МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ М. В. ЛОМОНОСОВА

Механико-математический факультет

экономический поток

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

3 курс

5-6 семестры

Лектор
к. ф.-м. н., доцент
М.В. Болдин
«___» ________ 2021 г.

Семинарист
к. ф.-м. н., ассистент
А.А. Муромская
«___» ________ 2021 г.

Москва, 2021 г.

Техническая информация

Данный PDF содержит примерную программу осеннего и весеннего семестров 3 курса по предмету «Математическая статистика».

Собрали и напечатали по мотивам лекций и семинаров студенты 3-го курса Конов Марк, Валерий Старцев, Гащук Елизавета.

Авторы выражают огромную благодарность лектору, кандидату ф.-м. наук, доценту Болдину Михаилу Васильевичу, а также семинаристу, кандидату ф.-м. наук, ассистенту Муромской Анастасии Андреевне за прочитанный курс по предмету «Математическая статистика».

Добавления и исправления принимаются на почты vkonov2@yandex.ru, sharfikeg@yandex.ru, gashchuk2011@mail.ru.

ПРИЯТНОГО ИЗУЧЕНИЯ

Содержание

1 Предварительные сведения

1.1 Мера, распределение

Пусть $\Omega = \{w\}$ - произвольное множество, а $\mathcal{F}$ - сигма-алгебра его подмножеств. Т.е. $\mathcal{F}$ такая система множеств, что:

1) $\Omega \in \mathcal{F}$
2) если $A \in \mathcal{F}$, то $\overline{A} := \Omega - \mathcal{F} \in \mathcal{F}$
3) если $A_1, A_2, \dots \in \mathcal{F}$, то $\bigcup\limits_i A_i \in \mathcal{F}$, $\bigcap\limits^i A_i \in \mathcal{F}$

Определение 1.1. *Пусть $\Omega = R$, а $\mathcal{F}$ - наименьшая сигма-алгебра, содержащая все интервалы (α, β). Такая $\mathcal{F}$ обозначается $\mathcal{B}(\mathbb{R})$ и называется* ***борелевской*** *сигма-алгеброй.*

Определение 1.2. *Мера μ, определенная на $\mathcal{F}$, называется* ***сигма-аддитивной****, если:*

1) это неотрицательная функция $\mu(A) \geq 0 \; \forall A \in \mathcal{F}$
2) она удовлетворяет условию сигма-аддитивности:

$$\mu(\sum_i A_i) = \sum_i \mu(A_i), \; A_i \in \mathcal{F}, \; A_i A_j = \emptyset \text{ при } i \neq j$$

Определение 1.3. *Мера μ называется* ***сигма-конечной****, если существуют множества $A_i \in \mathcal{F}$ такие, что $\bigcup\limits_i A_i = \Omega$ и $\mu(A_i) < \infty$.*

Считающая мера: пусть Ω - счетное, $\mathcal{F}$ - множество всех подмножеств Ω. Положим для $A \in \mathcal{F}$ $\mu(A) := \{\text{число точек } \Omega, \text{ попавших в } A\}$. Такая мера μ называется считающей, она сигма-конечна.

Лебегова мера: пусть $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$. Существует единственная мера μ на $\mathcal{B}(\mathbb{R})$ такая, что $\mu\left((\alpha, \beta]\right) = \beta - \alpha$. Эта мера Лебега, она сигма-конечна.

$(\Omega, \mathcal{F})$ - измеримое пространство, $(\Omega, \mathcal{F}, \mu)$ - пространство с мерой.

Определение 1.4. *Если $\mu(\Omega) = 1$, то μ - **вероятностная мера**, она обозначается через P. Тройка $(\Omega, \mathcal{F}, \mu)$ - **вероятностное пространство**.*

Определение 1.5. *Измеримая функция $\xi : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ называется **случайной величиной**. Измеримость означает, что:*

$$\forall B \in \mathcal{B}(\mathbb{R}) \;\; \xi^{-1}(B) := (w \;:\; \xi(w) \in B) \in \mathcal{F}$$

*Измеримая функция $\varphi : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ называется **борелевской**.*

Определение 1.6. *Рассмотрим случайную величину $\xi \in \mathbb{R}^1$. Для $x \in \mathbb{R}^1$ функция $F(x) = P(w : \xi(w) \le x) = P(\xi \le x)$ называется **функцией распределения**.*

Определение 1.7. *Мера $P_\xi(A) = P(w : \xi(w) \in A)$, $A \in \mathcal{B}(\mathbb{R})$, называется **распределением** случайной величины ξ.*

Тогда $F(x) = P_\xi\left((-\infty, x]\right)$, т.е. P_ξ определяет $F(x)$.

Обратно: $P(\alpha < \xi \le \beta) = F(\beta) - F(\alpha)$, и существует единственная вероятностная мера P_ξ такая, что $P_\xi\left((\alpha, \beta]\right) = F(\beta) - F(\alpha)$, т.е. $F(x)$ определяет P_ξ.

Определение 1.8. *Пусть на $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ задана сигма-конечная мера μ. Если существует борелевская функция $f(x)$, $f(x) \ge 0$, такая, что*

$$P_\xi(A) = \int\limits_A f(x)\mu(dx) \;\; \forall A \in \mathcal{B}(\mathbb{R})$$

*то $f(x)$ называется **плотностью** вероятности по мере μ.*

Если μ - мера Лебега, то $f(x)$ - обычная плотность случайной величины ξ, введенная на 2-ом курсе. Если же ξ дискретна со значениями $x_1, x_2, \ldots$, а μ - считающая мера, сосредоточенная в этих точках, то, очевидно:

$$P_\xi(A) = \int\limits_A P(\xi = x)\mu(dx) \; \forall A \in \mathcal{B}(\mathbb{R})$$

Последнее равенство означает, что у дискретной случайной величины ξ есть плотность вероятности $f(x) = P(\xi = x)$, $x = x_1, x_2, \ldots$ по считающей мере. При

$x \neq x_1, x_2, \ldots$ значения этой плотности не важны, их можно положить равными нулю.

Определение 1.9. ***Математическим ожиданием*** *случайной величины ξ называется число $E\xi = \int\limits_{\Omega} \xi(w)P(dw)$ в предположении, что $\int\limits_{\Omega} |\xi(w)|P(dw) < \infty$. Если $\int\limits_{\Omega} |\xi(w)|P(dw) = \infty$, то будем говорить, что $E\xi$ не существует.*

Если $f(x)$ - плотность вероятности случайной величины ξ по мере μ, а $\varphi(x)$ - борелевская функция, то:

$$E\varphi(x) = \int\limits_R \varphi(x)P_\xi(dx) = \int\limits_R \varphi(x)f(x)\mu(dx)$$

В частности, если ξ - абсолютно непрерывная случайная величина в терминологии 2-го курса (т.е. μ - мера Лебега), то в случае $\int\limits_R |\varphi(x)|f(x)dx < \infty$ пишем $E\varphi(x) = \int\limits_R \varphi(x)f(x)dx$.

Если ξ дискретна со значениями $x_1, x_2, \ldots$ и соответствующими вероятностями $p_1, p_2, \ldots$, то $E\varphi(\xi) = \sum\limits_{i \geq 1} \varphi(x_i)p_i$ (если ряд сходится абсолютно).

1.2 Случайные вектора

Обозначим $\mathcal{B}(\mathbb{R}^k)$ борелевскую сигма-алгебру подмножеств $\mathbb{R}^n$.

Определение 1.10. *Вектор $\xi = (\xi_1, \ldots, \xi_k)^T$ называется **k-мерным случайным вектором**, если ξ - измеримое отображение $\xi : (\Omega, \mathcal{F}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k)$*

Известно: ξ - случайный вектор тогда и только тогда, когда каждая компонента ξ_i - одномерная случайная величина.

Определение 1.11. *Функция распределения случайного вектора ξ: $F(x_1, \ldots, x_k) = P(\xi_1 \leq x_1, \ldots, \xi_k \leq x_k)$, $x_i \in \mathbb{R}$, а распределение $P_\xi(A) = P(w : \xi(w) \in A)$, $A \in \mathcal{B}(\mathbb{R}^k)$.*

Определение 1.12. *Плотность вероятности вектора ξ по мере μ (μ распределена на элементам $\mathcal{B}(\mathbb{R}^k)$) - борелевская функция $f(x) \geq 0$, $x = (x_1, \ldots, x_n)$, такая что:*

$$P_\xi(A) = \int\limits_A f(x)\mu(dx)\ \forall A \in \mathcal{B}(\mathbb{R}^k)$$

Определение 1.13. *Случайные величины* $\{\xi_1, \dots, \xi_k\}$ *независимы, если*

$$P(\xi_1 \in A_1, \dots, \xi_k \in A_k) = \prod_{i=1}^{k} P(\xi_i \in A_i) \; \forall A_i \in \mathcal{B}(\mathbb{R})$$

Предложение 1.1 (*необходимые и достаточные условия независимости*).

$$F(x_1, \dots, x_k) = F_{\xi_1}(x_1)F_{\xi_2}(x_2) \dots F_{\xi_k}(x_k) \; \forall (x_1, \dots, x_k)$$

$$f(x_1, \dots, x_k) = f_{\xi_1}(x_1) \dots f_{\xi_k}(x_k)$$

1.3 Сходимости случайных векторов

Пусть случайные векторы $\xi, \xi_1, \xi_2, \dots$ размера k со значениями в $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ определены на некотором вероятностном пространстве $(\Omega, \mathcal{F}, P)$. Пусть $|\cdot|$ означает евклидову норму вектора, т.е. $|\xi| = \sqrt{\xi_1^2 + \dots + \xi_k^2}$.

Определение 1.14. *Говорят, что последовательность* $\{\xi_n\}$ ***сходится слабо*** *к* ξ *(*$\xi_n \xrightarrow[n\to\infty]{W} \xi$*), если для любой непрерывной и ограниченной функции* $g : \mathbb{R}^k \to \mathbb{R}^1$ *имеет место сходимость:*

$$\int_{\mathbb{R}^k} g(x)P_n(dx) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^k} g(x)P(dx) \tag{1}$$

(P_n *и* P *- распределения* ξ_n *и* ξ *соотвественно)*

Определение 1.15. *Обозначим* $F_n(x)$ *и* $F(x)$, $x = (x_1, \dots, x_n)$, *функции распределения векторов* ξ_n *и* ξ. *Тогда сходимость* (1) *эквивалентна* ***сходимости в основном****:*

$$F_n(x) \Rightarrow F(x) \;\Leftrightarrow\; F_n(x) \to F(x) \;\; \forall x \in \mathbb{C}(\mathcal{F}) \tag{2}$$

Определение 1.16. *Пусть* $\varphi_n(t)$ *и* $\varphi(t)$, $t \in \mathbb{R}^k$, *будут характеристические функции* ξ_n *и* ξ, *т.е.* $\varphi(t) := Ee^{it^T\xi}$. *Тогда сходимость* (2) *эквивалентна сходимости:*

$$\varphi_n(t) \xrightarrow[n\to\infty]{} \varphi(t) \;\; \forall t \in \mathbb{R}^k \tag{3}$$

Определение 1.17. *Если выполнено любое из соотношений* (1) – (3), *будем писать:*

$$\xi_n \xrightarrow[n\to\infty]{d} \xi \tag{4}$$

и говорить, что $\{\xi_n\}$ *сходится к* ξ ***по распределению***.

Замечание. *Сходимость (4) не следует из сходимости* $\xi_{in} \xrightarrow{d} \xi_i$, $i = 1, \ldots, k$, *компонент векторов* ξ_n *и* ξ.

Определение 1.18. *Говорят, что последовательность* $\{\xi_n\}$ *сходится* ***по вероятности*** *к вектору* ξ ($\xi_n \xrightarrow[n\to\infty]{P} \xi$), *если:*

$$\forall \varepsilon > 0 \;\; P(|\xi_n - \xi| > \varepsilon) \xrightarrow[n\to\infty]{} 0 \tag{5}$$

Замечание. *Понятно, что сходимость (5) эквивалентна сходимости компонент* $\xi_{in} \xrightarrow{P} \xi_i$ *для всех* $i = 1, 2, \ldots, k$.

Замечание. *Сходимость по вероятности (5) влечет сходимость по распределению (4). Обратное верно только в частных случаях.*

Определение 1.19. *Говорят, что последовательность* $\{\xi_n\}$ ***сходится п.н.*** *(почти наверно или с вероятностью единица) и пишут* $\xi_n \xrightarrow[n\to\infty]{\text{п.н.}} \xi$, *если:*

$$P(w : \xi_n(w) \to \xi(w)) = 1 \tag{6}$$

Замечание. *Сходимость п.н. (6) влечет сходимость по вероятности (5). Значит верна следующая цепочка:*

$$\xi_n \xrightarrow[n\to\infty]{\text{п.н.}} \xi \;\Rightarrow\; \xi_n \xrightarrow[n\to\infty]{P} \xi \;\Rightarrow\; \xi_n \xrightarrow[n\to\infty]{d} \xi$$

Теорема 1.1 (**непрерывности**). *Пусть векторы* $\{\xi_n\}, \xi$ *определены на* $(\Omega, \mathcal{F}, P)$, ξ_n, ξ
$\mathbb{R}^k$. *Пусть* $A \in \mathcal{B}(\mathbb{R}^k)$ *и* $P(\xi \in A) = 1$ *(т.е.* A *- носитель* ξ*). Пусть борелевская* $H : \mathbb{R}^k \to \mathbb{R}^1$ *непрерывна на множестве* A. *Тогда:*

1) *если* $\xi_n \xrightarrow[n\to\infty]{d} \xi$, *то* $H(\xi_n) \xrightarrow[n\to\infty]{d} H(\xi)$
2) *если* $\xi_n \xrightarrow[n\to\infty]{P} \xi$, *то* $H(\xi_n) \xrightarrow[n\to\infty]{P} H(\xi)$
3) *если* $\xi_n \xrightarrow[n\to\infty]{\text{п.н.}} \xi$, *то* $H(\xi_n) \xrightarrow[n\to\infty]{\text{п.н.}} H(\xi)$

Доказательство. Докажем пункт 3. Два других пункта будут доказаны на практических занятиях.

Итак, в силу непрерывности функции $H(x)$ на A:

$$(w : \xi_n(w) \to \xi(w)) \cap (w : \xi(w) \in A) \subseteq (w : H(\xi_n(w))) \to H(\xi(w)) \;\Rightarrow$$
$$\Rightarrow\; 1 = P(\xi_n(w) \to \xi(w)) = P(\xi_n(w) \to \xi(w), \xi(w) \in A) \le P(H(\xi_n(w)) \to H(\xi(w)))$$

$\blacksquare$

1.4 ЗБЧ и ЦПТ

Пусть на $(\Omega, \mathcal{F}, P)$ задана бесконечная последовательность случайных величин $\xi_1, \xi_2, \ldots$

Определение 1.20. *Если $\{\xi_i\}$ независимы и одинаково распределены с конечным средним, $E|\xi_1| < \infty$, то*

$$n^{-1}\sum_{i=1}^{n}\xi_i \xrightarrow[n\to\infty]{п.н.} E\xi_1 \qquad (7)$$

Соотношение (7) - ***усиленный закон больших чисел Колмогорова***.

Определение 1.21. *Если $\{\xi_i\}$ некоррелированные случайные величины, может быть, разнораспределенные, но с одинаковым средним $m = E\xi_i$ и $D\xi_i \leq C < \infty$, то*

$$n^{-1}\sum_{i=1}^{n}\xi_i \xrightarrow[n\to\infty]{P} m = E\xi_i \qquad (8)$$

Соотношение (8) - ***слабый закон больших чисел***.

Определение 1.22. *Если $\{\xi_i\}$ - н.о.р.с.в., $E\xi_1 = m$, $0 < D\xi_1 = \sigma^2 < \infty$, то*

$$\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}(\xi_i - m) \xrightarrow[n\to\infty]{d} \xi \sim N(0,1) \qquad (9)$$

Соотношение (9) - ***центральная предельная теорема***, *точнее ее вариант, т.е.:*

$$n^{\frac{1}{2}}(\overline{\xi} - m) \xrightarrow[n\to\infty]{d} N(0,\sigma^2), \text{ где } \overline{\xi} := n^{-1}\sum_{i=1}^{n}\xi_i$$

2 Статистическая модель

2.1 Оценка среднего

Пример 2.1 (**оценка среднего**). *Будем предполагать, что на некотором вероятностном пространстве* $(\Omega, \mathcal{F}, P)$ *определена бесконечная последовательность* $X_1, X_2, \ldots$ *и* $X_1, \ldots, X_n$ *- ее первые n членов. Интересующий нас параметр, определяющий (в какой-то мере) срок службы, отождествим* $\theta = EX_1$.

Одна из стандартных статистических задач состоит в том, чтобы выяснить, чему равно θ. *Вот возможное решение. В силу УЗБЧ Колмогорова*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} \xrightarrow[n\to\infty]{п.н.} EX_1 = \theta$$

Возьмем n готовых изделий и проверим их. Пусть $x_1, x_2, \ldots, x_n$ *- сроки службы готовых изделий. Это реализации сл.в.* $X_1, \ldots, X_n$. *Естественно ожидать, что* $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ *при больших n окажется близким к* θ. *Это* ***задача точечного оценивания параметра****: пусть* $X_1, .., X_n-$ *случайные наблюдения;* $\overline{X}-$ *статистическая оценка(это случайная величина);* $\overline{x}$ *- реализация оценки, с ней обычно работают на практике.*

Ясно, что нужны оценки, которые в среднем близки к θ. *Тогда и реализации будут близки.*

Пусть в частности $P(X_1 \le t) = \begin{cases} 0, \ t \le 0 \\ 1 - e^{-\frac{t}{\theta}}, \ t > 0 \end{cases}$, *параметр* $\theta > 0$. *Т.е.* $X_1 \sim exp(\frac{1}{\theta})$ *и* $E_\theta X_1 = \theta$.

Тогда $\overline{X}$ ***оптимальна*** *при любом n > 0 в следующем смысле.*

1) $E_\theta \overline{X} = \frac{1}{n}\sum\limits_{i=0}^{n} E_\theta X_i = \theta \ \forall \theta > 0$ *- это свойство* ***несмещенности****. Качественно: реализации* $\overline{x}$ *группируются вокруг* θ*.*
2) $D_\theta \overline{X} \leq D_\theta \hat{\theta}_n$*,* $\theta > 0$ *и любой несмещенной оценки* $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$*. Качественно: реализации* $\overline{X}$ *в среднем лежат ближе к* θ*, чем у других* $\hat{\theta}_n$*.*

2.2 Проверка однородности данных

Пример 2.2 (**проверка однородности данных**)**.** *Пусть некоторый эксперимент проводится сначала* m *раз в условиях* A*, а затем* n *раз в условиях* B *(например, влияет ли некоторый препарат на на развитие растений, лекарство на анализы больного и т.д.).*
Будем считать x_i *реализациями н.о.р.с.в.* X_i *с функцией распределения* $X_1 \sim F_X(x) = P(X_1 \leq x)$*. Пусть* y_i *- реализации н.о.р.с.в.* Y_i*, ф.р.* $Y = F_Y(x)$*. Последовательности* x_i, y_i *независимы.*
Интерпретируем поставленную задачу как проверку гипотезы $H : F_X = F_Y$*. Предположение о том, что условие* B *дает другой результат интерпретируем как гипотезу (альтернативную к* H*)* $K : F_X \neq F_Y$*.*

Важно*: ни* F_X*, ни* F_Y *неизвестны!*
Оценкой F_X *возьмем* $\hat{F}_{mX}(x) = \frac{1}{m}\sum\limits_{i=1}^{m} I(X_i \leq x)$*,* $x \in \mathbb{R}$ *- это «хорошая» оценка, т.к. в силу УЗБЧ:* $\hat{F}_{mX}(x) \xrightarrow{п.н.} EI(X_1 \leq x) = F_X(x)$ *(у нас* $\{X_i\}$ *и* $\{Y_i\}$ *определены на одном* $(\Omega, \mathcal{F}, P)$*).*

Теорема 2.1 (**Глиненко-Кантелли**)**.**

$$\sup_x |\hat{F}_{mX}(x) - F_X(x)| \xrightarrow[m \to \infty]{п.н.} 0$$

Очевидно, если гипотеза H *верна, то величина* $D_{mn} := \sup\limits_x |\hat{F}_{mX}(x) - \hat{F}_{nY}(x)|$ *мала при больших* m, n*. Вот естественное правило:*

- *если* $D_{mn} \leq c$*, то* H *принять*
- *если* $D_{mn} > c$*, то* H *отвергнуть и принять* K

Но как выбрать константу c?

Лемма 2.1. *Пусть верна гипотеза* H *и* $F_X = F_Y = F$*. Пусть* F *непрерывна. Тогда распределение сл.в.* D_{mn} *не зависит от* $F(x)$ *при любом* x *и конечных* m, n*.*

Доказательство. *Докажем лемму при дополнительном предположении: $F(x)$ строго возрастает. Тогда при любом $t \in (0,1)$ существует $F^{-1}(t)$, и эта функция непрерывна и строго возрастает. Сделаем замену переменной $F(x) = t$, $x = F^{-1}(t)$. Тогда при $x \in \mathbb{R}$ переменная $t \in (0,1)$ и*

$$D_{mn} = \sup_{t\in(0,1)} |\hat{F}_{mX}(F^{-1}(t)) - \hat{F}_{nY}(F^{-1}(t))|.$$

Но $\hat{F}_{mX}(F^{-1}(t)) = \frac{1}{m}\sum_{i=1}^{m} I(X_i \le F^{-1}(t)) = \frac{1}{m}\sum_{i=1}^{m} I(F(X_i) \le t)$, *т.к.* $(X_i \le F^{-1}(t)) = (F(X_i) \le t)$. *Осталось заметить, что если* $X_i \sim F(x)$ *и* $F(x)$ *строго возрастает, то* $F(X_i) = \eta_i \sim R(0,1)$.
Действительно, $\forall t \in (0,1)$ $P(F(X_i) \le t) = P(X_i \le F^{-1}(t)) = F(F^{-1}(t)) = t$.
Значит $\hat{F}_{mX}(F^{-1}(t)) = \frac{1}{m}\sum_{i=1}^{m} I(\eta_i \le t)$, *где* η_i *- н.о.р.* $R(0,1)$ *сл.в., а тогда* $\hat{F}_{mX}(F^{-1}(t))$ *имеет ф.р., которая от* $F(x)$ *не зависит. Для* $\hat{F}_{nY}(F^{-1}(t))$ *имеем то же самое.* ■

Если D_{mn} *свободно от* $F(x)$ *(при* H*), то его можно вычислить при любых* m, n. *Например, полагая, что* $X_1, .., X_m$ *и* $Y_1, .., Y_n$ *распределены как* $R(0,1)$. *Но особенно красив ответ при* $m, n \to \infty$.

Теорема 2.2 (**Смирнова**). *Пусть* H *верна. Пусть* $F_X = F_Y = F$, *и* F *непрерывна. Тогда при* $\lambda > 0$

$$\lim_{m,n\to\infty} P\left(\sqrt{\frac{mn}{m+m}} D_{mn} < \lambda\right) = K(\lambda),$$

где $K(\lambda) = 1 - 2\sum_{j\ge1}(-1)^{j+1}e^{-2j^2\lambda^2}$ - ***функция распредления Колмогорова***.

Выберем малое $0 < \alpha < 1$ *и пусть* $\lambda_{1-\alpha}$ *такое число, что* $K(\lambda_{1-\alpha}) = 1 - \alpha$. *Число* $\lambda_{1-\alpha}$ *называют* ***квантилью уровня*** $1 - \alpha$.

Положим $c_\alpha(m,n) = \sqrt{\frac{m+n}{mn}}\lambda_{1-\alpha}$ *- это и есть искомая константа c!*

Правило: $\begin{cases} \text{если } D_{mn} \le c_\alpha(m,n), & \text{то } H \text{ принимаем} \\ \text{если } D_{mn} > c_\alpha(m,n), & \text{то принимаем } K \end{cases}$

Тогда вероятность ***ошибки первого рода****:*

$$P(K|H) = P(\sqrt{\frac{mn}{m+n}}D_{mn} > \lambda_{1-\alpha}) =$$

$$= 1 - P(\sqrt{\frac{mn}{m+n}}D_{mn} \leq \lambda_{1-\alpha}) \to 1 - K(\lambda_{1-\alpha}) = \alpha$$

Можно показать, что $P(H|K) \to 0$ *при* $m, n \to \infty$. *Это вероятность* ***ошибки 2-ого рода****.*

Итак, $\begin{cases} P(H|H) \to \ 1-\alpha \\ P(K|K) \to \ 1 \end{cases}$

(т.е. тест с большой вероятностью выберет правильную гипотезу!)

3 Теорема Гливенко-Кантелли. Метод подстановки

3.1 Теорема Гливенко-Кантелли

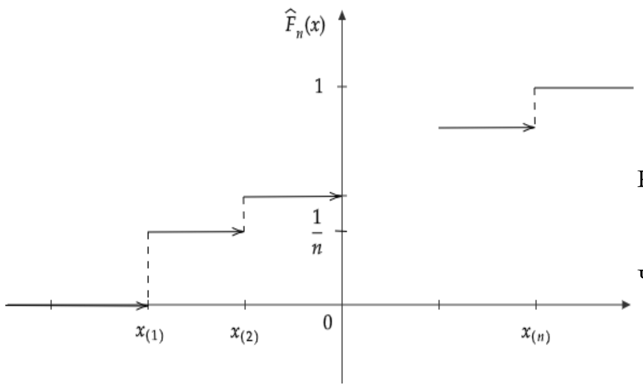
Пусть $X = (X_1, \ldots, X_n)$ - случайный вектор наблюдений. Дальше n будет расти. Поэтому предполагается, что на некотором вероятностном пространстве $(\Omega, \mathcal{F}, P)$ определена бесконечная последовательность н.о.р. случайных величин $X_1, X_2, \ldots$ с неизвестной функцией распределения $F(x)$. Наблюдение X содержит первые n компонент этой последовательности.

Наша цель - оценить $F(x) = P(X_1 \le x), x \in \mathbb{R}$. Зафиксируем $\omega \in \Omega$ и рассмотрим реализации $x_k = X_k(\omega),\ k = 1, \ldots, n$. Пусть $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$.

Определение 3.1. *Случайная величина $X_{(k)}$, равная на упомянутом ω $X_{(k)}(\omega) = x_{(k)},\ k = 1, 2, \ldots, n$ называется* ***к-ой порядковой статистикой****. Совокупность $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ называется* ***вариационным рядом****.*

Определение 3.2. *Оценкой $F(x)$ в точке x возьмем $\hat{F}_n(x) = \frac{1}{n}\sum\limits_{i=0}^{n} I(X_i \le x)$, где I - индикатор. $\hat{F}_n(x)$ называется* ***эмпирической функцией распределения****.*

Если $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ - реализация вариационного ряда, то график реализации $\hat{F}_n(x)$ такой:



При каждом ω $\hat{F}_n(x) = \hat{F}_n(x, \omega)$ - дискретная ф.р.

При фиксированном x $\hat{F}_n(x)$ - случайная величина.

В силу УЗБЧ: $\frac{1}{n}\sum\limits_{i=0}^{n} I(X_i \le x) \xrightarrow[n\to\infty]{\text{п.н.}} EI(X_1 \le x) = F(x)$.

В силу ЦПТ: $\frac{1}{n}(\hat{F}_n(x) - F(x)) = \frac{1}{\sqrt{n}}\sum\limits_{i=0}^{n}(I(X_i \le x) - F(x)) \xrightarrow[n\to\infty]{d} N(0, \hat{F}_n(x) - F^2(x))$, т.к. $DI(X_1 \le x) = EI^2(X_1 \le x) - (EI(X_1 \le x))^2 = F(x) - F^2(x)$.

Докажем следующую важнейшую теорему:

Теорема 3.1 (**Гливенко-Кантелли**). *Пусть* $X_1, \dots, X_n$ *- н.о.р.с.в.,* $X_1 \sim F(x)$. *Тогда*

$$\sup_x |\hat{F}_n(x) - F(x)| \xrightarrow[n\to\infty]{\text{п.н.}} 0$$

Доказательство. Пусть $F(x)$ непрерывна. Пусть $\varepsilon > 0$ - любое малое число, такое что $N = \frac{1}{\varepsilon}$ - целое. Выберем точки $-\infty = z_0 < z_1 < \dots < z_{N-1} < z_N = \infty$ так что $F(z_k) = \dfrac{k}{N}$, $k = 0, 1, \dots, N$. Для $x \in [z_k, z_{k+1})$ в силу монотонности $\hat{F}_n(x)$ имеем:

$$\hat{F}_n(x) - F(x) \le \hat{F}_n(z_{k+1}) - F(z_k) = \hat{F}_n(z_{k+1}) - F(z_{k+1}) + \varepsilon \le \max_k |\hat{F}_n(z_k) - F(z_k)| + \varepsilon$$

$$\hat{F}_n(x) - F(x) \ge \hat{F}_n(z_k) - F(z_{k+1}) = \hat{F}_n(z_k) - F(z_k) - \varepsilon \ge -\max_k |\hat{F}_n(z_k) - F(z_k)| + \varepsilon$$

Из двух последних неравенств получаем:

$$\sup_x |\hat{F}_n(x) - F(x)| \le \max_k |\hat{F}_n(z_k) - F(z_k)| + \varepsilon \qquad (1)$$

Пусть $A_k = \{\omega : \hat{F}_n(z_k) \to F(z_k)\}$. Тогда $P(A_k) = 1$. Пусть $A = \bigcap\limits_k A_k$. Тогда $\forall \omega \in A \ \max\limits_k |\hat{F}_n(z_k) - F(z_k)| \to 0$. Значит:

$$\forall \omega \in A \ \exists n_0 = n_0(\omega) : \ n > n_0 \quad \max_k |\hat{F}_n(z_k) - F(z_k)| < \varepsilon \qquad (2)$$

В силу (1) и (2) для этого ω при $n > n_0$ получаем, что:

$$\sup_x |\hat{F}_n(x) - F(x)| < 2\varepsilon \qquad (3)$$

Так как $P(A) = 1$ и ε произвольно, то (3) означает $\sup\limits_x |\hat{F}_n(x) - F(x)| \xrightarrow[n\to\infty]{\text{п.н.}} 0$. ■

Задача 3.1. *Доказать теорему Гливенко-Кантелли для разрывной $F(x)$.*

(см. [А.А.Боровков. Математическая статистика. Оценка параметров и проверка гипотез. М., Наука, 1984 г.])

3.2 Метод подстановки

Пусть надо оценить параметр $\theta = G(F)$, $G(\cdot)$ - функционал на множестве функций распределения. Естественная оценка подставновки $\hat{\theta}_n = G(\hat{F}_n)$.

Пример 3.1. *Пусть $E|X_1|^k < \infty$, $\nu_k = EX_1^k$, $k \in \mathbb{N}$. ν_k называют* ***к-ым начальным моментом****. Тогда $\nu_k = G(F) = \int\limits_{-\infty}^{\infty} x^k dF(x)$. Оценка подстановки для $\theta = \nu_k$:* $\hat{\theta}_n = \hat{\nu}_k = \int\limits_{-\infty}^{\infty} x^k d\hat{F}_n(x) = \frac{1}{n}\sum\limits_{i=1}^{n} X_{(i)}^k = \frac{1}{n}\sum\limits_{i=0}^{n} X_i^k$.

В силу УЗБЧ: $\hat{\nu}_k = \frac{1}{n}\sum\limits_{i=0}^{n} X_i^k \xrightarrow[n\to\infty]{п.н.} EX_1^k = \nu_k$.

Кроме того, при $EX_1^{2k} < \infty$ имеем в силу ЦПТ:

$$\sqrt{n}(\hat{\nu}_k - \nu_k) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i^k - \nu_k) \xrightarrow{d} N(0, \nu_{2k} - \nu_k^2),\ n \to \infty$$

Значит $(\nu_{2k} - \nu_k^2)^{-\frac{1}{2}}\sqrt{n}(\hat{\nu}_k - \nu_k) \to N(0,1)$. Отсюда:

$$\forall \varepsilon > 0\ P\left((\nu_{2k} - \nu_k^2)^{-\frac{1}{2}}|\sqrt{n}(\hat{\nu}_k - \nu_k)| \le \varepsilon\right) \to \Phi(\varepsilon) - \Phi(-\varepsilon) = 2\Phi(\varepsilon) - 1 \qquad (4)$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-\frac{t^2}{2}}dt$ - функция Лапласа. Асимптотическая нормальность позволила оценить в (4) точность оценки $\hat{\nu}_k$.

Пример 3.2 (**Выборочные квантили**)**.** *Для $0 < p < 1$ и любой (не обязательно непрерывной) функции распредления $F(x)$ полагают $F^{-1}(p) \equiv \sup x : F(x) \le p$. Величина $F^{-1}(p)$ называется* ***квантилью*** *функции распределения $F(x)$ и обозначается далее ξ_p.*

Если $F(x)$ непрерывна и строго возрастает, то $F(\xi_p) = p$.

Пусть $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ - вариационный ряд выборки $X_1, \ldots, X_n$. Оценка ξ_p по методу подстановки:

$$\hat{\xi}_p = \hat{F}_n^{-1}(p) = sup\{x : \hat{F}_n x \leq p\} = X_{([np]+1)}$$

Лемма 3.1. *Пусть функция распределения $F(x)$ непрерывна и строго возрастает. Тогда функционал $G(F) = \xi_p,\ \ 0 < p < 1$ непрерывен в равномерной метрике. Т.е., если последовательность ф.р. $F_n(x)$ такова, что $\sup\limits_x |F_n(x) - F(x)| \to 0$, то $G(F_n) \to G(F)$.*

Доказательство. $\forall\, \varepsilon > 0$ при $n > n_0(\varepsilon)$ имеем:

$$G(F_n) \equiv \xi_p^n = \sup x : F_n(x) \leq p = \sup x : F(x) \leq F(x) - F_n(x) + p \leq$$
$$\leq \sup x : F(x) \leq \sup_y |F_n(y) - F(y)| + p \leq \sup x : F(x) \leq p + \varepsilon = F^{-1}(p + \varepsilon)$$

Аналогично: $\xi_p^n \geq F^{-1}(p-\varepsilon),\ n > n_0$. Значит $F^{-1}(p-\varepsilon) \leq \xi_p^n \leq F^{-1}(p+\varepsilon),\ n > n_0$. Тогда $F^{-1}(p - \varepsilon) \leq \varliminf\limits_{n\to\infty} \xi_p^n \leq \varlimsup\limits_{n\to\infty} \xi_p^n \leq F^{-1}(p + \varepsilon)$. Функция $F^{-1}(t),\ 0 < t < 1$ непрерывна. Устремляя ε к нулю получим:

$$\xi_p = \varliminf_{n\to\infty} \xi_p^n \leq \varlimsup_{n\to\infty} \xi_p^n = \xi_p, \text{ т.е. } \lim_{n\to\infty} \xi_n^p = \xi_p$$

■

Следствие 3.1. *Если $F(x)$ непрерывна и строго возрастает, то $\hat{\xi}_p \xrightarrow[n\to\infty]{п.н.} \xi_p$. Это прямо следует из* ***теоремы Гливенко-Кантелли***.

Определение 3.3. *Величина $\xi_{\frac{1}{2}}$ называется* ***медианой****, а $\hat{\xi}_{\frac{1}{2}}$ -* ***выборочной медианой***.

Теорема 3.2. *Пусть $F(x)$ дифференциируема в точке $\xi_{\frac{1}{2}}$, и $g(\xi_{\frac{1}{2}}) \equiv F'(\xi_{\frac{1}{2}}) > 0$. Тогда:*

$$\sqrt{n}(\hat{\xi}_{\frac{1}{2}} - \xi_{\frac{1}{2}}) \xrightarrow{d} N(0, \frac{1}{4g^2(\xi_{\frac{1}{2}})}),\ \ n \to \infty.$$

3.3 Асимптотическая относительная эффективность оценок (АОЭ)

Асимптотически нормальные оценки можно сравнивать между собой.

Пусть по вектору наблюдений $X = (X_1, \ldots, X_n)$ оценивается параметр θ, и $\hat{\theta}_{1n}$ - его оценка. Пусть

$$\sqrt{n}(\hat{\theta}_{1n} - \theta) \xrightarrow[n\to\infty]{d} N(0, \sigma^2(\theta)) \tag{5}$$

Пусть есть другая оценка $\hat{\theta}_{2n}$, такая что:

$$\sqrt{n}(\hat{\theta}_{1n'} - \theta) \xrightarrow[n\to\infty]{d} N(0, \sigma^2(\theta)) \tag{6}$$

где $n' = n'(n) \xrightarrow[n\to\infty]{} \infty$.

Определение 3.4. ***Асимптотической относительной эффективностью*** *(АОЭ) оценки $\hat{\theta}_{1n}$ относительно оценки $\hat{\theta}_{2n}$ называется величина*

$$l_{1,2} \equiv \lim_{n\to\infty} \frac{n'(n)}{n}.$$

Пусть, например, $l_{1,2} = 3$. Тогда при больших n $n' \approx 3n$. Значит, для $\hat{\theta}_{2n}$ нужно в три раза больше наблюдений, чем для $\hat{\theta}_{1n}$, чтобы достичь одинаковой точности $\sigma^2(\theta)/n$. Оценка $\hat{\theta}_{1n}$ в три раза лучше оценки $\hat{\theta}_{2n}$.

Лемма 3.2. *Пусть* $\sqrt{n}(\hat{\theta}_{in} - \theta) \xrightarrow[n\to\infty]{d} N(0, \sigma_i^2(\theta)),\ \sigma_i^2(\theta) > 0,\ i = 1, 2.$ *Тогда АОЭ существует и равна*

$$l_{1,2} = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}.$$

Доказательство. Пусть $n' \sim \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}n,\ \ n \to \infty$. Тогда

$$\sqrt{n}(\hat{\theta}_{2n'} - \theta) = \sqrt{\frac{n}{n'}}\sqrt{n'}(\hat{\theta}_{2n'} - \theta) \xrightarrow{d} \frac{\sigma_2(\theta)}{\sigma_1(\theta)}\xi,\ \ \xi \sim N(0, \sigma_2^2(\theta))$$

(использовали **лемму Слуцкого**: если $\xi_n \xrightarrow{d} \xi,\ \eta_n \xrightarrow{d} c,$ то $\xi_n\eta_n \xrightarrow{d} c\xi$)

Значит: $\sqrt{n}(\hat{\theta}_{2n'} - \theta) \xrightarrow{d} N(0, \sigma_1^2(\theta))$. Получаем: $\lim\limits_{n\to\infty} \dfrac{n'(n)}{n} = \dfrac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$. ■

Пример 3.3 (Важный). *Пусть $X_i = \theta + \varepsilon_i,\ \ i = 1, .., n,\ \ \varepsilon_i -$ н.о.р.. Пусть $E\varepsilon_1 = 0,\ \ D\varepsilon_1 = \sigma^2 < \infty$. Тогда $EX_1 = \theta$, и оценкой θ можно взять $\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Знаем, что $\sqrt{n}(\overline{X} - \theta) \xrightarrow[n\to\infty]{d} N(0, \sigma^2)$, т.к. $\nu_2 - \nu_1^2 = \sigma^2$.*

Пусть теперь ε_1 имеют ф.р. $G(x)$ и существует плотность вероятности $g(x) = G'(x)$. Пусть $g(x) = g(-x)$ и $g(0) > 0$. Тогда $G(0) = \frac{1}{2}$. Значит ф.р. X_1 имеет вид: $P(X_1 \le \theta) = P(\theta + \varepsilon_1 \le \theta) = P(\varepsilon_1 \le 0) = \dfrac{1}{2}$, т.е. θ-медиана X_1.

Возьмем оценкой выборочную медиану $\hat{\xi}_{\frac{1}{2}}$. Тогда $\sqrt{n}(\hat{\xi}_{\frac{1}{2}} - \theta) \xrightarrow[n\to\infty]{d} N\left(0, \dfrac{1}{4g^2(0)}\right)$, т.к. плотность вероятности X_1 есть $g(x-\theta)$, и $g(x-\theta)|_{x=\xi_{\frac{1}{2}}=\theta} = g(0)$.

Значит, АОЭ выборочной медианы относительно выборочного среднего равна

$$l_{\hat{\xi}_{\frac{1}{2}}, \overline{X}} = \frac{\sigma^2}{\frac{1}{4g^2(0)}} = 4g^2(0)\sigma^2.$$

1) Если $\varepsilon_1 \sim N(0, \sigma^2)$, то $l_{\hat{\xi}_{\frac{1}{2}}, \overline{X}} = 4\left(\dfrac{1}{\sqrt{2\pi}\sigma}\right)^2 \sigma^2 = \dfrac{2}{\pi} \approx 0.637 < 1$. Т.е. если выборочную медиану построить по n наблюдениям, то ту же точность получим для $\overline{X}$ по $0.637n$ наблюдениям! $\overline{X}$ - лучше выборочной медианы в $\frac{\pi}{2}$ раз.
2) Пусть $\varepsilon_1 \sim Lap(\lambda)$, $\quad \lambda > 0$. Тогда $g(x) = \dfrac{\lambda}{2} e^{-\lambda|x|}$. $E\varepsilon_1 = 0, E\varepsilon_1^2 = \dfrac{2}{\lambda^2}$. $l_{\hat{\xi}_{\frac{1}{2}}, \overline{X}} = 4\left(\dfrac{\lambda}{2}\right)^2 \dfrac{2}{\lambda^2} = 2 > 1$. Отсюда, медиана в 2 раза лучше выборочного среднего.

4 Параметрическое оценивание

4.1 Оптимальные и несмещенные оценки

Пусть $X = (X_1, \ldots, X_n)$ - случайное наблюдение, т.е. случайным вектор со значениями в $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Пусть P_X - распределение X, т.е.:

$$P_X(A) = P(X \in A),\ A \in \mathcal{B}(\mathbb{R}^n)$$

Будем предполагать, что $P_X \in \left\{P_\theta \colon \theta \in \Theta \subseteq \mathbb{R}^1\right\}$.
Тройка $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{P_\theta \colon \theta \in \Theta\})$ - статистическая модель.
Распределение P_θ известно с точностью до параметра θ. Его надо оценить.

Определение 4.1. ***Оценка параметра*** *θ - это любая борелевскя функция $\hat{\theta}_n = \varphi(x) \in \mathbb{R}^1$, зависящая только от наблюдений, но не от θ.*

Определение 4.2. ***Качество оценки*** *будем характеризовать средне квадратическим риском:*

$$R_n(\hat{\theta}_n, \theta) := E_\theta(\hat{\theta}_n - \theta)^2$$

Напоминание 4.1.

$$E_\theta\varphi(x) = \int\limits_{\mathbb{R}^n} \varphi(x) P_\theta(dx)$$

Замечание. *Пусть наблюдение $X = (X_1, \ldots, X_n)$ имеет распределение P_X, и определено на вероятностном пространстве $(\Omega, \mathcal{F}, P)$. Обычно явный вид этого пространства в рассмотрении не участвует, но иногда его удобно конкретизировать. Например, пусть X имеет плотность по мере μ, т.е.:*

$$P_X(A) = \int\limits_A p(x)\mu(dx),\ a \in \mathcal{B}(\mathbb{R}^n)$$

Пусть $N_P = \{x \colon p(x) > 0\}$ *- носитель плотности. Тогда полагают:*

$$(\Omega, \mathcal{F}, P) = (N_P, \mathcal{B}(N_P), P_X), \quad X(w) = X(x) = x$$

Здесь $\mathcal{B}(N_p)$ *- сигма-алгебра борелевских подмножеств* N_P*. При таком выборе распределение случайного вектора* $X(x) = x$ *есть* P_X*.*

При $P_X \in \left\{P_\theta \colon \theta \in \Theta \subseteq \mathbb{R}^1\right\}$ *получаем:*

$$(\Omega, \mathcal{F}, P) = (N_P, \mathcal{B}(N_P), P_\theta) \text{ при некотором неизвестном } \theta \in \Theta$$

Определение 4.3. *Оценка* $\hat{\theta}_n$ *называется* ***оптимальной*** *(наилучшей) в средне квадратическом смысле, если:*

$$R_n(\hat{\theta}_n, \theta) \leq R_n(\tilde{\theta}_n, \theta)\ \forall \theta \in \Theta \text{ и } \forall \tilde{\theta}_n \text{ с конечной дисперсией} \tag{7}$$

НО: оптимальные оценки в смысле (7) существуют лишь в вырожденных случаях.

Действительно, положим $\tilde{\theta}_n = \theta_0 \in \Theta$. Тогда $R_n(\tilde{\theta}_n, \theta_0) = 0$, и если (7) верно, то $R_n(\hat{\theta}_n, \theta_0) = E_{\theta_0}(\hat{\theta}_n - \theta_0)^2 = 0$. Т.к. θ_0 может быть любой точкой Θ, получаем:

$$E_\theta(\hat{\theta}_n - \theta)^2 = 0\ \forall \theta \in \Theta$$

Значит, $\hat{\theta}_n = \hat{\theta}_n(X) = \theta$ п.н. по P_θ мере. Это и означает, что ситуация вырожденная, и наблюдение X полностью определяет θ.

Пример 4.1. $X = (X_1)$, $X_1 \sim R(\theta, \theta + 1)$, $\theta \in \Theta = \mathbb{N}$. *Тогда, если* $\hat{\theta}_n = [X_1]$, *то* $\hat{\theta}_n = \theta$ *п.н..*

Сузим класс оценко, и будем искать оптимальные внутри более узкого класса. Ради общности будем далее оценивать $\tau(\theta) \in \mathbb{R}^1$. Оценка $\hat{\tau}_n = \hat{\tau}_n(X) \in \mathbb{R}^1$. Тогда:

$$R_n(\hat{\tau}_n, \tau(\theta)) := E_\theta(\hat{\tau}_n - \tau(\theta))^2 =$$

$$= E_\theta(\hat{\tau}_n - E_\theta\hat{\tau}_n + E_\theta\hat{\tau}_n - \tau(\theta))^2 = D_\theta\hat{\tau}_n + (E_\theta\hat{\tau}_n - \tau(\theta))^2 \tag{1}$$

Определение 4.4. *Величина* $b(\theta) := E\hat{\theta}_n - \tau(\theta)$ *называется* ***смещением оценки*** $\hat{\tau}_n$ *в точке* θ*. Если* $b(\theta) = 0\ \forall \theta \in \Theta$, *то* $\hat{\tau}_n$ *называется* ***несмещенной*** *оценкой.*

Для несмещенной оценки в силу (1): $R_n(\hat{\tau}_n, \tau(\theta)) = D_\theta\hat{\tau}_n$.

