 \rightarrow \eta_2$ .

- (iii) В силу Сильного Закона Больших Чисел:

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{\|\eta_1 - \eta_2\| < \delta} |L_n(\eta_1) - L_n(\eta_2)| &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sup_{\|\eta_1 - \eta_2\| < \delta} |\ell(X_i, \eta_1) - \ell(X_i, \eta_2)| \\ &= \mathbb{E}_{\vartheta} \left[ \sup_{\|\eta_1 - \eta_2\| < \delta} |\ell(X, \eta_1) - \ell(X, \eta_2)| \right] \quad Q_{\vartheta} = \otimes_{i=1}^{\infty} P_{\vartheta}\text{-п.н.} \end{aligned}$$

Так как  $\Theta$  компакт, то  $\eta \mapsto \ell(X, \eta)$  (п.н.) равномерно непрерывная. Как следствие, последнее выражение сходится к нулю при  $\delta \rightarrow 0$  (используя снова мажорируемую сходимость).

*Шаг 2:* пусть  $A_n$  –  $k$ -мерный прямоугольник с вершинами  $\hat{\theta}_n$  и  $\vartheta$ . Поскольку  $\hat{\theta}_n \xrightarrow{Q_{\vartheta}} \vartheta$  и  $\vartheta \in \text{int}(\Theta)$ , то

$$Q_{\vartheta}(A_n \subset \text{int}(\Theta)) \rightarrow 1.$$

Также

$$\dot{L}_n(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^n \dot{\ell}(X_i, \hat{\theta}_n) = 0$$

по определению  $\hat{\theta}_n$ . По Теореме о среднем:

$$-\dot{L}_n(\vartheta) = \dot{L}_n(\hat{\theta}_n) - \dot{L}_n(\vartheta) = \ddot{L}_n(\tilde{\theta}_n)(\hat{\theta}_n - \vartheta)$$

для некоторого  $\tilde{\theta}_n \in A_n$ . Как и в Замечании 2.25:

$$\begin{aligned} \mathbb{E}[\dot{\ell}(X_i, \vartheta)] &= \int_{\mathcal{X}} \dot{\ell}(x, \vartheta) f(x, \vartheta) \mu(dx) \\ \ell = \log f &\rightarrow = \int_{\mathcal{X}} \dot{f}(x, \vartheta) \mu(dx) = 0. \end{aligned}$$

По определению

$$\text{Cov}(\dot{\ell}(X_i, \vartheta)) = I(f(\cdot, \vartheta)).$$

Тогда по Центральной Предельной Теореме:

$$\sqrt{n} \dot{L}_n(\vartheta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I(f(\cdot, \vartheta))).$$

*Шаг 3:* предположим, что

$$\ddot{L}_n(\tilde{\theta}_n) \xrightarrow{Q_{\vartheta}} -I(f(\cdot, \vartheta)). \quad (5.3)$$

Если (5.3) соблюдается, мы можем заключить, что

$$\lim_{n \rightarrow \infty} Q_{\vartheta}(\ddot{L}_n(\tilde{\theta}_n) \text{ обратима}) = 1.$$

Тогда, используя лемму Слуцкого, получаем

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \vartheta) &= -\ddot{L}_n(\tilde{\theta}_n)^{-1} \dot{L}_n(\vartheta) 1_{\{A_n \subset \text{int}(\Theta)\} \cap \{\ddot{L}_n(\tilde{\theta}_n) \text{ обратима}\}} + B_n \\ &\rightarrow I(f(\cdot, \vartheta))^{-1} \mathcal{N}(0, I(f(\cdot, \vartheta))) = \mathcal{N}(0, I(f(\cdot, \vartheta))^{-1}), \end{aligned}$$

так как  $B_n \xrightarrow{Q_\vartheta} 0$ .

*Шаг 4:* докажем (5.3). Используем равенство:

$$\ddot{\ell}(x, \vartheta) = \frac{\ddot{f}(x, \vartheta)}{f(x, \vartheta)} - \dot{\ell}(x, \vartheta)\dot{\ell}(x, \vartheta)^T.$$

Тогда имеет место:

$$\mathbb{E}_\vartheta[\ddot{\ell}(X, \vartheta)] + I(f(\cdot, \vartheta)) = \mathbb{E}_\vartheta\left[\frac{\ddot{f}(X, \vartheta)}{f(X, \vartheta)}\right] = 0,$$

как в Замечании 2.25. Из Закона Больших Чисел следует:

$$\ddot{L}_n(\vartheta) \xrightarrow{Q_\vartheta} -I(f(\cdot, \vartheta)).$$

Наконец, используем равенство

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} Q_\vartheta(\|\tilde{\theta}_n - \vartheta\| < \delta) = 1$$

и непрерывность  $\ddot{\ell}$  по  $\vartheta$ , чтобы завершить доказательство (5.3).

**Замечание 5.20.** Теорема 5.17 и Теорема 5.19 – это специальные случаи *оценок минимального контраста*, которые минимизируют функцию

$$\eta \mapsto \frac{1}{n} \sum_{i=1}^n k(X_i, \eta).$$

Можно получить похожие результаты при схожих предположениях, но асимптотическая эффективность достигается только при  $k = -\ell$ .

**Пример 5.21.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \text{Exp}(\lambda)$ . Найдём оценку максимального правдоподобия. Для этого

$$f_n(X, \lambda) = \lambda^n \exp\left(-\lambda \sum_{i=1}^n X_i\right) 1_{[0, \infty)}\left(\min_{i=1}^n X_i\right)$$

или

$$\ell_n(X, \lambda) = n \log(\lambda) - \lambda \sum_{i=1}^n X_i + \log 1_{[0, \infty)}\left(\min_{i=1}^n X_i\right)$$

должны быть максимизированы. Приравнявая любую из производных к нулю, получаем условие:

$$\frac{n}{\lambda} = \sum_{i=1}^n X_i,$$

то есть оценка:

$$\hat{\lambda}_n = \frac{1}{\bar{X}_n}.$$

Рассчитаем информацию Фишера:

$$\begin{aligned} \ell_1(X, \lambda) &= \log(\lambda) - \lambda X + \log(1_{[0, \infty)} X) \\ \Rightarrow \dot{\ell}_1(X, \lambda) &= -\left(X - \frac{1}{\lambda}\right) \\ \Rightarrow I(f(\cdot, \lambda)) &= \mathbb{E}_\lambda\left[\left(X - \frac{1}{\lambda}\right)^2\right]. \end{aligned}$$

По Теореме 5.19

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda^2),$$

но с другой стороны:

$$\sqrt{n}\left(\bar{X}_n - \frac{1}{\lambda}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\lambda^2}\right).$$

Используя Дельта-метод для  $g(x) = x^{-1}$ , получаем

$$\sqrt{n}(\bar{X}_n^{-1} - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda^2).$$

Также заметим, что

$$\text{Var}_\lambda(\hat{\lambda}_n) = n^{-1}\lambda^2 = (nI(f(\cdot, \lambda)))^{-1}$$

является следствием из Предположения 2.31 для экспоненциального семейства.

## Упражнения

**5.1.** Докажите следующий вариант леммы Слущкого: пусть  $(X_n)$  и  $(Y_n)$  – последовательности вещественнозначных случайных величин, такие что

$$X_n \xrightarrow{\mathcal{L}} X \quad \text{и} \quad Y_n \xrightarrow{\mathbb{P}} c,$$

где  $X$  – вещественнозначная случайная величина и  $c \in \mathbb{R}$ . Тогда имеет место:

$$X_n + Y_n \xrightarrow{\mathcal{L}} X + c.$$

*Подсказка:* последовательность  $(Z_n)$  сходится к  $Z$  тогда и только тогда, когда  $\mathbb{E}[f(Z_n)] \rightarrow \mathbb{E}[f(Z)]$  для любых равномерно непрерывных и ограниченных функций  $f: \mathbb{R} \rightarrow \mathbb{R}$ .

**5.2.** Пусть  $(X_n)$  – последовательность i.i.d. случайных величин с плотностью

$$f(x, \vartheta) = \vartheta x^{\vartheta-1} 1_{[0,1]}(x).$$

(i) Найдите оценку максимального правдоподобия  $\hat{\vartheta}$ .

(ii) Рассчитайте асимптотическое распределение  $\hat{\vartheta}$ .

*Подсказка:* для  $\vartheta > 0$ :

$$\int_0^1 \log(x) x^{\vartheta-1} dx = -\frac{1}{\vartheta}, \quad \int_0^1 (\log(x))^2 x^{\vartheta-1} dx = \frac{2}{\vartheta^3}.$$

**5.3.** *Коэффициент вариации* вероятностного распределения  $P$  определяется как

$$c_v(X) := \frac{\sqrt{\text{Var}(X)}}{\mathbb{E}[X]}, \quad X \sim P,$$

если соответствующие моменты существуют и  $\mathbb{E}[X] \neq 0$ . Пусть  $(X_n)$  – последовательность i.i.d. нормально распределенных случайных величин:  $X_i \sim \mathcal{N}(\mu, \sigma^2)$ , где  $\mu \in \mathbb{R}/\{0\}$  и  $\sigma^2 > 0$ .

(i) Найдите оценку  $\hat{c}_n(X)$  для  $c_v(X)$ , полученную методом моментов.

(ii) Выведите Центральную Предельную Теорему для оценки  $\hat{c}_n(X)$ .

**5.4.** Пусть  $(X_n)$  – последовательность i.i.d. равномерно распределенных случайных величин:  $X_i \sim \mathcal{U}[0, \vartheta]$ . Покажите, что  $(\prod_{i=1}^n X_i)^{\frac{1}{n}} - \frac{\vartheta}{e}$  – состоятельная оценка для  $\frac{\vartheta}{e}$ .

*Подсказка:* один из способов доказать сходимость по распределению, это показать, что

$$\sqrt{n} \left( \left( \prod_{i=1}^n X_i \right)^{\frac{1}{n}} - \frac{\vartheta}{e} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, \left( \frac{\vartheta}{e} \right)^2 \right).$$

**5.5.** Пусть  $(X_n)$  – последовательность i.i.d.  $\mathcal{N}(\mu, 1)$  распределенных величин. Пусть также  $a > 0$  и

$$\hat{\mu}_a = \begin{cases} \bar{X}_n, & |\bar{X}_n| \geq n^{-\frac{1}{4}}, \\ a\bar{X}_n, & \text{иначе.} \end{cases}$$

(i) Покажите, что

$$\sqrt{n}(\hat{\mu}_a - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v(\mu)),$$

где  $v(\mu) = 1$ , если  $\mu \neq 0$ , и  $v(\mu) = a^2$ , иначе.

(ii) Для каких  $a$   $\mu_a$  является эффективной?

(iii) Покажите, что существуют случаи, для которых имеет место:

$$v(\mu) \leq I_1^{-1}(\mu).$$

## Глава 6

# Основы тестирования

В этой главе мы проверяем гипотезы о неизвестном параметре  $\vartheta$ . Как и ранее, мы рассматриваем статистический эксперимент  $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ , где  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$ .

**Пример 6.1.** Обсудим (упрощенный) клинический эксперимент, в котором мы принимаем решение, лучше ли новоизобретенное лекарство В, чем известное лекарство А или нет. Предположим, что мы знаем по опыту предыдущих лет, что А имеет шанс излечения 65%. Новое лекарство В было протестировано на 100 подопытных и 80% выздоровели. Должны мы выбрать А или В? На языке математики мы проверяем:

$$H : p \leq 0.65 \quad \text{против} \quad K : p > 0.65,$$

где  $p$  – неизвестная вероятность излечения после принятия лекарства В.

**Определение 6.2.** Пусть  $\Theta = \Theta_H \cup \Theta_K$  – разделение параметрического пространства.

(i)  $\Theta_H$  называется *(нулевой) гипотезой*,  $\Theta_K$  называется *альтернативой*.

(ii) *Рандомизированный критерий* – измеримое отображение

$$\varphi : (\mathcal{X}, \mathcal{B}) \rightarrow ([0, 1], \mathcal{B}|_{[0, 1]}).$$

Функция  $\varphi(x)$  – это вероятность принятия решения, что  $\vartheta \in \Theta_K$ , после наблюдения  $x = X(\omega)$ . Множество таких критериев обозначим:

$$\Phi = \{\varphi \mid \varphi \text{ – рандомизированный критерий}\}.$$

(iii) Для критерия  $\varphi$  назовем  $\mathcal{K} = \{x \mid \varphi(x) = 1\}$  *критической областью*, а  $\mathcal{R} = \{x \mid \varphi(x) \in (0, 1)\}$  – *областью рандомизации*. Критерий  $\varphi$  называется *нерандомизированным*, если  $\mathcal{R} = \emptyset$ .

**Пример 6.3.** В ситуации в предыдущем примере мы знаем, что  $\bar{X}_n$  является UMVU-оценкой  $p$ . Разумно принять  $K$ , если значение  $\bar{X}_n$  достаточно большое, например:

$$\varphi(x) = \begin{cases} 1 & \bar{X}_n > 0.7 \\ 0 & \bar{X}_n \leq 0.7 \end{cases}$$

является разумным критерием.

