

Лекции
по курсу *”Математическая Статистика”*
Ибрагимов Д. Н.

Contents

Источники	7
1 Многомерное нормальное распределение	8
Замечание	8
1.1 Лемма 1	8
1.2 Определение 1	8
1.3 Лемма 2	8
1.4 Лемма 3	9
1.5 Лемма 4	9
note	9
1.6 Определение 2	9
Замечание	9
1.7 Лемма 5	10
Доказательство	10
Замечание	10
1.8 Определение 3	10
1.8.1 Доказательство леммы 3	10
1.8.2 Доказательство леммы 4	11
Замечание	11
2 Теорема о нормальной корреляции	12
2.1 Определение 1	12
2.2 Основные свойства условного М. О.	12
2.2.1 Свойство 1	12
2.2.2 Свойство 2	12
2.2.3 Свойство 3	12
2.2.4 Свойство 4	12
2.2.5 Свойство 5	12
2.3 Лемма 1	13
Доказательство	13
Замечание	13
2.4 Определение 2	13
2.5 Определение 3	13
2.6 Теорема 1	13
Доказательство	13
2.7 Теорема 2 (О нормальной корреляции)	14
Доказательство	14
Замечание	15
3 Виды сходимости последовательностей случайных величин	16
3.1 Определение 1	16
3.2 Определение 2	16
3.3 Определение 3	16
3.4 Определение 4	16
3.5 Пример 1	16
3.6 Пример 2	17

3.7	Пример 3	17
3.8	Пример 4	17
3.9	Пример 5	18
	Замечание	18
3.10	Лемма 1	18
	Доказательство	18
3.11	Лемма 2 (Неравенство Маркова)	18
	Доказательство	18
3.12	Следствие 1	19
	Доказательство	19
3.13	Следствие 2 (Неравенство Чебышёва)	19
	Доказательство	19
3.14	Лемма 3	19
	Доказательство	19
3.15	Теорема 1 (Бореля — Кантелли)	19
	Доказательство	19
3.16	Лемма 4	20
	Доказательство	20
	Замечание	20
4	Закон больших чисел	22
4.1	Определение 1	22
4.2	Определение 2	22
4.3	Теорема 1 (Закон Больших Чисел Чебышёва)	22
	Доказательство	22
4.4	Теорема 2 (Закон Больших Чисел Колмогорова)	22
	Замечание	22
4.5	Теорема 3	22
	Доказательство	22
4.6	Следствие 1	23
	Доказательство	23
4.7	Теорема 4	23
	Замечание	23
4.8	Следствие 2	23
	Доказательство	23
5	Центральная предельная теорема (ЦПТ)	25
	Замечание	25
5.1	Определение 1	25
5.2	Лемма 1	25
5.3	Лемма 2	25
	Доказательство	25
5.4	Доказательство леммы 5.1	26
	Замечание	26
5.5	Определение 2	26
5.6	Теорема 1 (Центральная предельная)	26
	Доказательство	26
5.7	Следствие 1 (Теорема Муавра - Лапласа)	27

Доказательство	27
5.8 Пример 1	27
5.9 Теорема 2 (Ляпунова)	28
Замечание	28
5.10 Теорема 3 (Неравенство Берри-Эссена)	28
Замечание	28
5.11 Пример	28
6 Выборка и ее характеристики	29
6.1 Определение 1	29
6.2 Определение 2	29
6.3 Определение 3	29
Замечание	29
6.4 Определение 4	29
6.5 Определение 5	29
Замечание	29
6.6 Определение 6	29
6.7 Определение 7	29
6.8 Лемма 1	30
Доказательство	30
6.9 Следствие 1	30
6.10 Определение 8	30
6.11 Теорема 1 (Мостеллера)	30
6.12 Определение 9	30
Замечание	30
6.13 Свойства $\hat{F}_n(x)$	31
Замечание	31
6.14 Определение 10	32
Замечание	32
<u>Выборочные моменты</u>	33
6.15 Определение 1	33
6.16 Определение 2	33
6.17 Определение 3	33
6.18 Свойства выборочных моментов	33
7 Основные распределения в статистике	35
<u>Точечные оценки</u>	35
7.1 Определение 1	35
7.2 Свойства распределения $\chi^2(n)$	35
7.3 Определение 2	36
7.4 Свойства распределения $t(n)$	36
7.5 Определение 3	37
7.6 Свойства распределения $F(n; m)$	38
7.7 Определение 4	39
7.8 Определение 5	39
Замечание	39
7.9 Определение 6	39
7.10 Определение 7	39

7.11	Определение 8	39
7.12	Определение 9	39
	Замечание	39
7.13	Определение 10	39
7.14	Пример	40
7.15	Теорема 1	40
	Доказательство	40
8	Эффективные оценки	41
8.1	Определение 1	41
	Замечание	41
8.2	Определение 2	41
	Замечание	41
8.3	Определение 3	41
8.4	Лемма 1	41
	Доказательство	42
8.5	Определение 4	42
	Замечание	42
8.6	Лемма 2	42
	Доказательство	42
8.7	Пример 1	42
8.8	Пример 2	43
8.9	Теорема 1 (Неравенство Рао-Крамера)	43
	Доказательство	43
8.10	Определение 5	43
	Замечание	44
8.11	Пример 3	44
9	Методы построения точечных оценок	45
	<u>Метод максимального правдоподобия</u>	45
9.1	Определение 1	45
	Замечание	45
9.2	Пример 1	45
9.3	Пример 2	46
	Замечание	46
9.4	Пример 3	46
	Замечание	46
9.5	Теорема 1	47
	<u>Метод моментов</u>	47
9.6	Определение 2	47
9.7	Определение 3	47
9.8	Теорема 2	47
	Доказательство	47
9.9	Пример 1	47
9.10	Пример 2	48
	Замечание	48

10 Интервальные оценки	49
10.1 Определение 1	49
10.2 Определение 2	49
10.3 Определение 3	49
Замечание	49
<u>Построение доверительного интервала на основе центральной статистики</u>	50
10.4 Определение 4	50
Замечание	50
10.5 Определение 5	50
10.6 Лемма 1	50
Доказательство	51
10.7 Теорема 1 (ФИШЕРА)	51
Доказательство	51
10.8 Пример 3	52
10.9 Пример 4	53
11 Проверка статистических гипотез	54
11.1 Определение 1	54
11.2 Определение 2	54
11.3 Определение 3	54
11.4 Примеры	54
Замечание	54
11.5 Определение 4	54
11.6 Определение 5	55
11.7 Определение 6	55
11.8 Определение 7	55
Замечание	55
11.9 Определение 8	55
11.10 Определение 9	55
Замечание	56
<u>Критерий согласия Колмогорова</u>	56
<u>Критерий согласия хи-квадрат Пирсона</u>	56
11.11 Теорема 1	56
Замечание	57
<u>Проверка гипотезы о значении параметра</u>	57
12 Метод наименьших квадратов	58
12.1 Определение 1	58
12.2 Определение 2	58
12.3 Определение 3	58
12.4 Теорема 1 (Гаусса-Маркова)	58
Доказательство	59
Замечание	60
<u>Нормальная регрессия</u>	60
12.5 Определение 4	60
12.6 Лемма 1	60
Доказательство	60
12.7 Лемма 2	60

Доказательство	60
12.8 Следствие 1	60
Доказательство	60
12.9 Лемма 3	60
Доказательство	61
12.10 Определение 5	61
12.11 Лемма 4	61
Доказательство	61
12.12 Лемма 5	61
Доказательство	61
12.13 Лемма 6	61
Доказательство	62
12.14 Лемма 7	62
Доказательство	62
12.15 Лемма 8	63
Доказательство	63
Замечание	63
<u>Критерий Фишера</u>	63

Источники

- Ивченко Г. И., Медведев Ю. И. “Математическая статистика”, изд. “Высшая школа”, 1984
- Кибзун А. И., Наумов А. В., Горяинова Е. Р. “Теория вероятностей и математическая статистика. Базовый курс с примерами и задачами”, изд “ФИЗМАТЛИТ”, 2013
- Панков А. Р., Платонов Е. Н. “Практикум по математической статистике”, изд. “МАИ”, 2006

1 Многомерное нормальное распределение

Замечание

Вектор $X = (X_1, \dots, X_n)^T$ называется **случайным**, если X_1, \dots, X_n — случайные величины (далее **с.в.**), определенные на одном вероятностном пространстве.

Через $M[X] = m_X$ обозначим **вектор математического ожидания**:

$$M[X] = m_X = \begin{pmatrix} M[X_1] \\ \vdots \\ M[X_n] \end{pmatrix}$$

Через K_x обозначим **ковариационную матрицу** с.в. X :

$$K_X = \begin{pmatrix} \text{cov}(X_1, X_1) & \dots & \text{cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \dots & \text{cov}(X_n, X_n) \end{pmatrix}$$

1.1 Лемма 1

Пусть $K_X \in \mathbb{R}^{n \times n}$ — ковариационная матрица с.в. X . Тогда:

1. $K_X \geq 0$, т.е. $\forall x \in \mathbb{R}^n \setminus \{0\}, x^T K_X x \geq 0$;
2. $K_X^T = K_X$

1.2 Определение 1

Случайный вектор $X = (X_1, \dots, X_n)^T$ называется **невыврожденным нормальным вектором**:

$$X \sim N(m_X, K_X)$$

если совместная плотность вероятности имеет вид:

$$f_X(x) = ((2\pi)^n \det K_X)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x - m_X)^T K_X^{-1}(x - m_X)\right\}$$

где $m_X \in \mathbb{R}^n, K_X \in \mathbb{R}^{n \times n}, K_X > 0, K_X^T = K_X$

1.3 Лемма 2

Пусть X — невырожденный нормальный вектор с параметрами m_X и K_X .

Тогда $M[X] = m_X$, а K_X — ковариационная матрица X .

Рассмотрим основные свойства многомерного нормального распределения.

1.4 Лемма 3

Пусть $X \sim N(m_X, K_X)$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.

Тогда:

$$\begin{aligned} Y = AX + b &\sim N(m_Y, K_Y), \\ m_Y &= Am_X + b, \\ K_Y &= AK_X A^T. \end{aligned}$$

1.5 Лемма 4

Пусть $X \sim N(m_X, K_X)$.

Тогда компоненты вектора X **независимы** тогда и только тогда, когда они *некоррелированы*.

note

Доказательство данных утверждений при помощи аппарата функций распределения и плотности довольно сложно. Поэтому рассмотрим аппарат характеристических функций.

1.6 Определение 2

Пусть $X = (X_1, \dots, X_n)^T$ — случайный вектор.

Тогда **характеристической функцией** называется:

$$\psi_X(\lambda) = M[e^{i\lambda^T X}] = \int_{\mathbb{R}^n} e^{i\lambda^T X} dF_X(x)$$

Замечание

Характеристическая функция определена для любого случайного вектора или с.в.

Если с.в **дискретная**, то:

$$\psi_X(\lambda) = \sum_{k=1}^{\infty} e^{i\lambda X_k} p_k$$

Если с.в **абсолютно непрерывная**, то

$$\psi_X(\lambda) = \int_{\mathbb{R}} e^{i\lambda X} f_X(x) dx$$

В этом случае $\psi_X(\lambda)$ является **преобразованием Фурье** f_X .

Поскольку преобразование Фурье взаимно однозначно, а f_X однозначно определяет распределение, то характеристическая функция $\psi_X(x)$ также однозначно определяет распределение с.в X .

Причем:

$$f_X(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-i\lambda^T X} \psi_X(\lambda) d\lambda$$

1.7 Лемма 5

Пусть X — случайный вектор, $A \in \mathbb{R}^{n \times n}$, $b \in \mathbb{R}^n$.

Тогда:

1. для $Y = AX + b$

$$\psi_Y(\lambda) = e^{i\lambda^T b} \psi_X(A^T \lambda)$$

2. компоненты вектора X **независимы** тогда и только тогда, когда

$$\psi_Y(\lambda) = \prod_{k=1}^n \psi_{X_k}(\lambda_k)$$

Доказательство

$$1. \psi_Y(\lambda) = M[e^{i\lambda^T Y}] = M[e^{i\lambda^T AX} e^{i\lambda^T b}] = e^{i\lambda^T b} M[e^{i(A^T \lambda)^T X}] = e^{i\lambda^T b} \psi_X(A^T \lambda)$$

$$2. \psi_X(\lambda) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} e^{i(\lambda_1 x_1 + \dots + \lambda_n x_n)} f_X(x_1, \dots, x_n) dx_1 \dots dx_n \stackrel{\text{H/3}}{=} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} e^{i\lambda_1 x_1} \dots e^{i\lambda_n x_n} \cdot f_{X_1}(x) \cdot \dots \cdot f_{X_n} dx_1 \dots dx_n = \int_{\mathbb{R}} e^{i\lambda_1 x_1} f_{x_1}(x_1) dx_1 \dots \int_{\mathbb{R}} e^{i\lambda_n x_n} f_{x_n}(x_n) dx_n = \prod_{k=1}^n \psi_{X_k}(\lambda_k)$$

■

Замечание

При помощи характеристической функции можно дать другое определение нормального распределения. В том числе для вырожденного K_X .

1.8 Определение 3

Случайный вектор X называется **нормальным**: $X \sim N(m_X, K_X)$, если:

$$\psi_X(\lambda) = \exp\{i\lambda^T m_X - \frac{1}{2} \lambda^T K_X \lambda\}$$

1.8.1 Доказательство леммы 3

В силу [Леммы 5, п.1](#)

$$\begin{aligned} \psi_Y(\lambda) &= e^{i\lambda^T b} \psi_X(A^T \lambda) = e^{i\lambda^T b} \exp\{i\lambda^T A m_x - \frac{1}{2} \lambda^T A K_X A^T \lambda\} = \\ &= \exp\{i\lambda^T \underbrace{(A m_x + b)}_{m_Y} - \frac{1}{2} \lambda^T \underbrace{(A K_X A^T)}_{K_Y} \lambda\} \end{aligned}$$

■

1.8.2 Доказательство леммы 4

Пусть X_i, \dots, X_n попарно некоррелированы. Тогда $\text{cov}(X_i, X_j) = 0$, $i \neq j$, т.е. :

$$K_x = \text{diag}(\sigma_{X_1}^2, \dots, \sigma_{X_n}^2) = \begin{pmatrix} \sigma_{X_1}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{X_n}^2 \end{pmatrix}$$

$$\psi_X(\lambda) = \exp\{i\lambda_1 m_{X_1} + \dots + i\lambda_n m_{X_n} - \frac{1}{2} \lambda^T K_X \lambda\} = \exp\{i\lambda_1 m_{X_1} + \dots + i\lambda_n m_{X_n} - \frac{1}{2}(\lambda_1^2 \sigma_{X_1}^2 + \dots + \lambda_n^2 \sigma_{X_n}^2)\} = \prod_{k=1}^n \exp\{i\lambda_k m_{X_k} - \frac{1}{2} \lambda_k^2 \sigma_{X_k}^2\} = \prod_{k=1}^n \psi_{X_k}(\lambda_k).$$

Откуда с учетом Леммы 5, п.1 X_1, \dots, X_n — н/з.

Пусть X_1, \dots, X_n — н/з. Тогда X_1, \dots, X_n попарно некоррелированы. ■

Замечание

Поскольку K_X — невырожденная, симметрическая и положительноопределенная, то существует $S \in \mathbb{R}^{n \times n}$ — ортогональная (т. е. $S^T = S^{-1}$) такая, что:

$$S^T K_X S = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$$

где $\lambda_i > 0$, $i = \overline{1, n}$

Определим матрицу $\Lambda^{-\frac{1}{2}} = \text{diag}(\lambda_1^{-\frac{1}{2}}, \dots, \lambda_n^{-\frac{1}{2}})$.

Рассмотрим вектор

$$Y = \Lambda^{-\frac{1}{2}} S^T (X - m_X)$$

Тогда $A = \Lambda^{-\frac{1}{2}} S^T$, $b = -\Lambda^{-\frac{1}{2}} S^T m_X$.

В силу Леммы 3:

$$m_Y = A m_X + b = \Lambda^{-\frac{1}{2}} S^T m_X - \Lambda^{-\frac{1}{2}} S^T m_X = 0,$$

$$K_Y = A K_X A^T = \Lambda^{-\frac{1}{2}} S^T K_X S \Lambda^{-\frac{1}{2}} = I,$$

т. е. $Y \sim N(0, I)$.

При помощи невырожденного линейного преобразования с.в. X может быть преобразован в стандартный нормальный вектор.

Верно и обратное:

$$X = m_X + S \Lambda^{\frac{1}{2}} Y,$$

откуда следует Лемма 2

2 Теорема о нормальной корреляции

2.1 Определение 1

Условным математическим ожиданием абсолютно непрерывного случайного вектора X относительно абсолютно непрерывного случайного вектора Y называется:

$$M[X | Y] = \int_{-\infty}^{+\infty} x f_{X|Y}(x | Y) dx,$$

$$\text{где } f_{X|Y}(x | Y) = \frac{f_z(x, Y)}{f_Y(Y)}, z = \begin{pmatrix} X \\ Y \end{pmatrix}$$

2.2 Основные свойства условного М. О.

