- *Для непрерывного случайного вектора*

$$D[X \,|\, Y = y_j] = \int\limits_{-\infty}^{+\infty} (x - M[X \,|\, Y = y_j])^2 f_X(x \,|\, Y = y_j)\, dx$$

# Весенний семестр
# Лекция №17, 11.02.2019

## 2.6 Предельные теоремы теории вероятностей

### 2.6.1 Неравенства Чебышева

**Теорема.** ***Первое неравенство Чебышева.***
*Пусть*

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

*Тогда*

$$\forall \varepsilon > 0 \;\; P\{X \geqslant \varepsilon\} \leqslant \frac{MX}{\varepsilon}$$

**Доказательство.** *(для случая непрерывной случайной величины* $X$*, для дискретной случайной величины* $X$ *доказательство аналогично).*

$$MX = \int\limits_{-\infty}^{+\infty} x\, f(x)\, dx = \langle \textit{свойство аддитивности определённого интеграла} \rangle =$$

$$= \int\limits_{0}^{\varepsilon} x\, f(x)\, dx + \int\limits_{\varepsilon}^{+\infty} x\, f(x)\, dx \geqslant \int\limits_{\varepsilon}^{+\infty} x\, f(x)\, dx \geqslant \langle \forall x \in [\varepsilon;\, \infty)\; x \geqslant \varepsilon \rangle \geqslant$$

$$\geqslant \varepsilon \int\limits_{\varepsilon}^{+\infty} f(x)\, dx = \varepsilon\, P\{X \geqslant \varepsilon\}$$

**Теорема.** ***Второе неравенство Чебышева.***
*Пусть*

1. $X$ *— случайная величина;*

*2.* $\exists MX$, $\exists DX$.

*Тогда*

$$\forall \varepsilon > 0 \ \ P\{|X - MX| \geqslant \varepsilon\} \leqslant \frac{DX}{\varepsilon^2}$$

**Доказательство.**

$$DX = M[(X - MX)^2] \geqslant$$

$\langle$*Рассмотрим случайную величину* $Y = (X - MX)^2 \geqslant 0$*; из первого неравенства Чебышева для* $Y$ *следует, что* $\forall \sigma > 0 \ P\{Y \geqslant \sigma\} \leqslant \dfrac{MY}{\sigma}$. *Используем это для* $\sigma = \varepsilon^2\rangle$

$$\geqslant \sigma P\{(X - MX)^2 \geqslant \sigma\} = \langle \sigma = \varepsilon^2 \rangle = \varepsilon^2 P\{(X - MX)^2 \geqslant \varepsilon^2\} = \varepsilon^2 P\{|X - MX| \geqslant \varepsilon\}$$

*Таким образом,*

$$DX \geqslant \varepsilon^2 P\{|X - MX| \geqslant \varepsilon\} \implies P\{|X - MX| \geqslant \varepsilon\} \leqslant \frac{DX}{\varepsilon^2}$$

**Задача.** *Значение предельно допустимого давления в пневмокамере ракеты составляет* 200 *Па. После проверки большого кол-ва ракет было установлено, что среднее значение давления в пневмокамерах составляет* 150 *Па. Оценить вероятность того, что давление в очередной проверенной камере ракеты превысит предельно допустимое значение, если также известно, что среднеквадратичное отклонение значения давления составляет* 5 *Па.*

**Решение.**

- *Пусть* $X$ *— случайная величина, принимающая значения, равные значениям давления в пневмокамерах этих ракет (в Па). Тогда* $MX = 150$, $DX = 5^2$.

- *Используем для оценки первое неравенство Чебышева, т. к.* $X \geqslant 0$*:*

$$P\{X \geqslant 200\} \leqslant \frac{MX}{200} = \frac{3}{4}$$

- *Используем второе неравенство Чебышева:*

$$P\{X \geqslant 200\} = P\{X - MX \geqslant 200 - 150\} = P\{X - MX \geqslant 50\} \leqslant$$
$$\leqslant P\left\{\{X - MX \geqslant 50\} + \{X - MX \leqslant -50\}\right\} = P\{|X - MX| \geqslant \underbrace{50}_{\varepsilon}\} \leqslant$$
$$\leqslant \langle \text{второе неравенство Чебышева} \rangle \leqslant \frac{DX}{50^2} = \frac{5^2}{50^2} = \frac{1}{100}$$

**Замечание 1.** *В результате решения примера получено:*

- *С использованием первого неравенства Чебышева:*

$$P\{X \geqslant 200\} \leqslant \frac{3}{4}$$

- *С использованием второго неравенства Чебышева:*

$$P\{X \geqslant 200\} \leqslant \frac{1}{100}$$

*Второе неравенство Чебышева даёт существенно более точный результат, т. к. дополнительно использует информацию о значении дисперсии.*

**Замечание 2.** *Очевидно, что*

1. *Если в первом неравенстве Чебышева $\varepsilon \leqslant MX$, то оно даёт тривиальную оценку: $P \leqslant 1$;*

2. *Если во втором неравенстве Чебышева $\varepsilon \leqslant \sqrt{DX}$, то оно также даёт тривиальную оценку.*

### 2.6.2 Виды сходимости последовательности случайных величин

Пусть $X_1$, $X_2$, ... — последовательность случайных величин. Мы изучим лишь два вида сходимости:

1. Сходимость по вероятности;

2. Слабая сходимость.

**Определение.** *Говорят, что последовательность случайных величин $X_1$, $X_2$, ... сходится по вероятности к случайной величине $Z$, если*

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

*Обозначается как $X_n \xrightarrow[n\to\infty]{P} Z$.*

**Определение.** *Говорят, что последовательность величин $X_1$, $X_2$, ... слабо сходится к случайной величине $Z$, если функциональная последовательность*

$$F_{X_1}(x),\, F_{X_2}(x),\, \ldots$$

*поточечно сходится к функции $F_Z(x)$ во всех точках непрерывности последней, т. е.*

$$(\forall x_0 \in \mathbb{R})((\text{функция } F_Z(x) \text{ непрерывна в } x_0) \implies \underbrace{F_{X_n}(x_0)}_{\text{числовая последовательность}} \xrightarrow[n\to\infty]{} F_Z(x)$$

*Обозначается как $F_{X_n}(x) \implies F_Z(x)$.*

### 2.6.3 Закон больших чисел

**Определение.** *Говорят, что последовательность $X_1$, $X_2$, ... удовлетворяет закону больших чисел, если*

$$\forall \varepsilon > 0 \ P\Big\{ \left| \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} m_i \right| \geqslant \varepsilon \Big\} \xrightarrow[n\to\infty]{} 0$$

*где $m_i = MX_i$, $i \in \mathbb{N}$.*

**Замечание 1.** *Рассмотрим случайную величину $\overline{X_n} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$.*

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

*Выполнение закона больших чисел для последовательности $X_1$, $X_2$, ... означает, что*

$$\forall \varepsilon > 0 \ P\{|\overline{X_n} - M\overline{X_n}| \geqslant \varepsilon\} \xrightarrow[n\to\infty]{} P$$

*т. е. при достаточно больших $n$ случайная величина $\overline{X_n}$ принимает детерминированное (т. е. неслучайное) значение $\frac{1}{n}\sum\limits_{i=1}^{n} m_i$ с вероятностью, близкой к 1. Таким образом, в этом случае случайная величина $\overline{X_n}$ при достаточно больших $n$ утрачивает случайный характер.*

**Замечание 2.** *Выполнение закона больших чисел для последовательности $X_1$, $X_2$, ... означает, что последовательность*

$$Y_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - m_i)$$

*при $n \to \infty$ сходится по вероятности к нулю, т. е.*

$$\forall \varepsilon > 0 \ P\{|Y_n - 0| \geqslant \varepsilon\} \xrightarrow[n\to\infty]{} 0$$

Вопрос: какой должна быть последовательность $X_1$, $X_2$, ..., чтобы она удовлетворяла закону больших чисел?