**Замечание 6.4.** При принятии решения могут произойти две ошибки:

- Ошибка первого рода: отклонить гипотезу  $H$ , когда она верна.
- Ошибка второго рода: принять гипотезу  $H$ , когда она неверна.

		Истина	
		$\Theta_H$	$\Theta_K$
Решение	$\Theta_H$	✓	Ошибка 2-го рода
	$\Theta_K$	Ошибка 1-го рода	✓

Обе ошибки могут произойти с определенными вероятностями.

**Пример 6.5.** В Примере 6.3 вероятность принятия  $K$ :

$$P_p(\varphi(X) = 1) = P_p(\bar{X}_n > 0.7).$$

На практике биномиальное распределение можно аппроксимировать нормальным:

$$P_p(\bar{X}_n > 0.7) = P_p\left(\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} > \frac{\sqrt{n}(0.7 - p)}{\sqrt{p(1-p)}}\right)$$

$$\text{ЦПТ} \rightarrow \approx P\left(\mathcal{N}(0, 1) > \frac{\sqrt{n}(0.7 - p)}{\sqrt{p(1-p)}}\right) = \Phi\left(\frac{\sqrt{n}(0.7 - p)}{\sqrt{p(1-p)}}\right).$$

Используя Лемму Слущкого, мы можем заменить  $p$  на  $\bar{X}_n$  в знаменателе. Пусть  $n = 100$  и  $\bar{X}_n = 0.8$ , тогда

$$P_p(\varphi(X) = 1) \approx \Phi(25(p - 0.7)).$$

Например, если  $p \leq 0.65$ :

$$P_p(\text{Ошибка первого рода}) \approx \begin{cases} 0, & p = 0.5 \\ 0.006, & p = 0.6 \end{cases}$$

Вероятность ошибки ограничена сверху:

$$P_p(\text{Ошибка первого рода}) \leq P_{0.65}(\text{Ошибка первого рода}) \approx \Phi(1.25) \approx 0.106.$$

Симметрично

$$P_p(\text{Ошибка второго рода}) \approx \begin{cases} 0, & p = 0.9 \\ 0.006, & p = 0.8 \\ 0.5, & p = 0.7 \end{cases}$$

Граница сверху:

$$P_p(\text{Ошибка второго рода}) \leq P_{0.65}(\text{Ошибка второго рода}) \approx 0.894.$$

**Замечание 6.6.** В идеале, мы хотим минимизировать вероятности обеих ошибок и выбрать оптимальный критерий. Проблема заключается в том, что критерии

$$\varphi_0(X) \equiv 0 \Rightarrow \begin{cases} P_p(\text{Ошибка первого рода}) = 0 \\ P_p(\text{Ошибка второго рода}) = 1 \end{cases}$$

$$\varphi_1(X) \equiv 1 \Rightarrow \begin{cases} P_p(\text{Ошибка первого рода}) = 1 \\ P_p(\text{Ошибка второго рода}) = 0 \end{cases}$$

являются оптимальными, если нужно минимизировать вероятность одной из ошибок, но не минимизирует вероятности обеих ошибок одновременно. На практике берут границу  $\alpha$  для вероятности ошибки первого рода и минимизируют вероятность ошибки второго рода по всем критериям. Обычно  $0.01 \leq \alpha \leq 0.1$ .

**Определение 6.7.** Пусть  $\varphi$  – критерий для  $H : \varphi \in \Theta_H$  против  $K : \vartheta \in \Theta_K$ .

(i) Функция

$$\beta_\varphi : \begin{cases} \Theta & \rightarrow [0, 1] \\ \vartheta & \mapsto \mathbb{E}_\vartheta[\varphi(X)] \end{cases}$$

называется **функцией мощности**  $\varphi$ .

(ii) Критерий  $\varphi$  называется **критерием с уровнем значимости**  $\alpha \in [0, 1]$ , если

$$\beta_\varphi(\vartheta) \leq \alpha \quad \forall \vartheta \in \Theta_H$$

Зададим множество таких критериев:

$$\Phi_\alpha = \{\varphi \in \Phi \mid \varphi \text{ – критерий с уровнем значимости } \alpha\}.$$

- (iii) Критерий  $\varphi$  называется **несмещенным с уровнем значимости**  $\alpha \in [0, 1]$ , если  $\varphi \in \Phi_\alpha$  и

$$\beta_\varphi(\vartheta) \geq \alpha \quad \forall \vartheta \in \Theta_K,$$

$$\Phi_{\alpha\alpha} = \{\varphi \in \Phi_\alpha \mid \varphi - \text{несмещенный критерий}\}.$$

**Замечание 6.8.**

- (i) Если  $\varphi$  – нерандомизированный критерий, то

$$\beta_\varphi(\vartheta) = P_\vartheta(\varphi(X) = 1)$$

– вероятность принятия  $K$ . В частности,

(a)  $\vartheta \in \Theta_H$ :  $\beta_\varphi(\vartheta)$  – вероятность ошибки 1-го рода.

(b)  $\vartheta \in \Theta_K$ :  $1 - \beta_\varphi(\vartheta)$  – вероятность ошибки 2-го рода.

Аналогичная интерпретация имеет место для рандомизированных критериев.

- (ii) Критерий  $\varphi$  ограничивает вероятность ошибки первого рода уровнем значимости  $\alpha \quad \forall \vartheta \in \Theta_H$ .
- (iii) Для несмещенного критерия  $\varphi$  вероятность принятия гипотезы  $K$  в случае, если  $\vartheta \in \Theta_K$ , не меньше, чем в случае, если  $\vartheta \in \Theta_H$ .

**Пример 6.9.** Функция мощности критерия из Примера 6.3 будет приблизительно равна:

$$\beta_\varphi : \begin{cases} [0, 1] & \rightarrow [0, 1] \\ p & \mapsto \beta_\varphi(p) \approx \Phi\left(\frac{\sqrt{n}(p-0.7)}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}\right). \end{cases}$$

Уровень значимости критерия  $\varphi$ :  $\alpha \approx 0.106$ .

**Определение 6.10.**

- (i) Критерий  $\varphi^* \in \Phi_\alpha$  называется **равномерно наиболее мощным (UMP) с уровнем значимости**  $\alpha$ , если

$$\beta_{\varphi^*}(\vartheta) = \sup_{\varphi \in \Phi_\alpha} \beta_\varphi(\vartheta) \quad \forall \vartheta \in \Theta_K.$$

- (ii) Критерий  $\varphi^* \in \Phi_{\alpha\alpha}$  называется **равномерно наиболее мощным несмещенным (UMPU) с уровнем значимости**  $\alpha$ , если

$$\beta_{\varphi^*}(\vartheta) = \sup_{\varphi \in \Phi_{\alpha\alpha}} \beta_\varphi(\vartheta) \quad \forall \vartheta \in \Theta_K.$$

**Теорема 6.11.** Пусть  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  – семейство распределений на  $(\mathcal{X}, \mathcal{B})$ , статистика  $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{T}, \mathcal{D})$  – достаточная для  $\vartheta$  и  $\varphi$  – критерий. Тогда существует критерий  $\psi \circ T$ , обладающий той же функцией мощности, что и  $\varphi$ , а именно:

$$\psi \circ T = \mathbb{E}[\varphi|T].$$

ДОКАЗАТЕЛЬСТВО: Пусть  $\psi(t) = \mathbb{E}[\varphi|T = t]$ . Во-первых,  $\psi(T) \in [0, 1]$ , так как это критерий. Также,

$$\beta_{\psi \circ T}(\vartheta) = \mathbb{E}_\vartheta[\psi \circ T] = \mathbb{E}_\vartheta[\mathbb{E}[\varphi|T]] = \mathbb{E}_\vartheta[\varphi] = \beta_\varphi(\vartheta).$$

**Замечание 6.12.**

- (i) Теорема 6.11 показывает, что для создания критерия всегда можно воспользоваться достаточной статистикой.

(ii) Простейший случай простых гипотез:

$$H : \vartheta \in \{\vartheta_0\} \quad \text{против} \quad K : \vartheta \in \{\vartheta_1\}, \quad \vartheta_0 \neq \vartheta_1.$$

Зададим доминирующую меру для  $P_{\vartheta_0}$  и  $P_{\vartheta_1}$ :

$$\mu = P_{\vartheta_0} + P_{\vartheta_1}.$$

Также, пусть

$$p_i = \frac{dP_{\vartheta_i}}{d\mu}$$

и

$$\frac{p_1}{p_0} = \begin{cases} \infty, & p_1 > 0, p_0 = 0 \\ \text{произвольное}, & p_1 = 0, p_0 = 0. \end{cases}$$

Имеют место следующие утверждения:

- (a) Величина  $\frac{p_1}{p_0}$  независима от  $\mu$  (вплоть до множеств меры нуль).
- (b)  $\frac{p_1}{p_0}$  – минимальная достаточная статистика для  $\vartheta$  (Пример 4.20).
- (c) UMP-критерий с уровнем значимости  $\alpha$  максимизирует

$$\beta_{\varphi}(\vartheta_1) = \mathbb{E}_{\vartheta_1}[\varphi(X)] = \int \varphi(x)p_1(x)\mu(dx)$$

при ограничении

$$\beta_{\varphi}(\vartheta_0) = \mathbb{E}_{\vartheta_0}[\varphi(X)] = \int \varphi(x)p_0(x)\mu(dx) \leq \alpha.$$

**Определение 6.13.** В случае простой гипотезы критерий  $\varphi$  называется **критерием Неймана-Пирсона (NP критерий)**, если существует константа  $c \in [0, \infty)$ , такая что

$$\varphi(x) : \begin{cases} 1, & p_1(x) > cp_0(x), \\ 0, & p_1(x) < cp_0(x). \end{cases}$$

**Теорема 6.14 (NP лемма).** Пусть  $\Theta = \{\vartheta_0, \vartheta_1\}$  и рассматриваются гипотезы

$$H : \vartheta \in \{\vartheta_0\} \quad \text{против} \quad K : \vartheta \in \{\vartheta_1\}.$$

- (i) NP критерий  $\varphi^*$  – UMP-критерий с уровнем значимости  $\alpha = \mathbb{E}_{\vartheta_0}[\varphi^*(X)]$ .
- (ii) Для любого  $\alpha \in [0, 1]$  существует NP критерий  $\varphi$ , такой что  $\mathbb{E}_{\vartheta_0}[\varphi(X)] = \alpha$ .
- (iii) Если  $\varphi'$  – UMP-критерий с уровнем значимости  $\alpha$ , то  $\varphi'$  – (п.н.) NP критерий. Если  $\mathbb{E}_{\vartheta_0}[\varphi'(X)] < \alpha$ , то  $\mathbb{E}_{\vartheta_1}[\varphi'(X)] = 1$ .

ДОКАЗАТЕЛЬСТВО:

- (i) Пусть  $\varphi^*$  – NP критерий с константой  $c^*$  и  $\varphi$  – другой критерий, такой что  $\beta_{\varphi}(\vartheta_0) \leq \alpha = \beta_{\varphi^*}(\vartheta_0)$ . Рассмотрим

$$\beta_{\varphi^*}(\vartheta_1) - \beta_{\varphi}(\vartheta_1) = \int (\varphi^* - \varphi)p_1 d\mu = \int (\varphi^* - \varphi)(p_1 - c^*p_0) d\mu + \int c^*p_0(\varphi^* - \varphi) d\mu.$$

Второй интеграл неотрицателен, так как его подинтегральное выражение:

$$c^*(\beta_{\varphi^*}(\vartheta_0) - \beta_{\varphi}(\vartheta_0)) \geq 0.$$

В первом интеграле:

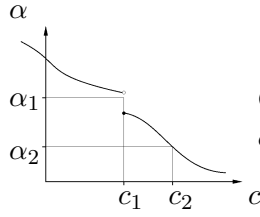
$$\varphi^* - \varphi > 0 \implies \varphi^* > 0 \implies p_1 \geq c^*p_0,$$

$$\varphi^* - \varphi < 0 \implies \varphi^* < 1 \implies p_1 \leq c^*p_0$$

$$\implies (\varphi^* - \varphi)(p_1 - c^*p_0) \geq 0 \text{ всегда.}$$



(ii) Для  $c \in \mathbb{R}$  зададим



$$\alpha(c) := P_{\vartheta_0} \left( \frac{p_1(X)}{p_0(X)} > c \right).$$

Очевидно, что  $1 - \alpha(c)$  — функция распределения. Для  $\alpha \neq 0$  выберем  $c^*$  таким образом, чтобы выполнялось соотношение

$$\alpha(c^*) \leq \alpha \leq \alpha(c^* -),$$

где  $\alpha(c-)$  обозначает левый предел в точке  $c$ . Для  $\alpha = 0$  пусть  $c^* = \infty$ . Зададим функции

$$\gamma^* : \begin{cases} \frac{\alpha - \alpha(c^*)}{\alpha(c^* -) - \alpha(c^*)}, & \text{если } \alpha(c^* -) - \alpha(c^*) > 0, \\ 0, & \text{иначе.} \end{cases}$$

и

$$\varphi^*(x) : \begin{cases} 1, & p_1(x) > c^* p_0(x), \\ \gamma^*, & p_1(x) = c^* p_0(x), \\ 0, & p_1(x) < c^* p_0(x). \end{cases}$$