2.2.1 Свойство 1

$$\boxed{M[C | Y] = C}$$

Доказательство

$$M[C | Y] = \int_{-\infty}^{+\infty} C f_{X|Y}(x | Y) dx = \frac{C \int_{-\infty}^{+\infty} f_z(x, Y) dx}{f_Y(Y)} = C \frac{f_Y(Y)}{f_Y(Y)} = C. \blacksquare$$

2.2.2 Свойство 2

$$\boxed{M[X \varphi(Y) | Y] = \varphi(Y) M[X | Y]}$$

Доказательство

$$\begin{aligned} M[\varphi(Y) X | Y] &= \int_{-\infty}^{+\infty} \varphi(Y) X f_{X|Y}(x | Y) dx = \varphi(Y) \int_{-\infty}^{+\infty} x f_{X|Y}(x | Y) dx = \\ &= \varphi(Y) M[X | Y]. \blacksquare \end{aligned}$$

2.2.3 Свойство 3

$$\boxed{M[\alpha X_1 + \beta X_2 | Y] = \alpha M[X_1 | Y] + \beta M[X_2 | Y]}$$

2.2.4 Свойство 4

Пусть X, Y — независимые. Тогда $\boxed{M[X | Y] = M[X]}$

Доказательство

$$M[X | Y] = \int_{-\infty}^{+\infty} x f_{X|Y}(x | Y) dx = \int_{-\infty}^{+\infty} x \frac{f_z(x, Y)}{f_Y(Y)} dx = \int_{-\infty}^{+\infty} x \frac{f_X(x) f_Y(Y)}{f_Y(Y)} dx = M[X]. \blacksquare$$

2.2.5 Свойство 5

$$\boxed{M[M[X | Y]] = M[X]} \quad (\text{формула повторного М. О.})$$

Доказательство

$$M[M[X | Y]] = \int_{-\infty}^{+\infty} M[X | Y] f_Y(y) dy = \int_{-\infty}^{+\infty} f_Y(y) \int_{-\infty}^{+\infty} x f_{X|Y}(x | y) dx dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \frac{f_Y(y) f_z(x, y)}{f_Y(y)} dy dx =$$

$$\int_{-\infty}^{+\infty} x \int_{-\infty}^{+\infty} f_z(x, y) dy dx = \int_{-\infty}^{+\infty} x f_X(x) dx = M[X]. \blacksquare$$

2.3 Лемма 1

Пусть X, Y — случайные векторы с конечными вторыми моментами. Тогда:

$$M[(X - \hat{X})\varphi(Y)^T] = 0$$

где $\hat{X} = M[X | Y]$

Доказательство

$$M[(X - \hat{X})\varphi(Y)^T] = M[X\varphi(Y)^T] - M[M[X | Y]\varphi(Y)^T] =$$

$$= \text{по Свойству 2} = M[X\varphi(Y)^T] - M[M[X\varphi(Y)^T | Y]] = \text{по Свойству 5} = M[X\varphi(Y)^T] -$$

$$M[X\varphi(Y)^T] = 0. \blacksquare$$

Замечание

Если рассмотреть евклидово пространство $\mathbb{L}_2(\Omega)$ со скалярным произведением:

$$(X, Y) = M[X \cdot Y]$$

то условное М. О. — *оператор ортогонального проектирования* X на подпространство, порождаемое Y .

2.4 Определение 2

Оценкой X по наблюдениям Y называется любая измеримая функция $\varphi(Y)$.

2.5 Определение 3

Оценка \hat{X} называется с.к.-оптимальной оценкой X , если для любой другой оценки \tilde{X} верно

$$M[|X - \hat{X}|^2] \leq M[|X - \tilde{X}|^2]$$

2.6 Теорема 1

$M[X | Y]$ — с.к.-оптимальная оценка X по наблюдениям Y .

Доказательство

$$M[|X - \tilde{X}|^2] = M[|X - \hat{X} + \hat{X} - \tilde{X}|^2] = M[|X - \hat{X}|^2] + 2M[(X - \hat{X})^T(\hat{X} - \tilde{X})] + M[|\hat{X} - \tilde{X}|^2] \stackrel{*}{=}$$

Поскольку по определению $\tilde{X} - \hat{X} = \varphi(Y)$, то в силу Леммы 2.1 $M[(X - \hat{X})^T(\tilde{X} - \hat{X})] = 0$.

$$\stackrel{*}{=} M[|X - \hat{X}|^2] + M[|\hat{X} - \tilde{X}|^2] \geq M[|X - \hat{X}|^2]. \blacksquare$$

2.7 Теорема 2 (О нормальной корреляции)

Пусть

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left(\begin{pmatrix} m_X \\ m_Y \end{pmatrix}, \begin{pmatrix} K_X & K_{XY} \\ K_{XY}^T & K_Y \end{pmatrix} \right)$$

Тогда

$$1. \text{Law}(X | Y) = N(\mu(Y), \Delta),$$

где

$$\mu(Y) = M[X | Y] = m_X + K_{XY} K_Y^{-1} (Y - m_Y)$$

$$\Delta = K_X - K_{XY} K_Y^{-1} K_{YX}$$

$$2. M[|X - \mu(Y)|^2] = \text{tr}(K_X - K_{XY} K_Y^{-1} K_{YX})$$

Доказательство

Рассмотрим линейное преобразование Y :

$$\mu(Y) = m_X + K_{XY} K_Y^{-1} (Y - m_Y)$$

В силу [Леммы 1.3](#)

$$X - \mu(Y) = (I - K_{XY} K_Y^{-1}) \begin{pmatrix} X \\ Y \end{pmatrix} - m_X + K_{XY} K_Y^{-1} m_Y \sim N(\mu, K)$$

$$\mu = (I - K_{XY} K_Y^{-1}) \begin{pmatrix} m_X \\ m_Y \end{pmatrix} - m_X + K_{XY} K_Y^{-1} m_Y = 0$$

$$K = (I - K_{XY} K_Y^{-1}) \begin{pmatrix} K_X & K_{XY} \\ K_{XY}^T & K_Y \end{pmatrix} \begin{pmatrix} I \\ -(K_{XY} K_Y^{-1})^T \end{pmatrix} =$$

$$= (K_X - K_{XY} K_Y^{-1} K_{XY}^T \quad K_{XY} - K_{XY} K_Y^{-1} K_Y) \begin{pmatrix} I \\ K_Y^{-1} K_{XY}^T \end{pmatrix} =$$

$$= K_X - K_{XY} K_Y^{-1} K_{XY}^T = \Delta$$

$$\text{cov}(X - \mu(Y), Y) = \text{cov}(X, Y) - \text{cov}(\mu(Y), Y) = \text{cov}(X, Y) - \text{cov}(K_{XY} K_Y^{-1} Y + m_X - K_{XY} K_Y^{-1} m_Y, Y) = \text{cov}(X, Y) - K_{XY} K_Y^{-1} \text{cov}(Y, Y) = K_{XY} - K_{XY} K_Y^{-1} K_Y = 0$$

т.е. $X - \mu(Y)$ и Y **некорреливаны**.

Тогда в силу [Леммы 1.4, п.2](#) $X - \mu(Y)$ и Y **независимы**. Построим характеристическую функцию условного распределения X относительно Y :

$$\psi_{X|Y}(\lambda | Y) = \int_{\mathbb{R}^n} e^{i\lambda^T X} f_{X|Y}(x | Y) dx = M[e^{i\lambda^T X} | Y] = M[e^{i\lambda^T (X - \mu(Y))} e^{i\lambda^T \mu(Y)} | Y] \stackrel{*}{=}$$

в силу [Свойства 2 УМО](#) и **независимости** $X - \mu(Y)$ и Y

$$\stackrel{*}{=} M[e^{i\lambda^T (X - \mu(Y))} | Y] \cdot M[e^{i\lambda^T \mu(Y)} | Y] = M[e^{i\lambda^T (X - \mu(Y))}] e^{i\lambda^T \mu(Y)} = \psi_{X - \mu(Y)}(\lambda) e^{i\lambda^T \mu(Y)} = \exp\{-\frac{1}{2} \lambda^T \Delta \lambda\} \cdot \exp\{i\lambda^T \mu(Y) - \frac{1}{2} \lambda^T \Delta \lambda\}$$

т.е. Условное распределение **нормальное**:

$$X(Y \sim N(\mu(Y), \Delta))$$

Вычислим с.к. ошибку:

$$M[|X - \mu(Y)|^2] = M[\Delta X_1^2 + \Delta X_2^2 + \dots + \Delta X_n^2] = \sum_{k=1}^n M[\Delta X_k^2] = \sum_{k=1}^n D[\Delta X_k] = \sum_{k=1}^n \Delta_{kk} = \text{tr} \Delta.$$

■

Замечание

1. Из [Теоремы о нормальной корреляции](#) следует, что в *гауссовском* случае с.к.-оптимальная оценка является *линейной*.
2. Если X и Y — *независимы*, то с.к.-оптимальная оценка — m_X .
3. С.к.-оптимальная оценка *несмещенная*, т.к. $M[X - \mu(Y)] = 0$.

3 Виды сходимости последовательностей случайных величин

3.1 Определение 1

Говорят, что $\{X_n\}_{n=1}^{\infty}$ образует *последовательность случайных величин*, если $\forall N \in \mathbb{N}$ X_n определены на одном вероятностном пространстве.

3.2 Определение 2

Говорят, что последовательность с.в. $\{X_n\}_{n \in \mathbb{N}}$ *сходится по вероятности* к с.в. X , если $\forall \varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|X_n - X| \leq \varepsilon) = 1$$

ИЛИ

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$$

3.3 Определение 3

Говорят, что последовательность с.в. $\{X_n\}_{n \in \mathbb{N}}$ *сходится почти наверное* к с.в. X , если

$$P(\{\omega : X_n(\omega) \not\rightarrow X(\omega)\}) = 0$$

ИЛИ

$$P(\{\omega : X_n(\omega) \xrightarrow{n \rightarrow \infty} X(\omega)\}) = 1$$

3.4 Определение 4

Говорят, что последовательность с.в. $\{X_n\}_{n \in \mathbb{N}}$ *сходится в среднем квадратическом* к с.в. X , если

$$M[|X_n - X|^2] \xrightarrow{n \rightarrow \infty} 0$$

3.5 Пример 1

Пусть $\{X_n\}_{n \in \mathbb{N}}$ — случайная последовательность.

$$X \sim \begin{pmatrix} 0 & n \\ 1 - \frac{1}{n^2} & \frac{1}{n^2} \end{pmatrix}$$

$$P(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) \neq 0\}) = P(\{\omega : \forall N \in \mathbb{N}, \exists n \geq N : X_n(\omega) = n\}) = P(\prod_{N=1}^{\infty} \sum_{n=N}^{\infty} \{\omega : X_n(\omega) = n\})^* =$$

$$1. \sum_{n=N+1}^{\infty} \{\omega : X_n(\omega) = n\} \subset \sum_{n=N}^{\infty} \{\omega : X_n(\omega) = n\}$$

$$2. P(\sum_{n=N}^{\infty} \{\omega : X_n(\omega) = n\}) \leq \sum_{n=N}^{\infty} P(\{\omega : X_n(\omega) = n\}) = \sum_{n=N}^{\infty} \frac{1}{n^2} \xrightarrow{n \rightarrow \infty} 0,$$

т.к. $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$

Тогда в силу аксиомы непрерывности $\stackrel{*}{=} 0$, т.е. $X_n \xrightarrow{\text{п.н.}} 0$

3.6 Пример 2

Рассмотрим ту же последовательность. Выберем $\varepsilon > 0$

$$P(|X_n - 0| \leq \varepsilon) = \begin{cases} 1, & \varepsilon \geq n \\ 1 - \frac{1}{n^2}, & \varepsilon \in (0; n) \end{cases} \xrightarrow{n \rightarrow \infty} 1$$

Тогда $X_n \xrightarrow{P} 0$

3.7 Пример 3

Рассмотрим ту же последовательность:

$$M[|X_n - 0|^2] = M[X_n^2] = 0^2(1 - \frac{1}{n^2}) + n^2 \frac{1}{n^2} = 1 \xrightarrow{n \rightarrow \infty} 1$$

Тогда $X_n \not\xrightarrow{\text{с.к.}} 0$

3.8 Пример 4

Пусть $f_{nk} : [0; 1] \rightarrow \{0; 1\}, n \in \mathbb{N}, k = \overline{1, n}$,

$$f_{n,k} = \begin{cases} 0, & t \notin [\frac{k-1}{n}; \frac{k}{n}], \\ 1, & t \in [\frac{k-1}{n}; \frac{k}{n}]. \end{cases}$$

Пусть $X \sim R(0; 1)$. Рассмотрим последовательность с.в. $X_{nk} = f_{nk}(X)$. $\forall \omega \in \Omega X(\omega) \in [0; 1]$. Тогда $\forall n \in \mathbb{N}, \exists k = \overline{1, n}$ такое, что $X(\omega) \in [\frac{k-1}{n}; \frac{k}{n}]$. Т.е. если $\varepsilon = \frac{1}{2}$, то $\forall n \in \mathbb{N}$ найдется $k = \overline{1, n}$ такой, что

$$|f_{nk}(X(\omega)) - 0| > \varepsilon$$

Тогда $X_{nk}(\omega) \not\xrightarrow{n \rightarrow \infty} 0$, т.е.

$$\{\omega : \lim_{n, k \rightarrow \infty} X_{nk}(\omega) = 0\} = \emptyset$$

$$X_{nk} \not\xrightarrow{\text{п.н.}} 0$$

При этом $\forall \varepsilon > 0$

$$R(|X_{nk} - 0| > \varepsilon) = \begin{cases} 0, & \varepsilon \geq 1, \\ P(X \in [\frac{k-1}{n}, \frac{k}{n}]), & \varepsilon \in (0; 1) \end{cases} = \begin{cases} 0, & \varepsilon \geq 1, \\ \frac{1}{n}, & \varepsilon \in (0; 1) \end{cases} \xrightarrow{n \rightarrow \infty} 0$$

$$X_{nk} \xrightarrow{P} 0$$

$$M[|X_{nk} - 0|^2] = M[X_{nk}] = M[f_{nk}(X)] = \int_0^1 f_{nk}(x) \underbrace{f_X(x)}_1 dx = \int_0^1 f_{nk}(x) dx = \int_{\frac{k-1}{n}}^{\frac{k}{n}} 1 dx = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0$$

$$X_{nk} \xrightarrow{\text{с.к.}} 0$$

3.9 Пример 5

Рассмотрим последовательность с.в. $Y_{n_1k} = nK_{n_1k}$ Тогда $Y_{n_1k} \xrightarrow{\text{п.н.}} 0$. $\forall \varepsilon > 0$.

$$P(|Y_{n_1k} - 0| > \varepsilon) = \begin{cases} 0, & \varepsilon \geq n, \\ P(X \in [\frac{k-1}{n}; \frac{k}{n}]), & \varepsilon \in (0; n) \end{cases} = \begin{cases} 0, & \varepsilon \geq n, \\ \frac{1}{n}, & \varepsilon \in (0; n) \end{cases} \xrightarrow{n \rightarrow \infty} 0$$

$$Y_{n_1k} \xrightarrow{P} 0$$

$$M[|Y_{n_1k} - 0|^2] = M[n^2 X_{nk}^2] = n^2 M[X_{nk}] = n \xrightarrow{P} \infty$$

$$Y_{n_1k} \not\xrightarrow{\text{с.к.}} 0$$

Замечание

Согласно определению для исследования на сходимость нужно знать совместное распределение с.в. X_n и X , а для случая *сходимости почти наверное* совместное распределение всей последовательности $\{X_n\}_{n \in \mathbb{N}}$ и X . Поэтому исследование на сходимость иначе, чем к детерминированной константе, довольно проблематично.

3.10 Лемма 1

Пусть $X_n \xrightarrow{\text{п.н.}} X$. Тогда $X_n \xrightarrow{P} X$.

Доказательство

$$0 = P(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) \neq X(\omega)\}) = P(\{\omega : \exists \varepsilon > 0 \forall N \in \mathbb{N} \exists n \geq N : |X_n(\omega) - X(\omega)| > \varepsilon\}) =$$

$$P(\sum_{\varepsilon > 0} \prod_{N=1}^{\infty} \sum_{n=N}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \geq P(\prod_{N=1}^{\infty} \sum_{n=N}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon'\}), \forall \varepsilon' > 0$$

$$\text{Тогда } 0 = P(\prod_{N=1}^{\infty} \sum_{n=N}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}),$$

$$\sum_{n=N+1}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\} \subset \sum_{n=N}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$$

В силу аксиомы непрерывности:

$$0 = \lim_{N \rightarrow \infty} P(\sum_{n=N}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \geq \lim_{N \rightarrow \infty} P(\{\omega : |X_N(\omega) - X(\omega)| > \varepsilon\})$$

$$X_N \xrightarrow{P} X$$

■

3.11 Лемма 2 (Неравенство Маркова)

Пусть $P(X \geq 0) = 1$, $M[X] < \infty$. Тогда $\forall \varepsilon > 0$

$$P(X > \varepsilon) \leq \frac{M[X]}{\varepsilon}$$

Доказательство

$$M[X] = \int_0^{+\infty} x dF_X(x) \geq \int_{\varepsilon}^{+\infty} x dF_X(x) \geq \varepsilon \int_{\varepsilon}^{+\infty} dF_X(x) = \varepsilon P(X > \varepsilon). \blacksquare$$

3.12 Следствие 1

Пусть $M[X^k] < \infty$. Тогда $\forall \varepsilon > 0$:

$$P(|X| > \varepsilon) \leq \frac{M[|X|^k]}{\varepsilon^k}$$

Доказательство

$$P(|X| > \varepsilon) = P(|X|^k > \varepsilon^k) \leq \frac{M[|X|^k]}{\varepsilon^k}. \blacksquare$$

3.13 Следствие 2 (Неравенство Чебышёва)

Пусть $M[X^2] < \infty$. Тогда

$$P(|X - M[X]| > \varepsilon) \leq \frac{D[X]}{\varepsilon^2}$$

Доказательство

$$P(|X - M[X]| > \varepsilon) \leq \frac{M[|X - M[X]|^2]}{\varepsilon^2} = \frac{D[X]}{\varepsilon^2}. \blacksquare$$

3.14 Лемма 3

Пусть $X_n \xrightarrow{\text{с.к.}} X$. Тогда $X_n \xrightarrow{P} X$.

Доказательство

$$P(|X_n - X| > \varepsilon) \leq \frac{M[|X_n - X|^2]}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0. \blacksquare$$

3.15 Теорема 1 (Бореля — Кантелли)

Пусть $A_1, \dots, A_n \subset \Omega, B = \prod_{N=1}^{\infty} \sum_{n=N}^{\infty} A_n$. Тогда

1. Если $\sum_{n=1}^{\infty} P(A_n) < \infty$, то $P(B) = 0$;
2. Если A_1, \dots, A_n независимы в совокупности и $\sum_{n=1}^{\infty} P(A_n) = \infty$, то $P(B) = 1$.