Пример 4.2. $X = (X_1, \ldots, X_n)$, $\{X_i\}$ *н.о.р.*, $X_1 \sim N(\theta, \sigma^2)$, $\theta \in \Theta = \mathbb{R}^1$, $\overline{X} = n^{-1}\sum\limits_{i=1}^{n} X_i$. *Тогда* $\overline{X}$ *- несмещенная оценка* $\tau(\theta) = \theta$*.*

Пример 4.3. *$X = (X_1, \dots, X_n)$, $\{X_i\}$ - н.о.р., $X_1 \sim Pois(\theta), \theta > 0$, т.е. $\Theta = \mathbb{R}^+$. Пусть $\tau(\theta) = \frac{1}{\theta}$. Условие несмещнности для $\tau_1(\hat{X_1})$:*

$$E_\theta \hat{\tau}_1(X_1) = \sum_{k \geq 0} \hat{\tau}_1(k) \frac{\theta^k}{k!} e^{-\theta} = \frac{1}{\theta} \; \forall \theta > 0$$

Значит:

$$\sum_{k \geq 0} \hat{\tau}_1(k) \frac{\theta^{k+1}}{k!} = e^\theta = \sum_{r \geq 0} \frac{\theta^r}{r!} \; \forall \theta > 0$$

Но это невозможно, т.е. все коэффициенты рядов должны совпадать, а слева коэффициенты при θ^0 есть ноль, а справа - единица.

Т.о. ***нет несмещенный оценок для*** *$\tau(\theta) = \frac{1}{\theta}$.*

Определение 4.5. *Несмещенная оценка с конечной дисперсией $\hat{\tau}_n = \hat{\tau}_n(X)$ функции $\tau(\theta)$ называется* ***с.к. оптимальной****, если:*

$$D_\theta \hat{\tau}_n \leq D_\theta \tilde{\tau}_n \; \forall \theta \in \Theta \text{ и } \forall \text{ несмещенной } \tilde{\tau}_n \text{ с конечной диспресией}$$

Замечание. *Иногда рассмтаривают класс $\mathbb{C}$ несмещенный оценок с конечной дисперсией и некоторым дополнительным условием, например:*

$$E_\theta \hat{\tau}_n{}^2 = \alpha < \infty \; \forall \theta \in \Theta$$

Тогда в определение надо добавить $\hat{\tau}_n, \tilde{\tau}_n \in \mathbb{C}$. Это с.к. оптимальность в $\mathbb{C}$.

4.2 Неравенство Рао-Крамера и информация Фишера

Пусть распределение P_θ имеет плотность $p(x, \theta)$ в абсолютно непрерывном случае по мере μ. Тогда:

$$E_\theta \varphi(x) = \int\limits_{N_P} \varphi(x) p(x, \theta) \mu(dx) = \begin{cases} \int\limits_{N_p} \varphi(x) p(x, \theta) dx \text{ в абс. непр. случае} \\ \sum\limits_i \varphi(x_i) P(X = x_i) \text{ в дискретном случае} \end{cases}$$

Условие 4.1 (**R**). *Перечислим ряд условий:*

(*i*) *Носитель $N_P = \{x \colon p(x, \theta) > 0\}$ не зависит от θ.*

(ii) Θ - *интервал, и* $\forall x \in N_P$ *существует производная* $\frac{\partial}{\partial\theta}\ln p(x,\theta)$ *при любом* $\theta \in \Theta$.

(iii) (a) *Верно равенство:*

$$\frac{\partial}{\partial\theta}\int\limits_{N_P} p(x,\theta)\mu(dx) = \int\limits_{N_P}\frac{\partial}{\partial\theta}p(x,\theta)\mu(dx) = 0\ \forall\theta\in\Theta$$

(b) *Верно соотношение:*

$$\tau'(\theta) = \frac{\partial}{\partial\theta}\int\limits_{N_P}\hat{\tau}_n(x)p(x,\theta)\mu(dx) = \int\limits_{N_P}\hat{\tau}_n(x)\frac{\partial}{\partial\theta}p(x,\theta)\mu(dx) = 0\ \forall\theta\in\Theta$$

(iv) *Существует величина:*

$$I(\theta) := E_\theta\left(\frac{\partial}{\partial\theta}\ln p(X,\theta)\right)^2,\quad 0 < I(\theta) < \infty$$

$I(\theta)$ *называется* ***информацией Фишера*** *о* θ, *содержащейся в* X.

Теорема 4.1 (**неравенство Рао-Крамера**). *Пусть* $\hat{\tau}_n(x)$ *- несмещенная оценка для* $\tau(\theta)$ *с конечной при всех* $\theta\in\Theta$ *диспресией. Пусть выполнено условие* (R). *Тогда:*

$$D_\theta\hat{\tau}_n(x) \geq \frac{(\tau'(\theta))^2}{I(\theta)}\ \forall\theta\in\Theta$$

Доказательство. В силу условия $(iii)(a)$:

$$E_\theta\frac{\partial}{\partial\theta}\ln p(X,\theta) = \int\limits_{N_P}(\ln p(x,\theta))\,p(x,\theta)\mu(dx) = \int\limits_{N_P}\frac{\partial}{\partial\theta}p(x,\theta)\mu(dx) = 0\ \forall\theta\in\Theta \quad (2)$$

В силу условия $(iii)(b)$ и (2):

$$\tau'(\theta) = \int\limits_{N_P}\hat{\tau}_n(x)\frac{\partial p(x,\theta)}{\partial\theta}\mu(dx) = E_\theta\left(\hat{\tau}_n(x)\frac{\partial}{\partial\theta}\ln p(X,\theta)\right) =$$

$$= E_\theta\left\{(\hat{\tau}_n(x)-\tau(\theta))\,\frac{\partial}{\partial\theta}\ln p(X,\theta)\right\} \quad (3)$$

В силу неравенства Коши-Буняковского: $\{E(\xi\eta)\}^2 \leq E\xi^2\cdot E\eta^2$.

Равенство достигается тогда и только тогда, когда $\eta \overset{\text{п.н.}}{=} a\xi$. Тогда из (3) следует:

$$(\tau'(\theta))^2 = \left\{E_\theta \left\{(\hat{\tau_n}(x) - \tau(\theta)) \frac{\partial}{\partial\theta} \ln p(x,\theta)\right\}\right\}^2 \leq$$
$$\leq E_\theta \left(\hat{\tau_n}(x) - \tau(\theta)\right)^2 \cdot E_\theta \left(\frac{\partial}{\partial\theta} \ln p(x,\theta)\right)^2$$

Т.е. получаем $D_\theta\hat{\tau_n}(x)I(\theta) \geq (\tau'(\theta))^2 \ \forall\theta \in \Theta$. Значит $D_\theta\hat{\tau_n}(x) \geq \dfrac{(\tau'(\theta))^2}{I(\theta)} \ \forall\theta \in \Theta$. ■

Замечание. *Пусть* $X = (X_1, \ldots, X_n)$ *и* $\{X_i\}$ *- н.о.р.,* $X_1 \sim f(x,\theta)$ *по мере* ν. *Тогда* $p(x_1, \ldots, x_n, \theta) \overset{\text{п.н.}}{=} \prod\limits_{i=1}^{n} f(x_i,\theta)$ *по мере* ν.

Предположим, что $\forall\theta \in \Theta$ имеем:

$$E_\theta \frac{\partial}{\partial\theta} \ln f(x,\theta) = 0 \text{ и } 0 < E_\theta \left(\frac{\partial}{\partial\theta} \ln f(X_1,\theta)\right)^2 < \infty$$

Определение 4.6. *Величина* $i(\theta) = E_\theta \left(\frac{\partial}{\partial\theta} \ln f(X_1,\theta)\right)^2$ *называется* ***информацией Фишера*** *о параметре* θ*, содержащейся в одном наблюдении* X_1.

Очевидно, что $i(\theta) = D_\theta \left(\dfrac{\partial}{\partial\theta} \ln f(X_1,\theta)\right)$. Имеем:

$$I(\theta) = E_\theta \left(\frac{\partial}{\partial\theta} \ln f(X_1,\theta)\right)^2 = E_\theta \left(\frac{\partial}{\partial\theta} \ln \prod_{i=1}^{n} f(X_i,\theta)\right)^2 = E_\theta \left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_i,\theta)\right)^2 =$$
$$= D_\theta \left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(X_i,\theta)\right)^2 = nD_\theta \left(\frac{\partial}{\partial\theta} \ln f(X_1,\theta)\right)^2 = ni(\theta)$$

Итак: $I(\theta) = ni(\theta)$ и неравенство Рао-Крамера имеет вид:

$$D_\theta\hat{\tau_n}(x) \geq \frac{(\tau'(\theta))^2}{ni(\theta)} \ \forall\theta \in \Theta$$

4.3 Эффикетивные оценки, необходимое и достаточное условия равенства в НРК

Обозначим $\mathbb{C}_R$ класс несмещенных оценок для $\tau(\theta)$ с конечной дисперсией и удовлетворяющих условию (R).

Определение 4.7. *Если для оценки $\hat{\tau}_n \in \mathbb{C}_R$ в неравенстве Рао-Крамера достигается равенство, т.е.*

$$D_\theta \hat{\tau}_n = \frac{(\tau'(\theta))^2}{I(\theta)} \; \forall \theta \in \Theta$$

то $\hat{\tau}_n$ называется ***эффективной*** *в $\mathbb{C}_R$.*

Тогда:

$$\forall \tilde{\tau}_n \in \mathbb{C}_R \; D_\theta \tilde{\tau}_n = \frac{(\tau'(\theta))^2}{I(\theta)} = D_\theta \hat{\tau}_n \; \forall \theta \in \Theta$$

Значит, эффективная в $\mathbb{C}_R$ оценка является оптимальной в $\mathbb{C}_R$.

Каковы условия равенства в неравенстве Рао-Крамера?

Определение 4.8. *Пусть вектор X имеет плотность $p(x,\theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$, относительно меры μ. Если эта плотность представима в виде:*

$$p(x,\theta) = exp\left\{\sum_{j=1}^{k} a_j(\theta)u_j(x) + b(\theta)\right\}\overline{h}(x), \; x \in N_P$$

то распределение P_θ вектора X принадлежит ***экспоненциальному семейству****.*

Обычно требуют, чтобы функции $a_0(\theta) = 1, a_1(\theta), \ldots, a_k(\theta)$ были именно независимы на Θ.

Задача 4.1. *Пусть $X = (X_1, \ldots, X_n)$, $\{X_i\}$ - н.о.р. Показать: если $X_1 \sim N(\theta, \sigma^2)$, $N(c_1, \theta)$, $N(\theta_1, \theta_2)$, $Exp(\theta)$, $Pois(\theta)$, $Bin(1, \theta)$, то распределение X принадлежит экспоненциальному семейству.*

Теорема 4.2 (**необходимое условие равенства в неравенстве Рао-Крамера**). *Пусть $\hat{\tau}_n$ - несмещенная оценка $\tau(\theta)$, $0 < D_\theta \hat{\tau}_n < \infty \; \forall \theta \in \Theta$. Пусть выполнено условие (R). Пусть функции $\frac{\partial}{\partial \theta} \ln p(x,\theta)$ для $x \in N_P$, $I(\theta)$ и $\tau'(\theta)$ непрерывны по θ. Тогда, если в неравенстве Рао-Крамера достигается равенство, то:*

$$p(x,\theta) = exp\left\{\sum_{j=1}^{k} a_j(\theta)u_j(x) + b(\theta)\right\}\overline{h}(x), \; x \in N_P, \; \theta \in \Theta \tag{4}$$

Замечание. *Теорема 11.1 означает, что если эффективная в $\mathbb{C}_R$ оценка для $\tau(\theta)$ существует, то $p(x,\theta)$ есть плотность из экспоненциального семейства специального вида* (4).

Доказательство. Из доказательства неравенства Рао-Крамера следует, что равенство в этом неравенстве достигается тогда и только тогда, когда при фиксированном $\theta \in \Theta$:

$$\hat{\tau_n}(x) - \tau(\theta) = a(\theta)\frac{\partial}{\partial\theta}\ln p(X,\theta) \ \ (P_\theta \text{ - п.н.}) \tag{5}$$

Таким образом, из последнего равенства надо получить (4). У нас $(\Omega, \mathcal{F}, P) = (N_P, \mathcal{B}(N_P), P_\theta),\ X(x) = x$. Поэтому упомянутое равенство (5) эквивалентно:

$$\hat{\tau_n}(x) - \tau(\theta) = a(\theta)\frac{\partial}{\partial\theta}\ln p(x,\theta) \text{ для } P_\theta\text{-п.в., } x \in N_P \tag{6}$$

При фиксированном θ соотношение (6) не выполнено при $x \in A_\theta,\ P_\theta(A_\theta) = 0$. При $x \in \overline{A_\theta}$ (6) выполнено, $P_\theta(\overline{A_\theta}) = 1$ ($A_\theta \in N_P,\ \overline{A_\theta} = N_P \setminus A_\theta$).

Рассмотрим (5). Домножим (5) на $\frac{\partial}{\partial\theta}\ln p(X,\theta)$ и возьмем среднее:

$$E_\theta\left\{\hat{\tau_n}(x)\frac{\partial}{\partial\theta}\ln p(X,\theta)\right\} = a(\theta)I(\theta) = \tau'(\theta)$$

(воспользовались условием (R))

Значит, $a(\theta) = \frac{\tau'(\theta)}{I(\theta)}$ - непрерывная функция, и $a(\theta) \neq 0$, т.к. $\tau'(\theta) \neq 0$ из-за условия $D_\theta\hat{\tau_n} > 0$.

Рассмотрим (6). $P_\theta(A_\theta) = \int\limits_{A_\theta} p(x,\theta)\mu(dx) = 0 \ \Rightarrow\ \mu(A_\theta) = 0$. Пусть $A = \bigcup\limits_{\theta\in\mathbb{Q}} A_\theta$, тогда $\mu(A) = 0$, но $\overline{A} = \bigcap\limits_{\theta\in\mathbb{Q}} \overline{A_\theta}$, и при $x \in \overline{A}$ соотношение (6) выполнено при всех рациональных θ. Но левая и правая части (6) непрерывны по θ. Значит, при $x \in \overline{A}$ (6) верно при всех θ. Тогда при любом $x \in \overline{A}$ из (6) следует:

$$\frac{\partial}{\partial\theta}\ln p(x,\theta) = \frac{\hat{\tau_n}(x)}{a(\theta)} - \frac{\tau(\theta)}{a(\theta)} \ \Rightarrow$$

$$\Rightarrow\ \ln p(x,\theta) = \hat{\tau_n}(x)\int\limits_{\theta_1}^{\theta}\frac{d\theta}{a(\theta)} - \int\limits_{\theta_1}^{\theta}\frac{\tau'(\theta)}{a(\theta)}d\theta + \ln p(x,\theta_1)$$

$$\int\limits_{\theta_1}^{\theta} \frac{d\theta}{a(\theta)} = A(\theta), \quad -\int\limits_{\theta_1}^{\theta} \frac{\tau'(\theta)}{a(\theta)} d\theta = B(\theta), \quad \ln p(x, \theta_1) = \overline{h}(x)$$

Отсюда: $p(x, \theta) = exp\{\hat{\tau}_n(x)A(\theta) + B(\theta)\}\overline{h}(x), \ x \in \overline{A}$. На множестве $A, \ \mu(A) = 0$, значения плотности вещественны. Т.о. (4) верно при всех $x \in N_P, \ \theta \in \Theta$. ■

Теорема 4.3 (**достаточное условие равенства в неравенстве Рао-Крамера**). *Пусть $\hat{\tau}_n$ - несмещенная оценка $\tau(\theta)$, $0 < D_\theta \hat{\tau}_n < \infty \ \forall \theta \in \Theta$. Пусть выполнено условие (R). Тогда, если:*

$$p(x, \theta) = exp\{\hat{\tau}_n(x)A(\theta) + B(\theta)\}\overline{h}(x), \ x \in N_P \tag{7}$$

то в неравенстве Рао-Крамера достигается равенство.

Доказательство. В силу (7) при $x \in N_p, \ \theta \in \Theta$:

$$\ln p(x, \theta) = \hat{\tau}_n(x)A(\theta) + B(\theta) + \ln \overline{h}(x)$$

Значит:

$$\frac{\partial}{\partial \theta} \ln p(x, \theta) = \hat{\tau}_n(x)A'(\theta) + B'(\theta) = A'(\theta)\left(\hat{\tau}_n(x) + \frac{B'(\theta)}{A'(\theta)}\right) =$$
$$= A'(\theta)(\hat{\tau}_n(x) - \tau(\theta)), \ x \in N_P, \ \theta \in \Theta$$

Последнее соотношение влечет (6), а значит и (5). ■

Итак, в силу теорем 2 и 3 равенство в неравенстве Рао-Крамера достигается лишь для плотностей

$$p(x, \theta) = exp\left\{\hat{\tau}_n(x)A(\theta) + B(\theta)\right\}\overline{h}(x), \ x \in N_P, \ \theta \in \Theta,$$
$$\text{причем } -\frac{B'(\theta)}{A'(\theta)} = \tau(\theta)$$

Это очень специальный вид плотности из экспоненциального семейства. Т.о., эффикетивных оценок мало.

Пример 4.4. $X = (X_1, \ldots, X_n)$, $\{X_i\}$ - *н.о.р.*, $X_1 \sim N(\theta, \sigma^2)$, $\theta \in\in \Theta = \mathbb{R}^1$. $\tau(\theta) = \theta$. *Найти эффективную оценку.*

Решение. Здесь $\tau(\theta) = \theta$.

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

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

Здесь $\hat{\tau_n}(x) = \overline{X},\ A(\theta) = \frac{n\theta}{\sigma^2},\ B(\theta) = \frac{n\theta^2}{2\theta^2},\ -\frac{B'(\theta)}{A'(\theta)} = \theta = \tau(\theta)$.
Прочие условия теоремы 3 выполнены. В силу теоремы 3 $\hat{\tau_n}(x) = \overline{X}$ - эффективная оценка $\tau(\theta) = \theta$. ■

Можно показать, что если некоторая функция $\tau(\theta)$ допускает эффективное оценивание $\hat{\tau_n}(x)$, то эффективно можно оценить еще функцию $a\tau(\theta)+b$ (a,b - константы) и никакие другие. Оценка - $a\hat{\tau_n}(x)+b$.

Значит, в последнем примере все функции, допускающие эффективное оценивание, имеют вид $\tau(\theta) = a\theta+b$, а их оценки $\hat{\tau_n}(x) = a\overline{X}+b$.

5 Оценивание в многопараметрическом случае

5.1 Основные понятия

Пусть $A = (a_{ij})_{i,j=1,2,\dots,m}$ - $m \times m$-матрица, $a_{ij} \in \mathbb{R}^1$.

- A - симметрическая (симметричная), если $A = A^T$
- симметрическая матрица A неотрицательно определена ($A \geq 0$), если $\alpha^T A\alpha \geq 0\ \forall \alpha \in \mathbb{R}^m$
  $A \geq 0\ \Leftrightarrow$ собственные числа $\lambda_i \geq 0,\ i = 1, 2, \dots, m$
- симметрическая $A > 0$, если $\alpha^T A\alpha > 0\ \forall \alpha \in \mathbb{R}^m,\ \alpha \neq 0$
  $A > 0\ \Leftrightarrow$ собственные числа $\lambda_i > 0,\ i = 1, 2, \dots, m$

Пусть случайный вектор ξ определен на $(\Omega, \mathcal{F}, P)$ и принимает значения в $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$, $\xi = (\xi_1, \dots, \xi_m)^T$.

- ξ - случайный вектор $\Leftrightarrow$ ξ_i - случайные величины, $i = 1, 2, \dots, m$
- $E\xi := (E\xi_1, \dots, E\xi_m)^2$, $E|\xi| < \infty\ \Leftrightarrow\ E|\xi_i| < \infty$
- $cov(\xi, \xi) = D\xi := E(\xi - E\xi)(\xi - E\xi)^T$
  $D\xi$ существует $\Leftrightarrow$ $D\xi_i < \infty$
- $D\xi = (D\xi)^T$, т.е. ковариационная матрица является симметрической
- $D\xi \geq 0$, т.е. $\alpha^T D\xi\alpha \geq 0\ \forall \alpha \in \mathbb{R}^m$

Пусть $X = (X_1, \dots, X_n)$ - случайное наблюдение со значениями в $(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$. Пусть $X \sim P_X$ - распределение. Будем предполагать далее, что $P_X \in \{P_\theta,\ \theta \in \Theta \subseteq \mathbb{R}^k\}$.

Необходимо оценить функцию $\tau(\theta) = (\tau_1(\theta), \dots, \tau_m(\theta))^T$. Оценка - $\hat{\tau}_n(X) = (\hat{\tau}_{1n}(X), \dots, \hat{\tau}_{mn}(X))^T$, скалярные борелевские функции $\hat{\tau}_{in}(X)$ не зависят от θ, но зависят от X.

Определение 5.1. *Оценка $\hat{\tau}_n(X)$ функции $\tau(\theta)$ называется* ***несмещенной****, если*

$$E_\theta \hat{\tau}_n(X) := (E_\theta \hat{\tau}_{1n}(X), \ldots, E_\theta \hat{\tau}_{mn}(X))^T = (\tau_1(\theta), \ldots, \tau_m(\theta))^T \ \forall \theta \in \Theta \subseteq \mathbb{R}^m$$

Ковариационная матрица несмещенной оценки: $D_\theta \hat{\tau}_n := E_\theta(\hat{\tau}_n - \tau(\theta))(\hat{\tau}_n - \tau(\theta))^T$ - это симметрическая неотрицательная определенная $(m \times m)$-матрица.

Определение 5.2. *Если $\hat{\tau}_n$ - несмещенная оценка для $\tau(\theta)$ с конечной ковариационной матрицей и*

$$D_\theta \hat{\tau}_n(X) \leq D_\theta \tilde{\tau}_n(X) \ \forall \theta \in \Theta \tag{1}$$

(где $\tilde{\tau}_n$ - любая несмещенная оценка $\tau(\theta)$ с конечной ковариационной матрицей), то $\hat{\tau}_n(X)$ называется ***оптимальной*** *в с.к. смысле.*

Неравенство (1) означает, что $\alpha^T D_\theta \hat{\tau}_n \alpha \leq \alpha^T D_\theta \tilde{\tau}_n \alpha \ \forall \alpha \in \mathbb{R}^n \ \forall \theta \in \Theta$.
Разумеется, если $\hat{\tau}_n$ - с.к. оптимальная оценка $\tau(\theta)$, то $\hat{\tau_{in}}$ - оптимальные оценки для $\tau_i(\theta)$.

Существует ли равномерная нижняя граница для $D_\theta \hat{\tau}_n$?

5.2 Многомерное неравенство Рао-Крамера

Многомерное неравенство Рао-Крамера

- Если $\theta = (\theta_1, \ldots, \theta_k)^T$, $\varphi(x, \theta) \in \mathbb{R}^1$, то

$$\frac{\partial}{\partial \theta} \varphi(x, \theta) := \left(\frac{\partial}{\partial \theta_1} \varphi(x, \theta), \ldots, \frac{\partial}{\partial \theta_k} \varphi(x, \theta) \right)^T$$

(вектор-столбец размера k)

- Если $\varphi(x, \theta) = (\varphi_1(x, \theta), \ldots, \varphi_m(x, \theta))^T$, то

$$\frac{\partial}{\partial \theta} \varphi(x, \theta) := \left(\frac{\partial}{\partial \theta} \varphi_1(x, \theta), \ldots, \frac{\partial}{\partial \theta} \varphi_m(x, \theta) \right) = \left(\frac{\partial}{\partial \theta_i} \varphi_j(x, \theta) \right)$$

($(k \times m)$-матрица)

Условие (RM)

(i) Θ - прямоугольник, т.е. $a_i < \theta_i < b_i, \ i = 1, 2, \ldots, k$

(ii) $X \sim p(x, \theta)$ по мере μ; носитель $N_p = \{x \colon p(x, \theta) > 0\}$ не зависит от θ, и $\forall x \in N_p$ существует $\frac{\partial}{\partial \theta} \ln p(x, \theta)$ при всех $\theta \in \Theta$

(iii) (a) $\frac{\partial}{\partial\theta}\int\limits_{N_p} p(x,\theta)\mu(dx) = \int\limits_{N_p} \frac{\partial p(x,\theta)}{\partial\theta}\mu(dx) = 0\ \forall\theta\in\Theta$

(b) $\left(\frac{\partial}{\partial\theta}\int\limits_{N_p} \hat{\tau_n}(x)p(x,\theta)\mu(dx)\right)^T = \int\limits_{N_p} \hat{\tau_n}(x)\left(\frac{\partial}{\partial\theta}p(x,\theta)\right)^T\mu(dx)\ \forall\theta\in\Theta,$

где $\left(\frac{\partial}{\partial\theta}\int\limits_{N_p} \hat{\tau_n}(x)p(x,\theta)\mu(dx)\right)^T$ - $(m\times k)$-матрица

(iv) если $I(\theta) := E_\theta\left(\frac{\partial}{\partial\theta}\ln p(x,\theta)\right)\left(\frac{\partial}{\partial\theta}\ln p(x,\theta)\right)^T$ - информация Фишера, то $0 < I(\theta) < \infty\ \forall\theta\in\Theta$, $\left(\frac{\partial}{\partial\theta}\ln p(x,\theta)\right)$ - $(k\times k)$-матрица.

$$I(\theta) = \left(E_\theta\frac{\partial}{\partial\theta_i}\ln p(x,\theta)\frac{\partial}{\partial\theta_j}\ln p(x,\theta)\right)_{i,j=1,2,\ldots,k}$$

Теорема 5.1 (**векторное неравенство Рао-Крамера**). *Пусть $\hat{\tau_n}(X)$ - несмещенная оценка $\tau(\theta)$ с конечной ковариационной матрицей $D_\theta\hat{\tau_n}(X)$. Пусть выполнено условие (RM). Тогда:*

$$D_\theta\hat{\tau_n}(X) \geq \left(\frac{\partial\tau(\theta)}{\partial\theta}\right)^T I^{-1}(\theta)\frac{\partial\tau(\theta)}{\partial\theta}\ \forall\theta\in\Theta$$

Если в этом неравенстве достигается равенство, то $\hat{\tau_n}(X)$ называется **эффективной** в классе $\mathbb{C}_{RM}$. Тогда $p(x,\theta) = exp\{\hat{\tau_n}^T(x)A(\theta)+B(\theta)\}h(x)$ для некоторых специальных $A(\theta), B(\theta)$, $x\in N_p$, $\theta\in\Theta$, т.е. распределение X принадлежит экспоненциальному семейству очень специального вида. (см. про матричное неравенство Коши-Буняковского в пар. 16, гл. 2, А.А. Боровков, Мат. стат. оценка пов., пров. гип.).

Пусть $X = (X_1,\ldots,X_n)$ - наблюдение, и $\{X_i\}$ - н.о.р.с.в. Пусть X_1 имеет плотность $f(x,\theta)$, $\theta\in\Theta\subseteq\mathbb{R}^k$, по мере ν.

Предположим, что при $x\in N_f$ существует $\frac{\partial}{\partial\theta}\ln f(x,\theta)$, $E_\theta\frac{\partial}{\partial\theta_i}\ln f(X_1,\theta) = 0$, $E_\theta\left\{\frac{\partial}{\partial\theta_i}\ln f(X_1,\theta)\right\}^2 < \infty$, $\theta\in\Theta$, $i = 1,2,\ldots,k$.

Тогда существует информация фишера о параметре θ, содержащаяся в одном наблюдении X_1 (матрица информации фишера):

$$I_1(\theta) := \left(E_\theta\frac{\partial}{\partial\theta_i}\ln f(x_1,\theta)\cdot\frac{\partial}{\partial\theta_j}f(x_1,\theta)\right),\ i,j = 1,2,\ldots,k$$

Поскольку $I_1(\theta)$ - ковариационная матрица вектора $\frac{\partial}{\partial\theta}\ln f(x,\theta)$, то $I_1(\theta)\geq 0\ \forall\theta\in$

Θ. Если $detI(\theta) \neq 0$, то $I_1(\theta) > 0$.
Рассуждая как в одномерном случае (т.е. при $k = 1$) получим: $I(\theta) = nI_1(\theta)$. Для н.о.р. наблюдений информация $I(\theta)$ есть сумма информаций $I_1(\theta)$. Тогад неравенство Рао-Крамера (2) приобретает вид:

$$D_\theta \hat{\tau_n}(x) \geq \left(\frac{\partial\tau(\theta)}{\partial\theta}\right)^T (nI_1(\theta))^{-1}\frac{\partial\tau(\theta)}{\partial\theta},\ \theta \in \Theta \tag{3}$$

Важный пример
Пусть $X = (X_1, \dots, X_n)$, $\{X_i\}$ - н.о.р.с.в., $n \geq 2$, $X_1 \sim N(\theta_1, \theta_2)$, где $\theta_1 \in \mathbb{R}^1$, $\theta_2 > 0$ (т.е. из $\mathbb{R}^+$). Пусть $\tau(\theta) = (\theta_1, \theta_2)^T$, оценка $\hat{\tau_n}(X) = (\overline{X}, S^2)^T$, где $\overline{X} = n^{-1}\sum\limits_{i=1}^n X_i$, $S^2 = \frac{1}{n-1}\sum\limits_{i=1}^n (X_i - \overline{X})^2$.

Определение 5.3. *Если $\xi_1, \dots, \xi_k$ - н.о.р. стандартные гауссовские $N(0,1)$ сл.в., то сл.в. $\eta_k := \xi_1^2 + \dots + \xi_k^2$ имеет распределение* ***хи-квадрат Пирсона*** *с k степенями свободы. Пишем $\eta_k \sim \chi^2(k)$.*

Очевидно, что $E\eta_k = kE\xi_1^2 = k$, $D\eta_k = kD\xi_1^2 = k\left(E\xi_1^4 - (E\xi_1^2)^2\right) = k(3-1) = 2k$.

Задача 5.1. *Пусть $\xi \sim N(0, \sigma^2)$. Проверить, что $E\xi^{2k} = 1 \cdot 3 \cdot \dots \cdot (2k-1)\sigma^{2k}$.*

Очевидно, что $\overline{X} \sim N(\theta_1, \frac{\theta_2}{n})$. Вскоре будет показано, что $\frac{(n-1)S^2}{\theta_2} \sim \chi^2(n-1)$, и величины $\overline{X}$ и S^2 независимы. Значит, $D_\theta\frac{(n-1)S^2}{\theta_2} = \frac{(n-1)^2}{\theta_2^2}D_\theta S^2 = 2(n-1)$, т.е. $D_\theta S^2 = \frac{2\theta_2^2}{n-1}$. Значит, ковариационная матрица $D_\theta\hat{\tau_n} = \begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n-1}\end{pmatrix}$.
Найдем информационную матрицу фишера $I_1(\theta) = (i_{ij}(\theta))$, $i, j = 1, 2$. Имеем:

$$f(x,\theta) = \frac{1}{\sqrt{2\pi\theta_2}}e^{-\frac{1}{2\theta_2}(x-\theta_1)^2}$$

$$\ln f(x,\theta) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\theta_2 - \frac{1}{2\theta_2}(x-\theta_1)^2$$

$$\frac{\partial \ln f(x,\theta)}{\partial\theta_1} = \frac{x-\theta_1}{\theta_2},\ \frac{\partial \ln f(x,\theta)}{\partial\theta_2} = \frac{(x-\theta_1)^2 - \theta_2}{2\theta_2^2}$$

$$i_{1,1}(\theta) = E_\theta\frac{(x_1-\theta_1)^2}{\theta_2^2} = \frac{1}{\theta_2},\ i_{2,2} = E_\theta\left\{\frac{(x-\theta_1)^2-\theta_2}{2\theta_2^2}\right\}^2 = \frac{1}{3\theta_2^2}D_\theta\frac{(x_1-\theta_1)^2}{\theta_2} = \frac{1}{2\theta_2^2}$$

$$i_{1,2}(\theta) = i_{2,1}(\theta) = E_\theta\frac{x_1-\theta_1}{\theta_2}\cdot\frac{(x_1-\theta_1)^2-\theta_2}{2\theta_2} = 0$$

$$I_1(\theta) = \begin{pmatrix} \frac{1}{\theta_2} & 0 \\ 0 & \frac{1}{2\theta_2^2}\end{pmatrix}$$

Т.к. $\frac{\partial\tau(\theta)}{\partial\theta} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E_2$, то неравенство Рао-Крамера имеет вид:

$$\begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n-1} \end{pmatrix} \geq \begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n} \end{pmatrix} \ \forall\theta \in \Theta$$

Неравенство верное, но равенства нет, т.е. $\hat{\tau}_n(X) = (\overline{X}, S^2)^T$ неэффективная оценка $\tau(\theta) = (\theta_1, \theta_2)^T$.

Далее покажем, что $\hat{\tau}_n(X)$ - оптимальная оценка для $\tau(\theta)$.

6 УМО и условные распределения

6.1 Определение условного математического ожидания

ШАГ 1:

Пусть имеются вероятностное пространство $(\Omega, \mathcal{F}, P)$, случайная величина $\xi : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, причем $E|\xi| < \infty$, где $E\xi = \int\limits_{\Omega} \xi(\omega) P(d\omega)$.

Пусть $A \in \mathcal{F}$. Тогда для $C \in \mathcal{F}$ $P(C|A) = \dfrac{P(CA)}{P(A)} = P_A(C)$.

Если $CA = \varnothing$, то $P_A(C) = 0$, а если $C \subset A$, то $P_A(C) = \frac{P(C)}{P(A)}$.

Имеем новое вероятностное пространство $(\Omega, \mathcal{F}, P_A)$.

Естественно положить $E(\xi|A) = \displaystyle\int\limits_{\Omega} \xi(\omega) P_A(d\omega)$. Тогда:

$$E(\xi|A) = \int\limits_{A} \xi(\omega) P_A(d\omega) + \int\limits_{\overline{A}} \xi(\omega) P_A(d\omega) = \frac{1}{P(A)} \int\limits_{A} \xi(\omega) P(d\omega)$$

Итак: $E(\xi|A) = \dfrac{1}{P(A)} \displaystyle\int\limits_{A} \xi(\omega) P(d\omega) = \dfrac{E(\xi I(A))}{P(A)} = \dfrac{E(\xi, A)}{P(A)}$.

ШАГ 2: Рассмотрим $(\Omega, \mathcal{F}, P)$, и пусть $\{A_1, A_2, ..\}$ - разбиение Ω, т.е.

$$A_1 + A_2 + .. = \Omega, \;\; A_i A_j = \varnothing \; i \neq j, \;\; P(A_i) > 0.$$

Рассмотрим сигма-алгебру $U = \sigma\{A_1, A_2, ..\}$. Элементы U - всевозможные объединения $A_1, A_2, ...$

Пусть ξ - сл.в., определенная на $(\Omega, \mathcal{F})$ со значениями в $(\mathbb{R}, B(\mathbb{R}))$, $E|\xi| < \infty$.

Определение 6.1. ***Условным математическим ожиданием*** *сл.вел.* ξ ***относительно*** U *называется случайная величина*

$$\hat{\xi} = E(\xi|U) = \sum_{k\geq 1} \frac{E(\xi, A_k)}{P(A_k)} I(A_k) \tag{1}$$

где $E(\xi, A_k) = \int\limits_{A_k} \xi(\omega) P(d\omega)$. *Т.е.* $E(\xi|U) = \sum\limits_{k\geq 1} E(\xi|A_k) I(A_k)$.

Имеем следующую лемму (без доказательства):

Лемма 6.1. *Пусть* U_ξ *- сигма-алгебра, порожденная* ξ. *Тогда сл.в.* η *является* U_ξ*- измеримой тогда и только тогда, когда* $\eta = \varphi(\xi)$ *для некторой борелевской* φ.

С помощью этого утверждения можно вывести два свойства УМО из (1):

1) Сл.в. $\hat{\xi}$ измерима относительно U.

   Доказательство. U порождается сл.в. $\xi = \sum\limits_{k\geq 1} c_k I(A_k)$, где все c_k различны. Тогда сл.в. η будет U-измерима тогда и только тогда, когда $\eta = \phi(\xi) = \sum\limits_{k\geq 1} b_k I(A_k)$. Но именно такой вид имеет $\hat{\xi}$ из (1). ■

2) $\forall A \in U \;\; E(\hat{\xi}, A) = E(\xi, A)$.

   Доказательство. Так как имеется представление $A = \sum\limits_k A_{jk}$, то

$$E(\hat{\xi}, A) = \sum_k E(\hat{\xi}, A_{jk}) = \sum_k E(\hat{\xi} I(A_{jk})) =$$
$$= \sum_k E(\frac{E(\xi I(A_{jk}))}{P(A_{jk})} I(A_{jk})) = \sum_k E(\xi I(A_{jk})) = E(\xi, A).$$

■

Лемма 6.2. *Свойства* 1) *и* 2) *выше однозначно определяют УМО и эквивалентны определению* (1).

Доказательство. Уже доказано, что определение (1) влечет свойства 1), 2).

Обратно, пусть для некоторой сл.в. $\hat{\xi}$ выполнены свойства 1) и 2) Тогда в силу 1): $\hat{\xi} = \sum\limits_{k\geq 1} c_k I(A_k)$. В силу 2): $E(\hat{\xi}, A_j) = E(\xi, A_j) = E(c_j I(A_j)) = c_j P(A_j)$, т.е. $c_j = E(\xi, A_j)/P(A_j)$. ■

ШАГ 3 (определение УМО для произвольных U):

Определение 6.2. *Пусть ξ есть сл.вел. (или сл. вектор) на $(\Omega, \mathcal{F}, P)$, и $U \subset \mathcal{F}$. Пусть $E|\xi| < \infty$. Тогда* ***условным математическим ожиданием*** *ξ* ***относисительно*** *U называется сл.в. (или сл.вектор той же размерности, что и ξ), которая обладает двумя свойствами:*

1) *$\hat{\xi}$ измерима относительно U;*
2) *$\forall A \in U \;\; E(\hat{\xi}, A) = E(\xi, A)$, т.е. $\int\limits_A \hat{\xi}(\omega) P(d\omega) = \int\limits_A \xi(\omega) P(d\omega)$.*

УМО будет обозначаться так же $E(\xi|U)$. Заметим, что если $E|\xi| < \infty$, то $E\hat{\xi} = \int\limits_\Omega \hat{\xi}(\omega) P(d\omega) = \int\limits_\Omega \xi(\omega) P(d\omega) = E\xi$ и, значит, $E|\hat{\xi}| < \infty$.

Теорема 6.1. *Если $E|\xi| < \infty$, то УМО $\hat{\xi}$ в определении 11.16 всегда существует и единственно с точностью до значений на множестве меры нуль.*

О производной Радона-Никодима

Определение 6.3. *Пусть на измеримом пространстве $(\Omega, \mathcal{F})$ заданы сигма-конечные меры μ и λ. Меру λ называют* ***абсолютно непрерывной относительно меры*** *μ, если из равенства $\mu(A) = 0, \; A \in \mathcal{F}$, следует $\lambda(A) = 0$. Пишут $\lambda \prec \mu$.*

Теорема 6.2 (**Радона-Никодима**). *Пусть на $(\Omega, \mathcal{F})$ заданы σ-конечные меры μ и λ. Тогда $\lambda \prec \mu$ тогда и только тогда, когда существует $\mathcal{F}$-измеримая функция $f(\omega) \geq 0$, для которой $\lambda(A) = \int\limits_A f(\omega)\mu(d\omega) \;\; \forall \; A \in \mathcal{F}$. Функция $f(\omega)$ единственна с точностью до значений на множестве μ-меры нуль. Функция $f(\omega)$ называется* ***производной Радона-Никодима*** *меры λ по мере μ. Пишут $f(\omega) = \frac{d\lambda}{d\mu}(\omega)$.*

Докажем теорему 6.1.

Доказательство.

1) ξ - скалярная сл.в., $\xi \geq 0$. Введем функцию множеств

$$Q(A) = \int\limits_A \xi(\omega)P(d\omega) = E(\xi, A), \quad A \in U$$

Тогда [Ширяев, Вероятность, гл. III, § 6]:

- a) $Q(A) \geq 0, \quad Q(\Omega) = E\xi < \infty$;
- b) Если $A = \sum\limits_i A_i, \quad A_iA_j = \varnothing \; i \neq j$, то $Q(A) = \sum\limits_i Q(A_i)$;
- c) Если $P(A) = 0$, то $Q(A) = 0$.

Свойства a)-c) означают, что $Q(A)$ есть конечная σ-аддитивная мера, и $Q \prec P$. В силу **теоремы Радона -Никодима** существует U-измеримая функция $\hat{\xi}$, такая что

$$Q(A) = \int\limits_A \xi(\omega)P(d\omega) = \int\limits_A \hat{\xi}(\omega)P(d\omega).$$

Эта функция почти наверное единственна.

2) Пусть ξ - скалярная и не обязательно неотрицательная. Тогда $\xi = \xi^+ - \xi^-$, где $\xi^+ = \max(0, \xi) \geq 0$, $\xi^- = \max(0, -\xi)$. Положим $\hat{\xi} = \hat{\xi^+} - \hat{\xi^-}$. Очевидно, $\hat{\xi}$ — U-измерима, т.к. $\hat{\xi^+}$ и $\hat{\xi^-}$ - U-измеримы. Далее:

$$E(\hat{\xi}, A) = E(\hat{\xi^+}, A) - E(\hat{\xi^-}, A) = E(\xi^+, A) - E(\xi^-, A) = E(\xi, A), \; A \in U$$

Наконец, если $\overline{\xi}$ - U-измерима и $E(\overline{\xi}, A) = E(\xi, A) \; \forall A \in U$, то $\overline{\xi} = \hat{\xi}$ почти наверное [Ширяев, Вероятность, гл. II, § 6]. Здесь использовалось предположение $E|\hat{\xi}| < \infty, \; E|\overline{\xi}| < \infty$.

3) Пусть $\xi = (\xi_1, \ldots, \xi_s)^T$. Тогда $\hat{\xi} = (\hat{\xi}_1, \ldots, \hat{\xi}_s)^T$. Измеримость следует из измеримости $\hat{\xi}_i, \; i = 1, \ldots, s$. Далее:

$$E(\hat{\xi}, A) = (E(\hat{\xi}_1, A), \ldots, E(\hat{\xi}_s, A))^T = (E(\xi_1, A), \ldots, E(\xi_s, A))^T = E(\xi, A)$$

Наконец, если $E(\overline{\xi}, A) = E(\hat{\xi}, A), \; \forall \, A \in F$, то $E(\overline{\xi_i}, A) = E(\hat{\xi}_i, A) \; \Rightarrow \; \overline{\xi_i} \overset{\text{п.н.}}{=} \hat{\xi}_i$, т.е. $\overline{\xi} \overset{\text{п.н.}}{=} \hat{\xi}$.

■

6.2 Свойства условного математического ожидания

Далее всегда $E|\xi| < \infty$. Следующие 6 утверждений верны и для скаляров и для векторов, причем соотношения $\xi_1 \leq \xi_2$, $\xi_n \uparrow \xi$ понимаются для векторов покомпонентно. Пусть также сигма-алгебра $U \subset \mathcal{F}$.

Утверждение 6.1. *Имеем следующие свойства:*

a) $E(c\xi|U) = cE(\xi|U)$ *п.н.*
b) $E(\xi_1 + \xi_2|U) = E(\xi_1|U) + E(\xi_2|U)$ *п.н.*
c) *Если* $\xi_1 \leq \xi_2$ *п.н.,* $E(\xi_1|U) \leq E(\xi_2|U)$ *п.н.*

Доказательство.

a) Напомним, что $\hat{\xi} = E(\xi|U)$ такая сл.величина (вектор), что:

$$\hat{\xi} \text{ - } U - \text{измеримая сл.в.} \tag{5}$$

$$E(\hat{\xi}, A) = E(\xi, A)\ \forall A \in U, \text{ т.е. } \int\limits_A \hat{\xi}(\omega)P(d\omega) = \int\limits_A \xi(\omega)P(d\omega) \tag{6}$$

Значит, надо показать, что:

1) $c\hat{\xi}$ - $U-$измерима,
2) $\forall\ A \in U\ \ E(c\hat{\xi}, A) = E(c\xi, A)$

Соотношение 1) очевидно, если $\hat{\xi}$ - U-измерима. Докажем 2): $E(c\hat{\xi}, A) = cE(\hat{\xi}, A) = cE(\xi, A) = E(c\xi, A)$.

b) Доказательство b) аналогично доказательству a).

c) Пусть $\hat{\xi}_i = E(\xi_i|U)$, $i = 1, 2$.

Тогда $\forall A \in U E(\hat{\xi}_1, A) = \int\limits_A \hat{\xi}_1 P(d\omega) = E(\xi_1, A) \leq E(\xi_2, A) = \int\limits_A \hat{\xi}_2 P(d\omega)$.

Значит, $\int\limits_A (\hat{\xi}_2 - \hat{\xi}_1)P(d\omega) \geq 0\ \forall A \in U$ и $\hat{\xi}_2 - \hat{\xi}_1 \geq 0$ п.н.

■

Утверждение 6.2. *Если сигма-алгебра U и сигма-алгебра $\sigma(\xi)$ независимы, то $E(\xi|U) = E\xi$ п.н.*

Доказательство. Надо проверить, что $E\xi$- вариант УМО. Так как константа U-измерима, то достаточно проверить, что $E(E\xi, A) = E(\xi, A)\ \ \forall\ A \in U$.
Имеем $E(\xi, A) = E(\xi I(A)) = E\xi P(A) = E(E\xi, A)$. ■

Утверждение 6.3 (**теорема о монотонной сходимости**). *Если п.н.* $0 \le \xi_n \uparrow \xi$, *то* $E(\xi_n|U) \uparrow E(\xi|U)$ *п.н.*

Доказательство. Из $\xi_{n+1} \ge \xi_n$ п.н. следует в силу пункта c) утврепждения 6.1, что $\hat{\xi}_{n+1} \ge \hat{\xi}_n \ge 0$ п.н. Значит (т.2, §4 , гл.2. [Ширяев, Вероятность]) существует U-измеримая случайная величина $\hat{\xi}$, такая что $\hat{\xi}_n \uparrow \hat{\xi}$ п.н. Почему $\hat{\xi}$ - УМО?

В силу теоремы о монотонной сходимости:

$$\forall A \in U \int\limits_A \hat{\xi}_n P(d\omega) \ \to \int\limits_A \hat{\xi} P(d\omega), \ \int\limits_A \xi_n P(d\omega) \ \to \int\limits_A \xi P(d\omega)$$

Т.к. левые части в двух последних равенствах совпадают (т.к. $\hat{\xi}_n$ - УМО для ξ), то совпадают правые. Т.е. $\int\limits_A \hat{\xi} P(d\omega) = \int\limits_A \xi P(d\omega)$. Значит, $\hat{\xi}$ - УМО для ξ. ■

Утверждение 6.4. *Если* η *- скалярная сл.в. и* U*-измерима,* $E|\xi| < \infty$, $E|\xi\eta| < \infty$, *то*

$$E(\xi\eta|U) = \eta E(\xi|U) \quad \text{п.н.}$$

Доказательство. Докажем в 4 шага.