Если  $c^* < \infty$ , то

$$\beta_{\varphi^*}(\vartheta_0) = \int_{\{p_1 > c^* p_0\}} dP_{\vartheta_0} + \int_{\{p_1 = c^* p_0\}} \gamma^* dP_{\vartheta_0} = \alpha(c^*) + \gamma^*(\alpha(c^* -) - \alpha(c^*)) = \alpha.$$

Если  $c^* = \infty$ , то

$$\beta_{\varphi^*}(\vartheta_0) = P_{\vartheta_0}(p_1(X) > \infty) = 0.$$

(iii) Пусть  $\varphi'$  — UMP-критерий с уровнем значимости  $\alpha$  и  $\varphi^*$  — критерий из (ii). Так как из (i) следует, что  $\varphi^*$  — UMP, то  $\beta_{\varphi^*}(\vartheta_1) = \beta_{\varphi'}(\vartheta_1)$ . Таким образом,

$$\int (\varphi^* - \varphi') p_1 d\mu = 0,$$

но

$$\int (\varphi^* - \varphi') p_1 d\mu = \int (\varphi^* - \varphi')(p_1 - c^* p_0) d\mu + \int c^* p_0 (\varphi^* - \varphi') d\mu = I + II$$

Так как  $\beta_{\varphi^*}(\vartheta_0) = \alpha \geq \beta_{\varphi'}(\vartheta_0)$ , второй интеграл неотрицателен. Как и в доказательстве (i), мы можем показать, что  $I \geq 0$ . Следовательно,  $I = II = 0$ . Зададим множество

$$S = \{x \mid \varphi'(x) \neq \varphi^*(x)\} \cap \{x \mid p_1(x) \neq c^* p_0(x)\}.$$

На множестве  $S$  имеет место неравенство

$$(\varphi^* - \varphi')(p_1 - c^* p_0) > 0$$

$\Rightarrow \mu(S) = 0$ . На его дополнении  $S^c$   $\varphi'$  — NP критерий. Так как  $II = 0$ , либо  $c^* = 0$ , либо  $\beta_{\varphi'}(\vartheta_0) = \beta_{\varphi^*}(\vartheta_0) = \alpha$ . Если  $\beta_{\varphi'}(\vartheta_0) < \alpha$ , то  $c^* = 0$  и  $\varphi^*(x) = 1$  для любого  $x$ , такого что  $p_1(x) > 0$ . Следовательно,

$$\beta_{\varphi^*}(\vartheta_1) = \int \varphi^* p_1 d\mu = \int p_1 d\mu = 1.$$

Утверждение следует из равенства  $\beta_{\varphi'}(\vartheta_1) = \beta_{\varphi^*}(\vartheta_1) = 1$ .

**Замечание 6.15.** NP критерий  $\varphi^*$  для  $H : \vartheta = \vartheta_0$  против  $K : \vartheta = \vartheta_1$  на дополнении множества  $S_{=} = \{x \mid p_1(x) = c^* p_0(x)\}$  определен единственным образом. На множестве  $S_{=}$  критерий может быть выбран таким образом, чтобы  $\beta_{\varphi^*}(\vartheta_0) = \alpha$ . Один из возможных способов показан в пункте (ii).

**Следствие 6.16.** Любой NP критерий  $\varphi^*$ , такой что  $\beta_{\varphi^*}(\vartheta_0) \in (0, 1)$  является несмещенным. В частности,

$$\alpha := \beta_{\varphi^*}(\vartheta_0) < \beta_{\varphi^*}(\vartheta_1).$$

ДОКАЗАТЕЛЬСТВО: Критерий вида  $\varphi \equiv \alpha$  имеет уровень значимости  $\alpha$ . Так как  $\varphi^*$  – UMP,  $\beta_{\varphi}(\vartheta_1) \leq \beta_{\varphi^*}(\vartheta_1)$ . Также заметим, что если  $\alpha = \beta_{\varphi^*}(\vartheta_1) < 1$ , то  $\varphi$  – UMP. Так как любой UMP-критерий является NP критерием, то имеет место равенство  $p_1(x) = c^* p_0(x)$  для почти всех  $x$ . Следовательно,  $c^* = 1$  и  $p_1 = p_0$   $\mu$ -п.н., и также  $P_{\vartheta_0} = P_{\vartheta_1}$  – мы приходим к противоречию.

**Пример 6.17.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$  с известным параметром  $\sigma^2$ . Рассмотрим гипотезы:

$$H : \mu = \mu_0 \quad \text{против} \quad K : \mu = \mu_1,$$

где  $\mu_0 < \mu_1$ . Плотность  $X_1, \dots, X_n$ :

$$p_j(x) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left( \sum_{i=1}^n X_i^2 - 2\mu_j \sum_{i=1}^n X_i + n\mu_j^2 \right) \right\}, \quad j = 0, 1.$$

Неравенство для отношения плотностей (или отношения правдоподобия), необходимое для создания NP критерия:

$$\frac{p_1(x)}{p_0(x)} = \exp \left\{ \frac{1}{\sigma^2} \sum_{i=1}^n x_i(\mu_1 - \mu_0) \right\} \cdot f(\sigma^2, \mu_1, \mu_0) > c^*,$$

где  $f(\sigma^2, \mu_1, \mu_0)$  – известная положительная константа. Это неравенство эквивалентно:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i > c,$$

где  $c$  – соответствующая константа. Таким образом, достаточно найти  $c$ , такую что

$$P_{\mu_0}(\bar{X}_n > c) = \alpha$$

или, что эквивалентно,

$$P_{\mu_0} \left( \underbrace{\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma}}_{\sim \mathcal{N}(0, 1)} > \frac{\sqrt{n}(c - \mu_0)}{\sigma} \right) = 1 - \Phi \left( \frac{\sqrt{n}(c - \mu_0)}{\sigma} \right) = \alpha.$$

Назовем величину  $u_\beta$   **$\beta$ -квантилем** стандартного нормального распределения  $\mathcal{N}(0, 1)$ , если  $\Phi(u_\beta) = \beta$ . Тогда

$$\frac{\sqrt{n}(c - \mu_0)}{\sigma} = u_{1-\alpha} \iff c = \mu_0 + u_{1-\alpha} \frac{\sigma}{\sqrt{n}},$$

и NP критерий:

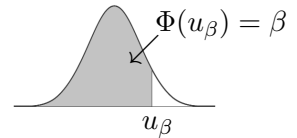
$$\varphi^*(x) = 1_{\{\bar{X}_n > \mu_0 + u_{1-\alpha} \frac{\sigma}{\sqrt{n}}\}}.$$

**Пример 6.18.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{U}[0, \vartheta]$  и мы проверяем

$$H : \vartheta = \vartheta_0 \quad \text{против} \quad K : \vartheta = \vartheta_1,$$

где  $\vartheta_0 < \vartheta_1$ . Плотности:

$$p_j(x) = \left( \frac{1}{\vartheta_j} \right)^n 1_{[0, \vartheta_j]}(x_{(n)}), \quad j = 0, 1,$$



и их отношение правдоподобия:

$$\frac{p_1(x)}{p_0(x)} = \begin{cases} \left(\frac{\vartheta_0}{\vartheta_1}\right)^n, & x_{(n)} \leq \vartheta_0, \\ \infty, & x_{(n)} \in (\vartheta_0, \vartheta_1], \\ \text{произвольное,} & \text{иначе.} \end{cases}$$

Далее

$$\alpha(c) = P_{\vartheta_0}(p_1(X) > cp_0(X)) = \begin{cases} 1, & c \leq \left(\frac{\vartheta_0}{\vartheta_1}\right)^n, \\ 0, & c > \left(\frac{\vartheta_0}{\vartheta_1}\right)^n. \end{cases}$$

Для любого  $\alpha \in (0, 1)$  имеет место

$$\alpha(c) \leq \alpha \leq \alpha(c-) \iff c = c^* = \left(\frac{\vartheta_0}{\vartheta_1}\right)^n.$$

НР лемма гласит, что

$$\varphi^*(x) = 1_{\left\{\frac{p_1(x)}{p_0(x)} > c^*\right\}} + \gamma(x)1_{\left\{\frac{p_1(x)}{p_0(x)} = c^*\right\}} = 1_{\{x_{(n)} > \vartheta_0\}} + \gamma(x)1_{\{x_{(n)} \leq \vartheta_0\}}.$$

Как выбрать  $\gamma(x)$ ? Возможностей много:

$$\begin{aligned} \varphi_1(x) &= 1_{\{x_{(n)} > c_1\}}, \quad c_1 = \vartheta_0 \sqrt[n]{1 - \alpha}, \\ \varphi_2(x) &= 1_{\{x_{(n)} > \vartheta_0\}} + 1_{\{x_{(n)} < c_2\}}, \quad c_2 = \vartheta_0 \sqrt[n]{\alpha}, \\ \varphi_3(x) &= 1_{\{x_{(n)} > \vartheta_0\}} + \alpha 1_{\{x_{(n)} \leq \vartheta_0\}}. \end{aligned}$$

**Замечание 6.19.** Простые гипотезы на практике не актуальны, но

- (i) Они дают интуитивное ощущение того, как нужно строить критерии. Во-первых, нужен т.н. **доверительный интервал**  $c(X) \subset \Theta$ , внутри которого неизвестный параметр лежит с вероятностью  $1 - \alpha$ . В Примере 6.17 мы использовали, что для  $c(X) = [\bar{X}_n - u_{1-\alpha} \frac{\sigma}{\sqrt{n}}, \infty)$  имеет место

$$P_{\mu_0}(\mu_0 \in c(X)) = P_{\mu_0}(\bar{X}_n \leq \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}) = 1 - \alpha.$$

Любой такой интервал  $c(X)$  может быть использован для построения критерия, например:

$$c'(X) = \left[ \bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right].$$

В дополнение, простые гипотезы показывают в какой стороне лежит альтернатива, и поэтому был выбран интервал  $c(X)$  в Примере 6.17.

- (ii) С помощью формальных результатов, таких как НР лемма, можно вывести более актуальные результаты.

**Определение 6.20.** Пусть  $\Theta \subset \mathbb{R}$ ,  $\mathcal{P} = \{P_{\vartheta} \mid \vartheta \in \Theta\}$  и  $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$  – статистика. Семейство  $\mathcal{P}$  называется **классом с монотонным (изотоническим) отношением правдоподобия**, если для любого  $\vartheta < \vartheta_1$  существует монотонно возрастающая функция  $H_{\vartheta_0, \vartheta_1} : \mathbb{R} \rightarrow [0, \infty)$ , такая что

$$\frac{p_{\vartheta_1}(x)}{p_{\vartheta_0}(x)} = H_{\vartheta_0, \vartheta_1}(T(x)) \quad P_{\vartheta_0} + P_{\vartheta_1}\text{-п.н.}$$

**Пример 6.21.**

- (i) В Примере 6.17

$$\frac{p_{\mu_1}(x)}{p_{\mu_0}(x)} = \exp \left\{ \frac{1}{\sigma^2} \sum_{i=1}^n x_i (\mu_1 - \mu_0) \right\} \cdot f(\sigma^2, \mu_1, \mu_0),$$

монотонно возрастает по  $\bar{X}_n$ . Это свойство может быть обобщено до однопараметрических экспоненциальных семейств.

(ii) В Примере 6.18

$$\frac{p_{\vartheta_1}(x)}{p_{\vartheta_0}(x)} = \left(\frac{\vartheta_0}{\vartheta_1}\right)^n 1_{[0, \vartheta_0]}(x_{(n)}) + \infty 1_{(\vartheta_0, \vartheta_1]}(X_{(n)})$$

монотонно возрастает по  $X_{(n)}$ .

**Теорема 6.22.** Пусть  $\mathcal{P} = \{P_{\vartheta} \mid \vartheta \in \Theta\}$  – класс с монотонным отношением правдоподобия по  $T$ ,  $\vartheta_0 \in \Theta$  и  $\alpha \in (0, 1)$ . Также, пусть

$$\varphi^*(x) = 1_{\{T(x) > c\}} + \gamma 1_{\{T(x) = c\}},$$

где

$$c := \inf\{t \mid P_{\vartheta_0}(T(X) > t) \leq \alpha\}$$

и

$$\gamma = \begin{cases} \frac{\alpha - P_{\vartheta_0}(T(X) > c)}{P_{\vartheta_0}(T(X) = c)}, & \text{если } P_{\vartheta_0}(T(X) = c) \neq 0 \\ 0, & \text{иначе.} \end{cases}$$

двусторонние гипотезы:

$$\overline{K \quad H \quad K}$$

односторонние гипотезы:

$$\overline{H \quad K}$$

Тогда

(i)  $\beta_{\varphi^*}(\vartheta_0) = \alpha$  и  $\varphi^*$  – UMP-критерий с уровнем значимости  $\alpha$  для односторонних гипотез:

$$H : \vartheta \leq \vartheta_0 \quad \text{против} \quad K : \vartheta > \vartheta_0.$$

(ii) Для любого  $\vartheta < \vartheta_0$  имеет место равенство

$$\beta_{\varphi^*}(\vartheta) = \inf\{\beta_{\varphi}(\vartheta) \mid \varphi \in \Phi \text{ и } \beta_{\varphi}(\vartheta_0) = \alpha\}.$$

(iii) Функция мощности  $\vartheta \mapsto \beta_{\varphi^*}(\vartheta)$  строго монотонно возрастает для любого  $\vartheta$ , такого что  $\beta_{\varphi^*}(\vartheta) \in (0, 1)$ .