Доказательство

$$1. P(B) = P\left(\prod_{N=1}^{\infty} \sum_{n=N}^{\infty} A_n\right)^* =$$

$$\begin{aligned} \text{т.к. } \sum_{n=N+1}^{\infty} A_n \subset \sum_{n=N}^{\infty} A_n, \text{ то по аксиоме непрерывности } & \stackrel{*}{=} \lim_{N \rightarrow \infty} P\left(\sum_{n=N}^{\infty} A_n\right) \leq \\ \lim_{N \rightarrow \infty} \sum_{n=N}^{\infty} P(A_n) = 0, \text{ т.к. } \sum_{n=1}^{\infty} P(A_n) < \infty & \end{aligned}$$

$$2. P(B) = \dots = \lim_{N \rightarrow \infty} P\left(\sum_{n=N}^{\infty} A_n\right) = \lim_{N \rightarrow \infty} (1 - P\left(\prod_{n=N}^{\infty} \bar{A}_n\right)) = 1 - \lim_{N \rightarrow \infty} P\left(\prod_{M=N}^{\infty} \prod_{n=N}^M \bar{A}_n\right)^* =$$

$$\begin{aligned}
 & \text{т.к. } \prod_{n=N}^{M+1} \bar{A}_n \subset \prod_{n=N}^M \bar{A}_n, \text{ то по аксиоме непрерывности} \\
 & \stackrel{*}{=} 1 - \lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} P\left(\prod_{n=N}^M \bar{A}_n\right) = 1 - \lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} \prod_{n=N}^M (1 - P(A_n)) = 1 - \lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} \prod_{n=N}^M e^{\ln(1-P(A_n))} = \\
 & 1 - \lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} e^{\sum_{n=N}^M \ln(1-P(A_n))} \stackrel{*}{\geq} \\
 & \text{т.к. } \ln(1-t) = -t - \frac{t^2}{2} - \frac{t^3}{3} - \frac{t^4}{4} - \dots < -t, \text{ то} \\
 & \stackrel{*}{\geq} 1 - \lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} e^{-\sum_{n=N}^M P(A_n)} = 1 - \lim_{n \rightarrow \infty} 0 = 1. \blacksquare
 \end{aligned}$$

3.16 Лемма 4

Пусть $X_n \xrightarrow{P} X$,

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty$$

Тогда $X_n \xrightarrow{\text{п.н.}} X$

Доказательство

В силу [Теоремы 3.1](#) $\forall \varepsilon > 0$

$$P\left(\prod_{N=1}^{\infty} \sum_{n=N}^{\infty} \{\omega : |X_n - X| > \varepsilon\}\right) = 0$$

Тогда

$$\begin{aligned}
 & P(\{\omega : \lim_{n \rightarrow \infty} X_n X_n(\omega) \neq X(\omega)\}) = P(\{\omega : \exists \varepsilon > 0, \forall N \in \mathbb{N} \exists n \geq N : |X_n(\omega) - X(\omega)| > \varepsilon\}) = \\
 & P(\{\omega : \exists M \in \mathbb{N}, \forall N \in \mathbb{N} \exists n \geq N : |X_n(\omega) - X(\omega)| > \frac{1}{M}\}) = P\left(\sum_{M=1}^{\infty} \prod_{N=1}^{\infty} \sum_{n=N}^{\infty} \{\omega : |X_n - X| > \frac{1}{M}\}\right) \leq \\
 & \sum_{M=1}^{\infty} P\left(\prod_{N=1}^{\infty} \sum_{n=N}^{\infty} \{\omega : |X_n - X| > \frac{1}{M}\}\right) = 0. \blacksquare
 \end{aligned}$$

Замечание

1.

$$X_n \xrightarrow{\text{с.к.}} X \implies X_n \xrightarrow{P} X$$

ИЛИ

$$X_n \xrightarrow{\text{п.н.}} X \implies X_n \xrightarrow{P} X$$

2. В силу *теоремы Рисса* (функциональный анализ) если $X_n \xrightarrow{P} X$, то существует подпоследовательность $\{X_{n_k}\}_{k \in \mathbb{N}} : X_{n_k} \xrightarrow{\text{п.н.}} X$

3. В силу *теоремы о мажорирующей сходимости*, если $X_n \xrightarrow{P} X$ и $\exists Y$ — с.в.: $|X_n| \leq Y, M[Y^2] < \infty$, то $X_n \xrightarrow{\text{с.к.}} X$

4. Также из функционального анализа известно, что операция предела (по мере, почти наверное, в средне квадратическом) замкнута относительно линейных операций и непрерывных преобразований.

4 Закон больших чисел

4.1 Определение 1

Выборкой объема n будем называть с.в. $Z_n = (X_1, \dots, X_n)^T$, где X_1, \dots, X_n — *независимые* с.в.

Через $F_k(x)$ обозначим *функцию распределения* k -го элемента выборки.

Если $F_k = F_1, k = 2, \bar{n}$, то выборка называется *однородной*.

4.2 Определение 2

Выборочным средним \bar{X}_n выборки Z_n называется $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$

4.3 Теорема 1 (Закон Больших Чисел Чебышёва)

Пусть Z_n — однородная выборка, $M[X_k^2] < \infty$.

Тогда $\bar{X}_n \xrightarrow{\text{с.к.}} m_X, \bar{X}_n \xrightarrow{P} m_X$

Доказательство

$$M[|\bar{X}_n - m_X|^2] = M[|\bar{X}_n - M[\bar{X}_n]|^2] = D[\bar{X}_n] = \frac{1}{n^2} D\left[\sum_{k=1}^n X_k\right] = \frac{nD[X_1]}{n^2} \xrightarrow{n \rightarrow \infty} 0$$

Т.е. по определению $\bar{X}_n \xrightarrow{\text{с.к.}} m_X$.

С учетом [Леммы 3.3](#) $\bar{X}_n \xrightarrow{P} m_X$. ■

4.4 Теорема 2 (Закон Больших Чисел Колмогорова)

Пусть Z_n — *однородная* выборка, $M[X_k] = m_X < \infty$.

Тогда $\bar{X}_n \xrightarrow{\text{п.н.}} m_X$

Замечание

Т.о. для *однородной* выборки \bar{X}_n сходится **почти наверное** и **по вероятности** к m_X , если оно существует, и в среднем квадратичном, если существует дисперсия.

4.5 Теорема 3

Пусть Z_n — *неоднородная* выборка, $M[X_k] = m_X < \infty, D[X_k] = D_k \leq D_{\max} < \infty$, где $k \in \mathbb{N}$

Тогда $\bar{X}_n \xrightarrow{\text{с.к.}} m_X, \bar{X}_n \xrightarrow{P} m_X$.

Доказательство

$$M[|\bar{X}_n - m_X|^2] = M[|\bar{X}_n - M[\bar{X}_n]|^2] = D[\bar{X}_n] = \frac{1}{n^2} \sum_{k=1}^n D_k \leq \frac{nD_{\max}}{n^2} = \frac{D_{\max}}{n} \xrightarrow{n \rightarrow \infty} 0$$

Т.о. $\bar{X}_n \xrightarrow{\text{с.к.}} m_X$. Тогда в силу [Леммы 3.3](#) $\bar{X}_n \xrightarrow{P} m_X$. ■

4.6 Следствие 1

Пусть Z_n — неоднородная выборка, $D[X_k] = D_k \leq D_{max} < \infty, k \in \mathbb{N}$.

Тогда $\bar{X}_n - \frac{1}{n} \sum_{k=1}^n m_{X_k} \xrightarrow{\text{с.к.}} 0, \bar{X}_n - \frac{1}{n} \sum_{k=1}^n m_{X_k} \xrightarrow{P} 0$.

Доказательство

$$\bar{X}_n - \frac{1}{n} \sum_{k=1}^n m_{X_k} = \frac{1}{n} \sum_{k=1}^n (X_k - m_{X_k}) = \frac{1}{n} \sum_{k=1}^n Y_k,$$

где $M[Y_k] = 0, D[Y_k] = D[X_k] \leq D_{max} < \infty$

Тогда \bar{Y}_k удовлетворяет условиям [Теоремы 4.3](#). ■

4.7 Теорема 4

Пусть Z_n — неоднородная выборка, $M[X_k] = m_X < \infty, D[X_k] = D_k < \infty$,

$$\sum_{k=1}^n \frac{D[X_k]}{k^2} < \infty$$

Тогда

$$\bar{X}_n \xrightarrow{\text{п.н.}} m_X$$

Замечание

Условие [Теоремы 4](#) более мягкое, чем условие [Теоремы 3](#). Пусть $D[X_k] \leq D_{max}, k \in \mathbb{N}$. Тогда

$$\sum_{k=1}^n \frac{D[X_k]}{k^2} \leq \sum_{k=1}^n \frac{D_{max}}{k^2} = \frac{\pi^2 D_{max}}{\sigma} < \infty$$

4.8 Следствие 2

Пусть $N(A)$ — число появления события A в серии из N независимых опытов. Тогда

$$\frac{N(A)}{N} \xrightarrow{\text{п.н.}} P(A), \frac{N(A)}{N} \xrightarrow{\text{с.к.}} P(A)$$

Доказательство

По условию $N(A) \sim Bi(N; P(A))$. Тогда $\exists X_1, \dots, X_n \sim Be(P(A))$ — независимые с.в.

При этом $M[X_1] = P(A), D[X_1] = P(A)(1 - P(A)) \leq \frac{1}{4}$.

Тогда в силу [Теоремы 4](#)

$$\frac{N(A)}{N} = \frac{1}{N} \sum_{k=1}^N X_k \xrightarrow{\text{с.к.}} P(A)$$

в силу [Теоремы 2](#)

$$\frac{N(A)}{N} = \frac{1}{N} \sum_{k=1}^N X_k \xrightarrow{\text{п.н.}} P(A)$$



5 Центральная предельная теорема (ЦПТ)

Замечание

Сходимости **с.к.**, **п.н.** и P в общем случае исследования предполагают либо знание совместного распределения элементов последовательности, либо наличие точкой функциональной зависимости от $\omega \in \Omega$.

Как правило, в теории вероятностей это неизвестно, а с.в. описываются при помощи их распределений, а не как функции. При этом если у двух величин совпадают распределения, то это вовсе не значит, что они равны.

Поэтому довольно важным является вид сходимости *по распределению*, т.е. в смысле “описательного инструмента” с.в.

5.1 Определение 1

Говорят, что последовательность с.в. $\{X_n\}_{n \in \mathbb{N}}$ **сходится по распределению** к с.в. X , если

$$F_{X_n}(x) \xrightarrow{n \rightarrow \infty} F_X(x), \forall x \text{ — точки непрерывности } F_X(x).$$

5.2 Лемма 1

Пусть $X_n \xrightarrow{P} X$. Тогда $X_n \xrightarrow{d} X$

5.3 Лемма 2

Пусть $X_n \xrightarrow{d} X, Y_n \xrightarrow{P} C$

Тогда $X_n + Y_n \xrightarrow{d} X + C$

Доказательство

Пусть $C = 0, x_0$ — точка **непрерывности** $F_X(x)$. Тогда $\forall \varepsilon > 0$

$$F_{X_n+Y_n}(x_0) = P(X_n+Y_n \leq x_0) = \underbrace{P(\{X_n+Y_n \leq x_0\}\{|Y_n| > \varepsilon\})}_{p_1} + \underbrace{P(\{X_n+Y_n \leq x_0\}\{|Y_n| \leq \varepsilon\})}_{p_2}$$

$$0 \leq p_1 \leq P(|Y_n| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$$

$$p_2 = P(\{X_n+Y_n \leq x_0\}\{-\varepsilon \leq Y_n \leq \varepsilon\}) \stackrel{*}{\leq}$$

$$\begin{cases} X_n+Y_n \leq x_0 \\ -\varepsilon \leq Y_n \leq \varepsilon \end{cases} \implies \begin{cases} X_n+Y_n \leq x_0 \\ -\varepsilon \leq -Y_n \leq \varepsilon \end{cases} \implies X_n \leq x_0 + \varepsilon$$

$$\stackrel{*}{\leq} P(X_n \leq x_0 + \varepsilon) = F_{X_n}(x_0 + \varepsilon)$$

$$p_2 \geq P(\{\varepsilon + X_n \leq x_0\}\{-\varepsilon \leq Y_n \leq \varepsilon\}) \stackrel{*}{\geq}$$

$$P(AB) = P(A) - P(A \setminus B) \geq P(A) - P(\overline{B})$$

$$\stackrel{*}{\geq} P(\varepsilon + X_n \leq x_0) - P(|Y_n| > \varepsilon) = F_{X_n}(x_0 - \varepsilon) - P(|Y_n| > \varepsilon)$$

$$\text{Т.о. } F_{X_n}(x_0 - \varepsilon) - P(|Y_n| > \varepsilon) \leq F_{X_n+Y_n}(x_0) \leq F_{X_n}(x_0 + \varepsilon) + p_1$$

Выберем $\varepsilon > 0$ так, чтобы $(x_0 - \varepsilon; x_0 + \varepsilon)$ было областью непрерывности $F_X(x)$

Тогда $\lim_{n \rightarrow \infty} F_{X_n}(x_0 \pm \varepsilon) = F_X(x_0 \pm \varepsilon)$

$$F_{X_n}(x_0 - \varepsilon) \leq \lim_{n \rightarrow \infty} F_{X_n + Y_n}(x_0) \leq \overline{\lim_{n \rightarrow \infty}} F_{X_n + Y_n}(x_0) \leq F_X(x_0 + \varepsilon)$$

Возьмем предел по $\varepsilon \rightarrow 0$. В силу непрерывности $F_X(x)$ в x_0 $\lim_{\varepsilon \rightarrow 0} F_X(x_0 \pm \varepsilon) = F_X(x_0)$

Откуда $F_X(x_0) = \lim_{\varepsilon \rightarrow 0} F_{X_n + Y_n}(x_0)$, т.е.

$$X_n + Y_n \xrightarrow{d} X$$

Пусть $C \neq 0$. Тогда $Y_n - C = \tilde{Y} \xrightarrow{P} 0$,

$$X_n + C = \tilde{X}_n \xrightarrow{d} X + C = \tilde{X}$$

Получаем $X_n + Y_n = \tilde{X}_n + \tilde{Y}_n \xrightarrow{d} \tilde{X} = X + C$. ■

5.4 Доказательство леммы 5.1

$X_n = (X_n - X) + X$, где $X \xrightarrow{d} X$, $X_n - X \xrightarrow{P} 0$. В силу Леммы 5.2 $X_n = (X_n - X) + X \xrightarrow{d} X + 0 = X$. ■

Замечание

Из теории преобразования Фурье следует, что $X_n \xrightarrow{d} X$ тогда и только тогда, когда $\psi_{X_n}(\lambda) \xrightarrow{n \rightarrow \infty} \psi_X(\lambda)$, $\lambda \in \mathbb{R}$

5.5 Определение 2

Последовательность с.в. $\{X_n\}_{n \in \mathbb{N}}$ называется **асимптотически нормальной**, если $X_n \xrightarrow{d} X$, где $X \sim N(m; \sigma^2)$

5.6 Теорема 1 (Центральная предельная)

Пусть $\{X_n\}_{n \in \mathbb{N}}$ — последовательность независимых и одинаково распределенных с.в., причем

$$M[X_1] = m_X, D[X_1] = \sigma_X^2$$

Тогда

$$S_n = \frac{\sum_{k=1}^n X_k - nm_X}{\sigma_X \sqrt{n}} \xrightarrow{d} N(0; 1)$$

Доказательство

Обозначим через $Y_k = \frac{X_k - m_X}{\sigma_X \sqrt{n}}$

Тогда $\sum_{k=1}^n Y_k$, где Y_1, \dots, Y_n — независимые с.в.

В силу Леммы 1.5

$$\psi_{S_n}(\lambda) = \psi_Y(\lambda, \lambda, \dots, \lambda) = \prod_{k=1}^n \psi_{Y_k}(\lambda) = \psi_{Y_1}^n(\lambda) = \psi^n\left(\frac{\lambda}{\sigma_X \sqrt{n}}\right)$$

где $\psi(\lambda)$ — **характеристическая функция** $X_k - m_X$.

$$\psi(0) = M[e^{i0(X_k - m_X)}] = M[1] = 1$$

$$\psi'(0) = M[i(X_k - m_X)e^{i0(X_k - m_X)}] = M[X_k - m_X]i = 0$$

$$\psi''(0) = -M[(X_k - m_X)^2 e^{i0(X_k - m_X)}] = -M[(X_k - m_X)^2] = -\sigma_X^2$$

Тогда согласно *формуле Тейлора*

$$\psi(\lambda) = \psi(0) + \psi'(0)\lambda + \frac{\psi''(0)}{2}\lambda^2 + o(\lambda^2) = 1 - \frac{\sigma_X^2}{2}\lambda^2 + o(\lambda^2)$$

$$\text{Рассмотрим } \ln \psi_{S_n}(\lambda) = n \ln \psi\left(\frac{\lambda}{\sigma_X \sqrt{n}}\right) = n \ln\left(1 - \frac{\lambda^2}{2n} + o\left(\frac{\lambda^2}{\sigma_X^2 n}\right)\right) = n\left(-\frac{\lambda^2}{2n} + o\left(\frac{\lambda^2}{\sigma_X^2 n}\right) + o\left(\frac{\lambda^2}{\sigma_X^2 n}\right)\right)$$

$$= n\left(-\frac{\lambda^2}{2n} + o\left(\frac{\lambda^2}{2n}\right)\right) = -\frac{\lambda^2}{2} + \frac{o\left(\frac{\lambda^2}{\sigma_X^2 n}\right)}{\frac{\lambda^2}{\sigma_X^2 n}} \cdot \frac{\lambda^2}{2} \xrightarrow{n \rightarrow \infty} -\frac{\lambda^2}{2},$$

$$\psi_{S_n}(\lambda) \xrightarrow{n \rightarrow \infty} e^{-\frac{\lambda^2}{2}}$$

где $\psi_Y(\lambda) = e^{-\frac{\lambda^2}{2}}$ по определению является *характеристической функцией $Y \sim N(0; 1)$

Тогда $S_n \xrightarrow{d} Y \sim N(0; 1)$. ■

5.7 Следствие 1 (Теорема Муавра - Лапласа)

Пусть $X_n \sim Bi(n; p)$

Тогда

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0; 1)$$

Доказательство

Т.к. $X_n \sim Bi(n; p)$, то существуют **независимые** $\tilde{X}_1, \dots, \tilde{X}_n \sim Be(p)$ такие, что

$$X_n = \sum_{k=1}^n \tilde{X}_k, M[\tilde{X}_k] = p = m_X, D[\tilde{X}_k] = p(1-p) = \sigma_X^2$$

Тогда в силу [Теоремы 1](#):

$$\frac{\sum_{k=1}^n \tilde{X}_k - nm_X}{\sigma_X \sqrt{n}} \xrightarrow{d} N(0; 1). \blacksquare$$

5.8 Пример 1

Вычислить вероятность того, что при $n = 1000$ подбрасываниях монета упадет “орлом” от 400 до 600 раз.

Пусть X — число выпавших “орлов”. Тогда $X \sim Bi(1000; \frac{1}{2})$. По *формуле Бернулли*

$$P(X \in [400; 600]) = \sum_{k=400}^{600} C_{1000}^k \frac{1}{2^{1000}}$$

Оценим данную величину с помощью *ЦПТ*.