**Теорема.** *(Чебышева, достаточное условие того, что последовательность удовлетворяет закону больших чисел)*

*Пусть*

1. *$X_1$, $X_2$, ... — последовательность независимых случайных величин;*
2. *$\exists MX_i = m_i$, $\exists DX_i = \sigma_i^2$, $i \in \mathbb{N}$;*

3. *Дисперсии случайных величин $X_1$, $X_2$, ... ограничены в совокупности, т. е. $\exists C > 0\ \sigma_i^2 \leqslant C,\ i \in \mathbb{N}$.*

*Тогда последовательность $X_1$, $X_2$, ... удовлетворяет закону больших чисел.*

**Доказательство.**

- *Рассмотрим $\overline{X_n} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i,\ n \in \mathbb{N}$. Тогда*

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

$$D\left[\overline{X_n}\right] = D\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} D\left[\sum_{i=1}^{n} X_i\right] = \langle X_i \text{ независимы}\rangle =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2$$

- *Применим к случайной величине $\overline{X_n}$ второе неравенство Чебышева:*

$$P\left\{|\overline{X_n} - M\overline{X_n}| \geqslant \varepsilon\right\} \leqslant \frac{D\overline{X_n}}{\varepsilon^2}$$

*Таким образом,*

$$P\Big\{\left|\overline{X_n} - \frac{1}{n}\sum_{i=1}^{n} m_i\right| \geqslant \varepsilon\Big\} \leqslant \frac{1}{\varepsilon^2}\cdot\frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 \tag{13}$$

$$\sum_{i=1}^{n}\sigma_i^2 \leqslant \langle \sigma_i^2 \leqslant C\rangle \leqslant \sum_{i=1}^{n} C = nC$$

$$0 \leqslant P\Big\{\left|\overline{X_n} - \frac{1}{n}\sum_{i=1}^{n} m_i\right| \geqslant \varepsilon\Big\} \leqslant \frac{C}{\varepsilon^2}\cdot\frac{1}{n^2}\cdot n = \overbrace{\frac{C}{\varepsilon^2}\cdot\frac{1}{n}}^{\nearrow\ 0,\ n\to\infty}$$

*При $n \to \infty$ по теореме о двух миллиционерах*

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

*т. е. последовательность $X_1$, $X_2$, ... удовлетворяет закону больших чисел.*

**Замечание 1.** *Если последовательность $X_1$, $X_2$, ... удовлетворяет условию теоремы Чебышева, то, разумеется, она удовлетворяет закон больших чисел. При этом говорят, что эта последовательность удовлетворяет закону больших чисел в форме Чебышева.*

**Замечание 2.** *Условие 3 теоремы Чебышева можно ослабить, заменив его на условие 3′:*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} DX_i = 0$$

*Из оценки 13 следует, что в этом случае последовательность $X_1$, $X_2$, … также удовлетворяет закону больших чисел.*

**Следствие 1.** *Пусть*

1. *Выполнены условия теоремы Чебышева;*
2. *Все случайные величины $X_i$ одинаково распределены (обозначим $m \equiv m_i \equiv MX_i$).*

   *Тогда*

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

**Доказательство.** *Т. к. $m_i \equiv m$, то $\frac{1}{n}\sum\limits_{i=1}^{n} m_i = m$ и используем закон больших чисел в форме Чебышева.*

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

*Пусть*

1. *Проводится $n$ испытаний по схеме Бернулли с вероятностью успеха $p$;*
2. $$r_n = \frac{\{\text{число «успехов» в этой серии}\}}{n}$$

   *($r_n$ называют относительной (наблюдательной) частотой успеха)*

*Тогда $r_n \xrightarrow[n\to\infty]{P} P$.*

**Доказательство.**

- *Введём случайные величины $X_i$, $i = \overline{1; n}$,*

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

*Тогда*

1. *Закон распределения случайной величины $X_i$:*

$$\begin{array}{ccc} X_i & 0 & 1 \\ P & q = 1 - p & p \end{array}$$

*Таким образом, все $X_i$ одинаково распределены;*

2. $DX_i \equiv pq$ *— ограничены в совокупности;*
3. $X_i$ *независимы, т. к. отдельные испытания в схеме Бернулли независимы;*
4. *Таким образом, последовательность $X_1$, $X_2$, … удовлетворяет первому следствию из теоремы Чебышева и для неё справедливо*

$$\forall \varepsilon > 0\ P\Big\{\Big|\frac{1}{n} \underbrace{\sum_{i=1}^{n} X_i}_{=r_n,\ \text{общее число успехов из } n \text{ испытаний}} - \underbrace{m}_{=P}\Big| \geqslant \varepsilon\Big\} \xrightarrow[n\to\infty]{} 0$$

$$\forall \varepsilon > 0\ P\{|r_n - P| \geqslant \varepsilon\} \xrightarrow[n\to\infty]{} 0$$

*т. е.* $r_n \xrightarrow[n\to\infty]{P} P$.

# Лекция №18, 18.02.2019

### 2.6.4 Центральная предельная теорема

Пусть:

1. $X_1$, $X_2$, … — последовательность независимых случайных величин;
2. Все случайные величины из этой последовательности одинаково распределены;
3. $\exists MX_i = m$, $\exists DX_i = \sigma^2$, $i \in \mathbb{N}$.

Рассмотрим случайную величину $\overline{X_n} = \frac{1}{n}\sum\limits_i X_i$, т. е. $\overline{X}_1 = X_1$, $\overline{X}_2 = \frac{1}{2}(X_1 + X_2)$ и т. д. Тогда $M\overline{X_n} = m$, $n \in \mathbb{N}$; $D\overline{X_n} = \frac{\sigma^2}{n}, n \in \mathbb{N}$. Рассмотрим случайную величину

$$Y_n = \frac{\overline{X_n} - M\overline{X_n}}{\sqrt{D\overline{X_n}}} = \frac{\overline{X_n} - m}{\left(\dfrac{\sigma}{\sqrt{n}}\right)},\ n \in \mathbb{N}$$

Для неё $MY_n = 0$, $DY_n = 1$.

**Теорема.** ***Центральная предельная теорема.***
*Пусть выполнены условия 1-3.*
*Тогда последовательность величин $Y_n$ при $n \to \infty$ слабо сходятся к случайной величине $Z$, имеющей стандартное нормальное распределение, т. е.*

$$\forall x \in \mathbb{R}\ F_Y(x) \xrightarrow[n\to\infty]{} F_Z(x)$$

*где $Z \sim N(0,1)$, т. е.* $F_Z(x) = \Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{\frac{-t^2}{2}}\, dt$.