(iv) Для любого  $\vartheta' \in \Theta$   $\varphi^*$  – UMP-критерий с уровнем значимости  $\alpha' = \mathbb{E}_{\vartheta'}[\varphi^*(X)]$  для гипотез

$$H' : \vartheta \leq \vartheta' \quad \text{против} \quad K' : \vartheta > \vartheta'.$$

ДОКАЗАТЕЛЬСТВО:

(i) Если  $P_{\vartheta_0}(T(X) = c) = 0$ , то

$$\beta_{\varphi^*}(\vartheta_0) = P_{\vartheta_0}(T(X) > c) = \alpha.$$

Если  $P_{\vartheta_0}(T(X) = c) > 0$ , то мы подбираем  $\gamma$  таким образом, чтобы имело место равенство

$$\beta_{\varphi^*}(\vartheta_0) = P_{\vartheta_0}(T(X) > c) + \gamma P_{\vartheta_0}(T(X) = c) = \alpha.$$

Пусть  $\vartheta_0 < \vartheta_1$  и  $H_{\vartheta_0, \vartheta_1}(T(x)) = p_{\vartheta_1}(x)/p_{\vartheta_0}(x)$ . Вследствие монотонности

$$H_{\vartheta_0, \vartheta_1}(T(x)) \leq H_{\vartheta_0, \vartheta_1}(c) = s \implies T(x) \leq c$$

и

$$\varphi^*(x) = \begin{cases} 1, & H_{\vartheta_0, \vartheta_1}(x) > s, \\ 0, & H_{\vartheta_0, \vartheta_1}(x) < s. \end{cases}$$

Таким образом  $\varphi^*$  – NP критерий с уровнем значимости  $\alpha$  и по NP лемме

$$\beta_{\varphi^*}(\vartheta') = \sup\{\beta_{\varphi}(\vartheta_1) \mid \varphi \in \Phi \text{ и } \beta_{\varphi}(\vartheta_0) = 1 - \alpha\}.$$

Так как  $\varphi^*$  не зависит от выбора  $\vartheta_1$ , это соотношение имеет место для любого  $\vartheta_1 > \vartheta_0$ . Наконец, пусть  $\varphi'(x) = 1 - \varphi^*(x)$ . Используя те же рассуждения, что и выше, можно показать, что

$$\beta_{\varphi'}(\vartheta_2) = \sup\{\beta_{\varphi}(\vartheta_2) \mid \varphi \in \Phi \text{ и } \beta_{\varphi}(\vartheta_0) = \alpha\} \quad \forall \vartheta_2 < \vartheta_0.$$

Так как критерий вида  $\bar{\varphi} \equiv \alpha$  удовлетворяет равенству  $\beta_{\bar{\varphi}}(\vartheta_0) = \alpha$ , мы заключаем, что

$$1 - \beta_{\varphi^*}(\vartheta_2) = \beta_{\varphi'}(\vartheta_2) \geq \beta_{1-\bar{\varphi}}(\vartheta_2) = 1 - \beta_{\bar{\varphi}}(\vartheta_2) = 1 - \alpha.$$

Следовательно,  $\beta_{\varphi^*}(\vartheta_2) \leq \alpha$  и  $\varphi^* \in \Phi_{\alpha}$ .

- (ii) Утверждение следует непосредственно, так как  $\beta_{\varphi'} = 1 - \beta_{\varphi^*}$ .
- (iii) Следует из Следствия 6.16, так как для любых  $\vartheta_1 < \vartheta_2$   $\varphi^*$  – НР критерий.
- (iv) Доказывается, используя схожие аргументы, что и в доказательстве пункта (i).

**Пример 6.23.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$  с известным параметром  $\sigma^2$ . Из Примера 6.17 мы знаем, что плотности

$$p_\mu(x) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\}$$

обладают монотонным отношением правдоподобия по  $T(x) = \bar{x}_n$ . Из Теоремы 6.22: UMP-критерий с уровнем значимости  $\alpha$  для

$$H: \mu \leq \mu_0 \quad \text{против} \quad K: \mu > \mu_0$$

имеет вид

$$\varphi^*(x) = 1_{\{\bar{x}_n > c\}} + \gamma 1_{\{\bar{x}_n = c\}}.$$

Так как  $P_{\mu_0}(T(X) = c) = 0$ , то  $\gamma = 0$  и выбрать  $c$  так, что  $P_{\mu_0}(\bar{X}_n > c) = \alpha \iff c = \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}$ . UMP-критерий

$$\varphi^*(x) = 1_{\{\bar{X}_n > \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}\}}$$

называется *односторонним критерием Гаусса*.

#### Замечание 6.24.

- (i) Существует эвристика, как получить односторонний критерий Гаусса: так как  $\bar{X}_n$  UMVU-оценка для  $\mu$ , разумной стратегией будет принятие гипотезы  $K$ , если  $\bar{X}_n$  достаточно большое. Следовательно, критерий должен иметь вид:

$$\varphi(x) = 1_{\{\bar{X}_n > c\}}.$$

Выбираем  $c$ , контролируя вероятность ошибки первого рода. Для любого  $\mu \leq \mu_0$  имеет место

$$\beta_\varphi(\mu) = P_\mu(\bar{X}_n > c) = P_\mu\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} > \frac{\sqrt{n}(c - \mu)}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}(c - \mu)}{\sigma}\right) \leq 1 - \Phi\left(\frac{\sqrt{n}(c - \mu_0)}{\sigma}\right).$$

Мы должны удостовериться, что:

$$1 - \Phi\left(\frac{\sqrt{n}(c - \mu_0)}{\sigma}\right) \leq \alpha,$$

иначе говоря:

$$c \geq \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}.$$

Мы берем  $c = \mu_0 + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}$ , чтобы вероятность ошибки первого рода была равна  $\alpha$ .

- (ii) Этот метод позволяет построить критерий, однако ничего не говорит о его оптимальности. Что важно, он может быть применен в более общих ситуациях, например в случае неизвестного параметра  $\sigma^2$ . В такой ситуации можно воспользоваться оценкой

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Как и выше, мы получаем:

$$\beta_\varphi(\mu) = P_\mu\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\hat{\sigma}_n} > \frac{\sqrt{n}(c - \mu)}{\hat{\sigma}_n}\right) = 1 - F_{t_{n-1}}\left(\frac{c - \mu}{\sqrt{\hat{\sigma}_n^2/n}}\right)$$

из Следствия 2.21, где  $F_{t_{n-1}}$  – функция распределения  $t_{n-1}$ . Разумный выбор константы:

$$c = \mu_0 + \frac{\hat{\sigma}_n}{\sqrt{n}} t_{n-1, 1-\alpha},$$

где  $t_{n-1, 1-\alpha}$  –  $1-\alpha$ -квантиль распределения Стьюдента с  $n-1$  степенями свободы. Критерий

$$1_{\{\bar{x}_n > \mu_0 + \frac{\hat{\sigma}_n}{\sqrt{n}} t_{n-1, 1-\alpha}\}}$$

называется *односторонним  $t$ -критерием*.

**Замечание 6.25.** В общем случае не существует УМР-критериев для

$$H : \vartheta = \vartheta_0 \quad \text{против} \quad K : \vartheta \neq \vartheta_0,$$

так как этот критерий должен быть оптимальным для всех

$$H' : \vartheta = \vartheta_0 \quad \text{против} \quad K' : \vartheta = \vartheta_1,$$

где  $\vartheta_0 \neq \vartheta_1$ . В случае монотонного отношения правдоподобия оптимальный критерий будет

$$\varphi(x) = 1_{\{T(x) > c\}} + \gamma(x) 1_{\{T(x) = c\}}$$

для  $\vartheta_1 > \vartheta_0$  и

$$\varphi'(x) = 1_{\{T(x) < c'\}} + \gamma'(x) 1_{\{T(x) = c'\}}$$

для  $\vartheta_1 < \vartheta_0$ , что невозможно.

**Теорема 6.26.** Пусть  $\mathcal{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$  – однопараметрическое экспоненциальное семейство с  $\mu$ -плотностью

$$p_\vartheta(x) = c(\vartheta) h(x) \exp(Q(\vartheta)T(x))$$

с возрастающей функцией  $Q$ . Тогда существует УМРУ-критерий для

$$H : \vartheta \in [\vartheta_1, \vartheta_2] \quad \text{против} \quad K : \vartheta \notin [\vartheta_1, \vartheta_2],$$

а именно

$$\varphi^*(x) = \begin{cases} 1, & \text{если } T(x) \notin [c_1, c_2], \\ \gamma_i, & \text{если } T(x) = c_i, \\ 0, & \text{если } T(x) \in (c_1, c_2), \end{cases}$$

где константы  $c_i$ ,  $\gamma_i$  определяются из равенства

$$\beta_\varphi(\vartheta_1) = \beta_\varphi(\vartheta_2) = \alpha$$

ДОКАЗАТЕЛЬСТВО: Теорема 3.7.1 в [4].

**Замечание 6.27.** Похожие утверждения имеют место для  $k$ -параметрических экспоненциальных семейств.

## Упражнения

**6.1.** Пусть  $X_1, \dots, X_n$  – i.i.d. нормально распределенные случайные величины:  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  с известным математическим ожиданием  $\mu \in \mathbb{R}$  и неизвестной дисперсией  $\sigma^2 > 0$ . Постройте UMP-критерий с уровнем значимости  $\alpha \in (0, 1)$  для гипотез

$$H : \sigma^2 < \sigma_0^2 \quad \text{против} \quad K : \sigma \geq \sigma_0^2.$$

**6.2.** Пусть мы наблюдаем  $X \in (0, 1)$ . Постройте UMP-критерий с уровнем значимости  $\alpha$  для

$$H : \text{плотность } X: f(x) = 4x1_{(0,1/2)}(x) + (4 - 4x)1_{(1/2,1)}(x)$$

против альтернативы

$$K : X \sim \mathcal{U}(0, 1).$$

**6.3.** Пусть  $X$  – случайная величина с плотностью

$$f_{\vartheta}(x) = \frac{2(\vartheta - x)}{\vartheta^2} 1_{(0, \vartheta)}(x).$$

Постройте UMP-критерий с уровнем значимости  $\alpha$  для гипотез

$$H : \vartheta = \vartheta_0 \quad \text{против} \quad K : \vartheta = \vartheta_1$$

для  $\vartheta_1 < \vartheta_0$ .

**6.4.** Пусть  $X_1, \dots, X_n$  – i.i.d. экспоненциально распределенные случайные величины:  $X_i \sim \text{Exp}(\vartheta)$ , т.е. плотность  $X_i$ :

$$f_{\vartheta}(x) = \vartheta \exp(\vartheta x) 1_{[0, \infty)}(x).$$

(i) Покажите, что распределение  $X = (X_1, \dots, X_n)$  обладает монотонным отношением правдоподобия.

(ii) Постройте UMP-критерий для гипотез

$$H : \vartheta < \vartheta_0 \quad \text{против} \quad K : \vartheta \geq \vartheta_0.$$

(iii) Рассчитайте критическую область для этого критерия для  $\vartheta_0 = 1$ ,  $n = 10$  и  $\alpha = 0.05$ .

*Подсказка:* 5%-квантиль гамма распределения  $\Gamma(10, 1)$  приблизительно равен 5.43.

**6.5.** Рассмотрим еще раз ситуацию из Упражнения 6.4.

(i) Рассмотрим гипотезы

$$H : \vartheta = \vartheta_0 \quad \text{против} \quad K : \vartheta \neq \vartheta_0.$$

Обрисуйте построение доверительного интервала  $[\bar{X}_n - \alpha, \bar{X}_n + \alpha]$  с уровнем значимости  $\alpha$  для  $\gamma(\vartheta) = 1/\vartheta$ . Как он поможет для построения критерия для гипотез выше?

(ii) Используйте аппроксимацию нормальным распределением, чтобы построить доверительный интервал из (i) для  $n = 100$  и  $\alpha = 0.1$ .

**6.6.** Продолжим рассмотрение Упражнения 6.4.

(i) Пусть  $n = 1$ . Постройте UMPU-критерий с уровнем значимости  $\alpha$  для гипотез

$$H : \vartheta \in [1, 2] \quad \text{против} \quad K : \vartheta \notin [1, 2].$$

(ii) Теоретически мы можем воспользоваться Теоремой 6.26, чтобы построить UMPU-критерий для гипотез

$$H : \vartheta = \vartheta_0 \quad \text{против} \quad K : \vartheta \neq \vartheta_0.$$

Какие проблемы могут возникнуть? Как их можно обойти?





