1 Неравенства Чебышева.

Первое неравенство Чебышева. *Пусть*

1. X *– случайная величина*
2. $X \geq 0$*(т.е.* $P\{X < 0\} = 0$*)*
3. $\exists MX$

Тогда $\forall \varepsilon > 0 \quad P\{X \geq \varepsilon\} \leq \frac{MX}{\varepsilon}$

Доказательство. Для случая непрерывной случайной величины X(для случая дискретной случайной величины X доказательство аналогично)
$MX = \int_{-\infty}^{+\infty} xf(x)dx = |X \geq 0| = \int_0^{+\infty} xf(x)dx = \int_0^{\varepsilon} xf(x)dx + \int_{\varepsilon}^{+\infty} xf(x)dx \geq \int_{\varepsilon}^{+\infty} xf(x)dx \geq |x \in [\varepsilon, +\infty) \to x \geq \varepsilon| \geq \varepsilon \int_{\varepsilon}^{+\infty} f(x)dx \geq \varepsilon P\{X \geq \varepsilon\}$

$$\int_{\varepsilon}^{+\infty} f(x)dx = P\{X \geq \varepsilon\}$$

Таким образом,

$$MX \geq \varepsilon P\{X \geq \varepsilon\} \Longrightarrow P\{X \geq \varepsilon\} \leq \frac{MX}{\varepsilon}$$

□

Второе неравенство Чебышева. *Пусть*

1. X *– случайная величина*
2. $\exists MX, \quad \exists DX$

Тогда $\quad \forall \varepsilon > 0 \quad P\{|X - MX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2}$

Доказательство.

1. Рассмотрим случайную величину $Y = (X - MX)^2$
2. Из первого неравенства Чебышева для Y следует, что $\forall \delta > 0$ $P\{Y \geq \delta\} \leq \frac{MY}{\delta}$
3. Используем $P\{Y \geq \delta\} \leq \frac{MY}{\delta}$ для $\delta = \varepsilon^2$

$$DX = M[(X-MX)^2] \geq \delta P\{(X-MX)^2 \geq \delta\} = \varepsilon^2 P\{(X-MX)^2 \geq \varepsilon^2\} = \varepsilon^2 P\{|X-MX| \geq \varepsilon\}$$

Таким образом, $DX \geq \varepsilon^2 P\{|X - MX| \geq \varepsilon\} \Longrightarrow P\{|X - MX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2}$ □

2 Сходимость. Закон больших чисел.

Сходимость по вероятности и слабая сходимость для последовательности случайных величин. Закон больших чисел. *Пусть* $X_1, \ldots, X_n, \ldots$ *– последовательность случайных величин.*

Определение 2.1. Говорят, что последовательность случайных величин $X_1, \ldots, X_n, \ldots$ сходится по вероятности к случайной величине Z, если

$$\forall \varepsilon > 0 \quad P\{|X_n - Z| \geq \varepsilon\} \xrightarrow[n\to\infty]{} 0$$

$$X_n \xrightarrow[n\to\infty]{P} Z$$

Определение 2.2. Говорят, что последовательность случайных величин $X_1, \dots, X_n, \dots$ слабо сходится к случайной величине Z, если функциональная последовательность $F_{X_1}(x)$, $F_{X_2}(x), \dots$ поточечно сходится к функции $F_Z(x)$ во всех точках непрерывности последней, т.е.

$$(\forall x_0 \in \mathbb{R})(F_Z(x) \text{ непрерывна в } x_0) \Longrightarrow F_{X_n}(x_0) \xrightarrow[n\to\infty]{} F_z(x_0)$$

Закон больших чисел.

Определение 2.3. Говорят, что последовательность $X_1, \dots, X_n, \dots$ удовлетворяет закону больших чисел, если

$$\forall \varepsilon > 0 \quad P\{|\tfrac{1}{n}\textstyle\sum_{i=1}^n X_i - \tfrac{1}{n}\sum_{i=1}^n m_i| \geq \varepsilon\} \xrightarrow[n\to\infty]{} 0$$

где $m_i = MX_i, \quad i \in N$

Закон больших чисел в форме Чебышева. *Пусть*

1. $X_1, \dots, X_n, \dots$ *– последовательность независимых случайных величин*
2. $\exists MX_i = m_i \quad \exists DX_i = \sigma_i^2, \quad i \in N$
3. *Дисперсия случайных величин* $X_1, \dots, X_n, \dots$ *ограничена в совокупности, то есть*

$$\exists c > 0 \quad \sigma_i^2 \leq c, \quad i \in N$$

Тогда последовательность $X_1, \dots, X_n, \dots$ *удовлетворяет закону больших чисел.*

Доказательство. 1. Рассмотрим

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i, \quad n \in N$$

Тогда

$$M[\overline{X_n}] = \frac{1}{n}\sum_{i=1}^n m_i$$

$$D[\overline{X_n}] = D[\frac{1}{n}\sum_{i=1}^n X_i] = \frac{1}{n^2}D[\sum_{i=1}^n X_i] = \frac{1}{n^2}\sum_{i=1}^n DX_i = \frac{1}{n^2}\sum_{i=1}^n \sigma_i^2$$

2. Применим к случайной величине $\overline{X_n}$ второе неравенство Чебышева

$$P\{|\overline{X_n} - M\overline{X_n}| \geq \varepsilon\} \leq \frac{D\overline{X_n}}{\varepsilon^2}$$

Таким образом,

$$P\{|\overline{X_n} - \frac{1}{n}\sum_{i=1}^n m_i| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^n \sigma_i^2$$

$$\sum_{i=1}^n \sigma_i^2 \leq \sum_{i=1}^n c = nc$$

$$0 \leq P\{|\overline{X_n} - \frac{1}{n}\sum_{i=1}^n m_i| \geq \varepsilon\} \leq \frac{c}{\varepsilon^2 n^2} \cdot n = \frac{c}{\varepsilon^2 n}$$