**Доказательство.** *Без доказательства.*

**Замечание.** *В этом случае говорят, что случайные величины $Y_n$ имеют асимптотически стандартное нормальное распределение.*

**Пример.** *ЭВМ проводят суммирование $10^4$ чисел, каждое из которых округлено до $10^{-4}$. Считая, что ошибки округления отдельных слагаемых независимы и имеют равномерное распределение на отрезке $[-0.5{\cdot}10^4;\ 0.5{\cdot}10^{-4}]$ найти диапазон, в котором с вероятностью в $0.95$ будет заключена ошибка результата суммирования.*

**Решение.**

1. *Пусть $X_i$ — ошибка округления $i$-ого слагаемого, $i = \overline{1;\ 10^4}$;*
   $X_i \sim R[-0.5 \cdot 10^{-4};\ 0.5 \cdot 10^{-4}] \implies MX_i = 0; DX_i = (0.5 \cdot 10^{-4} - (-0.5 \cdot 10^{-4}))^2 \cdot \frac{1}{12} = \underbrace{\frac{1}{12}}_{\approx \frac{1}{10}} 10^{-8} \approx 10^{-9}$

2. *Тогда погрешность суммы всех таких слагаемых $S = \sum\limits_i X_i$, где $n = 10^4$.*

3. *Рассмотрим случайную величину $Y_n = \frac{\overline{X_n} - MX_i}{\sqrt{D\overline{X_n}/n}} = \frac{\overline{X_n}}{\sqrt{10^{-9}/10^4}} = \frac{\overline{X_n}}{\sqrt{10^{-13}}}$, $n = 10^4$.*

   *Т. к. $n \gg 1$, то в соответствии с ЦПТ можно полагать, что*
   $$Y_n \underbrace{\sim}_{\text{(приближённо)}} N(0,\ 1)$$

4. *$u_{0.975}$ — квантиль уровня $0.975$ стандартного нормального распределения, из таблицы $u_{0.975} = 1.96$ Таким образом, $P\{|Y_n| < u_0.975\} = 0.95$. Таким образом, с вероятностью $0.95$ можно утверждать, что*
   $\left|\frac{\overline{X_n}}{\sqrt{10^{-13}}}\right| < 1.96 \iff \left|\frac{1}{n}\overline{X_n}\right| < 1.96 \cdot \sqrt{10^{-13}} \iff \left|\overline{X_n} = \frac{1}{n}S\right| \iff |\sigma| \leqslant 1.96 \cdot n \cdot \sqrt{10^{-13}} \approx 6.2 \cdot 10^{-3}$

***Ответ.*** *С вероятностью $0.95$ можно утверждать, что погрешность результата не превышает $6.2 \cdot 10^{-3}$.*

**Теорема.** *Центральная теорема Муавра-Лапласа.*
*Пусть*

1. *Проводится большое число испытаний $n$ по схеме Бернулли с вероятностью успеха $p$;*

2. *$k$ — число успехов в этой серии.*

*Тогда*
$$P\{k_1 \leqslant k \leqslant k_2\} = \Phi(x_2) - \Phi(x_1)$$

где $x_i = \frac{k_i - np}{\sqrt{npq}}$, $i = \overline{1;\,2}$, $q = 1 - p$, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$.

**Доказательство.**

1. *Пусть $X_i$ — случайная величина, принимающая значения 0 или 1 в соответствии с правилом*

$$X_i = \begin{cases} 1, & \text{если в } i\text{-ом испытании был успех;} \\ 0 & \text{иначе} \end{cases}, \; i = \overline{1;\,n}$$

   *Тогда*

   (a) *Случайные величины $X_1, \ldots, X_n$ независимы;*

   (b) *$MX_i = p$, $DX_i = pq$, $i \in \mathbb{N}$;*

   (c) *$X_i$ одинаково распределены.*

2.

$$P\{k_1 \leqslant k \leqslant k_2\} = P\{k_1 \leqslant \sum_{i=1}^{n} X_i \leqslant k_2\} = P\left\{\frac{k_1}{n} - p \leqslant \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{=\overline{X_n}} - p \leqslant \frac{k_2}{n} - p\right\} =$$

$$= P\{\underbrace{\frac{\frac{k_1}{n} - p}{\sqrt{pqn}}}_{=X_1} \leqslant \underbrace{\frac{\overline{X_n} - \overbrace{p}^{=MX_i}}{\underbrace{\sqrt{pqn}}_{=DX_i}}}_{=Y_n} \leqslant \underbrace{\frac{\frac{k_2}{n} - p}{\sqrt{pqn}}}_{=X_2}\} =$$

$$= \langle \text{в соответствии с ЦПТ (т. к. } n \gg 1) \; Y_n \sim N(0,\,1) \text{ приближённо} \rangle =$$
$$= \Phi(x_2) - \Phi(x_1)$$

**Задача.** *В эксперименте Пирсона по подбрасыванию симметричной монеты 24000 раз герб выпал 12012 раз. Какова вероятность того, что при повторном испытании (повторении эксперимента) абсолютное отклонение относительной частоты от $\frac{1}{2}$ окажется таким же или больше?*

**Решение.**

1. *Зададим схему Бернулли, успех — выпадение герба, $p = q = \frac{1}{2}$;*

2.

$$A = \{ \text{ относительное отклонение частоты выпадения герба от } \frac{1}{2}$$
$$\text{будет не меньше, чем в эксперименте Пирсона } \}$$

*Тогда событие* $\overline{A}$ *будет*

$$A = \{ \text{ относительное отклонение частоты выпадения герба от } \frac{1}{2}$$
$$\text{будет меньше, чем в эксперименте Пирсона } \}$$

$$P\{\overline{A}\} = P\left\{\left|\frac{k}{n} - \frac{1}{2}\right| < \frac{12012}{n}\right\} = P\{11988 < k < 12012\} \approx$$
$$\approx \langle \text{интегральной теоремой Муавра-Лапласа, т. к. } n = 24000 \gg 1 \rangle \approx$$
$$\approx \Phi_0(x_2) - \Phi_0(x_1) =$$
$$= \left\langle x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{12012 - 12000}{\sqrt{24000 \cdot \frac{1}{4}}} \approx 0.155,\ x_1 = -x_2 \right\rangle =$$
$$= 2\,\Phi_0(0.155) \approx 0.123$$

*Таким образом,* $P(A) = 1 - P(\overline{A}) = 0.877.$

# 3 Математическая статистика

## 3.1 Основные понятия выборочной теории

### 3.1.1 Основные определения

Теория вероятностей является одной из ветвей «чистой» математики, которая строится дедуктивно, исходя из определённого набора аксиом. Математическая статистика является разделом прикладной математики, которая строится индуктивно от наблюдений к гипотезе. При этом выводы и аргументация математической статистики базируется на результатах теории вероятностей.

**Пример 1.** ***Типовые задачи теории вероятностей:*** *известно, что при одном подбрасывании симметричной монеты герб выпадает с вероятностью $p$. Какова вероятность того, что при $n$ подбрасываниях герб выпадет $n$ раз?*

**Пример 2.** ***Типовая задача математической статистики:*** *известно, что при $n$ подбрасываниях монеток герб выпал $n$ раз. Чему равна вероятность $p$ выпадения герба при одном подбрасывании?*

Основная задача математической статистики — разработка методов получения научно обоснованных выводов о массовых явлениях или процессах по данным наблюдения или экспериментов. Эти выводы относятся не к отдельным экспериментам, а а являются утверждениями относительно вероятностных характеристик изучаемых явлений или процессов.

**Замечание.** *Упрощённо можно считать, что «общая» задача математической статистики формулируется следующим образом:*

***Дано:*** *имеется случайная величина $X$, закон распределения которой неизвестен или известен не полностью. Требуется по результатам серии наблюдений сделать выводы о законе распределения этой случайной величины.*

При этом различают первую и вторую задачи математической статистики.

**Первая основная задача Математической статистики.** Есть случайная величина $X$, закон распределения которой неизвестен. Требуется найти закон распределения случайной величины $X$ (требуется проверить гипотезу о законе распределения случайной величины $X$).

**Вторая основная задача математической статистики.** Есть некоторая случайная величина $X$, закон распределения которой известен с точностью до вектора параметров (таким образом, известен общий вид закона распределения случайной величины $X$, но неизвестны значения некоторых числовых параметров этого закона). Нужно найти или оценить значения этих параметров.