В силу [Теоремы 1](#) и [Следствия](#)

$$\frac{X - n\frac{1}{2}}{\sqrt{n\frac{1}{2}(1 - \frac{1}{2})}} \xrightarrow{d} N(0; 1)$$

Тогда

$$P(400 \leq X \leq 600) = P\left(\frac{400-500}{\sqrt{250}} \leq \frac{X-nm_X}{\sigma_X\sqrt{n}} \leq \frac{600-500}{\sqrt{250}}\right) \approx \Phi_0\left(\frac{100}{5\sqrt{10}}\right) - \Phi_0\left(-\frac{100}{5\sqrt{10}}\right) = 2\Phi_0(2\sqrt{10}) \approx 1.$$

5.9 Теорема 2 (Ляпунова)

Пусть $\{X_n\}_{n \in \mathbb{N}}$ — последовательность *независимых* с.в., $M[X_n] = m_{X_n}, D[X_n] = \sigma_{X_n}^2, M[|X_n - m_{X_n}|^3] = C_n^3 < \infty$

При этом $\frac{(\sum_{k=1}^n C_k^3)^{\frac{1}{3}}}{(\sum_{k=1}^n \sigma_{X_k}^2)^{\frac{1}{2}}} \xrightarrow{n \rightarrow \infty} 0$ (*Условие Ляпунова*)

Тогда $\frac{(\sum_{k=1}^n X_k - M[\sum_{k=1}^n X_k])}{\sqrt{D[\sum_{k=1}^n X_k]}} \xrightarrow[n \rightarrow \infty]{d} \sim N(0; 1)$

Замечание

Для аппроксимации точности использования ЦПТ используется **неравенство Берри-Эссеена**

5.10 Теорема 3 (Неравенство Берри-Эссеена)

Пусть $\{X_n\}_{n \in \mathbb{N}}$ — последовательность *независимых* и *одинаково распределенных* с.в.

$$M[X_n] = m_X, D[X_n] = \sigma_X^2, M[|X_n - m_X|^3] = \rho < \infty$$

Тогда $\forall x \in \mathbb{R}$ и $n \in \mathbb{N}$

$$\left| P\left(\frac{\sum_{k=1}^n X_k - nm_X}{\sigma_X\sqrt{n}}\right) - \left(\frac{1}{2} + \Phi_0(x)\right) \right| \leq \frac{C_0\rho}{\sigma_X^3\sqrt{n}}$$

Замечание

Точное значение константы C **неизвестно**.

По текущим данным (2010 г.) $C_0 \leq 0.4784$

5.11 Пример

Оценим точность решения в предыдущем примере: $\sigma_X^2 = \frac{1}{4}, n = 1000, m_X = \frac{1}{2}, X_n \sim \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

$$\rho = M[|X_n - m_X|^3] = |0 - \frac{1}{2}|^3 \cdot \frac{1}{2} + |1 - \frac{1}{2}|^3 \cdot \frac{1}{2} = \frac{1}{8}$$

Тогда погрешность составит для двухстороннего нер-ва:

$$\frac{2C_0\rho}{\sigma_X^3\sqrt{n}} = \frac{2 \cdot 0.4784 \cdot \frac{1}{8}}{\frac{1}{64} \cdot \sqrt{1000}} \approx 0.03$$

6 Выборка и ее характеристики

6.1 Определение 1

Выборкой называется $Z_n = (X_1, \dots, X_n)^T$ независимый вектор с.в. Если все X_1, \dots, X_n — одинаково распределены, а $F(x)$ — функция распределения, то говорят, что Z_n — **однородная** выборка, порожденная распределением $F(x)$

6.2 Определение 2

Реализацией выборки $Z_n \in \mathbb{R}^n$ называется *неслучайный* вектор $z_n = Z_n(\omega)$, состоящий из реализаций элементов выборки $X_k, k = \overline{1, n}$.

6.3 Определение 3

Множество S всех возможных реализаций выборки Z_n называют **выборочным пространством**

Замечание

Обычно распределение, порождающее выборку, известно неточно.

$$F_X = F_X(x; \theta)$$

Задача состоит в построении оценки θ по элементам выборки.

6.4 Определение 4

С.в. $\varphi(Z_n)$, где $\varphi : S \rightarrow \mathbb{R}$ — измерима, называется **статистикой**.

6.5 Определение 5

k -ой порядковой статистикой называется k -е по величине значение элемента выборки $Z_n = (X_1, \dots, X_n)^T$ и обозначается $X^{(k)}$

Замечание

$X^{(k)}$ является функцией от всей выборки, т.к. при различных $\omega \in \Omega$ $X^{(k)}$ будет совпадать по значению с разными X_i .

6.6 Определение 6

Набор порядковых статистик $X^{(1)}, \dots, X^{(n)}$ называется **вариационным рядом**.

6.7 Определение 7

$$X^{(1)} = \min_{k=\overline{1, n}} X_k, X^{(n)} = \max_{k=\overline{1, n}} X_k.$$

6.8 Лемма 1

Пусть однородная выборка Z_n порождена распределением $F(x)$. Тогда функция распределения $X^{(k)}$ имеет вид:

$$F_{(k)}(x) = P(X^{(k)} \leq x) = \sum_{i=k}^n C_n^i (F(x))^i (1 - F(x))^{n-i}$$

Доказательство

Рассмотрим с.в. Y , равную числу элементов выборки, не превосходящих x . Тогда $Y \sim Bi(n; F(x))$.

$$F_{(k)}(x) = P(X^{(k)} \leq x) = P(Y \geq k) = \sum_{i=k}^n C_n^i (F(x))^i (1 - F(x))^{n-i}. \blacksquare$$

6.9 Следствие 1

$$\begin{aligned} F_{(1)}(x) &= 1 - (1 - F(x))^n, \\ F_{(n)}(x) &= (F(x))^n. \end{aligned}$$

6.10 Определение 8

Выборочным квантилем уровня $\alpha \in (0; 1)$ называется *порядковая статистика* $X^{([n\alpha]+1)}$

6.11 Теорема 1 (Мостеллера)

Пусть X — абсолютно непрерывная с.в., x_α — точка гладкости $f_X(x)$, $f_X(x_\alpha) > 0$

Тогда $(X^{([n\alpha]+1)} - x_\alpha) \sqrt{\frac{nf_X^2(x_\alpha)}{p(1-p)}} \xrightarrow{d} N(0; 1)$

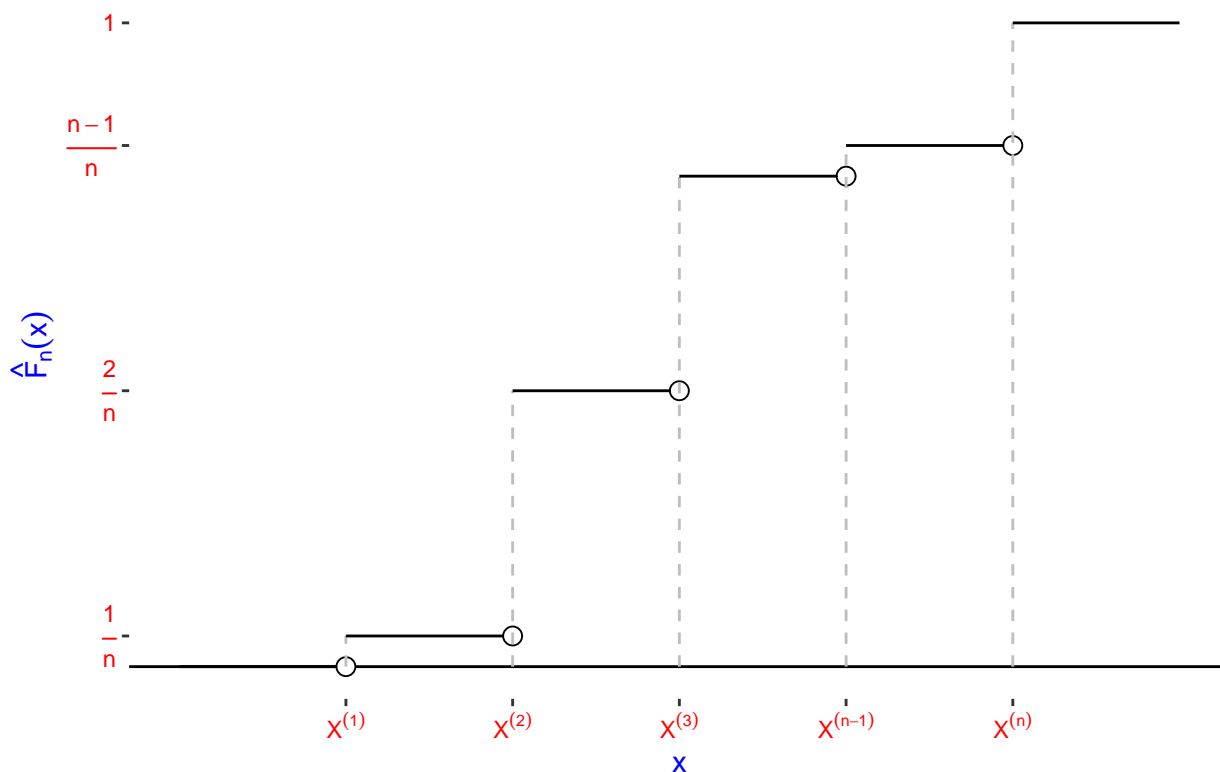
6.12 Определение 9

Выборочной функцией распределения называется статистика $\hat{F}_n(x)$:

$$\hat{F}_n(x) = \begin{cases} \frac{1}{n} \max\{k = \overline{1, n} : X^{(k)} \leq x\}, & x \geq X^{(1)}, \\ 0, & x < X^{(1)} \end{cases}$$

Замечание

Фактически $\hat{F}_n(x)$ — частота события $\{X \leq x\}$, которая используется для оценки вероятности $F(x) = P(X \leq x)$.



6.13 Свойства $\hat{F}_n(x)$

$$1. \quad n \cdot \hat{F}_n(x) \sim Bi(n; F(x))$$

$$2. \quad M[\hat{F}_n(x)] = F(x)$$

$$3. \quad \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{\text{п.н.}} 0$$

(Теорема Гливенко - Кантелли)

$$4. \quad M[(\hat{F}_n(x) - F(x))^2] = \frac{F(x)(1-F(x))}{n} \leq \frac{1}{4n}$$

$$5. \quad |\hat{F}_n(x) - F(x)| \xrightarrow{\text{с.к.}} 0$$

$$6. \quad \frac{\hat{F}_n(x) - F(x)}{\sqrt{F(x)(1-F(x))}} \sqrt{n} \xrightarrow{d} N(0; 1)$$

(Следует из теоремы Муавра - Лапласа)

Замечание

$\hat{F}_n(x)$ при увеличении n равномерно приближается к $F(x)$, при этом точность приближения можно оценить при помощи свойств 4 и 6.

Гистограмма. На основе реализации вариационного ряда построим разбиение $\mathbb{R} - \infty = t_0 < t_1 < t_2 < \dots < t_l < t_{l+1} = +\infty$,
 $t_1 \leq x^{(1)}, t_l > x^{(n)}$.

Как правило, длина интервалов разбиения выбирается *одинаковой*:

$$h_k = t_{k+1} - t_k = \frac{t_l - t_1}{l - 1}, k = \overline{1, l - 1}$$

Вычислим частоту попадания элементов выборки в k -й интервал:

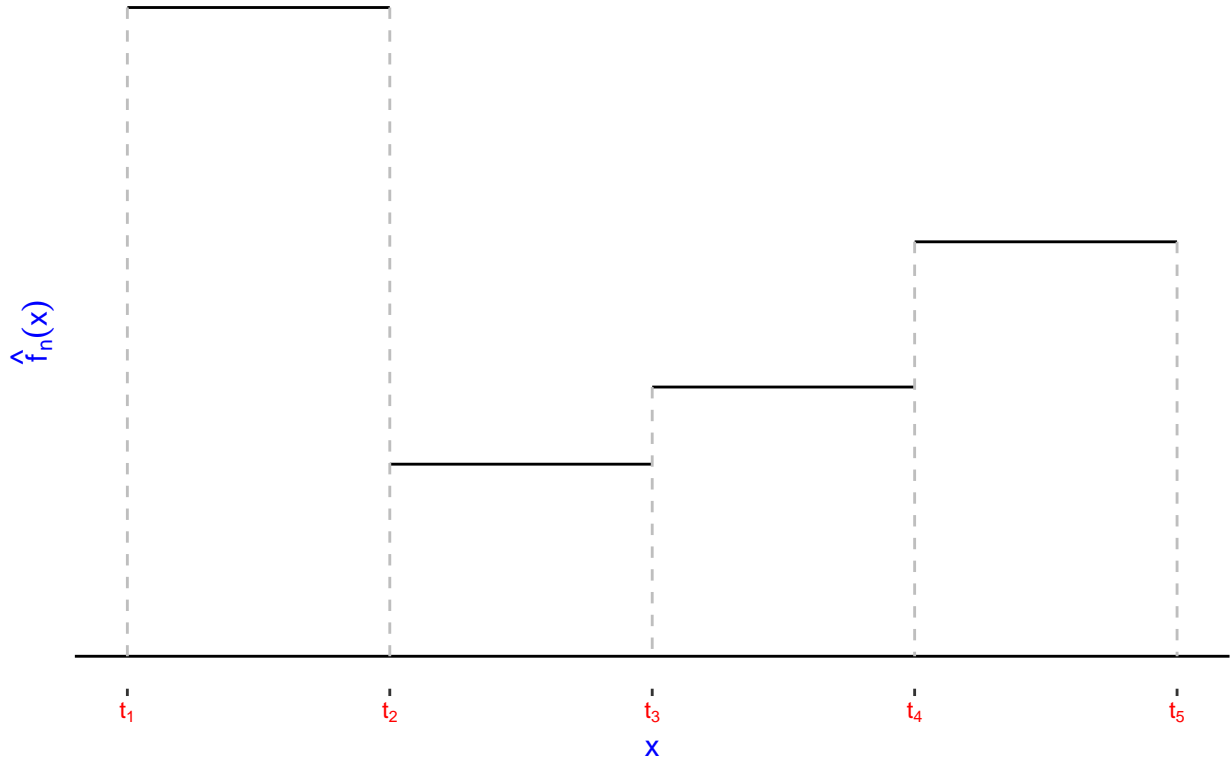
$\hat{p}_k = \frac{n_k}{n}$, где n_k — число элементов выборки, попавших в $[t_k; t_{k+1})$, $k = \overline{0, l}$

Заметим, что $\hat{p}_0 = \hat{p}_l = 0$.

6.14 Определение 10

Гистограммой называется функция:

$$\hat{f}_n(x) = \begin{cases} 0, & x \in (t_0; t_1) \cup [t_l; t_{l+1}), \\ \hat{p}_k, & x \in (t_k; t_{k+1}), k = \overline{1, l - 1} \end{cases}$$



Замечание

Если плотность вероятности $f_X(x)$ непрерывна и ограничена, а число разрядов гистограммы l_n удовлетворяет условию: $l_n \rightarrow +\infty, \frac{n}{l_n} \rightarrow +\infty$,
 то

$$\hat{f}_n(x) \xrightarrow{P} f_X(x)$$

Т.е. гистограмма является статистической аппроксимацией функции плотности вероятности.

Выборочные моменты

6.15 Определение 1

Выборочным начальным и центральным моментами называется соответственно статистики:

$$\hat{\nu}_r(n) = \frac{1}{n} \sum_{k=1}^n X_k^r$$

И

$$\hat{\mu}_r(n) = \frac{1}{n} \sum_{k=1}^n (X_k - \hat{\nu}_1(n))^r$$

6.16 Определение 2

Выборочным средним и выборочной дисперсией называются соответственно статистики:

$$\bar{X}_n = \hat{\nu}_1(n) = \frac{1}{n} \sum_{k=1}^n X_k$$

И

$$\hat{d}_X(n) = \hat{S}^2(n) = \hat{\mu}_2(n) = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2$$

6.17 Определение 3

Пусть $Z_n = (X_1, \dots, X_n)^T$ и $V_n = (Y_1, \dots, Y_n)^T$ — выборки, порожденные распределениями F_X и F_Y соответственно. Тогда *выборочным коэффициентом корреляции* называется:

$$\hat{r}_{XY} = \frac{\sum_{k=1}^n (X_k - \bar{X}_n)(Y_k - \bar{Y}_n)}{n\sqrt{\hat{d}_X \cdot \hat{d}_Y}}$$

6.18 Свойства выборочных моментов

$$1. \quad \boxed{M[\hat{\nu}_r(n)] = \nu_r, r \in \mathbb{N}}$$

Доказательство

$$M[\hat{\nu}_r(n)] = \frac{1}{n} \sum_{k=1}^n M[X_k^r] = \frac{1}{n} n \nu_r. \quad \blacksquare$$

$$2. \quad \boxed{\text{Если } M[X^r] < \infty, \text{ то } \hat{\nu}_r(n) \xrightarrow{\text{п.п.}} \nu_r}$$

Доказательство

$$3. \quad \boxed{\text{Если } M[X^r] < \infty, \text{ то } \hat{\mu}_r(n) \xrightarrow{\text{п.п.}} \mu_r}$$

Доказательство

$$\begin{aligned}\hat{\mu}_r(n) &= \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^r = \frac{1}{n} \sum_{k=1}^n \sum_{i=0}^r C_r^i X_k^i (-\hat{X}_n)^{r-i} = \sum_{i=0}^r C_r^i (-\hat{X}_n)^{r-i} \left(\frac{1}{n} \sum_{k=1}^n X_k^i \right) = \\ \sum_{i=0}^r C_r^i (-\hat{X}_n)^{r-i} \hat{\nu}_i(n) &\xrightarrow[\text{CB-BO (2)}]{\text{П.Н.}} \sum_{i=0}^r C_r^i (-\nu_1)^{r-i} \hat{\nu}_i = \dots = \mu_i. \blacksquare\end{aligned}$$

$$4. \quad \boxed{D[\bar{X}_n] = \frac{1}{n} D[X]}$$

Доказательство

С учетом независимости X_1, \dots, X_n

$$D[X\bar{X}_n] = D\left[\frac{1}{n} \sum_{k=1}^n X_k\right] = \frac{1}{n^2} \sum_{k=1}^n D[X_k] = \frac{nD[X]}{n^2}. \blacksquare$$

$$5. \quad \boxed{M[\hat{d}_X(n)] = \frac{n-1}{n} D[X]}$$

Доказательство

$$\begin{aligned}M[\hat{d}_X(n)] &= \frac{1}{n} \sum_{k=1}^n nM[(X_k - \bar{X}_n)^2] = M[(X_1 - \bar{X}_n)^2] = D[X_1 - \bar{X}_n]^2 = D\left[\frac{n-1}{n} X_1 - \right. \\ \left. \sum_{k=2}^n \frac{1}{n} \bar{X}_k\right] &\stackrel{\text{H.3}}{=} \frac{n-1}{n} D[X_1] + \frac{1}{n^2} \sum_{k=2}^n D[X_k] = D[X] \frac{n^2 - 2n + 1 + n - 1}{n^2} = D[X] \cdot \frac{n-1}{n}. \blacksquare\end{aligned}$$

$$6. \quad \boxed{\frac{\bar{X}_n - m_X}{\sigma_X} \sqrt{n} \xrightarrow{d} N(0; 1)}$$

Доказательство

Т.к. X_1, \dots, X_n независимые, $M[X_k] = m_X$, $D[X_k] = \sigma_X^2$, $k \in \mathbb{N}$, то в силу [Теоремы 5.1](#):

$$\frac{\sum_{k=1}^n X_k - nm_X}{\sigma_X \sqrt{n}} = \frac{\frac{1}{n} \sum_{k=1}^n X_k - m_X}{\frac{\sigma_X}{\sqrt{n}}} \xrightarrow{d} N(0; 1)$$

■

$$7. \quad \boxed{\frac{\hat{d}_X(n) - \sigma_X^2}{\sqrt{\mu_4 - \mu_2}} \sqrt{n} \xrightarrow{d} N(0; 1)}$$

7 Основные распределения в статистике

Точечные оценки

7.1 Определение 1

Пусть X_1, \dots, X_n — независимые с.в.