При $n \to \infty \quad \frac{c}{\varepsilon^2 n} \to 0$. По теореме о двух милиционерах

$$P\{|\overline{X_n} - \frac{1}{n}\sum_{i=1}^{n} m_i| \geq \varepsilon\} \xrightarrow[n\to\infty]{} 0$$

то есть последовательность $X_1, \dots, X_n, \dots$ удовлетворяет закону больших чисел. $\square$

Следствие 1. *Пусть*

1. *выполнены условия теоремы Чебышева*
2. *все случайные величины X_i одинаково распределены(обозначим $m_i \equiv m = MX_i$)*

Тогда

$$\forall \varepsilon > 0 \quad P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - m| \geq \varepsilon\} \xrightarrow[n\to\infty]{} 0$$

Доказательство. Так как $m_i \equiv m$, то $\frac{1}{n}\sum_{i=1}^{n} m_i = m$ и используем закон больших чисел в форме Чебышева. $\square$

Следствие 2. *Закон больших чисел в форме Бернулли.*
Пусть

1. *проводится n испытаний по схеме Бернулли с вероятностью успеха p*
2. $r_n = \frac{\textit{количество наступлений успеха}}{n}$ *– относительная(наблюденная) частота успеха*

Тогда

$$r_n \xrightarrow[n\to\infty]{P} p$$

Доказательство. 1. Введем случайные величины $X_i, \quad i = \overline{1, m}$,

$$X_i = \begin{cases} 1, & \text{если в i-м испытании произошёл успех} \\ 0, & \text{иначе} \end{cases}$$

Тогда

- Закон распределения X_i

X_i	0	1
p	q	p

Таким образом, все X_i одинаково распределены, $MX_i = p, \quad DX_i = pq$

- $DX_i \equiv pq \Longrightarrow$ ограничены в совокупности
- X_i независимы, так как отдельные испытания в схеме испытаний Бернулли независимы

2. Таким образом, последовательность $X_1, \dots, X_n, \dots$ удовлетворяет следствию 1 из теоремы Чебышева и для нее справедливо

$$\forall \varepsilon > 0 \quad P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - m| \geq \varepsilon\} \xrightarrow[n\to\infty]{} 0$$

$$\forall \varepsilon > 0 \quad P\{|r_n - p| \geq \varepsilon\} \xrightarrow[n\to\infty]{} 0\text{, то есть } r_n \xrightarrow[n\to\infty]{P} p$$

$\square$

3 Центральная предельная теорема

Пусть выполнены следующие 3 условия:

1. $X_1, \dots, X_n, \dots$ – последовательность независимых случайных величин
2. все случайные величины $X_i, \quad i \in N$ одинаково распределены
3. $\exists MX_i = m, \quad \exists DX_i = \sigma^2, \quad i \in N$

Рассмотрим случайную величину

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad M\overline{X_n} = m, \quad D\overline{X_n} = \frac{\sigma^2}{n}, \quad n \in N$$

Рассмотрим случайную величину

$$Y_n = \frac{\overline{X_n} - M\overline{X_n}}{\sqrt{D\overline{X_n}}} = \frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}}$$

Центральная предельная теорема. *Пусть выполнены условия 1-3. Тогда последовательность случайных величин Y_n при $n \to \infty$ слабо сходится к случайной величине Z, имеющей стандартное нормальное распределение, то есть*

$$\forall x \in \mathbb{R} \quad F_{Y_n} \xrightarrow[n\to\infty]{} F_Z(x),$$

где

$$Z \sim N(0,1), \quad F_Z(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt.$$

Интегральная теорема Муавра-Лапласа. *Пусть*

1. *проводится большое число испытаний по схеме Бернулли с вероятностью успеха p*
2. *k – число успехов в этой серии*

Тогда

$$P\{k_1 \le k \le k_2\} \approx \Phi(x_2) - \Phi(x_1), \quad x_i = \frac{k_i - np}{\sqrt{npq}}, \quad i = \overline{1,2}, \quad q = 1-p, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{\frac{-t^2}{2}}dt$$

Доказательство.

1. Пусть X_i – случайная величина, принимающая значения 0 или 1 в соответствии с правилом

$$X_i = \begin{cases} 1, & \text{если в i-м испытании произошёл успех} \\ 0, & \text{иначе} \end{cases}$$

Тогда

- Случайные величины $X_1, \dots, X_n, \dots$ независимы
- $MX_i = p$, $DX_i = pq$, $i \in N$
- X_i одинаково распределены

2. $P\{k_1 \le k \le k_2\} = P\{k_1 \le \sum_{i=1}^{n} X_i \le k_2\} = P\{\frac{k_1}{n} - p \le \frac{1}{n}\sum_{i=1}^{n} X_i - p \le \frac{k_2}{n} - p\} = P\{\frac{k_1/n-p}{\sqrt{\frac{pq}{n}}} \le \frac{\overline{X_n}-p}{\sqrt{\frac{pq}{n}}} \le \frac{k_2/n-p}{\sqrt{\frac{pq}{n}}}\} \approx \Phi(x_2) - \Phi(x_1)$

$\square$

4 Математическая статистика

Определение 4.1. Множество возможных значений случайной величины X называют генеральной совокупностью.

Определение 4.2. Случайной выборкой из генеральной совокупности X называют случайный вектор $\vec{X} = (X_1, \ldots, X_n)$, где $X_1, \ldots, X_n$ – независимые в совокупности случайные величины, каждая из которых имеет то же распределение, что и X. При этом n называется объёмом случайной выборки.