**Пример.** ***Вторая основная задача МС.*** *Известно, что случайная величина $X$ имеет нормальное распределение, но значения параметров $m$ и $\sigma$ неизвестны. Требуется по результатам наблюдений оценить значения этих параметров. $\vec{\theta} = (m, \sigma^2)$ — вектор неизвестных параметров.*

**Определение.** *Множество возможных значений случайной величины $X$ называют генеральной совокупностью.*

**Определение.** ***Случайной*** *выборкой из генеральной совокупности $X$ называется 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 $n$ называют объёмом случайной выборки.*

**Замечание.** *Всюду далее векторы будут обозначаться стрелочкой $\vec{X}$,но не чёрточкой $\overline{X}$. Черта будет обозначать выборочное среднее: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.*

**Замечание 2.** *Пусть $F(t)$ — функция распределения случайной величины $X$. Тогда функция распределения случайной выборки $\vec{X}$ объёма $n$ из генеральной совокупности $X$:*

$$F_{\vec{X}}(t_1, \ldots, t_n) = P\{X_1 < t_1, \ldots, X_n < t_n\} = \langle X_i \text{ — независимы } \rangle =$$
$$= P\{X_1 < t_1\} \cdot P\{X_2 < t_2\} \cdot \ldots \cdot P\{X_n < t_n\} = \langle X_i \sim X \rangle =$$
$$=?\langle \text{ далее додумано редактором } \rangle \prod_{i=1}^{n} F(t_i)$$

**Определение.** *Любую возможную реализацию $\vec{x} = (x_1, \ldots, x_n)$ случайной выборки $\vec{X}$ называются выборкой («просто» выборкой) из генеральной совокупности $X$. При этом число $x_k$ называют $k$-ым элементом выборки $\vec{x}$.*

**Определение.** *Множество всех возможных значений случайной выборки $\vec{x}$ называют выборочным пространством и обозначают $X_n$.*

**Определение.** *Любую функцию $g(\vec{X})$ случайной выборки $\vec{X}$ называют статистикой или выборочной характеристикой.*

Закон распределения $F_g(\vec{t})$ называются выборочным законом распределения статистики $g$.

**Замечание 1.** *Значение $g(\vec{x})$ статистики $g$ на выборке $\vec{x}$ называется её выборочным значением.*

**Замечание 2.** *Пусть $\vec{x}$ — реализация случайной выборки $\vec{X}$. При этом $\vec{x} = (x_1, \ldots, x_n)$, это позволяет считать, что случайная величина $X$ моделируется дискретной случайной величиной, закон (ряд) распределения которой имеет следующий вид:*

$$\begin{array}{ccccc} \vec{x} & x_1 & x_2 & \ldots & x_n \\ P & \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{array}$$

*Математическое ожидание такой дискретной случайной величины $m = \frac{1}{n}\sum x_i$, дисперсия $\sigma^2 = \frac{1}{n}\sum(x_i - m)^2$*

*Эти соображения приводят к следующему определению.*

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

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

**Определение.** *Выборочной дисперсией называют вот статистику*

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

# Лекция №19, 25.02.2019

**Определение.** *Выборочным начальным моментом порядка $k$ называют статистику*

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

**Определение.** *Центральным выборочным моментом порядка $k$ называют статистику*[41]

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

**Замечание.** *Очевидно, что* $\overline{X} = \hat{m}_1(\vec{X})$, $\hat{\sigma}^2(\vec{X}) = \hat{\nu}_2(\vec{X})$.

### 3.1.2 Предварительная обработка результатов эксперимента

Пусть $\vec{X}$ — случайная выборка из генеральной совокупности $X$, а $\vec{x}$ — её реализация (выборка).

#### 3.1.2.1 Вариационный ряд

Расположим элементы выборки $\vec{x}$ в порядке неубывания, $i$-ый элементы полученной последовательности обозначим $x_{(i)}$, $i = \overline{1;\,n}$: $x_{(1)} \leqslant \ldots \leqslant x_{(n)}$.

**Замечание.** $x_{(1)}$ *— наименьший член этого ряда,* $x_{(n)}$ *— наибольший член этого ряда.*

**Определение.** *Последовательность* $x_{(1)}, \ldots, x_{(n)}$ *называют вариационным рядом выборки* $\vec{x}$. *При этом* $x_{(i)}$ *называют $i$-ым членом вариационного ряда.*

**Определение.** *Вариационным рядом случайной выборки* $\vec{X} = (X_1, \ldots, X_n)$ *называется последовательность случайных величин* $X_{(1)}, \ldots, X_{(n)}$, *где для каждой реализации* $\vec{x}$ *случайной выборки* $\vec{X}$ *случайная величина* $X_{(i)}$ *принимает значение, равное $i$-ому члену вариационного ряда для выборки* $\vec{x}$.

[41] $\nu$ — строчная буква «ню» греческого алфавита. — Прим. ред.

**Замечание 1.** $P\{X_{(i)} \leqslant X_{(i+1)}\} = 1,\ i = \overline{1;\, n-1}$

**Замечание 2.** *Пусть $F(x)$ — функция распределения случайной величины $X$; тогда*

$$F_{X_{(n)}}(x) = P\{X_{(n)} < x\} = \left\langle X_{(n)} = \max_{i=\overline{1;\,n}}\{X_i\} \right\rangle =$$
$$= P\{X_1 < x,\, X_2 < x,\, \ldots,\, X_n < x\} = \langle X_{(i)} \text{ — независимые}\rangle =$$
$$= P\{X_1 < x\} \cdot \ldots \cdot P\{X_n < x\} = F(x) \cdot \ldots \cdot F(x) = [F(x)]^n$$

$$F_{X_{(1)}} = P\{X_{(1)} < x\} = 1 - P\{X_{(1)} \geqslant x\} = \left\langle X_{(1)} = \min_{i=\overline{1;\,n}\{X_i\}} \right\rangle =$$
$$= 1 - P\{X_1 \geqslant x\} \cdot \ldots \cdot P\{X_n \geqslant x\} =$$
$$= 1 - (1 - P\{X_1 < x\}) \cdot \ldots \cdot (1 - P\{X_n < x\}) = 1 - (1 - F(x))^n$$

#### 3.1.2.2 Статистический ряд

Среди элементов выборки $\vec{x} = (x_1,\, \ldots,\, x_n)$ некоторые могут совпадать. Это возможно в случае, когда случайная величина $X$ является дискретной или в случае, если $X$ — непрерывная случайная величина, но при проведении эксперимента полученные значения округлялись.

Пусть $z_{(1)}, \ldots, z_{(m)}$ — все попарно различные значения, встречающиеся в выборке $\vec{x}$, причём $z_{(1)} < \ldots < z_{(m)}$.

**Определение.** *Статистическим рядом, отвечающий выборке $\vec{x}$, называется таблица*

$$\begin{array}{ccccc} z_{(1)} & \ldots & z_{(i)} & \ldots & z_{(m)} \\ n_1 & \ldots & n_i & \ldots & n_m \end{array}$$

*Здесь $n_i$ — кол-во элементов выборки $\vec{x}$, которые имеют значение $z_{(i)}$.*

**Замечание 1.** $n_1 + ... + n_m = n$;

**Замечание 2.** *$n_i$ называют частотой значения $z_{(i)}$, а $\frac{n_i}{n}$ — относительной частотой этого значения.*

#### 3.1.2.3 Эмпирическая и выборочная функции распределения

Пусть $\vec{x} = (x_1,\, \ldots,\, x_n)$ — выборка из генеральной совокупности $X$.