1) Если $\eta = I(B),\ \ B \in U$, то утверждение верно. Действительно, $\eta E(\xi|U)$ - U-измерима, и $\forall A \in U$

$$\int\limits_A E(I(B)\xi|U)P(d\omega) = \int\limits_A I(B)\xi P(d\omega) = \int\limits_{AB} \xi P(d\omega) =$$

$$= \int\limits_{AB} E(\xi|U)P(d\omega) = \int\limits_A I(B)E(\xi|U)P(d\omega).$$

Значит (см. §6, гл.2 [Ширяев, Вероятность]) $E(I(B)\xi|U) = I(B)E(\xi|U)$ п.н.

2) Значит, утверждение верно для простых U-измеримых функций $\eta = \sum\limits_{i=1}^{k} c_i I(B_i)$, $B_i \in U$ в силу линейности УМО.
3) Пусть $\xi \ge 0,\ \ \eta \ge 0$. Возьмем последовательность простых U-измеримых $0 \le \eta_n \uparrow \eta$ (см теорему 1 в §4 гл.II [Ширяев, Вероятность]).

Тогда в силу шага 2) имеем:

$$E(\eta_n\xi|U) = \eta_n E(\xi|U) \uparrow \eta E(\xi|U) \quad \text{п.н.} \tag{7}$$

Отсюда в силу утверждения 6.3, т.к. $\eta_n\xi \uparrow \eta\xi$, имеем:

$$E(\eta_n\xi|U) \uparrow E(\eta\xi) \quad \text{п.н.} \tag{8}$$

В соотношениях (7),(8) левые части совпадают, значит совпадают и правые части, т.е. $E(\eta\xi|U) = \eta E(\xi|U)$ п.н..

4) Пусть ξ и η произвольные. Тогда:

$$E(\xi\eta|U) = E((\eta^+ - \eta^-)(\xi^+ - \xi^-)|U) = \eta^+ E(\xi^+|U) - \eta^- E(\xi^+|U) -$$
$$-\eta^+ E(\xi^-|U) + \eta^- E(\xi^-|U) = \eta^+ E(\xi|U) - \eta^- E(\xi|U) = \eta E(\xi|U) \ \text{ п.н.}$$

■

Утверждение 6.5 (**формула полной вероятности**). *$EE(\xi|U) = E\xi$.*

Доказательство. $EE(\xi|U) = \int\limits_\Omega E(\xi|U)P(d\omega) = \int\limits_\Omega \xi P(d\omega) = E\xi.$ ■

Утверждение 6.6 (**формула последовательного усреднения**). *Если $U \subset U_1 \subset F$, то $E(\xi|U) = E(E(\xi|U_1)|U)$ п.н.*

Доказательство. Если $A \in U$, то $A \in U_1$. Значит, $\forall\ A \in U$

$$\int\limits_A E(E(\xi|U_1)|U)P(d\omega) = \int\limits_A E(\xi|U_1)P(d\omega) = \int\limits_A \xi P(d\omega) = \int\limits_A E(\xi|U)P(d\omega).$$

Значит $E(E(\xi|U_1)|U) = E(\xi|U)$ п.н. U-измеримость $E(E(\xi|U_1)|U)$ - следствие определения. ■

Утверждение 6.7. *Пусть ξ- скалярная сл. вел., $E\xi^2 < \infty$. Пусть H_U есть множество U-измеримых сл.величин с конечным вторым моментом. Тогда решение задачи $E(\xi(\omega) - a(\omega))^2 \to \min\limits_{a(\omega)\in H_U}$ есть $a^*(\omega) = E(\xi|U)$.*

Доказательство. Имеем:

$$E(\xi - a(\omega))^2 = E(\xi - E(\xi|U))^2 + 2E(\xi - E(\xi|U))(E(\xi|U) - a(\omega)) +$$
$$+ E(E(\xi|U) - a(\omega))^2$$

Для среднего члена имеем:

$$2E(\xi - E(\xi|U))(E(\xi|U) - a(\omega)) = 2EE(\xi - E(\xi|U))(E(\xi|U) - a(\omega)|U) =$$
$$= 2E(E(\xi|U) - a(\omega))E(\xi - E(\xi|U)|U) = 0$$

Т.е. $E(\xi - a(\omega))^2 = E(\xi - E(\xi|U))^2 + E(E(\xi|U) - a(\omega))^2$. Минимум последнего выражения достигается при $a^*(\omega) = E(\xi|U)$. Осталось проверить, что $E(a^*(\omega))^2 = E(E(\xi|U))^2 < \infty$ при $E\xi^2 < \infty$.

Задача 6.1. *Доказать* ***неравенство Коши-Буняковского****:*

$$\textit{при } E\xi^2 < \infty \ \ (E(\xi|U))^2 \le E(\xi^2|U) \ \ \textit{п.н.}$$

Тогда в силу неравенства Коши-Буняковского имеем:

$$E(E(\xi|U))^2 \le EE(\xi^2|U) = E\xi^2 < \infty$$

■

Пример 6.1 (**оптимальный среднеквадратический прогноз в авторегрессии**)**.** *Пусть S_t, $t = 0, 1, \ldots$ - стоимости ценных бумаг в момент времени t. Введем логарифмические приращения $u_t := \ln \dfrac{S_t}{S_{t-1}} = \ln S_t - \ln S_{t-1}$. Для описании динамики последоватльности $\{u_t\}$ используют стохастические разностные уравнения. Например, $AR(1)$ уравнение имеет вид:*

$$u_t = \beta u_{t-1} + \varepsilon_t, \ t = 1, 2, \ldots, \ \beta \in \mathbb{R}^1$$

$\{\varepsilon_t\}$ - н.о.р.с.в., $E\varepsilon_1 = 0$, $0 < D\varepsilon_1 = \sigma^2 < \infty$. Начальное значение u_0 от $\{\varepsilon_t\}$ не зависит, $Eu_0 = 0$, $Eu_0^2 < \infty$. Параметры β и σ^2 обычно неизвестны, распределение сл.в. ε_1 тоже неизвестны.

Из $AR(1)$ уравнения следует:

$$u_t = \varepsilon_t + \beta(\varepsilon_{t-2} + \beta u_{t-2}) = \varepsilon_t + \beta\varepsilon_{t-2} + \beta^2 u_{t-2} =$$
$$= \ldots = \varepsilon_t + \beta\varepsilon_{t-1} + \ldots + \beta^{t-1}\varepsilon_1 + \beta^t u_0$$

Поэтому $Eu_0 = 0$, $Du_t = \sigma^2(1 + \beta^2 + \ldots + \beta^{2(t-1)}) + \beta^{2t}Eu_0^2 < \infty$. Сл.величина ε_{t+1} от $u_t, u_{t-1,\ldots,u_1}$ не зависит.

Пусть $u_1, \ldots, u_n$ - наблюдения, $\mathcal{F}_n = \sigma\{u_1, \ldots, u_n\}$. Оптимальный среднеквадратический прогноз ненаблюдаемой величины u_{n+1} по наблюдениям - есть *решение* u^*_{n+1} *задачи*

$$E(u_{n+1}-u^*_{n+1})^2 \to \min,\ u^*_{n+1}\text{ - }\mathcal{F}_n - \textit{измерима},\ E(u^*_{n+1})^2<\infty$$

Тогда $u^*_{n+1}=\varphi(u_1,\dots,u_n)=E(u_{n+1}|\mathcal{F}_n)$ *в силу утверждения 6.7. Имеем:*

$$E(u_{n+1}|F_n)=E(\beta u_n+\varepsilon_{t+1}|F_n)=E(\beta u_n|F_n)+E(\varepsilon_{n+1}|F_n)=\beta u_n+E\varepsilon_{n+1}=\beta u_n$$

Пусть L^2 *будет множество сл.в. с конечным вторым моментом. Отождествим сл.в., равные п.н. Получим множество классов эквивалентных сл.в. Для класса* $\tilde{\xi}$ *норма* $|\tilde{\xi}|:=\left(E\xi^2\right)^{\frac{1}{2}}$, ξ *- любой элемент* $\tilde{\xi}$.

Множество классов эквивалентных сл.в. с такой нормой есть банаховское линейное пространство $\mathbb{L}^2$. *Для* $\xi,\eta\in\mathbb{L}^2$ *можно ввести скалярное произведение* $(\xi,\eta):=E\xi\eta$. *Если* $(\xi,\eta)=0$, *то* $\xi\perp\eta$.

Если H_U *есть линейное подпространство сл.в. в* $\mathbb{L}^2$, *которые* U*-измеримы, то решение задачи* $E\left(\xi-a(\omega)\right)^2\to\min\limits_{a(\omega)\in H_U}$ *есть* $a^*=proj_{H_U}\xi$. *В силу утверждения 6.7 имеем:*

$$a^*(\omega)=proj_{H_U}\xi=E(\xi|U)$$

Это соотношение может служить определением УМО $E(\xi|U)$ *в случае* $E\xi^2<\infty$. *Но у нас только* $E|\xi|<\infty$!

6.3 УМО и условные распределения относительно сл.в.

Пусть на $(\Omega,\mathcal{F},P)$ заданы сл. векторы $\xi\in\mathbb{R}^s,\ \eta\in\mathbb{R}^k$. Пусть $U_\eta=\sigma\{\omega:\eta(\omega)\in B, B\in\mathcal{B}(\mathbb{R}^k)\}$.

Определение 6.4. $E(\xi|\eta):=E(\xi|U_\eta)$ - ***УМО*** ξ ***относительно*** η. *Т.к.* $E(\xi|\eta)$ *-* U_η*-измерима, то* $E(\xi|\eta)=m(\eta)$ *для некоторой борелевской функции* $m(y)$, $y\in\mathbb{R}^k$, *обозначается* $E(\xi|\eta=y)$ *или короче* $E(\xi|y)$.

Ясно, что $m(\eta)$ такая функция, что (она автоматически U_η-измерима):

$$\int\limits_W m(\eta)P(d\omega)=\int\limits_W \xi P(d\omega)\ \forall B\in\mathcal{B}(\mathbb{R}^k),\ W=\{\omega:\eta(\omega)\in B\} \tag{9}$$

Делая замену $\eta(\omega)=y$ и заменяя $m(\eta)$ на $E(\xi|y)$, получим для (9) эквивалентное выражение:

$$\int\limits_B E(\xi|y)P_\eta(dy) = \int\limits_W \xi P(d\omega)\ \forall B \in \mathcal{B}(\mathbb{R}^k),\ W = \{\omega : \eta(\omega) \in B\} \tag{10}$$

Итак, $E(\xi|y)$ - борелевская функция, удовлетворяющая (10). Она определена на множестве P_η-вероятности ноль.

Определение 6.5. *Поскольку $P(C) = EI(C)$, $C \in \mathcal{F}$, то естественно определение $P(C|\eta) := E\left(I(C)|U_\eta\right)$, $C \in \mathcal{F}$. Тогда $P(C|\eta) = g_C(\eta)$. Функцию $g_C(y)$, $y \in \mathbb{R}^k$ обозначим $P(C|\eta = y)$ или короче $P(C|y)$.*

Ясно, что $g_C(\eta)$ такая функция, что (это переписанное соотношение (9)):

$$\int\limits_W g_C(\eta)P(d\omega) = \int\limits_W I(C)P(d\omega)\ \forall B \in \mathcal{B}(\mathbb{R}^k),\ W = \{\omega : \eta(\omega) \in B\} \tag{11}$$

Делая в (11) замену $\eta(\omega) = y$, получим, что $P(C|y)$ такая функция, что:

$$\int\limits_B P(C|y))P_\eta(dy) = P(C, \eta \in B)\ \forall B \in \mathcal{B}(\mathbb{R}^k) \tag{12}$$

Функция $P(C|y)$ определена с точностью до значений на множестве P_η-вероятности ноль.

Определение 6.6. *Функция $P(A|y)$ множества $A \in \mathcal{B}(\mathbb{R}^s)$ и $y \in \mathbb{R}^k$ называется* ***условным распределением ξ при условии $\eta = y$****, если выполнены два условия:*

1) *При каждом фиксированном A $P(A|y)$ есть условная веростность события $C = (\omega : \xi(\omega \in A))$ при условии $\eta = y$, т.е. $P(A|y) = P(\xi(\omega) \in A|\eta = y)$.*
2) *Для любого $y \in \mathbb{R}^k$, за исключением, быть может, множества y-ов P_η-веростности ноль, $P(A|y)$ есть распределение веростностей по A, т.е. выполняется счетная аддитивность по A.*

Условное распределение существует (см. [Ширяев, Вероятность, §7, гл. II]). Условное распределение обозначается также $P(\xi \in A|y)$, $P(\xi \in A|\eta = y)$.

Определение 6.7. *Пусть скалярная функция $f(x|y) \geq 0$ и измерима по паре (x, y) (т.е. борелевская). Если для P_η-п.в. $y \in \mathbb{R}^k$ условное распределение*

$$P(\xi \in A|\eta = y) = \int\limits_{x \in A} f(x|y)\mu(dx)\ \forall A \in \mathcal{B}(\mathbb{R}^s)$$

то $f(x|y)$ *называется* ***условной плотностью*** ξ ***при условии*** $\eta = y$ *относительно меры* μ.

Замечание. *Если* $f(x|y)$ *- неотрицательная борелевская функция, удовлетворяющая условию:*

$$\forall A \in \mathcal{B}(\mathbb{R}^s),\ B \in \mathcal{B}(\mathbb{R}^k) \int\limits_{y \in B} \left(\int\limits_{x \in A} f(x|y)\mu(dx) \right) P_\eta(dy) = P(\xi \in A, \eta \in B) \qquad (13)$$

то $f(x|y)$ *- условная плотность вероятность. Действительно, если* (13) *выполнено, то в силу* (12) $\int\limits_{x \in A} f(x|y)\mu(dx)$ *есть условная вероятность* $P(\xi \in A|\eta = y)$. *Но этот интеграл счетно-аддитивен по* A, *т.е.* $\int\limits_{x \in A} f(x|y)\mu(dx)$ *есть условное распределение* ξ *при условии* $\eta = y$, *но тогда* $f(x|y)$ *- условная плотность!*

Замечание. *Пусть на* $\mathbb{R}^s$ *и* $\mathbb{R}^k$ *заданы меры* μ *и* λ. *Произведение этих мер называется такая мера* $\mu \times \lambda$ *на* $\mathbb{R}^s \times \mathbb{R}^k$, *что*

$$\forall A \in \mathcal{B}(\mathbb{R}^s),\ B \in \mathcal{B}(\mathbb{R}^k)\ \mu \times \lambda(A \times B) = \mu(A) \times \lambda(B)$$

Если μ *и* λ *- меры Лебега, то* $\mu \times \lambda$ *- мера Лебега на* $\mathbb{R}^s \times \mathbb{R}^k = \mathbb{R}^{s+k}$.
Если μ *и* λ *- считающие меры, то* $\mu \times \lambda$ *- «считающая мера на* $\mathbb{R}^{s+k}$*».*

Теорема 6.3. *Если совместное распределение* ξ *и* η *в* $\mathbb{R}^s \times \mathbb{R}^k$ *имеет плотность* $f(x, y)$ *относительно меры* $\mu \times \lambda$, *то функция*

$$f(x|y) = \begin{cases} \dfrac{f(x,y)}{q(y)}, \ q(y) \neq 0 \\ 0, \ q(y) = 0 \end{cases} \qquad (14)$$

есть условная плотность вероятности ξ *при условии* $\eta = y$. *Здесь*

$$q(y) = \int\limits_{\mathbb{R}^s} f(x, y)\mu(dx) \qquad (15)$$

есть плотность η относительно меры λ. Кроме того:

$$E(\xi|\eta = y) = \int\limits_{\mathbb{R}^s} x f(x|y)\mu(dx) \text{ для } P_\eta\text{-п.в. } y \tag{16}$$

Доказательство.

- Докажем (15).

$$P(\eta \in A) = P(\eta \in A, \xi \in \mathbb{R}^s) = \iint\limits_{y\in A, x\in\mathbb{R}^s} f(x,y)\mu(dx)\lambda(dy) =$$

$$= \Big|\text{по теореме Фубини}\Big| = \int\limits_{y\in A}\left(\int\limits_{x\in\mathbb{R}^s} f(x,y)\mu(dx)\right)\lambda(dy)$$

Отсюда $q(y) = \int\limits_{\mathbb{R}^s} f(x,y)\mu(dx)$ есть плотность η относительно λ.

- Докажем (14). Т.к. $f(x|y) \geq 0$, $f(x|y)$ - борелевская и η имеет плотность $q(y)$, достаточно проверить (13). Для $f(x|y)$ из (14) имеем:

$$\int\limits_{y\in B, q(y)\neq 0}\left(\int\limits_{x\in A}\frac{f(x,y)}{q(y)}\mu(dx)\right)q(y)\lambda(dy) = \iint\limits_{x\in A, y\in B} f(x,y)\mu(dx)\lambda(dy) =$$
$$= P(\xi \in A, \eta \in B)$$

Значит, (13) выполянется, и $f(x|y)$ из (14) есть условная плотность вер-ти.

- Докажем (16). Достаточно проверить (10) с $E(\xi|\eta = y)$ из (16). Имеем:

$$\int\limits_{y\in B}\left(\int\limits_{\mathbb{R}^s} x f(x|y)\mu(dx)\right)q(y)\lambda(dy) = \int\limits_{y\in B, q(y)\neq 0}\left(\int\limits_{\mathbb{R}^s} x\frac{f(x,y)}{q(y)}\mu(dx)\right)q(y)\lambda(dy) =$$

$$= \iint\limits_{x\in\mathbb{R}^s, y\in B} x f(x,y)\mu(dx)\lambda(dy) = E\left(\xi I(y \in B)\right) = \int\limits_{(\omega:\eta(\omega)\in B)} \xi P(d\omega)\ \forall B \in \mathcal{B}(\mathbb{R}^k)$$

Т.е. (10) с $E(\xi|\eta = y)$ из (16) верно и (16) доказано.

∎

Пример 6.2. *Пусть* ξ *и* η *дискретные векторы со значениями* $X = (x_1, x_2, \dots)$ *и* $Y = (y_1, y_2, \dots)$. *Тогда* $f(x,y) = P(\xi = x, \eta = y)$, $q(y) = P(\eta = y)$. *Для* $y \in Y$ $f(x|y) = \dfrac{P(\xi = x, \eta = y)}{P(\eta = y)} = P(\xi = x | \eta = y)$. *При* $y \notin Y$ $f(x|y) = 0$. *Условное распределение:*

$$P(\xi \in A | \eta = y) = \int\limits_A f(x|y)\mu(dx) = \sum_{x_i \in A} P(\xi = x_i | \eta = y),\ y \in Y$$

При $y \notin Y$ $P(\xi \in A|\eta = y) = 0$. *Наконец:*

$$E(\xi|y) = \int\limits_{\mathbb{R}^s} x f(x|y)\mu(dx) = \sum_i x_i P(\xi = x_i|\eta = y),\ y \in Y$$

При $y \notin Y$ $E(\xi|y) = 0$.

Пример 6.3. *Говорят, что вектор* (ξ, η) *имеет двумерный невырожденный гауссовский закон, если:*

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho_{\xi\eta}^2}\sigma_\xi\sigma_\eta} exp\Big\{ -\frac{1}{2(1-\rho_{\xi\eta}^2)}\Big[\frac{(x-m_\xi)^2}{\sigma_\xi^2} + \frac{(y-m_\eta)^2}{\sigma_\eta^2} - \\ -2\rho_{\xi\eta}\frac{(x-m_\xi)(y-m_\eta)}{\sigma_\xi\sigma_\eta}\Big]\Big\}$$

Здесь $m_\xi, m_\eta \in \mathbb{R}$; $\sigma_\xi^2, \sigma_\eta^2 > 0$; $|\rho_{\xi\eta}| < 1$.
Прямым вычислением показывается:

$$f_\eta(y) \sim N(m_\eta, \sigma_\eta^2),\ f_\xi(x) \sim N(m_\xi, \sigma_\xi^2),\ \rho_{\xi\eta} = corr(\xi,\eta) = \frac{cov(\xi,\eta)}{\sigma_\xi\sigma_\eta}$$

$$f(x|y) \sim N\left(E(\xi|\eta = y), D(\xi|\eta = y)\right),$$

$$\text{где } E(\xi|\eta = y) = m_\xi + \rho_{\xi\eta}(y - m_\eta),\ D(\xi|\eta = y) = (1-\rho_{\xi\eta}^2)\sigma_\xi^2$$

При $\rho_{\xi\eta} = 0$ ξ *и* η *независимы!*

7 Достаточные статистики и оптимальные оценки

7.1 Определение достаточной статистики

Пример 7.1. *Пусть* $X = (X_1, \dots, X_n)$ *- н.о.р.,* $X_1 \sim Pois(\theta)$, $\theta > 0$. *Т.е.:*

$$f(y, \theta) = P_\theta(X_1 = y) = \frac{\theta^y}{y!}e^{-\theta}, \ \ y \in \mathcal{X} = \{0, 1, 2, \dots\}$$

Пусть $T(X) = X_1 + \dots + X_n$. *Тогда* $T(X) \sim Pois(n\theta)$. *Обозначим реализацию* X *через* $x = (x_1, \dots, x_n)$. *Тогда условное распределение:*

$$P_\theta\left(X \in A | T(X) = t\right) = \sum_{x \in A} f_\theta(x|t), \ \text{где } f_\theta(x|t) = \begin{cases} P_\theta\left(X = x | T(X) = t\right), \ t \in \mathcal{X} \\ 0, \ t \notin \mathcal{X} \end{cases}$$

Покажем, что *условное распределение не зависит от* θ. *Достаточно проверить, что* $f_\theta(x|t)$ *не зависит от* θ. *Дальше* $t \in \mathcal{X}$.

a) *Если* $T(x) = \sum\limits_{i=1}^{n} x_i \neq t$, *то* $f_\theta(x|t) = \dfrac{P\left(X = x, T(X) = t\right)}{P(T(X) = t)} = \emptyset$

b) *Если* $T(x) = t$, *то:*

$$f_\theta(x|t) = \frac{P\left(X = x, T(X) = T(x)\right)}{P_\theta(T(X) = t)} = \frac{P_\theta(X = x)}{P_\theta(T(X) = t)} = \frac{\prod\limits_{i=1}^{n} P_\theta(X_i = x_i)}{P_\theta(T(X) = t)} =$$

$$= \frac{\theta^{\sum\limits_{i=1}^{n} x_i} e^{-n\theta} t!}{x_1! \dots x_n! (n\theta)^t e^{-n\theta}} = \frac{t!}{x_1! \dots x_n!} \cdot \frac{1}{n^t} \ \text{- не зависит от } \theta$$

Покажем, что $I^X(\theta) = I^T(\theta)$.

$$i(\theta) = E_\theta\left(\frac{\partial}{\partial\theta}\left(\underbrace{\ln\frac{\theta^{X_1}}{X_1!}e^{-\theta}}_{=f(X_1,\theta)}\right)\right)^2 = E_\theta\left(\frac{X_1}{\theta}-1\right)^2 = \frac{D_\theta X_1}{\theta^2} = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

Значит $I^X(\theta) = ni(\theta) = \dfrac{n}{\theta}$.

$$I^T(\theta) = E_\theta\left(\frac{\partial}{\partial\theta}\left(\ln\frac{(n\theta)^T}{T!}e^{-n\theta}\right)\right)^2 = E_\theta\left(\frac{T}{\theta}-n\right)^2 = \frac{D_\theta T}{\theta^2} = \frac{n\theta}{\theta^2} = \frac{n}{\theta}$$

Т.е. получаем, что $I^X(\theta) = I^T(\theta)$. *Т.е.* $T(X) = X_1 + \dots + X_n$ *содержит всю информацию Фишера от* θ*, что и* X!.

Пусть $X = (X_1, \dots, X_n)$ - наблюдение, $P_X \in \left\{P_\theta \colon \theta \in \Theta \subseteq \mathbb{R}^k\right\}$. Скалярные или векторные функции $T(X)$ называется статистиками.

Определение 7.1. *Статистика* $T(X)$ *называется* ***достаточной*** *для параметра* θ*, если существует вариант условного распределения* $P_\theta\left(X \in A | T(X) = t\right)$*, не зависящий от* θ *при любом* t.

Замечание. *Сделаем несколько комментариев:*

1) *Условное распределение* $P_\theta\left(X \in A | T(X) = t\right)$ *можно интерпретировать как распределение* X *на поверхности* $T(x) = t$*. Тогда, если* $T(X)$ *- достаточная, то знание, где выборочная точка* X *находится на поверзности, не дает никакой дополнительной информации о параметре* θ*. Т.е. вся информация о параметре* θ *содержится в* $T(X)$*. Для построения оценки* θ *достаточно знать* $T(X)$*, остальные данные, содержащиеся в* X*, бесполезны.*
2) *Если* $T(X)$ *- некоторая (необязательно достаточная) статистика, то при некоторых условиях регулярности (см. [А.А.Боровков. Матем. статист., §17, теорема 1])* $I^T(\theta) \le I^X(\theta)\ \forall\theta \in \Theta$*, где* $I^T(\theta)$ *и* $I^X(\theta)$ *-* $(k \times k)$*-матрицы. Равенство здесь достигается тогда и только тогда, когда* T *- достаточная. Это точный смысл слов: «достаточная статистика содержит всю информацию о параметре* θ*», применительно к информации Фишера.*

7.2 Критерий факторизации Неймана-Фишера

Теорема 7.1 (**критерий факторизации Неймана-Фишера**). *Пусть X имеет плотность $p(x,\theta)$ относительно меры μ. Тогда $T(X)$ будет достаточной статистикой для θ тогда и только тогда, когда*

$$p(x,\theta) = \psi(T(x),\theta)\hbar(x) \text{ для } \mu\text{-п.в. } x \tag{1}$$

Здесь $\psi \geq 0$ и $\hbar \geq 0$ зависят только от своих аргументов, $\psi(s,\theta)$ измерима по s, $\hbar(x)$ измерима по x.

Следствие 7.1. *Если T достаточная, и $u = \varphi(v)$ взаимоодназначна и измерима в обе стороны, то $T_1 = \varphi(T)$ тоже достаточна.*

Доказательство. $\psi(T,\theta) = \psi(\varphi^{-1}(T_1),\theta)$. ■

Значит, представление (1) неоднозначно!

Докажем теорему 7.1.

Доказательство. Докажем для дискретного X. Пусть X имеет носитель $\mathcal{X} = (x_1, x_2, \dots)$, не зависящий от θ. Тогда $\mathcal{X}_T = (T(x_1), T(x_2), \dots)$. Условное распределение:

$$P_\theta\left(X \in A | T(X) = t\right) = \begin{cases} \sum\limits_{x_i \in A} P_\theta\left(X = x_i | T(X) = t\right), & t \in \mathcal{X}_t \\ 0, & t \notin \mathcal{X}_t \end{cases}$$

Значит, условное распределение не зависит от θ тогда и только тогда, когда условная вероятность $P_\theta(X = x|T(X) = t)$ не зависит от θ для всех $t \in \mathcal{X}_t$, $x \in \mathcal{X}$. Но при $x \in \mathcal{X}$, $t \in \mathcal{X}_t$:

$$P_\theta\left(X = x | T(X) = t\right) = \frac{P_\theta(X = x, T(X) = t)}{P_\theta(T(X) = t)} =$$

$$= \begin{cases} \dfrac{P_\theta \overbrace{(X = x, T(X) = T(x))}^{(X=x)}}{P_\theta(T(X) = T(x))}, & T(x) = t \\ 0, & T(x) \neq t \end{cases} = \begin{cases} \dfrac{P_\theta(X = x)}{P_\theta(T(X) = T(x))}, & T(x) = t \\ 0, & T(x) \neq t \end{cases}$$

Итак, $T(X)$ - достаточная тогда и только тогда, когда

$$\frac{P_\theta(X = x)}{P_\theta(T(X) = T(x))} \text{ не зависит от } \theta \ \forall x \in \mathcal{X} \tag{2}$$

1) Пусть выполнено (1). Тогда, учитывая $P_\theta(X=x)=p(x,\theta)$, имеем:

$$\frac{P_\theta(X=x)}{P_\theta(T(X)=T(x))}=\frac{\psi(T(x),\theta)\hbar(x)}{\sum\limits_{y:T(y)=T(x)}p(y,\theta)}=\frac{\psi(T(x),\theta)\hbar(x)}{\sum\limits_{y:T(y)=T(x)}\psi(T(y),\theta)\hbar(y)}=\frac{\hbar(x)}{\sum\limits_{y:T(y)=T(x)}\hbar(y)}$$

Это выражение от θ не зависит, т.е. T - достаточная (выполнено (2)).

2) Наоборот, пусть выполнено (2). Обозначая дробь в (2) через $\hbar(x)$, получим:

$$P_\theta(X=x)=\hbar(x)P_\theta(T(X)=T(x)),\ \text{т.е.}$$
$$p(x,\theta)=\hbar(x)\sum_{y:T(y)=T(x)}p(y,\theta)=\hbar(x)\psi(T(x),\theta)$$

■

Примеры

1) $T(X)=X$ всегда достаточная статистика, ее называют тривиальной. Здесь $\psi(T(x),\theta)=p(x,\theta)$, $\hbar(x)=1$.
2) $X=(X_1,\dots,X_n)$, $\{X_i\}$ - н.о.р., $X_1\sim Pois(\theta)$, $\theta>0$.

$$p(x,\theta)=\prod_{i=1}^n P_\theta(X_i=x_i)=\frac{\theta^{\sum x_i}}{x_1!\dots x_n!}e^{-n\theta},\ x=(x_1,\dots,x_n)$$

$T(X)=\sum\limits_{i=1}^n X_i$ - достаточная, $\psi(T(x),\theta)=\theta^{\sum x_i}e^{-n\theta}$, $\hbar(x)=\dfrac{1}{x_1!\dots x_n!}$.

3) $X=(X_1,\dots,X_n)$, $\{X_i\}$ - н.о.р., $X_1\sim R(0,\theta)$, $\theta>0$.

$$p(x,\theta)=\begin{cases}\frac{1}{\theta^n},\ x_{(1)}\geq 1,\ x_{(n)}\leq\theta\\ 0,\ \text{в противном случае}\end{cases}=\frac{1}{\theta^n}I(x_{(n)}\leq\theta)\cdot I(x_{(1)}\geq 0)$$

Здесь $\psi(T(x),\theta)=\frac{1}{\theta^n}I(x_{(n)}\leq\theta)$ с $T(x)=x_{(n)}$, $\hbar(x)=I(x_{(1)}\geq 0)$, $x=(x_1,\dots,x_n)$, $x_{(1)}\leq x_{(2)}\leq\dots\leq x_{(n)}$.

4) $X=(X_1,\dots,X_n)$, $\{X_i\}$ - н.о.р., $X_1\sim N(\theta_1,\theta_2)$ с $\theta_1\in\mathbb{R}^1$ и $\theta_2>0$. Здесь $\theta=(\theta_1,\theta_2)^T$.

Пусть $T(X)=\left(\sum\limits_{i=1}^n X_i,\sum\limits_{i=1}^n X_i^2\right)^2$, тогда $T(X)$ - достаточная статистика для

θ. Действительно:

$$p(x,\theta) = \left(\frac{1}{\sqrt{2\pi\theta_2}}\right)^n e^{-\frac{1}{2\theta_2}\sum\limits_{i=1}^{n}(x_i-\theta_1)^2} =$$

$$= \left(\frac{1}{\sqrt{2\pi\theta_2}}\right)^n exp\left\{-\frac{1}{2\theta_2}\left[\sum x_i^2 - 2\theta_1\sum x_i + n\theta_1^2\right]\right\} = \psi(T(x),\theta)\hbar(x),\ \hbar(x) = 1$$

Пусть $T_1 = \left(\overline{X}, S^2\right)$, где $\overline{X} = n^{-1}\sum\limits_i X_i$, $S^2 = \dfrac{1}{n-1}\sum(X_i - \overline{X})^2$. Поскольку $S^2 = \dfrac{n}{n-1}\left(n^{-1}\sum X_i^2 - \overline{X}^2\right)$, то отображение $T(x) \leftrightarrow T_1(x)$ взаимнооднозначно и измеримо в обе стороны. Значит, $T_1(x) = \left(\overline{X}, S^2\right)^T$ - достаточная статистика.

7.3 Теорема Блекуэла-Рао-Колмогорова

Теорема 7.2 (Блекуэл-Рао-Колмогоров). *Пусть $T = T(X)$ - достаточная статистика, и $\hat{\tau}_n$ - несмещенная оценка $\tau(\theta) \in \mathbb{R}^m$ с конечной ковариационной матрицей. Тогда функция $\tau_n^* := E_\theta(\hat{\tau}_n|T)$ обладает свойствами:*

1) *τ_n^* - несмещенная оценка $\tau(\theta)$ с конечной ковариационной матрицей.*
2) *τ_n^* зависит от X лишь через $T(X)$.*
3) *$D_\theta\tau_n^* \le D_\theta\hat{\tau}_n$ при всех $\theta \in \Theta$. Равенство возможно тогда и только тогда, когда $\hat{\tau}_n = \tau_n^*$ P_θ-п.н. при всех $\theta \in \Theta$.*

Доказательство. Докажем 1) и 2). Если T - достаточная статистика, то борелевская функция $m(t) = E_\theta(\hat{\tau}_n|T(x) = t$ от θ не зависит, т.к. условное распределение X при условии $T(x) = t$ от θ не зависит. Значит, и $m(T(X)) = E_\theta(\hat{\tau}_n|T(X)) = \tau_n^*(X)$ от θ не зависит (т.е. τ_n^* - оценка), а от X зависит лишь через $T(X)$.
Далее, $E_\theta\tau_n^* = E_\theta E_\theta(\hat{\tau}_n|T) = E_\theta\hat{\tau}_n = \tau(\theta)$ при всех θ в силу формулы полной вероятности и несмещенности $\hat{\tau}_n(\theta)$.
Наконец, если $E_\theta\hat{\tau}_{in}^2 < \infty$, то (см. доказательство утверждения 6.7):

$$E_\theta\left(\tau_{in}^*\right)^2 = E_\theta\left(E_\theta\left(\hat{\tau}_{in}|T\right)\right)^2 \le E_\theta E_\theta\left(\hat{\tau}_{in}^2|T\right) = E_\theta\hat{\tau}_{in}^2 < \infty$$

(использовали неравенство Коши-Буняковского)

Докажем 3). Достаточно проверить, что:

$$\alpha^T D_\theta \tau_n^* \alpha \le \alpha^T D_\theta \hat{\tau}_n \alpha \ \forall \alpha \in \mathbb{R}^m,\ \forall \theta \in \Theta$$

Но также мы имеем:

$$\alpha^T D_\theta^* \alpha = \alpha^T E(\tau^* - \tau)(\tau^* - \tau)^T \alpha = E_\theta \left(\alpha^T \tau_n^* - \alpha^T \tau\right)^2 = D_\theta(\alpha^T \tau_n^*)$$

Значит, достаточно проверить, что:

$$D_\theta(\alpha^T \tau_n^*) \le D_\theta(\alpha^T \hat{\tau}_n) \ \forall \alpha \in \mathbb{R}^m,\ \forall \theta \in \Theta$$

Но также имеем:

$$D(\alpha^t \hat{\tau}_n) = E_\theta \left(\alpha^T \hat{\tau}_n - \alpha^T \tau - \alpha^t \tau_n^* + \alpha^T \tau_n^*\right) =$$
$$= E_\theta \left(\alpha^T \hat{\tau}_n - \alpha^T \tau_n^*\right)^2 + E_\theta \left(\alpha^T \tau_n^* - \alpha^T \tau\right)^2 + 2E_\theta \left(\alpha^T \hat{\tau}_n - \alpha^T \tau_n^*\right)\left(\alpha^T \tau_n^* - \alpha^T \tau\right)$$

Заметим, что:

$$2E_\theta E_\theta \left(\left(\alpha^T \hat{\tau}_n - \alpha^T \tau_n^*\right)\left(\alpha^T \tau_n^* - \alpha^T \tau\right)|T\right) =$$
$$= 2E_\theta \left(\left(\alpha^T \hat{\tau}_n - \alpha^T \tau_n^*\right) \underbrace{E_\theta \left(\alpha^T \tau_n^* - \alpha^T \tau | T\right)}_{=0}\right) = 0$$

Значит:

$$D(\alpha^t \hat{\tau}_n) = D_\theta(\alpha^T \tau_n^*) + E_\theta \left(\alpha^T \hat{\tau}_n - \alpha^T \tau_n^*\right)^2$$

Следовательно, $D(\alpha^t \hat{\tau}_n) \ge D_\theta(\alpha^T \tau_n^*)$ и равенство возможно, если $\alpha^T \hat{\tau}_n = \alpha^T \tau_n^*$ P_θ-п.н. $\forall \theta \in \Theta$. Это равносильно $\hat{\tau}_n = \tau_n^*$ P_θ-п.н. $\forall \theta \in \Theta$. ■

Следствие 7.2. *Пусть T - достаточная статистика, и существует оптимальная оценка $\hat{\tau}_n(X)$ для функции $\tau(\theta) \in \mathbb{R}^m$. Тогда $\hat{\tau}_n(X) = \varphi(T)$ P_θ-п.н. для некоторой борелевской функции φ.*

Доказательство. Если $\hat{\tau}_n$ - оптимальная, то и $\tau_n^* = E_\theta(\hat{\tau}_n|T)$ тоже. Тогда $D_\theta \tau_n^* = D_\theta \hat{\tau}_n$ $\forall \theta \in \Theta$. Значит, в силу пункта 3) теоремы 7.2 получаем, что $\tau_n^* = \hat{\tau}_n$ P_θ-п.н., но $\tau_n^* = \varphi(T)$. ■

Итак, если T - достаточная статистика, то оптимальную оценку $\hat{\tau}_n = \varphi(T)$ функции $\tau(\theta) \in \mathbb{R}^m$ можно искать как решения **уравнения несмещенности**:

$$E_\theta \varphi(T) = \tau(\theta) \ \forall \theta \in \Theta \tag{3}$$

Определение 7.2. *Статистика $T(X)$ называется* ***полной****, если из равенства $E_\theta \varphi(T(X)) = 0\ \forall\theta \in \Theta \subseteq \mathbb{R}^k$ следует, что $\varphi(T(X)) = 0$ п.н. по P_θ-вероятности $\forall\theta \in \Theta$.*

Пусть $T(X)$ имеет плотность $q(x,\theta)$ по мере μ. Тогда $\varphi(T(X)) = 0$ P_θ-п.н. тогда и только тогда, когда

$$\forall\theta \in \Theta\ 0 = P_\theta\left(\varphi(T(X)) \neq 0\right) = E_\theta I\left(\varphi(T(X)) \neq 0\right) = \int\limits_{N_q^\theta} I(\varphi(s) \neq 0) q(s,\theta)\mu(ds)$$

Отсюда следует, что $I(\varphi(s) \neq 0) = 0$ μ-п.в. на N_q^θ, т.е.:

$$\varphi(s) = 0 \ \mu\text{-п.в. на } N_q^\theta \tag{4}$$

1) Пусть $T(X)$ дискретна с множеством значений $\mathcal{X}_t^\theta$. Тогда (4) (т.е. полнота) эквивалентна условию $\varphi(s) = 0\ \forall s \in \mathcal{X}_t^\theta$. Т.е. $\varphi(s)$ равна нулю на множетсве значений T.
2) Пусть $T(X)$ абсолютно непрерывна по мере Лебега. Тогда (4) эквивалентна условию $\varphi(s) = 0$ п.в. по мере Лебега на N_q^θ $\forall\theta \in \Theta$.

7.4 Оптимальные оценки при полной статистике и лемма Лемана-Шеффе

Лемма 7.1 (**об оптимальных оценках при наличии полной достаточной статистики**). *Пусть T - полная достаточная для θ статистика. Тогда:*

1) *Если уравнение несмещенности (3) имеет решение, то оно P_θ-п.н. единственно. Если это решение имеет конечную ковариационную матрицу, то это опатимальная оценка $\tau(\theta)$.*
2) *Если уравнение несмещенности (3) не имеет решений, то нет оптимальных оценков для $\tau(\theta)$. Более того, нет даже несмещенных оценок $\tau(\theta)$.*
3) *Если $\hat{\tau}_n$ - несмещенная оценка $\tau(\theta)$ с конечной ковариационной матрицей, то $\tau_n^* := E_\theta(\hat{\tau}_n | T)$ есть оптимальная оценка для $\tau(\theta)$.*

Доказательство. Докажем утверждения по пунктам.

1) Пусть $\varphi(T)$ и $\varphi_1(T)$ - два решения уравнения (3). Тогда $E_\theta\left[\varphi(T) - \varphi_1(T)\right] = 0\ \forall\theta \in \Theta$. В силу полноты статистики T $\varphi(T) = \varphi_1(T)$ P_θ-п.н. $\forall\theta \in \Theta$. Т.е. решение (3) P_θ-п.н. единственно, но если решение одно (т.е. есть одна

несмещенная оценка вида $\varphi(T)$) и имеет конечную ковариацию, то это и есть оптимальная оценка.

2) Если бы существовала несмещенная оценка $\hat{\tau}_n(X)$, то $\varphi(T) := E_\theta(\hat{\tau}_n|T)$ тоже была бы несмещенной оценкой $\tau(\theta)$, т.к. $E_\theta\varphi(T) = E_\theta E_\theta(\hat{\tau}_n|T) = E_\theta\hat{\tau}_n = \tau(\theta)\ \forall\theta\in\Theta$. Но тогда $\varphi(T)$ - решение (3) - получаем противоречие.
3) $E_\theta\tau_n^* = E_\theta E_\theta(\hat{\tau}_n|T) = E_\theta\hat{\tau}_n = \tau(\theta)$. Т.е. $\tau_n^* = \tau_n^*(T)$ удовлетворяет (3) и имеет конечную ковариационную матрицу. Осталось применить пункт 1).

■

Лемма 7.2 (Лемана-Шеффе). *Пусть T - полная достаточная статистика, а борелевская функция $g(T)$ имеет конечную ковариационную матрицу. Тогда $g(T)$ есть оценка своего математического ожидания $\tau(\theta) = E_\theta g(T)$.*

Утверждение леммы 7.2 следует из пункта 1) леммы 7.1, т.к. $g(T)$ удовлетворяет уравнению несмещенности (3).

Пример 7.2. *Пусть $X = (X_1, \dots, X_n)$, $\{X_i\}$ - н.о.р., $X_1 \sim Pois(\theta)$, $\theta > 0$. Указать полную достаточную статистику и найти оптимальные оценки функций $\tau_1(\theta) = \theta^2$, $\tau_2(\theta) = \theta$.*

Решение. Мы знаем, что $T(X) = \sum\limits_{i=1}^{n} X_i$ - достаточная статистика, $T(X) \sim Pois(n\theta)$.

- Проверим полноту $T(X)$. Если $E_\theta\varphi(T) = 0\ \forall\theta > 0$, то ёто эквивалентно условию $\sum\limits_{k\geq 0}\varphi(k)\dfrac{(n\theta)^k}{k!}e^{-n\theta} = 0\ \forall\theta > 0$. Т.е. $\sum\limits_{k\geq 0}\dfrac{\varphi(k)n^k}{k!}\theta^k = 0\ \forall\theta > 0$. Но если степенной ряд на невырожденном множетсве равен тождественно нулю, то его коэффициенты равны нулю, т.е. $\dfrac{\varphi(k)n^k}{k!} = 0$, $\varphi(k) = 0$. Таким образом, полнота доказана.
- Уравнение несмещенности для $\tau_1(\theta)$:

$$E_\theta\varphi(T) = \tau_1(\theta),\ \sum_{k\geq 0}\varphi(k)\frac{(n\theta)^k}{k!}e^{-n\theta} = \theta^2$$

$$\sum_{k\geq 0}\frac{\varphi(k)n^k}{k!}\theta^k = \theta^2 e^{n\theta} = \sum_{s\geq 0}\frac{n^s}{k!}\theta^{s+2} = \sum_{l\geq 2}\frac{n^{l-2}}{(l-2)!}\theta^l\ \forall\theta > 0$$

Значит, $\varphi(0) = \varphi(1) = 0$. Для $k \geq 2$ $\dfrac{\varphi(k)n^k}{k!} = \dfrac{n^{k-2}}{(k-2)!}$, т.е. $\varphi(k) = \dfrac{k(k-1)}{n^2}$,

$\varphi(T) = \dfrac{T(T-1)}{n^2}$. Очевидно, что $E_\theta\varphi^2(T) < \infty$, т.е. $\varphi(T)$ - оптимальная оценка $\varphi_1(\theta)$.

- $\hat{\tau}_n(X) = X_1$ - несмещенная оценка для $\tau_2(\theta) = \theta$. Оптимальная оценка для $\tau_2(\theta) = \theta$ есть $\varphi(T) = E_\theta(X_1|\sum X_i) = \overline{X}$. Т.к. $D_\theta\overline{X} = \dfrac{\theta}{n} < \infty$, то $\overline{X}$ - оптимальная оценка для $\tau_2(\theta) = \theta$.

■

Пример 7.3. $X = (X_1, \ldots, X_n)$, $\{X_i\}$ - *н.о.р.*, $X_1 \sim R(0,\theta)$, $\theta > 0$. *Построить оптимальную оценку* $\tau(\theta) = \theta$.

Решение. Знаем, что $X_{(n)} = T(X)$ - достаточная статистика. Здесь $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ - вариационный ряд.

- Докажем, что $T(X) = X_{(n)}$ - полная статистика.

$$F_T(x) = P_\theta(X_{(n)} \le x) = P_\theta(X_1 \le x, \ldots, X_n \le x) =$$
$$= \prod_{i=1}^{n} P(X_i \le x) = \left(\frac{x}{\theta}\right)^n \text{ при } 0 \le x \le \theta$$

Плотность вероятности (существует!):

$$f_T(x,\theta) = \begin{cases} F_t'(x,\theta) = \frac{nx^{n-1}}{\theta^n} \text{ при } 0 \le x \le \theta \\ 0, \text{ при прочих } x \end{cases}$$

Если $E_\theta\varphi(T) = \int\limits_0^\theta \varphi(x) f_T(x,\theta)dx = \int\limits_0^\theta \varphi(x)\dfrac{nx^{n-1}}{\theta^n}dx = 0\ \forall\theta > 0$, то $\varphi(\theta)\theta^{n-1} = 0$ п.в., т.е. $\varphi(\theta) = 0$ п.в. для $\theta > 0$. Таким образом, статистика $T(X) = X_{(n)}$ - полная!

- Уравнение несмещенности для $\tau(\theta) = \theta$ имеет вид $E_\theta\varphi(T) = \int\limits_0^\theta \varphi(x)\dfrac{nx^{n-1}}{\theta^n}dx = \theta$, т.е. $\int\limits_0^\theta \varphi(x)nx^{n-1}dx = \theta^{n+1}$, $\varphi(\theta)\theta^{n-1} = (n+1)\theta^n$ п.в. Отсюда получаем: $\varphi(X_{(n)}) = \dfrac{n+1}{n}X_{(n)}$.