Определение 4.3. Любую возможную реализацию $\vec{x} = (x_1, \ldots, x_n)$ случайной выборки $\vec{X}$ называют выборкой из генеральной совокупности X. При этом число x_k называется k-м элементом выборки $\vec{x}$.

Определение 4.4. Вариационным рядом, построенным по выборке $\vec{x}$, называется кортеж $(x_{(1)}, \ldots, x_{(n)})$, где $x_{(1)}, \ldots, x_{(n)}$ – элементы выборки $\vec{x}$, расположенные в порядке неубывания.

Определение 4.5. Пусть $F(x)$ – функция распределения случайной величины X. Тогда функция распределения случайной выборки $\vec{X}$ объема n из совокупности X:

$$F_{\vec{X}}(t_1, \ldots, t_n) = F(t_1) \cdot \ldots \cdot F(t_n)$$

$$P\{X_1 < t_1, \ldots, X_n < t_n\} = P\{X_1 < t_1\} \cdot \ldots \cdot P\{X_n < t_n\} = F(t_1) \cdot \ldots \cdot F(t_n)$$

$F_{x_{(n)}}(x) = P\{x_{(n)} < x\} = P\{X_1 < x, \ldots, X_n < x\} = P\{X_1 < x\} \cdot \ldots \cdot P\{X_n < x\} = F(x) \cdot \ldots \cdot F(x) = [F(x)]^n$

$F_{x_{(1)}}(x) = P\{x_{(1)} < x\} = 1 - P\{X_1 \geq x\} = 1 - P\{X_1 \geq x, \ldots, X_n \geq x\} = 1 - (1 - P\{X_1 < x\}) \cdot \ldots \cdot (1 - P\{X_n < x\}) = 1 - (1 - F(x))^n$

Определение 4.6. Любую функцию $g(\vec{X})$ случайной выборки $\vec{X}$ называют статистикой.

Определение 4.7. Выборочным начальным моментом порядка k называют статистику:

$$\hat{m}_k(\vec{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

Определение 4.8. Центральным выборочным моментом порядка k называют статистику

$$\hat{\nu}_k(\vec{X}) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^k$$

Определение 4.9. Выборочным средним (выборочным математическим ожиданием) называют статистику

$$\hat{m}(\vec{X}) = \overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Определение 4.10. Выборочной дисперсией называют статистику

$$\hat{\sigma^2}(\vec{X}) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Замечание. *Выборочное среднее является несмещённой оценкой своего теоретического аналога, а выборочная дисперсия – нет.*

Доказательство. $\hat{m}(\vec{X}) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
$M[\hat{m}(\vec{X})] = M[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}M[\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} MX_i = |X_i \sim X| = \frac{1}{n}\sum_{i=1}^{n} m_i = m$ □

Определение 4.11. Эмпирической функцией распределения, отвечающей выборке $\vec{x}$ называют функцию

$$F_n\colon \mathbb{R} \to \mathbb{R}, \quad F_n(x) = \frac{n(x, \vec{x})}{n},$$

где

- $\vec{x} = (x_1, \ldots, x_n)$ — выборка из генеральной совокупности $\vec{X}_n$;
- $n(x, \vec{x})$ — количество элементов выборки $\vec{x}$, которые меньше x.

Определение 4.12. Выборочной функцией распределения, отвечающей случайной выборке $\vec{X}$, называется функция:

$$\hat{F}_n(x) = \frac{n(x, \vec{X})}{n},$$

где $n(x, \vec{X})$ – случайная величина, которая для каждой реализации $\vec{x}$ случайной выборки $\vec{X}$ принимает значение, равное $n(x, \vec{x})$.

Теорема о сходимости выборочной функции распределения.. *Для любого фиксированного $x \in \mathbb{R}$ $\hat{F}_n(x)$ сходится по вероятности к значению $F(x)$ теоретической функции распределения случайной величины X:*

$$\forall x \in \mathbb{R} \qquad \hat{F}_n(x) \xrightarrow[n\to\infty]{P} F(x)$$

Доказательство. $\hat{F}_n(x)$ – относительная частота успеха в серии из n испытаний по схеме Бернулли с вероятностью успеха p.
В соответствии с законом больших чисел в форме Бернулли

$$\hat{F}_n(x) \xrightarrow[n\to\infty]{P} p, \text{ но } p = P\{X < x\} = F(x)$$

□

Определение 4.13. Интервальным статистическим рядом называют таблицу:

J_1	$\ldots$	J_m
n_1	$\ldots$	n_m

Здесь n_i – количество элементов выборки $\vec{x}$, принадлежащих J_i.

Определение 4.14. Эмпирической плотностью распределения случайной выборки $\vec{X}_n$ называют функцию

$$f_n(x) = \begin{cases} \frac{n_i}{n\,\Delta}, & x \in J_i,\ i = \overline{1, m}; \\ 0, & \text{иначе.} \end{cases}, \quad \text{где}$$

- J_i, $i = \overline{1; m}$, — полуинтервал из $J = [x_{(1)}, x_{(n)}]$, где

$$x_{(1)} = \min\{x_1, \ldots, x_n\}, \qquad x_{(n)} = \max\{x_1, \ldots, x_n\}; \tag{1}$$

при этом все полуинтервалы, кроме последнего, не содержат правую границу т. е.

$$J_i = [x_{(1)} + (i-1)\Delta, x_{(1)} + i\Delta), \quad i = \overline{1, m-1}; \tag{2}$$
$$J_m = [x_{(1)} + (m-1)\Delta, x_{(1)} + m\Delta]; \tag{3}$$

- m — количество полуинтервалов интервала $J = [x_{(1)}, x_{(n)}]$;
- Δ — длина полуинтервала J_i, $i = \overline{1, m}$ равная

$$\Delta = \frac{x_{(n)} - x_{(1)}}{m} = \frac{|J|}{m};$$

- n_i — количество элементов выборки в полуинтервале J_i, $i = \overline{1, m}$;
- n — количество элементов в выборке.