## Глава 7

# Асимптотические свойства критериев

В этой главе пусть  $X^{(n)} = (X_1, \dots, X_n)^T$  – вектор с распределением  $\mathcal{P}^n = \{P_{\vartheta}^n \mid \vartheta \in \Theta\}$ . Также, пусть функция

$$\varphi_n : \begin{cases} \mathcal{X}_n & \rightarrow [0, 1] \\ x^{(n)} & \mapsto \varphi_n(x^{(n)}) \end{cases}$$

будет критерием для:

$$H : \vartheta \in \Theta_H \quad \text{против} \quad K : \vartheta \in \Theta_K.$$

### Определение 7.1.

- (i) Последовательность  $(\varphi_n)$  имеет **асимптотический уровень значимости**  $\alpha$ , если

$$\lim_{n \rightarrow \infty} \sup_{\vartheta \in \Theta_H} \beta_{\varphi_n}(\vartheta) \leq \alpha.$$

- (ii) Последовательность  $(\varphi_n)$  называется **состоятельной**, если

$$\lim_{n \rightarrow \infty} \beta_{\varphi_n}(\vartheta) = 1 \quad \forall \vartheta \in \Theta_K.$$

**Замечание 7.2.** Чтобы критерий имел смысл, оба свойства должны соблюдаться: уровень значимости должен быть по меньшей мере асимптотическим и с возрастанием объема выборки вероятность ошибки второго рода должна уменьшаться.

**Пример 7.3.** Пусть  $X_1, \dots, X_m$  i.i.d.  $\sim \mathcal{N}(\mu_1, \sigma^2)$  и  $Y_1, \dots, Y_n$  i.i.d.  $\sim \mathcal{N}(\mu_2, \tau^2)$  – две независимые выборки. Мы хотим проверить гипотезы:

$$H : \mu_1 \leq \mu_2 \quad \text{против} \quad K : \mu_1 > \mu_2.$$

Мы принимаем  $K$ , если  $\bar{Y}_n$  “намного меньше”, чем  $\bar{X}_m$ .

- (i) Допустим,  $\sigma^2 = \tau^2$ , но дисперсия неизвестна. Из Леммы 2.20 мы знаем, что:

$$\bar{X}_m - \bar{Y}_n = \mathcal{N}\left(\mu_1 - \mu_2, \sigma^2 \left(\frac{1}{m} + \frac{1}{n}\right)\right)$$

и

$$\hat{\sigma}_{m,n}^2 = \frac{1}{m+n-2} \left( \sum_{i=1}^m (X_i - \bar{X}_m)^2 + \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \right) \sim \frac{\sigma^2}{m+n-2} \chi_{m+n-2}^2.$$

Если  $\mu_1 = \mu_2$ , то

$$T_{m,n} = \sqrt{\frac{mn}{m+n}} \frac{\bar{X}_m - \bar{Y}_n}{\hat{\sigma}_{m,n}} \sim t_{m+n-2},$$

таким образом критерий уровня значимости  $\alpha$ :

$$\varphi_{m,n}(x) = 1_{\{T_{m,n} > t_{m+n-2, 1-\alpha}\}}.$$

Такой критерий называется **двухвыборочным  $t$ -критерием**.

(ii) Допустим,  $\sigma^2 \neq \tau^2$ . Тогда:

$$\bar{X}_m - \bar{Y}_n = \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma^2}{m} + \frac{\tau^2}{n}\right).$$

Оценка дисперсии:

$$\hat{s}_{m,n}^2 = \frac{1}{m} \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_m)^2 + \frac{1}{n} \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2.$$

Распределение случайной величины

$$T_{m,n}^* = \frac{\bar{X}_m - \bar{Y}_n}{\hat{s}_{m,n}}$$

неизвестно (проблема Беренса-Фишера). Из ЦПТ мы знаем, что

$$\frac{\bar{X}_m - \bar{Y}_n - (\mu_1 - \mu_2)}{\hat{s}_{m,n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

если  $m \rightarrow \infty$ ,  $n \rightarrow \infty$  и  $\frac{m}{n} \rightarrow \lambda \in (0, \infty)$ . Пусть

$$\varphi_{m,n}^*(x) = 1_{\{T_{m,n}^* > u_{1-\alpha}\}},$$

тогда:

$$\beta_{\varphi_{m,n}^*}(\mu_1, \mu_2) = P_{\mu_1, \mu_2}(T_{m,n}^* > u_{1-\alpha}) = P_{\mu_1, \mu_2}\left(\frac{\bar{X}_m - \bar{Y}_n - (\mu_1 - \mu_2)}{\hat{s}_{m,n}} > \frac{-(\mu_1 - \mu_2)}{\hat{s}_{m,n}} + u_{1-\alpha}\right)$$

$$\xrightarrow{m \rightarrow \infty, n \rightarrow \infty, \frac{m}{n} \rightarrow \lambda} \begin{cases} 0, & \mu_1 < \mu_2, \\ \alpha, & \mu_1 = \mu_2, \\ 1, & \mu_1 > \mu_2. \end{cases}$$

Мы видим, что  $\varphi_{m,n}^*$  состоятельная и имеет асимптотический уровень значимости  $\alpha$ .

**Замечание 7.4.** Общий принцип построения критериев для

$$H : \vartheta \in \Theta_H \quad \text{против} \quad K : \vartheta \in \Theta_K$$

– это **метод отношения правдоподобия**. Допустим,  $f_n(x^{(n)}, \vartheta)$  – плотность  $P_{\vartheta}^n$  по некоторой мере  $\mu$ . Тогда **отношение правдоподобия**:

$$\lambda(x^{(n)}) = \frac{\sup_{\vartheta \in \Theta_H} f_n(x^{(n)}, \vartheta)}{\sup_{\vartheta \in \Theta} f_n(x^{(n)}, \vartheta)}$$

и **критерий отношения правдоподобия**:

$$\varphi_n(x^{(n)}) = 1_{\{\lambda(x^{(n)}) < c\}}.$$

Как правило,  $c$  выбирается таким образом, чтобы:

$$\sup_{\vartheta \in \Theta_H} P_{\vartheta}(\lambda(X^{(n)}) < c) \leq \alpha.$$

Распределение  $\lambda(X^{(n)})$ , тем не менее, может быть оценено лишь асимптотически.

**Условия 7.5.**

(i) Допустим  $\Theta \subset \mathbb{R}^d$  и существуют  $\Delta \subset \mathbb{R}^c$  и  $h : \Delta \rightarrow \Theta$ , такие что

$$(a) \quad \Theta_H = h(\Delta),$$

(b)  $h \in C^2(\Delta, \Theta)$ ,

(c) Матрица Якоби  $h$  – матрица полного ранга.

(ii) Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim P_\vartheta$  и в обоих семействах  $\mathcal{P}_\vartheta = \{P_\vartheta \mid \vartheta \in \Theta\}$  и  $\mathcal{P}_h = \{P_{h(\eta)} \mid \eta \in \Delta\}$  условия Теоремы 5.19 соблюдаются.

**Пример 7.6.** Пусть  $X_1, \dots, X_n$  i.i.d.  $\sim \mathcal{N}(\mu_1, \sigma^2)$  и  $Y_1, \dots, Y_n$  i.i.d.  $\sim \mathcal{N}(\mu_2, \sigma^2)$  – независимые выборки. Допустим, мы хотим проверить эквивалентность математических ожиданий:

$$H : \mu_1 = \mu_2 \quad \text{против} \quad K : \mu_1 \neq \mu_2.$$

Тогда  $\Theta \in \mathbb{R}^2 \times \mathbb{R}^+$ ,  $\Delta = \mathbb{R} \times \mathbb{R}^+$  и

$$h : \begin{cases} \Delta & \rightarrow \Theta \\ (\mu, \sigma^2)^T & \mapsto (\mu, \mu, \sigma^2)^T. \end{cases}$$

Матрица Якоби  $h$ :

$$D = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

– матрица полного ранга.

**Теорема 7.7.** Если условия 7.5 соблюдены, то

$$T_n = -2 \log \lambda(X^{(n)}) = 2(\log f_n(X^{(n)}, \hat{\theta}_n) - \log f_n(X^{(n)}, h(\hat{\eta}_n))) \xrightarrow{\mathcal{L}} \chi_{d-c}^2,$$

если  $\vartheta \in \Theta_H$ .

**ДОКАЗАТЕЛЬСТВО:** Как и прежде пусть

$$\ell(x, \vartheta) = \log f(x, \vartheta),$$

где  $f$  – функция плотности  $X_1$ . Сначала рассмотрим:

$$\begin{aligned} T_n^{(1)} &= 2(\log f_n(X^{(n)}, \hat{\theta}_n) - \log f_n(X^{(n)}, \vartheta)) \\ &= 2 \sum_{i=1}^n \left( \ell(X_i, \hat{\theta}_n) - \ell(X_i, \vartheta) \right) \\ &= 2(\hat{\theta}_n - \vartheta)^T \sum_{i=1}^n \dot{\ell}(X_i, \vartheta) + (\hat{\theta}_n - \vartheta)^T \sum_{i=1}^n \ddot{\ell}(X_i, \tilde{\vartheta}_n)(\hat{\theta}_n - \vartheta) \\ &= 2(\hat{\theta}_n - \vartheta)^T \left( \sum_{i=1}^n \dot{\ell}(X_i, \vartheta) + \sum_{i=1}^n \ddot{\ell}(X_i, \tilde{\vartheta}_n)(\hat{\theta}_n - \vartheta) \right) - (\hat{\theta}_n - \vartheta)^T \sum_{i=1}^n \ddot{\ell}(X_i, \tilde{\vartheta}_n)(\hat{\theta}_n - \vartheta) \end{aligned}$$

для некоторой  $\tilde{\theta}_n$  между  $\hat{\theta}_n$  и  $\vartheta$ . Используя обозначения из Теоремы 5.19, запишем первое слагаемое в виде

$$2n(\hat{\theta}_n - \vartheta)^T \underbrace{(\dot{L}_n(\vartheta) - \ddot{L}_n(\vartheta)(\hat{\theta}_n - \vartheta))}_{=0}$$

Также по Теореме 5.19:

$$T_n^{(1)} = -\sqrt{n}(\hat{\theta}_n - \vartheta)^T \ddot{L}_n(\tilde{\vartheta}_n) \sqrt{n}(\hat{\theta}_n - \vartheta),$$

где

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \vartheta)^T &\xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}(f(\cdot, \vartheta))), \\ \ddot{L}_n(\tilde{\vartheta}_n) &\xrightarrow{\mathbb{P}} -I(f(\cdot, \vartheta)), \\ \sqrt{n}(\hat{\theta}_n - \vartheta) &\xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}(f(\cdot, \vartheta))). \end{aligned}$$

Если  $X \sim \mathcal{N}_d(0, \Sigma)$  и  $\Sigma > 0$ , то

$$X^T \Sigma X \sim \mathcal{X}_d^2.$$

Следовательно,  $T_n^{(1)} \xrightarrow{\mathcal{L}} A = \mathcal{X}_d^2$ . Таким же образом,

$$T_n^{(2)} = 2(\log f_n(X^{(n)}, h(\hat{\eta}_n)) - \log f_n(X^{(n)}, h(\eta))) \xrightarrow{\mathcal{L}} B = \mathcal{X}_c^2.$$

Если выполняется  $H$ , то  $\vartheta = h(\eta)$  и

$$T_n = T_n^{(1)} - T_n^{(2)} \xrightarrow{\mathcal{L}} A - B = \mathcal{X}_{d-c}^2,$$

так как  $A$  и  $B$  независимы.

### Замечание 7.8.

(i) Теорема 7.7 показывает, что

$$\varphi_n(X^{(n)}) = \begin{cases} 1, & -2 \log \lambda(X^{(n)}) > \mathcal{X}_{d-c, 1-\alpha}^2, \\ 0, & \text{иначе} \end{cases}$$

является критерием с асимптотическим уровнем  $\alpha$  для

$$H : \vartheta \in \Theta_H \quad \text{против} \quad K : \vartheta \in \Theta_K.$$

(ii) Последовательность  $(\vartheta_n)$  состоятельная, так как

$$\begin{aligned} -\frac{2}{n} \log(\lambda(X^{(n)})) &= \frac{2}{n} \sum_{i=1}^n \left( \ell(X_i, \hat{\theta}_n) - \ell(X_i, h(\hat{\eta}_n)) \right) \\ &\xrightarrow{Q_\vartheta} 2\mathbb{E}_\vartheta[\ell(X, \vartheta) - \ell(X, h(\eta))] \\ &= 2KL(\vartheta|h(\eta)) > 0, \end{aligned}$$

если  $\vartheta \neq h(\eta)$  (если  $\vartheta \in \Theta_K$ ). Следовательно,

$$-2 \log(\lambda(X^{(n)})) \xrightarrow{Q_\vartheta} \infty.$$

**Пример 7.9 (Критерий Бартлетта).** Пусть  $X_{ij} \sim \mathcal{N}(\mu_i, \sigma_i^2)$ ,  $i = 1, \dots, r$  и  $j = 1, \dots, n_i$ , где  $n_i \rightarrow \infty$  с одинаковой скоростью. Мы проверяем равенство дисперсий:

$$H : \sigma_1^2 = \dots = \sigma_r^2 \quad \text{против} \quad K : \sigma_i^2 \neq \sigma_j^2 \text{ для некоторых } i \neq j.$$

Здесь  $\Theta = \mathbb{R}^r \times (\mathbb{R}^+)^r$ ,  $\Delta = \mathbb{R}^r \times \mathbb{R}^+$  и

$$h((x_1, \dots, x_r, y)^T) = (x_1, \dots, x_r, y, \dots, y)^T.$$

Оценка максимального правдоподобия:

$$\hat{\theta}_n = (\mu_1, \dots, \mu_r, \hat{s}_1^2, \dots, \hat{s}_r^2),$$

где  $\mu_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} =: \bar{X}_i$  и  $\hat{s}_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$ . В этом случае

$$f_n(X^{(n)}, \hat{\vartheta}_n) = \prod_{i=1}^r (2\pi e \hat{s}_i^2)^{-\frac{n_i}{2}}.$$

Если нулевая гипотеза верна, то оценка МП максимизирует

$$f_n(X^{(n)}, \hat{\eta}_n) = \prod_{i=1}^r (2\pi \sigma^2)^{-\frac{n_i}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \right\}.$$

Задав  $n = \sum_{i=1}^r n_i$ , получаем

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - X_{i\cdot})^2 = \sum_{i=1}^r \frac{n_i}{n} \hat{s}_i^2.$$

Тогда

$$f_n(X^{(n)}, \hat{\eta}_n) = \prod_{i=1}^r (2\pi e \hat{\sigma}^2)^{-\frac{n_i}{2}} = (2\pi e \hat{\sigma}^2)^{-\frac{n}{2}}$$

и тестовая статистика:

$$T_n = -2 \log \lambda(X^{(n)}) = n \log \hat{\sigma}^2 - \sum_{i=1}^r n_i \log \hat{s}_i^2.$$

Критерий Бартлетта:

$$\varphi_n(X^{(n)}) = \begin{cases} 1, & T_n > \chi_{r-1, 1-\alpha}^2, \\ 0, & \text{иначе.} \end{cases}$$

**Пример 7.10 (Критерий независимости).** Даны две статистические характеристики  $A$  и  $B$  (пол, возраст, образование, доход), где  $A$  состоит из  $r$  факторов и  $B$  – из  $s$  факторов. Всего наблюдается  $n$  лиц.