**Определение.** *Эмпирической функцией распределения, отвечающей выборке $\vec{x}$, называется функция $F_n(x) = \frac{h(x,\vec{x})}{n}$ где $h(x,\, \vec{x})$ — кол-во элементов выборки $\vec{x}$, которые имеют значение, меньшее $x$.*

**Замечание 1.** *$F_n(x)$ обладает всеми свойствами функции распределения:*

1. *$F_n$ — неубывающая;*

2. $$\lim_{x\to-\infty} F_n(x) = 0,\ \lim_{x\to+\infty} F_n(x) = 1$$

3. *$F_n(x)$ — непрерывна слева в каждой точке $x \in \mathbb{R}$. При этом $F_n(x)$ кусочно-постоянна и изменяется скачками в точках $x = z_{(i)}$.*

   *Если все элементы выборки $\vec{x}$ попарно различны, то*

$$F_n(x) = \begin{cases} 0, & x \leqslant x_{(1)} \\ \frac{i}{n}, & x_{(i)} < x \leqslant x_{(i+1)},\ i = \overline{1;\, n-1} \\ 1, & x > x_{(n)} \end{cases}$$

**Замечание 2.** *Эмпирическая функция распределения позволяет интерпретировать выборку $\vec{x}$ как реализацию дискретной случайной величины $\tilde{X}$, закон распределения которой имеет следующий вид:*

$$\begin{matrix} \tilde{X} & z_{(1)} & \dots & z_{(m)} \\ P & \frac{n_1}{n} & \dots & \frac{n_m}{n} \end{matrix}$$

*Здесь $n_i$ — частота значения $z_{(i)}$.*

*В дальнейшем это позволит рассматривать числовые характеристики случайной величины $\tilde{X}$ как приближённые значения числовых характеристик случайной величины $X$.*

**Определение.** *Выборочной функцией распределения, отвечающей случайной выборке $\hat{X}$, называется функция*

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

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

**Замечание 1.** *Зафиксированное некоторое значение $x \in \mathbb{R}$,*

$$P\{X_i < x\} = \langle X_i \text{ независимса и одинаково распределена с } X \rangle = \\ = \underbrace{P\{X < x\}}_{\text{обозначим } p \text{ (для данного фиксированного } x)}$$

**Замечание 2.** *Какие в принципе значения может принимать функция $\hat{F}_n(x)$ при фиксированном $x$?*

$$0,\, \frac{1}{n},\, \frac{2}{n},\, \dots,\, \frac{n-1}{n},\, 1$$

*Тогда* $\hat{F}_n(x) = \frac{k}{n} \iff n(x,\, \vec{X}) = k$, *поэтому*

$$P\{\hat{F}_n(x) = \tfrac{k}{n}\} = P\{n(x,\, \vec{X}) = k\} =$$
$$= P\{\textit{среди случайных величин } X_1,\, \ldots,\, X_n \textit{ ровно } k \textit{ приняли значение меньше, чем } x\} =$$
$$= P\{\textit{в серии из } n \textit{ испытаний по схеме Бернулли с вероятностью успеха } p$$
$$\textit{произошло ровно } k \textit{ успехов } \} = C_n^k p^k q^{n-k}$$

*где* $p = P\{X_i < x\}$, $q = 1 - p$.

*Таким образом, для каждого фиксированного* $x$ *случайная величина* $n \cdot \hat{F}_n$ *имеет биномиальное распределение, т. е.* $n \cdot \hat{F}_n \sim B(n,\, p)$.

**Теорема.** *Для любого фиксированного* $x \in \mathbb{R}$ *последовательность* $\hat{F}_n(x)$ *сходится по вероятности к значению* $F(x)$ *теоретической (т. е. «истинной») функции распределения случайной величины* $X$*:*

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

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

#### 3.1.2.4 Интервальный статистический ряд

Выше было введено понятие статистического ряда, однако при больших объёмах выборки ($n > 50$) его использование не очень удобно ввиду громоздкости. В этом случае отрезок $[x_{(1)};\, x_{(n)}]$ разбивают на $m$ равновеликих интервалов и для каждого интервала указывают кол-во принадлежащих ему элементов выборки. При этом, как правило, $m$ определяют как:

$m = [\log_2 n] + 1$ — кол-во интервалов, $\triangle = \frac{x_{(n)} - x_{(1)}}{m}$ — ширина одного интервала.

Интервалы:

$$J_i = [x_{(1)} + (i-1)\triangle;\, x_{(1)} + i\triangle),\ i = \overline{1;\, m-1}$$
$$J_m = [x_{(1)} + (m-i)\triangle;\, \underbrace{x_{(1)} + m\triangle}_{x_{(n)}}]$$

Обратите внимание на границы интервалов.

**Определение.** *Интервальным статистическим рядом называют таблицу*

$$\begin{array}{ccccc} J_1 & \ldots & J_i & \ldots & J_m \\ n_1 & \ldots & n_i & \ldots & n_m \end{array}$$

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

#### 3.1.2.5 Эмпирическая плотность и гистограмма

**Определение.** *Эмпирической плотностью, отвечающей выборке $\vec{x}$, называют функцию*

$$f_n(x) = \begin{cases} \frac{n_i}{n\triangle}, & x \in J_i \\ 0 & \text{иначе} \end{cases}$$

*Здесь $\triangle$, $J_i$ выбираются из интервального ряда.*

**Замечание.**

$$\int\limits_{-\infty}^{+\infty} f_n(x)\,dx = \sum_{i=1}^{m} (\text{площадей прямоугольников}) = \sum_{i=1}^{m} \frac{n_i}{n \cdot \cancel{\triangle}} \cdot \cancel{\triangle} = \frac{1}{n} \cancelto{n}{\sum_{i=1}^{m} n_i} = 1$$

*Эмпирическая функция плотности обладает всеми свойствами «обычной» функции плотности и по этой причине является её статистическим аналогом.*

*Аналогично случаю выборочной функции распределения можно показать, что при $n \to \infty$*

$$f_n(x) \to f(x)$$

*где $f$ — функция плотности случайной величины $X$ (если $X$ — непрерывная).*

**Определение.** *График эмпирической функции плотности называется гистограммой.*

#### 3.1.2.6 Полигон частот

Предположим, что для данной выборки $\vec{x}$ построена гистограмма.

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

# Лекция №20, 04.03.2019

## 3.2 Основные распределения, используемые в математической статистике

### 3.2.1 Гамма-функция Эйлера

**Определение.** *Гамма-функцией Эйлера называется функция вида*

$$\Gamma(\alpha) = \int\limits_{0}^{+\infty} x^{\alpha-1} e^{-x}\,dx \tag{14}$$

Будем рассматривать $\Gamma$-функцию для вещественных значений $\alpha$.

Интеграл в правой части 14 является несобственным и сходится при $\alpha > 0$.

#### 3.2.1.1 Некоторые свойства гамма-функции

1. $\Gamma(\alpha)$ является бесконечно число раз дифференцируемой функцией при $\alpha \in (0;\,+\infty)$, причём

$$\Gamma^k(\alpha) = \int\limits_0^{+\infty} (\ln x)^k x^{\alpha-1} e^{-x}\,dx$$

2. $$\Gamma(\alpha+1) = \alpha\Gamma(\alpha),\ \alpha > 0$$

3. $$\Gamma(1) = \int\limits_0^{+\infty} x^{1-1}e^{-x}dx = \int\limits_0^{+\infty} e^{-x}dx = 1$$

4. Из свойств 2, 3:

$$\Gamma(n+1) = n!,\ n \in \mathbb{N}_0$$

**Замечание.** *В силу свойств 2 и 4 часто говорят, что $\Gamma$-функция является обобщением понятия факториала на случай вещественных значений аргумента.*

5. $$\Gamma(\tfrac{1}{2}) = \int\limits_0^{+\infty} x^{\frac{1}{2}-1}e^{-x}\,dx = \int\limits_0^{+\infty} \frac{1}{\sqrt{x}}e^{-x}\,dx = 2\int\limits_0^{+\infty} e^{-x}d\sqrt{x} = \langle \sqrt{x} = t\rangle =$$
$$= 2\int\limits_0^{+\infty} e^{-t^2}\,dt = \sqrt{\pi}$$

6. $$\Gamma(n+\tfrac{1}{2}) = \langle\text{свойство 2}\rangle = (n-\tfrac{1}{2})\cdot\Gamma(n-\tfrac{1}{2}) = \langle\text{свойство 2}\rangle =$$
$$= (n-\tfrac{1}{2})\cdot(n-\tfrac{3}{2})\cdot\Gamma(n-\tfrac{3}{2}) =$$
$$= (n-\tfrac{1}{2})\cdot(n-\tfrac{3}{2})\cdot\ldots\cdot\tfrac{1}{2}\cdot\overset{\sqrt{\pi}}{\cancel{\Gamma(\tfrac{1}{2})}} = \tfrac{(n-1)\cdot(n-3)\cdot\ldots\cdot3\cdot1}{2^n}\sqrt{\pi},\ n \in \mathbb{N}_0$$

7. Примерный график функции $\Gamma(x)$...