Определение 4.15. График функции $f_n(x)$ называют гистограммой.

Определение 4.16. Полигоном частот для выборки $\vec{x}$ называется ломанная, звенья которой соединяют середины верхних сторон прямоугольников гистограммы.

5 Точечные оценки.

Пусть X – случайная величина, общий закон распределения которой известен, но неизвестны значения одного или нескольких параметров этого закона. Пусть θ – неизвестный параметр закона распределения случайной величины X.

Определение 5.1. Точечной оценкой параметра θ называется статистика $\hat{\theta}(\vec{X})$, выборочное значение которой принимается в качестве значения параметра θ: $\theta := \hat{\theta}(\vec{X})$.

Качество используемой точечной оценки $\hat{\theta}(\vec{X})$ параметра θ характеризуют следующие свойства:

1. несмещенность
2. состоятельность
3. эффективность

Определение 5.2. Точечная оценка $\hat{\theta}(\vec{X})$ параметра θ называется несмещенной, если

$$\exists M[\hat{\theta}(\vec{X})] = \theta$$

Доказать, что выборочная дисперсия является смещённой оценкой дисперсии.

Доказательство.

- X – случайная величина
- $\sigma^2 = DX$
- $\hat{\sigma^2}(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

$M[\hat{\sigma^2}(\vec{X})] = M[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2] = \frac{1}{n}M[\sum_{i=1}^{n}(X_i - \overline{X})^2] = \frac{1}{n}\sum_{i=1}^{n}M[(X_i - \overline{X})^2] = \frac{1}{n}\sum_{i=1}^{n}M[(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j)^2] = \frac{1}{n}\sum_{i=1}^{n}M[((X_i - m) - \frac{1}{n}\sum_{j=1}^{n}(X_j - m))^2] = \frac{1}{n}\sum_{i=1}^{n}M[(X_i - m)^2 - \frac{2}{n}\sum_{j=1}^{n}(X_j - m)(X_i - m) + \frac{1}{n^2}(\sum_{j=1}^{n}(X_j - m))^2] = \frac{1}{n}\sum_{i=1}^{n}\{M[(X_i - m)^2] - \frac{2}{n}\sum_{j=1}^{n}M[(X_i - m)(X_j - m)] + \frac{1}{n^2}\sum_{j=1}^{n}M[(X_j - m)^2 + \frac{1}{n^2}\sum_{j,k=1,k\neq j}^{n}M[(X_k - m)(X_j - m)]] = \frac{1}{n}\sum_{i=1}^{n}\{\sigma^2 - \frac{2}{n}\sigma^2 + \frac{1}{n^2}\sum_{j=1}^{n}\sigma^2\} = \frac{1}{n} \cdot n\{\sigma^2 - \frac{2}{n}\sigma^2 + \frac{1}{n}\sigma^2\} = \sigma^2 - \frac{1}{n}\sigma^2 = \sigma^2(1 - \frac{1}{n}) = \sigma^2\frac{n-1}{n} \neq \sigma^2$ □

Определение 5.3. Статистику $S^2(\vec{X})$ называется исправленной выборочной дисперсией и равна
$S^2(\vec{X}) = \frac{n}{n-1}\hat{\sigma^2}(\vec{X}) = M[S^2] = M[\frac{n}{n-1}\sigma^2] = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n}{n-1}M[\hat{\sigma^2}] = \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot \sigma^2 = \sigma^2$,
то есть $S^2(\vec{X})$ является несмещённой оценкой дисперсии.

Определение 5.4. Оценка $\hat{\theta}$ называется состоятельной оценкой, если

$$\hat{\theta}(\vec{X}) \xrightarrow[n\to\infty]{P} \theta,$$

где n – объем выборки.

Замечание. *Условие из определения можно записать в виде:*

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P\{|\hat{\theta}(\vec{X}) - \theta| < \varepsilon\} = 1$$

Пример 1. *Пусть X – случайная величина, $\exists MX = m$.*

- *последовательность $X_1, \dots, X_n, \dots$ независима и одинаково распределена*
- $\exists MX_i = m, \quad \exists DX_i = \sigma^2$
- *из предыдущих пунктов следует, что $X_1, \dots, X_n, \dots$ удовлетворяет закону больших чисел в форме Чебышева*

$$\forall \varepsilon > 0 \quad P\{|\overline{X} - m| < \varepsilon\} \xrightarrow[n\to\infty]{} 1$$
$$\overline{X} \xrightarrow[n\to\infty]{P} m$$

Пример 2. *Пусть*

1. $X \sim N(m, \sigma^2)$, *m и σ^2 неизвестны*
2. $\hat{m}(\vec{X}) = X_1$ *– результат первого наблюдения – точечная оценка для m*

Покажем, что $\hat{m}$ – несостоятельная оценка.
Зафиксируем $\varepsilon > 0$
$P\{|\hat{m}(\vec{X}) - m| < \varepsilon\} = P\{|X_1 - m| < \varepsilon\} =/X_1 \sim X \sim N(m, \sigma^2)/= P\{m - \varepsilon < X_1 < m + \varepsilon\} = \Phi_0(\frac{m+\varepsilon-m}{\sigma}) - \Phi_0(\frac{m-\varepsilon-m}{\sigma}) = 2\Phi_0(\frac{\varepsilon}{\sigma}) \neq 1$, *если* $\frac{\varepsilon}{\sigma} \neq +\infty$.
Тогда $P\{|\hat{m}(\vec{X}) - m| < \varepsilon\} \xrightarrow[n\to\infty]{} 1$.