Тогда

$$Y = \sum_{k=1}^n X_k^2 \sim \chi^2(n)$$

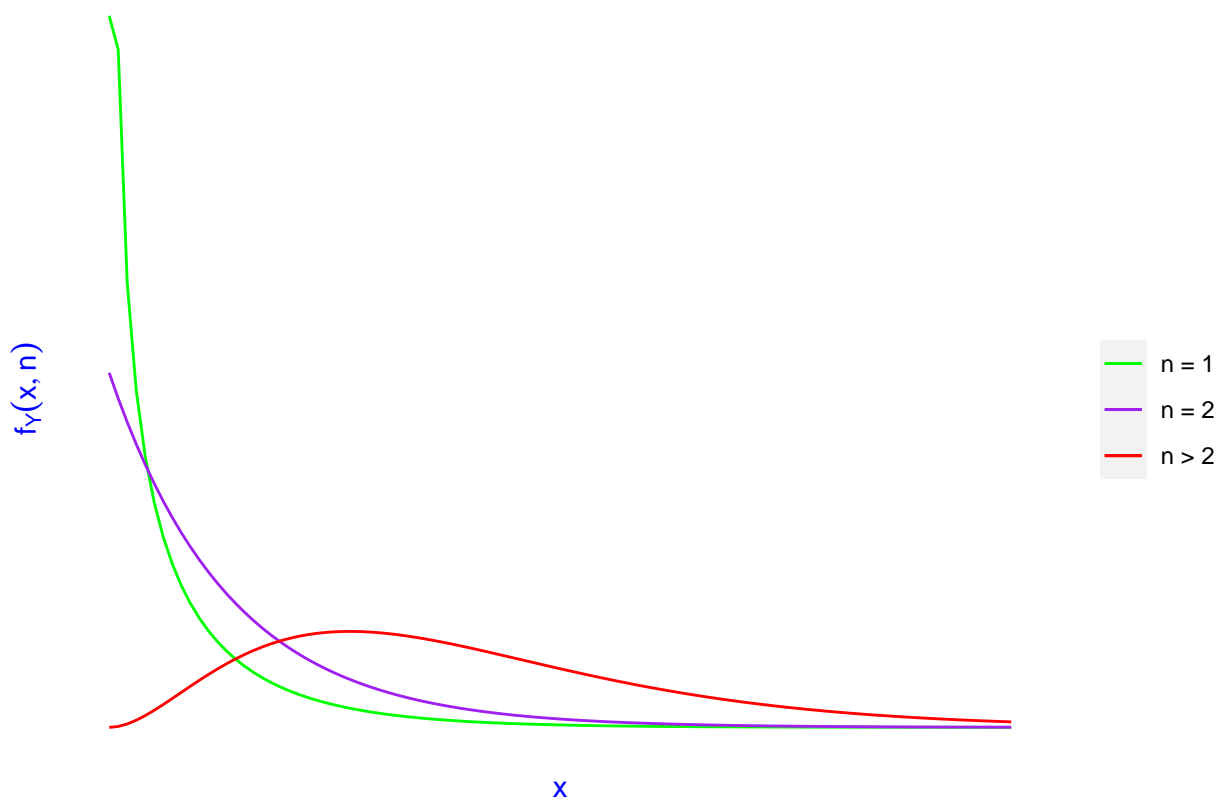
имеет *распределение хи-квадрат* с n степенями свободы.

7.2 Свойства распределения $\chi^2(n)$

1. Y имеет плотность вероятности

$$f_Y(x; n) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

где $\Gamma(z) = \int_0^{+\infty} y^{z-1} e^{-y} dy$ — гамма-функция.



2. Y имеет характеристическую функцию

$$\psi_Y(\lambda) = (1 - 2\lambda i)^{\frac{n}{2}}$$

3. $Y \sim \chi^2(n), M[Y] = n, D[Y] = 2n$

4. Пусть $n \geq 3$ $Y_1 \sim \chi^2(n_1), \dots, Y_k \sim \chi^2(n_k)$. Тогда

$$\sum_{k=1}^n Y_i \sim \chi^2\left(\sum_{k=1}^n n_i\right)$$

5.

$$\frac{Y - n}{\sqrt{2n}} \xrightarrow{d} N(0; 1)$$

6. Пусть Z_n порожденная распределением $N(m_X, \sigma_X^2)$. Тогда если

$$\hat{d}_X = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2,$$

то $\frac{n\hat{d}_X}{\sigma_X^2} \sim \chi^2(n-1)$.

(Доказательство приведено далее в теореме Фишера)

7.3 Определение 2

Пусть $X_0, X_1, \dots, X_n \sim N(0; 1)$ — независимые с.в.. Тогда с.в.

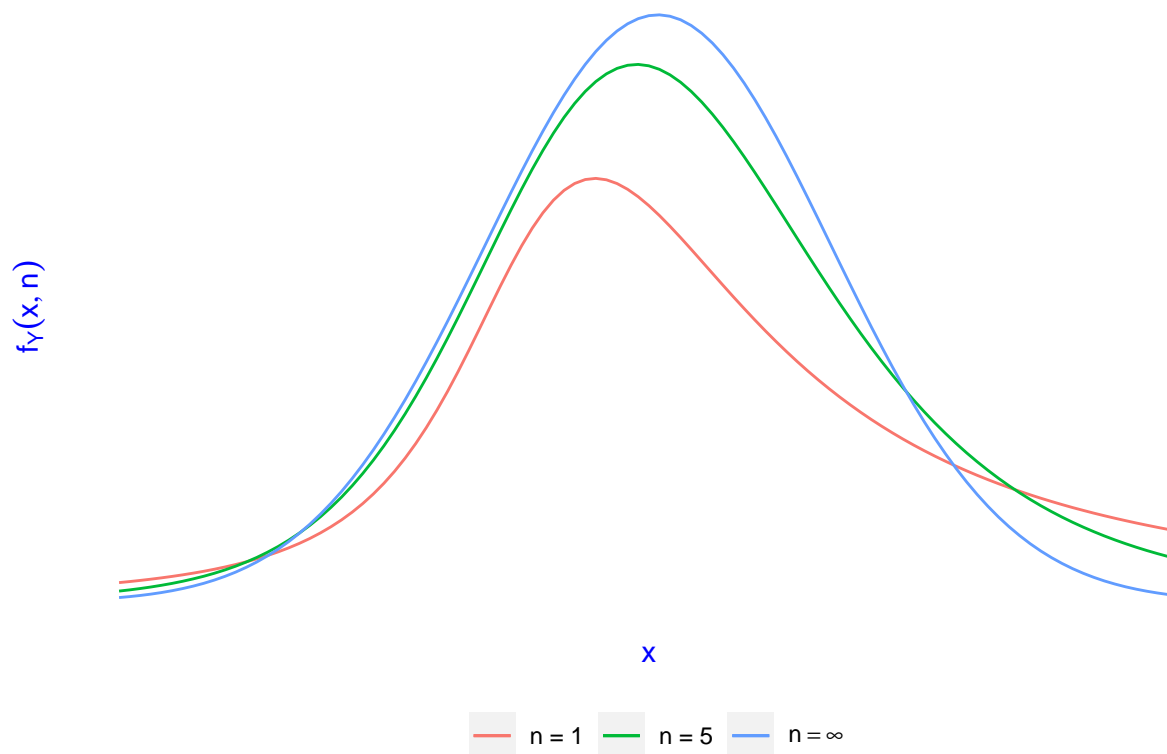
$$Y = \frac{X_0}{\sqrt{\frac{1}{n} \sum_{k=1}^n X_k^2}} \sim t(n)$$

имеем **распределение Стьюдента** с n степенями свободы.

7.4 Свойства распределения $t(n)$

1. Y имеет плотность вероятности

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



2. $Y \sim t(n), M[Y] = 0, n \geq 2$

$$D[Y] = \frac{n}{n-2}, n > 2$$

3. $t(1) = C$ — *распределение Коши*

4. $Y \sim t(n), Y \xrightarrow{d} N(0; 1)$

5. Пусть Z_n порождена распределением $N(m_X; \sigma_X^2)$. Тогда

$$\frac{\bar{X}_n - m_X}{\sqrt{\hat{d}_X}} \cdot \sqrt{n-1} \sim t(n-1)$$

(Доказательство приведено далее в теореме Фишера)

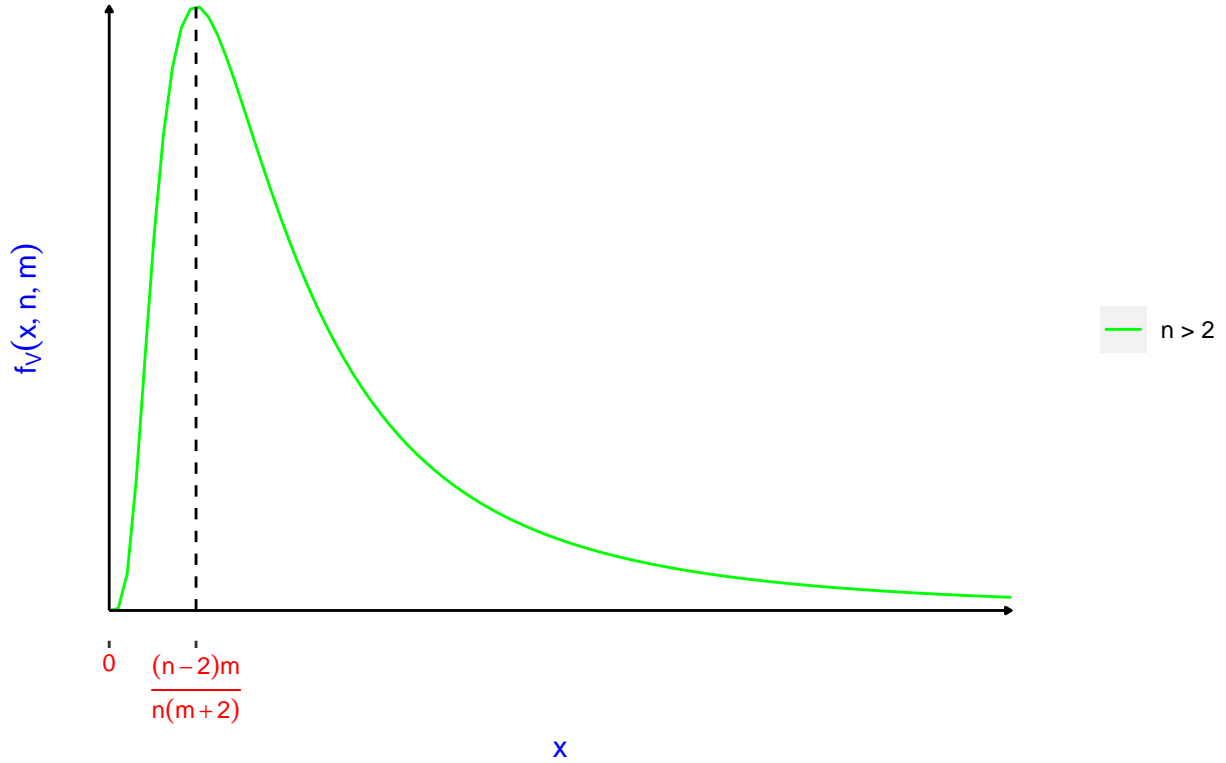
7.5 Определение 3

Пусть $X \sim \chi^2(n), Y \sim \chi^2(m)$ — независимые с.в. Тогда с.в. $V = \frac{Xm}{Yn} \sim F(n; m)$ имеет **Распределение Фишера** с n и m степенями свободы.

7.6 Свойства распределения $F(n; m)$

1. V имеет плотность вероятности

$$f_V(x, n, m) = \begin{cases} \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2}) \cdot \Gamma(\frac{m}{2})} n^{\frac{n}{2}} m^{\frac{m}{2}} \frac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{n+m}{2}}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$



2. $V \sim F(n; m), M[V] = \frac{m}{m-2}, m > 2,$

$$D[V] = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}, m > 4$$

3. Пусть $Z_n = (X_1, \dots, X_n)^T$ и $W_m = (Y_1, \dots, Y_m)^T$ — однородные выборки, порожденные распределениями $N(m_X; \sigma^2)$ и $N(m_Y; \sigma^2)$.

Тогда если Z_n и W_m независимы,

$$V = \frac{\frac{1}{n-1} \sum_{k=1}^n n(X_k - \bar{X}_n)^2}{\frac{1}{m-1} \sum_{k=1}^m n(Y_k - \bar{Y}_m)^2} \sim F(n-1; m-1)$$

(Данный факт следует из свойства б распределения $\chi^2(n)$)

7.7 Определение 4

Параметром $\theta \in \mathbb{R}^n$ **распределения** с.в. X называется любая числовая характеристика, входящая в $F_x(x, \theta)$ явно.

7.8 Определение 5

Точечной оценкой неизвестного параметра θ называется произвольная статистика $\hat{\theta}(Z_n)$.

Замечание

Оценка $\hat{\theta}(Z_n)$ является с.в.

На практике используется ее реализация.

7.9 Определение 6

Оценка $\hat{\theta}(Z_n)$ называется **несмещенной**, если

$$M[\hat{\theta}] = \theta$$

7.10 Определение 7

Оценка $\hat{\theta}(Z_n)$ называется **состоятельной**, если

$$\hat{\theta} \xrightarrow{P} \theta$$

7.11 Определение 8

Оценка $\hat{\theta}(Z_n)$ называется **сильно состоятельной**, если

$$\hat{\theta} \xrightarrow{\text{п.н.}} \theta$$

7.12 Определение 9

Оценка $\hat{\theta}(Z_n)$ называется **с.к. состоятельной**, если

$$\hat{\theta} \xrightarrow{\text{с.к.}} \theta$$

Замечание

1. Из свойств \bar{X}_n следует, что \bar{X}_n — **несмещенная** и **сильносостоятельная** оценка m_X .
2. Из свойств \hat{d}_X следует, что \hat{d}_X — **смещенная** и **сильносостоятельная** оценка m_X .

7.13 Определение 10

Несмещенная оценка $\hat{\theta}^*(Z_n)$ называется **эффективной**, если $\forall \hat{\theta}(Z_n)$ — **несмещенной** оценки верно, что

$$D[\hat{\theta}^*(Z_n)] \leq D[\hat{\theta}(Z_n)]$$

7.14 Пример

Пусть $M[\bar{X}] < \infty$. Тогда \bar{X} — *сильно состоятельная* оценка m_X . Если $D[\bar{X}] < \infty$, то \bar{X} — *с.к.-состоятельная* оценка m_X

(Доказательство следует из ЗБЧ)

7.15 Теорема 1

Пусть $\hat{\theta}_1, \hat{\theta}_2$, — *с.к.-оптимальные* оценки параметра θ . Тогда

$$\hat{\theta}_1 = \hat{\theta}_2$$

Доказательство

Т.к. $\hat{\theta}_1$ и $\hat{\theta}_2$ **оптимальны**, то $D[\hat{\theta}_1] = D[\hat{\theta}_2] = d$.

Пусть $\hat{\theta}_3 = \frac{1}{2}(\hat{\theta}_1 + \hat{\theta}_2)$. Тогда

$$D[\hat{\theta}_3] = \frac{1}{4}D[\hat{\theta}_1 + \hat{\theta}_2] = \frac{1}{4}(D[\hat{\theta}_1] + D[\hat{\theta}_2] + 2\text{cov}(\hat{\theta}_1, \hat{\theta}_2)) = \frac{1}{2}(d + \text{cov}(\hat{\theta}_1, \hat{\theta}_2)) \leq \frac{1}{2}(d + |\text{cov}(\hat{\theta}_1, \hat{\theta}_2)|) \leq \frac{1}{2}(d + \sqrt{D[\hat{\theta}_1] \cdot D[\hat{\theta}_2]}) = d$$

Тогда в силу оптимальности $\hat{\theta}_1$ и $\hat{\theta}_2$

$$D[\hat{\theta}_3] = d$$

$$d = D[\hat{\theta}_3] = \frac{1}{2}(d + \text{cov}(\hat{\theta}_1, \hat{\theta}_2)), \quad \text{cov}(\hat{\theta}_1, \hat{\theta}_2) = d = \sqrt{D[\hat{\theta}_1] \cdot D[\hat{\theta}_2]},$$

т.е. в неравенстве Коши-Буняковского достигается равенство. Следовательно,

$$\hat{\theta}_1 = \alpha \hat{\theta}_2 + \beta, \alpha > 0$$

$$\begin{cases} M[\hat{\theta}_1] = \alpha M[\hat{\theta}_2] + \beta \\ D[\hat{\theta}_1] = \alpha^2 D[\hat{\theta}_2] \end{cases} \implies \begin{cases} \theta = \alpha \theta + \beta \\ d = \alpha^2 d \end{cases} \implies \begin{cases} \beta = 0 \\ \alpha = 1 \end{cases}$$

Окончательно, $\hat{\theta}_1 = \hat{\theta}_2$. ■

8 Эффективные оценки

Обозначим через $f(x; \theta)$ плотность вероятности с.в. X , порождающей **выборку** в *абсолютно непрерывном случае* или **функцию** в *дискретном случае*. В силу критерия независимости функция

$$L(z_n; \theta) = \prod_{k=1}^n f(x_k; \theta)$$

является *плотностью вероятности* с.в. Z_n .

8.1 Определение 1

$L(z_n; \theta)$ при фиксированном $z_n \in S$ и переменной θ называется **функцией правдоподобия**.

Замечание

Далее будем полагать, что $\theta \in \mathbb{R}^1$.

8.2 Определение 2

Распределение с.в. X называется **регулярным**, если

1. $\sqrt{f(x; \theta)}$ — дифференцируема по θ почти для всех x .
2. $i(\theta) = \int_{-\infty}^{+\infty} \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 f(x; \theta) dx$ конечна, непрерывна по θ и положительна.