■

8 Гауссовская линейная модель

Определение 8.1. *Случайный вектор $\xi = (\xi_1, \dots, \xi_n)^T$ имеет* ***гауссовское (нормальное) распределение****, если его характеристическая функция имеет вид:*

$$\varphi_\xi(t) = Ee^{it^T\xi} = e^{it^Ta - \frac{1}{2}t^TKt}$$

где $t = (t_1, \dots, t_n)^T \in \mathbb{R}^n$, $a = (a_1, \dots, a_n)^T$, $K = (k_{ij})$ с $i, j = 1, \dots, n$, $k_{ij} \in \mathbb{R}$, $K = K^T$, $K \geqslant 0$, $i^2 = -1$.
Обозначение: $\xi \sim N(a, K)$.

8.1 Свойства Гауссовского закона

1) $\xi_j \sim N(a_j, k_{jj})$
Доказательство. Действительно, $\varphi_{\xi_j}(t_j) = \varphi(0, \dots, t_j, 0, \dots, 0) = e^{it_ja_j - \frac{1}{2}t_j^2k_{jj}}$, а это – характеристическая функция $N(a_j, k_{jj})$ ■
2) $k_{ij} = Cov(\xi_i, \xi_j)$, $k_{ii} = D\xi_i$, K – **ковариационная матрица** ξ.
Доказательство. Из свойства 1) следует, что $a = E\xi$ и, если $\tilde{\xi} = \xi - a$, то $E\tilde{\xi} = 0$. $\varphi_{\tilde{\xi}}(t) = e^{-it^Ta}\varphi_\xi(t) = e^{-\frac{1}{2}t^TKt}$. Отсюда имеем:

$$E\tilde{\xi}_i\tilde{\xi}_j = Cov(\xi_i, \xi_j) = -\frac{\partial^2\varphi_{\tilde{\xi}}(0)}{\partial t_i \partial t_j} = k_{ij}$$

■
3) Если $\xi \sim N(a, K)$, то $\{\xi_1, \dots, \xi_n\}$ независимы тогда и только тогда, когда K – диагональная матрица.

Доказательство. K – диагональная тогда и только тогда, когда:

$$\varphi_\xi(t) = e^{it^T a - \frac{1}{2}\sum\limits_{t=1}^{n} k_{ii}t_i^2} = \prod_{i=1}^{n} e^{it_i a_i - \frac{1}{2}k_{ii}t_i^2} = \prod_{i=1}^{n} \varphi_{\xi_i}(t)$$

Это есть необходимое и достаточное условие независимости. ■

Следствие 8.1. *Если $\eta \sim N(0, E_n)$, то $\{\eta_1, \ldots, \eta_n\}$ – н.о.р. $N(0,1)$ сл.в.*

4) Если $\xi \sim N(a, K)$, а $\eta = A\xi + b$, где A – матрица размера $(m \times n)$, $b = (b_1, \ldots, b_m)^T \in \mathbb{R}^m$, то $\eta \sim N(Aa + b, AKA^T)$
Доказательство.

$$\varphi_\eta(s) = Ee^{is^T(A\xi+b)} = e^{is^T b}Ee^{i(A^T s)^T \xi} = e^{is^T b}e^{i(A^T s)^T a - \frac{1}{2}(A^T s)^T K (A^T s)} =$$
$$= e^{is^T(Aa+b) - \frac{1}{2}s^T(AKA^T)s} \text{ – х.ф., соответствующая сл.в.} \sim N(Aa + b, AKA^T)$$

. ■

Пусть C такая ортогональная матрица ($CC^T = E_n$, т.е. стр. орт.), что $CKC^T = D$, где $D = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_n \end{pmatrix}$, $d_i \geqslant 0$ – собственные числа K.

Тогда $K = C^T DC$. Положим $D^{\frac{1}{2}} := \begin{pmatrix} \sqrt{d_1} & 0 & \ldots & 0 \\ 0 & \sqrt{d_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{d_n} \end{pmatrix}$, $K^{\frac{1}{2}} := C^T D^{\frac{1}{2}} C$.

Тогда имеем следующие свойства:

- $K^{\frac{1}{2}} = (K^{\frac{1}{2}})^T$;
- $K^{\frac{1}{2}} K^{\frac{1}{2}} = K$;
- Если $K > 0$, то $K^{\frac{1}{2}} K K^{-\frac{1}{2}} = E_n$;
- $det(K^{\frac{1}{2}}) = (det K)^{\frac{1}{2}}$

5) Если $\xi \sim N(a, K)$ и $K > 0$, то существует плотность вероятности по мере Лебега $p_\xi(x) = \dfrac{1}{(2\pi)^{\frac{n}{2}} det(K^{\frac{1}{2}})} e^{-\frac{1}{2}(x-a)^T K^{-1}(x-a)}$.
Доказательство. Положим $\eta := K^{-\frac{1}{2}}(\xi - a)$. Тогда $E\eta = K^{-\frac{1}{2}}E(\xi - a) = 0$, $E\eta\eta^T = K^{-\frac{1}{2}}E(\xi - a)(\xi - a)^T K^{-\frac{1}{2}} = K^{-\frac{1}{2}} K K^{-\frac{1}{2}} = E_n$.
В силу свойства 2) имеем $\eta_n \sim N(0, E_n)$, а также:

$$\xi = K^{\frac{1}{2}}\eta + a \tag{1}$$

Задача 8.1. *Если Z и Y – случайные векторы размерности n, A – матрица $(n \times n)$ такая, что $det(A) \neq 0$, $Z = AY + a$, то:*

$$p_z = \frac{1}{det(A)} p_y(A^{-1}(x-a)) \tag{2}$$

Если $\eta \sim N(0, E_n)$, то

$$p_\eta(x) = \prod_{i=1}^{n} p_{\eta_i}(x_i) = \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{1}{2}\sum\limits_{i=1}^{n} x_i^2} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}x^T x}$$

. В силу (1), (2) имеем

$$p_\xi(x) = \frac{1}{(2\pi)^{\frac{n}{2}} det(K^{\frac{1}{2}})} p_\eta(K^{-\frac{1}{2}}(x-a)) = \frac{1}{(2\pi)^{\frac{n}{2}}(detK)^{\frac{1}{2}}} e^{-\frac{1}{2}(x-a)^T K^{-1}(x-a)}$$

. ■

Напоминание 8.1. *Из раздела 4. Если $\{\xi_1, \ldots, \xi_k\}$ – н.о.р. $N(0,1)$ сл.в., то $\eta_k = \xi_1^2 + \ldots + \xi_k^2$ имеет распределение χ^2 Пирсона (хи-квадрат Пирсона) с k степенями свободы: $\eta_k \sim \chi^2(k)$. Тогда $E\eta_k = k$, $D\eta_k = 2k$.*

Лемма 8.1. *Пусть $\xi \sim N(0, \sigma^2 E_n), \sigma^2 > 0$. Тогда:*

1) *если C – ортогональная матрица размера $n \times n$, а $\eta = C\xi$, то:*

$$\eta \sim N(0, \sigma^2 E_n), \text{ т.е. } \xi \overset{d}{=} \eta$$

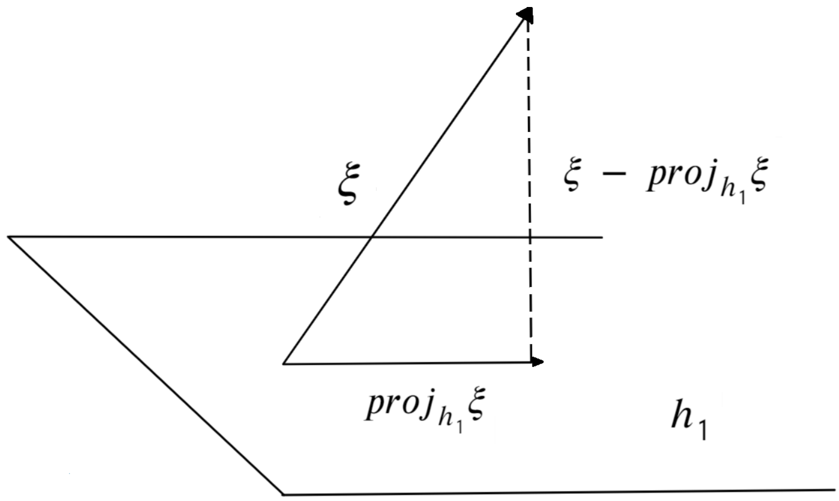
2) *если h_1, h_2 – линейные подпространства $\mathbb{R}^n$, $h_1 \perp h_2$, то $\{proj_{h_i}\xi\}$, $i = 1, 2$ являются независимыми гауссовскими векторами:*

$$E\ proj_{h_i}\xi = 0,\ \frac{1}{\sigma^2}|proj_{h_i}\xi|^2 \sim \chi^2(dim\ h_i)$$

Доказательство. Докажем по пунктам.

1) $E\eta = EC\xi = CE\xi = 0$, $E\eta\eta^T = C(\sigma^2 E_n)C^T = \sigma^2 E_n$. Значит, $\eta \sim N(0, \sigma^2 E_n)$ по свойству 4.
2) Выберем в $\mathbb{R}^n$ ортонормированный базис $\underbrace{e_1, \ldots, e_p,}_{\text{базис } h_1} \underbrace{e_{p+1}, \ldots, e_{p+m},}_{\text{базис } h_2} \ldots, e_n$. Тогда $dim(h_1) = p$. Имеем:

$$proj_{h_1}\xi = \sum_{i=1}^{p} \underbrace{(\xi^T e_i)}_{=\eta_i} e_i = \sum_{i=1}^{p} \eta_i e_i \tag{3}$$



Действительно, $proj_{h_1}\xi = \sum\limits_{i=1}^{p} b_i e_i$. Также $(\xi - proj_{h_1}\xi^T e_j = 0)$ для $j = 1, \ldots, p$, $\xi^T e_j = b_j = \eta_j$, тогда (3) верно. Аналогично получаем:

$$proj_{h_2}\xi = \sum_{i=p+1}^{p+m} \eta_i e_i \tag{4}$$

Если $\eta = (\eta_1, \ldots, \eta_n)^T = (e_1^T, \ldots, e_n^T)^T \xi$, то $(e_1^T, \ldots, e_n^T)^T$ – ортонормированная матрица. В силу пункта 1) этой леммы $\eta \sim N(0, \sigma^2 E_n)$. По (3), (4) $proj_{h_1}\xi$ и $proj_{h_2}\xi$ независимы, так как определяются $\eta_1, \ldots, \eta_p$ и $\eta_{p+1}, \ldots, \eta_{p+m}$ соответственно.

Очевидно, что $\{proj_{h_i}\xi\}$, $i = 1, 2$ – гауссовские векторы, т.е. получаются линейным преобразованием гауссовского вектора η. Например:

$$proj_{h_1}\xi = \underbrace{(e_1, \ldots, e_p)}_{\text{матрица } (n\times p)} (\eta_1, \ldots, \eta_p)^T$$

Так же очевидно, что $E\, proj_{h_1}\xi = 0$, т.к. $E\eta = 0$. Наконец:

$$\frac{1}{\sigma^2}|proj_{h_1}\xi|^2 = \frac{1}{\sigma^2}\left|\sum_{i=1}^{p} \eta_i e_i\right|^2 = \frac{1}{\sigma^2}\sum_{i=1}^{p} \eta_i^2 = \sum_{i=1}^{p}\left(\frac{\eta_i}{\sigma}\right)^2 \sim \chi^2(p = dim\,(h_1))$$

■

8.2 Линейная Гауссовская модель.

$X \sim N(l, \sigma^2 E_n)$, $\sigma^2 > 0, \sigma^2$ – неизвестно, $l \in h$ – неизвестно. h – известное линейное подпространство $\mathbb{R}^n$, $dim h = p < n$. Если $\varepsilon := x - l$, то:

$$X = l + \varepsilon,\ \varepsilon \sim N(0, \sigma^2 E_n), l \in h \tag{5}$$

Неизвестный параметр $\theta \in \mathbb{R}^{n+1}, \theta^T = (l^T, \sigma^2)$. Пусть $h^\perp$ – ортогональное дополнение к h в $\mathbb{R}^n$, то есть множество векторов из $\mathbb{R}^n$, перпендикулярных h. Тогда $dim(h^\perp) = n - p$, и имеем:

$$\forall x \in \mathbb{R}^n\ x = proj_h x + proj_{h^\perp} \tag{6}$$

Найдем достаточную статистику для θ. Плотность X засчет (5), (6) есть

$$p(x, l, \sigma^2) = (\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{1}{2\sigma^2}|x-l|^2} \overset{\text{в силу}(6)}{=} (\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{1}{2\sigma^2}|(proj_h x - l) + proj_{h^\perp}|^2} =$$
$$= (\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{1}{2\sigma^2}(|proj_h x - l|^2 + |proj_{h^\perp}|^2)} = \psi(T(x), \theta) h(x)$$

где $T(x) = \left((proj_h x)^T, |proj_{h^\perp}|^2\right)^T$, $h(x) = 1$. В силу критерия факторизации $T(x)$ – достаточная статистика. Примем без доказательства, что это – полная статистика.

1) Оптимальная оценка l. $proj_h X = proj_h l = proj_h \varepsilon$, тогда $E\ proj_h X = l + E\ proj_h \varepsilon = l$ в силу пункта 2) леммы 7.1. Итак, $proj_h X$ есть функция полной достаточной статистики $E proj_h X = l$. По лемме 7.2 Лемана-Шеффе $\hat{l^n} := proj_h X$ – оптимальная оценка l.
2) Оптимаьная оценка σ^2. $proj_{h^\perp} X = proj_{h^\perp} l + proj_{h^\perp} \varepsilon = proj_{h^\perp} \varepsilon$.
В силу пункта 2) леммы 7.1 имеем:

$$\frac{1}{\sigma^2}|proj_{h^\perp} X|^2 = \frac{1}{\sigma^2}|proj_{h^\perp}\varepsilon|^2 \sim \chi^2(n-p) \tag{7}$$

Значит, $E\frac{1}{\sigma^2}|proj_{h^\perp}\varepsilon|^2 = n - p$, $E\frac{1}{n-p}|proj_{h^\perp} X|^2 = \sigma^2$. В силу леммы 7.2 Лемана-Шеффе $\hat{s_n}^2 := \frac{1}{n-p}|proj_{h^\perp} X|^2$ – оптимальная оценка σ^2. Кроме того, по (7) имеем:

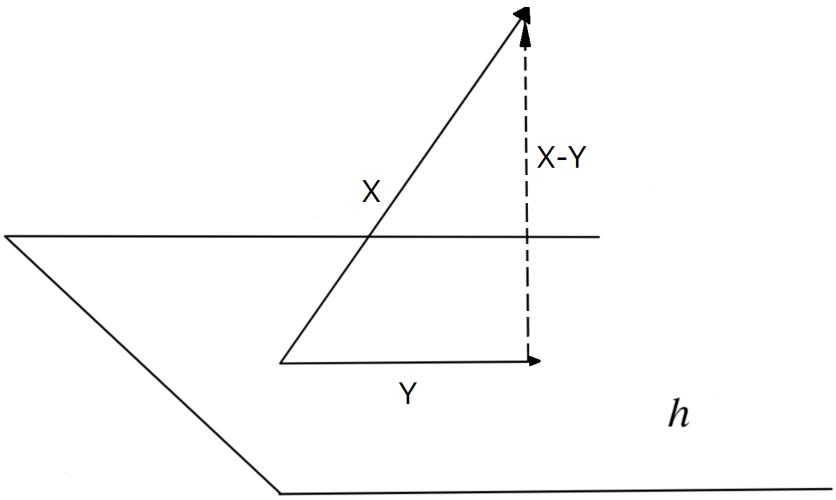
$$\frac{(n-p)\hat{s_n}^2}{\sigma^2} = \frac{1}{\sigma^2}|proj_{h^\perp} X|^2 \sim \chi^2(n-p) \tag{8}$$

3) Независимость $\hat{l_n}$ и $\hat{s_n}^2$. В силу пункта 2) леммы 7.1 $proj_h X = l + proj_h \varepsilon$ и $proj_{h^\perp} X = proj_{h^\perp} \varepsilon$ независимы, значит $\hat{l_n}$ и $\hat{s_n}^2$ независимы.

В силу леммы 7.2 Лемана-Шеффе $(\hat{l_n}^T, \hat{s_n}^2)^T$ – оптимальная оценка вектора $\theta = (l^T, \sigma^2)^T$.

$$\hat{l_n} = proj_h X = \underset{Y \in h}{argmin} |X - Y| = \underset{Y \in h}{argmin} |X - Y|^2 = \underset{Y \in h}{argmin} \sum_{i=1}^{n} (X_i - Y_i)^2$$

Поэтому $\hat{l_n}$ и $\hat{s_n}^2 = \dfrac{1}{n-p} |X - proj_h X|^2$ называются **оценками по методу наименьших квадратов**.



Выберем в h базис, пусть столбцы матрицы размера $(n \times p)$ $Z = \{z_{ij}\}$, $i = 1, \ldots, n, j = 1, \ldots, p$ будут базисными векторами. Тогда для некоторого $c = (c_1, \ldots, c_p)^T$ имеем $l = Zc$, и линейная модель (5) получает вид:

$$X = Zc + \varepsilon \tag{9}$$

Если z_i^T – строки матрицы Z, то (9) эквивалентно:

$$X_i = z_i^T c + \varepsilon_i, \; i = 1, \ldots, n \tag{10}$$

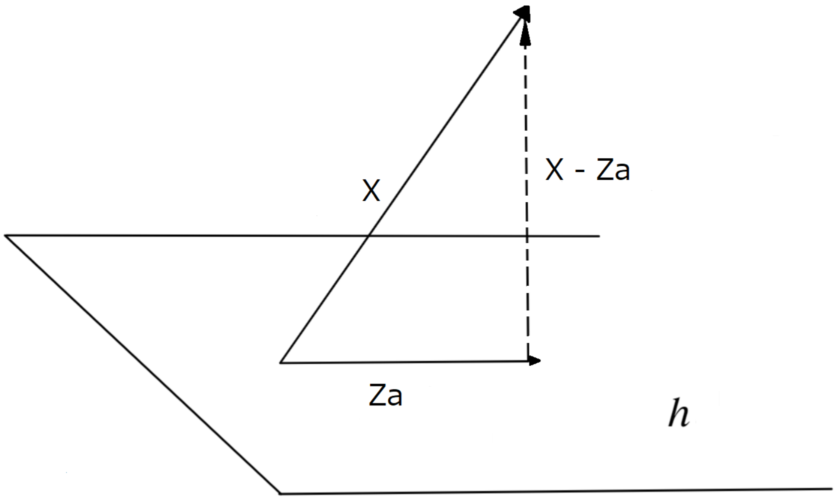
Соотношения (9) и (10) задают **гауссовскую линейную регрессию**, это – еще одна форма записи линейной модели (5).

В (9) неизвестный параметр θ, $\theta^T = (c^T, \sigma^2), dim(\theta) = p + 1$. Матрица Z – **регрессионная матрица**, она известна. X – наблюдение. *Надо оценить* θ, то есть σ^2.

Определение 8.2. ***Оценкой наименьших квадратов (н.к.)*** $\hat{c_n}$ ***вектора*** c *называется решение задачи* $|X - Z\alpha|^2 = \sum\limits_{i=1}^{n}(X_i - z_i^T\alpha)^2 \longrightarrow \min\limits_{\alpha\in\mathbb{R}^p}$, *то есть* $\hat{c_n} = \operatorname*{argmin}\limits_{\alpha\in\mathbb{R}^p}|X - Z\alpha|$.

Пусть h – линейное пространство столбцов Z, то есть столбцы Z – базисные векторы h. Ясно, что $|X - Z\alpha|^2$ достигает минимума при $\alpha = \hat{c_n}$ таком, что $X - Z\hat{c_n} \perp h$, то есть $Z^T(X - Z\hat{c_n}) = 0$, $Z^TX = Z^TZ\hat{c_n}$, тогда получаем:

$$\hat{c_n} = (Z^TZ)^{-1}Z^TX \tag{11}$$



Отметим еще раз: $Z\hat{c_n} = proj_hX$. Матрица Z^TZ в (11) невырождена, так как при $\alpha \neq 0$ $\alpha^TZ^TZ\alpha = |Z\alpha|^2 > 0$ из-за независимости столбцов Z.

Оценка н.к. для σ^2: $\hat{s_n}^2 = \frac{1}{n-p}|X - Z\hat{c_n}|^2$. Разумеется:

$$\hat{s_n}^2 = \frac{1}{n-p}\sum_{i=1}^{n}(x_i - z_i^T\hat{c_n})^2, \quad \hat{s_n}^2 = \frac{1}{n-p}|proj_{h^\perp}X|^2$$

Определение 8.3. *Вектор* $X - Z\hat{c_n} = (\hat{\varepsilon_1}, \ldots, \hat{\varepsilon_n})^T$ *называют* ***вектором остатков***, *а* $\hat{s_n}^2$ – ***остаточной дисперсией.***

Теорема 8.1. *Имеем несколько утверждений:*

1) $\hat{c_n} \sim N(c, \sigma^2(Z^TZ)^{-1})$, $\frac{(n-p)\hat{s_n}^2}{\sigma^2} \sim \chi^2(n-p)$, $E_{c,\sigma^2}\hat{s_n}^2 = \sigma^2$, $D_{c,\sigma^2}\hat{s_n}^2 = \frac{2\sigma^4}{n-p}$.
2) $\hat{c_n}$ *и* $\hat{s_n}^2$ *независимы.*

3) *Оценки* $\hat{c_n}$ *и* $\hat{s_n}^2$ *– оптимальные оценки* c *и* σ^2 *соответственно.*

Доказательство.

1) Распределение $\hat{c_n}$. В силу (11) $\hat{c_n}$ есть линейное преобразование гауссовского вектора X. В силу свойства 4 для гауссовких векторов, $\hat{c_n}$ – гауссовский вектор.

$$E_{c,\sigma^2}\hat{c_n} = E_{c,\sigma^2}(Z^TZ)^{-1}Z^TX = (Z^TZ)^{-1}Z^TE_{c,\sigma^2}X = (Z^TZ)^{-1}Z^TZc = c$$

То есть $\hat{c_n}$ – несмещенная оценка.

$$D_{c,\sigma^2}\hat{c_n} = E_{c,\sigma^2}(\hat{c_n} - c)(\hat{c_n} - c)^T = E_{c,\sigma^2}(Z^TZ)^{-1}Z^T(\varepsilon\varepsilon^T)Z(Z^TZ)^{-1} = \\ = (Z^TZ)^{-1}Z^T(\sigma^2E_n)Z(Z^TZ)^{-1} = \sigma^2(Z^TZ)^{-1}$$

Итак, $\hat{c_n} \sim N(0, \sigma^2(Z^TZ)^{-1})$.
Распределение $\hat{s_n}^2$. Модель (9) эквивалентна $X = l + \varepsilon$, где $l = Zc$, $l \in h$, h – пространство столбцов Z. В силу (13) и (8) $\dfrac{(n-p)\hat{s_n}^2}{\sigma^2} \sim \chi^2(n-p)$.
Значит, $E_{c,\sigma^2}\hat{s_n}^2 = \dfrac{\sigma^2}{n-p}\eta_{n-p} = \sigma^2$. То есть $\hat{s_n}^2$ – несмещенная оценка σ^2.

2) В силу (12) имеем:

$$\hat{c_n} = (Z^TZ)^{-1}Z^TZ\hat{c_n} = (Z^TZ)^{-1}Z^Tproj_hX \tag{14}$$

Т.к. $proj_hX$ и $proj_{h^\perp}X$ независимы, то $\hat{c_n}$ и $\hat{s_n}^2$ независимы.

3) Докажем оптимальность $\hat{c_n}$, оптимальность $\hat{s_n}^2$ доказывается аналогично. Уже имеем, что $\hat{c_n}$ – несмещенная. Надо доказать, что:

$$D_{c,\sigma^2}\hat{c_n} = D_{l,\sigma^2}\widehat{c_n}\ \forall l,\ \sigma^2 > 0 \tag{15}$$

где $\widehat{c_n}$ – любая несмещенная оценка c.
У нас $l = Zc$, то есть $Z^Tl = Z^TZc$, $c = (Z^TZ)^{-1}Z^Tl$. Имеем взаимно однозначное отображение $c \longleftrightarrow l$, $l \in h$, $c \in \mathbb{R}^p$. Иогда левая часть (15) в силу (14) равна $D_{c,\sigma^2}(Z^TZ)^{-1}Z^Tproj_hX$ – ковариация функции полной достаточной статистики $\left((proj_hx)^T, |proj_{h^\perp}|^2\right)^T$ для параметра $(l^T, \sigma^2)^T$. Правая часть (15) есть $E_{l,\sigma^2}\widehat{c_n}$. В силу леммы 7.2 Лемана-Шеффе $D_{l,\sigma^2}(Z^TZ)^{-1}Zproj_hX \leq D_{l,\sigma^2}\widehat{c_n}\ \forall l \in h,\ \sigma^2$. Значит, (15) – верно.

■

8.3 Пример(Гауссовская выборка.)

Пример 8.1. *Пусть $X = (X_1, \dots, X_n)^T$, где $\{X_i\}$ – н.о.р., $X_1 \sim N(a, \sigma^2)$. Построим оптимальные оценки a и σ^2, исследуем их свойства.*

Пусть $\varepsilon_i = X_i - a,\ i = 1, \dots, n$. Тогда:

$$X_i = a + \varepsilon_i,\ i = 1, \dots, n;\ \{\varepsilon_i\} \text{ – н.о.р., } \varepsilon_1 \sim N(0, \sigma^2) \tag{16}$$

Уравнение (16) – частный случай линейной регрессии (10), где $z_i^T = 1,\ c = a,\ p = 1$. Значит, оптимальная оценка для a – о.н.к., которая получается решением задачи $\sum\limits_{i=1}^{n}(X_i - \alpha)^2 \longrightarrow \min\limits_{\alpha}$. Эта задача эквивалентна решениюуравнения $-2\sum\limits_{i=1}^{n}(X_i - \alpha) = 0$, корень – $\hat{\alpha_n} = \bar{X}$.

Оптимальная оценка для σ^2 – остаточная дисперсия: $\hat{s_n}^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2 = S^2$. Матрица $Z = (z_1^T, \dots, z_n^T)^T = (1, \dots, 1)$, линейное пространство столбцов матрицы Z – линейное пространство, натянутое на вектор $(1, \dots, 1)^T$, $l = (a_1, \dots, a_n)^T$, оптимальная оценка l – $\hat{l} = Z\hat{c_n} = (\bar{X}, \dots, \bar{X})^T$. Из теоремы 8.1 $\bar{X} \sim N(a, \dfrac{\sigma^2}{n})$, $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

$\bar{X}$ и S^2 независимы. $D_{a,\sigma^2}S^2 = \dfrac{2\sigma^2}{n-1}$. Ковариационная матрица вектора $\hat{\theta}_n = (\bar{X}, S^2)^T$ есть $D_{a,\sigma^2}\hat{\theta}_n = \begin{pmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{2\sigma^2}{n-1} \end{pmatrix}$. Напомним, ***матрица информации Фишера*** *(смотри раздел 5) равна $I(\theta) = \begin{pmatrix} \dfrac{n}{\sigma^2} & 0 \\ 0 & \dfrac{n}{2\sigma^2} \end{pmatrix}$, поэтому:*

$$D_{a,\sigma^2}\hat{\theta}_n = \begin{pmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{2\sigma^2}{n-1} \end{pmatrix} > I^{-1}(\theta) = \begin{pmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{2\sigma^2}{n} \end{pmatrix}$$

Значит, оценка $(\bar{X}, S^2)^T$ является оптимальной оценкой $(a, \sigma^2)^T$, но не является эффективной в $C_{\mathbb{R}}$.

Определение 8.4. *Пусть $\xi_0, \ldots, \xi_k$ – н.о.р. $N(0,1)$ сл.в. Случайная величина*

$$t_k = \frac{\xi_0}{\sqrt{\frac{1}{k}(\xi_1^1 + \ldots + \xi_k^2)}}$$

имеет ***распределение Стьюдента с k степенями свободы****.*

То есть $t_k = \dfrac{\xi_0}{\sqrt{\frac{1}{k}\eta_k}}$, где $\xi_0 \sim N(0,1)$, $\eta_k \sim \chi^2(k)$, ξ_0 и η_k независимы.

Поскольку $\dfrac{n^{\frac{1}{2}}(\bar{X} - a)}{\sigma} \sim N(0,1)$, а $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, то:

$$\frac{n^{\frac{1}{2}}(\bar{X} - a)}{\sigma}\sqrt{\frac{1}{n-1}\frac{(n-1)S^2}{\sigma^2}} = \frac{n^{\frac{1}{2}}(\bar{X} - a)}{S} \sim S(n-1)$$

Величина $\dfrac{n^{\frac{1}{2}}(\bar{X} - a)}{S}$ называется **стьюдентовой дробью**.

9 Введение в доверительное оценивание

Пусть наблюдение $X = (X_1, \dots, X_n)$, $X \sim P_\theta$, $\theta \in \Theta \in \mathbb{R}^1$, Θ – интервал. Пусть $T_1(X) \leq T_2(X)$, $(T_1(X), T_2(X)) \subseteq \Theta$.

Определение 9.1. *Если $P_\theta(T_1(X) < \theta < T_2(X)) \geq 1 - \alpha \ \forall \theta \in \Theta$, то случайный интервал $(T_1(X), T_2(X))$ называется **доверительным интервалом уровня** $1 - \alpha$, $0 < \alpha < 1$.*

Интервал $(T_1(X), T_2(X))$ можно понимать как интервальную оценку (в отличие от точечной) параметра θ. Он покрывает неизвестное θ с вероятностью, не меньшей, чем $1 - \alpha$.

9.1 Доверительные интервалы для параметров Гауссовских выборок

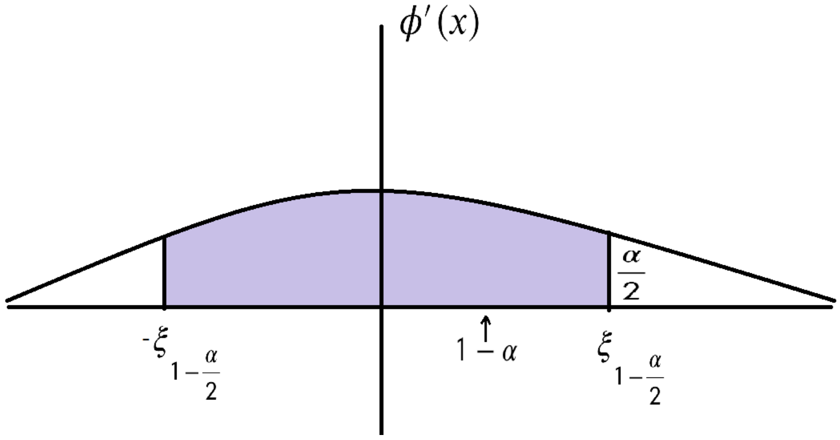
Доверительный интервал для среднего a при известной дисперсии σ^2

$X = (X_1, \dots, X_n)$, $\{X_i\}$ – н.о.р., $X_1 \sim N(a, \sigma^2)$.

Оптимальная оценка для a – $\bar{X} \sim N(a, \frac{\sigma^2}{n})$. Значит, $\frac{n^{\frac{1}{2}}(\bar{X} - a)}{\sigma} \sim N(0, 1)$. Пусть $\phi(x)$ – функция распределения $N(0,1)$, то есть $\phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\inf}^{x} e^{-\frac{t^2}{2}} dt$. Пусть $\xi_\alpha : \phi(\xi_\alpha) = \alpha$, $0 < \alpha < 1$.

Тогда $\forall a \ P_a(|\frac{n^{\frac{1}{2}}(\bar{X}-a)}{\sigma}| < \xi_{1-\frac{\alpha}{2}}) = 1-\alpha$, то есть с вероятностью $1-\alpha$:

$$-\xi_{1-\frac{\alpha}{2}} < \frac{n^{\frac{1}{2}}(\bar{X}-a)}{\sigma} < \xi_{1-\frac{\alpha}{2}}, \ \bar{X} - \frac{\sigma\xi_{1-\frac{\alpha}{2}}}{\sqrt{n}} < a < \bar{X} + \frac{\sigma\xi_{1-\frac{\alpha}{2}}}{\sqrt{n}}$$



$$\xi_{1-\frac{\alpha}{2}} = -\xi_{\frac{\alpha}{2}}$$

Интервал:

$$(\bar{X} - \frac{\sigma\xi_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \bar{X} + \frac{\sigma\xi_{1-\frac{\alpha}{2}}}{\sqrt{n}}) \tag{1}$$

называется **доверительным интервалом уровня** $1-\alpha$, его длина $ln = \frac{2\sigma\xi_{1-\frac{\alpha}{2}}}{\sqrt{n}}$ю

Свойство 9.1. *Имеем несколько свойств:*

1) $\alpha \to 0 \ \Rightarrow \ \xi_{1-\frac{\alpha}{2}} \to \inf \ \Rightarrow \ ln \to \inf$.

   Замечание. *Обычно* $\alpha = 0.05, \ 0.01, \ldots$ *и* α *фиксировано.*

2) $n \to \inf \ \Rightarrow \ ln \to 0$.
3) $\sigma \to 0 \ \Rightarrow \ ln \to 0$.

Замечание. *Доверительных интервалов много. Например,* $\frac{X_1 - \theta}{\sigma} \sim N(0,1)$, *и на этой статистике можно построить доверительный интервал уровня* $1-\alpha$.

Какой доверительный интервал наилучший?

Определение 9.2. *Доверительный интервал* (T_1, T_2) *уровня* $1-\alpha$ *называется* ***несмещенным***, *если* $P_\theta(T_1 < \theta' < T_2) \leq 1-\alpha \ \forall\theta, \ \theta'$ *таких, что* $\theta \neq \theta'$. *То есть вероятность накрытия неверного параметра всегда* ***не больше*** *вероятности накрытия верного параметра.*

Определение 9.3. *Несмещенный доверительный интервал* (T_1, T_2) *уровня* $1-\alpha$ *называется* ***наиболее точным****, если он минимизирует вероятность* $P_\theta(T_1 < \theta' < T_2)$ $\forall \theta,\ \theta'$ *таких, что* $\theta \neq \theta'$ *в классе всех* ***несмещенных*** *доверительных интервалов* (T_1, T_2) *уровня* $1-\alpha$.

Можно показать, что (1) является наиболее точным несмещенным доверительным интервалом уровня $1-\alpha$, то есть *оптимальным*.

Доверительный интервал для среднего a при неизвестной дисперсии σ^2

$X = (X_1, \ldots, X_n)$, $\{X_i\}$ – н.о.р., $X_1 \sim N(a, \sigma^2)$, $n \geq 2$.

Определение 9.4. *Если* $\xi_0, \ldots, \xi_k$ *– н.о.р. N(0, 1) сл.в., то сл.в.*

$$t_k = \frac{\xi_0}{\sqrt{\frac{1}{k}(\xi_1^2 + \ldots + \xi_k^2)}}$$

имеет ***распределение Стьюдента с к степенями свободы****. Обозначение:* $t_k \sim S(k)$.

Очевидно, $t_k = \dfrac{\xi_0}{\sqrt{\frac{1}{k}\eta_k}}$, где ξ_0, η_k независимы, $\eta_k \sim \chi^2(k)$.

Пусть $S_k(x) := P(t_k \leq x)$ – ф.р. t_k. Пусть $S_k(t_\alpha(k)) = \alpha$, $0 < \alpha < 1$, то есть $t_\alpha(k)$ – **квантиль уровня α ф.р. $S_k(x)$**. Поскольку $t_k \overset{d}{\ } t_k$, то плотность вероятности $S_k(x)'$ – четная функция. Значит, $-t_\alpha(x) = t_{1-\alpha}(x)$.

Известно, что $\bar{X}$ и s^2 – оптимальные оценки a, σ^2, $\bar{X}$ и s^2 независимы,

$$\frac{n^{\frac{1}{2}}(\bar{X} - a)}{\sigma} \sim N(0,1),\ \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

Здесь $s^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})$. Значит:

$$\frac{n^{\frac{1}{2}}(\bar{X} - a)}{\sigma} \Big/ \sqrt{\frac{1}{n-1}\frac{(n-1)s^2}{\sigma^2}} = \frac{n^{\frac{1}{2}}(\bar{X} - a)}{s} \sim S(n-1).$$

Значит:

$$P_{a,\sigma^2}(|\frac{n^{\frac{1}{2}}(\bar{X} - a)}{s}|) < t_{1-\frac{\alpha}{2}}(n-1) = 1 - \alpha.$$

Получаем доверительный интервал уровня $1-\alpha$:

$$\bar{X} - \frac{st_{1-\frac{\alpha}{2}}(n-1)}{\sqrt{n}} < a < \bar{X} + \frac{st_{1-\frac{\alpha}{2}}(n-1)}{\sqrt{n}}$$

Доверительный интервал для дисперсии σ^2 при неизвестном среднем a

$X = (X_1, \ldots, X_n),\ \{X_i\}$ – н.о.р., $X_1 \sim N(a, \sigma^2),\ n \geq 2$.

Знаем, что $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$. Обозначим за $\chi_{(n-1)}(x)$ ф.р. $\chi^2(n-1),\ x_\alpha(n-1)$ – квантиль уровня α, то есть $\chi_{(n-1)}(x_\alpha(n-1)) = \alpha,\ 0 < \alpha < 1$. Тогда $P_{a,\sigma^2}(x_{\frac{\alpha}{2}} < \dfrac{(n-1)s^2}{\sigma^2} < x_{1-\frac{\alpha}{2}}(n-1)) = 1-\alpha$. Доверительный интервал уровня $1-\alpha$:

$$\frac{(n-1)s^2}{x_{1-\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)s^2}{x_{\frac{\alpha}{2}}(n-1)}$$

9.2 Оценивание параметров линейной регрессии

Если $\eta_k \sim \chi^2(k),\ \upsilon_m \sim \chi^2(m),\ \eta_k$ и υ_m независимы, то сл.в. $f_{k,m} = \dfrac{\frac{1}{k}\eta_k}{\frac{1}{m}\upsilon_m}$ имеет распределение Фишера с (k, m) степенями свободы. Пишем $f_{k,m} \sim F(k, m)$. Пусть $F_{k,m}(x)$ – ф.р., то есть $F_{k,m}(f_\alpha(k, m)) = \alpha,\ 0 < \alpha < 1$.

Лемма 9.1. *Если* $\xi \in \mathbb{R}^k,\ \xi \sim N(0, \Sigma),\ \Sigma > 0,$ *то* $\sigma^T \Sigma^{-1} \sigma \sim \chi^2(k)$.

Доказательство.

$$\sigma^T \Sigma^{-1} \sigma = (\Sigma^{-\frac{1}{2}}\xi)^T (\Sigma^{-\frac{1}{2}}\xi) = |\Sigma^{-\frac{1}{2}}\xi|^2$$

При этом $\eta := \Sigma^{-\frac{1}{2}}\xi \sim N(0, E_k)$, так как η – гаусс. Тогда:

$$E\eta = \Sigma^{-\frac{1}{2}} E\eta = 0,\ Cov(\eta, \eta) = E\eta\eta^T = \Sigma^{-\frac{1}{2}} \Sigma \Sigma^{-\frac{1}{2}} = E_k,$$
$$\text{т.е. } |\Sigma^{-\frac{1}{2}}\xi|^2 = |\eta|^2 = \eta_1^1 + \ldots + \eta_k^2 \sim \chi^2(k)$$

∎

Рассмотрим регрессию $X = Zc + \varepsilon,\ \varepsilon \sim N(0, \sigma^2 E_n)$. Пусть $\hat{c_n}$ – о.н.к. для c, $\hat{s_n}^2$ – о.н.к. для σ^2.

Тогда известно, $\hat{c_n} \sim N(c, \sigma^2(Z^TZ)^{-1})$, $\frac{(n-p)\hat{s_n}}{\sigma^2} \sim \chi^2(n-p)$, $\hat{c_n}$ и $\hat{s_n}^2$ независимы. Значит, в силу леммы 9.1:

$$\frac{1}{\sigma^2}(\hat{c_n}-c)^T(Z^TZ)(\hat{c_n}-c) \sim \chi^2(p),$$

$$f_{p,n-p} := \frac{\frac{1}{p}\frac{1}{\sigma^2}(\hat{c_n}-c)^T(Z^TZ)(\hat{c_n}-c)}{\frac{1}{n-p}\frac{(n-p)\hat{s_n}^2}{\sigma^2}} = \frac{(\hat{c_n}-c)^T(Z^TZ)(\hat{c_n}-c)}{p\hat{s_n}^2} \sim F(p, n-p)$$

Значит:

$$P_{c,\sigma^2}((\hat{c_n}-c)^T(Z^TZ)(\hat{c_n}-c) \le p\hat{s_n}^2 f_{1-\alpha}(p, n-p)) = 1-\alpha$$

Доверительный эллипсоид уровня $1-\alpha$:

$$\{c: \ (\hat{c_n}-c)^T(Z^TZ)(\hat{c_n}-c) < p\hat{s_n}^2 f_{1-\alpha}(p, n-p)\}$$

Он накрывает неизвестный c с вероятностью $1-\alpha$.

Пусть $c = (c_1, \dots, c_p)^T$, $\hat{c_n} = (\hat{c_1}, \dots, \hat{c_p})^T$, тогда $\hat{c_{in}} \sim N(c_i, \sigma^2 a_{ii})$, где $(Z^TZ)^{-1} = (a_{ij}),\ i, j = 1, \dots, p$. Так как $\hat{c_{in}}$ и $\hat{s_n}^2$ независимы, то

$$t_{n-p} := \frac{\hat{c_{in}} - c_i}{\sqrt{\sigma^2 a_{ii}}} / \sqrt{\frac{1}{n-p}\frac{(n-p)\hat{s_n}^2}{\sigma^2}} = \frac{\hat{c_{in}} - c_i}{\hat{s_n}\sqrt{a_{ii}}} \sim S(n-p).$$

Доверительный интервал для c_i уровня $1-\alpha$:

$$\hat{c_{in}} - \hat{s_n}\sqrt{a_{ii}}t_{1-\frac{\alpha}{2}}(n-p) < c_i < \hat{c_{in}} + \hat{s_n}\sqrt{a_{ii}}t_{1-\frac{\alpha}{2}}(n-p)$$

.

9.3 Ассимптотический доверительный интервал

Пусть для неизвестного параметра $\theta \in \Theta \in \mathbb{R}^1$ существует ассимптотически нормальная оценка $\hat{\theta_n}$, то есть:

$$n^{\frac{1}{2}}(\hat{\theta_n} - \theta) \xrightarrow{d} N(0, \sigma^2(\theta)),\ n \to \inf \tag{2}$$

.

Предположим, что $\sigma^2(\theta) > 0\ \forall\theta \in \Theta$ и $\sigma^2(\theta)$ непрерывна по θ. В силу (2) $\hat{\theta}_n - \theta = n^{-\frac{1}{2}} n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \xrightarrow{P} 0,\ n \to \inf$, то есть $\hat{\theta}_n$ – состоятельная оценка θ. Значит:

$$\hat{\sigma_n}^2 := \sigma^2(\theta),\ n \to \inf \tag{3}$$

В силу (2), (3) $\dfrac{n^{\frac{1}{2}}(\hat{\theta}_n - \theta)}{\hat{\sigma_n}} \xrightarrow{d} N(0,1),\ n \to \inf$. Значит, $P_\theta(|\dfrac{n^{\frac{1}{2}}(\hat{\theta}_n - \theta)}{\hat{\theta}_n}| < \xi_{1-\frac{\alpha}{2}}) \to 1-\alpha,\ n \to \inf$.

Ассимптотический доверительный интервал уровня $1-\alpha$ имеет вид:

$$\hat{\theta}_n - \frac{\hat{\sigma_n}\xi_{1-\frac{\alpha}{2}}}{\sqrt{n}} < \theta < \hat{\theta}_n + \frac{\hat{\sigma_n}\xi_{1-\frac{\alpha}{2}}}{\sqrt{n}}$$

Он накрывает неизвестный параметр θ прмерно с вероятностью $1-\alpha$ при больших n.

9.4 Примеры

Пример 9.1. *$X = (X_1, \ldots, X_n),\ \{X_i\}$ – н.о.р., $X_1 \sim Pois(\theta),\ \theta > 0$. Тогда $n^{\frac{1}{2}}(\bar{X} - \theta) \xrightarrow{d} N(0,\theta)$, а т.к. $\bar{X} \xrightarrow{P} \theta$, то $\dfrac{n^{\frac{1}{2}}(\bar{X}-\theta)}{\sqrt{\bar{X}}} \xrightarrow{d} N(0,1)$. Ассимптотический доверительный интервал уровня $1-\alpha$:*

$$\bar{X} - \frac{\sqrt{\bar{X}}\xi_{1-\frac{\alpha}{2}}}{\sqrt{n}} < \theta < \bar{X} + \frac{\sqrt{\bar{X}}\xi_{1-\frac{\alpha}{2}}}{\sqrt{n}}$$

Пример 9.2. *$X_1 \sim R(0,\theta),\ \theta > 0$.*

$$E_\theta X_1 = \frac{\theta}{2},\ D_\theta X_1 = \frac{\theta^2}{12},\ 2\bar{X} \xrightarrow{P} \theta,\ \frac{n^{\frac{1}{2}}(\bar{X} - \frac{\theta}{2})}{\frac{\theta}{2}\sqrt{3}} \xrightarrow{d} N(0,1),\ n \to \inf$$

Значит:

$$\frac{\sqrt{3n}(2\bar{X}-\theta)}{\theta} \xrightarrow{d} N(0,1),\ \frac{\sqrt{3n}(2\bar{X}-\theta)}{2\bar{X}} \xrightarrow{d} N(0,1)$$

Ассимптотический доверительный интервал:

$$2\bar{X} - \frac{2\bar{X}\xi_{1-\frac{\alpha}{2}}}{\sqrt{3n}} < \theta < 2\bar{X} + \frac{2\bar{X}\xi_{1-\frac{\alpha}{2}}}{\sqrt{3n}}$$

10 Ассимптотически оптимальные оценки.

10.1 Сходимости, лемма Слуцкого

Пусть случайные величины $\xi_n, \xi \in \mathbb{R}^n$ определены на колмогоровой тройке (Ω, F, P). $F_n(x)$ - функция распределения ξ_n, $\varphi_n(t)$ - характеристическая функция, Q_n - распределение (мера на множестве борелевских подмножеств).