Определение 5.5. Оценка $\hat{\theta}$ называется эффективной оценкой для параметра θ, если:

1. $\hat{\theta}$ – несмещенная оценка для θ
2. $\hat{\theta}$ обладает наименьшей дисперсией среди всех несмещенных оценок θ

Замечание. *Иногда говорят об эффективной оценке в классе оценок Θ.*
Если Θ – некоторое множество несмещенных оценок для θ, то оценка $\hat{\theta} \in \Theta$ называется эффективной оценкой для θ в классе Θ, если $\hat{\theta}$ обладает наименьшей дисперсией среди всех оценок класса Θ.

$$\forall \tilde{\theta} \in \Theta \quad D\hat{\theta} \le D\tilde{\theta}$$

Доказать, что выборочное среднее является эффективной оценкой для m в классе линейных оценок. *Пусть X – случайная величина, $\exists MX = m, \quad \exists DX = \sigma^2$, так что выборочное среднее $\overline{X}$ является эффективной оценкой для m в классе линейных оценок.*

Доказательство. 1. Линейная оценка имеет вид:

$$\hat{m}(\vec{X}) = \lambda_1 X_1 + \dots + \lambda_n X_n, \quad \lambda_j \in \mathbb{R}$$

2. Так как оценка должна быть несмещенной

$$M[\hat{m}(\vec{X})] = M[\sum_{i=1}^{n} \lambda_i X_i] = \sum_{i=1}^{n} \lambda_i M X_i = \sum_{i=1}^{m} \lambda_i m = m \sum_{i=1}^{n} \lambda_i$$

Требуется $M[\hat{m}(\vec{X})] = m \Longrightarrow \sum_{i=1}^{n} \lambda_i = 1$. Подберем в линейной оценке $\hat{m}(\vec{X}) = \sum_{i=1}^{n} \lambda_i x_i$ параметр λ_i так, чтобы $D[\hat{m}(\vec{X})]$ было минимальным среди значений дисперсии всевозможных линейных оценок.

$$D[\hat{m}] = D[\sum_{i=1}^{n} \lambda_i X_i] = |X_i| = \sum_{i=1}^{n} \lambda_i^2 D X_i = \sigma_i^2 \sum_{i=1}^{n} \lambda_i^2$$

Поиск условий экстремума.

$$\begin{cases} f(\lambda_1 \ldots, \lambda_n) = \sum_{i=1}^{n} \lambda_i^2 \longrightarrow min \\ \vdots \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases}$$

Составим функцию Лагранжа

$$L(\lambda_1, \ldots, \lambda_n, \mu) = f(\lambda_1, \ldots, \lambda_n) \cdot \mu(\sum_{i=1}^{n} \lambda_i - 1)$$

Необходимое условие экстремума

$$\begin{cases} \frac{\partial L}{\partial \lambda_1} = 2\lambda_1 - \mu = 0 \\ \vdots \\ \frac{\partial L}{\partial \lambda_n} = 2\lambda_n - \mu = 0 \\ \frac{\partial L}{\partial \mu} = -(\sum_{i=1}^{n} \lambda_i - 1) = 0 \end{cases}$$

$\lambda_i = \frac{\mu}{2}, \quad i = \overline{1, n}.$

$\sum_{i=1}^{n} \frac{\mu}{2} = 1, \quad \frac{\mu n}{2} = 1 \Longrightarrow \mu = \frac{2}{n} \Longrightarrow \quad \lambda_i = \frac{1}{n}.$

Можно, проверив достаточное условие экстремума, показать, что $(\frac{1}{n}, \ldots, \frac{1}{n})$ является условным минимумом $f(\lambda_1, \ldots, \lambda_n)$ таким образом, линейная оценка с наименьшей дисперсией

$$\hat{m}(\vec{X}) = \sum_{i=1}^{n} \frac{1}{n} X_i = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$$

Соответствующее значение дисперсии

$$D[\hat{m}(\vec{X})]|_{(\lambda_1, \ldots, \lambda_n) = (\frac{1}{n}, \ldots, \frac{1}{n})} = \sigma^2 (\sum_{i=1}^{n} \lambda_i^2)|_{(\lambda_i = \frac{1}{n})} = \frac{\sigma^2}{n^2}$$

□

Единственность эффективной оценки. *Пусть $\hat{\theta}_1(\vec{X})$ и $\hat{\theta}_2(\vec{X})$ – две эффективные оценки θ. Тогда*

$$\hat{\theta}_1(\vec{X}) = \hat{\theta}_2(\vec{X})$$

Доказательство. Рассмотрим оценку

$$\hat{\theta} = \frac{1}{2}[\hat{\theta}_1 + \hat{\theta}_2]$$

$M\hat{\theta} = M[\frac{1}{2}(\hat{\theta}_1+\hat{\theta}_2)] = \frac{1}{2}[M\hat{\theta}_1+M\hat{\theta}_2] = |\hat{\theta}_1$ и $\hat{\theta}_2$ эффективные, а следовательно несмещенные$|=$ $\frac{1}{2}[\theta + \theta] = \theta$, то есть $\hat{\theta}$ так же является несмещенной оценкой для θ.

$D\hat{\theta} = \frac{1}{4}D[\hat{\theta}_1 + \hat{\theta}_2] = \frac{1}{4}[D\hat{\theta}_1 + D\hat{\theta}_2 + 2cov(\hat{\theta}_1, \hat{\theta}_2)] =$ |обозначим $D\hat{\theta}_1 = a^2 = D\hat{\theta}_2$| $= \frac{1}{2}[a^2 + cov(\hat{\theta}_1, \hat{\theta}_2)]$ (*)

$|cov(\hat{\theta}_1, \hat{\theta}_2)| \leq \sqrt{D\hat{\theta}_1 D\hat{\theta}_2} = a^2$

Таким образом, $D\hat{\theta} \leq$ |см.(*)|$\leq \frac{1}{2}[a^2 + a^2] = a^2$.(**)

$\hat{\theta}$ – несмещенная оценка для θ, а $\hat{\theta}_1, \hat{\theta}_2$ эффективные оценки $\Longrightarrow D\hat{\theta}_1 = D\hat{\theta}_2 \leq D\hat{\theta}$. С учетом (**) $D\hat{\theta} = a^2$.