		Фактор $B$		
		1	...	s
Фактор $A$	1	$X_{11}$	...	$X_{1s}$
	...	...	...	...
	$r$	$X_{r1}$	...	$X_{rs}$
	Сумма	$X_{\cdot 1}$	...	$X_{\cdot s}$
		Сумма		
		$X_{1\cdot} = \sum_{j=1}^s X_{1j}$		
		...		
		$X_{r\cdot}$		
		$n$		

Модель: мультиномиальное распределение

$$(X_1, \dots, X_n)^T \sim \mathcal{M}(n, p_{11}, \dots, p_{rs}),$$

где  $\sum_{ij} p_{ij} = 1$ . Плотность распределения:

$$f_n(x^{(n)}, p) = P_p(X_{ij} = x_{ij}) = \frac{n!}{\prod_{i,j=1}^{r,s} x_{ij}!} \prod_{i,j=1}^{r,s} (p_{ij})^{x_{ij}},$$

где  $x_{ij} = \{0, \dots, n\}$  и  $\sum_{i,j=1}^{r,s} x_{ij} = n$ . Оценка максимального правдоподобия:

$$\hat{p}_{ij} = \frac{X_{ij}}{n}$$

(по аналогии с биномиальным распределением) и

$$f_n(X^{(n)}, \hat{p}) = \frac{n!}{\prod_{i,j=1}^{r,s} X_{ij}!} \prod_{i,j=1}^{r,s} \left( \frac{X_{ij}}{n} \right)^{X_{ij}}.$$

Допустим, мы хотим проверить независимость между характеристиками:

$$H : p_{ij} = p_i q_j \quad \forall i, j \quad \text{против} \quad K : p_{ij} \neq p_i q_j \quad \text{для некоторых } i \neq j,$$

где  $p_i = p_{i\cdot} = \sum_{j=1}^s p_{ij}$  и  $q_j = p_{\cdot j} = \sum_{i=1}^r p_{ij}$ . Здесь  $d = rs - 1$ ,  $c = r + s - 2$  и  $d - c = (r - 1)(s - 1)$ . Если нулевая гипотеза выполняется, то

$$f_n(X^{(n)}, p, q) = \frac{n!}{\prod_{i,j=1}^{r,s} X_{ij}!} \prod_{i,j=1}^{r,s} (p_i q_j)^{X_{ij}} = \frac{n!}{\prod_{i,j=1}^{r,s} X_{ij}!} \prod_i p_i^{X_{i\cdot}} \prod_j q_j^{X_{\cdot j}}.$$

МП оценки:

$$\hat{p}_i = \frac{X_{i\cdot}}{n} \quad \text{и} \quad \hat{q}_j = \frac{X_{\cdot j}}{n}$$

и функция правдоподобия:

$$f_n(X^{(n)}, \hat{p}, \hat{q}) = \frac{n!}{\prod_{i,j=1}^{r,s} X_{ij}!} \prod_{i,j=1}^{r,s} \left( \frac{X_{i\cdot} X_{\cdot j}}{n^2} \right)^{X_{ij}}.$$

Мы получаем:

$$T_n = -2 \log \lambda(X^{(n)}) = 2 \sum_{i=1}^r \sum_{j=1}^s X_{ij} \log \left( \frac{X_{ij}}{X_{i\cdot} X_{\cdot j}} \right)$$

и **критерий независимости**  $\chi^2$ :

$$\varphi_n(X^{(n)}) = \begin{cases} 1, & T_n > \chi_{(r-1)(s-1), 1-\alpha}^2, \\ 0, & \text{иначе.} \end{cases}$$

Используя разложение Тейлора (до второго порядка) и Закон Больших Чисел, получаем асимптотический эквивалент:

$$\tilde{T}_n = \sum_{i=1}^r \sum_{j=1}^s \frac{\left( X_{ij} - \frac{X_{i\cdot} X_{\cdot j}}{n} \right)^2}{X_{i\cdot} X_{\cdot j}} n.$$

Обычно

$$V_n^2 = \frac{\tilde{T}_n}{n(\min(r, s) - 1)} \xrightarrow{\mathbb{P}} \frac{\sum_{i=1}^r \sum_{j=1}^s \frac{(p_{ij} - p_{i\cdot} p_{\cdot j})}{p_{i\cdot} p_{\cdot j}}}{\min(r, s) - 1}$$

используется в качестве меры зависимости между  $A$  и  $B$ . Например, рассмотрим следующие характеристики:

		Годовой доход				
		1	2	3	4	Сумма
Число детей	0	2161	3577	2184	1636	9558
	1	2755	5081	2222	1052	11110
	2	936	1753	640	306	3635
	3	225	419	96	38	778
	4	39	98	31	14	182
Сумма		6116	10928	5173	3046	25263

Тогда

$$\tilde{T}_n = 568.566 \quad \text{и} \quad \chi_{12, 0.95}^2 = 21.026$$

и гипотеза о независимости принимается с уровнем значимости 5%. Тем не менее,  $V_n = 0.087$ , что показывает слабую зависимость.

## Упражнения

**7.1.** Пусть  $X_1, \dots, X_n$  – i.i.d. случайные величины с распределением Бернулли:  $X_i \sim \text{Bin}(1, p)$ . Мы проверяем гипотезы

$$H : p \leq p_0 \quad \text{против} \quad K : p > p_0,$$

где  $p_0 \in (0, 1)$ .

- (i) Докажите, что последовательность критериев с уровнем значимости  $\alpha \in (0, 1)$ , полученных из аппроксимации нормальным распределением:

$$\varphi_n(x) = 1 \quad \Leftrightarrow \quad \bar{x}_n > p_0 \sqrt{\frac{p_0(1-p_0)}{n}} u_{1-\alpha},$$

имеет асимптотический уровень значимости  $\alpha$ .

- (ii) Покажите, что последовательность  $(\varphi_n)$  – состоятельная.

**7.2.** В Примере 7.6.

- (i) Найдите оценку максимального правдоподобия для  $\Delta$  и  $\Theta$ .  
(ii) Рассчитайте  $T = -2 \log \lambda(Z)$ , где  $Z = (X_1, \dots, X_m, Y_1, \dots, Y_n)$

**7.3.** Пусть  $Y_1, \dots, Y_n$  независимые случайные величины, такие что

$$Y_i = ax_i + \varepsilon_i,$$

где  $a$  – неизвестный параметр,  $x_1, \dots, x_n$  фиксированы,  $\varepsilon_1, \dots, \varepsilon_n$  – i.i.d. и  $\varepsilon_i \sim \mathcal{N}(0, 1)$ .

- (i) Найдите оценку максимального правдоподобия  $\hat{a}$  для  $a$ .  
(ii) Найдите критерий отношения правдоподобия для

$$H : a = 0 \quad \text{против} \quad K : a \neq 0.$$

**7.4.** Пусть  $X_1, \dots, X_n$  – i.i.d. случайные величины с гамма-распределением:  $X_i \sim \Gamma(\alpha, \beta)$ , т.е. их плотность:

$$f_{\alpha, \beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$$

где  $\alpha, \beta > 0$ .

- (i) Найдите оценку максимального правдоподобия  $\hat{b}$  для  $b$ .  
(ii) Найдите критерий отношения правдоподобия для

$$H : b = b_0 \quad \text{против} \quad K : b \neq b_0.$$

**7.5.** Пусть  $(X_i)_{i \in \mathbb{N}}$  – последовательность i.i.d. случайных величин с неизвестной плотностью  $p$ . Мы хотим протестировать

$$H : p = p_0 \quad \text{против} \quad K : p = p_1,$$

где  $\{x \in \mathbb{R} \mid p_1(x) > 0\} \subset \{x \in \mathbb{R} \mid p_0(x) > 0\}$ . Используя лемму Неймана-Пирсона, УМП-критерий с уровнем значимости  $\alpha$  определяется критической областью  $\prod_{i=1}^n r(X_i) \geq C_n(\alpha)$ , где  $r(x) = p_1(x)/p_0(x)$  и  $C_n(\alpha)$  – константа, зависящая от  $n$  и  $\alpha$ .

- (i) Докажите, что критическая область может быть записана в виде:

$$\frac{1}{\sqrt{n}} \left( \sum_{i=1}^n \log r(X_i) - \mathbb{E}_{p_0}[\log r(X_i)] \right) \geq k_n(\alpha)$$

для подходящей  $k_n(\alpha)$ .

- (ii) Покажите, что

$$k_n \rightarrow \sqrt{\text{Var}_{p_0}(\log r(X_i))} u_{1-\alpha}.$$

- (iii) Используя Лемму 5.16, покажите, что последовательность критериев состоятельная для  $p_1 \neq p_0$ .





## Глава 8

# Линейная модель

**Пример 8.1 (Линейная регрессия).** Предположим, что  $X$  и  $Y$  связаны следующим соотношением:

$$Y = b_0 + b_1 X,$$

и мы хотим оценить значения  $b_0$  и  $b_1$ . На практике рассматривается не строгая линейная зависимость, а соотношение вида

$$Y_i = b_0 + b_1 X_i + \varepsilon_i \quad \forall i = 1, \dots, n,$$

где  $\varepsilon_i$  – ошибка, такая что  $\mathbb{E}[\varepsilon_i] = 0$  и  $\text{Var}(\varepsilon_i) = \sigma^2 > 0$ . В векторной записи:  $Y = Xb + \varepsilon$  или

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

**Пример 8.2 (Анализ дисперсии).** Рассмотрим эксперимент: разделим  $n$  животных на  $a$  групп по типу питания. В каждой  $i$ -й группе  $n_i$  животных. Смоделируем вес каждого животного:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \dots, a, \quad j = 1, \dots, n_i.$$

В векторной записи:  $Y = X\mu + \varepsilon$  или

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \vdots \\ Y_{a1} \\ \vdots \\ Y_{an_a} \end{pmatrix} = \begin{pmatrix} \mathbb{1}_{n_1} & 0 & \dots & 0 \\ 0 & \mathbb{1}_{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbb{1}_{n_a} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_a \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \vdots \\ \varepsilon_{a1} \\ \vdots \\ \varepsilon_{an_a} \end{pmatrix}$$

Вопрос: как оценить  $\mu_i$  и протестировать  $\mu_1 = \dots = \mu_a$ ?

**Определение 8.3.** Пусть  $X \in \mathbb{R}^{n \times k}$  и  $b \in \mathbb{R}^k$ ,  $n > k$  и пусть  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$  –  $n$ -мерный случайный вектор.

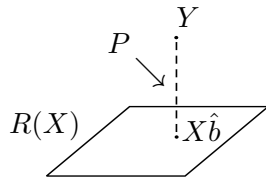
- (i) Если  $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 \mathbb{1}_n)$ , то  $Y = Xb + \varepsilon$  называется *линейной моделью с предположением нормальности (LMN)*.
- (ii) Пусть  $Z \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_n)$ . Если

$$\mathbb{E}[\varepsilon_{i_1} \varepsilon_{i_2} \varepsilon_{i_3} \varepsilon_{i_4}] = \mathbb{E}[Z_{i_1} Z_{i_2} Z_{i_3} Z_{i_4}] \quad \forall i_j \in \{1, \dots, n\},$$

то  $Y = Xb + \varepsilon$  называется *линейной моделью с предположением о моментах (LMM)*.