Определение 10.1. *Говорят, что F_n сходится к F **в основном** ($F_n(x) \Rightarrow F(x)$), если $F_n(x) \to F(x)$ $\forall x \in \mathbb{C}(F)$.*

Определение 10.2. *Говорят, что Q_n сходится **слабо** к Q ($Q_n \xrightarrow{W} Q$), если для любой непрерывной и ограниченной функции $g : \mathbb{R}^k \to \mathbb{R}^1$*

$$\int\limits_{\mathbb{R}^k} g(x)Q_n(dx) \to \int\limits_{\mathbb{R}^1} g(x)Q(dx) \;\Leftrightarrow\; Eg(\xi_n) \to Eg(\xi)$$

Теорема 10.1. *Следующие условия эквивалентны:*

1. $F_n \Rightarrow F$
2. $Q_n \xrightarrow{W} Q$
3. $\varphi_n(t) \to \varphi(t)$ $\forall t \in \mathbb{R}^k$

Определение 10.3. *Если выполнено одно из условий 1) – 3) из предыдущей теоремы, то говорят, что ξ_n сходится к ξ **по распределению** ($\xi_n \xrightarrow{d} \xi$).*

Теорема 10.2 (**о наследовании сходимости**). *Пусть $\xi_n, \xi \in \mathbb{R}^k$ и отображение $H : \mathbb{R}^k \to \mathbb{R}^1$ непрерывно. Тогда:*

1. *если $\xi_n \xrightarrow{d} \xi$, то $H(\xi_n) \xrightarrow{d} H(\xi)$*
2. *если $\xi_n \xrightarrow{P} \xi$, то $H(\xi_n) \xrightarrow{P} H(\xi)$*

Лемма 10.1 (**Слуцкого**). *Пусть* $\xi_n, \xi, \eta_n, a \in \mathbb{R}^1$ *и* $\xi_n \xrightarrow{d} \xi, \eta_n \xrightarrow{P} a$. *Тогда:*

I) $\xi_n + \eta_n \xrightarrow{d} \xi + a$

II) $\xi_n \eta_n \xrightarrow{d} a\xi$

Доказательство. Достаточно, чтобы была следующая сходимость:

$$(\xi_n, \eta_n)^T \xrightarrow{d} (\xi, a)^T \qquad (1)$$

Действительно, если (1) верно, то при $H(x, y) = x + y$ по теореме 1.2 получаем I, а при $H(x, y) = xy$ получаем II.

Докажем (1) : $(\xi_n, \eta_n)^T \xrightarrow{d} (\xi, a)^T$. Проверим, что характеристическая функция $(\xi_n, \eta_n)^T$ сходится к характеристической функции $(\xi, a)^T$:

$$|Ee^{it\xi_n + is\eta_n} - Ee^{it\xi + isa}| \leq |Ee^{it\xi_n + is\eta_n} - Ee^{it\xi_n + isa}| + |Ee^{it\xi_n + isa} - Ee^{it\xi + isa}| = \alpha_n + \beta_n$$

$$\alpha_n \leq E|e^{it\xi_n}(e^{is\eta_n} - e^{isa})| = E|e^{is\eta_n} - e^{isa}| = Eg(\eta_n)$$

$$g(x) = |e^{isx} - e^{isa}| \text{ - непрерывная и ограниченная, } \eta_n \xrightarrow{d} a \ \Rightarrow$$

$$\text{по теореме 1.2 } \ Eg(\eta_n) \to Eg(a) = 0 \ \Rightarrow \ \alpha_n \to 0$$

$$\beta_n = |Ee^{isa}(e^{it\xi_n} - e^{it\xi})| = |e^{isa}E(e^{it\xi_n} - e^{it\xi})| = |Ee^{it\xi_n} - Ee^{it\xi}| \to 0, \text{ т.к. } \xi_n \xrightarrow{d} \xi.$$

Т.о. $\varphi_n(t) \to \varphi(t)$. ■

10.2 Асимптотически нормальные, состоятельные оценки, асимптотический доверительный интервал

Пусть наблюдение $X \sim P_\theta$, $\theta \in \Theta \subseteq \mathbb{R}^k$. $\hat{\theta}_n$ - оценка θ.

Определение 10.4. *Если* $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \Sigma(\theta))$ $\forall \theta \in \Theta$ *и* $0 < \Sigma(\theta) < \infty$, *то* $\hat{\theta}_n$ *называется* ***асимптотически нормальной оценкой***.

Определение 10.5. *Если* $\hat{\theta}_n \xrightarrow{P} \theta$ $\forall \theta \in \Theta$, *то* $\hat{\theta}_n$ *называется* ***состоятельной оценкой***.

Пусть $\theta \in \Theta \subseteq \mathbb{R}^1$, т.е. θ и $\hat{\theta}_n$ - скаляры.

Если $\hat{\theta}_n$ - состоятельная оценка θ, то при больших n $\hat{\theta}_n \simeq \theta$ с вероятностью, близкой к единице.

Если $\hat{\theta}_n$ - асимптотически нормальная оценка θ, то есть $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta)),\ 0 < \sigma^2(\theta) < \infty\ \forall \theta \in \Theta$, то:

1. $\hat{\theta}_n$ - состоятельная оценка θ, т.к. $\hat{\theta}_n - \theta = \frac{1}{\sqrt{n}}\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{P} 0$ в силу пункта II леммы Слуцкого.
2. скорость сходимости $\hat{\theta}_n$ к θ есть $\mathcal{O}(\sqrt{n})$
3. при больших n случайную величину $\sqrt{n}(\hat{\theta}_n - \theta)$ можно рассматривать как гауссовскую величину.

   Пример 10.1. *Пусть $\sigma^2(\theta)$ - непрерывная функция, а θ неизвестно. Тогда:*

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\hat{\theta}_n)} = \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\theta)} \cdot \frac{\sigma(\theta)}{\sigma(\hat{\theta}_n)}$$

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\theta)} \xrightarrow{d} N(0,1),\ \frac{\sigma(\theta)}{\sigma(\hat{\theta}_n)} \xrightarrow{P} 1 \Rightarrow$$

$$\Rightarrow \frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\hat{\theta}_n)} \xrightarrow{d} \eta \sim N(0,1) \text{ в силу пункта } II) \text{ леммы Слуцкого.}$$

$$\text{Значит: } P_\theta\left(\left|\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\hat{\theta}_n)}\right| < \xi_{1-\frac{\alpha}{2}}\right) \to P(|\eta| < \xi_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

   Т.е. примерно с вероятность $1-\alpha$ выполнено неравенство:

$$\hat{\theta}_n - \frac{1}{\sqrt{n}}\sigma(\hat{\theta}_n)\xi_{1-\frac{\alpha}{2}} < \theta < \hat{\theta}_n + \frac{1}{\sqrt{n}}\sigma(\hat{\theta}_n)\xi_{1-\frac{\alpha}{2}}$$

   Это называется ***асимптотическим доверительным интервалом*** *для θ уровня $1-\alpha$.*

4. асимтотические гауссовские оценки можно сравнивать между собой: если $\sqrt{n}(\hat{\theta}_{in} - \theta) \xrightarrow{d} N(0, \sigma_i^2(\theta))\ \forall i = 1, 2, \ldots$, то можно определить **асимтотическую нормальную эффективность**:

$$e_{1,2} = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}$$

$$\text{напоминание: } e_{1,2} = \lim_{n\to\infty}\frac{n'(n)}{n}, \text{ где } \begin{cases} \sqrt{n}(\hat{\theta_{1n}} - \theta) \xrightarrow{d} N(0, \sigma_1^2(\theta)) \\ \sqrt{n}(\hat{\theta_{2n'}} - \theta) \xrightarrow{d} N(0, \sigma_1^2(\theta)) \end{cases}$$

Вопрос: существует ли θ_n^* такая, что $e_{\theta_n^*,\hat{\theta}_n}(\theta) \geq 1 \ \forall \hat{\theta}_n \ \forall \theta \in \Theta$?

Если есть, то θ_n^* требует не больше наблюдений, чем любая $\hat{\theta}_n$, чтобы достичь одинаковой с $\hat{\theta}_n$ точности.

Предельная дисперсия $\sqrt{n}(\theta_n^* - \theta)$ должна быть не больше асимптотической дисперсии $\sqrt{n}(\hat{\theta}_n - \theta)$ для любой асимптотической гауссовской оценки $\hat{\theta}_n$.

10.3 Теорема Бахадура, асимптотически эффективная оценка

Теорема 10.3 (Бахадура)**.** *Пусть $X_1, \ldots, X_n$ - н.о.р.с.в., X_1 имеет плотность вероятности $f(x,\theta), \theta \in \Theta \subseteq \mathbb{R}^1$ по мере ν. Пусть выполнены условия:*

(i) *θ - интервал*
(ii) *носитель $N_f = \{x \colon f(x,\theta) > 0\}$ не зависит от θ*
(iii) *$\forall x \in N_f$ плотность $f(x,\theta)$ дважды непрерывно дифференцируема по θ*
(iv) *интеграл $\int f(x,\theta)\nu(dx)$ можно дважды дифференцировать по θ, внося знак дифференцирования под знак интеграла*
(v) *информация Фишера $0 < i(\theta) < \infty \ \forall \theta \in \Theta$*
(vi) *$|\frac{\partial^2}{\partial\theta^2} \ln f(x,\theta)| \leq M(x) \ \forall x \in N_f, \theta \in \Theta$ и $E_\theta M(X_1) < \infty$*

Тогда если $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma^2(\theta))$, то $\sigma^2(\theta) \geq \frac{1}{i(\theta)}$ всюду за исключением множества Лебеговой меры нуль.

Замечание. *Если вдобавок $\sigma^2(\theta)$ и $i(\theta)$ непрерывны, то $\sigma^2(\theta) \geq \frac{1}{i(\theta)} \ \forall \theta \in \Theta$.*

Определение 10.6. *Если $\theta, \hat{\theta}_n \in \mathbb{R}^1$ и $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \frac{1}{i(\theta)}), n \to \infty, \forall \theta \in \Theta$, причем $0 < i(\theta) < \infty$, то $\hat{\theta}_n$ называется **асимтотически эффективной оценкой** (асимптотически оптимальной оценкой).*

10.4 Правдоподобие, экстремальное свойство правдоподобия

Пусть далее $X = (X_1, \ldots, X_n)$, $X \sim P_\theta, \theta \in \Theta \subseteq \mathbb{R}^1$.

Условие (А)

(i) θ - интервал, $P_{\theta_1} \neq P_{\theta_2}$ при $\theta_1 \neq \theta_2$

(ii) $X_1, \ldots, X_n$ - н.о.р.с.в., X_1 имеет плотность вероятности $f(x, \theta)$ по мере ν, носитель $N_f = \{x \colon f(x, \theta) > 0\}$ не зависит от θ.

Плотность вектора X есть $p(x, \theta) = \prod\limits_{i=1}^{n} f(x_i, \theta)$.

Определение 10.7. *Функция $p(x, \theta)$ как функция θ при фиксированном x называется* ***правдоподобием****.*

Определение 10.8. *Функция $Ln(x, \theta) = \ln p(x, \theta) = \sum\limits_{i=1}^{n} \ln f(x_i, \theta)$ называется* ***логарифмическим правдоподобием****.*

Пусть θ_0 - истинное значение параметра.

Теорема 10.4 (**экстремальное свойство правдоподобия**)**.** *Пусть выполнено условие (А), пусть $E_{\theta_0} |\ln f(x_1, \theta)| < \infty \;\Rightarrow\; P_{\theta_0}(p(x, \theta_0) > p(x, \theta)) \xrightarrow[n \to \infty]{} 1$, когда $\theta_0 \neq \theta$.*

Доказательство.

$$p(x, \theta_0) > p(x, \theta) \;\Leftrightarrow\; \ln p(x, \theta_0) > \ln p(x, \theta) \;\Leftrightarrow$$

$$\Leftrightarrow\; \eta_n := \frac{1}{n} \sum_{i=1}^{n} \ln\left(\frac{f(x_i, \theta)}{f(x_i, \theta_0)}\right) < 0, \text{ где } \ln\left(\frac{f(x_i, \theta)}{f(x_i, \theta_0)}\right) \text{ - борелевские функции.}$$

Т.е. надо показать, что $P_{\theta_0}(\eta_n < 0) \to 1$

$$\text{но } \eta_n = \frac{1}{n} \sum_{i=1}^{n} \ln\left(\frac{f(x_i, \theta)}{f(x_i, \theta_0)}\right) \xrightarrow{P} E_\theta \ln\left(\frac{f(x_1, \theta)}{f(x_1, \theta_0)}\right)$$

(в силу слабого ЗБЧ в форме Чебышева).

Неравенство Йенсена: пусть $g(x)$ выпуклая снизу борелевская функция, $E|\xi| < \infty$, $E|g(\xi)| < \infty \;\Rightarrow\; g(E\xi) \leq Eg(\xi)$. Если ξ не является почти наверно константой и g строго выпукла, то неравенство строгое.

Функция $-\ln x$ строго выпукла и $\frac{f(x_1,\theta)}{f(x_1,\theta_0)}$ не является почти наверно константой в силу пункта (i) условия (A). Тогда в силу неравенства Йенсена получаем:

$$E_{\theta_0} \ln \frac{f(x_1,\theta)}{f(x_1,\theta_0)} < \ln E_{\theta_0} \frac{f(x_1,\theta)}{f(x_1,\theta_0)} =$$

$$= \ln \int\limits_{N_f} \frac{f(x,\theta)}{f(x,\theta_0)} f(x,\theta_0)\nu(dx) = \ln 1 = 0 \text{ (из условия нормировки)}$$

Но если η_n сходится по вероятности к отрицательному числу, то

$$P_{\theta_0}(\eta_n < 0) \to 1.$$

■

10.5 Оценка максимального правдоподобия, состоятельность решения уравнения правдоподобия, обобщенный корень уравнения правдоподобия

В силу теоремы 2.2 естественно брать оценкой то значение θ, которое максимизирует $p(x,\theta)$ при данном x.

Определение 10.9. *Случайная величина $\hat{\theta}_n \in \Theta$ называется* ***оценкой максимального правдоподобия****, если*

$$p(x,\hat{\theta}_n) = \max_{\theta\in\Theta} p(x,\theta) \ \Leftrightarrow \ Ln(x,\hat{\theta}_n) = \max_{\theta\in\Theta} Ln(x,\theta)$$
$$\text{т.о. ОМП } \hat{\theta}_n = arg \max_{\theta\in\Theta} Ln(x,\theta)$$

Определение 10.10. *Если θ - интервал, а $Ln(x,\theta)$ - гладкая по θ функция, то θ удовлетворяет* ***уравнению правдоподобия****:*

$$\frac{\partial}{\partial\theta} Ln(x,\theta) = 0 \tag{2}$$

Теорема 10.5 (**о состоятельности решения уравнения правдоподобия**). *Пусть выполнено условие (A), пусть $\forall x \in N_f$ существует непрерывная производная $f'_\theta(x,\theta)$ $\Rightarrow$ уравнение правдоподобия (2) с вероятностью, стремящейся* *к единице при* $n \to \infty$*, имеет решение, принадлежащее* Θ*. При этом среди всех таких решений* (2) *есть такой корень* $\hat{\theta}_n$*, что он явялется состоятельной оценкой* θ_0*.*

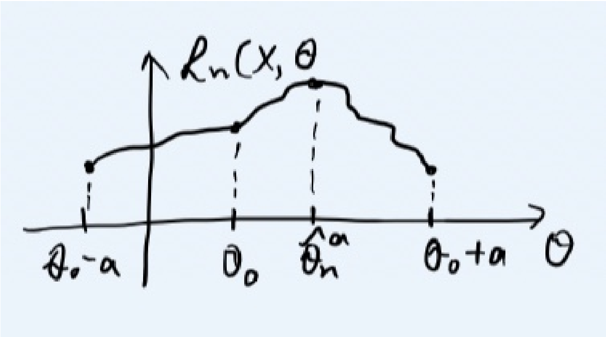
Доказательство. Пусть $S_n = \{w\}$, при которых уравнение (2) имеет решение для $\theta \in \Theta$. Теорема 2.3 утверждает:

1. $P_{\theta_0}(S_n) \to 1$
2. существует такое решение $\hat{\theta}_n \in \Theta$, что $\forall \varepsilon > 0$ $P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \varepsilon, S_n) \xrightarrow[n\to\infty]{} 1$

1. выберем малое $a > 0$ так, что $(\theta_0 - a, \theta_0 + a) \subseteq \Theta$. Тогда:

$$S_n^a = \{w\colon Ln(x,\theta_0) > Ln(x,\theta_0 - a), Ln(x,\theta_0) > Ln(x,\theta_0 + a)\}$$

   В силу теоремы 2.2 $P_{\theta_0}(S_n^a) \to 1$. При $w \in S_n^a$ функция $Ln(x,\theta)$ имеет локальный максимум $\hat{\theta}_n^a$ в интервале $(\theta_0 - a, \theta_0 + a)$, значит $\frac{\partial}{\partial\theta}Ln(x,\hat{\theta}_n^a) = 0 \Rightarrow P_{\theta_0}(S_n) \geq P_{\theta_0}(S_n^a) \to 1$, т.к. $S_n^a \subseteq S_n \Rightarrow$ доказали пункт 1.



2. $\forall n$ при $w \in S_n$ может существовать целое множество корней $\{\theta_n^*\}$. Выберем в этом множестве корень $\hat{\theta}_n$, ближайший к θ_0 (инфимум в множестве корней). Это можно сделать, т.к. функция $\frac{\partial}{\partial\theta}Ln(x,\theta)$ непрерывна по θ, и последовательность корней есть тоже корень. Этот корень $\hat{\theta}_n$ и есть состоятельная оценка θ, покажем это.
   Т.к. $S_n^\varepsilon \subseteq S_n$, $(w : |\hat{\theta}_n^\varepsilon - \theta_0| < \varepsilon) \subseteq (w : |\hat{\theta}_n - \theta_0| < \varepsilon)$, то для любого малого $\varepsilon > 0$:

$$P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \varepsilon, S_n) \geq P_{\theta_0}(|\hat{\theta}_n^\varepsilon - \theta_0| < \varepsilon, S_n^\varepsilon) \qquad (3)$$

$$\text{Но } P_{\theta_0}(|\hat{\theta}_n^\varepsilon - \theta_0| < \varepsilon, S_n^\varepsilon) = P_{\theta_0}(S_n^\varepsilon) \to 1 \Rightarrow$$

$$\Rightarrow \text{ в силу (3) } P_{\theta_0}(|\hat{\theta}_n - \theta_0| < \varepsilon, S_n) \to 1 \Rightarrow \text{ пункт (2) доказан.}$$

■

Замечание. *Определим следующую величину:*

$$\theta_n^* = \begin{cases} \text{сост. корень } \hat{\theta}_n \text{ ур-ния правдоподобия, если он} \exists \\ \theta', \ \theta' \in \Theta, \ \text{в противном случае.} \end{cases} \tag{4}$$

Тогда случайная величина θ_n^* всегда определена и $\theta_n^* \xrightarrow{P} \theta_0$, т.к.

$$P(|\theta_n^* - \theta_0| < \varepsilon) = P(|\hat{\theta}_n - \theta_0| < \varepsilon, S_n) + P(|\theta' - \theta| < \varepsilon, \overline{S_n}) \to 1.$$

Ясно, что $\frac{\partial}{\partial\theta} Ln(x, \theta_n^*) = o(1)$, т.к. производная отлична от нуля только на $\overline{S_n}$.

Определение 10.11. *Будем называть θ_n^* **обощенным корнем уравнения правдоподобия** (2).*

Теорема 10.6 (**об асимптотической эффективности состоятельного решения**). *Пусть $X = (X_1, \ldots, X_n)$, $\{X_i\}$ - н.о.р.с.в. Удовлетворяются условия теоремы Бахадура, в которых условия (iii) и (vi) заменены на предположение о третьей, а не второй производной, т.е. $|\frac{\partial^3}{\partial\theta^3} \ln f(x, \theta)| < M(x)\ \forall x \in N_f, \theta \in \Theta$ и $E_{\theta_0} M(x_1) < \infty$. Тогда, если θ_n^* - обощенный состоятельный корень уравнения правдоподобия из (4), то $\sqrt{n}(\theta_n^* - \theta_0) \xrightarrow{d} N(0, \frac{1}{i(\theta)})$, т.е. θ_n^* - асимптотически эффективная оценка.*

Доказательство. Будем обозначать $\frac{\partial}{\partial\theta} Ln(x, \theta), \frac{\partial^2}{\partial\theta^2} Ln(x, \theta), \ldots$ через $Ln'(\theta), Ln^{(2)}(\theta), \ldots$

Для фиксированного X в силу формулы Тейлора и замечания из предыдущей лекции:

$$\overline{o_p}(1) = Ln'(\theta_n^*) = Ln'(\theta_0) + Ln^{(2)}(\theta_0)(\theta_n^* - \theta_0) + \frac{1}{2} Ln^{(3)}(\tilde{\theta_n})(\theta_n^* - \theta_0)^2, \ \tilde{\theta_n} \in (\theta_0, \theta_n^*)$$

После преобразований получаем выражение:

$$\sqrt{n}(\theta_n^* - \theta_0) = -\frac{n^{-\frac{1}{2}} Ln'(\theta_0) + \overline{o_p}(1)}{n^{-1} Ln^{(2)}(\theta_0) + \frac{1}{2n} Ln^{(3)}(\tilde{\theta_n})(\theta_n^* - \theta_0)} \tag{5}$$

Рассмотрим числитель (5):

$$n^{-\frac{1}{2}} Ln'(\theta_0) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{f'_\theta(x_i, \theta_0)}{f(x_i, \theta_0)} \xrightarrow{d} \xi \sim N(0, i(\theta_0)) \tag{6}$$

Действительно:

$$E_{\theta_0}\frac{f'_\theta(x_1,\theta_0)}{f(x_1,\theta_0)} = \int\limits_{N_f}\frac{f'_\theta(x,\theta_0)}{f(x,\theta_0)}f(x,\theta_0)\nu(dx) = 0$$

где N_fноситель плотности вероятности f, $f \geq 0$

$$D_{\theta_0}\frac{f'_\theta(x_1,\theta_0)}{f(x_1,\theta_0)} = E_{\theta_0}(\frac{\partial}{\partial\theta}\ln f(x_1,\theta_0))^2 = i(\theta_0)$$

Получаем, что величины $\{\frac{f'_\theta(x_i,\theta_0)}{f(x_i,\theta_0)}, i = 1,2,\ldots,n\}$ - н.о.р. и соотношение (6) следует из ЦПТ. Т.о. в силу леммы Слуцкого числитель (5) сходится по вероятности к $N(0,i(\theta_0))$.

Рассмотрим знаменатель (5):

$$n^{-1}Ln^{(2)}(\theta_0) = n^{-1}\sum_{i=1}^{n}\left[\frac{f_\theta^{(2)}(x_i,\theta_0)}{f(x_i,\theta_0)} - (\frac{f'_\theta(x_i,\theta_0)}{f(x_i,\theta_0)})^2\right] \xrightarrow{P} -i(\theta_0) \qquad (7)$$

$\dfrac{f_\theta^{(2)}(x_i,\theta_0)}{f(x_i,\theta_0)} - (\dfrac{f'_\theta(x_i,\theta_0)}{f(x_i,\theta_0)})^2$ - производная от 1-ой производной по правилу Лейбница

Действительно, в силу ЗБЧ - хотелось бы применить слабый ЗБЧ в форме Чебышева, но там нужна дисперсия, поэтому воспользуемся ЗБЧ в форме Колмогорова, из которого получим сходимость почти наверно:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{f_\theta^{(2)(x,\theta_0)}}{f(x_i,\theta_0)} \xrightarrow[\text{если } \exists \text{ МО}]{P} E_{\theta_0}\frac{f_\theta^{(2)}(x_1,\theta_0)}{f(x_1,\theta_0)} = \int\limits_{N_f}\frac{f_\theta^{(2)}(x,\theta_0)}{f(x,\theta_0)}f(x,\theta_0)\nu(dx) = 0$$

$$\left(\frac{f_\theta^{(2)(x,\theta_0)}}{f(x_i,\theta_0)} \text{ - борелевские функции, н.о.р.с.в.}\right)$$

$$\frac{1}{n}\sum_{i=1}^{n}(\frac{f'_\theta(x_i,\theta_0)}{f(x_i,\theta_0)})^2 \xrightarrow{P} E_{\theta_0}(\frac{\partial}{\partial\theta}\ln f(x_1,\theta_0))^2 = i(\theta_0)$$

Применяя лемму Слуцкого, получаем (7) (сходимость к $-i(\theta_0)$).

Рассмотрим второе слагаемое в знаменателе (5):

$$\left|\frac{1}{2n}Ln^{(3)}(\tilde{\theta_n})(\theta_n^* - \theta_0)\right| \le \frac{1}{n}|\theta_n^* - \theta_0| \cdot \frac{1}{n}\sum_{i=1}^{n} M(x_i)$$

$$\left(Ln^{(3)}(\tilde{\theta_n}) \le \sum_{i=1}^{n} M(x_i) \text{ (из условия)}\right)$$

$$|\theta_n^* - \theta_0| \xrightarrow{P} 0, \; \frac{1}{n}\sum_{i=1}^{n} M(x_i) \xrightarrow{P} M \text{ - число (в силу ЗБЧ)}$$

Тогда в силу леммы Слуцкого:

$$\left|\frac{1}{2n}Ln^{(3)}(\tilde{\theta_n})(\theta_n^* - \theta_0)\right| \xrightarrow{P} 0 \tag{8}$$

В силу (7) и (8) и леммы Слуцкого знаменатель в (5) сходится по вероятности к $-i(\theta_0)$.
Значит, по лемме Слуцкого вся дробь (5) сходится по распределению к $\frac{1}{i(\theta_0)}\xi \sim N(0, \frac{i(\theta_0)}{i^2(\theta_0)}) = N(0, \frac{1}{i(\theta_0)})$, где ξ - гауссовская случайная величина. ∎

10.6 ОМП для векторного параметра

Пусть $X = (X_1, \ldots, X_n)$ - н.о.р., $X_1 \sim f(x, \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^k$, Θ - открытое множество.

Логарифмическое правдоподобие имеет вид:

$$Ln(x, \theta) = \sum_{i=1}^{n} \ln f(x_i, \theta)$$

Система уравнений правдоподобия имеет вид:

$$\frac{\partial Ln(x, \theta)}{\partial \theta_i} = 0, \; i = 1, 2, \ldots, k \tag{9}$$

При условиях регулярности, похожих на условия теоремы 3.1, показывается:

1. с вероятностью, стремящейся к 1 при $n \to \infty$, система уравнений (9) имеет такое решение $\hat{\theta}_n \in \Theta$, что $\hat{\theta}_n$ сходится к истинному значению θ_0

2. соответствующая оценка θ_n^* асимптотически нормальна, а именно:

$$\sqrt{n}(\theta_n^* - \theta_0) \xrightarrow[n\to\infty]{d} N(o, I^{-1}(\theta_0))$$

$$I(\theta) > 0 \text{ - } \textbf{матрица информации Фишера, } I(\theta) = (I_{ij}(\theta))$$

$$I_{ij}(\theta) = E_\theta\left(\frac{\partial \ln f(x,\theta)}{\partial\theta_i} \cdot \frac{\partial \ln f(x,\theta)}{\partial\theta_j}\right)$$

10.7 АЭО для интервала

Пример 10.2. *Пусть $X = (X_1, \ldots, X_n)$, где $\{X_i\}$ - н.о.р., $X_1 \sim N(\theta, \sigma^2)$, $a < \theta < b$, т.е. $\Theta = (a, b)$. a и b - известные конечные числа, θ неизвестно, σ^2 известно. Необходимо построить АЭО.*

Решение. Построим АЭО θ_n^* для θ.

$$p(x,\theta) = (\frac{1}{\sqrt{2\pi}\sigma})^n e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(x_i-\theta)^2} \text{ - плотность вероятности гауссовской сл. в.}$$

$$Ln(x,\theta) = \ln(\frac{1}{\sqrt{2\pi}\sigma})^n - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2 \text{ - парабола ветвями вниз}$$

Уравнение правдоподобия имеет вид:

$$\frac{\partial Ln(x,\theta)}{\partial\theta} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta) = 0$$

Решение существует и единственно - это $\overline{X}$, причем в точке $\theta = \overline{X}$ функция $Ln(x,\theta)$ достигает максимума, т.к.:

$$\frac{\partial^2 Ln(x,\overline{X})}{\partial\theta^2} = -\frac{1}{\sigma^2} < 0$$

Т.о., если $a < \overline{X} < b$, то ОМП существует с вероятностью, стремящейся к 1 и равна $\overline{X}$, в противном случае ОМП не существует.

Если положить:

$$\theta_n^* = \begin{cases} \overline{X},\ a < \overline{X} < b \\ \frac{a+b}{2} \text{ (любое число)},\ \overline{X} \notin (a,b) \end{cases} \tag{9}$$

то в силу теоремы 3.1 (условия выполнены) θ_n^* - АЭО, т.е.:

$$\sqrt{n}(\theta_n^* - \theta_0) \xrightarrow{d} N(0,\sigma^2),\ i(\theta) = \frac{1}{\sigma^2} \tag{10}$$

(также (10) можно проверить непосредственно) ■

Замечание. *Если $\theta \in [a,b]$, то по тоереме Вейерштрасса непрерывная функция на компакте достигает максимума и минимума. Тогда:*

$$\theta_n^* = \begin{cases} \overline{X},\ \overline{X} \in (a,b) \\ a,\ \overline{X} < a \\ b,\ \overline{X} > b \end{cases}$$

В $\theta = a$ или $\theta = b$ нет асимптотической гауссовости, поэтому надо рассматривать только открытые множества.

11

Проверка статистических гипотез

Пусть $X = (X_1, \dots, X_n)$ (н.о.р.) имеет плотность вер-ти по мере μ, $\theta \in \Theta \subseteq \mathbb{R}^k$.

Определение 11.1. *Предположение вида $H_0 : \theta \in \Theta_0$, где $\Theta_0 \subseteq \Theta$, называется **параметрической гипотезой**.*

Определение 11.2. ***Альтернатива** $H_1 : \theta \in \Theta_1, \Theta_1 = \Theta \setminus \Theta_0$.*

Определение 11.3. *Если $\Theta_0(\Theta_1)$ состоит из одной точки, то гипотеза H_0 (альтернатива H_1) называется **простой**. В противном случае $H_0(H_1)$ - **сложная**.*

Постановка задачи

Необходимо построить правило (**статистический критерий**), который позволяет заключить, согласуется ли X с H_0 или нет.

Правило

Выберем в множестве значений x вектора X (либо $x = \mathbb{R}^n$, либо $x = N_p \subseteq \mathbb{R}^n$) подмножество S.

- если $X \in S$, то H_0 отвергается и принимается H_1
- если $X \in \overline{S} = X \setminus S$, то H_0 принимается

Определение 11.4. *Множество S называется **критическим множеством** или **критерием**, $\overline{S}$ - **область принятия гипотезы**.*

Возможные ошибки

Определение 11.5. ***Ошибка 1-го рода** - принять H_1, когда верна H_0. Вероятность ошибки 1-го рода: $\alpha = P(H_1|H_0)$.*

Определение 11.6. ***Ошибка 2-го рода** - принять H_0, когда верна H_1. Вероятность ошибки 2-го рода: $\beta = P(H_0|H_1)$.*

Определение 11.7. ***Мощностью критерия*** *S называется функция $W(S,\theta) = W(\theta) := P_\theta(X \in S)$, т.е. мощность есть вероятность отвергнуть H_0, когда параметр равен θ.*

$$\alpha = \alpha(\theta) = W(\theta), \theta \in \Theta_0$$
$$\beta = \beta(\theta) = 1 - W(\theta), \theta \in \Theta_1$$

Замечание. *Обычно H_0 более важна $\Rightarrow$ рассматривают критерии такие, что:*

$$\alpha(\theta) = W(\theta) = P_\theta(X \in S) \le \alpha \; \forall \theta \in \Theta_0$$

Определение 11.8. *Число α называют* ***уровнем значимости*** *критерия и пишут S_α.*

Определение 11.9. *Если критерий $S_\alpha^* \in \{S_\alpha\}$ и $\forall \theta \in \Theta_1$ и $\forall S_\alpha$ $W(S_\alpha*, \theta) \ge W(S_\alpha, \theta)$, то критерий S_α^* называется* ***РНМ-критерием*** *(равномерно наиболее мощным).*

Если $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$, т.е. H_0 и H_1 - простые, то задача отыскания РНМ-критерия заданного уровня α имеет вид:

$$P_{\theta_0}(X \in S_\alpha^*) \le \alpha$$
$$P_{\theta_1}(X \in S_\alpha^*) \ge P_{\theta_1}(X \in S_\alpha) \; \forall S_\alpha$$

11.1 Лемма Неймана-Пирсона

Положим для краткости:

$$p_0(x) := p(x,\theta_0), p_1(x) = p(x,\theta_1), \; E_0 = E_{\theta_0}, \; E_1 = E_{\theta_1}$$

Введем множество: $S(\lambda) = \{x \colon p_1(x) - \lambda p_0(x) > 0\}$, $\lambda > 0$.

Теорема 11.1 (**лемма Неймана-Пирсона**). *Пусть для некоторого $\lambda > 0$ и критерия R выполнено:*

(a) $P_0(X \in R) \le P_0(X \in S(\lambda))$

Тогда:

(b) $P_1(X \in R) \le P_1(X \in S(\lambda))$
(c) $P_1(X \in S(\lambda)) \ge P_0(X \in S(\lambda))$

Замечание. $x \in S(\lambda) \ \Leftrightarrow \ \frac{p_1(x)}{p_0(x)} > \lambda$

Определение 11.10. *Т.к. $p_1(x)$ и $p_0(x)$ - правдоподобие, то критерий $S(\lambda)$ называется* ***критерием отношения правдоподобия Неймана-Пирсона****.*

Замечание. *Утверждение (c) для $S(\lambda)$ означает, что*

$$P(H_1|H_1) \geq P(H_1|H_0) \ \Leftrightarrow \ W(S(\lambda), \theta_1) \geq W(S(\lambda), \theta_0)$$

Это свойства называется ***несмещенностью*** *критерия $S(\lambda)$.*

Доказательство. Для краткости обозначим $S(\lambda) = S$.

Пусть $I_R(x) = \begin{cases} 1, x \in R \\ 0, x \notin R \end{cases}$ Тогда:

$$(a) \ \Leftrightarrow \ E_0 I_R(x) \leq E_0 I_S(x) \tag{1}$$

Докажем (b). Верно неравенство:

$$I_R(x)[p_1(x) - \lambda p_0(x)] \leq I_S(x)[p_1(x) - \lambda p_0(x)] \tag{2}$$

Действительно:

- если $p_1(x) - \lambda p_0(x) > 0$, то $I_S(x) = 1$ и (2) очевидно
- если $p_1(x) - \lambda p_0(x) \leq 0$, то правая часть (2) есть ноль, а левая меньше либо равна нуля

Итак, (2) верно. Интегрируем (2) по $x \in \mathbb{R}^n$, получаем:

$$E_1 I_R(x) - \lambda E_0 I_R(x) \leq E_1 I_S(x) - \lambda E_0 I_S(x)$$

$$E_1 I_S(x) - E_1 I_R(x) \geq \lambda [E_0 I_S(x) - E_0 I_R(x)] \tag{3}$$

$$(E_0 I_S(x) - E_0 I_R(x) \geq 0)$$

В силу (3) и (1) и условия $\lambda > 0$ получаем $E_1 I_S(x) \geq E_1 I_R(x)$, т.е. (b) доказано.

Докажем (c).

1) Пусть $\lambda \geq 1$. Из определения S: $p_1(x) > p_0(x) \ \forall x \in S \ \Rightarrow$

$$\Rightarrow \ P_0(X \in S) = \int\limits_{\mathbb{R}^n} I_S(x) p_0(x) \mu(dx) \leq \int\limits_{\mathbb{R}^n} I_S(x) p_1(x) \mu(dx) = P_1(X \in S)$$

т.е. $P(H_1|H_0) \leq P(H_1|H_1)$

2) Пусть $\lambda < 1$, $p_1(x) < p_0(x)\ \forall x \in \overline{S}\ \Rightarrow$

$$\Rightarrow P_1(X \in \overline{S}) = \int\limits_{\mathbb{R}^n} I_{\overline{S}}(x) p_1(x)\mu(dx) \leq \int\limits_{\mathbb{R}^n} I_{\overline{S}}(x) p_0(x)\mu(dx) = P_0(X \in \overline{S})$$

$$\text{т.е. } 1 - P_1(X \in S) \leq 1 - P_0(x \in S)\ \Rightarrow\ P_1(X \in S) \geq P_0(X \in S)$$

■

11.2 Пример построения НМ-критерия

Пример 11.1. $X = (X_1, \dots, X_n)$, $\{X_i\}$ - *н.о.р.,* $X_1 \sim N(\theta, \sigma^2)$, *дисперсия* σ^2 *известна. Построим НМ-критерий для проверки:* $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1, \theta_1 > \theta_0 \end{cases}$

Уровень значимости возьмем α.

Доказательство.

1) Имеем:

$$p_0(x) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i - \theta_0)^2}$$

$$p_1(x) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i - \theta_1)^2}$$

$$S(\lambda) = \{x\colon p_1(x) - \lambda p_0(x) > 0\}\ \Leftrightarrow$$

$$\Leftrightarrow\ exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(X_i - \theta_1)^2 - (X_i - \theta_0)^2]\} > \lambda\ \Leftrightarrow$$

$$\Leftrightarrow\ \sum_{i=1}^{n}[(X_i - \theta_1)^2 - (X_i - \theta_0)^2] < \lambda_1,\ \lambda_1 = -2\sigma^2 \ln\lambda\ \Leftrightarrow$$

$$\Leftrightarrow\ (\theta_0 - \theta_1)\sum_{i=1}^{n} X_i < \lambda_2\ \Leftrightarrow$$

$$\Leftrightarrow\ \sum_{i=1}^{n} X_i > \tilde{\lambda},\ \tilde{\lambda} = \tilde{\lambda}(\lambda, n, \sigma^2, \theta_0, \theta_1)$$

Итак:

$$S(\lambda) = \left\{x\colon \sum_{i=1}^{n} X_i > \tilde{\lambda}\right\} \text{ при некотором } \tilde{\lambda}$$

2) Определим $\tilde{\lambda} = \tilde{\lambda_\alpha}$ из уравнения:

$$\alpha = P_{\theta_0}(X \in S(\tilde{\lambda_\alpha})) = P_{\theta_0}(\sum_{i=1}^n X_i > \tilde{\lambda_\alpha})$$

Тогда:

$$\alpha = P_{\theta_0}(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^n (X_i - \theta_0) > \frac{\tilde{\lambda_\alpha} - n\theta_0}{\sqrt{n}\sigma}) = 1 - \Phi(\frac{\tilde{\lambda_\alpha - n\theta_0}}{\sqrt{n}\sigma})$$

т.к. $\dfrac{\sum\limits_{i=1}^n (X_i - \theta_0)}{\sqrt{n}\sigma} \underset{H_0}{\sim} N(0,1)$. Значит, $\Phi(\frac{\tilde{\lambda_\alpha - n\theta_0}}{\sqrt{n}\sigma}) = 1 - \alpha$, $\frac{\tilde{\lambda_\alpha - n\theta_0}}{\sqrt{n}\sigma} = \xi_{1-\alpha}$

$\xi_{1-\alpha}$ - квантиль норм. закона уровня α

$\Phi(\dots)$ - функция Лампласа

Итак:

$$\tilde{\lambda_\alpha} = n\theta_0 + \sqrt{n}\sigma\xi_{1-\alpha}$$

3) Положим:

$$S_\alpha^* = \left\{x \colon \sum_{i=1}^n X_i > \tilde{\lambda_\alpha}\right\} \Rightarrow \begin{cases} P_{\theta_0}(X \in S_\alpha^*) = \alpha \\ \forall S_\alpha \; P_{\theta_0}(X \in S_\alpha) \le \alpha = P_{\theta_0}(X \in S_\alpha^*) \end{cases}$$

Значит выполнено условие (a) леммы Неймана-Пирсона, тогда в силу пункта (b) этой леммы:

$$P_{\theta_1}(X \in S_\alpha) \le P_{\theta_1}(X \in S_\alpha^*)$$

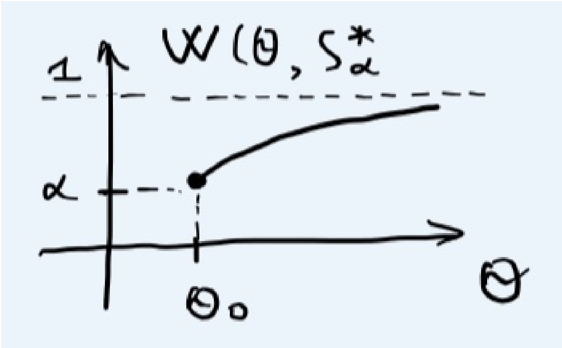
Таким образом, S_α^* - НМ-критерий.

Т.к. S_α^* не зависит от θ_1, то S_α^* - РНМ-критерий для $\begin{cases} H_0 : \theta = \theta_1 \\ H_1^+ : \theta > \theta_1 \end{cases}$

Мощность критерия S_α^* для H_0 при альтернативе H_1^+:

$$W(\theta, S_\alpha^*) = P_\theta(\sum_{i=1}^n X_i > n\theta_0 + \sqrt{n}\sigma\xi_{1-\alpha}) =$$

$$= P_{\theta_0}\left(\frac{\sum\limits_{i=1}^n (X_i - \theta_0)^2}{\sqrt{n}\sigma} > \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} + \xi_{1-\alpha}\right) = 1 - \Phi(\xi_{1-\alpha} + \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma})$$



■

11.3 Связь между доверительным оцениванием и проверкой гипотез

Определение 11.11. *Случайное подмножество $\Theta^* = \Theta^*(X, \alpha) \subseteq \Theta$ называется* ***доверительным интервалом*** *уровня $1 - \alpha, 0 < \alpha < 1$, если*

$$P_\theta(\theta \in \Theta^*(X, \alpha)) \geq 1 - \alpha \; \forall \theta \in \Theta$$

Теорема 11.2. *Докажем два пункта:*

1. *пусть $\forall \theta_0 \in \Theta$ гипотеза $H_0 : \theta = \theta_0$ при альтернативе $H_1 : \theta \neq \theta_0$ имеет $S_\alpha(\theta_0)$ критерий уровня α. Пусть $\Theta^*(X, \alpha) = \big\{\theta \colon x \in \overline{S_\alpha}(\theta_0)\big\}$. Тогда $\Theta^*(X, \alpha)$ - доверительное множество уровня $1 - \alpha$.*
2. *Если $\Theta^*(X, \alpha)$ - доверительное множество уровня $1 - \alpha$, то $\overline{S_\alpha}(\theta_0) = \{x \colon \theta_0 \in \Theta^*(X, \alpha)\}$ - область принятия гипотезы H_0.*

Замечание. *Пункт 2) означает, что если θ_0 попало в доверительное множество $\Theta^*(X, \alpha)$, то H_0 надо принимать.*

Доказательство.

1. $P_\theta(\theta \in \Theta^*(X, \alpha)) = P_\theta(x \in \overline{S_\alpha}(\theta)) = 1 - P_\theta(x \in S_\alpha(\theta)) \geq 1 - \alpha \; \forall \theta \in \Theta$, т.к. $P_\theta(x \in S_\alpha(\theta)) \leq \alpha$.
2. $P_{\theta_0}(x \in S_\alpha(\theta_0)) = 1 - P_{\theta_0}(x \in \overline{S_\alpha}(\theta_0)) = 1 - P_{\theta_0}(\theta_0 \in \Theta^*(X, \alpha)) \leq 1 - (1 - \alpha) = \alpha$, т.к. $P_{\theta_0}(\theta_0 \in \Theta^*(X, \alpha)) \geq 1 - \alpha$, т.е. $S_\alpha(\theta_0)$ - критерий уровня α.