Из (*) вытекает, что $a^2 = \frac{1}{2}[a^2 + cov(\hat{\theta}_1, \hat{\theta}_2)] \Longrightarrow cov(\hat{\theta}_1, \hat{\theta}_2) = a^2$, т.е. $cov(\hat{\theta}_1, \hat{\theta}_2) = \sqrt{D\hat{\theta}_1 D\hat{\theta}_2} \Longrightarrow$ |по свойству ковариации| $\Longrightarrow \hat{\theta}_1$ и $\hat{\theta}_2$ связаны положительной линейной зависимостью, то есть $\hat{\theta}_1 = k\hat{\theta}_2 + b (k > 0)$(***)
Из (***) следует, что $D\hat{\theta}_1 = k^2 D\hat{\theta}_2 \Longrightarrow k^2 = 1 \Longrightarrow k = 1$. Тогда $\hat{\theta}_1 = \hat{\theta}_2 + b \Longrightarrow M\hat{\theta}_1 = M\hat{\theta}_2 + b \Longrightarrow b = 0$.
Таким образом, $\hat{\theta}_1 = \hat{\theta}_2$. □

Пусть

- X – непрерывная случайная величина
- $f(t, \theta)$ – функция плотности распределения вероятностей случайной величины X
- $\vec{X} = (X_1, \ldots, X_n)$ – случайная выборка из генеральной совокупности X

Тогда функция плотности распределения случайного вектора $\vec{X}$:

$$f_{\vec{X}}(t_1, \ldots, t_n, \theta) = f(t_1, \theta) \cdot \ldots \cdot f(t_n, \theta)$$

Обозначим $(t_1, \ldots, t_n) = \vec{T}$.

Определение 5.6. Величина $I(\theta) = M\{[\frac{\partial lnf(\vec{T},\theta)}{\partial\theta}]^2\}$ называется количеством информации по Фишеру(в серии из n наблюдений).

Замечание. *Ниже иногда будет нужно дифференцировать по параметру под знаком интеграла:*

$$\frac{\partial}{\partial\theta}\int_G \phi(\vec{T}, \theta)d\vec{T} = \int_G \frac{\partial\phi(\vec{T}, \theta)}{\partial\theta}d\vec{T}$$

Параметрические модели, для которых справедлив такой переход, будем называть регулярными.

Неравенство Рао-Крамера. *Пусть*

1. *рассматривается регулярная модель*
2. *$\hat{\theta}(\vec{X})$ – несмещенная точечная оценка параметра θ закона распределения случайной величины X*

Тогда

$$D\hat{\theta}(\vec{X}) \geq \frac{1}{I(\theta)},$$

где $I(\theta)$ – количество информации по Фишеру.

Доказательство. 1. Обозначим: $G = \{t \in \mathbb{R} : f(t, \theta) > 0\}$
Тогда

$$\int_{\mathbb{R}^n} f_{\vec{X}}(\vec{T}, \theta) d\vec{T} = \int_{G^n} f_{\vec{X}}(\vec{T}, \theta) d\vec{T} = 1$$

2. Продифференцируем подчеркнутое равенство по θ:
Правая часть: $\frac{\partial 1}{\partial \theta} = 0$
Линейная часть: $\frac{\partial}{\partial \theta} \int_G f_{\vec{X}}(\vec{T}, \theta) d\vec{T} = |$модель является регулярной$| = \int_G \frac{\partial f_{\vec{X}}(\vec{T},\theta)}{\partial \theta} d\vec{T} =$ $|\frac{\partial lny}{\partial \theta} = \frac{1}{y} \frac{\partial y}{\partial \theta} \Longrightarrow \frac{\partial y}{\partial \theta} = y \frac{\partial lny}{\partial \theta}| = \int_G \frac{\partial lnf_{\vec{X}}(\vec{T}.\theta)}{\partial \theta} = f_{\vec{X}}(\vec{T}.\theta) d\vec{T} = M[\frac{\partial lnf_{\vec{X}}(\vec{T},\theta)}{\partial \theta}] = 0(*)$

3. Так как $\hat{\theta}(\vec{X})$ - несмещенная оценка для θ, то $\theta = M[\hat{\theta}(\vec{X})] = \int_G \hat{\theta}(\vec{T}) f_{\vec{X}}(\vec{T}, \theta) d\vec{T}$
Продифференцируем полученное равенство по θ: Левая часть: $\frac{\partial \theta}{\partial \theta} = 1$
Правая часть: $\frac{\partial}{\partial \theta} \int_G \hat{\theta}(\vec{T}) f(\vec{T}, \theta) d\vec{T} =|$модель является регулярной$| \int_{G^n} \hat{\theta}(\vec{T}) \frac{\partial f_{\vec{X}}(\vec{T},\theta)}{\partial \theta} d\vec{T} =| \frac{\partial y}{\partial \theta} =$ $y \frac{\partial lny}{\partial \theta}| = \int_G \hat{\theta}(\vec{T}) \frac{\partial lnf_{\vec{X}}(\vec{T},\theta)}{\partial \theta} f_{\vec{X}}(\vec{T}, \theta) d\vec{T} = M[\hat{\theta}(\vec{X}) \cdot \frac{\partial lnf_{\vec{X}}(\vec{X},\theta)}{\partial \theta}]$.
Таким образом,

$$M[\hat{\theta}(\vec{X}) \frac{\partial lnf_{\vec{X}}(\vec{X}, \theta)}{\partial \theta}] = 1 \quad (**)$$