   ...находится вот здесь. Смотрите только значения $x > 0$.

### 3.2.2 Гамма-распределение

**Определение.** *Говорят, что случайная величина $\xi$ имеет гамма-распределение с параметрами $\lambda$ и $\alpha$, если её функция плотности имеет вид*

$$f_\xi(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0 & \textit{иначе} \end{cases}$$

Обозначается как $\xi \sim \Gamma(\lambda,\, \alpha)$.

**Замечание.** *Экспоненциальное распределение*

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0 & \textit{иначе} \end{cases}$$

*является частным случаем гамма-распределения для значения $\alpha = 1$, т. е. $Exp(\lambda) = \Gamma(\lambda,\, 1)$.*

**Теорема.** *Пусть*

1. $\xi_1 \sim \Gamma(\lambda,\, \alpha_1)$;
2. $\xi_2 \sim \Gamma(\lambda,\, \alpha_2)$;
3. $\xi_1,\, \xi_2$ — *независимые.*

   *Тогда*

$$\xi_1 + \xi_2 \sim \Gamma(\lambda,\, \alpha_1 + \alpha_2)$$

**Следствие.** *Если случайные величины $\xi_i \sim \Gamma(\lambda,\, \alpha_i),\ i = \overline{1;\, n}$, то*

$$\xi_1 + \ldots + \xi_n \sim \Gamma(\lambda,\, \alpha_1 + \ldots + \alpha_2)$$

### 3.2.3 Распределение Рэлея

Пусть

1. $\xi \sim N(0,\, \sigma^2)$ — нормальная случайная величина с нулевым математическим ожиданием;
2. $\eta = \xi^2$. ([42])

**Определение.** *В этом случае говорят, что $\eta$ имеет распределение Рэлея с параметром $\sigma$. Можно показать, что в этом случае*

$$f_\eta(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}\sigma} e^{-\frac{x}{2\sigma^2}}, & x > 0 \\ 0 & \textit{иначе} \end{cases}$$

**Замечание.** *Распределение Рэлея является частным случаем гамма-распределения для $\lambda = \frac{1}{2\sigma^2}$, $\alpha = \frac{1}{2}$, т. е. $\eta \sim \Gamma(\frac{1}{2\sigma^2},\, \frac{1}{2})$.*

[42] $\eta$ — строчная буква «э́та» греческого алфавита. — Прим. ред.

### 3.2.4 Распределение хи-квадрат

Пусть

1. $\xi_1, \ldots, \xi_n$ — независимые случайные величины;
2. $\xi_i \sim N(0,\, 1), i = \overline{1;\, n}$;
3. $\eta = \xi_1^2 + \ldots + \xi_n^2$.

**Определение.** *В этом случае говорят, что случайная величина $\eta$ имеет распределение хи-квадрат с $n$ степенями свободы.*

Обозначение $\eta \sim \chi^2(n)$. ([43])

**Замечание 1.** *Если $\xi_i \sim N(0,\, 1)$, то $\xi_i^2 \sim \Gamma(\frac{1}{2},\, \frac{1}{2})$. Тогда по свойству $\Gamma$-распределения $\eta = \xi_1^2 + \ldots + \xi_n^2 \sim \Gamma(\frac{1}{2},\, \frac{n}{2})$.*

**Замечание 2.** *Если случайные величины $\eta_1, \ldots, \eta_m$ независимы и $\eta_i \sim \chi^2(n_i)$, $i = \overline{1;\, m}$, то $\eta_1 + \ldots + \eta_m \sim \Gamma(\frac{1}{2};\, \frac{n_1 + \ldots + n_m}{2}) = \chi^2(n_1 + \ldots + n_m)$.*

**Замечание 3.** *Пусть $\eta \sim \chi^2(n)$. Тогда график функции плотности случайной величины $\eta$ выглядит примерно так (см. случай для $n \gg 4$).*

### 3.2.5 Распределение Фи́шера

Пусть

1. $\xi_1$ и $\xi_2$ — независимы;
2. $\xi_i \sim \chi^2(n_i),\ i = \overline{1;\, 2}$;
3. $\zeta = \frac{n_2 \xi_1}{n_1 \xi_2}$. ([44])

**Определение.** *В этом случае говорят, что случайная величина $\zeta$ имеет распределение Фи́шера со степенями свободы $n_1$ и $n_2$.*

Обозначение $\zeta \sim F(n_1,\, n_2)$.

**Замечание 1.** *Если $\zeta \sim F(n_1,\, n_2)$, то $\frac{1}{\zeta} \sim F(n_2,\, n_1)$.*

**Замечание 2.** *Можно показать, что*

$$f_\zeta(x) = \begin{cases} c \cdot \frac{x^{\frac{n_1}{2} - 1}}{(1 + \frac{n_1}{n_2} x)^{\frac{n_1 + n_2}{2}}}, & x > 0 \\ 0 & \text{иначе} \end{cases}$$

[43] $\chi$ — строчная буква «хи» греческого алфавита. — Прим. ред.

[44] $\zeta$ — строчная буква «дзета» греческого алфавита. — Прим. ред.

*где*

$$c = \frac{\left(\frac{n_1}{n_2}\right)^{\left(\frac{n_1}{2}\right)}}{B\left(\frac{n_1}{2},\, \frac{n_2}{2}\right)}$$

*где*

$$B(p,\, q) = \int\limits_0^1 x^{p-1}(1-x)^{q-1}\, dx$$

*— бета-функция Эйлера.*

### 3.2.6 Распределение Стью́дента

Пусть

1. $\xi_1,\, \xi_2$ — независимые случайные величины;
2. $\xi_1 \sim N(0,\, 1),\ \xi_2 \sim \chi^2(n)$;
3. $\zeta = \frac{\xi_1}{\sqrt{\xi_2}}\sqrt{n}$.

**Определение.** *В этом случае говорят, что случайная величина $\zeta$ имеет распределение Стью́дента с $n$ степенями свободы.*

Обозначение $\zeta \sim St(n-1)$.

**Замечание 1.** *Можно показать, что*

$$f_\zeta(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}},\ x \in \mathbb{R}$$

**Замечание 2.** *Если $\zeta_1 \sim St(n_1)$, $\zeta_2 \sim St(n_2)$, где $n_1 < n_2$, то графики имеют вид...*
*...этот.*

## 3.3 Точечные оценки

### 3.3.1 Основные понятия

**Задача.** ***Вторая задача математической статистики.***
*Пусть $X$ — случайная величина, общий вид закона распределения которой известен, но неизвестны значения одного или нескольких параметров этого закона. Требуется найти (оценить) значения этих параметров.*

Для простоты будем считать, что функция распределения случайной величины $X$ зависит от одного неизвестного параметра $\theta \in \mathbb{R}$, т. е. $F(x,\, \theta)$ — функция распределения случайной величины $X$.