■

Пример 11.2. *Пусть $X = (X_1, \ldots, X_n)$, $\{X_i\}$ - н.о.р., $X_1 \sim N(\theta, \sigma^2)$, $\theta \in \mathbb{R}^1$. Построим критерий для $H_0 : \theta = \theta_0$ против $H_1 : \theta \neq \theta_0$. Уровень значимости будет α, $0 < \alpha < 1$.*

Доказательство. Построим доверительное множество для θ уровня $1-\alpha$. Пусть $\overline{X} = n^{-1}\sum\limits_{i=1}^{n} X_i$ - оптимальная оценка θ, тогда:

$$\frac{n^{\frac{1}{2}}(\overline{X}-\theta)}{\sigma} \sim N(0,1), \quad P_\theta\left(\left|\frac{n^{\frac{1}{2}}(\overline{X}-\theta)}{\sigma}\right| < \xi_{1-\frac{\alpha}{2}}\right) = 1-\alpha, \ \Phi(\xi_{1-\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}$$

$$\text{т.е. } \Theta^*(X,\alpha) = \left\{\theta\colon \left|\frac{n^{\frac{1}{2}}(\overline{X}-\theta)}{\sigma}\right| < \xi_{1-\frac{\alpha}{2}}\right\}$$

В силу замечания к теореме 11.2:

$$S_\alpha(\theta_0) = \left\{x\colon \left|\frac{n^{\frac{1}{2}}(\overline{X}-\theta)}{\sigma}\right| \geq \xi_{1-\frac{\alpha}{2}}\right\}$$

$S_\alpha(\theta_0)$ - есть критическое множество для H_0

$$\text{мощность } W(\theta) = P_\theta(x \in S_\alpha(\theta_0)) = P_\theta\left(\left|\frac{n^{\frac{1}{2}}(\overline{X}-\theta)}{\sigma}\right| \geq \xi_{1-\frac{\alpha}{2}}\right) =$$

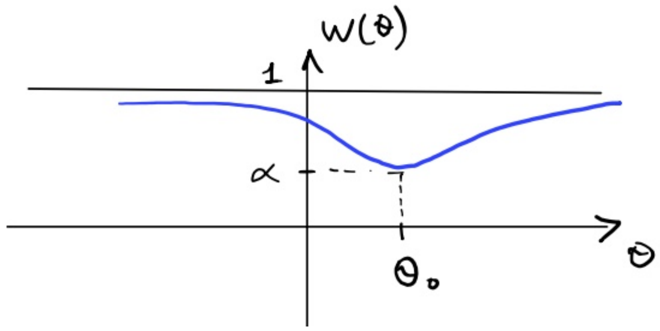
$$= 1 - P_\theta\left(\left|\frac{n^{\frac{1}{2}}(\overline{X}-\theta)}{\sigma}\right| < \xi_{1-\frac{\alpha}{2}}\right) =$$

$$= 1 - P_\theta\left(-\xi_{1-\frac{\alpha}{2}} + \frac{n^{\frac{1}{2}}(\theta_0-\theta)}{\sigma} \leq \frac{n^{\frac{1}{2}}(\overline{X}-\theta)}{\sigma} \leq \xi_{1-\frac{\alpha}{2}} + \frac{n^{\frac{1}{2}}(\theta_0-\theta)}{\sigma}\right) =$$

$$\left(\frac{n^{\frac{1}{2}}(\overline{X}-\theta)}{\sigma} \sim N(0,1)\right)$$

$$= 1 - \left[\Phi\left(\xi_{1-\frac{\alpha}{2}} + \frac{n^{\frac{1}{2}}(\theta_0-\theta)}{\sigma}\right) - \Phi\left(-\xi_{1-\frac{\alpha}{2}} + \frac{n^{\frac{1}{2}}(\theta_0-\theta)}{\sigma}\right)\right] =$$

$$= \Phi\left(\xi_{1-\frac{\alpha}{2}} + \frac{n^{\frac{1}{2}}(\theta_0-\theta)}{\sigma}\right) + \Phi\left(\xi_{1-\frac{\alpha}{2}} + \frac{n^{\frac{1}{2}}(\theta-\theta_0)}{\sigma}\right)$$



$$W'(\theta_0) = 0$$

$$W(\theta) \xrightarrow[n\to\infty]{} 1 \ \forall\theta \neq \theta_0$$

Т.е. $S_\alpha(\theta_0)$ состоятелен против любой фиксированной инициативы. ■

11.4 Критерий Фишера (F-критерий) в гауссовской линейной регрессии

Определение 11.12. *Если $\xi \sim N(0,1)$, $\eta_k \sim \chi^2(k)$, ξ и η_k независимы, константа $\mu \in \mathbb{R}^1$, то*

$$\text{сл.в. } t_k(\mu) \stackrel{\mathrm{d}}{=} \frac{\xi + \mu}{\sqrt{\frac{1}{k}\eta_k}} \sim S(k,\mu)$$

имеет ***нецентральное распределение Стьюдента*** *с k степенями свободы и параметром нецентральности μ.*

Определение 11.13. *Если $\xi_i \sim N(a_i,1), i = 1,2,\dots,k$ и $\{\xi_1,\dots,\xi_k\}$ независимы, $\triangle^2 = a_1^2 + \dots + a_k^2$, то*

$$\text{сл.в. } \eta_k(\triangle) \stackrel{\mathrm{d}}{=} \xi_1^2 + \dots + \xi_k^2 \sim \chi^2(k,\triangle^2)$$

имеет ***нецентральное распределение хи-квадрат*** *с k степенями свободы и параметром нецентральности $\triangle^2$.*

Определение 11.14. *Если $\eta_k \sim \chi^2(k,\triangle^2)$, $\nu_m \sim \chi^2(m)$, и η_k и ν_m независимы, то*

$$\text{сл.в. } f_{k,m}(\triangle) \stackrel{\mathrm{d}}{=} \frac{\frac{1}{k}\eta_k}{\frac{1}{m}\nu_m} \sim F(k,m,\triangle^2)$$

имеет ***нецентральное распределение Фишера*** *с (k,m) степенями свободы и параметром нецентральности $\triangle^2$.*

Лемма 11.1. *Докажем два пунтка:*

1. *распределение сл.в. $\eta_k \sim \chi^2(k,\triangle^2)$ зависит лишь от $\triangle$, но не от $a_1,\dots,a_k$, а именно*
$$\eta_k \stackrel{\mathrm{d}}{=} (z_1 + \triangle)^2 + z_2^2 + \dots + z_k^2$$
где $\{z_1,\dots,z_k\}$ - н.о.р. $N(0,1)$ сл.в.
2. *если вектор $\xi \in \mathbb{R}^k$, $\xi \sim N(a,\Sigma)$, $\Sigma > 0$, то*
$$\xi^T\Sigma^{-1}\xi \sim \chi^2(k,\triangle^2),\ \triangle^2 = a^T\Sigma^{-1}a$$

Доказательство.

1. По определению:

$$\eta_k = \eta_k(\triangle) \overset{\mathrm{d}}{=} \xi_1^2 + \dots + \xi_k^2$$
$$\{\xi_1, \dots, \xi_k\} \text{ - независимые } N(a_i, 1) \text{ сл.в.}$$

Пусть $\xi = (\xi_1, \dots, \xi_k)^T$.

Ортогональная матрица $C = \begin{pmatrix} \frac{a_1}{\triangle} & \cdots & \frac{a_k}{\triangle} \\ \dots & \dots & \dots \end{pmatrix}$, $\nu = C\xi$.

Тогда $\eta_k \overset{\mathrm{d}}{=} |\xi|^2 = |\nu|^2$, т.к. C - ортогональная, но

$$\nu = C \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} + C\overset{o}{\xi} = \begin{pmatrix} \triangle \\ 0 \\ \vdots \\ 0 \end{pmatrix} + Z$$

где $\overset{o}{\xi} = \xi - E\xi,\ Z = C\overset{o}{\xi} \sim N(0, E_k),$ таким образом:

$$\eta_k \overset{\mathrm{d}}{=} |\nu|^2 = (z_1 + \triangle)^2 + z_2^2 + \dots + z_k^2$$

2. $\xi^T \Sigma^{-1} \xi = |\Sigma^{-\frac{1}{2}}\xi|^2,$ причем $\Sigma^{-\frac{1}{2}}\xi \sim N(\Sigma^{-\frac{1}{2}}a, E_k)$, тогда:

$$|\Sigma^{-\frac{1}{2}}\xi|^2 \sim \chi^2(k, \triangle^2),\ \triangle^2 = |\Sigma^{-\frac{1}{2}}a|^2 = a^T \Sigma^{-1} a$$

■

Лемма 11.2. *Случайная величина $t_k(\mu)$ обладает следующим свойством стохастической упорядоченности:*

(4) при $\mu_2 > \mu_1$ и при $\forall x \in \mathbb{R}^1$ $P(t_k(\mu_2) > x) > P(t_k(\mu_1) > x)$
(5) $P(\eta_k(\triangle_2) > x) > P(\eta_k(\triangle_1) > x),\ \triangle_2 > \triangle_1$
(6) $P(f_{k,m}(\triangle_2) > x) > P(f_{k,m}(\triangle_1) > x),\ \triangle_2 > \triangle_1$

Доказательство. Заметим, что если ξ и η - независимые сл. вел. и $E|\varphi(\xi, \eta)| < \infty$, φ – борелевская функция, то

$$E\varphi(\xi, \eta) = E\{(E\varphi(\xi, \eta))\,|_{y=\eta}\} \tag{7}$$

В силу (7):

$$P(t_k(\mu_2) > x) = P\left(\frac{\xi+\mu_2}{\sqrt{\frac{1}{k}\eta_k}} > x\right) = E\mathbb{I}\left(\xi > x\sqrt{\frac{1}{k}\eta_k} - \mu_2\right) =$$

$$= E\left\{1 - \mathbb{I}\left(\xi \leq x\sqrt{\frac{1}{k}\eta_k} - \mu_2\right)\right\} = 1 - E\left\{(E\mathbb{I}(\xi \leq y))\,|_{y=x\sqrt{\frac{1}{k}\eta_k}-\mu_2}\right\} =$$

$$= 1 - E\Phi\left(x\sqrt{\frac{1}{k}\eta_k} - \mu_2\right) > 1 - E\Phi\left(x\sqrt{\frac{1}{k}\eta_k} - \mu_1\right) = P\left(t_k(\mu_1) > x\right)$$

$$\left(\text{т.к. } E\Phi\left(x\sqrt{\frac{1}{k}\eta_k} - \mu_2\right) < E\Phi\left(x\sqrt{\frac{1}{k}\eta_k} - \mu_1\right) \text{ в силу возрастания } \Phi\right)$$

■

11.5 Построение доверительного множества для линейной гауссовской модели

Пусть $X = Zc + \varepsilon$, где $X = (X_1, \dots, X_n)^T$ - наблюдение.
Z - $(n \times p)$-матрица регрессора, $rkZ = p,\ p < n$.
$\varepsilon \sim N(0, \sigma^2 E_n),\ c = (c_1, \dots, c_p)^T$, c и σ^2 неизвестны.

Рассмотрим новый вектор $\beta = Ac$, A - $(k \times p)$-матрица, $rkA = k \leq p$, т.е. строки A линейно независимы. Построим для β доверительное множество уровня $1 - \alpha$.
Решение. Пусть $\hat{c_n}$ - о.н.к. для c (также оптимальная).
Пусть $\hat{S_n}^2$ - о.н.к. для σ^2. Пусть $\hat{\beta_n} = A\hat{c_n}$ - оптимальная оценка для β.
Т.к. $\hat{c_n} \sim N(c, \sigma^2(Z^TZ)^{-1})$, то $\hat{\beta_n} \sim N(Ac, \sigma^2 D)$, где $Ac = \beta,\ D = A(Z^TZ)^{-1}A^T$.

Заметим, что $D > 0$, т.к. для $\alpha \in \mathbb{R}^k,\ \alpha \neq 0,\ \alpha^T D\alpha = (A^T\alpha)^T(Z^TZ)^{-1}(A^T\alpha) > 0$, поскольку $(Z^TZ)^{-1} > c,\ A^T\alpha \neq 0$ при $rkA = k$ и $\alpha \neq 0$.

В силу пункта 2) леммы 11.1: $\frac{1}{\sigma^2}(\hat{\beta_n} - \beta)^T D^{-1}(\hat{\beta_n} - \beta) \sim \chi^2(k)$

Т.к. $\frac{(n-p)\hat{S}_n^{\,2}}{\sigma^2} \sim \chi^2(n-p)$, $\hat{\beta}_n$ и $\hat{S}_n^{\,2}$ независимо, то

$$f_{k,n-p}(X,\beta) := \frac{\frac{\frac{1}{k}(\hat{\beta}_n-\beta)D^{-1}(\hat{\beta}_n-\beta)}{\sigma^2}}{\frac{\frac{1}{n-p}(n-p)\hat{S}_n^{\,2}}{\sigma^2}} = \frac{(\hat{\beta}_n-\beta)D^{-1}(\hat{\beta}_n-\beta)}{k\hat{S}_n^{\,2}} \sim F(k,n-p) \Rightarrow$$

$$P_{\beta,\sigma^2}\left((\hat{\beta}_n-\beta)D^{-1}(\hat{\beta}_n-\beta) \leq k\hat{S}_n^{\,2} f_{1-\alpha}(k,n-p)\right) = 1-\alpha$$

$f_{1-\alpha}(k,n-p)$ - квантиль уровня $1-\alpha$ распределения $F(k,n-p)$

Доверительное множество для β уровня $1-\alpha$:

$$\Theta^*(X,\alpha) = \left\{\beta\colon (\hat{\beta}_n-\beta)D^{-1}(\hat{\beta}_n-\beta) < k\hat{S}_n^{\,2} f_{1-\alpha}(k,n-p)\right\}$$

(**доверительный эллипсоид**)

■

Рассмотрим теперь проверку гипотезы $H_0 : \beta = \beta_0$ против $H_1 : \beta \neq \beta_0$.

Определение 11.15. *H_0 называется **линейной гипотезой**, т.к. $\beta = Ac$ получается линейным преобразованием c.*

В силу замечания 11.3 H_0 надо принимать, если $\beta_0 \in \Theta^*(X,\alpha)$, т.е. область принятия H_0:

$$\overline{S_\alpha}(\beta_0) = \{x\colon f_{k,n-p}(x,\beta_0) \leq f_{1-\alpha}(k,n-p)\}$$

т.е. критическое множество (критерий уровня α):

$$S_\alpha(\beta_0) = \{x\colon f_{k,n-p}(x,\beta_0) > f_{1-\alpha}(k,n-p)\} \tag{7}$$

Определение 11.16. *Критерий (7) называется **критерием Фишера (F-критерие** а $f_{k,n-p}(x,\beta_0)$ - **статистикой F-критерия**.*

Рассмотрим поведение F-критерия при альтернативе H_1: при H_1 в силу пункта 2 леммы 11.1:

$$f_{k,n-p}(x,\beta_0) = \frac{\frac{\frac{1}{k}(\hat{\beta}_n-\beta)D^{-1}(\hat{\beta}_n-\beta)}{\sigma^2}}{\frac{\frac{1}{n-p}(n-p)\hat{S}_n^{\,2}}{\sigma^2}},\ \frac{(\hat{\beta}_n-\beta)D^{-1}(\hat{\beta}_n-\beta)}{\sigma^2} \sim \chi^2(k,\triangle^2)$$

$$\frac{(n-p)\hat{S}_n^{\,2}}{\sigma^2} \sim \chi^2(n-p),\ \text{тогда } f_{k,n-p}(x,\beta_0) \sim F(k,n-p,\triangle^2)$$

Параметр нецентральности: $\triangle^2 = \frac{1}{\sigma^2}(\beta - \beta_0)D^{-1}(\beta - \beta_0)$

Мощность F-критерия:

$$W(\beta, S_\alpha(\beta_0)) = P_{\beta,\sigma^2}(f_{k,n-p}(x, \beta_0) > f_{1-\alpha}(k, n-p)) = 1 - F_{k,n-p}(f_{1-\alpha}(k, n-p), \triangle^2)$$
$$\text{при } \beta = \beta_0 \;\; W(\beta_0, S_\alpha(\beta_0)) = \alpha$$

Свойства 11.1. *Свойства мощности:*

1. $\triangle = \triangle(\beta) > 0 = \triangle(\beta_0)$ *при* $\beta \neq \beta_0$, *тогда из соотношения (6) леммы 11.2:*

   $$P_{\beta,\sigma^2}(f_{k,n-p}(x, \beta_0) > f_{1-\alpha}(k, n-p)) > P_{\beta_0,\sigma^2}(f_{k,n-p}(x, \beta_0) > f_{1-\alpha}(k, n-p)) = \alpha$$

   т.е. при $\beta \neq \beta_0$ $P(H_1|H_1) > P(H_1|H_0)$,
   т.е. F-критерий ***несмещенный***
2. *мощность* $W(\beta, S_\alpha(\beta_0))$ *строго монотонна по* $\triangle$ *из соотношения (8)*

11.6 Пример определения порядка регрессии

Пусть $c^T = (c^T_{(1)}, c^T_{(2)})$, $c_{(1)}$ - m-вектор, $c_{(2)}$ - $p - m$-вектор, $1 \leq m \leq p$.
$H_0 : c_{(2)} = 0$ (т.е. порядок $\leq m$), $H_1 : c_{(2)} \neq 0$.

Пусть матрица $A = \begin{array}{c} \begin{array}{cccccc} \underline{1} & \underline{\cdots} & \underline{m} & \underline{m+1} & \underline{\cdots} & \underline{p} \end{array} \\ \begin{pmatrix} 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 1 \end{pmatrix} \end{array}$, тогда $Ac = c_{(2)}$ и
H_0 эквивалентна гипотезе $Ac = 0 = \beta_0 - p - m$-вектор.

Пусть $\hat{c_n}^T = (\hat{c_{(1)}}_n^T, \hat{c_{(2)}}_n^T)$, где $\hat{c_{(1)}}_n^T$ – m-вектор, $\hat{c_{(2)}}_n^T$ – $p - m$-вектор. Тогда $\hat{\beta_n} = A\hat{c_n} = \hat{c_{(2)}}_n$.

$$\text{Если } (Z^TZ)^{-1} = \begin{array}{r} {\scriptstyle 1\times m}| \\ {\scriptstyle (m+1)\times p}| \end{array} \begin{array}{c} \begin{array}{cc} \underline{1\times m} & \underline{(m+1)\times p} \end{array} \\ \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \end{array},$$
$$\text{то } D = A(Z^TZ)^{-1}A^T = B_{22}.$$

Тогда:

$$f_{p-m,n-p}(x, \beta_0 = 0) = \frac{\hat{c_{(2)}}_n^T B_{22}^{-1} \hat{c_{(2)}}_n}{(p-m)\hat{S_n}^2}$$

- при H_0 $f_{p-m,n-p}(x,0) \sim F(p-m, n-p)$
- H_0 отвергается, если $f_{p-m,n-p}(x,0) > f_{1-\alpha}(p-m, n-p)$, т.е.

$$S_\alpha(0) = \left\{ x \colon \frac{\hat{c_{(2)}}_n^T B_{22}^{-1} \hat{c_{(2)}}_n}{(p-m)\hat{S_n}^2} > f_{1-\alpha}(p-m, n-p) \right\} \tag{9}$$

- при H_1 $f_{p-m,n-p}(x,0) \sim F(p-m, n-p, \triangle^2)$, где

$$\triangle^2 = \frac{\hat{c_{(2)}}^T B_{22}^{-1} \hat{c_{(2)}}}{\sigma^2} \tag{10}$$

- критерий (9) несмещенный, т.е. $P(H_1|H_1) > P(H_1|H_0) = \alpha$. Его мощность:

$$W(c_{(2)}, S_\alpha(0)) = P_{c_{(2)},\sigma^2}(f_{p-m,n-p}(x,0) > f_{1-\alpha}(p-m, n-p)) =$$
$$= 1 - F_{p-m,n-p}(f_{1-\alpha}(p-m, n-p), \triangle^2)$$

т.е. мощность строго возрастает по $\triangle^2$

Параметр нецентральности $\triangle^2$ определен в (10)

11.7 Пример проверки однородности двух выборок

Пусть $X = (X_1, \ldots, X_m)$, $Y = (Y_1, \ldots, Y_m)$ - две независимые гауссовские выборки, т.е. $\{X_i\}$ - н.о.р., $X_1 \sim N(a, \sigma^2)$, $\{Y_j\}$ - н.о.р., $Y_1 \sim N(b, \sigma^2)$. Совокупности $\{X_i\}$ и $\{Y_j\}$ независимы, $m + n > 2$. Дисперсии $DX_1 = DY_1 = \sigma^2$ и неизвестны, средние a и b неизвестны.

Проверим гипотезу $H_0 : a = b$ против $H_1 : a \neq b$ по наблюдениям X и Y. Гипотеза H_0 называется **гипотезой однородности**.

При $DX_1 \neq DX_2$ описанная задача называется **проблемой Беренса-Фишера**.

Итак:

$$\begin{cases} X_i = a + \varepsilon_i, \ i = 1, 2, \ldots, m, \ \ \varepsilon_i := X_i - a \\ Y_j = b + \tilde{\varepsilon_j}, \ j = 1, 2, \ldots, n, \ \ \tilde{\varepsilon_j} := Y_j - b \end{cases}$$

Тогда $\varepsilon_1, \ldots, \varepsilon_m, \tilde{\varepsilon_1}, \ldots, \tilde{\varepsilon_n}$ - н.о.р. $N(0, \sigma^2)$ с.в.
Пусть $\tilde{X} = (X_1, \ldots, X_m, Y_1, \ldots, Y_n)^T$, $c = (a, b)^T$, $\varepsilon^T = (\varepsilon_1, \ldots, \varepsilon_m, \tilde{\varepsilon_1}, \ldots, \tilde{\varepsilon_n})^T$

$$Z = \begin{array}{r} 1| \\ \vdots| \\ m| \\ {\scriptstyle m+1}| \\ \vdots| \\ {\scriptstyle m+n}| \end{array} \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \Rightarrow \tilde{X} = Zc + \varepsilon \tag{11}$$

Соотношение (11) определяет гауссовскую линейную регрессию.

Положим $A = (1, -1) \ \Rightarrow \ Ac = a - b = \beta$.

$$\begin{aligned} H_0 &: Ac = a - b = \beta = 0 \ (= \beta_0) \\ H_1 &: Ac = a - b \neq 0 \ (\text{т.е. } \beta \neq 0) \end{aligned}$$

Тогда, о.н.к. для вектора c - это решение задачи.

$$\sum_{i=1}^{m}(X_i - a)^2 + \sum_{j=1}^{n}(Y_j - b)^2 \to \min_{a,b} \tag{12}$$

Задача (12) эквивалентна системе уравнений:

$$\begin{cases} -2\sum\limits_{i=1}^{m}(X_i - a) = 0 \\ -2\sum\limits_{j=1}^{n}(Y_j - b) = 0 \end{cases}$$

Решения этой системы $\hat{a_n} = \overline{X}$, $\hat{b_n} = \overline{Y}$ - оптимальные оценки a и b. $\hat{c_n} = (\overline{X}, \overline{Y})^T$ - оптимальная оценка c. Оптимальная оценка для σ^2:

$$\hat{S_{m+n}}^2 = \frac{1}{m+n-2}\left[\sum_{i=1}^{m}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2\right]$$

Тогда $\hat{\beta}_n = A\hat{c_n} = \overline{X} - \overline{Y}$.

$$Z^T Z = \begin{array}{c} \begin{array}{cccccc} \underline{1} & \underline{\ldots} & \underline{m} & \underline{m+1} & \underline{\ldots} & \underline{m+n} \end{array} \\ \begin{pmatrix} 1 & \ldots & 1 & 0 & \ldots & 0 \\ 0 & \ldots & 0 & 1 & \ldots & 1 \end{pmatrix} \end{array} \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} m & 0 \\ 0 & n \end{pmatrix}$$

$$D = A(Z^T Z)^{-1} A^T = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{m} & 0 \\ 0 & \frac{1}{n} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{n} + \frac{1}{m}$$

Значит:

$$f_{1,m+n-2}(X, \beta_0 = 0) = \frac{(\overline{X} - \overline{Y})^2}{\left(\frac{1}{n} + \frac{1}{m}\right) \hat{S}_{m+n}^{2}}$$

F-критерий для H_0 имеет вид:

$$S_\alpha(0) = \left\{x \in \mathbb{R}^{m+n} \colon f_{1,m+n-2}(X, 0) > f_{1-\alpha}(1, m+n-2)\right\}$$

При H_0 (т.е. при $a = b$) $f_{1,m+n-2}(X, 0) \sim F(1, m+n-2)$.
При H_1 (т.е. при $a \neq b$) $f_{1,m+n-2}(X, 0) \sim F(1, m+n-2, \triangle^2)$.
Параметр нецентральности:

$$\triangle^2 = \triangle^2(\beta) = \triangle^2(a-b) = \frac{(a-b)^2}{\sigma^2 \left(\frac{1}{m} + \frac{1}{n}\right)}$$

1. если $|a-b|$ возрастает, то мощность F-теста тоже возрастает
2. если $\sigma \to 0$ или $\frac{1}{m} + \frac{1}{n} \to 0$, то мощность возрастает

11.8 Проверка простой гипотезы в схеме Бернулли

Пусть проводится n независимых испытаний и в каждом испытании возможны m исходов, $m \geq 2$, $A_1, \ldots, A_m$ таких, что $A_i A_j = \emptyset$ при $i \neq j$, $\sum A_i = \Omega$. $P(A_j) = p_j > 0$, $\sum\limits_{j=1}^{m} p_j = 1$. Пусть $\nu = (\nu_1, \ldots, \nu_m)^T$, а ν_j - число появлений A_j в n опытах $X \Rightarrow \sum\limits_{j=1}^{m} \nu_j = n$.

По вектору наблюдений ν необходимо проверить гипотезу H_0:

$H_0 : p_j = p_j^0,\ j = 1, \ldots, m.$

Альтернатива $H_1 : p_j \neq p_j^0$ хотя бы при одном j.

Подчеркнем, что H_0 - простая гипотеза, т.к. она полностью определяет распределение вектора ν, а именно при H_0:

$$P(\nu_1 = k_1, \ldots, \nu_m = k_m) = \frac{n!}{k_1! \ldots k_m!}(p_1^0)^{k_1} \ldots (p_m^0)^{k_m} -$$

это полиномиальное распределение $\Pi(n, p_1^0, \ldots, p_m^0)$

Статистика хи-квадрат Пирсона для H_0 имеет вид:

$$\chi_n^2 := \sum_{j=1}^m \frac{(\nu_j - np_j^0)^2}{np_j^0}$$

Поведение при альтернативе: очевидно, что

$$\chi_n^2 = n \sum_{j=1}^m \frac{(\frac{\nu_j}{n} - p_j^0)^2}{p_j^0}$$

В силу теоремы Бернулли:

$$\frac{\nu_j}{n} \xrightarrow{P} p_j \ \Rightarrow\ \sum_{j=1}^m \frac{(\frac{\nu_j}{n} - p_j^0)^2}{p_j^0} \xrightarrow{P} \sum_{j=1}^m \frac{(p_j - p_j^0)^2}{p_j^0} > 0 \text{ при } H_1$$

(по теореме о наследовании сходимости по вероятности)

Значит, при H_1 $\chi_n^2 \xrightarrow[n\to\infty]{P} \infty$, поэтому большие значения χ_n^2 свидетельствуют против H_0. Но что такое большие значения?

Теорема 11.3 (**Пирсона**). *При H_0 и $n \to \infty$ $\chi_n^2 \xrightarrow{d} \chi^2(m-1)$.*

Правило 11.1. *Если $\chi_n^2 > \chi_{1-\alpha}(m-1)$, то H_0 отвергаем и принимаем H_1. Если $\chi_n^2 \leq \chi_{1-\alpha}(m-1)$, то принимаем H_0.*

Тогда:

$$P(H_1|H_0) = P(\chi_n^2 > \chi_{1-\alpha}(m-1)) \to \alpha$$

$$P(H_0|H_1) = P(\chi_n^2 \leq \chi_{1-\alpha}(m-1)|H_1) \xrightarrow[n\to\infty]{} 0$$

$$\text{т.е.} \begin{cases} P(H_0|H_0) \to 1-\alpha \\ P(H_1|H_1) \to 1 \end{cases}$$

Вероятность принять правильную гипотезу близка к единице.

11.9 Проверка простой гипотезы о виде функции распределения

Пусть $X = (X_1, \dots, X_n)$, $\{X_i\}$ - н.о.р., $X_1 \sim F(x)$.
$H_0 : F(x) = F_0(x)$, F_0 полностью известна.
$H_1 : F(x) = F_1(x)$ и $F_1(x) \neq F_0(x)$.

Разобьем носитель X_1 на непересекающиеся отрезки $\triangle_1, \dots, \triangle_m$ так, что $(m \geq 2)$ $X_1 \in \triangle_1 \bigcup \dots \bigcup \triangle_m$.

$$p_j^0 := P(x_1 \in \triangle_j | H_0) = \int\limits_{\triangle_j} dF_0(x) > 0 \; \forall j \;\Rightarrow\; \sum_{j=1}^{m} p_j^0 = 1$$

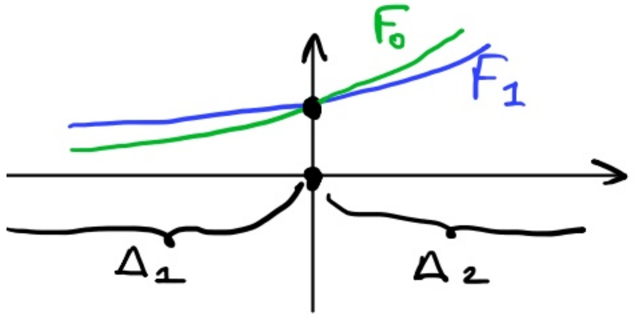
С каждой величиной X_i свяжем испытание с исходами $A_1, \dots, A_m$, причем A_j происходит тогда и только тогда, когда $X_i \in \triangle_j$.
При H_0 $P(A_j) = p_j^0$, тогда наблюдения $X_1, \dots, X_n$ порождают полиномиальную схему независимых испытаний.
Пусть ν_j - число исхода A_j в этих испытаниях, т.е. число наблюдений среди $X_1, \dots, X_n$, попавших в $\triangle_j$.
В силу теоремы Пирсона при H_0:

$$\chi_n^2 := \sum_{j=1}^{m} \frac{(\nu_j - np_j^0)^2}{np_j^0} \xrightarrow{d} \chi^2(m-1)$$

Правило 11.2. *H_0 будем отвергать, если $\chi_n^2 > \chi_{1-\alpha}(m-1) \;\Rightarrow\; P(H_1|H_0) \xrightarrow[n\to\infty]{} \alpha$.*

Если верна H_1 и хотя бы при одном j $p_j := P(X_1 \in \triangle_j | H_1) = \int\limits_{\triangle_j} dF_1(x) \neq p_j^0$, то $P(H_0|H_1) = P(\chi_n^2 < \chi_{1-\alpha}(m-1)|H_0) \to 0$

Замечание. *Если $F_0 \not\equiv F_1$, но $p_j = p_j^0 \; \forall j$, то $P(H_0|H_1) = P(H_0|H_0) \to 1 - \alpha \neq 0$. Например:*



$$m = 2$$
$$P(X_1 \in \triangle_1 | H_0) = F_0(0) =$$
$$= P(X_1 \in \triangle_1 | H_1) = F_1(0)$$
$$P(X_1 \in \triangle_2 | H_0) = 1 - F_0(0) =$$
$$= P(X_1 \in \triangle_2 | H_1) = 1 - F_1(0)$$

Доказательство. (доказательство теоремы Пирсона)

Покажем, что вектор $\nu = (\nu_1, \dots, \nu_m)^T$ асимптотически нормален, т.е.

$$n^{\frac{1}{2}}(n^{-1}\nu - p) \xrightarrow{d} N(0, P - pp^T) \tag{13}$$

где $P = \begin{pmatrix} p_1^0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & p_m^0 \end{pmatrix}$, $p = (p_1^0, \dots, p_m^0)^T$.

Для этого введем вектора $X_1, \dots, X_n$, где $X_i = (0, \dots, 0, 1, 0, \dots, 0)^T$ с 1 на j-ом месте, если в i-ом испытании произошло A_j, тогда: $\nu = \sum\limits_{i=1}^{n} X_i$,

$$n^{\frac{1}{2}}(n^{-1}\nu - p) = n^{\frac{1}{2}}\sum_{i=1}^{n}(X_i - p) \tag{14}$$

В (14) $\{X_i\}$ - н.о.р., $EX_1 = p$.

$$cov(X_1, X_1) = E(X_1 - p)(X_1 - p)^T = EX_1X_1^T - pp^T = P - pp^T$$

Поэтому соотношение (13) следует из представления (14) и ЦПТ.
Матрица $P - pp^T$ вырождена, т.к. сумма r столбцов есть ноль:
если $e = (1, \dots, 1)^T$, то $(P - pp^T)e = p - p(p^Te) = p - p = 0$.

Пусть $P^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\sqrt{p_1^0}} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{p_m^0}} \end{pmatrix}$, $\xi_n := n^{\frac{1}{2}}P^{-\frac{1}{2}}(n^{-1}\nu - p)$.

В силу теоремы о наследовании слабой сходимости и соотношения (13):

$$(H(x) = P^{-\frac{1}{2}}x,\ x \in \mathbb{R}^m)$$
$$\xi_n \xrightarrow{d} N(0, P^{-\frac{1}{2}}(P - pp^T)(P^{-\frac{1}{2}})^T)$$

Но $P^{-\frac{1}{2}}(P - pp^T)(P^{-\frac{1}{2}})^T = E_m - ZZ^T$, где $Z = (\sqrt{p_1^0}, \dots, \sqrt{p_m^0})^T$. Тогда:

$$\xi_n \xrightarrow{d} N(0, E_m - ZZ^T) \tag{15}$$

Пусть ортогональная матрица $U = \begin{pmatrix} \sqrt{p_1^0} & \dots & \sqrt{p_m^0} \\ \dots & \dots & \dots \end{pmatrix}$. Тогда:

$$U(E_m - ZZ^T)U^T = E_m - (UZ)(UZ)^T = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} =$$

$$= \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = \tilde{E}_1$$

В силу (15) и теоремы о наследовании слабой сходимости:

$$U\xi_n \xrightarrow{d} N(0, \tilde{E}_1) \tag{16}$$

т.е. $U\xi_n \xrightarrow{d} (0, \eta_2, \dots, \eta_m)^T,\ \{\eta_2, \dots, \eta_m\}$ - нез. $N(0,1)$ с.в.

Из (16) опять в силу теоремы о наследовании слабой сходимости:

$$|U\xi_n|^2 \xrightarrow{d} \eta_2^2 + \dots + \eta_m^2 \sim \chi^2(m-1) \tag{17}$$

Осталось заметить, что:

$$|U\xi_n|^2 = |\xi_n|^2 = \sum_{j=1}^m \left[\frac{1}{\sqrt{p_j^0}} n^{\frac{1}{2}} (n^{-1}\nu_j - p_j^0) \right]^2 = \sum_{j=1}^m \frac{(\nu_j - np_j^0)^2}{np_j^0} = \chi_n^2$$

Последнее равенство и соотношение (17) доказывают теорему Пирсона. ■

11.10 Проверка сложной гипотезы в схеме испытаний Бернулли

Пусть проводится n независимых испытаний, исходы $A_1, \dots, A_m$, $\nu = (\nu_1, \dots, \nu_m)^T$ - вектор наблюдений.
Пусть $H_0 : p(A_j) = p_j(\theta),\ \theta \in \Theta \subseteq \mathbb{R}^k,\ k < m-1$.

Условия регулярности:

(i) $\sum\limits_{j=1}^{m} p_j(\theta) = 1 \ \forall \theta \in \Theta$

(ii) $p_j(\theta) \geq c > 0$ для $j = 1, 2, \dots, m$ и $\exists \frac{\partial p_j(\theta)}{\partial \theta_l}, \ \frac{\partial^2 p_j(\theta)}{\partial \theta_l \partial \theta_r}$

(iii) $rk \left(\frac{\partial p_j(\theta)}{\partial \theta_l} \right) = k \ \forall \theta \in \Theta, \frac{\partial p_j(\theta)}{\partial \theta_l}$ - $m \times k$-матрица

В качестве оценки θ при H_0 будем использовать мультиноминальные оценки максимального правдоподобия:

$$P(\nu_1 = k_1, \dots, \nu_m = k_m) = \frac{n!}{k_1! \dots k_m!} p_1^{k_1}(\theta) \dots p_m^{k_m}(\theta)$$

Логарифм правдоподобия:

$$Ln(\nu, \theta) = \ln \frac{n!}{\nu_1! \dots \nu_m!} + \sum_{j=1}^{m} \nu_j \ln p_j(\theta)$$

О.м.п. (мультиномиальная): $Ln(\nu, \theta) \to \max\limits_{\theta \in \Theta}$.

Теорема 11.4 (**Фишера**). *Пусть верна H_0, пусть выполнены условия $(i) - (iii)$, пусть θ_n - мультиномиальная о.м.п., $\hat{\chi_n}^2 := \sum\limits_{j=1}^{m} \frac{(\nu_j - np_j(\hat{\theta_n}))^2}{np_j(\hat{\theta})}$.*

Тогда $\hat{\chi_n}^2 \xrightarrow{d} \chi(m-k-1)$.

Правило 11.3. *Если $\hat{\chi_n}^2 > \chi_{1-\alpha}(m-k-1)$, то H_0 отвергаем, тогда $P(\overline{H_0}|H_0) \to \alpha$.*

11.11 Проверка независимости признаков

Пусть объект классифицирован по двум признакам A и B, $A = \{A_1, \dots, A_s\}$, $B = \{B_1, \dots, B_r\}$, причем $s, r > 1$.
Проводится n опытов, пусть ν_{ij} - число объектов, имеющих признаки $A_i B_j$, пусть $p_{ij} = P(A_i B_j)$.

Гипотеза независимости $H_0 : p_{ij} = p_{i\cdot} p_{\cdot j}$ для положительных $p_{i\cdot}$ и $p_{\cdot j}$ таких, что $\sum\limits_{i=1}^{s} p_{i\cdot} = 1, \ \sum\limits_{j=1}^{r} p_{\cdot j} = 1.$

При H_0 логарифимческое правдоподобие:

$$Ln(\nu, p_{i\cdot}, p_{\cdot j}) = \ln \frac{n!}{\prod\limits_{ij} \nu_{ij}} + \sum_{i=1}^{s}\sum_{j=1}^{r} \nu_{ij}(\ln p_{i\cdot} + \ln p_{\cdot j})$$

Максимизируя эту функцию по $p_{i\cdot}, p_{\cdot j}$ при условиях $\sum\limits_{i=1}^{s} p_{i\cdot} = 1$ и $\sum\limits_{j=1}^{r} p_{\cdot j} = 1$, находим оценки:

$$\hat{p_{i\cdot}} = \frac{\nu_{i\cdot}}{n},\ \hat{p_{\cdot j}} = \frac{\nu_{\cdot j}}{n},\ \text{где } \nu_{i\cdot} = \sum_{j=1}^{r} \nu_{ij},\ \nu_{\cdot j} = \sum_{i=1}^{s} \nu_{ij}$$

Статистика хи-квадрат имеет вид:

$$\hat{\chi_n}^2 = \sum_{i=1}^{s}\sum_{j=1}^{r} \frac{(\nu_{ij} - n\hat{p_{i\cdot}}\hat{p_{\cdot j}})^2}{n\hat{p_{i\cdot}}\hat{p_{\cdot j}}}$$

При H_0 $\hat{\chi_n}^2 \xrightarrow{d} \chi\left((s-1)(r-1)\right)$, т.к. число разбиений $m-k-1 = sr-(s+r-2)-1 = (s-1)(r-1)$.

Правило 11.4. *Если $\hat{\chi_n}^2 > \chi_{1-\alpha}\left((s-1)(r-1)\right)$, то H_0 отвергается, асимптотический уровень теста равен α.*

Числовой пример (W.H. Gilby, Biometrika)
1725 школьников классифицировали (1) в соответствии с качеством их одежды и (2) в соответствии с их умственными способностями. Использовали следующую градацию:

градация	характеристика
A	умственно отсталый
B	медлительный и тупой
C	тупой
D	медлительный, но умный
E	достаточно умный
F	явно способный
G	очень способный

H_0 : признаки независимы.

как одевается?	**А и В**	**C**	**D**	**E**	**F**	**G**	**Сумма**
очень хорошо	33	48	113	209	194	39	636
хорошо	41	100	202	255	138	15	751
сносно	39	58	70	61	33	4	265
очень плохо	17	13	22	10	10	1	73
сумма	130	219	407	535	375	59	1725

Здесь $\chi_n^2 = 174.92 > \chi_{0.999}(15) = 37.697$, здесь $(s-1)(r-1) = (4-1)(6-1) = 15$, т.е. H_0 отвергается.

12 Введение в робастное оценивание

Схема, предложенная Мартином-Йохаи (Martin, Yohai, 1986), имеет вид:

$$y_t = u_t + z_t^\gamma \xi_t,\ t = 1, 2, \ldots, n$$

- здесь $\{u_t\}$ - «полезный сигнал» (временной ряд)
- $\{z_t^\gamma\}$ - н.о.р.с.в., $z_1^\gamma \sim Bin(1, \gamma)$ с $0 \le \gamma \le 1$, γ - уровень засорения
- ξ_t - н.о.р.с.в. - грубые выбросы, ξ_1 имеет распределение $\mu_\xi \in M_\xi$, распределение μ_ξ неизвестно, а множество M_ξ известно
- последовательности $\{u_t\}, \{z_t^\gamma\}, \{\xi_t\}$ независимы между собой

Пусть $y_1, \ldots, y_n$ - наблюдения, и распределение вектора $Y_n = (y_1, \ldots, y_n)$ зависит от неизвестного параметра β. Пусть $\hat{\beta}_n$ - некоторая оценка β.

Основное предположение
При любом $0 \le \gamma \le 1$ существует предел $\hat{\beta}_n \xrightarrow[n\to\infty]{P} \theta_\gamma,\ \theta_0 = \beta$, т.е. $\hat{\beta}_n$ состоятельна.

Определение 12.1. *Если существует предел*

$$IF(\theta_\gamma, \mu_\xi) := \lim_{\gamma\to+0} \frac{\theta_\gamma - \theta_0}{\gamma}$$

то $IF(\theta_\gamma, \mu_\xi)$ называется ***функционалом влияния*** *(influence function) оценки $\hat{\beta}_n$.*

Если функционал влияния существует, то

$$\theta_\gamma = \theta_0 + IF(\theta_\gamma, \mu_\xi)\gamma + \overline{o}(\gamma),\ \gamma \to +0$$

т.е. $IF(\theta_\gamma, \mu_\xi)$ характеризует главный линейный по γ член в разложении по γ асимптотического смещения $\theta_\gamma - \theta_0 = \theta_\gamma - \beta$.

Определение 12.2. *Величина* $CES(\theta_\gamma, M_\xi) := \sup\limits_{\mu_\xi \in M_\xi} |IF(\theta_\gamma, \mu_\xi)|$ *(cross error sensivity) называется* ***чувствительностью*** *оценки* $\hat{\beta}_n$ *к засорениям (выбросам).*

Если $CES(\theta_\gamma, M_\xi) < \infty$, то главный член по γ асимптотического смещения $IF(\theta_\gamma, \mu_\xi)\gamma$ равномерно мал при малых γ.

Определение 12.3. *Если* $CES(\theta_\gamma, M_\xi) < \infty$, *то оценка* $\hat{\beta}_n$ *называется* ***робастной*** *по смещению, или* ***B-робастной****.*

12.1 Пример о выборочном среднем

$$\begin{cases} u_t = a + \varepsilon_t, \ \varepsilon_t \text{ - ошибки измерений} \\ y_t = u_t + z_t^\gamma \xi_t, \ t = 1, 2, \dots, n, \quad E|\xi_1| < \infty \end{cases}$$

$\{\varepsilon_t\}$ - н.о.р.с.в., $E\varepsilon_1 = 0 \ \Rightarrow \ Eu_t = a$.

Возьмем оценкой a эмпирическое среднее $\overline{y} = n^{-1}\sum\limits_{t=1}^{n} y_t$ - о.н.к. $(\sum\limits_{t=1}^{n}(y_t - \theta)^2 \to \min\limits_{\theta})$, тогда:

$$\overline{y} \xrightarrow{P} E(u_1 + z_1^\gamma \xi_1)$$

(по теореме Колмогорова)

$$E(u_1 + z_1^\gamma \xi_1) = a + \gamma E\xi_1 = \theta_\gamma^{LS}$$

Функция θ_γ^{LS} (least square) определена при всех γ.

$$\frac{\partial \theta_\gamma^{LS}}{\partial \gamma} = E\xi_1 = IF(\theta_\gamma, \mu_\xi)$$

Если M_1 - класс распределений с конечным первым моментом, то

$$CES(\theta_\gamma^{LS}, M_1) = \sup_{\mu_\xi \in M_1} |E\xi_1| = \infty$$

т.е. $\overline{y}$ - не B-робастна на классе M_1.

12.2 Пример о выборочной медиане

Пусть

$$u_t = a + \varepsilon_t, \ t = 1, 2, \dots, n \tag{1}$$

где $\{\varepsilon_t\}$ - н.о.р.с.в., $\varepsilon_t \sim G(x)$ и ф.р. $G(x)$ неизвестна, $G(0) = \frac{1}{2}$. Тогда ф.р. u_t есть $F(x) = G(x-a)$, т.е. $F(a) = \frac{1}{2}$. Итак, ноль - медиана $G(x)$, а a - медиана $F(x)$.

Если ε_t имеет симметричное относительно нуля распределение (т.е. $\varepsilon_t \overset{\text{d}}{=} -\varepsilon_t$, что для непрерывной $G(x)$ равносильно условию $G(x) + G(-x) = 1$ при всех x), то условия выше выполняются автоматически.

Т.о. при сформулированных условиях оценку медианы можно использовать как оценку математического ожидания.

Пусть $u_{(1)} \le u_{(2)} \le \dots \le u_{(n)}$ будет вариационный ряд наблюдений $u_1, \dots, u_n$.

Определение 12.4. *Величина* $\hat{m}_n = \begin{cases} u_{(k+1)}, & n = 2k-1 \\ \frac{u_{(k+1)}+u_{(k)}}{2}, & n = 2k \end{cases}$ $, k = 1, 2, \dots$ *называется* ***выборочной медианой*** *наблюдений* $u_1, \dots, u_n$.

Мы знаем, что если $G(x)$ дифференцируема в нуле, и $g(0) = G'(0) > 0$, то для выборочной медианы справедлива асимптотическая нормальность:

$$n^{\frac{1}{2}}(\hat{m}_n - a) \xrightarrow[n\to\infty]{d} N\left(0, \frac{1}{4g^2(0)}\right)$$

Если в (1) $\{\varepsilon_t\}$ - н.о.р., $E\varepsilon_t = 0,\ 0 < E\varepsilon_t^2 = \sigma^2 < \infty$, то $n^{\frac{1}{2}}(\overline{u} - a) \xrightarrow[n\to\infty]{d} N(0, \sigma^2)$. Значит АОЭ выборочной медианы относительно $\overline{u}$ равна $e_{\hat{m}_n, \overline{X}} = 4g^2(0)\sigma^2$.

Изучим В-робастность выборочной медианы. Пусть:

$$\begin{cases} u_t = a + \varepsilon_t \\ y_t = u_t + z_t^\gamma \xi_t,\ t = 1, 2, \dots, n \end{cases}$$

$$\hat{m}_n^y = \begin{cases} y_{(k+1)}, n = 2k-1 \\ \frac{y_{(k)}+y_{(k+1)}}{2}, n = 2k \end{cases}$$

Теорема 12.1. *Пусть существует производная* $g(x) = G'(x)$, $g(x)$ *непрерывна и ограничена,* $g(0) > 0$, $G(0) = \frac{1}{2}$. *Тогда:*

1) $\hat{m}_n^y \xrightarrow[n\to\infty]{P} \theta_\gamma^m$, $\theta_0 = a$

2) существует функционал влияния выборочной медианы

$$IF(\theta_\gamma^m, \mu_\xi) = \frac{1 - 2EG(-\xi_1)}{2g(0)}$$

3) чувствительность выборочной медианы на классе всех возможных распределений M_ξ

$$GES(\theta^m_\gamma, M_\xi) = \sup_{\mu_\xi \in M_\xi} |IF(\theta^m_\gamma, \mu_\xi)| = \frac{1}{2g(0)} < \infty$$

т.е. выборочная медиана B-робастна.

Доказательство.

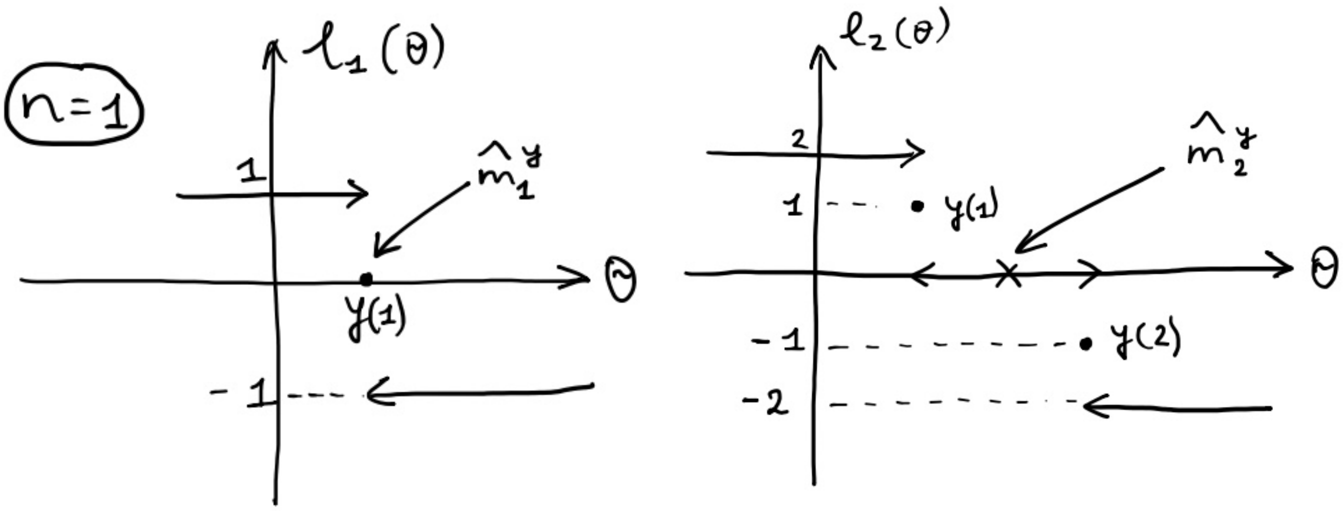
ШАГ 1.