4. Умножим обе части (*) на θ:

$$M[\frac{\partial lnf_{\vec{X}}(\vec{X}, \theta)}{\partial \theta}] = 0 \quad (***)$$

Вычтем из (**) равенство (***):

$$M[\frac{\partial lnf_{\vec{X}}(\vec{X}, \theta)}{\partial \theta} (\hat{\theta}(\vec{X}) - \theta)] = 1$$

Возведем обе части равенства в квадрат:
$1 = \{M[\frac{\partial lnf_{\vec{X}}(\vec{X}),\theta}{\partial \theta} (\hat{\theta}(\vec{X}) - \theta)]\}^2 = \{\int_{G^n} \frac{\partial lnf_{\vec{X}}(\vec{T},\theta)}{\partial \theta} (\hat{\theta}(\vec{T}) - \theta) f_X(\vec{T}, \theta) dT\}^2 = \{(a(\vec{T}), b(\vec{T}))\}^2 \leq$ $(a(\vec{T}), a(\vec{T})) \cdot (b(\vec{T}), b(\vec{T})) = \int_{G^n} [\frac{\partial lnf_{\vec{X}}(\vec{T},\theta)}{\partial \theta}]^2 f_{\vec{X}}(\vec{T}, \theta) d\vec{T} \cdot \int_{G^n} (\hat{\theta} - \theta)^2 f_X(\vec{T}, \theta) d\vec{T} =$ $M[(\frac{\partial lnf_{\vec{X}}(\vec{X},\theta)}{\partial \theta})^2] \cdot M[(\hat{\theta}(\vec{X}) - \theta)^2] = I(\theta) \cdot D[\hat{\theta}]$
Таким образом,

$$1 \leq I(\theta) D(\hat{\theta}) \Longrightarrow D(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

□

Показать, что выборочное среднее является эффективной оценкой нормальной случайной величины при известной дисперсии. *Пусть $X \sim N(\theta, \sigma^2)$, где θ неизвестно, σ^2 – известно. Показать, что $\hat{\theta}(\vec{X}) = \overline{X}$ является эффективной оценкой по Рао-Крамеру.*

Доказательство.

$$D\hat{\theta}(\vec{X}) = \frac{1}{I(\theta)}$$

$$D\hat{\theta} = D\overline{X} = D[\frac{1}{n} \sum_{i=1}^{n} X_i] = \frac{1}{n^2} \sum_{i=1}^{n} DX_i = |X_i \sim X| = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 = \frac{\sigma^2}{n}$$

$$I(\theta) = M[(\frac{\partial lnf_{\vec{X}}(\vec{X}, \theta)}{\partial \theta})^2]$$

$$f_{\vec{X}}(\vec{X},\theta) = f(X_1,\theta)\cdot\dots\cdot f(X_n,\theta)$$

$$f(X_i,\theta) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(X-\theta)^2}{2\sigma^2}}, \quad x\in\mathbb{R}$$

$$f_{\vec{X}}(\vec{X},\theta) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\theta)^2}$$

$$\ln f_{\vec{X}}(\vec{X},\theta) = \ln(\frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}) - \frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\theta)^2$$

$$(\frac{\partial \ln f_{\vec{X}}}{\partial\theta})^2 = \frac{1}{\sigma^4}(\sum_{i=1}^n (X_i-\theta) + 2\sum_{i=1}^n (X_i-\theta)(X_j-\theta))$$

$$I(\theta) = \frac{1}{\sigma^4}[M[\sum_{i=1}^n (X_i-\theta)^2]] + 2\sum_{i=1}^n M[(X_i-\theta)(X_j-\theta)] = \frac{n}{\sigma^2}$$

$$D(\hat{\theta}) * I(\theta) = \frac{\sigma^2}{n}\cdot\frac{n}{\sigma^2} = 1$$

□

6 Методы построения точечных оценок

6.1 Метод моментов

Пусть

1. X – случайная величина, закон распределения которой известен с точностью до вектора $\vec{\theta} = (\theta_1,\dots,\theta_r)$ неизвестных параметров

2. У случайной величины $X \quad \exists r$ первых моментов

Для построения точечных оценок параметров $\theta_1,\dots,\theta_r$ с использованием метода моментов необходимо сделать следующее:

1. найти выражения для r первых моментов теоретических моментов случайной величины X(так как функция распределения случайной величины X зависит от параметров $\theta_1,\dots,\theta_r$, то и теоретические моменты также будут зависеть от этих параметров):
$m_1(\theta_1,\dots,\theta_r) = M[X]$
$\vdots$
$m_n(\theta_1,\dots,\theta_r) = M[X^r]$

2. Нужно приравнять выражения для теоретических моментов к их выборочным аналогам:
$$\begin{cases} m_1(\theta_1,\dots,\theta_r) &= \hat{m}(\vec{X}) \\ \vdots \\ m_r(\theta_1,\dots,\theta_r) &= \hat{m}_r(\vec{X}) \end{cases}$$

Решаем полученную систему относительно неизвестных параметров:
$$\begin{cases} \theta_1 &= \hat{\theta}(\vec{X}) \\ \vdots \\ \theta_r &= \hat{\theta}_r(\vec{X}) \end{cases}$$

Пример 3. $X \sim Exp(\lambda, \alpha)$
$f(x) = \begin{cases} \lambda e^{-\lambda(x-\alpha)} & \textit{если } x > \alpha \\ 0 & \textit{иначе} \end{cases}$
Найдем точечные моменты:

$$m_1 = MX = \alpha + \frac{1}{\lambda}, \quad m_2 = M[X^2] = DX = \frac{1}{\lambda^2}$$

Система:
$\begin{cases} m_1 &= \alpha + \frac{1}{\lambda} = \hat{m}_1(\vec{X}) = \overline{X} \\ m_2 &= \frac{1}{\lambda^2} = S^2(\vec{X}) = \hat{\nu}_2(\vec{X}) \end{cases}$
$\hat{\lambda}(\vec{X}) = \frac{1}{S} = \frac{\sqrt{n-1}}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}}$
$\hat{\alpha}(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{\sqrt{n-1}}\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}$

6.2 Метод максимального правдоподобия

Пусть X – случайная величина, закон распределения которой известен с точностью до вектора $\hat{\theta} = (\theta_1, \ldots, \theta_r)$ неизвестных параметров.