**Определение.** *Статистику $\hat{\theta}(\vec{X})$ называют точечной оценкой параметра $\theta$, если её выборочное значение принимается в качестве значения параметра $\theta$, т. е.*

$$\theta := \hat{\theta}(\vec{x})$$

**Пример.** *Пусть $X$ — случайная величина, $m = MX$ и оно неизвестно; рассмотрим следующие точечные оценки для $m$:*

$$\hat{m}_1(\vec{X}) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\hat{m}_2(\vec{X}) = \tfrac{1}{2}[X_{(1)} + X_{(n)}]$$

$$\hat{m}_3(\vec{X}) = \begin{cases} X_{(\frac{n+1}{2})}, & \text{если } n \text{ — нечётное,} \\ \frac{1}{2}[X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}] & \text{иначе} \end{cases}$$

*($\hat{m}_3(\vec{X})$ — средний член вариационного ряда)*

$$\hat{m}_4(\vec{X}) = X_1$$

*($\hat{m}_4(\vec{X})$ — результат первого наблюдения)*

$$\hat{m}_5(\vec{X}) = e^{(\overline{X})^2} \cdot \sin(X_1^3 + 2 \cdot X_2^3 + \ldots + n \cdot X_n^3)$$

*Понятно, что некоторые из этих оценок более удачны, некоторые менее, некоторые абсурдны. Для сравнения качества точечных оценок используются следующие их свойства:*

1. *Несмещённость;*
2. *Состоятельность;*
3. *Эффективность.*

# Лекция №21, 11.03.2019

**Замечание.** *Будет ли некоторая статистика $\hat{\theta}(\vec{X})$ служить точечной оценкой для параметра $\theta$, зависит от нашего желания: если мы используем $\hat{\theta}(\vec{x})$ в качестве значения параметра $\theta$, то эта статистика будет являться точечной оценкой для $\theta$. В противном случае — не будет.*

### 3.3.2 Несмещённость точечной оценки

Пусть

1. $F(x, \theta)$ — функция распределения случайной величины $X$ с неизвестным параметром $\theta$;

2. $\hat{\theta}(\vec{X})$ — точечная оценка для $\theta$.

**Определение.** *Точечная оценка $\hat{\theta}(\vec{X})$ называется несмещённой оценкой для $\theta$, если $\exists M[\hat{\theta}(\vec{X})] = \theta$.*

**Пример 1.** *Рассмотрим оценки математического ожидания точечных оценок из предыдущего примера:*

$$M[\hat{m}_1(\vec{X})] = M[\tfrac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} MX_i = \frac{1}{n}\sum_{i=1}^{n} m = m$$

*т. е. $\hat{m}_1$ — несмещённая оценка для $m$.*

$$M[\hat{m}_4(\vec{X})] = M[X_1] = \langle X_1 \sim X \rangle = m$$

*т. е. $\hat{m}_4$ — несмещённая оценка для $m$.*

**Пример 2.** *Пусть $X$ — случайная величина, также $\exists DX = \sigma^2$ и она неизвестна. Можно показать, что выборочная дисперсия $\hat{\sigma}^2(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ является смещённой оценкой дисперсии: $M[\hat{\sigma}^2(\vec{X})] = \frac{n-1}{n}\sigma^2$ (доказано на семинаре).*

По этой причине рассмотрим статистику

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

для которой

$$M[S^2(\vec{X})] = \frac{n}{n-1}M[\hat{\sigma}^2] = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2$$

Таким образом, $S^2(\vec{X})$ является несмещённой оценкой дисперсии.

**Определение.** *Статистика $S^2(\vec{X})$ называется исправленной выборочной дисперсией.*

### 3.3.3 Состоятельность точечной оценки

Пусть $X$ — случайная величина, $F(x, \theta)$ — функция распределения случайной величины $X$, $\theta$ — неизвестный параметр, $\hat{\theta}(\vec{X})$ — точечная оценка для $\theta$.

**Определение.** *Оценка $\hat{\theta}(\vec{X})$ называется состоятельной оценкой параметра $\theta$, если*

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

*где $n$ — объём выборки $\vec{X}$.*

**Замечание 1.** *Условие из определения точечной оценки можно записать в виде: $\forall\epsilon > 0 \lim\limits_{n\to\infty} P\{|\hat{\theta}(\vec{X}) - \theta| - \epsilon\} = 1$*

**Замечание 2.** *Если оценка $\hat{\theta}(\vec{X})$ не является состоятельной, то она никому не нужна.*

**Пример.** *Пусть*

1. *$X$ — случайная величина;*
2. *$\exists MX = m$;*
3. *$\exists DX = \sigma^2$.*

*Покажем, что выборочное среднее $\overline{X}$ является состоятельной оценкой для $m$.*

1. *Последовательность $X_1$, $X_2$, …, $X_n$, … является последовательностью независимых одинаково распределённых случайных величин;*
2. *$\exists MX_i = m$, $\exists DX_i = \sigma^2$, $i \in \mathbb{N}$;*
3. *Из 1, 2 следует, что последовательность $X_1$, $X_2$ и т. д. удовлетворяет закону больших чисел в форме Чебышева, из чего следует*

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

*т. е. $\overline{X} \xrightarrow[n\to\infty]{P} m$.*

**Замечание 1.** *Состоятельность оценки $\overline{X}$ для $m$ можно доказать и не предполагая существование конечной дисперсии.*

**Замечание 2.** *Можно показать, что статистика $S^2(\vec{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ является состоятельной оценкой дисперсии случайной величины $X$, если эта случайная величина имеет момент до четвёртого порядка включительно.*

**Замечание 3.** *Более общо: если случайная величина $X$ имеет начальные и центральные моменты, то их выборочные аналоги являются их состоятельными оценками. При этом все выборочные моменты порядка два и выше являются смещёнными оценками соответствующих теоретических моментов.*

**Пример.** *Пусть*

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

*Покажем, что $\hat{m}$ не является состоятельной оценкой. Зафиксируем $\epsilon > 0$:*

$$P\{|\hat{m}(\vec{X}) - m| < \epsilon\} = P\{|X_1 - m| < \epsilon\} =$$
$$= P\{m - \epsilon < X_1 < m + \epsilon\} = \langle X_1 \sim X \sim N(m, \sigma^2)\rangle =$$
$$= \Phi_0(\tfrac{m+\epsilon-m}{\sigma}) - \Phi_0(\tfrac{m-\epsilon-m}{\sigma}) = 2\Phi_0(\tfrac{\epsilon}{\sigma}) \neq 1, \text{ если } \frac{\epsilon}{\sigma} \neq +\infty$$

*Поэтому $P\{|\hat{m}(\vec{X}) - m| < \epsilon\} \cancel{\xrightarrow[n\to\infty]{}} 1$.*

### 3.3.4 Эффективность точечной оценки

1. Дисперсия характеризует разброс значений случайной величины относительно её математического ожидания: чем больше дисперсия, тем больше разброс.

   Если $Y$, $Z$ — две случайные величины, $MY = MZ = m$, но если $DY < DZ$, то качественно эту ситуацию можно проиллюстрировать следующим образом:

   *<Здесь был рисунок, но в нём не было никакого откровения.>*

2. Пусть $\hat{\theta}(\vec{X})$ — несмещённая оценка параметра $\theta$. $\vec{X}$ — случайный вектор, $\hat{\theta}(\vec{X})$ — скалярная функция случайного вектора, т. е. $\hat{\theta}(\vec{X})$ является случайной величиной. Возможно, что значения $\hat{\theta}(\vec{X})$ кучно группируются около математического ожидания. Возможно, что менее кучно. С точки зрения качества точечной оценки первая ситуация предпочтительнее.

Пусть $\hat{\theta}_1(\vec{X})$, $\hat{\theta}_2(\vec{X})$ — несмещённые оценки для параметра $\theta$.