Замечание

Далее будем предполагать, что выборка Z_n порождена *регулярным распределением*.

8.3 Определение 3

Случайная величина

$$U(Z_n; \theta) = \frac{\partial \ln L(Z_n; \theta)}{\partial \theta} = \sum_{k=1}^n \frac{\partial \ln f(x; \theta)}{\partial \theta}$$

называется **вкладом выборки** Z_n

8.4 Лемма 1

Пусть распределение *регулярное*. Тогда

$$M[U(Z_n; \theta)] = 0$$

Доказательство

В силу условия нормировки

$$\int_{\mathbb{R}^n} L(z_n; \theta) d_{X_1}, \dots, d_{X_n} = 1$$

С учетом условий регулярности

$$0 = \frac{\partial}{\partial \theta} \int_{\mathbb{R}^n} L(z_n; \theta) d_{X_1}, \dots, d_{X_n} = \int_{\mathbb{R}^n} \frac{\partial L(z_n; \theta)}{\partial \theta} d_{X_1}, \dots, d_{X_n} = \int_{\mathbb{R}^n} \frac{\partial \ln L(z_n; \theta)}{\partial \theta} L(z_n; \theta) d_{X_1}, \dots, d_{X_n} = M[U(Z_n; \theta)]. \blacksquare$$

8.5 Определение 4

Информацией Фишера о параметре θ , содержащейся в выборке Z_n , называют

$$I_n(\theta) = D[U(Z_n; \theta)] \stackrel{\text{регул.}}{=} M[U^2(Z_n; \theta)]$$

$i(\theta) = M[(\frac{\partial \ln f(x; \theta)}{\partial \theta})^2]$ называется **количеством информации Фишера, содержащимся в одном наблюдении**.

Замечание

Из определения $U(Z_n; \theta)$ и независимости элементов выборки следует, что $I_n(\theta) = n \cdot i(\theta)$, т.е. количество информации **вырастает пропорционально объему выборки**.

8.6 Лемма 2

Пусть $f(x; \theta)$ дважды непрерывно дифференцируема по θ . Тогда

$$i(\theta) = -M[\frac{\partial^2 \ln f(X_1; \theta)}{\partial \theta^2}]$$

Доказательство

$U(X_1; \theta) = \frac{\partial \ln f(X_1; \theta)}{\partial \theta}$. С учетом **Леммы 1**

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} M[U(X_1; \theta)] = \frac{\partial}{\partial \theta} \int_{-\infty}^{+\infty} \frac{\partial \ln f(x; \theta)}{\partial \theta} f(x; \theta) dx = \int_{-\infty}^{+\infty} \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} f(x; \theta) dx + \int_{-\infty}^{+\infty} \frac{\partial \ln f(x; \theta)}{\partial \theta} \cdot \\ &\frac{\partial f(x; \theta)}{\partial \theta} dx = \int_{-\infty}^{+\infty} \frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} f(x; \theta) dx + \int_{-\infty}^{+\infty} (\frac{\partial \ln f(x; \theta)}{\partial \theta})^2 f(x; \theta) dx = M[\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2}] + i(\theta). \blacksquare \end{aligned}$$

8.7 Пример 1

Пусть $X \sim N(\theta; \sigma^2)$.

$$f(x; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(x - \theta)^2}{2\sigma^2}\}$$

$$U(X_1; \theta) = \frac{\partial}{\partial \theta} (-\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(x - \theta)^2}{2\sigma^2}) = \frac{X - \theta}{\sigma^2}$$

$$\frac{\partial^2 \ln f(x; \theta)}{\partial \theta^2} = -\frac{1}{\sigma^2}$$

С учетом [Леммы 2](#)

$$i(\theta) = -M\left[-\frac{1}{\sigma^2}\right] = \frac{1}{\sigma^2}$$

8.8 Пример 2

Рассмотрим нерегулярную модель.

$$X \sim R(0; \theta)$$

Здесь из множества $\int_0^\theta \frac{1}{\theta} dx = 1$ не следует, что $\int_0^\theta \frac{\partial}{\partial \theta} \left(\frac{1}{\theta}\right) dx = 0$, т.к. при дифференцировании по θ появляется еще одно слагаемое:

$$\frac{\partial}{\partial \theta} \int_0^\theta \frac{1}{\theta} dx = \frac{1}{\theta} + \int_0^\theta \frac{\partial}{\partial \theta} \left(\frac{1}{\theta}\right) dx$$

8.9 Теорема 1 (Неравенство Рао-Крамера)

Пусть распределение $F(x; \theta)$, порождающее выборку Z_n *регулярно*. Тогда для любой несмещенной оценки $\hat{\theta}$ верно неравенство:

$$D[\hat{\theta}(Z_n)] \geq \frac{1}{I_n(\theta)} = \frac{1}{ni(\theta)}$$

При этом равенство достигается лишь в том случае, если $\hat{\theta}(Z_n) - \theta = a(\theta) \cdot U(Z_n; \theta)$, где $a(\theta)$ — некоторая функция от θ

Доказательство

В силу несмещенности $\hat{\theta}$

$$M[\hat{\theta}] = \int_{\mathbb{R}^n} (z_n) L(z_n; \theta) dx_1, \dots, dx_n = \theta$$

В силу *регулярности* и [Леммы 1](#)

$$1 = \frac{\partial}{\partial \theta}(\theta) = \frac{\partial}{\partial \theta} M[\hat{\theta}] = \int_{\mathbb{R}^n} \hat{\theta}(z_n) \frac{\partial L(z_n; \theta)}{\partial \theta} dx_1, \dots, dx_n = \int_{\mathbb{R}^n} \hat{\theta}(z_n) \frac{\partial \ln L(z_n; \theta)}{\partial \theta} \cdot L(z_n; \theta) dx_1, \dots, dx_n =$$

$$M[\hat{\theta} U(Z_n; \theta)] = M[(\hat{\theta} - \theta)(U(Z_n; \theta) - \theta)] + \theta \cdot M[U(Z_n; \theta)] = \text{cov}(\hat{\theta}, U(Z_n; \theta))$$

Откуда с учетом *неравенства Коши-Буняковского*

$$1^2 \leq D[\hat{\theta}] \cdot D[U(Z_n; \theta)] = D[\hat{\theta}] \cdot I_n(\theta)$$

Причем *равенство достигается* в том и только том случае, когда $\hat{\theta} = a(\theta)U(Z_n; \theta) + b(\theta)$
Но с учетом [Леммы 1](#) $b(\theta) = 0$.

8.10 Определение 5

Оценка $\hat{\theta}^*(Z_n)$, для которой достигается равенство в неравенстве Рао-Крамера называется **эффективной**

Замечание

В силу [Теоремы 1](#) эффективная оценка является *оптимальной*. А с учетом [Теоремы 7.1](#) эффективная оценка **единственна**

8.11 Пример 3

Пусть $X \sim N(\theta; \sigma^2)$

$$U(Z_n; \theta) = \sum_{k=1}^n \frac{\partial \ln f(X_k; \theta)}{\partial \theta} = \sum_{k=1}^n \frac{X_k - \theta}{\sigma^2} = \frac{n}{\sigma^2} (\bar{X}_n - \theta)$$

Т.о. $a(\theta) = \frac{\sigma^2}{n}$. Тогда $a(\theta)U(Z_n; \theta) = \bar{X}_n - \theta$, откуда следует, что \bar{X}_n — *эффективная оценка*.

9 Методы построения точечных оценок

Метод максимального правдоподобия

9.1 Определение 1

Оценкой максимального правдоподобия θ называют

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} L(Z_n; \theta)$$

Замечание

1. В силу *монотонности* функции $\ln x$ справедливо представление:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} L(Z_n; \theta) = \underset{\theta}{\operatorname{argmax}} \ln L(Z_n; \theta)$$

2. Если $L(Z_n; \theta)$ — **гладкая** и **максимум** по θ достигается **внутри множества** возможных значений θ , то θ можно вычислить из *уравнения правдоподобия*

$$U(Z_n; \theta) = \frac{\partial \ln L(z_n; \theta)}{\partial \theta} = 0$$

3. Из [Теоремы 8.1](#) следует, что $\hat{\theta}$ **будет также эффективной оценкой**.

9.2 Пример 1

Рассмотрим случайную величину $X \sim N(m; \sigma^2)$ с неизвестными m и σ^2 :

$$\theta = (m; \sigma^2)^T$$

$$L(Z_n; \theta) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(X_k - \theta_1)^2}{2\theta_2}}$$

$$\ln L(Z_k; \theta) = \sum_{k=1}^n n \left(-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \theta_2 - \frac{(X_k - \theta_1)^2}{2\theta_2} \right)$$

$$\begin{cases} \frac{\partial \ln L(Z_n; \theta)}{\partial \theta_1} = \sum_{k=1}^n \frac{(X_k - \theta_1)}{\theta_2} = \frac{1}{\theta_2} (n\bar{X}_n - n\theta_1) = 0 \\ \frac{\partial \ln L(Z_n; \theta)}{\partial \theta_2} = \sum_{k=1}^n \left(\frac{(X_k - \theta_1)^2}{2\theta_2^2} - \frac{1}{2\theta_2} \right) = \frac{1}{2\theta_2} (n\hat{d}_X(n) - n\theta_2) = 0 \\ \hat{\theta}_1 = \bar{X}_n \\ \hat{\theta}_2 = \hat{d}_X(n) \end{cases}$$

9.3 Пример 2

Пусть $X \sim R(\theta_1; \theta_2)$. В этом случае $L(Z_n; \theta)$ **не является** непрерывной:

$$L(Z_n; \theta) = \prod_{k=1}^n f_X(X_k; \theta) = \begin{cases} 0, & \exists k = \overline{1, n} : X_k \notin [\theta_1; \theta_2] \\ \frac{1}{(\theta_2 - \theta_1)^n}, & \forall k = \overline{1, n} : X_k \in [\theta_1; \theta_2] \end{cases}$$

Тогда оценка максимального правдоподобия **не может быть вычислена из уравнений правдоподобия**. Хотя $\hat{\theta}$ существует:

$$L(Z_n; \theta) = \begin{cases} 0, & \min_{k=\overline{1, n}} X_k < \theta_1 \text{ или } \max_{k=\overline{1, n}} X_k > \theta_2 \\ \frac{1}{(\theta_2 - \theta_1)^n}, & \theta_1 \leq \min_{k=\overline{1, n}} X_k \leq \max_{k=\overline{1, n}} X_k \leq \theta_2 \end{cases}$$

Откуда $L(Z_n; \theta)$ *возрастает* по θ_1 и *убывает* по θ_2

Тогда $\hat{\theta}_1 = \min_{k=\overline{1, n}} X_k = X^{(1)}$, $\hat{\theta}_2 = \max_{k=\overline{1, n}} X_k = X^{(n)}$

Замечание

Для МП-оценок выполняется **принцип инвариантности**:

Пусть $g(\theta)$ — *биективное* отображение.

Тогда МП-оценка $\hat{g}(Z_n) = g(\hat{\theta}(Z_n))$.

Действительно,

$$\sup_{\theta} L(Z_n; \theta) = \sup_g L(Z_n; g^{-1}(g))$$

Тогда $g^{-1}(\hat{g}) = \hat{\theta}$, т.е. $\hat{g} = g(\hat{\theta})$

9.4 Пример 3

Пусть $X \sim N(\theta_1; \theta_2)$. Требуется оценить $F_X(x_0) = \frac{1}{2} + \Phi_0\left(\frac{x_0 - \theta_1}{\theta_2}\right)$. Рассмотрим *биективное отображение*

$$g(\theta_1; \theta_2) = \begin{pmatrix} \frac{1}{2} + \Phi_0\left(\frac{x_0 - \theta_1}{\theta_2}\right) \\ \theta_2 \end{pmatrix}$$

Тогда

$$\hat{g}(Z_n) = \begin{pmatrix} \frac{1}{2} + \Phi_0\left(\frac{x_0 - \bar{X}_n}{\sqrt{\hat{d}_X(n)}}\right) \\ \sqrt{\hat{d}_X(n)} \end{pmatrix}$$

$X_0 \in \mathbb{R}$

Замечание

Для решения уравнений правдоподобия часто используются *численные методы*

9.5 Теорема 1

Пусть распределение с.в. X , порождающей выборку Z_n , *регулярно*.

Функция правдоподобия $L(z_n; \theta)$ имеет единственный достижимый максимум по $\theta \forall z_n \in S, n \in \mathbb{N}$. Тогда

1. МП-оценка $\hat{\theta}$ *состоятельна*;

2. Если $|\frac{\partial^k f(x; \theta)}{\partial \theta^k}| \leq g_k(x), \forall \theta$, где

$\int_{\mathbb{R}} g_1(x) dx < \infty, \int_{\mathbb{R}} g_2(x) dx < \infty, \int_{\mathbb{R}} g_3(x) f(x; \theta) dx \leq C \leq \infty$, а функция $i(\theta) = \int_{\mathbb{R}} (\frac{\partial \ln f(x; \theta)}{\partial \theta})^2 f(x; \theta) dx$ конечна и положительна $\forall \theta$, то

2.1 $M[\hat{\theta}] \rightarrow \theta$ (*асимптотически несмещенность*)

2.2 $\hat{\theta}$ *сильно состоятельна*;

2.3 $\sqrt{ni(\theta)}(\hat{\theta} - \theta) \xrightarrow{d} N(0; 1)$

Асимптотическая нормальная

Метод моментов

9.6 Определение 2

Пусть $\theta = (\theta_1, \dots, \theta_r)^T$, а для распределения $F_X(x; \theta)$, порождающего выборку $Z_n, M[X^r] < \infty$

Тогда

$$\begin{cases} \nu_1(\theta) = \hat{\nu}_1(n) \\ \vdots \\ \nu_r(\theta) = \hat{\nu}_r(n) \end{cases}$$

называется *системой метода моментов*.

9.7 Определение 3

Решение системы метода моментов

$$\hat{\theta}_i = \varphi_i(\hat{\nu}_1(n), \dots, \hat{\nu}_r(n)), i = \overline{1, r}$$

называется *оценкой метода моментов*.

9.8 Теорема 2

Пусть функции $\varphi_1, \dots, \varphi_r$, определяющие оценку метода моментов *непрерывные и биективные*. Тогда оценка метода моментов *состоятельна*.

Доказательство

Доказательство следует из **состоятельности** статистик $\hat{\nu}_i(n)$.

9.9 Пример 1

$X \sim N(\theta_1; \theta_2)$. Тогда $\hat{\theta}_1 = \bar{X}_n, \hat{\theta}_2 = \hat{d}_X(n)$

9.10 Пример 2

$X \sim R(\theta_1; \theta_2)$. Тогда

$$\begin{cases} \bar{X}_n = \frac{\theta_1 + \theta_2}{2} \\ \hat{d}_X(n) = \frac{(\theta_2 - \theta_1)^2}{12} \end{cases} \quad \begin{cases} \theta_1 + \theta_2 = 2\bar{X}_n \\ \theta_2 - \theta_1 = 2\sqrt{\hat{d}_X(n)}\sqrt{3} \end{cases} \quad \begin{cases} \theta_1 = \bar{X}_n - \sqrt{3\hat{d}_X(n)} \\ \theta_2 = \bar{X}_n + \sqrt{3\hat{d}_X(n)} \end{cases}$$

Замечание

Метод моментов **трудноприменим**, если *теоретические моменты не удается вычислить явно*.

10 Интервальные оценки

10.1 Определение 1

Пусть выполнено условие:

$$P(\hat{\theta}_1(Z_n) \leq \theta \leq \hat{\theta}_2(Z_n)) = 1 - \alpha$$

для некоторого распределения параметром $F_x(x; \theta)$. Тогда интервал $[\hat{\theta}_1(Z_n); \hat{\theta}_2(Z_n)]$ — называется **доверительным интервалом** или **интервальной оценкой** параметра θ уровня надежности $1 - \alpha$.

10.2 Определение 2

Доверительный интервал называется **центральный**, если верно условие:

$$P(\theta \leq \hat{\theta}_1(Z_n)) = P(\theta \geq \hat{\theta}_2(Z_n)) = \frac{\alpha}{2}$$

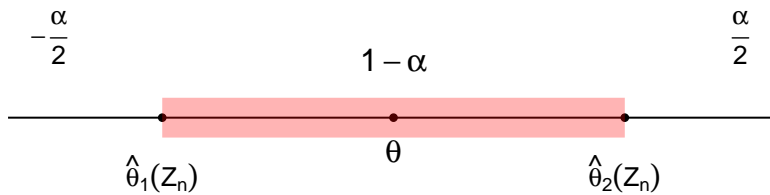
10.3 Определение 3

Доверительный интервал называется **правосторонним** или **левосторонним**, если выполнено условие:

$$P(\theta \geq \hat{\theta}_2(Z_n)) = \alpha \text{ или } P(\theta \leq \hat{\theta}_1(Z_n)) = \alpha$$

Замечание

1. Можно визуализировать интервальную оценку следующим образом:



2. Интервальную оценку можно рассматривать в качестве оценки погрешности точечной оценки $\hat{\theta}(Z_n)$, если строить доверительный интервал в форме:

$$[\hat{\theta}(Z_n) - \varepsilon_1; \hat{\theta}(Z_n) + \varepsilon_2]$$

где $\varepsilon_1, \varepsilon_2 > 0$.

Построение доверительного интервала на основе центральной статистики

10.4 Определение 4

Статистика $G(Z_n; \theta)$ называется **центральной**, если G непрерывна и строго монотонна по θ , $\forall z_n \in S$, а ее распределение **не зависит** от θ .

Если распределение $F_G(Z_n; \theta)$ известно, то можно подобрать числа g_1, g_2 такие, что $P(g_1 \leq G(Z_n; \theta) \leq g_2) = 1 - \alpha$.

Например, для центрального доверительного интервала $g_1 = G_{\frac{\alpha}{2}}, g_2 = G_{1-\frac{\alpha}{2}}$ — квантили уровней $\frac{\alpha}{2}$ и $1 - \frac{\alpha}{2}$.

Если $G(Z_n; \theta)$ монотонно возрастает по θ , то можно перейти к эквивалентному неравенству:

$$P(G^{-1}(Z_n; g_1) \leq \theta \leq G^{-1}(Z_n; g_2)) = 1 - \alpha$$

Тогда границы интервальной оценки имеют вид:

$$\hat{\theta}_1(Z_n) = G^{-1}(Z_n; g_1), \hat{\theta}_2(Z_n) = G^{-1}(Z_n; g_2)$$

Замечание

Построение центральных статистик довольно сложно. В связи с чем зачастую используют **асимптотические распределения**, которые в силу ЦПТ тесно связаны с нормальным распределением.

10.5 Определение 5

Матрица $C \in \mathbb{R}^{n \times n}$ называется **ортогональной**, если $C^T = C^{-1}$.