Выборочная медиана $\hat{m}^y_n$ удовлетворяет уравнению:

$$l_n(\theta) := n^{-1}\sum_{t=1}^{n} sign(y_t - \theta) = 0 \qquad (2)$$

где $sign(x) = \begin{cases} -1, x < 0 \\ 0, x = 0 \\ 1, x > 1 \end{cases}$

Справедливость (2) легко понять из рисунков:



Так бывает всегда: при нечетном n решение уравнения (2) всегда существует и единственно - это $\hat{m}^y_n$; при четном n решений целый интервал и $\hat{m}^y_n$ - его середина.

В силу ЗБЧ при любом θ и любом $0 \le \gamma \le 1$:

$$l_n(\theta) = n^{-1}\sum_{t=1}^{n} sign(y_t - \theta) \xrightarrow[n\to\infty]{P} Esign(y_1 - \theta) =: \Lambda_M(\gamma, \theta)$$

Задача 12.1. *Пусть ξ, η – независимые случайные векторы, причем η – дискретный вектор со значениями $\eta_1, \eta_2, \ldots$. Необходимо проверить, что*

$$E\varphi(\xi,\eta) = \sum_{k\geq 1} E\varphi(\xi,\eta_k)P(\eta=\eta_k) = \sum_{k\geq 1} E(\varphi(\xi,\eta)|H_k)P(H_k),$$

где гипотеза $H_k = (\eta = \eta_k)$.

Найдем удобный вид для $\Lambda_M(\gamma,\theta)$. Имеем

$$\Lambda_M(\gamma,\theta) = E(1-2I(y_1-\theta\leq 0)) = 1-2EI(\varepsilon_1 \leq \theta - a - z_1^\gamma\xi_1) = 1-2EG(\theta-a-z_1^\gamma\xi_1), \tag{3}$$

т.к. $sign(x) = 1-2I(x<0)$ при $x\neq 0$.

Чтобы упростить (3), введем две гипотезы: $H_1 = (z_1^\gamma = 0)$, $H_2 = (z_2^\gamma = 1)$. Тогда, используя задачу, получим из (3), что

$$\Lambda_M(\gamma,\theta) = 1 - 2(1-\gamma)G(\theta-a) - 2\gamma EG(\theta - a - \xi_1).$$

Функция $\Lambda_M(\gamma,\theta)$ определена при всех γ, θ, в том числе для отрицательных γ.

ШАГ 2.

Функция $\Lambda_M(\gamma,\theta)$ в окрестности точки (0, a) удовлетворяет всем предположениям теоремы о неявной функции. А именно:

1. $\Lambda_M(0,a) = 1-2G(0) = 0$;
2. Существуют и непрерывнф по паре (γ,θ) фнкции $\dfrac{\partial\Lambda_M(\gamma,\theta)}{\partial\gamma}$, $\dfrac{\partial\Lambda_M(\gamma,\theta)}{\partial\theta}$;
3. $\dfrac{\partial\Lambda_M(\gamma,\theta)}{\partial\theta} = -2g(0)\neq 0.$

Значит, в окрестности точки (0, a) определена функция $\theta_m(\gamma) = \theta_\gamma^m$ такая, что $\Lambda_M(\gamma,\theta_\gamma^m) = 0$. Кроме того, $\theta_0^m = a$, $\theta_\gamma^m \longrightarrow \theta_0$ при $\gamma\longrightarrow 0$. Функция θ_γ^m дифференцируема в точке $\gamma=0$, и

$$\frac{d\theta_\gamma^m}{d\gamma}|_{\gamma=0} = -(\frac{\partial\Lambda_M(0,a)}{\partial\theta})^{-1}\frac{\partial\Lambda_M(0,a)}{\partial\gamma} = \frac{1-2EG(-\xi_2)}{2g(0)} \tag{4}$$

ШАГ 3.

Покажем, что

$$\hat{m_n^y} \longrightarrow \theta_\gamma^m,\ n\longrightarrow\infty \tag{5}$$

Тогда из (4), (5) будет следовать, что функционал влияния выборочной медианы равен

$$IF(\theta_\gamma^m,\mu_\xi) = \frac{1-2EG(-\xi_1)}{2g(0)} \tag{6}$$

Модуль числителя в (6) не больше 1, причем, если ξ_1 неслучайно и $\xi_1 \longrightarrow +\infty$, то числитель стремится к 1. Значит,

$$GES(\theta_\gamma^m, M_\xi) = \sup_{\mu_\xi \in M_\xi} |IF(\theta_\gamma^m, \mu_\xi)| = \frac{1}{2g(0)}.$$

Итак, докажем (5). Имеем при малых γ (γ фиксированно!) и θ вблизи а:

$$\frac{\partial \Lambda_M(\gamma, \theta)}{\partial \theta} = -2(1-\gamma)g(\theta - a) - 2\gamma E g(\theta - a - \xi_1) < 0$$

То есть $\Lambda(\gamma, \theta)$ убывает по θ. Значит,

$$\begin{cases} \Lambda_M(\gamma, \theta_\gamma^m - \Delta) > 0 \\ \Lambda_M(\gamma, \theta_\gamma^m + \Delta) < 0 \end{cases},$$

но

$$\begin{cases} l_n(\theta_\gamma^m - \Delta) \xrightarrow{P} \Lambda_M(\gamma, \theta_\gamma^m - \Delta) > 0 \\ l_n(\theta_\gamma^m + \Delta) \xrightarrow{P} \Lambda_M(\gamma, \theta_\gamma^m + \Delta) < 0 \end{cases} \tag{7}$$

Функция $l_n(\theta)$ монотонно убывает (точнее, не возрастет) по θ. В силу (7) с вероятностью сколь угодно близкой к единице при достаточно больших n все корни уравнения $l_n(\theta) = 0$ лежат в интервале $(\theta_\gamma^m - \Delta, \theta_\gamma^m + \Delta)$. Выборочная медиана $\hat{m_n^y}$ тоже лежит в этом интервале! Поскольку $\Delta > 0$ любое, получаем, что

$$\hat{m_n^y} \xrightarrow{P} \theta_\gamma^m, \; n \longrightarrow \infty.$$

■

12.3 Нахождение функционала влияния в общем случае

Пусть оценка $\hat{\beta}_n$ ищется как корень уравнения

$$l_n(\theta) := n^{-1} \sum_{t=1}^{n} \varphi(\mathcal{I}_n, \theta) = 0 \tag{1}$$

Пусть выполнены следующие условия:

$$(i) l_n(\theta) = n^{-1}\sum_{t=1}^{n}\varphi(\mathcal{I}_n,\theta) \xrightarrow{P} \Lambda(\gamma,\theta) \text{ при всех } |\theta-\beta|<\delta,\ 0\geq\gamma<\gamma_0$$

$$(ii)\Lambda(0,\beta)=0$$

(iii)Пусть $\Lambda(\gamma,\theta)$ можно продолжить на отрицательные малые γ так, что при $|\theta-\beta|<\delta,\ |\gamma|<\gamma_0$ существуют и непрерывны по паре (γ,θ) функции

$$\frac{\partial\Lambda(\gamma,\theta)}{\partial\gamma},\ \frac{\partial\Lambda(\gamma,\theta)}{\partial\theta}$$

$$(iV)\text{Пусть } \lambda(\beta):=\frac{\partial\Lambda(\gamma,\theta)}{\partial\theta}\neq 0$$

Теорема 12.2. *Пусть выполнены условия (i)-(iV), функции $\varphi(\mathcal{I}_n,\theta)$ непрерывны по θ. Тогда уравнение (1) с вероятностью, стремящейся к 1 при $n\longrightarrow\infty$, имеет при достаточно малых $\gamma\geq 0$ такое решение $\hat{\beta}_n$, что соответствующая оценка $\hat{\beta}_n\xrightarrow{P}\theta_\gamma,\ \theta_0=0,$ и $\exists$ функционал влияния*

$$IF(\theta_\gamma,\mu_\xi)=-(\lambda(\beta))^{-1}\frac{\partial}{\partial\gamma}\Lambda(0,\beta).$$

12.4 M - оценка медианы

Пусть

$$\begin{cases} u_t=a+\varepsilon_t\\ y_t=u_t+z_t^g\xi_t,\end{cases}$$

где $\{\varepsilon_t\}$ – н.о.р., $\varepsilon_1\sim g(x)=G'(x),\ g(x)=g(-x)$. Тогда a – медиана ф.р. сл.в. u_1. Будем искать оценку a (обозначим ее $\hat{a_n}$) как корень уравнения

$$(9)\ \sum_{t=1}^{n}\psi(y_t-\theta)=0$$

Такая оценка называется **M - оценкой**. Вчастности, при $\psi(x)=x,\ \hat{a_n}=\bar{y}$, при $\psi(x)=sign(x)\ \hat{a_n}=\hat{m_n}^y$.

Пусть выполнены условия:

1. $\psi(x)$ – нечетная строго возрастающая функция, $\lim\limits_{x\to+\infty}\psi(x)=c_1>0,$
   $\lim\limits_{x\to-\infty}\psi(x)=c_2<\infty$

2. $\exists$ непрерывная и ограниченная $\psi'(x)$, $E\psi'(\varepsilon_1) \neq 0$

Тогда уравнение (9) всегда имеет единственное решение. Условия 1, 2 выполнены, например, для $\psi(x) = \arctan(x)$, $\lim\limits_{x\to+\infty}\psi(x) = \frac{\pi}{2}$, $\lim\limits_{x\to-\infty}\psi(x) = -\frac{\pi}{2}$, $\psi'(x) = \frac{1}{x^2+1}$.

Найдем функционал влияния и чувствительность М - оценки. Используем теорему 12.2. Проверим условия:

$$n^{-1}\sum_{t=1}^{n}\psi(y_t-\theta) \xrightarrow{P} E\psi(y_1-\theta) =: \Lambda(\gamma,\theta) \text{ при всех } \theta,\ 0\le\gamma\le 1 \qquad (i)$$

Введем гипотезы $H_1 = (z_1^\gamma = 0)$, $H_2 = (z_2^\gamma = 1)$. Тогда:

$$\Lambda(\gamma,\theta) = \sum_{i=2}^{n} E(\psi(\underbrace{\varepsilon_1 + a + z_1^\gamma\xi_1}_{=y_1} - \theta) \mid H_i) \times P(H_i) =$$
$$= (1-\gamma)E\psi(\varepsilon_1 + a - \theta) + \gamma E\psi(\varepsilon_1 + \xi_1 + a - \theta)$$

$$\Lambda(0,a) = E\psi(\varepsilon_1) = 0 \qquad (ii)$$

т.к. $\Lambda(\gamma,\theta)$ определена при всех γ и θ. $\frac{\partial\Lambda(\gamma,\theta)}{\partial\gamma}$, $\frac{\partial\Lambda(\gamma,\theta)}{\partial\theta}$ существуют при условиях (i), (ii) и непрерывна по паре (γ,θ). В частности,

$$\frac{\partial\Lambda(\gamma,\theta)}{\partial\gamma} = -E\psi(\varepsilon_1) + E\psi(\varepsilon_1+\xi_1) = 0 + E\psi(\varepsilon_1+\xi_1).$$

$$\frac{\partial\Lambda(0,a)}{\partial\theta} = -E\psi'(\varepsilon_1) \neq 0 \qquad (iii)$$

В силу теоремы 12.2 $\hat{a_n} \xrightarrow{P} \theta_\gamma$, $\theta_0 = a$,

$$IF(\theta_\gamma, \mu_\xi) = \frac{E\psi(\varepsilon_1+\xi_1)}{E\psi'(\varepsilon_1)},$$
$$GES(\theta_\gamma, M_\xi) \le \frac{max(|c_1|,|c_2|)}{E\psi'(\varepsilon_1)} < \infty,$$

M_ξ – класс всех вер. распределений.

13 Статистический анализ AR моделей

Пусть $\ldots, S_{-1}, S_0, S_1, \ldots$ – стоимости ценных бумаг, например, акций. Величины $u_t = \log \frac{S_t}{S_{t-1}} = \log S_t - \log S_{t-1}$ называются **логарифмическими приращениями** и для описания их поведения часто используют стохастические разностные уравнения.

Например, *AR(p)* - уравнение имеет вид

$$u_t = \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_p u_{t-p} + \varepsilon_t,\ t \in \mathbb{Z},\ \{\varepsilon_t\} \text{ – н.о.р.сл.в., } E\varepsilon_1 = 0,$$
$$0 < E\varepsilon_1^2 = \sigma^2 < \infty,\ \beta_1, \ldots, \beta_p \in \mathbb{R}^1 \text{ – неизвестные коэффициенты авторегрессии,}$$
$$\beta_p \neq 0.$$

Иногда удобно рассматривать *AR(p)* - уравнение для $t = 1, 2, \ldots$ при начальных условиях $w_{1-p}, \ldots, w_n$.

ARCH(p) - уравнение имеет вид

$$u_t = \sigma_t \varepsilon_t,\ \text{где } \sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_p u_{t-p}^2,\ t \in \mathbb{Z},\ \alpha_0 > 0,\ \alpha_i \geq 0,\ \alpha_p > 0,$$
$$\{\varepsilon_t\} \text{ – н.о.р., } E\varepsilon_1 = 0,\ E\varepsilon_1^2 = 1.$$

13.1 Метод максимального правдоподобия и метод наименьших квадратов в авторегрессии

AR(1) - модель

$$u_t = \beta u_{t-1} + \varepsilon_t,\ t = 1, 2, \dots,\ \{\varepsilon_t\} \text{ – н.о.р.сл.в., } E\varepsilon_1 = 0,$$
$$0 < E\varepsilon_1^2 = \sigma^2 < \infty, \beta \in \mathbb{R}^1,\ u_0 = 0.$$
$$\text{Тогда } u_t = \beta(\beta u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \varepsilon_t + \beta\varepsilon_{t-1} + \beta^2 u_{t-2} = \dots =$$
$$= \varepsilon_t + \beta\varepsilon_{t-1} + \dots + \beta^{t-1}\varepsilon_1.$$

1. *Стационарный случай* $|\beta| < 1$.
   $u_t \xrightarrow{\text{с.к.}} u_t^0 := \sum\limits_{j\geq 0} \beta^j \varepsilon_{t-j}$(это строго стационарная последовательность, стационарная по Хинчину в широком смысле) и ряд с.к. сходится, так как $E(u_t - u_t^0)^2 = E(\sum\limits_{j\geq t} \beta^j\varepsilon_{t-j})^2 = E\varepsilon_1^2 \sum\limits_{j\geq t}\beta^{2j} = \mathcal{O}(\beta^{2t}) = \mathcal{O}(1),\ t \longrightarrow \infty.$
2. *Критический случай (неустойчивая авторегрессия)* $|\beta| = 1$.
3. *Взрывающаяся авторегрессия* $|\beta| > 1$.
   $Du_t = D\sum\limits_{j=0}^{t-1}\beta^j\varepsilon_{t-j}$ (как дисперсия суммы независимых) $= E\varepsilon_1^2\sum\limits_{j=0}^{t-1}\beta^{2t} =$
   $= \dfrac{E\varepsilon_1^2(1-\beta^{2t})}{1-\beta^2}$ (по формуле геометрической прогрессии) $= \mathcal{O}(\beta^{2t}) \longrightarrow \infty$ при $t \longrightarrow \infty$ экспоненциально быстро. Мы знаем, что оптимальный с.к. прогноз u_{n+1} по $u_1, \dots, u_n$ есть $\tilde{u_{n+1}} = \beta u_n$
   Надо уметь оценивать β. Пусть $\varepsilon_1 \sim g(x)$ – плотность вероятности по мере Лебега. Положим:

$$\varepsilon := (\varepsilon_1, \dots, \varepsilon_n)^T,\ u = (u_1, \dots, u_n)^T, B = \begin{pmatrix} 1 & 0 & \dots & \dots & 0 \\ -\beta & 1 & \dots & \dots & 0 \\ 0 & -\beta & \dots & \dots & 0 \\ \dots & \dots & \ddots & \ddots & \dots \\ 0 & \dots & \dots & -\beta & 1 \end{pmatrix}$$

Тогда из (1) $\varepsilon = Bu$, имеем (2) $u = B^{-1}\varepsilon$. Плотность вероятности вектора ε

есть $g_\varepsilon(x_1, \ldots, x_n) = \prod\limits_{i=1}^{n} g(x_i)$. Тогда плотность вероятности вектора u в силу (2):

$$g_u(y, B) = \frac{1}{|B^{-1}|} g_\varepsilon(By) = \prod_{t=1}^{n} g(y_t - \beta y_{t-1}),\ y = (y_1, \ldots, y_n),\ y_0 = 0.$$

О.м.п. для β – решение задачи

$$\log g_u(u, \theta) = \sum_{t=1}^{n} \log g(u_t - \theta u_{t-1}) \to \underset{\theta \in \mathbb{R}^1}{max} \tag{3}$$

Для гладкой g уравнение максимального правдоподобия

$$\sum_{t=1}^{n} u_{t-1} \frac{g'(u_t - \theta u_{t-1})}{g(u_t - \theta u_{t-1})} = 0 \tag{4}$$

Пример 13.1 ($\varepsilon \sim N(0, \sigma^2)$)**.** *Тогда* $g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$ *и задача (3) имеет вид*

$$\sum_{t=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u_t - \theta u_{t-1})^2}{2\sigma^2}} \to \underset{\theta \in \mathbb{R}^1}{max}.$$

Последняя задача эквивалентна следующей:

$$\sum_{t=1}^{n} (u_t - \theta u_{t-1})^2 \longrightarrow \underset{\theta \in \mathbb{R}^1}{min} \tag{5}$$

Решение (5) – о.м.п.:

$$\hat{\beta}_{n,Mh} = \frac{\sum\limits_{t=1}^{n} u_{t-1} u_t}{\sum\limits_{t=1}^{n} u_{t-1}^2} \tag{6}$$

Если мы не предполагаем гауссовость ε_1*, то решение задачи (5) есть о.н.к.*

$$\hat{\beta}_{n,hS} = \frac{\sum\limits_{t=1}^{n} u_{t-1} u_t}{\sum\limits_{t=1}^{n} u_{t-1}^2} \tag{7}$$

Оценка $\hat{\beta}_{n,Mh}$ *– параметрическая,* $\hat{\beta}_{n,hS}$ *– непараметрическая.*

Пример 13.2 ($\varepsilon \sim Lap(\lambda)$). *Тогда* $g(x) = \frac{\lambda}{2}e^{-\lambda|x|}$, $\lambda > 0$. *Задача* (5) *имеет вид* $\sum\limits_{t=1}^{n} \log \frac{\lambda}{2} e^{-\lambda|u_t - \theta u_{t-1}} \longrightarrow \max\limits_{\theta \in \mathbb{R}^1}$, *что эквивалентно задаче*

$$\sum_{t=1}^{n} |u_t - \theta u_{t-1}| \longrightarrow \min_{\theta \in \mathbb{R}^1} \tag{8}$$

Решение (8) – о.м.п. $\hat{\beta}_{n,Mh}$. *Если распределение* ε_1 *неизвестно, то решение (8) – о.н.м.* $\hat{\beta}_{n,hD}$.

Замечание. *Оценка* $\hat{\beta}_{n,hD}$ *не выписывается явно!*

13.2 Случай гауссовских $\{\varepsilon_t\}$, $\varepsilon \sim N(0,1)$, теорема о предельном распределении о.м.п. в AR(1)

Теорема 13.1. *Пусть*

$$d_n^2(\beta) = \begin{cases} \dfrac{n}{1-\beta^2}, \ |\beta| < 1 \\ \dfrac{n^2}{2}, \ |\beta| = 1 \\ \dfrac{\beta^{2n}}{(\beta^2-1)^2}, \ |\beta| > 1 \end{cases}$$

Покажем, что $d_n^2(\beta) \sim \mathbb{J}_n(\beta)\ n \longrightarrow \infty$, $\mathbb{J}_n(\beta)$ *– информация Фишера о параметре* β, *содержащаяся в* $u_1, \ \ldots, u_n$

Доказательство. Если $u = (u_1, \ldots, u_n)$, $y = (y_1, \ldots, y_n)$, то плотность вероятности

$$g_u(y, \beta) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum\limits_{t=1}^{n}(y_t - \beta y_{t-1})^2},$$

а потому

$$\mathbb{J}_n(\beta) = E_\beta(\frac{\partial}{\partial\beta}\log(g_u(y,\beta))^2 = E_\beta(\frac{\partial}{\partial\beta}(-\frac{1}{2}\sum_{t=1}^{n}(u_t-\beta u_{t-1})^2)) =$$

$$= E_\beta(\sum_{t=1}^{n}u_{t-1}(u_t-\beta u_{t-1}))^2 = E_\beta(\sum_{t=1}^{n}u_{t-1}\varepsilon_t)^2 = \sum_{t=1}^{n}E_\beta u_{t-1}^2 = \sum_{t=1}^{n-1}E_\beta u_t^2,$$

но $u_t = \sum\limits_{j=1}^{t-1}\beta^j\varepsilon_{t-j}$, и

$$Eu_t^2 = E(\sum_{j=1}^{t-1}\beta^j\varepsilon_{t-j})^2 = \sum_{j=0}^{t-1}\beta^{2j} = \begin{cases}\dfrac{1-\beta^{2t}}{1-\beta^2}, & |\beta|\neq 1\\ t, & |\beta|=1\end{cases}$$

Значит,

$$\mathbb{J}_n(\beta) = \begin{cases}\dfrac{n-1}{1-\beta^2}-\dfrac{\beta^2(1-\beta^{\frac{2}{n-1}})}{(1-\beta^2)^2}, & |\beta|\neq 1\\ \dfrac{(n-1)(1+(n-1))}{2}, & |\beta|=1\end{cases}$$

Отсюда

$$\mathbb{J}_n(\beta) \sim \begin{cases}\dfrac{n}{1-\beta^2}, & |\beta|<1\\ \dfrac{n^2}{2}, & |\beta|=1\\ \dfrac{\beta^{2n}}{(\beta^2-1)^2}, & |\beta|>1\end{cases} = d_n^2(\beta)$$

■

13.3 Случай гауссовских $\{\varepsilon_t\}$, $\varepsilon\sim N(0,1)$, теорема о предельном распределении о.м.п. в AR(1) при гауссовских инновациях при случайной нормировке

Распределение Коши с параметрами (0,1) обозначим $\mathbb{K}$, то есть $f(x)=\dfrac{1}{\pi}\dfrac{1}{1+x^2}$. Пусть $w(s),\ s\in[0,1]$, – *стандартный винеровский процесс.*

Напоминание 13.1.

Определение 13.1. *Случайный процесс w_t, $t \geq 0$, называется **винеровским процессом**, если:*

1. $w_0 = 0$ *почти наверное*
2. w_t *– процесс с независимыми приращениями*
3. $w_t - w_s \sim N(0, \sigma^2(t-s))\ \forall\ 0 \leq s < t \leq \infty$

Обозначим за $H(\beta)$, $|\beta| = 1$, распределение случайной величины

$$\beta \frac{w^2(1) - 1}{2^{\frac{3}{2}} \int\limits_0^1 w^2(s)ds}.$$

$u_t = \beta u_{t-1} + \varepsilon_t$, $t = 1, 2, \dots$, $\beta \in \mathbb{R}$.

Теорема 13.2. *Пусть $\{\varepsilon_t\}$ – н.о.р.сл.в., $\varepsilon_1 \sim N(0,1)$. Тогда*

$$d_n(\beta)(\hat{\beta}_{n,Mh} - \beta) \xrightarrow[n\to\infty]{d} \begin{cases} N(0,1), & |\beta| < 1 \\ H(\beta), & |\beta| = 1 \\ \mathbb{K}(0,1), & |\beta| > 1 \end{cases}$$

Доказательство.

$$\hat{\beta}_{n,Mh} = \frac{\sum\limits_{t=1}^n u_{t-1}u_t}{\sum\limits_{t=1}^n u_{t-1}^2} = \frac{\sum\limits_{t=1}^n u_{t-1}(\beta u_{t-1} + \varepsilon_t)}{\sum\limits_{t=1}^n u_{t-1}^2} = \beta + \frac{\sum\limits_{t=1}^n \varepsilon_t u_{t-1}}{\sum\limits_{t=1}^n u_{t-1}^2}.$$

Положим для краткости $M_n := d_n^{-1}(\beta) \sum\limits_{t=1}^n \varepsilon_t u_{t-1}$, $V_n := d_n^{-2}(\beta) \sum\limits_{t=1}^n u_{t-1}^2$, Тогда

$$d_n(\beta)(\hat{\beta}_{n,Mh} - \beta) = \frac{M_n}{V_n}.$$

Пусть $f_n(t,s)$ – совместная характеристическая функция M_n, V_n. Тогда(см. [Rao Statist., 1978, v.6, pp 185 - 190]). Рао доказал, что

$$f_n(t,s) \to f(t,s) = \begin{cases} e^{is - \frac{t^2}{2}}, & |\beta| < 1 \\ (1 + t^2 - 2is)^{-\frac{1}{2}}, & |\beta| > 1 \end{cases} \tag{9}$$

1. $|\beta| < 1$. Тогда $f_n(t, s)$ есть характеристическая функция вектора $(\xi, 1)^T$, где $\xi \sim N(0, 1)$. Действительно:

$$\varphi(t, s) = Ee^{i(t\xi + s1))} = e^{is}\varphi_\xi(t) = e^{is - \frac{t^2}{2}}.$$

Теорема 13.3 (**Теорема о наследовании слабой сходимости**). *Пусть сл.в. $S_n \xrightarrow{d} S$, $n \to \infty$, S_n, $S \in \mathbb{R}^k$, а $H : \mathbb{R}^k \to \mathbb{R}$ – борелевская функция, непрерывная на множестве A таком, что $P(S \in A) = 1$. Тогда $H(S_n) \xrightarrow{d} H(S)$, $n \to \infty$*

У нас в силу (9) $(M_n, V_n)^T \xrightarrow{d} (\xi, 1)^T$ (из сходимости х.ф. имеем сходимость по распределению). Если $H(x, y) = \frac{x}{y}$, то $H(x, y)$ непрерывна при у > 0. Можно взять $A = \{y : y > 0\}$, $P((\xi, 1)^T \in A) = 1$. В силу теоремы о наследовании слабой сходимости:

$$d_n(\beta)(\hat{\beta}_{n,Mh} - \beta) = \frac{M_n}{V_n} = H(M_n, V_n) \xrightarrow{d} H(\xi, 1) = \xi.$$

2. $|\beta| > 1$. Тогда f(t, s) есть х.ф. вектора $(\xi\eta, \eta^2)^T$, где $\xi, \eta \sim N(0, 1)$, ξ, η независимы. Действительно,

$$\varphi(t, s) = Ee^{i(t\xi\eta + s\eta^2))} = EE(e^{i(t\xi\eta + s\eta^2))}|\eta), \text{ т.е. считаем, что } \eta \text{ – константа,}$$

$$is\eta^2 - \eta\text{-измеримая, поэтому} = Ee^{is\eta^2}E(e^{i(t\eta)\xi)}|\eta) = e^{is\eta^2}\varphi_\xi(t\eta),\ \varphi_\xi \text{ – х.ф. } \xi :=$$

$$:= e^{is\eta^2}e^{-\frac{t^2\eta^2}{2}} = Ee^{i(s + \frac{it^2}{2})\eta^2}, \text{ т.к. } \eta^2 \sim \chi^2(1) \text{ (х.ф. для хи-квадрат: } Ee^{ilx_1^2} =$$

$$= (1 - 2il)^{-\frac{1}{2}}) = (1 - 2is + \frac{2t^2}{2})^{-\frac{1}{2}} = (1 - 2is + t^2)^{-\frac{1}{2}} = \varphi(t, s).$$

Значит, $(M_n, V_n)^T \xrightarrow{d} (\xi\eta, \eta^2)^T$ и:

$$d_n(\beta)(\hat{\beta}_{n,Mh} - \beta) = \frac{M_n}{V_n} \xrightarrow{d} \frac{\xi\eta}{\eta^2} = \frac{\xi}{\eta} \sim K(0, 1)$$

3. $\beta = 1$, случай $\beta = -1$ аналогичен. Тогда:

$$M_n = \frac{\sqrt{2}}{n}\sum_{t=1}^{n}\varepsilon_t u_{t-1}, V_n = \frac{2}{n^2}\sum_{t=1}^{n}u_{t-1}^2$$

Далее, $u_t = u_{t-1} + \varepsilon_t = \varepsilon_1 + \ldots + \varepsilon_t$. Введем винеровский последовательный процесс:

$$W_n(s) := n^{-\frac{1}{2}} \sum_{i \le ns} \varepsilon_i, \; s \in [0,1],$$

$$W_n(s) = 0 \text{ при } 0 \le s \le \frac{1}{n}.$$

Тогда $n^{-\frac{1}{2}} u_{t-1} = W_n(\frac{t-1}{n})$. Пусть $\Delta W_n(\frac{t}{n}) := W_n(\frac{t}{n}) - W_n(\frac{t-1}{n}) = \frac{\varepsilon_t}{\sqrt{n}}$, тогда

$$M_n = \sqrt{2} \sum_{t=1}^{n} W_n(\frac{t-1}{n}) \Delta W_n(\frac{t}{n}),$$

$$V_n = 2 \sum_{t=1}^{n} W_n(\frac{t-1}{n})^2 \frac{1}{n}.$$

Пусть $U_n = (W_n(\frac{1}{n}), W_n(\frac{2}{n}), \dots, W_n(\frac{n}{n}))^T = (\frac{\varepsilon_1}{\sqrt{n}}, \frac{\varepsilon_1+\varepsilon_2}{\sqrt{n}}, \dots, W_n(\frac{\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_n}{\sqrt{n}}))^T$. Это есть гауссовский вектор со средним 0, $Cov(W_n(\frac{i}{n}), W_n(\frac{j}{n})) = \frac{min(i,j)}{n}$. Действительно:

$$U_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

Для $i \le j$ имеем:

$$Cov(W_n(\frac{i}{n}), W_n(\frac{j}{n})) = E(\frac{1}{n} \sum_{t=1}^{i} \varepsilon_t \sum_{k=1}^{j} \varepsilon_k), (\text{ т.к. } E\varepsilon_s\varepsilon_p = 0) = \frac{1}{n} E(\sum_{t=1}^{i} \varepsilon_t)^2 =$$
$$= \frac{i}{n} = \frac{min(i,j)}{n}.$$

Введем вектор $U = (W(\frac{1}{n}), W(\frac{2}{n}), \dots, W(\frac{n}{n}))^T$, где $W(s)$ – стандартный винеровский. U – гауссовский вектор со средним 0, $Cov(W(\frac{i}{n}), W(\frac{j}{n})) = \frac{min(i,j)}{n}$. Значит,

$$U_n \stackrel{d}{=} U, \text{ следовательно, } \forall \text{ борелевской } \varphi: \; \varphi(U_n) \stackrel{d}{=} \varphi(U) \qquad (10)$$

Действительно, пусть $\xi \stackrel{d}{=} \eta, \; \xi\eta \in \mathbb{R}^k$. Тогда $f(\xi) \stackrel{d}{=} f(\eta)$, т.к. $P(f(\xi) \in A) =$

$P(\xi \in f^{-1}(A)) = P(\eta \in f^{-1}(A)) = P(f(\eta) \in A)$. Пусть

$$\bar{M}_n = \sqrt{2}\sum_{t=1}^{n} W(\frac{t-1}{n})\Delta W(\frac{t}{n}),$$
$$\bar{V}_n = 2\sum_{t=1}^{n} W(\frac{t-1}{n})^2\frac{1}{n}.$$

Из (10) следует, что

$$\frac{M_n}{V_n} \stackrel{d}{=} \frac{\bar{M}_n}{\bar{V}_n}, \tag{11}$$

т.к. M_n, V_n – борелевские функции от U_n, $\bar{M}_n, \bar{V}_n$ – борелевские функции от U. Но $\bar{M}_n \stackrel{\text{с.к.}}{\to} \sqrt{2}\int\limits_0^1 W(s)dW(s)$, $\bar{V}_n \stackrel{\text{с.к.}}{\to} 2\int\limits_0^1 W^2(s)ds$. Значит, $(M_n, V_n)^T \stackrel{d}{\to}$ $(\sqrt{2}\int\limits_0^1 W(s)dW(s), \int\limits_0^1 W^2(s)ds)^T$, следовательно,

$$\frac{\bar{M}_n}{\bar{V}_n} \to \frac{\sqrt{2}\int\limits_0^1 W(s)dW(s)}{\int\limits_0^1 W^2(s)ds} = \frac{W^2(1)-1}{2^{\frac{3}{2}}\int\limits_0^1 W^2(s)ds}. \tag{12}$$

Поскольку $d_n(\beta)(\hat{\beta}_{n,Mh}-\beta) = \frac{\bar{M}_n}{\bar{V}_n}$, соотношения (11), (12) влекут утверждение теоремы.

■

Теорема 13.4. $\{\varepsilon_t\}$ *– н.о.р.* $N(0,1)$*. Тогда*

$$\sqrt{\sum_{t=1}^{n} u_{t-1}^2}(\hat{\beta}_{n,Mh}-\beta) \stackrel{d}{\to} \begin{cases} N(0,1), \ |\beta| \neq 1, \\ \tilde{H}(\beta), \ |\beta| = 1 \end{cases}$$

Здесь $\tilde{H}(\beta)$ *– распределение случайной величины*

$$\frac{\sqrt{2}\int\limits_0^1 W(s)dW(s)}{2\sqrt{\int\limits_0^1 W^2(s)ds}} = \frac{W^2(1)-1}{\sqrt{\int\limits_0^1 W^2(s)ds}}.$$

Доказательство.

$$\sqrt{\sum_{t=1}^{n} u_{t-1}^2}(\hat{\beta}_{n,Mh} - \beta) = \frac{M_n}{\sqrt{V_n}},$$

где $M_n = d_n^{-1}(\beta) \sum\limits_{t=1}^{n} \varepsilon_t u_{t-1}, \; V_n = d_n^{-2}(\beta) \sum\limits_{t=1}^{n} u_{t-1}^2$

1. $|\beta| < 1$. Тогда $(M_n, V_n)^T \xrightarrow{d} (\xi, 1)^T \Longrightarrow \dfrac{M_n}{\sqrt{V_n}} \xrightarrow{d} \dfrac{\xi}{\sqrt{1}} \sim N(0,1)$.
2. $|\beta| > 1$. Тогда $(M_n, V_n)^T \xrightarrow{d} (\xi\eta, \eta^2)^T \Longrightarrow \dfrac{M_n}{\sqrt{V_n}} \xrightarrow{d} \dfrac{\xi\eta}{\sqrt{\eta^2}} = \xi sign(\eta) \sim N(0,1)$.
3. $|\beta| = 1$. Тогда $(M_n, V_n)^T \xrightarrow{d} (\frac{1}{\sqrt{2}}(W^2(1) - 1), 2\int\limits_0^1 W^2(s)ds)^T \Longrightarrow \dfrac{M_n}{\sqrt{V_n}} \xrightarrow{d} \dfrac{W^2(1) - 1}{2\sqrt{\int\limits_0^1 W^2(s)ds}}$.

■

13.4 Об оценке наименьших квадратов в авторегрессии

Если $\{\varepsilon_t\}$ в AR(1) уравнении

$$u_t = \beta u_{t-1} + \varepsilon_t, \; u_0 = 0, \; t = 1, 2, \ldots, \; \beta \in \mathbb{R}^1, \tag{13}$$

есть н.о.р. $N(0,1)$ случайные величины, то о.н.к. - решение задачи

$$\sum_{t=1}^{n} (u_t - \theta u_{t-1})^2 \to \min_{\theta \in \mathbb{R}^1}.$$

Если же $\{\varepsilon_t\}$ – н.о.р. с неизвестным распределением, то задача (14) определяет о.н.к.

$$\hat{\beta}_{n,hS} = \frac{\sum\limits_{t=1}^{n} u_{t-1} u_t}{\sum\limits_{t=1}^{n} u_{t-1}^2}.$$

О.н.к. $\hat{\beta}_{n,hS}$ – непараметрическая!

13.5 Теорема об AR(1) с $|\beta| < 1$, существование, единственность и свойства стационарного решения

Теорема 13.5 (**Теорема об AR(1) с $|\beta| < 1$, существование, единственность и свойства стационарного решения**). *Пусть* $u_t = \beta u_{t-1} + \varepsilon_t,\ |\beta| < 1,\ t \in \mathbb{Z}$. *Если* $\{\varepsilon_t\}$ *– н.о.р.,* $E\varepsilon_1 = 0,\ 0 < E\varepsilon_1^2 < \infty$, *то*

$$n^{\frac{1}{2}}(\hat{\beta}_{n,hS} - \beta) \xrightarrow{d} N(0, 1 - \beta^2),\ n \to \infty$$

Замечание. *Сделаем несколько замечаний:*

1. *Если* $|\beta| = 1$, *то при* $E\varepsilon_1 = 0,\ 0 < E\varepsilon_1^2 < \infty,\ \{\varepsilon_t\}$ *– н.о.р. в схеме (13), то*

$$d_n(\beta)(\hat{\beta}_{n,hS} - \beta) \xrightarrow{d} \tilde{H}(\beta),\ n \to \infty$$

2. *Если* $|\beta| > 1$, *то при* $E\varepsilon_1 = 0,\ 0 < E\varepsilon_1^2 < \infty,\ \{\varepsilon_t\}$ *– н.о.р. в схеме (13), то*

$$d_n(\beta)(\hat{\beta}_{n,hS} - \beta) \xrightarrow{d} \frac{\sum\limits_{j\geq 1} \beta^{-j}\varepsilon_j}{\sum\limits_{j\geq 1} \beta^{-j}\varepsilon'_j},$$

$\{\varepsilon_t\},\ \{\varepsilon'_t\}$ *– независимые последовательности с н.о.р. компонентами.*

Рассмотрим стационарное AR(1) уравнение

$$u_t = \beta u_{t-1} + \varepsilon_t,\ t \in \mathbb{Z},\ |\beta| < 1, \tag{15}$$

$\{\varepsilon_t\}$ – независимые н.о.р., $E\varepsilon_1 = 0,\ 0 < E\varepsilon_1^2 < \infty$.

Определение 13.2. *Любая последовательность* $\{u_t\}$*, для которой в (15) левая часть равна правой почти наверное, называется решением уравнения (15).*

Определение 13.3 (**Стационарность в узком смысле (строгая стационарность)**). *Случайный процесс* $\{x(t)\}$ *называется стационарным случайным процессом в узком смысле, если* $\forall\ n \in N,\ \forall\ t_i, \tau:\ t_i, t_i + tau \in T$ *выполнено условие*

$$F(x_{t_1}, \dots, x_{t_n}) = F(x_{t_1+\tau}, \dots, x_{t_n+\tau})$$

Определение 13.4 (**Стационарность в широком смысле**). *Случайный процесс* $\{x(t)\}$ *называется стационарным случайным процессом в широком смысле, если*

1. $x(t) \in L_2(d\mathbb{P}) \; \forall \, t \in T$
2. $\forall \, t, s \in T, \; \forall \, h: \; t+h, s+h \in T$ *выполнены условия*

$$Ex(t+h) = Ex(t), \; Cov(x(t+h), x(s+h)) = Cov(x(t), x(s))$$

Напоминание 13.2.

Для доказательства теоремы 13.5 сформулируем следующую теорему.

Теорема 13.6. *При* $|\beta| < 1 \; \exists$ *почти наверное единственное строго стационарное решение уравнения (15). Оно имеет вид*

$$u_t = \sum_{j \geq 0} \beta^j \varepsilon_{t-j}, \tag{16}$$

ряд сходится в средне квадратическом (с.к.), то есть сходится в L^2*. Решение (16) является также стационарным в широком смысле, причем*

$$Eu_t = 0, \; R(\tau) = Cov(u_t, u_{t+\tau}) = \frac{\sigma^2 \beta^{|\tau|}}{1 - \beta^2}$$

Доказательство. Доказательство теоремы 13.6.

1. Существование.

   Пусть $u_t^{(n)} := \sum\limits_{j=0}^{n} \beta^j \varepsilon_{t-j}$ – астная сумма ряда (16). Ряд с.к. сходится, если для некоторой сл.в. S_t, $ES_t^2 < \infty \; \exists \lim_{n \to \infty} u_t^{(n)} = S_t$, S_t есть сумма ряда (т.е. $E|u_t^{(n)} - S_t|^2 \to 0, \; n \to \infty$). Из критерия Коши извесно, что эта с.к. сходимость эквивалентна с.к. фундаментальности, т.е.

$$\lim_{n,m \to \infty} E|u_t^{(n)} - u_t^{(m)}|^2 = 0.$$

   Пусть $l = min(n, m), \; k = max(n, m)$. Тогда

$$E|u_t^{(n)} - u_t^{(m)}| = E|\sum_{j=l+1}^{k} \beta^j \varepsilon_{t-j}|^2 = \delta^2 \sum_{j=l+1}^{k} \beta^j \to 0,$$

   т.к. $l, k \to \infty, \; |\beta| < 1$. Значит, ряд с.к. сходится. Имеем почти наверное

$$u_t = \sum_{j \geq 0} \beta^j \varepsilon_{t-j} = \varepsilon_t + \beta \sum_{j \geq 1} \beta^{j-1} \varepsilon_{t-j} = \varepsilon_t + \beta \sum_{s \geq 0} \beta^s \varepsilon_{t-s-1} = \varepsilon_t + \beta u_{t-1}.$$

Значит, $\{u_t\}$ из (16) есть решение (15).

2. Строгая стационарность.

Пусть $U(\tau) = (u_{t_1+\tau},\ u_{t_k+\tau})$. Надо показать, что $U(\tau) \stackrel{d}{=} U(0)$. Пусть $U_n(\tau) = (u^{(n)}_{t_1+\tau},\ u^{(n)}_{t_k+\tau})$.

Задача 13.1. *Если $\{\xi_t\}$ – строго стационарная последовательность, а $\eta_t = f(\xi_t, \ldots, \xi_{t-k})$, f – борелевская, то $\{\eta_t\}$ – строго стационарная последовательность.*

В силу этой задачи $\{u^{(n)}_t\}$ – строго стационарная последовательность (т.к. $\{\varepsilon_t\}$ – строго стационарная последовательность, $\{u^{(n)}_t\}$ – последовательность частных сумм), то есть распределение вектора $U_n(\tau)$ от τ не зависит. Но

$$U_n(\tau) \xrightarrow{d} U(\tau),\ n \to \infty, \tag{17}$$

т.к. $u^{(n)}_t \xrightarrow{\text{с.к.}} u_t$. Значит, в силу (17), распределение $U(\tau)$ от τ не зависит.

3. Единственность.

Пусть $\{\tilde{u}_t\}$ – любое строго стационарное решение (15). Тогда почти наверное $\tilde{u}_t = \beta\tilde{u}_{t-1} + \varepsilon_t = \varepsilon_t + \beta\varepsilon_{t-1} + \ldots + \beta^{k-1}\varepsilon_{t-k+1} + \beta^k\tilde{u}_{t-k}$. Имеем

$$P(|\beta^k\tilde{u}_{t-k}| > \delta) = \beta^k\tilde{u}_0| > \delta) \to 0,\ k \to \infty,$$

т.к. распределение $\tilde{u}_k$ не зависит от времени, $|\beta| < 1$. Знаем, что $u^{(k)}_t \xrightarrow{\text{с.к.}} u_t = \sum\limits_{j\geq 0} \beta^j\varepsilon_{t-j},\ Eu_t^2 < \infty$. Значит, $u^{(k)}_t + \beta^k\tilde{u}_{t-k} \xrightarrow{P} u_t,\ k \to \infty$, так как, раз $u^{(k)}_t \xrightarrow{\text{с.к.}} u_t$, то $u^{(k)}_t \xrightarrow{P} u_t$, $\beta^k\tilde{u}_{t-k} \xrightarrow{P} 0$. Следовательно, п.н. $\tilde{u}_t = \lim_{k\to\infty}(u^k_t + \beta^k\tilde{u}_{t-k}) = u_t = \sum\limits_{j\geq 0} \beta^j\varepsilon_{t-j}$.

4. Стационарность в широком смысле.

Последовательность $\{u_t\}$ из (16) стационарна в широком смысле, так как она стационарна в узком смысле и есть моменты до 2-ого порядка. Тогда из (15) $Eu_t = \beta Eu_{t-1} + E\varepsilon_t,\ E\varepsilon_t = 0, Eu_t = Eu_0$, так как она строго стационарна, тогда возьмем $u_0 :\ (1-\beta)Eu_0 = 0 \Longrightarrow Eu_0 = 0$. Найдем дисперсию:

$$Eu_t^2 = \beta^2 Eu_{t-1}^2 + 2\beta E(u_{t-1}\varepsilon_t) + E\varepsilon_t^2, \text{ то есть } (1-\beta^2)Eu_0^2 = (1-\beta^2)R(0) = \\ = E\varepsilon_t^2 = \delta^2 \Longrightarrow R(0) = \frac{\delta^2}{1-\beta^2}.$$

Для $\tau > 0$ $Eu_{t+\tau}u_t = \beta Eu_{t+\tau-1}u_t + E\varepsilon_{t+\tau}u_t . E\varepsilon_{t+\tau}u_t = E\varepsilon_{t+\tau}Eu_t = 0$, т.к. $\varepsilon_{t+\tau},\ u_t$ независимы (взяли $u_{t+\tau} = \beta u_{t+\tau-1} + \varepsilon_{t+\tau}$, умножили равенство на u_t, взяли мат.ожидание от равентсва). Получаем, что

$$R(\tau) = \beta R(\tau - 1),\ R(0) = \frac{\sigma^2}{1-\beta^2} \Longrightarrow R(\tau) = \frac{\sigma^2\beta^\tau}{1-\beta^2},$$

а т.к. $R(\tau)$ – четная, то $\forall\ \tau\ R(\tau) = \dfrac{\sigma^2\beta^{|\tau|}}{1-\beta^2}$.

∎

13.6 Замечания о последовательностях с сильным перемешиванием (с.п.)