(iii)  $X$  называется **матрицей плана**.

**Замечание 8.4.** Если ранг матрицы  $X$  равен  $r$  и



$$R(X) = \{Xb \mid b \in \mathbb{R}^k\} \in \mathbb{R}^n$$

её диапазон, то наивной оценкой  $Xb$  является ортогональная проекция  $Y$  на  $R(X)$ . Тогда любой вектор  $\hat{b}$  такой, что  $P(Y) = X\hat{b}$  является приемлемой оценкой  $b$ .

**Определение 8.5.** Мы назовем  $G \in \mathbb{R}^{n \times m}$  **обобщенной обратной матрицей** к матрице  $A \in \mathbb{R}^{m \times n}$ , если

$$AGA = A.$$

Зададим множество обобщенных обратных к  $A$  матриц:

$$A^- = \{G \mid AGA = A\}.$$

**Замечание 8.6.** Мы будем писать  $A^-$  вместо  $G$ , если действительность формулы не зависит от выбора обобщенной обратной матрицы. Например,  $AA^-A = A$ .

**Пример 8.7.** Пусть

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Обобщенные обратные матрицы, например:

$$G_1 = \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{и} \quad G_2 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

**Лемма 8.8 (Включение диапазона).** Пусть  $X \in \mathbb{R}^{n \times k}$  и  $V \in \mathbb{R}^{n \times s}$ . Тогда:

$$(i) \quad R(X) \subset R(V) \iff VV^-X = X.$$

(ii) Если (i) соблюдается и  $V \geq 0$  (также  $n = s$ ), то

$$(a) \quad X^TV^-X = 0,$$

$$(b) \quad R(X^T) = R(X^TV^-X).$$

ДОКАЗАТЕЛЬСТВО:

(i) Напомним, что

$$R(X) \subset R(V) \iff X = VW.$$

" $\Leftarrow$ "  $\checkmark$ .

" $\Rightarrow$ " Пусть  $X = VW$  и  $G$  – обобщенная обратная матрица  $V$ . Тогда

$$VGX = VGVW = VW = X.$$

(ii) (a) Пусть  $G$  – обобщенная обратная матрица  $V$ . Тогда вследствие симметричности  $V$

$$X^TGX = W^TV^TGVW = W^TVW \geq 0.$$

(b) Напомним некоторые теоремы из линейной алгебры:

- i. Матрица  $V$  симметричная и неотрицательно определенная  $\Rightarrow$  её собственные числа  $\lambda_i$  вещественные и неотрицательные  $\Rightarrow V = \sum_{i=1}^n \lambda_i z_i z_i^T$  ( $z_i$  – ортонормальные собственные вектора)  $\Rightarrow V^\alpha = \sum_{i=1}^n \lambda_i^\alpha z_i z_i^T$  для  $\alpha \geq 0$  и  $V^{\alpha+\beta} = V^\alpha V^\beta$ .

ii.  $r(A) = r(A^T A)$  и  $r(A \cdot B) \leq \min\{r(A), r(B)\}$ . В частности,  $r(VW) = r(W^T VW)$ , так как

$$r(VW) = r(V^{\frac{1}{2}} V^{\frac{1}{2}} W) \leq r(V^{\frac{1}{2}} W) = r(W^T VW) \leq \min\{r(W^T), r(VW)\}.$$

Используя (а), мы получаем:

$$r(X^T V^- X) = r(W^T VW) = r(VW) = r(VV^- X) = r(X).$$

**Теорема 8.9.** В линейной модели  $Y = Xb + \varepsilon$  ортогональная проекция  $Y$  на  $R(X)$  и её ортогональное дополнение

$$R(X)^\perp = \{Z \mid Z^T X = 0\}$$

задаются как

$$P = X(X^T X)^- X^T \quad \text{и} \quad R = \mathbb{I}_n - P$$

соответственно.

**ДОКАЗАТЕЛЬСТВО:** Мы докажем только для  $P$ . Для ортогональной проекции имеют место равенства  $P^T = P$  и  $P^2 = P$ . Во-первых,

$$(X^T X)(X^T X)^- X^T = VV^- X^T = X^T$$

по Лемме 8.8(i) и (ii)(b). Тогда

$$P^2 = X(X^T X)^- X^T X(X^T X)^- X^T = X(X^T X)^- X^T = P$$

и  $P^T = P$  следует из Леммы 8.8 (ii)(a). Наконец,  $P(Y) \in R(X)$ , так как  $P$  вида  $XA$ , где  $A = (X^T X)^- X^T$ . Также,

$$P(Xb) = X(X^T X)^- X^T Xb = Xb.$$

### Замечание 8.10.

(i) Разумными оценками для  $Xb$  и  $\sigma^2$  являются

$$X\hat{b} = P(Y) = X(X^T X)^- X^T Y$$

и

$$\hat{\sigma}^2 = \frac{\|Y - X\hat{b}\|_2^2}{n - r} = \frac{\|RY\|_2^2}{n - r} = \frac{Y^T RY}{n - r},$$

где выбор знаменателя  $n - r$  будет обоснован в Следствии 8.13.

(ii) В общем случае, не существует единственной оценки для  $b$ . Однако если матрица  $X^T X$  обратима ( $r = r(X) = k$ ), то

$$X^T X\hat{b} = X^T X(X^T X)^{-1} X^T Y = X^T Y$$

и

$$\hat{b} = (X^T X)^{-1} X^T Y.$$

Заметим также, что

$$\mathbb{E}[\hat{b}] = (X^T X)^{-1} X^T \mathbb{E}[Y] = (X^T X)^{-1} X^T Xb = b.$$

**Пример 8.11.** Пусть  $Y_1, \dots, Y_n$  i.i.d.  $\sim \mathcal{N}(\mu, \sigma^2)$ , такие что:

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \mu + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} = X\mu + \varepsilon,$$

где  $\varepsilon_1, \dots, \varepsilon_n$  i.i.d.  $\sim \mathcal{N}(0, \sigma^2)$ . Так как

$$X^T X = (1 \ \dots \ 1) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = n,$$

то

$$P = X(X^T X)^{-1} X^T = \frac{1}{n} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} (1 \ \dots \ 1).$$

Таким образом,

$$PY = \frac{1}{n} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \sum_{i=1}^n Y_i = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \bar{Y}_n$$

является оценкой  $X\mu = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \mu$ . В частности,  $\bar{Y}_n$  – оценка  $\mu$ . Также,

$$\hat{\sigma}^2 = \frac{\|Y - X\hat{\mu}\|_2^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2.$$

**Лемма 8.12.** Пусть  $Y$  –  $n$ -мерная случайная величина с математическим ожиданием  $\mathbb{E}[Y] = \mu$  и дисперсией  $\text{Var}(Y) = V \geq 0$ . Также, пусть  $A, B \in \mathbb{R}^{n \times n}$ . Тогда:

(i)  $\mathbb{E}[Y^T A Y] = \mu^T A \mu + \text{tr}(A V)$ .

(ii) Пусть моменты  $Y$  до четвертого порядка совпадают с моментами нормального распределения, тогда

(a)  $\text{Cov}(Y, Y^T A Y) = 2V A \mu$ ,

(b)  $\text{Cov}(Y^T A Y, Y^T B Y) = 2\text{tr}(A V B V)$ , если  $\mu = 0$ .

ДОКАЗАТЕЛЬСТВО: Упражнение.

**Следствие 8.13.** В LMM  $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$ .

ДОКАЗАТЕЛЬСТВО: По определению и по Лемме 8.12:

$$\mathbb{E}[\hat{\sigma}^2] = \frac{\mathbb{E}[Y^T R Y]}{n-r} = \frac{1}{n-r} (\mu^T R \mu + \text{tr}(\sigma^2 R)),$$

где  $\mu = Xb$ . Так как  $R$  – ортогональная проекция  $R(X)^\perp$ , то  $RXb = 0$ . Следовательно,

$$\mathbb{E}[\hat{\sigma}^2] = \sigma^2 \frac{\text{tr}(R)}{n-r}.$$

Любая ортогональная проекция  $Q$  удовлетворяет равенству

$$Q = A \cdot \text{diag}(\lambda_i) \cdot A^T.$$

Так как  $Q$  идемпотентна ( $Q^2 = Q$ ), то все собственные числа равны либо 0, либо 1. Число единиц совпадает с рангом  $Q$ . Как следствие,

$$\text{tr}(R) = r(R) = n - r.$$

**Теорема 8.14 (Теорема Гаусса-Маркова).** Рассмотрим LMM с  $r(X) = k$ :

- (i) Оценки  $\hat{b}$  и  $\hat{\sigma}^2$  несмещенные и некоррелированные.
- (ii) Оценка  $\hat{b}$  – **лучшая линейная несмещенная оценка (BLUE)**  $b$ , то есть для любого вектора  $\tilde{b} = LY$ , такого что  $\mathbb{E}[\tilde{b}] = b$ :

$$\text{Var}(\tilde{b}) \geq \text{Var}(\hat{b}) = \sigma^2(X^T X)^{-1}.$$

- (iii) Оценка  $\hat{\sigma}^2$  – **лучшая квадратичная несмещенная оценка (BQUE)**  $\sigma^2$ , то есть для любого вектора  $\tilde{\sigma}^2 = Y^T A Y$ , такого что  $\mathbb{E}[\tilde{\sigma}^2] = \sigma^2$ :

$$\text{Var}(\tilde{\sigma}^2) \geq \text{Var}(\hat{\sigma}^2).$$

ДОКАЗАТЕЛЬСТВО:

- (i) Несмещенность следует из Замечания 8.10 (ii) и Следствия 8.13. Также,

$$\text{Cov}(\hat{b}, \hat{\sigma}^2) = \frac{1}{n-k} (X^T X)^{-1} X^T \text{Cov}(Y, Y^T R Y)$$

и из Леммы 8.12 (ii) (a) следует, что

$$\text{Cov}(Y, Y^T R Y) = 2\sigma^2 \mathbb{I}_n R X b = 0,$$

так как  $R X b = 0$ .

- (ii) Если  $\tilde{b}$  несмещенная, то

$$\mathbb{E}[\tilde{b}] = L \mathbb{E}[Y] = L X b = b \quad \forall b \in \mathbb{R}^k.$$

Следовательно,  $L X = \mathbb{I}_k$ . Тогда

$$\begin{aligned} 0 &\leq ((X^T X)^{-1} X^T - L)((X^T X)^{-1} X^T - L)^T \\ &= (X^T X)^{-1} - (X^T X)^{-1} X^T L^T - L X (X^T X)^{-1} + L L^T \\ &= L L^T - (X^T X)^{-1}. \end{aligned}$$

Наконец,

$$\text{Var}(\tilde{b}) = \text{Var}(LY) = L \text{Var}(Y) L^T = \sigma^2 L L^T \geq \sigma^2 (X^T X)^{-1} = \text{Var}(\hat{b}).$$

- (iii) Может быть доказано аналогично (ii), используя Лемму 8.12 (ii) (b).

**Лемма 8.15.** Пусть  $Y \sim \mathcal{N}(0, V)$  и  $A \in \mathbb{R}^{p \times n}$  и  $B \in \mathbb{R}^{q \times n}$ . Тогда

- (i)  $A Y$  и  $B Y$  независимы, если  $A V B^T = 0$ .
- (ii) Пусть  $q = n$  и  $B$  ортогональная проекция. Тогда  $Y^T A Y$  и  $B Y$  независимы, если  $A V B = 0$ .

ДОКАЗАТЕЛЬСТВО:

- (i) Свойство нормального распределения.

- (ii) Как в Лемме 2.20.

**Теорема 8.16.** В LMM с  $r(X) = k$  оценка  $(\hat{b}, \hat{\sigma}^2)^T$  UMVU для  $(b, \sigma^2)^T$  и обе оценки независимы.

ДОКАЗАТЕЛЬСТВО: Независимость следует из Леммы 8.15, где  $A = R$  и  $B = P$ .

UMVU: функция плотности случайной величины  $Y \sim \mathcal{N}(Xb, \sigma^2 \mathbb{I}_n)$ :

$$f(y) = c(\sigma^2) \exp \left\{ -\frac{1}{2\sigma^2} \|y - Xb\|_2^2 \right\} = \tilde{c}(\sigma^2, b) \exp \left\{ -\frac{1}{2\sigma^2} y^T y - 2b^T X^T y \right\}.$$

Следовательно, мы получаем  $(k+1)$ -мерное экспоненциальное семейство. Статистики  $Y^T Y$  и  $X^T Y$  являются достаточными (Теорема 4.22) и полными (Теорема 4.25) для оценки  $(b\sigma^2)^T$ . Статистика  $(\hat{b}, \hat{\sigma}^2)^T$  также достаточная и полная (Замечание 4.31). Теорема 4.28 завершает доказательство.