Требуется оценить (найти) значение вектора θ.

Определение 6.1. Функцией правдоподобия, отвечающей случайной выборке $\hat{X}(X_1, \ldots, X_n)$, называется функция

$$L(\hat{X}, \hat{\theta}) = p(X_1, \hat{\theta}) \cdot \cdots \cdot p(X_n, \theta),$$

где

- $p(X_i, \vec{\theta}) = P\{X = X_i\}$, если X – дискретная случайная величина
- $p(X_i, \vec{\theta}) = f(X_i, \vec{\theta})$, где f – плотность распределения непрерывной случайной величины X

В методе максимального правдоподобия в качестве точечной оценки вектора параметров $\hat{\theta}$ используют то значение, которое доставляет функции правдоподобия максимальное значение. Таким образом, оценка максимального правдоподобия $\forall x \in \chi_n \quad L(\vec{X}, \hat{\vec{\theta}}) \geq L(\vec{X}, \vec{\theta}), \quad \vec{\theta} \in \Theta$.

$$\hat{\vec{\theta}} = argmax_{\hat{\theta}} L(\vec{X}, \vec{\theta})$$

Для построения точечной оценки необходимо решить задачу

$$L(\vec{X}, \vec{\theta} \longrightarrow max_{\vec{\theta}) \in \Theta},$$

вместо которой чаще решают задачу:

$$lnL(\vec{X}, \vec{\theta}) \longrightarrow max_{\vec{\theta} \in \Theta}$$

Если для функции lnL выполнены соответствующие условия, то для нахождения значения $\vec{\theta}$ можно использовать систему уравнений:
$\begin{cases} \frac{\partial lnL(\vec{X}, \vec{\theta})}{\partial \theta_1} = 0 \\ \vdots \\ \frac{\partial lnL(\vec{X}, \vec{\theta})}{\partial \theta_r} = 0 \end{cases}$

7 Доверительные интервалы

Определение 7.1. γ– доверительным интервалом(доверительным интервалом уровня γ) для параметра θ называется пара статистик

$$\underline{\theta(\vec{X})}, \overline{\theta(\vec{X})}, \text{ таких, что}$$
$$P\{\theta \in (\underline{\theta(\vec{X})}, \overline{\theta(\vec{X})})\} = \gamma$$

Пусть

1. θ – неизвестный параметр закона распределения случайной величины X
2. $g(\vec{X}, \theta)$ – некоторая статистика

Определение 7.2. Статистику $g(\vec{X}, \theta)$ будем называть центральной, если закон ее распределения не зависит от θ, то есть

$$F_g(X, \theta) \equiv F_g(X), \text{ где } F_g \text{ – функция распределения случайной величины } g$$

Общий алгоритм. *Пусть*

1. *X – случайная величина, закон распределения которой зависит от неизвестного параметра θ*
2. *$g(\vec{X}, \theta)$ – центральная статистика*
3. *$g(X, \theta)$ является монотонно возрастающей с увеличением параметра θ*
4. *$F_g(X, \theta)$ также монотонно возрастает с увеличением θ*
5. *$\alpha_1 > 0, \alpha_2 > 0$ и таковы, что $\alpha_1 + \alpha_2 = 1 - \gamma$*

$\gamma = P\{q_{\alpha_1} < g(\vec{X}, \theta) < q_{1-\alpha_2}\} =$ */g монотонно возрастает с ростом θ/* $= P\{g^{-1}(\vec{X}, q_{\alpha_1}) < \theta < g^{-1}(\vec{X}, q_{1-\alpha_2})\}$.

Частные случаи. *$X \sim N(m, \sigma^2)$, где*

- *m – неизвестно*
- *σ^2 – известно*

$g(\vec{X}, m) = \frac{m-\overline{X}}{\sigma}\sqrt{n} \sim N(0,1)$, *то есть $g(\vec{X}, m)$ – центральная статистика.*
$\alpha_1 = \alpha_2 = \frac{1-\gamma}{2} = /1 - \gamma = \alpha/ = \frac{\alpha}{2}$
$\gamma = P\{-q_{1-\frac{\alpha}{2}} < g(\vec{X}, m) < q_{1-\frac{\alpha}{2}}\} = P\{-q_{1-\frac{\alpha}{2}} < \frac{m-\overline{X}}{\sigma}\sqrt{n} < q_{1-\frac{\alpha}{2}}\} = P\{\overline{X} - \frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}} < m < \overline{X} + \frac{\sigma q_{1-\frac{\alpha}{2}}}{\sqrt{n}}\}$
Если неизвестны оба параметра m и σ^2, то при построении доверительных интервалов для этих параметров:

$$g(\vec{X}, m) = \frac{m-\overline{X}}{S(\vec{X})}\sqrt{(n)} \sim St(n-1)$$
$$g(\vec{X}, m) = \frac{S(\vec{X})^2}{\sigma^2}(n-1) \sim \chi^2(n-1)$$