Определение 13.5 (**Условие сильного перемешивания**). *Пусть $\{u_t\}$, $t \in \mathbb{Z}$, – строго стационарная последовательность. Если*

$$\alpha(\tau) := \sup_{\substack{A \in M_{-\infty}^0, \\ B \in M_\tau^\infty}} |P(AB) - P(A)P(B)| \to 0,\ \tau \to \infty,$$

то $\{u_t\}$ удовлетворяет условию сильного перемешивания с коэффициентом перемешивания $\alpha(\tau)$. Здесь $M_a^b = \sigma(u_t, a \le t \le b)$.

Пример 13.3. *Приведем несколько примеров:*

1. *$\{\varepsilon_t\}$ – н.о.р.сл.в., здесь $\alpha(\tau) = 0,\ \tau > 0$, т.к. $\{\varepsilon_t\}$ независимы;*
2. *(Скользящее среднее порядка q) $u_t = \varepsilon_t + \alpha_1\varepsilon_{t-1} + \ldots + \alpha_q\varepsilon_{t-q}$, $\{\varepsilon_t\}$ – н.о.р., здесь $\alpha(\tau) = 0,\ \tau > q$;*
3. *(ARIMA(p, 0, 0)) $u_t = \beta_1 u_{t-1} + \ldots + \beta_p u_{t-p} + \varepsilon_t$, $\{\varepsilon_t\}$ – н.о.р., ε_1 имеет Лебегову плотность вероятности, $E\varepsilon_1 = 0$, $E\varepsilon_1^2 < \infty$, строго стационарное решение $\{u_t\}$ удовлетворяет условию с.п. с коэффициентом $\alpha(\tau) \le c\lambda^\tau$, $0 < \lambda < 1$. Это результата Mokkadem, 1998.*

Задача 13.2. *Если $\{u_t\}$ удовлетворяет условию с.п. с коэффициентом $\alpha(\tau)$, а $\eta_t = f(u_t, \ldots, u_{t-k})$, то η_t удовлетворяет условию с.п. с коэффициентом $\alpha_\tau \le \alpha(t - \tau)$, $\tau > k$, f – борелевская (пояснение устное: просто сигма-алгебры сдвигаются на τ).*

ЗБЧ (док-ва не было)

Если $\{u_t\},\ t \in \mathbb{Z},$ – строго стационарная последовательность с с.п., $E|u_1| < \infty \Longrightarrow n^{-1}\sum\limits_{t=1}^{n} u_t \overset{\text{п.н.}}{\to} Eu_1,\ n \to \infty.$

ЦПТ (Ибрагимов, Ленник, независимые и стационарные сл.в., Т 18.5.3., у нас док-ва не было)

Пусть $\{u_t\},\ t \in \mathbb{Z},$ – строго стационарная последовательность с с.п., $E|u_1| = 0,\ E|u_1|^{2+\delta} < \infty,$ при некотором $\delta > 0$. Пусть $\sum\limits_{\tau \geq 1}(\alpha(\tau))^{\frac{2}{2+\delta}} < \infty$, тогда:

1. Ряд $\Delta^2 = Eu_0^2 + 2\sum\limits_{\tau\geq 1} Eu_0u_\tau$ сходится абсолютно;
2. Если $\Delta^2 > 0$, то $n^{-\frac{1}{2}}\sum\limits_{t=1}^{n} u_t \overset{d}{\to} N(0, \Delta)$.

Следствие 13.1. *Если $\{u_t\}$ удовлетворяет условию с.п., $Eu_1 = m,\ E|u_1 - m|^{2+\delta} < \infty,\ \sum\limits_{\tau\geq 1}(\alpha(\tau))^{\frac{2}{2+\delta}} < \infty,$ то Δ^2 из пункта 1 ЦПТ можно переписать как $\bar{\Delta}^2 = Du_0 + 2\sum\limits_{\tau\geq 1} R(\tau)$,то при $\bar{\Delta} > 0$ по ЦПТ имеем*

$$\sup_x |P(n^{\frac{1}{2}}(\bar{u} - m) \leq x) - \varphi()\frac{x}{\bar{\Delta}}| \to 0,\ n \to \infty.$$

Напоминание 13.3. *Хотим доказать теорему 13.5. Мы рассматриваем AR(1)-модель*

$$u_t = \beta u_{t-1} + \varepsilon_y,\ t \in \mathbb{Z}, \tag{18}$$

в которой $\{\varepsilon_t\}$ – н.о.р., $E\varepsilon_1 = 0,\ 0 < \sigma^2 = E\varepsilon_1^2 < \infty,\ |\beta| < 1$. Пусть функция распределения $\varepsilon_1 = G(x),\ g(x) = G(x)$ – плотность, $G(x), g(x)$ – неизвестны. Пусть наблюдения $u_0, u_1, \ldots, u_n$ – выборка из стационарного решения AR(1)-уравнения. В качестве оценки неизвестного параметра β берем о.н.к, которая получена из решения задачи

$$\sum_{t=1}^{n}(u_t - \theta u_{t-1})^2 \to \min_\theta.$$

Обозначили эту оценку $\hat{\beta}_{n,hS}$. Очевидно, что

$$\hat{\beta}_{n,hS} = \frac{\sum\limits_{t=1}^{n} u_t u_{t-1}}{\sum\limits_{t=1}^{n} u_{t-1}^2}. \tag{18$'$}$$

аша ближайшая цель – доказать теорему 13.5, в силу которой при $|\beta| < 1,\ E\varepsilon_1 = 0,\ 0 < E\varepsilon_1^2 < \infty$ *имеем*

$$n^{\frac{1}{2}}(\hat{\beta}_{n,hS} - \beta) \xrightarrow{d} N(0, 1 - \beta^2),\ n \to \infty$$

13.7 Доказательство теоремы об AR(1) с $|\beta| < 1$

Теорема 13.7 (Теорема об AR(1) с $|\beta| < 1$, существование, единственность и свойства стационарного решения.)**.** *Пусть* $u_t = \beta u_{t-1} + \varepsilon_t,\ |\beta| < 1,\ t \in \mathbb{Z}$. *Если* $\{\varepsilon_t\}$ *– н.о.р.,* $E\varepsilon_1 = 0,\ 0 < E\varepsilon_1^2 < \infty$, *то*

$$n^{\frac{1}{2}}(\hat{\beta}_{n,hS} - \beta) \xrightarrow{d} N(0, 1 - \beta^2),\ n \to \infty$$

Доказательство. Доказательство теоремы 13.7. Сначала проверим условия ЦПТ.

Предположим, что $E|\varepsilon_1|^{2+\delta} < \infty$ при некотором $\delta > 0$. Пусть также $\exists$ плотность вероятности для $\varepsilon_1:\ g(x)$ по мере Лебега.

1. При $|\beta| < 1$ в силу теоремы 13.6 существует строго стационарное решение уравнения AR(1), оно имеет вид

$$u_t = \sum_{j\geq 1} \beta^j \varepsilon_{t-j},$$

ряд сходится в L^2. Покажем, что этот ряд сходится также в $L^{2+\delta}$, тогда $E|u_1|^{2+\delta} < \infty$.
Справедливо неравенство Миньковского: если $E|\xi|^{2+\delta} < \infty,\ E|\eta|^{2+\delta} < \infty,\ \delta > 0$, то

$$\{E|\xi + \eta|^{2+\delta}\}^{\frac{1}{2+\delta}} \leq \{E|\xi|^{2+\delta}\}^{\frac{1}{2+\delta}} + \{E|\eta|^{2+\delta}\}^{\frac{1}{2+\delta}},$$

это было неравенство треугольника. Рассмотрим частную сумму $S_n = \sum\limits_{j=1}^{n} \beta_j \varepsilon_{t-j}$.

$$\{E|S_n - S_m|^{2+\delta}\}^{\frac{1}{2+\delta}} = \{E|\sum_{j=min(n,m)+1}^{max(n,m)} \beta_j \varepsilon_{t-j}|^{2+\delta}\}^{\frac{1}{2+\delta}} \leq$$

$$\leq \sum_{j=min(n,m)+1}^{max(n,m)} \{E|\beta_j \varepsilon_{t-j}|^{2+\delta}\}^{\frac{1}{2+\delta}} = E\{|\varepsilon_1|^{2+\delta}\}^{\frac{1}{2+\delta}} \sum_{j=min(n,m)+1}^{max(n,m)} |\beta|^j \to 0,$$

при $|\beta| < 1$ $(min(m, n) \to \infty)$.

Значит, последовательность частных сумм $\{S_n\}$ – фундаментальная последовательность, и ряд $u_t = \sum\limits_{j\geq 1} \beta^j \varepsilon_{t-j}$ ряд сходится в $L^{2+\delta} \Longrightarrow E|u_1|^{2+\delta} < \infty$.

2. Знаем, что

$$\hat{\beta}_{n,hS} = \frac{\sum\limits_{t=1}^{n} u_t u_{t-1}}{\sum\limits_{t=1}^{n} u_{t-1}^2} \overset{\text{п.н.}}{=} \beta + \frac{\sum\limits_{t=1}^{n} \varepsilon_t u_{t-1}}{\sum\limits_{t=1}^{n} u_{t-1}^2},$$

$$n^{\frac{1}{2}}(\hat{\beta}_{n,hS} - \beta) = n^{\frac{1}{2}} \frac{\sum\limits_{t=1}^{n} \varepsilon_t u_{t-1}}{\sum\limits_{t=1}^{n} u_{t-1}^2}.$$

3. В силу результатов Mokkadem 13.3 последовательность $\{u_t\}$ удовлетворяет условию с.п. с $\alpha(\tau) \leq c\lambda^\tau$, $0 < \lambda < 1$. Последовательность $\{\varepsilon_t u_{t-1} = (u_t - \beta u_{t-1})u_{t-1} = f(u_t, u_{t-1})\}$ тоже удовлетворяет условию с.п. с $\alpha'(\tau) \leq c'\lambda^\tau \Longrightarrow$

$$\sum_{\tau\geq 1} (\alpha'(\tau))^{\frac{\delta}{2+\delta}} \leq \sum_{\tau \geq 1} (c'\lambda^\tau)^{\frac{\delta}{2+\delta}} = \frac{(c'\lambda)^{2+\delta}}{1-\lambda^{2+\delta}} < \infty.$$

$$E\varepsilon_t u_{t-1} = E\varepsilon_t E u_{t-1} = 0,\ E|\varepsilon_t u_{t-1}|^{2+\delta} = E|\varepsilon_t|^{2+\delta} E|u_{t-1}|^{2+\delta} < \infty,$$

тгда в силу ЦПТ для последовательности с с.п. имеем

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} \varepsilon_t u_{t-1} \xrightarrow{d} N(0, \Delta^2),$$

где

$$\Delta^2 = E(\varepsilon_1 u_0)^2 + 2\sum_{\tau\geq 1} \underbrace{E(\varepsilon_1 u_0 \varepsilon_{1+\tau} u_\tau)}_{=0} = E\varepsilon_1^2 E u_0^2$$

4. $n^{-1} \sum\limits_{t=1}^{n} u_{t-1}^2 \xrightarrow{\text{п.н.}} E u_0^2$ в силу ЗБЧ для последовательности с с.п.

Значит,

$$n^{\frac{1}{2}}(\hat{\beta}_{n,hS} - \beta) \xrightarrow{d} \frac{1}{E u_0^2} N(0, E\varepsilon_1^2 E u_0^2),$$

$$\frac{E\varepsilon_1^2 E u_0^2}{(E u_0^2)^2} = \frac{E\varepsilon_1^2}{E u_0^2} = 1 - \beta^2.$$

■

13.8 Ассимптотические доверительные интервалы

В силу (18')

$$\frac{n^{\frac{1}{2}}(\hat{\beta}_{n,hS}-\beta)}{\sqrt{1-\hat{\beta}^2_{n,hS}}} = \frac{n^{\frac{1}{2}}(\hat{\beta}_{n,hS}-\beta)}{\sqrt{1-\beta^2}}\frac{\sqrt{1-\beta^2}}{\sqrt{1-\hat{\beta}^2_{n,hS}}} \xrightarrow{d} N(0,1),\ n\to\infty,$$

т.к.

$$\frac{n^{\frac{1}{2}}(\hat{\beta}_{n,hS}-\beta)}{\sqrt{1-\beta^2}} \xrightarrow{d} N(0,1),\ \frac{\sqrt{1-\beta^2}}{\sqrt{1-\hat{\beta}^2_{n,hS}}} \xrightarrow{P} 1.$$

Применим лемму Слуцкого. Пусть $\xi_{1-\frac{\alpha}{2}}$ – квантиль уровня $1-\frac{\alpha}{2}$ функции распределения $\Phi(x)\sim N(0,1)$, тогда

$$P(|\frac{n^{\frac{1}{2}}(\hat{\beta}_{n,hS}-\beta)}{\sqrt{1-\hat{\beta}^2_{n,hS}}}|<\xi_{1-\frac{\alpha}{2}})\to 1-\alpha,\ n\to\infty,$$

т.е. при больших n примерно с вероятностью $1-\alpha$

$$\hat{\beta}_{n,hS}-\sqrt{\frac{1-\hat{\beta}^2_{n,hS}}{n}}\xi_{1-\frac{\alpha}{2}}<\beta<\hat{\beta}_{n,hS}+\sqrt{\frac{1-\hat{\beta}^2_{n,hS}}{n}}\xi_{1-\frac{\alpha}{2}}$$

Получили доверительный интервал для β уровня $1-\alpha$.

13.9 Проверка гипотез

Проверим гипотезу $H_0:\beta=\beta_0$ против альтенативы $H_1:\beta\neq\beta_0$. Критическое множество

$$S_\alpha=\{u_0,u_1,\ldots:|\frac{n^{\frac{1}{2}}(\hat{\beta}_{n,hS}-\beta_0)}{\sqrt{1-\hat{\beta}^2_{n,hS}}}|<\xi_{1-\frac{\alpha}{2}}\}.$$

тогда, очевидно, что $P(H_1|H_0)\to\alpha$, а т.к. при H_1 :

$$\frac{n^{\frac{1}{2}}(\hat{\beta}_{n,hS}-\beta_0)}{\sqrt{1-\beta_0}} = \frac{n^{\frac{1}{2}}(\hat{\beta}_{n,hS}-\beta_0)}{\sqrt{1-\beta}} + \frac{n^{\frac{1}{2}}(\beta-\beta_0)}{\sqrt{1-\beta_0^2}} \xrightarrow{P} \infty,\ n\to\infty,$$

то $P(H_0|H_1)$. Значит,

$$\begin{cases} P(H_0|H_0) \to 1-\alpha \\ P(H_1|H_1) \to 1, \ n \to \infty \end{cases}$$

Вероятность принять правильную гипотезу близка к единице!

13.10 О робастности о.н.к.

$u_t = \beta u_{t-1} + \varepsilon_t, \ t \in \mathbb{Z}, \ |\beta| < 1, \ \{\varepsilon_t\}$ – н.о.р., $E\varepsilon_1 = 0, \ 0 < \varepsilon_1^2 < \infty$. Пусть наблюдаются $y_t = u_t + z_t^\gamma \xi_t, \ t = 0, 1, \ldots, n, \ \{u_t\}$ – стационарное решение, $\{z_t^\gamma\}$ – н.о.р., $z_1^\gamma \sim Br(\gamma), \ 0 \le \gamma \le 1, \ \{\xi_t\}$ – н.о.р., $\xi_1 \sim \mu_\xi, \ \mu_\xi \in M_2 : \ E\xi_1^2 < \infty$, последовательности $\{\xi_t\}, \{u_t\}, \{z_t^\gamma\}$ независимы между собой. Пусть

$$\hat{\beta}^y_{n,hS} = \frac{\sum\limits_{t=1}^n y_{t-1}y_t}{\sum\limits_{t=1}^n y_{t-1}^2}$$

– о.н.к., построенная по засоренным данным $\{y_t\}$. Найдем ее функционал влияния.

Первый способ. Предположим дополнительно, что у ε_1 $\exists$ плотность, вероятности $g(x) = G'(x)$. Тогда последовательность $\{u_t\}$ удовлетворяет условию с.п., а т.к. $\{z_t^\gamma \xi_t\}, \ t \in \mathbb{Z}$, – последовательность н.о.р.сл.в., которые не зависят от $\{u_t\}$, то $\{y_t\}$ строго стационарная последовательность с с.п. Кроме того, $E|y_1| < \infty$, т.к.

$$Ey_1^2 = E(u_1 + z_1^\gamma \xi_1)^2 = Eu_1^2 + 2\underbrace{Eu_1}_{=0}Ez_1^\gamma E\xi_1 + E(z_1^\gamma\xi_1)^2 = Eu_1^2 + \gamma E\xi_1^2 < \infty.$$

Значит, в силу ЗБЧ для последовательностей с с.п.

$$\hat{\beta}^y_{n,hS} = \frac{n^{-1}\sum\limits_{t=1}^n y_{t-1}y_t}{n^{-1}\sum\limits_{t=1}^n y_{t-1}^2} \xrightarrow{P} \theta_\gamma^{hS} = \frac{Ey_0y_1}{Ey_0^2} = \frac{E(u_0 + z_0^\gamma\xi_0)(u_1 + z_1^\gamma\xi_1)}{E(u_0 + z_0^\gamma\xi_0)^2} = \frac{Eu_0u_1 + \gamma^2(E\xi_0)^2}{Eu_0^2 + \gamma E\xi_0^2}$$

$$\Longrightarrow IF(\theta_\gamma^{hS}, \mu_\xi) = \frac{d\theta_\gamma^{hS}}{d\gamma}|_{\gamma=0} = -\beta(1-\beta^2)\frac{E\xi_0^2}{E\varepsilon_1^2}.$$

Если M_2 – множество распределений с конечным вторым моментом, то при $\beta \ne 0$ $GES(\theta_\gamma^{hS}, M_2) = \sup\limits_{\mu_\xi \in M_2} |IF(\theta_\gamma^{hS}, \mu_\xi)| = \infty \Longrightarrow$ о.н.к. неробастна!

Второй способ. Предположим, что ε_1 имееет плотность вероятности $g(x)$ по норме Лебега. Тогда $\{y_t\}$ удовлетворяет условию с.п. Оценка $\hat{\beta}^y_{n,hS}$ – корень уравнения

$$l_{n,hS}(\theta) = n^{-1}\sum_{t=1}^{n} y_{t-1}(y_t - \theta y_{t-1}) = 0.$$

1. $l_{n,hS}(\theta) \xrightarrow{P} Ey_0(y_1 - \theta y_0),\ \forall\theta,\ 0 \le \gamma \le 1$. Т.е. $\Lambda_{hS}(\gamma,\theta) = Ey_0(y_1 - \theta y_0)$. Пусть $H_{00} = (z_0^\gamma = 0, z_1^\gamma = 0), H_{01} = (z_0^\gamma = 0, z_1^\gamma = 1), H_{10} = (z_0^\gamma = 1, z_1^\gamma = 0), H_{00} = (z_0^\gamma = 1, z_1^\gamma = 1)$. Тогда

$$\Lambda_{hS}(\gamma,\gamma) = \sum_{i,j=0}^{1} E(y_0(y_1 - \theta y_0)|H_{ij})P(H_{ij}) = (1-\gamma)^2 Eu_0(u_1 - \theta u_0) +$$
$$+(1-\gamma)\gamma Eu_0(u_1 + \xi_1 - \theta u_0) + \gamma(1-\gamma)E(u_0 + \xi_0)(u_1 - \theta u_0 - \theta\xi_0) +$$
$$+\gamma^2 E(u_0 + \xi_0)(u_1 + \xi_1 - \theta u_0 - \theta\xi_1) \Longrightarrow$$

   $\Lambda_{hS}(\gamma,\theta)$ определена $\forall\ \gamma,\theta$.
2. $\Lambda_{hS}(0,\beta) = Eu_0(u_1 - \beta u_0) = Eu_0\varepsilon_1 = 0$.
3. $\dfrac{\partial\Lambda_{hS}(\gamma,\theta)}{\partial\gamma}, \dfrac{\partial\Lambda_{hS}(\gamma,\theta)}{\partial\theta}$ существуют и непрерывны по паре $(\gamma,\theta),\ \gamma,\theta \in \mathbb{R}^1$,

$$\frac{\partial\Lambda_{hS}(\gamma,\theta)}{\partial\gamma} = -\beta E\xi_0^2,\ \frac{\partial\Lambda_{hS}(\gamma,\theta)}{\partial\theta} = -Eu_0^2.$$

4. $\lambda(\beta) = -Eu_0^2 = -\dfrac{E\varepsilon_1^2}{1-\beta^2} < 0$, т.к. $\varphi(\mathbb{I},\theta)$ непрерывна, то

$$IF(\theta_\gamma^{hS}, \mu_\xi) = -(\frac{\partial\Lambda_{hS}(0,\beta)}{\partial\theta})^{-1}\frac{\partial\Lambda_{hS}(0,\beta)}{\partial\gamma} = -\frac{-\beta E\xi_0^2}{-\dfrac{E\varepsilon_1^2}{1-\beta^2}} = -\beta(1-\beta^2)\frac{E\xi_0^2}{E\varepsilon_1^2}.$$

   Очевидно, при $\beta \ne 0$ $GES(\theta_\gamma^{hS}, M_2) = \infty$, т.е. $\hat{\beta}^y_{n,hS}$ не β-робастна.

Задача 13.3. *$u_t = \beta u_{t-1} + \varepsilon_t,\ t \in \mathbb{Z},\ |\beta| < 1,\ \beta \ne 0$ $\{\varepsilon_t\}$ – н.о.р., $E\varepsilon_1 = 0,\ 0 < \varepsilon_t^2 < \infty$.*

$y_t = u_t + z_t^\gamma \xi_t$. Оценка $\hat{\beta}_n$ ищется как корень уравнения

$$\sum_{t=1}^{n} y_{t-2}(y_t - \theta y_{t-1}) = 0$$

1. *Будет ли оценка $\hat{\beta}_n$ β-робастной?*

2. *Чему равен функционал влияния второго порядка?*

13.11 О процедурах наименьших квадратов в AR(p)

Определение 13.6. *Авторегрессия порядка p (AR(p) - модель) описывается стохастическим разностным уравнением*

$$u_t = \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_p u_{t-p} + \varepsilon_t,\ t \in \mathbb{Z}. \tag{20}$$

$\{\varepsilon_t\}$ – *н.о.р.,* $E\varepsilon_1 = 0,\ 0 < \varepsilon_1^2 = \sigma^2 < \infty,\ \beta_1, \ldots, \beta_p \in \mathbb{R}$ – *неизвестные коэффициенты авторегрессии.*

Определение 13.7. *Уравнение*

$$x^p = \beta_1 x^{p-1} + \ldots + \beta_p \tag{21}$$

называется характеристическим уравнением, соответствующим уравнению (20).

Теорема 13.8. *Пусть корни характеристического уравнения (21) по модулю меньше 1. Тогда уравнение (20) имеет почти наверное единственное строго стационарное решение. Ряд*

$$u_t = \sum_{j \geq 0} \gamma_j varepsilon_{t-j} \tag{22}$$

ходится с.к. Коэффициенты $\{\gamma_t\}$ *определяются рекурентным соотношением*

$$\gamma_j = \beta_1 \gamma_{j-1} + \ldots + \beta_p \gamma_{j-p}$$

$j = 1, 2, \ldots,\ \gamma_0 = 1,\ \gamma_j = 0$ *при* $j < 0,\ |\gamma_j| \leq c\lambda^j,\ 0 < \lambda < 1.$

Ряд (22) также определяет стационарную в широком смысле последовательность с нулевым средним.

Доказательство. Для p = 1 условие теоремы совпадает с утверждением 13.6. Для p > 1доказательство идейно не отличается, тно технически громоздко, мы его опускаем. ■

Следствие 13.2. *Поскольку из (22)* $u_t = u_t(\varepsilon_t, \varepsilon_{t-1}, \ldots)$, *то сл.в.* ε_{t+1} *не зависит от множества сл.в.* $\{u_t, u_{t-1}, \ldots\}$.

13.12 Прогнозирование

Пусть наблюдения $u_1, u_2, \ldots, u_n$ будут выборкой из стационарного решения (22). Пусть $n \geq p$. птимальный с.к. прогноз ненаблюдаемой величины u_{n+1} по наблюдениям $u_1, u_2, \ldots, u_n$ есть решение задачи

$$E|u_{n+1} - \varphi(u_1, u_2, \ldots, u_n)|^2 \to \underset{\text{бор.ф. } \varphi\colon E\varphi^2(u_1,\ldots,u_n)<\infty}{min}$$

Мы знаем, что решение этой задачи есть

$$u^*_{n+1} = \varphi^*(u_1, u_2, \ldots, u_n) = E(u_{n+1}|u_1, u_2, \ldots, u_n).$$

Имеем:

$$E(u_{n+1}|u_1, u_2, \ldots, u_n) = E(\sum_{j=1}^{p} \beta_j u_{n+1-j} + \varepsilon_{n+1}|u_1, u_2, \ldots, u_n) =$$

$$= E(\sum_{j=1}^{p} \beta_j u_{n+1-j}|u_1, u_2, \ldots, u_n) + E(\varepsilon_{n+1}|u_1, u_2, \ldots, u_n) = \beta_1 u_n + \ldots + \beta_p u_{n+1-p},$$

т.к. $E(\varepsilon_{n+1}|u_1, u_2, \ldots, u_n) = E(\varepsilon_{n+1}) = 0$ в силу замечания.

Итак, оптимальный с.к. прогноз

$$u^*_{n+1} = \beta_1 u_n + \ldots + \beta_p u_{n+1-p}.$$

Чтобы построить u^*_{n+1} надо оценить $\beta_1, \ldots, \beta_p$. Построим о.н.к. неизвестных $\beta_1, \ldots, \beta_p$ по наблюдениям $u_p, \ldots, u_{n-p}$. Положим, $\tilde{u}_{t-1} := (u_{t-1}, \ldots, u_{t-p})^T$, $\beta = (\beta_1, \ldots, \beta_p)^T \Longrightarrow$ (20) имеет вид

$$u_t = \tilde{u}_{t-1}\beta + \varepsilon_t,\ t \in \mathbb{Z}.$$

О.н.к. вектора β – решение задачи

$$\sum_{t=1}^{n} (u_t - \tilde{u}^T_{t-1}\theta)^2 \to \min_{\theta\in\mathbb{R}^1}, \tag{23}$$

$\theta = (\theta_1, \ldots, \theta_p)^T$. Решение задачи (23) совпадает с решением системы уравнений

$$\sum_{t=1}^{n} u_{t-j}(u_t - \tilde{u}^T_{t-1}\theta) = 0,\ j = 1, \ldots, p.$$

Запишем эту систему в векторном виде

$$\sum_{t=1}^{n} \tilde{u}_{t-1}(u_t - \tilde{u}_{t-1}^T \theta) = 0. \tag{23'}$$

Решение (23') есть

$$\hat{\beta}_{n,hS} = (\sum_{t=1}^{n} \tilde{u}_{t-1}\tilde{u}_{t-1}^T)^{-1} \sum_{t=1}^{n} \tilde{u}_{t-1}u_t.$$

Если $p \times p$-матрица $\sum\limits_{t=1}^{n} \tilde{u}_{t-1}\tilde{u}_{t-1}^T$ вырождена, то полагаем, что $\hat{\beta}_{n,hS} = 0$.

Оценка наименьших квадратов в AR(p). Формулировка теоремы об ассимптотической нормальности

Теорема 13.9 (Теорема об ассимптотической нормальности.). *Пусть $\{\varepsilon_t\}$ – н.о.р.сл.в., $E\varepsilon_1 = 0,\ 0 < \varepsilon_2^2 = \sigma^2 < \infty$. усть корни характеристического уравнения (21) по модулю меньше 1. Тогда*

$$n^{\frac{1}{2}}(\hat{\beta}_{n,hS} - \beta) \xrightarrow{d} N(0, \sigma^2\mathcal{K}^{-1}),\ n \to \infty,$$

где $\mathcal{K} = E\tilde{u}_0\tilde{u}_0^T > 0$.

Следствие 13.3. *В 13.9 речь идет о сходимости по распределению вектора $\xi_n = n^{\frac{1}{2}}(\hat{\beta}_{n,hS} - \beta)$ к гауссовскому вектору $\xi \sim N(0, \sigma^2\mathcal{K}^{-1})$. Напомним, что если $\xi_n, \xi \in \mathbb{R}^p$, то $\xi_n \xrightarrow{d} \xi$, если*

$$\int\limits_{\mathbb{R}^p} g(x)dP_n(x) \to \int\limits_{\mathbb{R}^p} g(x)dP(x) \tag{24}$$

$\forall$ непрерывной и ограниченной функции $g : \mathbb{R}^p \to \mathbb{R}^1$. $P_n,\ P$ – распределения векторов ξ_n, ξ. Можно проверить, что это определение равносильно следующему: для любой непрерывной и ограниченной функции $g : \mathbb{R}^p \to C$ выполнено (24).Разумеется, (24) означает, что $Eg(\xi_n) \to Eg(\xi)$. Мы знаем, что если $\xi_n \xrightarrow{d} \xi$, то для любого постоянного вектора $\lambda \in \mathbb{R}^p$

$$\lambda^T\xi_n \xrightarrow{d} \lambda^T\xi \tag{25}$$

Действительно, функция $\pi(x) := \lambda^T x$ непрерывна и (25) следует из Теоремы о наследовании слабой сходимости.

Верно и обратное: если выполнено соотношение (25), то $\xi_n \xrightarrow{d} \xi$. *Действительно, функция* $g(x) = e^{ix}$, $x \in \mathbb{R}^1$ *– непрерывная и ограниченная функция* $\mathbb{R} \to C$. *Тогда из (25) следует, что*

$$Eg(\xi_n) = Ee^{i\lambda^T \xi_n} \to Ee^{i\lambda^T \xi} = Eg(\xi) \ \forall\ \lambda \in \mathbb{R}^p.$$

Последние соотношения означают, что характеристическая функция вектора $\xi_n - -Ee^{i\lambda^T \xi_n}$ *и* $\forall\ \lambda$ *ходится к характеристической функции* $Ee^{i\lambda^T \xi}$ *вектора* $\xi \Longrightarrow \xi_n \xrightarrow{d} \xi$.

Лемма 13.1 (Прием Крамера-Уолда.)**.** *Если сл. векторы* $\xi_n, \xi \in \mathbb{R}^p$, *то* $\xi_n \xrightarrow{d} \xi \Longleftrightarrow \lambda^T \xi_n \xrightarrow{d} \lambda^T \xi \ \forall \lambda \in \mathbb{R}^p$.

Этот прием сводит сходимость по распределению векторов $\xi_n \xrightarrow{d} \xi$ к сходимости скаляров $\forall \lambda\ \lambda^T \xi_n \xrightarrow{d} \lambda^T \xi$.

Следствие 13.4. *Пусть* $\{\xi_t\}$ *– случайные вектора,* $\xi_n \in \mathbb{R}^k$. *Будем писать* $\xi_n = o_p(1)$ *(о-малое от 1 по вероятности),* $n \to \infty$, *если* $\xi_n \xrightarrow{P} 0$. *Будем говорить, что последовательность* $\{\xi_t\}$ *ограничена по вероятности (*$\xi_n = O_p(1)$*), если* $\forall\ \varepsilon > 0 \exists A = A(\varepsilon) : \ \sup\limits_n P(|\xi_n| > A) < \varepsilon$.

Задача 13.4.
1. *Если* $\xi_n = O_p(1)$, *а* $\eta_n = o_p(1)$, *то* $\xi_n \eta_n = o_p(1)$, $\xi_n \in \mathbb{R}^k$, $\eta_n \in \mathbb{R}^1$.
2. *Если* $\xi_n \xrightarrow{d} \xi$, *то* $\xi_n = O_p(1)$, $\xi_n, \xi \in \mathbb{R}$.

Оценка наименьших квадратов в AR(p). Доказательство теоремы об ассимптотической нормальности

Доказательство. Предположим, что $E|\varepsilon_1|^{2+\delta} < \infty$ для некоторого $\delta > 0$, $\exists g(x)$ – плотность вероятности по мере Лебега.

1. Покажем, что матрица $\mathcal{K} = E\tilde{u}_0 \tilde{u}_0^T > 0$. Для $\alpha \in \mathbb{R}^p$, $\alpha \neq 0$, имеем $\alpha^T \mathcal{K} \alpha = E(\alpha^T \tilde{u}_0)(\tilde{u}_0^T \alpha) = E|\alpha^T \tilde{u}_0|$. Но ряд

$$u_t = \sum_{s \geq 0} \gamma_s \varepsilon_{t-s} = \varepsilon_t + \sum_{s \geq 1} \gamma_s \varepsilon_{t-s}$$

сходится в $L^{2+\delta} \Longrightarrow$

$$\alpha^T \tilde{u}_0 = \alpha^T (u_{-1}, \ldots, u_{-p})^T = \alpha_1 \varepsilon_{-1} + \sum_{s \geq 1} \gamma_s \varepsilon_{-1} + \alpha_2 u_{-2} + \ldots + \alpha_p u_{-p}.$$

Сл.в. ε_{-1} абсолютно непрерывна и не зависит от остальных слагаемых. Значит, при $\alpha_1 \neq 0$ $\alpha^T \tilde{u}_0$ абсолютно непрерывна, $P(\alpha^T \tilde{u}_0 \neq 0) = 1$, и $E|\alpha^T \tilde{u}_0|^2 > 0$. Если $\alpha_1 = 0$, то повторим рассуждения для первой ненулевой компоненты вектора α.

2. Рассмотрим вектор $\mathcal{K}^{-1} n^{-\frac{1}{2}} \sum\limits_{t=1}^{n} \tilde{u}_{t-1}\varepsilon_t$. Покажем, что он сходится по распределению к $N(0, \sigma^2 \mathcal{K}^{-1})$. Для этого достаточно показать, что

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} \tilde{u}_{t-1}\varepsilon_t \xrightarrow{d} N(0, \sigma^2 \mathcal{K})$$

и применить Теорему о наследовании слабой сходимости. В силу леммы 13.1 достаточно проверить, что при $\forall\ \lambda \in \mathbb{R}^p$.

$$n^{-\frac{1}{2}} \sum_{t=1}^{n} \underbrace{\lambda^T \tilde{u}_{t-1}\varepsilon_t}_{=\eta_t} = n^{-\frac{1}{2}} \sum_{t=1}^{n} \eta_t \xrightarrow{d} N(0, \sigma^2 \lambda^T \mathcal{K} \lambda)$$

Последовательность η_t – функция от $\{u_t\}$, это строго стационарная последовательность с с.п., коэффициент перед $\alpha(\tau) \leq c\lambda^\tau,\ 0 < \lambda < 1$,

$$E\eta_t = E(\lambda^T \tilde{u}_{t-1}) E\varepsilon_t = 0,\ E|\eta|^{2+\delta} = E|\lambda^T \tilde{u}_{t-1}|^{2+\delta} E|\varepsilon_t|^{2+\delta} < \infty,$$

т.к. по условию $E|\varepsilon_1|^{2+\delta} < \infty$ (см. док-во теоремы 13.7), т.к.

$$\{E|\lambda^T \tilde{u}_{t-1}|^{2+\delta}\}^{\frac{1}{2+\delta}} \leq \sum_{i=1}^{p} |\lambda_i| \{E|\tilde{u}_1|^{2+\delta}\}^{\frac{1}{2+\delta}} < \infty$$

в силу неравенства Миньковского. Кроме того, при $t < s$

$$E\eta_t\eta_s = E\{(\lambda^T \tilde{u}_{t-1}\varepsilon_t) \times (\lambda^T \tilde{u}_{s-1})\}\varepsilon_s = 0.$$

Кроме того, $\sum\limits_{\tau \geq 1} (\alpha(\tau))^{\frac{2}{2+\delta}} < \infty$. силу ЦПТ для последовательностей с с.п.

$$n^{\frac{1}{2}} \sum_{t=1}^{n} \eta_t \xrightarrow{d} N(0, E\eta_0^2),\ E\eta_0^2 = \sigma^2 \lambda^T \mathcal{K} \lambda.$$

Соотношение (26) доказано $\Longrightarrow$

$$\mathcal{K}^{-1}n^{-\frac{1}{2}}\sum_{t=1}^{n}\tilde{u}_{t-1}\varepsilon_t \xrightarrow{d} N(0,\sigma^2\mathcal{K}^{-1}). \tag{27}$$

3. Пусть $\mathcal{K}_n := n^{-1}\sum\limits_{t=1}^{n}\tilde{u}_{t-1}\tilde{u}_{t-1}^T$. Если $det(\mathcal{K}_n) > 0$, то

$$\hat{\beta}_{n,hS} = \mathcal{K}_n^{-1}n^{-1}\sum_{t=1}^{n}\tilde{u}_{t-1}u_t = \mathcal{K}_n^{-1}n^{-1}\sum_{t=1}^{n}\tilde{u}_{t-1}(\tilde{u}_{t-1}^T\beta+\varepsilon_t) = \beta+$$

$$+\mathcal{K}_n^{-1}n^{-1}\sum_{t=1}^{n}\tilde{u}_{t-1}\varepsilon_t \Longrightarrow$$

$$\text{при невырожденном } \mathcal{K}_n : \ n^{\frac{1}{2}}(\hat{\beta}_{n,hS}-\beta) = \mathcal{K}_n^{-1}n^{-\frac{1}{2}}\sum_{t=1}^{n}\tilde{u}_{t-1}\varepsilon_t.$$

В силу ЗБЧ для последовательностей с с.п.

$$\mathcal{K}_n \xrightarrow{\text{п.н.}} \mathcal{K} = E\tilde{u}_0\tilde{u}_0^T > 0,\ det(\mathcal{K}) > 0 \Longrightarrow det(\mathcal{K}_n) > 0 \xrightarrow{\text{п.н.}} det(\mathcal{K}) > 0,$$

если $S_n := \{\omega :\ det(\mathcal{K}_n) > 0\}$, то $P(S_n) \to 1$. Напомним:

$$\begin{cases} \mathcal{K}_n^{-1}n^{-\frac{1}{2}}\sum\limits_{t=1}^{n}\tilde{u}_{t-1}u_t,\ \omega \in S_n, \\ 0,\ \omega \in \bar{S}_n. \end{cases}$$

Покажем, что

$$\gamma_n := n^{\frac{1}{2}}(\hat{\beta}_{n,hS}-\beta) - \mathcal{K}_n^{-1}n^{-\frac{1}{2}}\sum_{t=1}^{n}\tilde{u}_{t-1}\varepsilon_t \xrightarrow{P} 0. \tag{28}$$

Из (27), (28) следует, что

$$n^{\frac{1}{2}}(\hat{\beta}_{n,hS}-\beta) \xrightarrow{d} N(0,\sigma^2\mathcal{K}^{-1}).$$

Действительно, если ξ_n, η_n – юбые случайные векторы, такие, что $\xi_n = \eta_n + \alpha_n,\ \eta_n \xrightarrow{d} \xi,\ \alpha_n = o_p(1) \Longrightarrow \forall\ \lambda \in \mathbb{R}^p \lambda\xi_n = \lambda\eta_n + \lambda^T\alpha_n \xrightarrow{d} \lambda^T\xi$ в силу леммы Слуцкого. Значит, $\xi_n \xrightarrow{d} \xi$ в силу леммы 13.1.
Пусть $S_n^{\Delta} := \{\omega :\ det(\mathcal{K}_n) \geq \Delta = \frac{1}{2}det(\mathcal{K})\} \Longrightarrow P(\bar{S_n^{\Delta}}) \leq P(|det(\mathcal{K}_n) -$

$det(\mathcal{K})| > \Delta) \to 0 \Longrightarrow P(S_n^\Delta) \to 1$. Далее, $||$ – Евклидова норма матрицы или вектора. На S_n^Δ

$$n^{\frac{1}{2}}(\hat{\beta}_{n,hS} - \beta) = \mathcal{K}_n^{-1} n^{-\frac{1}{2}} \sum_{t=1}^{n} \tilde{u}_{t-1}\varepsilon_t \Longrightarrow \forall\, \delta > 0$$

$$P(|\gamma_n| > \delta, S_n^\Delta + \bar{S_n^\Delta}) = P(|(\mathcal{K}_n^{-1} - \mathcal{K}^{-1}) n^{-\frac{1}{2}} \sum_{t=1}^{n} \tilde{u}_{t-1}\varepsilon_t| > \delta, S_n^\Delta) + P(|\gamma_n| > \delta, \bar{S_n^\Delta}).$$

Вторая вероятность не больше $P(\bar{S_n^\Delta}) \to 0$, первая не больше, чем

$$P(|\mathcal{K}_n^{-1}||\mathcal{K} - \mathcal{K}_n||\mathcal{K}^{-1}||n^{-\frac{1}{2}} \sum_{t=1}^{n} \tilde{u}_{t-1}\varepsilon_t| > \delta, S_n^\Delta). \qquad (29)$$

В (29) $|\mathcal{K}_n - \mathcal{K}| \xrightarrow{P} 0$, $|\mathcal{K}^{-1}|$ – конечное число, $|n^{-\frac{1}{2}} \sum\limits_{t=1}^{n} \tilde{u}_{t-1}\varepsilon_t| = O_p(1)$. Почему $|\mathcal{K}_n^{-1}| = O_p(1)$? Рассмотрим $|\mathcal{K}_n^{-1}|$ на S_n^Δ. Элемент a_{ij}^n матрицы $\mathcal{K}_n^{-1}$ имеет вид:

$$a_{ij}^n = \frac{A_{ij}^n}{det(\mathcal{K}_n)},\ A_{ij}^n \text{ – алгебраическое дополнение } a_{ij}^n.$$

На S_n^Δ $|a_{ij}^n| \le \dfrac{|A_{ij}^n|}{\Delta}$. Пусть B_n – $p \times p$ матрица с элементами $b_{ij}^n = \dfrac{|A_{ij}^n|}{\Delta}$, а $B:\ b_{ij} = \dfrac{|A_{ij}|}{\Delta}$, A_{ij} – алгебраическое дополнение a_{ij} в $\mathcal{K}$. Т.к. $b_{ij}^n \xrightarrow{\text{п.н.}} b_{ij}$, то $|B_n| \xrightarrow{\text{п.н.}} |B| \Longrightarrow |\mathcal{K}_n^{-1}| \le |B_n| = O_p(1) \Longrightarrow$ вероятность в (29) не больше

$$P(|B_n||\mathcal{K} - \mathcal{K}_n||\mathcal{K}^{-1}||n^{-\frac{1}{2}} \sum_{t=1}^{n} \tilde{u}_{t-1}\varepsilon_t| > \delta) \to 0.$$

Соотношение (28) доказано.

■

13.13 Проверка гипотез о порядке авторегрессии

Пусть $\beta^T = (\beta_1^T, \beta_2^T)$, $\beta_1 - -$ m-вектор, $m < p$, $\beta_2 - -(p-m)$-вектор. $H_0 : \beta_2 = 0$, $H_1 : \beta_2 \neq 0$. Гипотеза H_0 означает, что порядок авторегрессии не больше m.

Лемма 13.2. *Пусть* $\xi_n, \xi \in \mathbb{R}^p$, $\xi_n \xrightarrow{d} \xi \sim N(a, \Sigma)$, $\Sigma > 0$. *Пусть оценка* $\hat{\Sigma}_n \xrightarrow{P} \Sigma$,

$$\tilde{\Sigma}_n^{-1} = \begin{cases} \hat{\Sigma}_n, & \text{если } \hat{\Sigma}_n \text{ невырождена;} \\ E_p, & \text{иначе.} \end{cases} \tag{13.1}$$

Тогда

$$(\xi - a)^T \tilde{\Sigma}_n^{-1} (\xi - a) \xrightarrow{d} \chi^2(p).$$

Возьмем оценкой матрицы $\mathcal{K} = E\tilde{u}_0\tilde{u}_0^T$ матрицу $\mathcal{K}_n := n^{-1}\sum\limits_{t=1}^{n} \tilde{u}_{t-1}\tilde{u}_{t-1}^T \Longrightarrow \mathcal{K}_n \xrightarrow{\text{п.н.}} \mathcal{K} > 0$. Пусть $\hat{\beta}_{n,hS} = (\hat{\beta}_{1n}^T, \hat{\beta}_{2n}^T)^T$, $\hat{\beta}_{1n}$ – m-вектор, $\hat{\beta}_{2n}$ – (p-m)-вектор. силу теоремы 13.9

$$n^{\frac{1}{2}}(\hat{\beta}_{2n} - \beta_2) \xrightarrow{d} N(0, \sigma^2 B_{22}),$$

где

$$\mathcal{K}^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}. \tag{13.2}$$

Тогда при $H_0 : \beta_2 = 0$ в силу леммы 13.2 для

$$\tilde{\mathcal{K}}_n^{-1} = \begin{pmatrix} \tilde{B_{11}} & \tilde{B_{12}} \\ \tilde{B_{21}} & \tilde{B_{22}} \end{pmatrix} = \begin{cases} \mathcal{K}_n^{-1}, & \text{если } \mathcal{K} \text{ невырождена;} \\ E_p, & \text{если } \mathcal{K} \text{ вырождена;} \end{cases} \tag{13.3}$$

имеем

$$\frac{n\hat{\beta}_{2n}^T \tilde{B}_{22}^{-1} \hat{\beta}_{2n}}{\sigma^2} \xrightarrow{d} \chi(p - m).$$

Задача 13.5. *Пусть* $\hat{s}_n^2 := n^{-1}\sum\limits_{t=1}^{n}(u_t - \tilde{u}_{t-1}^T \hat{\beta}_{n,hS})^2$. *Необходимо показать, что* $\hat{s}_n^2 \xrightarrow{P} \sigma^2$.

Тестовая статистика для H_0:

$$t_n := \frac{n\hat{\beta}_{2n}^T \tilde{B}_{22}^{-1} \hat{\beta}_{2n}}{\hat{s}_n^2}.$$

При H_0 $t_n \xrightarrow{d} \chi^2(p - m)$, критическое множество $t_n > \chi_{1-\alpha}(p - m)$, $\chi_{1-\alpha}(p - m)$ – квантиль уровня $1 - \alpha \Longrightarrow P(H_1|H_0) \to \alpha$, $P(H_0|H_1) \to 0$, т.е.

$$\begin{cases} P(H_0|H_0) \to 1 - \alpha, \\ P(H_1|H_1) \to 1. \end{cases}$$

Литература

[1] Курс лекций М.В.Болдина, механико-математический факультет МГУ им. М.В.Ломоносова, 2020-2021 гг.

[2] Курс семинаров А.А.Муромской, механико-математический факультет МГУ им. М.В.Ломоносова, 2020-2021 гг.

[3] Ивченко Г.И., Медведев Ю.И. Математическая статистика. М.б Высшая школа, 1984.

[4] Боровков А.А., Математическая статистика. М., Наука. 1984.

[5] Болдин М.В., Симонова Г.И., Тюрин Ю.Н. Знаковый анализ линейных моделей. М., Наука, 1997.

[6] Ширяев А.Н. Основы стохастической финансовой математики. Факты. Модели. М., Фазис, 1998.

[7] Леман Э. Теория точечного оценивания. М., Наука, 1991.

[8] Ширяев А.Н. Вероятность, 5-ое изд. МЦНМО, М., 2011.