**Пример 8.17 (Метод наименьших квадратов).** Рассмотрим линейную регрессию:

$$Y_i = b_0 + b_1 X_i + \varepsilon_i.$$

Если

$$X = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}$$

– матрица полного ранга, то  $\hat{b} = (X^T X)^{-1} X^T Y$ . Пусть  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  и  $\bar{X}_n^2 = \frac{1}{n} \sum_{i=1}^n X_i^2$ , тогда:

$$X^T X = n \begin{pmatrix} 1 & \bar{X}_n \\ \bar{X}_n & \bar{X}_n^2 \end{pmatrix}.$$

Обратная матрица

$$(X^T X)^{-1} = \frac{1}{\sum_{i=1}^n (X_i - \bar{X}_n)^2} \begin{pmatrix} \bar{X}_n^2 & -\bar{X}_n \\ -\bar{X}_n & 1 \end{pmatrix}$$

существует, если все  $X_i$  принимают различные значения. Также,

$$X^T Y = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix}.$$

Наконец,  $\hat{b} = (\hat{b}_0, \hat{b}_1)^T$ , где

$$\begin{aligned} \hat{b}_0 &= \bar{Y}_n - \hat{b}_1 \bar{X}_n, \\ \hat{b}_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sum_{i=1}^n (X_i - \bar{X}_n)^2} \end{aligned}$$

и

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{b}_0 - \hat{b}_1 X_i)^2.$$

**Замечание 8.18.** Часто мы заинтересованы не в  $b$ , но в  $K^T b$  для некоторого  $K \in \mathbb{R}^{k \times s}$ . Если  $r(X) = k$ , то разумной оценкой будет:

$$K^T \hat{b} = K^T (X^T X)^{-1} X^T Y.$$

Даже если  $r(X) \neq k$ , но  $R(K) \subset R(X^T)$ , то оценка

$$K^T \hat{b} = K^T (X^T X)^- X^T Y$$

единственна по Лемме 8.8 и

$$\mathbb{E}[K^T \hat{b}] = K^T (X^T X)^- X^T X b = K^T b.$$

**Определение 8.19.** Пусть  $K \in \mathbb{R}^{k \times s}$  и  $r(K) = s$ . Тогда мы назовем  $K^T b$  *оцениваемым*, если  $R(K) \subset R(X^T)$ .

**Пример 8.20.** Допустим, мы тестируем

$$H_0 : K^T b = 0 \quad \text{против} \quad H_1 : K^T b \neq 0.$$

Не должно быть ситуации, в которой одновременно  $Xb_1 = Xb_2$  и  $K^T b_1 \neq K^T b_2$ , так как  $b$  может быть получено только из  $Xb$  в модели  $Y = Xb + \varepsilon$ . Другими словами, если мы зададим множество

$$N(A) = \{y \mid Ay = 0\},$$

то

$$Xb_1 = Xb_2 \implies K^T b_1 = K^T b_2$$

эквивалентно

$$N(X) \subset N(K^T) \iff R(X^T)^\perp \subset R(K)^\perp \iff R(K) \subset R(X^T).$$

**Теорема 8.21.** Пусть  $K^T b$  оцениваем. Тогда:

(i) В LMM  $K^T \hat{b}$  – BLUE для  $K^T b$ , где

$$\text{Var}(K^T \hat{b}) = K^T \sigma^2 (X^T X)^{-1} K \in \mathbb{R}^{s \times s}.$$

(ii) В LMM  $K^T \hat{b}$  – UMVU для  $K^T b$ .

ДОКАЗАТЕЛЬСТВО: Как в Теоремах 8.14 и 8.16.

**Пример 8.22.**

(i) Рассмотрим линейную регрессию:  $Y_i = b_0 + b_1 x_i + \varepsilon_i$ ,  $i = 1, \dots, n$ , где  $\mathbb{E}[\varepsilon_i] = 0$  и  $\text{Var}(\varepsilon_i) = \sigma^2 > 0$ . В векторной записи:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Допустим, мы заинтересованы в гипотезах:

$$H_0 : b_0 = 0 \quad \text{против} \quad H_1 : b_0 \neq 0,$$

тогда мы выбираем  $K = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  и, например,

$$X = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix}.$$

Очевидно, что  $R(X) \subset R(K^T)$ . Также,

$$X^T X = \begin{pmatrix} n & 0 \\ 0 & 0 \end{pmatrix}.$$

не обратима, но мы можем взять

$$G = \frac{1}{n} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

как обобщенную обратную матрицу и  $K^T \hat{b}$  становится:

$$K^T G X^T Y = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}_n.$$

(ii) Рассмотрим анализ дисперсий для  $a = 3$ :

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad (i = 1, 2, 3, j = 1, \dots, n_i).$$

В векторной записи:

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \\ Y_{31} \\ \vdots \\ Y_{3n_3} \end{pmatrix} = \begin{pmatrix} \mathbb{1}_{n_1} & 0 & 0 \\ 0 & \mathbb{1}_{n_2} & 0 \\ 0 & 0 & \mathbb{1}_{n_3} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{2n_2} \\ \varepsilon_{31} \\ \vdots \\ \varepsilon_{3n_3} \end{pmatrix}.$$

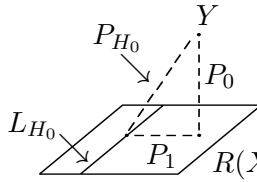
Если мы хотим проверить:

$$H_0 : \mu_1 = \mu_2 = \mu_3 \quad \text{против} \quad H_1 : \mu_i \neq \mu_j \text{ для некоторых } i \neq j,$$

то мы можем выбрать

$$K^T \mu = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \end{pmatrix}.$$

**Замечание 8.23.** Мы рассматриваем



где  $R(K) \subset R(X^T)$ . В таком случае решающей величиной будет являться расстояние от  $Y$  до пространства гипотезы:

$$L_{H_0} = \{XB \mid K^T b = 0, b \in \mathbb{R}^k\} \subset R(X).$$

Ортогональная проекция  $Y$  на  $L_{H_0}$ :

$$P_{H_0} = P_0 - P_1,$$

где

$$P_0 = X(X^T X)^{-1} X^T$$

и

$$P_1 = X(X^T X)^{-1} K(K^T(X^T X)^{-1} K)^{-1} K^T(X^T X)^{-1} X^T.$$

$P_1$  также является ортогональной проекцией. При построении критерия разумно опираться на расстояние между  $P_0 Y$  и  $P_{H_0} Y$ . По теореме Пифагора:

$$\|(\mathbb{I}_n - P_{H_0})Y\|_2^2 - \|(\mathbb{I}_n - P_0)Y\|_2^2 = \|P_1 Y\|_2^2 = Y^T P_1 Y.$$

Последнее равенство следует из свойства идемпотентности матрицы  $P_1$ .

**Теорема 8.24.** Пусть  $Y \sim \mathcal{N}(\mu, \sigma^2 \mathbb{I}_n)$  и  $P \in \mathbb{R}^{n \times n}$ , где  $P^T = P$ . Тогда  $P$  – ортогональная проекция тогда и только тогда, когда

$$Q = \frac{(Y - \mu)^T P (Y - \mu)}{\sigma^2} \sim \chi_{r(P)}^2.$$

ДОКАЗАТЕЛЬСТВО:

" $\Rightarrow$ " Если  $P^2 = P$ , то существует  $A \in \mathbb{R}^{n \times n}$ , где  $A^T A = A A^T = \mathbb{I}_n$ , такая что

$$A^T P A = \begin{pmatrix} \mathbb{I}_n & 0 \\ 0 & 0 \end{pmatrix},$$

где  $r = r(P)$ . Мы знаем, что  $Z = A^T(Y - \mu) \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_n)$ . Тогда

$$Q = \frac{(Y - \mu)^T P (Y - \mu)}{\sigma^2} = \frac{Z^T A^T P A Z}{\sigma^2} = \sum_{i=1}^r \left( \frac{Z_i}{\sigma} \right)^2 \sim \chi_r^2.$$

" $\Leftarrow$ " Так как  $P^T = P$ , существует матрица  $B$ , такая что  $B^T B = B B^T = \mathbb{I}_n$  и

$$B^T P B = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n),$$

где  $\lambda_i$  – вещественнозначные собственные числа  $P$ . Подставляя  $X = B^T(Y - \mu) \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_n)$ , получаем

$$Q = \frac{1}{\sigma^2} X^T B^T P B X.$$



Так как  $Q \sim \mathcal{X}_r^2$ , то её характеристическая функция:

$$\mathbb{E}[\exp\{itQ\}] = (1 - 2it)^{-r/2},$$

и также:

$$\begin{aligned} \mathbb{E}[\exp\{itQ\}] &= \mathbb{E}\left[\exp\left\{\frac{it}{\sigma^2} X^T B^T P B X\right\}\right] = \mathbb{E}\left[\exp\left\{it \sum_{j=1}^n \lambda_j \left(\frac{X_j}{\sigma}\right)^2\right\}\right] \\ &= \prod_{j=1}^n \mathbb{E}\left[\exp\left\{it \lambda_j \left(\frac{X_j}{\sigma}\right)^2\right\}\right] = \prod_{j=1}^n (1 - 2it \lambda_j). \end{aligned}$$

Поскольку полином однозначно определяется его линейными множителями,  $\lambda_1 = \dots = \lambda_r = 1$  и  $\lambda_j = 0 \ \forall j > r$ . В частности,

$$P^2 = B \Lambda B^T B \Lambda B^T = B^T \Lambda^2 B^T = B \Lambda B^T = P.$$

**Замечание 8.25.** Пусть  $Y \sim \mathcal{N}(\mu, \sigma^2 \mathbb{I}_n)$  и  $P$  – ортогональная проекция, где  $r(P) = r$ . Задав

$$A^T P A = \begin{pmatrix} \mathbb{1}_r & 0 \\ 0 & 0 \end{pmatrix},$$

мы получаем

$$\tilde{Q} = \frac{Y^T P Y}{\sigma^2} = \frac{\tilde{Z}^T A^T P A \tilde{Z}}{\sigma^2},$$

где

$$\tilde{Z} = A^T Y \sim \mathcal{N}(A^T \mu, \sigma^2 \mathbb{1}_n).$$

Можно показать, что распределение

$$\tilde{Q} = \sum_{i=1}^r \left(\frac{\tilde{Z}_i}{\sigma}\right)^2$$

зависит только от  $r$  и

$$\delta^2 = \sum_{i=1}^r \left(\frac{(A^T \mu)_i}{\sigma}\right)^2 = \frac{\mu^T P \mu}{\sigma^2}.$$

Оно называется смещенным распределением  $\mathcal{X}^2$  с  $r$  степенями свободы и смещением  $\delta^2$ . Обозначение:  $\tilde{Q} \sim \mathcal{X}_{r, \delta^2}^2$ .

**Определение 8.26.** Пусть  $X \sim \mathcal{X}_m^2$  и  $Y \sim \mathcal{X}_n^2$  независимы.

(i) Распределение случайной величины

$$F = \frac{nX}{mY}$$

называется ***F-распределением с  $t$  и  $n$  степенями свободы***. Обозначение  $F \sim F_{m,n}$ .

(ii) Если  $X \sim \mathcal{X}_{m, \sigma^2}^2$ , то

$$F = \frac{nX}{mY} \sim F_{m,n, \sigma^2}.$$

**Теорема 8.27 (F-критерий в LMN).** В LMN пусть  $R(K) \subset R(X^T)$ ,  $t = r(K)$  и  $r = r(X)$ . Тогда

(i)

$$F = \frac{\frac{1}{t} \|P_1 Y\|_2^2}{\frac{1}{n-r} \|R Y\|_2^2} \sim F_{t, n-r, \delta^2},$$

где

$$\delta^2 = \frac{1}{\sigma^2} (K^T b)^T (K^T (X^T X)^- K)^- K^T b$$

и

$$R = \mathbb{I} - P_0$$

как прежде.

(ii)

$$\varphi(Y) = \begin{cases} 1, & F > F_{t,n-r,1-\alpha} \\ 0, & \text{иначе.} \end{cases}$$

– **F-критерий** для

$$H_0 : K^T b = 0 \quad \text{против} \quad H_1 : K^T b \neq 0$$

с уровнем значимости  $\alpha$ .

ДОКАЗАТЕЛЬСТВО: Достаточно показать (i). Из Замечания 8.25 следует:

$$\frac{\|P_1 Y\|_2^2}{\sigma^2} = \frac{Y^T P_1 Y}{\sigma^2} \sim \chi_{t,\delta^2}^2$$

и

$$\delta^2 = \frac{(Xb)^T P_1 Xb}{\sigma^2} = \frac{(K^T b)^T (K^T (X^T X)^{-1} K)^{-1} K^T b}{\sigma^2},$$

исходя из определения  $P_1$ . Аналогично  $\frac{Y^T R Y}{\sigma^2} \sim \chi_{n-r}$ , так как  $RXb = 0$ . Лемма 8.15 и  $P_1 R = 0$  завершают доказательство.

# Литература

- [1] Bauer H. Maß- und Integrationstheorie. Berlin: Walter de Gruyter & Co., 1992.
- [2] Tripathi G. A matrix extension of the Cauchy-Schwarz inequality. *Econom. Lett.* 63 (1), 1 – 3. 1999.
- [3] Bickel P. J., Doksum K. A. Mathematical statistics. San Francisco, Calif.-Düsseldorf-Johannesburg: Holden-Day, Inc., 2001.
- [4] Lehmann E. L., Romano J. P. Testing statistical hypotheses. New York: Springer, 2005.
- [5] Witting H., Müller-Funk U. Mathematische Statistik. II. Stuttgart: B. G. Teubner, 2001.