10.6 Лемма 1

Пусть $C \in \mathbb{R}^{n \times n}$ — ортогональная матрица, $C_{1j} = \frac{1}{\sqrt{n}}, j = \overline{1, n}$. Тогда

1. $\sum_{k=1}^n C_{jk} \cdot C_{ik} = 0, j \neq i; i, j = \overline{1, n}$
2. $\sum_{k=1}^n C_{ik}^2 = 1, i = \overline{1, n}$
3. Если $y = Cx$, то $\|y\| = \|x\|$
4. $\sum_{k=1}^n C_{ik} = 0, i = \overline{2, n}$

Доказательство

Пункты 1 и 2 следуют из тождества

$$C^T \cdot C = I$$

$$3. \|y\|^2 = (y, y) = (Cx, Cx) = (x, C^T Cx) = (x, x) = \|x\|^2$$

4. Рассмотрим свойство 1 для $i = 1$. Тогда

$$0 = \sum_{k=1}^n C_{1k} \cdot C_{ik} = \sum_{k=1}^n \frac{1}{\sqrt{n}} C_{ik} = \frac{1}{\sqrt{n}} \sum_{k=1}^n C_{ik}$$

■

10.7 Теорема 1 (ФИШЕРА)

Пусть Z_n — однородная выборка, порожденная распределением $N(m_X; \sigma_X^2)$. Тогда

1. $\frac{\bar{X}_n - m_X}{\sigma_X} \sqrt{n} \sim N(0; 1)$
2. $\frac{n\hat{d}_X(n)}{\sigma_X^2} \sim \chi(n-1)$
3. $\frac{\bar{X}_n - m_X}{\sqrt{\hat{d}_X(n)}} \sqrt{n-1} \sim t(n-1)$
4. \bar{X}_n и $\hat{d}_X(n)$ **независимы**

Доказательство

$$.1) \frac{\bar{X}_n - m_X}{\sigma_X} \sqrt{n} = \sum_{k=1}^n \frac{1}{\sigma_X \sqrt{n}} X_k - \frac{m_X \sqrt{n}}{\sigma_X} = \left(\frac{1}{\sigma_X \sqrt{n}} + \frac{1}{\sigma_X \sqrt{n}} + \dots + \frac{1}{\sigma_X \sqrt{n}} \right) \cdot Z_n - \frac{m_X \sqrt{n}}{\sigma_X}.$$

Т.к. $Z_n \sim N\left(\begin{pmatrix} m_X \\ \vdots \\ m_X \end{pmatrix}; \sigma_X^2 I\right)$, то в силу [Леммы 1.3](#)

$$\frac{\bar{X}_n - m_X}{\sigma_X} \sqrt{n} \sim N(m; \sigma^2), \text{ где } m = \left(\frac{1}{\sigma_X \sqrt{n}} \dots \frac{1}{\sigma_X \sqrt{n}} \right) \begin{pmatrix} m_X \\ \vdots \\ m_X \end{pmatrix} - \frac{m_X \sqrt{n}}{\sigma_X} = 0$$

$$\sigma^2 = \left(\frac{1}{\sigma_X \sqrt{n}} \dots \frac{1}{\sigma_X \sqrt{n}} \right) \sigma^2 I \cdot \begin{pmatrix} \frac{1}{\sigma_X \sqrt{n}} \\ \vdots \\ \frac{1}{\sigma_X \sqrt{n}} \end{pmatrix} = 1$$

.2) Построим матрицу $C \in \mathbb{R}^{n \times n}$ так, чтобы $C_{1j} = \frac{1}{\sqrt{n}}, j = \overline{1, n}$; остальные строки выберем так, чтобы векторы $\begin{pmatrix} C_{11} \\ \vdots \\ C_{1n} \end{pmatrix}, \begin{pmatrix} C_{21} \\ \vdots \\ C_{2n} \end{pmatrix}, \dots, \begin{pmatrix} C_{n1} \\ \vdots \\ C_{nn} \end{pmatrix} \in \mathbb{R}^n$

образовывали *ортонормированный* базис, что возможно, т.к. вектор $\begin{pmatrix} C_{11} \\ \vdots \\ C_{1n} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} \\ \vdots \\ \frac{1}{\sqrt{n}} \end{pmatrix}$

всегда можно дополнить до *ортонормированного* базиса.

Тогда C удовлетворяет условиям [Леммы 1](#).

Рассмотрим $Y = \frac{1}{\sigma_X} CZ_n$. Тогда $Y \sim N(m_Y; K_Y)$, где в силу [Леммы 1.3](#) и [Леммы 1, п.4](#):

$$m_Y = \frac{1}{\sigma_X} C(\sigma_X^2 I) \frac{1}{\sigma_X} C^T = C I C^T = C C^T = I$$

Т.о. Y_1, \dots, Y_n — **некоррелированные** нормальные с.в.

Тогда в силу [Леммы 1.4](#) Y_1, \dots, Y_n *независимы*

$$\frac{n\hat{d}_X(n)}{\sigma_X^2} = \frac{1}{\sigma_X^2} \sum_{k=1}^n (X_k - \bar{X}_n)^2 = \frac{1}{\sigma_X^2} \sum_{k=1}^n (X_k^2 - 2X_k \bar{X}_n + \bar{X}_n^2) = \frac{1}{\sigma_X^2} \left(\sum_{k=1}^n X_k^2 - 2n\bar{X}_n \left(\frac{1}{n} \sum_{k=1}^n X_k \right) + n\bar{X}_n^2 \right) = \frac{1}{\sigma_X^2} \sum_{k=1}^n X_k^2 - \frac{n}{\sigma_X^2} \bar{X}_n^2 =$$

$$Y_1 = \frac{1}{\sigma_X \sqrt{n}} (X_1 + \dots + X_n) = \frac{\sqrt{n}}{\sigma_X} \bar{X}_n. \text{ Тогда с учетом } \textcolor{blue}{\text{Леммы 1}}$$

$$= \frac{1}{\sigma_X^2} \|Z_n\|^2 - Y_1^2 = \frac{1}{\sigma_X^2} \|CZ_n\|^2 - Y_1^2 = \left\| \frac{1}{\sigma_X} Z_n \right\|^2 - Y_1^2 = \|Y\|^2 - Y_1^2 = \sum_{k=1}^n Y_k^2 - Y_1^2 = \sum_{k=2}^n Y_k^2, \text{ где}$$

$Y_k \sim N(0; 1), k = \overline{2, n}$. Тогда по определению

$$\frac{n\hat{d}_X(n)}{\sigma_X^2} = \sum_{k=2}^n Y_k^2 \sim \chi^2(n-1)$$

.4) Заметим, что $\bar{X}_n = \frac{\sigma_X}{\sqrt{n}} Y_1 = \varphi(Y_1)$, где Y_1 и $(Y_2, \dots, Y_n)^T$ **независимы**. Тогда **независимы** также $\varphi(Y_1)$ и $\psi(Y_2, \dots, Y_n)$, т.е. \bar{X}_n и $\hat{d}_X(n)$

$$.3) \frac{\bar{X}_n - m_X}{\sqrt{\hat{d}_X(n)}} \sqrt{n-1} = \frac{\frac{\sigma_X}{\sqrt{n}} Y_1 - m_X}{\sqrt{\frac{\sigma_X^2}{n} \sum_{k=2}^n Y_k^2}} \sqrt{n-1} = \frac{Y_1 - \frac{m_X \sqrt{n}}{\sigma_X}}{\sqrt{\frac{1}{n-1} \sum_{k=2}^n Y_k^2}} \sim t(n-1), \text{ т.к.}$$

$$Y_1 - \frac{m_X \sqrt{n}}{\sigma_X} \sim N(0; 1) \text{ и } Y_1, (Y_2, \dots, Y_n)^T \text{ независимы. } \blacksquare$$

10.8 Пример 3

Построить интервальную оценку θ по однородной выборке Z_n , порожденной распределением $N(\theta; \sigma_X^2)$, если σ_X^2 *неизвестна*, уровня надежности $1 - \alpha$

В силу [Теоремы Фишера](#)

$$G(Z_n; \theta) = \frac{\bar{X}_n - \theta}{\sqrt{\hat{d}_X(n)}} \sqrt{n-1} \sim (n-1)$$

Тогда

$$1 - \alpha = P\left(t_{\frac{\alpha}{2}}(n-1) \leq \frac{\bar{X}_n - \theta}{\sqrt{\hat{d}_X(n)}} \sqrt{n-1} \leq t_{1-\frac{\alpha}{2}}(n-1)\right) = P\left(\bar{X}_n - \sqrt{\frac{\hat{d}_X(n)}{n-1}} t_{1-\frac{\alpha}{2}}(n-1) \leq \theta \leq \bar{X}_n + \sqrt{\frac{\hat{d}_X(n)}{n-1}} t_{1-\frac{\alpha}{2}}(n-1)\right)$$

Тогда интервальная оценка имеет вид:

$$\left[\bar{X}_n - \sqrt{\frac{\hat{d}_X(n)}{n-1}} t_{1-\frac{\alpha}{2}}(n-1); \bar{X}_n + \sqrt{\frac{\hat{d}_X(n)}{n-1}} t_{1-\frac{\alpha}{2}}(n-1) \right]$$

Заметим, что \bar{X}_n — *точечная оценка* θ — лежит в середине интервала. При этом длина интервала $2\sqrt{\frac{\hat{d}_X(n)}{n-1}} t_{1-\frac{\alpha}{2}}(n-1)$ **возрастает** по σ_X^2 и **убывает** по α и n .

10.9 Пример 4

Пусть выборка Z_n порождена распределением $R(0; \theta)$. Пусть

$$G(Z_n; \theta) = \left(\frac{X^{(n)}}{\theta}\right)^n$$

$$F_G(x) = P\left(\left(\frac{X^{(n)}}{\theta}\right)^n \leq x\right) = P(X^{(n)} \leq \theta \sqrt[n]{x}) = F_{(n)}(\theta \sqrt[n]{x})^*$$

в силу [Следствия 6.1](#)

$$^* = (F_X(\theta \sqrt[n]{x}))^n = \begin{cases} 1^n, & \theta \sqrt[n]{x} \geq \theta, \\ 0^n, & 0 \leq 0, \\ -\left(\frac{\theta \sqrt[n]{x} - 0}{\theta - 0}\right)^n, & 0 \leq (0; \theta) \end{cases} = \begin{cases} 1, & x \geq 1, \\ 0, & x \leq 0, \\ x, & 0 < x < 1 \end{cases} \implies G(Z_n; \theta) \sim R(0; 1)$$

Тогда $G_{\frac{\alpha}{2}} = \frac{\alpha}{2}, G_{1-\frac{\alpha}{2}} = 1 - \frac{\alpha}{2},$

$$1 - \alpha = P\left(\frac{\alpha}{2} \leq \left(\frac{X^{(n)}}{\theta}\right)^n \leq 1 - \frac{\alpha}{2}\right) = P\left(X^{(n)}\left(1 - \frac{\alpha}{2}\right)^{-\frac{1}{n}} \leq \theta \leq X^{(n)}\left(\frac{\alpha}{2}\right)^{-\frac{1}{n}}\right)$$

Интервальная оценка имеет вид:

$$\left[X^{(n)}\left(1 - \frac{\alpha}{2}\right)^{-\frac{1}{n}}; X^{(n)}\left(\frac{\alpha}{2}\right)^{-\frac{1}{n}}\right]$$

Заметим, что точечная МП-оценка $\hat{\theta}(Z_n) = X^{(n)}$ **не лежит внутри** интервала.

11 Проверка статистических гипотез

11.1 Определение 1

Статистической гипотезой H называется любое предположение относительно закона распределения с.в. X , порождающей выборку Z_n .

11.2 Определение 2

Проверяемая гипотеза H_0 называется *основной*. Конкурирующая с H_0 гипотеза H_1 называется *альтернативой*.

11.3 Определение 3

Статистическим критерием или *критерием согласия* называется правило, в соответствии с которым по реализации выборки $z_n = Z_n(\omega)$ принимается или отвергается гипотеза H_0 .

11.4 Примеры

Гипотеза о виде распределения

Пусть F_X — распределение, порождающее выборку Z_n , F — некоторая заданная функция распределения, \mathbb{F} — некоторый заданный класс функций распределения.

Например, $F(x) = \frac{1}{2} + \Phi(x-1)$, $\mathbb{F} = \{F(x) = \frac{1}{2} + \Phi(\frac{x-m}{\sigma}) : m \in \mathbb{R}, \sigma > 0\}$

Тогда можно рассмотреть

$H_0 : F_X(x) = F(x), x \in \mathbb{R}$ или $H_0 : F_X \in \mathbb{F}$

Гипотеза однородности

Пусть задано k выборок $Z_{n_1}^1, \dots, Z_{n_k}^k$.

Требуется определить, образуют ли они единую однородную выборку.

Если $F_i(x)$ — функция распределения, порождающая выборку $Z_{n_i}^i$, то

$$H_0 : F_1(x) = \dots = F_k(x)$$

Гипотеза независимости

Пусть Z_n^1, Z_n^2 — *однородные* выборки, порождаемые с.в. X и Y соответственно с функциями распределения F_X и F_Y .

Через $F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y); x, y \in \mathbb{R}$

Замечание

Для проверки гипотезы H_0 используются специальные статистики $T(Z_n)$, характеризующие отклонение наблюдения от теоретического предположения.

11.5 Определение 4

Обозначим через G *множество всех возможных значений статистики* T :

$$G = \{t : t = T(z_n), z_n \in S\}$$

11.6 Определение 5

Построим разбиение G на *доверительную область* $G_{\circ\alpha}$ и *критическую область* $G_{|\alpha}$ так, чтобы

$$G_{\circ\alpha} \cup G_{|\alpha} = G$$

$$G_{\circ\alpha} \cap G_{|\alpha} = \emptyset$$

$P(T(z_n) \in G_{|\alpha} | H_0) \leq \alpha, P(T(Z_n) \in G_{\circ\alpha} | H_0) \geq 1 - \alpha$ для некоторого $\alpha \in (0; 1)$.

Если H_0 принимается в случае, когда $T(Z_n) \in G_{|\alpha}$, то α называется *уровнем значимости критерия*.

11.7 Определение 6

Ошибкой первого рода называется событие, когда H_0 верна, но она отвергается $T(z_n) \in G_{|\alpha}, H_0$ верно.

11.8 Определение 7

Ошибкой второго рода называется событие, когда H_0 неверно, но она принимается $T(z_n) \in G_{\circ\alpha}, H_0$ неверно.

Замечание

1. Если H_0 принимается, то это не значит, что H_0 верно. Это значит, что наблюдения $z_n = Z_n(\omega)$ *согласуется* с гипотезой H_0 .
2. Для построения доверительной и критической областей необходимо знать

$$Law(T(Z_n) | H_0)$$

а также задать уровень значимости α

3. Структура критической области $G_{|\alpha}$ определяется альтернативой H_1

11.9 Определение 8

Пусть \mathbb{F} — множество всех распределений $T(Z_n)$, удовлетворяющих альтернативе H_1 . Тогда *функцией мощности критерия* $W : \mathbb{F} \rightarrow [0; 1]$ называется

$$W(F) = P(T(Z_n) \in G_{|\alpha} | Law(T(Z_n)) = F)$$

Значение $W(F)$ называется *мощностью критерия при альтернативе F* .

11.10 Определение 9

Критерий называется *несмещенным*, если $W(F) > \alpha, \forall F \in \mathbb{F}$.

Замечание

1. Для построения $W(F)$ необходимо знать $Law(T(Z_n)|F)$, что не всегда доступно.
2. Качество критерия определяется функцией мощности: чем выше мощность, тем чаще критерий отвергает H_0 , если она *неверна*.

Критерий согласия Колмогорова

$$H_0 : F_X(x) = F(x), x \in \mathbb{R}$$

$$H_1 : F_X(x_0) \neq F(x_0) \text{ для некоторого } x_0 \in \mathbb{R}$$

$$T(Z_n) = \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)|$$

Данный критерий применяется для *непрерывных* $F(x)$.

$Law(T(Z_n)|H_0)$ известно. Для *малых* $n \in \mathbb{N}$ соответствующие квантили берутся из таблиц. При больших $n \in \mathbb{N}$ используют *асимптотическую аппроксимацию*:

$$F_{\sqrt{n}T(Z_n)|H_0}(x) \xrightarrow{n \rightarrow \infty} K(x) = \begin{cases} \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2 x^2}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$K(x)$ — *функция распределения Колмогорова*

Критерий согласия хи-квадрат Пирсона

Пусть для выборки Z_n построено разбиение \mathbb{R} :

$-\infty = t_0 < t_1 < t_2 < \dots < t_l < t_{l+1} = +\infty$, где

$t_1 \leq X^{(1)}, t_l > X^{(n)}$.

На построенном разбиении построим гистограмму:

$$\hat{f}_n(x) = \begin{cases} 0, & x \in (t_0, t_1) \cup [t_l, t_{l+1}), \\ \hat{p}_k, & x \in [t_k, t_{k+1}), k = \overline{1, l-1} \end{cases}$$

Пусть $p_k = P(X \in [t_k, t_{k+1})|H_0) = F(t_{k+1}) - F(t_k), k = \overline{0, l}$.

Статистика критерия Пирсона имеет вид:

$$T_{\chi^2}(Z_n) = n \cdot \sum_{k=0}^l \frac{(\hat{p}_k - p_k)^2}{p_k}$$

где $\hat{p}_0 = \hat{p}_l = 0$.

11.11 Теорема 1

Пусть $p_k \in (0; 1), k = \overline{0, l}$.

Тогда

$$Law(T_{\chi^2}(Z_n)|H_0) \xrightarrow{n \rightarrow \infty} \chi^2(l)$$

Замечание

1. **Высокая точность** приближения законом $\chi^2(l)$ достигается при $l \geq 5$ и $n \geq 50$.
2. Если у $F(x; \theta_1, \dots, \theta_5)$ теоретической функции распределения есть s неизвестных параметров, то

$$Law(T_{\chi^2}(Z_n)|H_0) \xrightarrow{n \rightarrow \infty} \chi^2(l-s)$$

Проверка гипотезы о значении параметра

Рассматривается выборка Z_n , порожденная распределением $N(m_X; \sigma_X^2)$

Рассмотрим $H_0 : m_X = m_0$ против

$H_1 : m_X \neq m_0, H_2 : m_X > m_0, H_3 : m_X < m_0$

1 случай: G_X^2 известна

$$T(Z_n) = \frac{(\bar{X}_n - m_0)}{\sigma_X} \sqrt{n}, Law(T(Z_n)|H_0) = N(0; 1)$$

$$G_{1\alpha} = (-\infty; u_{\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$$

$$G_{2\alpha} = (u_{1-\alpha}; +\infty), G_{3\alpha} = (-\infty; U_\alpha)$$

2 случай: G_X^2 неизвестна

$$T(Z_n) = \frac{(\bar{X}_n - m_0)}{\sqrt{\hat{d}_X(n)}} \sqrt{n-1}, Law(T(Z_n)|H_0) = t(n-1)$$

$$G_{1\alpha} = (-\infty; t_{\frac{\alpha}{2}}(n-1)) \cup (t_{1-\frac{\alpha}{2}}(n-1); +\infty)$$

$$G_{2\alpha} = (t_{1-\alpha}(n-1); +\infty), G_{3\alpha} = (-\infty; t_\alpha(n-1))$$

Рассмотрим $H_0 : \sigma_X^2 = \sigma_0^2$ против

$H_1 : \sigma_X^2 \neq \sigma_0^2$

1 случай: m_X известна

$$T(Z_n) = \frac{\sum_{k=1}^n (X_k - m_X)}{\sigma_0^2}, Law(T(Z_n)|H_0) = \chi^2(n)$$

2 случай: m_X неизвестна

$$T(Z_n) = \frac{n\hat{d}_X(n)}{\sigma_0^2}, Law(T(Z_n)|H_0) = \chi^2(n-1)$$

12 Метод наименьших квадратов

В рамках регрессионного анализа рассматривается задача восстановления зависимости $Y = \varphi(x)$ по набору наблюдений Y_1, \dots, Y_n , которые предполагаются **защищенными**.

12.1 Определение 1

Линейной регрессионной моделью называется класс линейных по набору неизвестных параметров $\theta \in \mathbb{R}^s$ функций:

$$\varphi(x; \theta) = \theta_1 \varphi_1(x) + \dots + \theta_s \varphi_s(x)$$

12.2 Определение 2

Схемой Гаусса-Маркова называется модель наблюдения линейной регрессионной модели при наличии случайных ошибок наблюдения:

$$Y_k = \theta_1 \varphi_1(x_k) + \dots + \theta_s \varphi_s(x_k) + \varepsilon_k, k = \overline{1, n}$$

ИЛИ

$$Y = X\theta + E, \text{ где } X = \begin{pmatrix} \varphi_1(x_1) & \dots & \varphi_s(x_1) \\ \vdots & \ddots & \vdots \\ \varphi_1(x_n) & \dots & \varphi_s(x_n) \end{pmatrix} \in \mathbb{R}^{n \times s}$$

$$\theta = (\theta_1, \dots, \theta_s)^T \in \mathbb{R}^s, Y = (Y_1, \dots, Y_n)^T \in \mathbb{R}^n$$

$$E = (\varepsilon_1, \dots, \varepsilon_n)^T \in \mathbb{R}^n$$

где X называется **регрессионной матрицей**. Предполагается, что

$$M[E] = 0, K_E = \sigma^2 \cdot I$$

12.3 Определение 3

МНК-оценкой вектора θ называется

$$\hat{\theta}(Y) = \underset{\theta}{\operatorname{argmin}} (Y - X\theta)^T (Y - X\theta) = \underset{\theta}{\operatorname{argmin}} \left(\sum_{k=1}^n (Y_k - \theta_1 \varphi_1(x_k) - \dots - \theta_s \varphi_s(x_k))^2 \right)$$

12.4 Теорема 1 (Гаусса-Маркова)

Пусть матрица $X \in \mathbb{R}^{n \times s}$ такая, что $\det(X^T X) \neq 0$

Тогда

1. МНК-оценка $\hat{\theta}(Y)$ существует, **единственна** и определяется соотношением:

$$\hat{\theta}(Y) = (X^T X)^{-1} X^T Y$$

2. МНК-оценка $\hat{\theta}(Y)$ **несмещенная** обладает **наименьшей дисперсией** в классе линейных по Y и **несмещенных** оценок по координатам.

3. Ковариационная матрица $\hat{\theta}$ имеет вид

$$K_{\hat{\theta}} = \sigma^2 (X^T X)^{-1}$$

Доказательство

1.

$$J(\theta) = (Y - X\theta)^T (Y - X\theta) = Y^T Y - 2Y^T X\theta + \theta^T X^T X\theta \rightarrow \min_{\theta}$$

$J(\theta)$ — квадратичная функция. При этом $H_J = 2X^T X$ — матрица Гессе невырождена и положительно определена:

Пусть $x \in \mathbb{R}^s \setminus \{0\}$. Тогда $x^T (2X^T X)x = 2(Xx, Xx) > 0$.

Т.о. $J(\theta)$ имеем единственный экстремум — точку минимума, которая может быть найдена из необходимых условий:

$$\nabla_{\theta} J(\theta) = -(2Y^T X)^T + 2X^T \cdot X\theta = 0$$

$$\hat{\theta} = (X^T X)^{-1} X^T Y$$

$$2. M[\hat{\theta}] = M[(X^T X)^{-1} X^T Y] = (X^T X)^{-1} X^T M[X\theta + E] = (X^T X)^{-1} X^T X\theta + (X^T X)^{-1} X^T M[E] = \theta$$

Рассмотрим произвольную несмещенную линейную оценку $\tilde{\theta} = AY$, где $A \in \mathbb{R}^{s \times n}$
Тогда $\forall \theta \in \mathbb{R}^s$

$$\theta = M[\tilde{\theta}] = M[AX\theta + AE] = AX\theta + AM[E] = AX\theta$$

Откуда $AX = I$

$$K_{\hat{\theta}} = M[(\tilde{\theta} - \theta)(\tilde{\theta} - \theta)^T] = M[(AY - \theta)(AY - \theta)^T] = [(AX\theta - AX - \theta)(AX\theta - AE - \theta)^T] \stackrel{AX=I}{=} M[AEE^T A^T] = AK_E A^T = \sigma^2 AA^T \quad (*)$$

$$\begin{aligned} AA^T &= (A - ((X^T X)^{-1} X^T) + ((X^T X)^{-1} X^T))(A - (X^T X)^{-1} X^T + (X^T X)^{-1} X^T)^T = \\ &= (A - (X^T X)^{-1} X^T)(A - (X^T X)^{-1} X^T)^T + (A - (X^T X)^{-1} X^T)X(X^T X)^{-1} + (X^T X)^{-1} X^T(A - \\ &= (X^T X)^{-1} X^T)^T + (X^T X)^{-1} X^T X(X^T X)^{-1} = (A - (X^T X)^{-1} X^T)(A - (X^T X)^{-1} X^T)^T + \\ &+ AX(X^T X)^{-1} - (X^T X)^{-1} X^T X(X^T X)^{-1} + (X^T X)^{-1} X^T A^T - (X^T X)^{-1} X^T X(X^T X)^{-1} + \\ &+ (X^T X)^{-1} \stackrel{AX=I}{=} (A - (X^T X)^{-1} X^T)(A - (X^T X)^{-1} X^T)^T + (X^T X)^{-1} \end{aligned}$$

т.к. $(A - (X^T X)^{-1} X^T)(A - (X^T X)^{-1} X^T)^T \geq 0$, то все ее диагональные элементы, которые и определяют $D[\theta_k], k = \overline{1, s}$, неотрицательны. Тогда минимальное значение $D[\theta_k], \forall k = \overline{1, s}$ достигается, если

$$(A - (X^T X)^{-1} X^T)(A - (X^T X)^{-1} X^T)^T = 0, \text{ т.е.}$$

$$A = (X^T X)^{-1} X^T$$

ИЛИ

$$\hat{\theta} = \hat{\theta}$$

3. Если положить $A = (X^T X)^{-1} X^T$ то из (*) следует, что

$$K_{\hat{\theta}} = \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}. \blacksquare$$

Замечание

[Теорема 1](#) гарантирует свойства МНК-оценки для любых коррелированных шумов.

Нормальная регрессия**12.5 Определение 4**

Схема Гаусса-Маркова называется *нормальной регрессией*, если $E \sim N(0; \sigma^2 I)$.

12.6 Лемма 1

В нормальной регрессии

$$\hat{\theta} \sim N(\theta; \sigma^2 (X^T X)^{-1})$$

Доказательство

Доказательство следует из [Леммы 1.3](#) и [Теоремы 1 п.2, п.3](#). ■

12.7 Лемма 2

В нормальной регрессии МНК-оценка и МП-оценка *совпадают*

Доказательство

В силу [Леммы 1.3](#) $Y \sim N(X\theta; \sigma^2 I)$

Тогда функция правдоподобия по определению имеет вид

$$L(y; \theta) = f_Y(y; \theta) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\sigma^2 I)}} \exp\left\{-\frac{1}{2\sigma^2} (y - X\theta)^T (y - X\theta)\right\}$$

откуда $\operatorname{argmax}_{\theta} L(Y; \theta) = \operatorname{argmin}_{\theta} (Y - X\theta)^T (Y - X\theta) = \hat{\theta}$. ■

12.8 Следствие 1

В нормальной регрессии МНК-оценка $\hat{\theta}$ *эффективна*, т.е. *оптимальна* в классе всех *несмещенных* оценок.

Доказательство

Доказательство следует из [Леммы 2](#) и [Теоремы 8.1](#). ■

12.9 Лемма 3

В нормальной регрессии МП-оценка σ^2 имеет вид

$$\hat{\sigma}^2 = \frac{1}{n} \|Y - X\hat{\theta}\|^2$$

Доказательство

$$\ln L(Y; \theta; \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{(Y-X\theta)^T(Y-X\theta)}{2\sigma^2},$$

$$\begin{cases} \frac{\partial \ln L(Y; \theta; \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{(Y-X\theta)^T(Y-X\theta)}{2(\sigma^2)^2} = 0, \\ \nabla_{\theta} \ln L(Y; \theta; \sigma^2) = -\frac{1}{2\sigma^2}(-2X^T Y + 2X^T X\theta) = 0 \end{cases}$$

Откуда $\hat{\theta} = (X^T X)^{-1} X^T Y$ — МНК-оценка,

$$\hat{\sigma}^2 = \frac{1}{n} (Y - X\hat{\theta})^T (Y - X\hat{\theta})$$

■

12.10 Определение 5

Вектором остатков называется

$$\hat{E} = Y - X\hat{\theta}$$

12.11 Лемма 4

В нормальной регрессии $\hat{E} \sim N(0; \sigma^2(I - X(X^T X)^{-1} X^T))$

Доказательство

$$\hat{E} = Y - X\hat{\theta} = X\theta + E - X(X^T X)^{-1} X^T (X\theta + E) = X\theta + E - X(X^T X)^{-1} X^T X\theta - X(X^T X)^{-1} X^T E = (I - X(X^T X)^{-1} X^T) E.$$

Тогда в силу [Леммы 1.3](#) $m_{\hat{E}} = 0$

$$K_{\hat{E}} = (I - X(X^T X)^{-1} X^T) \sigma^2 I (I - X(X^T X)^{-1} X^T)^T = \sigma^2 (I - 2X(X^T X)^{-1} X^T + X(X^T X)^{-1} X^T X(X^T X)^{-1} X^T) = \sigma^2 (I - (X(X^T X)^{-1} X^T)). \blacksquare$$

12.12 Лемма 5

В нормальной регрессии $\hat{\theta}$ и \hat{E} **независимы**

Доказательство

В силу [Леммы 1.4](#) достаточно показать, что $\hat{\theta}$ и \hat{E} *некоррелированы*.

$$K_{\hat{\theta}\hat{E}} = \text{cov}(\hat{\theta}, Y - X\hat{\theta}) = \text{cov}(\hat{\theta}, Y) - \text{cov}(\hat{\theta}, \hat{\theta}) \cdot X^T \stackrel{*}{=}$$

с учетом [Леммы 1](#) и [Теоремы 1](#)

$$\stackrel{*}{=} \text{cov}((X^T X)^{-1} X^T Y, Y) - \sigma^2 (X^T X)^{-1} X^T = (X^T X)^{-1} X^T \text{cov}(X\theta + E, X\theta + E) - \sigma^2 (X^T X)^{-1} X^T = (X^T X)^{-1} X^T \sigma^2 I - \sigma^2 (X^T X)^{-1} X^T = 0. \blacksquare$$

12.13 Лемма 6

В нормальной регрессии

$$\frac{\|\hat{E}\|^2}{\sigma^2} = n \cdot \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-s)$$

Доказательство

Пусть $E_0 = \frac{E}{\sigma}$ Тогда $E_0 \sim N(0; I)$

$$\frac{\|\hat{E}\|^2}{\sigma^2} = \frac{1}{\sigma^2} \|Y - X(X^T X)^{-1} X^T Y\|^2 = \frac{1}{\sigma^2} \|X\theta + E - X(X^T X)^{-1} X^T X\theta - X(X^T X)^{-1} X^T E\|^2 =$$

$$\frac{1}{\sigma^2} \|(I - X(X^T X)^{-1} X^T)E\|^2 = E_0^T A E_0, \text{ где}$$

$$A = (I - X(X^T X)^{-1} X^T)^T (I - X(X^T X)^{-1} X^T) = I - 2X(X^T X)^{-1} X^T + X(X^T X)^{-1} X^T X(X^T X)^{-1} X^T =$$

$$I - X(X^T X)^{-1} X^T. \text{ Откуда следует, что } A^2 = A.$$

Т.к. $A^T = A$ и $A \geq 0$, то все собственные значения $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ неотрицательны, а переход в нормальный жорданов базис описывается ортогональным преобразованием C :

$$C^T = C^{-1}, A = C^T \text{diag}(\lambda_1, \dots, \lambda_n) C$$

Тогда

$$A^2 = C^T \text{diag}(\lambda_1^2, \dots, \lambda_n^2) C$$

Из $A^2 = A$ следует, что $\lambda_i = \lambda_i^2, i = \overline{1, n}$, т.е. $\lambda_i \in \{0; 1\}$

Тогда

$$\Lambda = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix} = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & 1 & \vdots \\ 0 & \dots & 0 \end{pmatrix}$$

где $m = \text{rg} A$

$$\frac{\|\hat{E}\|^2}{\sigma^2} = E_0^T C^T \cdot \Lambda C E_0 = \sum_{i=1}^n \tilde{\varepsilon}_i^2, \text{ где с учетом Леммы 1.3}$$

$$C E_0 = (\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n) \sim N(0, C I C^T) = N(0; I)$$

Т.о. по определению $\frac{\|\hat{E}\|^2}{\sigma^2} \sim \chi^2(m)$

Т.к. $\text{tr}(AB) = \text{tr}(BA), \forall A, B^T \in \mathbb{R}^{n \times m}$, то

$$\text{tr}(A) = \text{tr}(C^T \Lambda C) = \text{tr}(C C^T \Lambda) = \text{tr}(\Lambda) = m = \text{rg} A$$

$$m = \text{tr}(I - X(X^T X)^{-1} X^T) = n - \text{tr}(X(X^T X)^{-1} X^T) = n - \text{tr}(X^T X (X^T X)^{-1}) = n - \text{tr} \begin{pmatrix} I \\ s \times s \end{pmatrix} =$$

$n - s$. ■

12.14 Лемма 7

В нормальной регрессии

$$\frac{\hat{\theta}_k - \theta_k}{\|\hat{E}\| \sqrt{\alpha_k}} \cdot \sqrt{n - s} \sim t(n - s)$$

где α_k — элемент на главной диагонали матрицы $(X^T X)^{-1}$

Доказательство

Из Леммы 1 следует, что $\hat{\theta}_k \sim N(\theta_k, \sigma^2 \alpha_k)$

Тогда

$$\frac{\hat{\theta}_k - \theta_k}{\sigma \sqrt{\alpha_k}} \sim N(0; 1)$$

С учетом определения распределения Стюдента и Леммы 5, и Леммы 6

$$\frac{\hat{\theta}_k - \theta_k}{\|\hat{E}\| \sqrt{\alpha_k}} \cdot \sqrt{n - s} = \frac{\frac{\hat{\theta}_k - \theta_k}{\sigma \sqrt{\alpha_k}}}{\sqrt{\frac{1}{n-s} \frac{\|\hat{E}\|^2}{\sigma^2}}} \sim t(n - s). \quad \blacksquare$$

12.15 Лемма 8

В нормальной регрессии

$$\frac{\varphi(x; \hat{\theta}_k) - \varphi(x; \theta_k)}{\|E\| \sqrt{\alpha_k}} \cdot \sqrt{n-s} \sim t(n-s), \text{ где}$$

$$\alpha(x) = (\varphi_1(x), \dots, \varphi_s(x))(X^T X)^{-1} \begin{pmatrix} \varphi_1(x) \\ \vdots \\ \varphi_s(x) \end{pmatrix}, x \in \mathbb{R}$$

Доказательство

Из [Леммы 1](#) следует, что

$$\varphi(x; \hat{\theta}) = \hat{\theta} \varphi_1(x) + \dots + \hat{\theta}_s \varphi_s(x) \sim N(\varphi(x; \theta); \alpha(x) \sigma^2)$$

Тогда

$$\frac{\varphi(x; \hat{\theta}_k) - \varphi(x; \theta_k)}{\sigma \sqrt{\alpha_k}} \sim N(0; 1).$$

С учетом лемм 5 и 6, определения Стьюдента

$$\frac{\varphi(x; \hat{\theta}_k) - \varphi(x; \theta_k)}{\|E\| \sqrt{\alpha_k}} = \frac{\frac{\varphi(x; \hat{\theta}_k) - \varphi(x; \theta_k)}{\sigma \sqrt{\alpha_k}}}{\sqrt{\frac{1}{n-s} \frac{\|E\|^2}{\sigma^2}}} \sim t(n-s). \blacksquare$$

Замечание

Леммы 6, 7, 8 позволяют построить доверительные интервалы для $\theta_1, \dots, \theta_s, \sigma^2, \varphi(x; \theta)$.

Для построения точечной оценки $\theta_1, \dots, \theta_s$ и $\varphi(x; \theta)$ используется [Теорема 1](#), для точечной оценки — [Лемма 3](#).

Зачастую в качестве $\varphi_k(x)$ рассматривают $\varphi_k(x) = x^{k-1}$. Однако порядок многочлена s , как правило, несмещен. Для его определения можно использовать **критерий Фишера**:

Критерий Фишера

$$H_0 : \theta_k = 0$$

$$H_1 : \theta_k \neq 0$$

$$T(Y) = \frac{\hat{\theta}_k^2}{\alpha_k \|\hat{E}\|^2} (n-s)$$

где α_k — k -й элемент на главной диагонали матрицы $(X^T X)^{-1}$.

$$Law(T(Y)|H_0) = F(1, n-s)$$

$$G_{\circ\alpha} = [0; F_{1-\alpha}(1; n-s))$$

$$G_{|\alpha} = [F_{1-\alpha}(1; n-s); +\infty)$$