**Определение.** *Оценка $\hat{\theta}_1$ называется более эффективной, чем оценка $\hat{\theta}_2$, если*

1. *$\exists D\hat{\theta}_1$, $\exists D\hat{\theta}_2$.*
2. *$D\hat{\theta}_1 < D\hat{\theta}_2$.*

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

1. *$\hat{\theta}$ является несмещённой оценкой для $\theta$;*

2. *$\hat{\theta}$ обладает наименьшей дисперсией среди всех несмещённых оценок для $\theta$.*

**Замечание.** *Иногда говорят об эффективной оценке в классе оценок $\Theta$. Если $\Theta$ — некоторое множество несмещённых оценок для параметра $\theta$, то оценка $\hat{\theta} \in \Theta$ называется эффективной оценкой для $\theta$ в классе $\Theta$, если $\hat{\theta}$ обладает наименьшей дисперсией среди всех оценок класса $\Theta$, т. е. $\forall \tilde{\theta} \in \Theta \ D\hat{\theta} \leqslant \tilde{\theta}$.*

**Пример.** *Пусть*

1. *$X$ — случайная величина;*
2. *$\exists MX = m$, $\exists DX = \sigma^2$.*

*Покажем, что выборочное среднее $\overline{X}$ является эффективной оценкой для $m$* в классе линейных оценок.

1. *Линейная оценка имеет вид:*

$$\hat{m}(\vec{X}) = \lambda_1 X_1 + \ldots + \lambda_n X_n$$

*где $\lambda_j \in \mathbb{R}$, $j = \overline{1;\, n}$.*

2. *Т. к. оценка должна быть несмещённой, то*

$$M[\hat{m}(\vec{X})] = M\left[\sum_{i=1}^{n} \lambda_i X_i\right] = \sum_{i=1}^{n} \lambda_i M X_i = \langle X_i \sim X,\ i = \overline{1;\, n} \rangle =$$
$$= \sum_{i=1}^{n} \lambda_i m = m \sum_{i=1}^{n} \lambda_i$$

*Требуем: $M[\hat{m}(\vec{X})] = m \implies \sum\limits_{i=1}^{n} \lambda_i = 1$.*

3. *Подберём в линейной оценке $\hat{m}(\vec{X}) = \sum\limits_{i=1}^{n} \lambda_i X_i$ коэффициенты $\lambda_i$ таким образом, чтобы $D[\hat{m}(\vec{X})]$ была линейной среди значений дисперсии всевозможных линейных оценок.*

$$D[\hat{m}] = D[\sum_{i=1}^{n} \lambda_i X_i] = \langle X_i \text{ — независимы} \rangle = \sum_{i=1}^{n} \lambda_i^2 D X_i = \sigma^2 \sum_{i=1}^{n} \lambda_i^2$$

*Таким образом, приходим к значению поиска условия экстремума:*

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

**Замечание.** *Если бы экстремум был безусловный, то необходимые условия экстремума:*

$$\begin{cases} f(x_1,\, \ldots,\, x_n) \to min \\ \frac{\delta f}{\delta x_i} = 0,\ i = \overline{1;\, n} \end{cases}$$

*Если условный:*

$$\begin{cases} f(x_1,\, \ldots,\, x_n) \to min \\ \varphi(x_1,\, \ldots,\, x_n) = 0 \end{cases}$$

*Составим функцию Лагранжа:*

$$L(\lambda_1,\, \ldots,\, \lambda_n,\, \mu) = f(\lambda_1,\, \ldots,\, \lambda_n) - \mu(\sum_{i=1}^{n} \lambda_i - 1) = (\lambda_1^2 + \ldots + \lambda_n^2) - \mu(\lambda_1 + \ldots + \lambda_{n-1})$$

*Необходимые условия экстремума:*

$$\begin{cases} \frac{\delta L}{\delta \lambda_1} = 2\lambda_1 - \mu = 0 \\ \ldots \\ \frac{\delta L}{\delta \lambda_n} = 2\lambda_n - \mu = 0 \\ \frac{\delta L}{\delta \mu} = -(\sum\limits_{i=1}^{n} \lambda_i - 1) = 0 \end{cases}$$

*Из этих условий (кроме последнего) следует* $\lambda_i = \frac{\mu}{2},\ i = \overline{1;\, n}$*.*

*Используем последнее условие:*

$$\sum_{i=1}^{n} \underbrace{\frac{\mu}{2}}\lambda_i = 1$$

$\frac{\mu n}{2} = 1,\ \mu = \frac{2}{n}$.

*Тогда* $\lambda_i = \frac{1}{n},\ i = \overline{1;\, n}$*.*

*Можно, проверив достаточное условие экстремума, показать, что точка* $(\frac{1}{n},\, \ldots,\, \frac{1}{n})$ *является условием минимума функции* $f(\lambda_1,\, \ldots,\, \lambda_n)$*. 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:*

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

**Теорема.** ***О единственности эффективной оценки.***
*Пусть $\hat{\theta}_1(\vec{X})$ и $\hat{\theta}_2(\vec{X})$ — две эффективных оценки для параметра $\theta$. Тогда*

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

Это равенство является равенством для двух случайных величин $\hat{\theta}_1$ и $\hat{\theta}_2$, поэтому оно понимается в вероятностном смысле:

Без док-ва

$$P\{\vec{X} \in \{\vec{x} \colon \hat{\theta}_1(\vec{x}) \neq \hat{\theta}_2(\vec{x})\}\} = 0$$

~~**Доказательство.**~~

1. *Рассмотрим оценку $\hat{\theta} = \frac{1}{2}[\hat{\theta}_1 + \hat{\theta}_2]$.*

$$M\hat{\theta} = M[\tfrac{1}{2}(\hat{\theta}_1 + \hat{\theta}_2)] = \tfrac{1}{2}[M\hat{\theta}_1 + M\hat{\theta}_2] =$$
$$= \langle \hat{\theta}_1 \text{ и } \hat{\theta}_2 \text{ — эффективные и, следовательно, несмещённые } \rangle = \tfrac{1}{2}[\theta + \theta] = \theta$$

*т. е. $\hat{\theta}$ также является несмещённой оценкой для $\theta$.*

$$D\hat{\theta} = \tfrac{1}{4}D[\hat{\theta}_1 + \hat{\theta}_2] = \tfrac{1}{4}[D\hat{\theta}_1 + D\hat{\theta}_2 + 2\operatorname{cov}(\hat{\theta}_1, \hat{\theta}_2)] =$$

*$\langle$ Обозначим: $D\hat{\theta}_1 = a^2 = D\hat{\theta}_2$ (из эффективности оценки). Т. к. $\hat{\theta}_1$ и $\hat{\theta}_2$ — эффективные, то они имеют одинаковую дисперсию, которая является наименьшей среди дисперсий всех несмещённых оценок для $\theta$ $\rangle$*

$$= \tfrac{1}{2}[a^2 + \operatorname{cov}(\hat{\theta}_1, \hat{\theta}_2)] \tag{15}$$

2. 

$$|\operatorname{cov}(\hat{\theta}_1, \hat{\theta}_2)| \leqslant \sqrt{\underbrace{D\hat{\theta}_1}_{=a^2} \cdot \underbrace{D\hat{\theta}_2}_{=a^2}} = a^2$$

*Таким образом,*

$$D\hat{\theta} \leqslant \langle \text{ см. 15 } \rangle \leqslant \tfrac{1}{2}[a^2 + a^2] = a^2 \tag{16}$$

*$\hat{\theta}$ — несмещённая оценка для $\theta$, а $\hat{\theta}_1$, $\hat{\theta}_2$ — эффективные оценки $\Longrightarrow$ $D\hat{\theta}_1 = D\hat{\theta}_2 = a^2 \leqslant D\hat{\theta}$.*

*С учётом 16: $D\theta = a^2$.*

3. *15 $\Longrightarrow$*

$$\underbrace{a^2}_{D\hat{\theta}} = \tfrac{1}{2}[a^2 + \operatorname{cov}(\hat{\theta}_1, \hat{\theta}_2)] \Longrightarrow \operatorname{cov}(\hat{\theta}_1, \hat{\theta}_2) = a^2$$

*т. е.*

$$\operatorname{cov}(\hat{\theta}_1,\, \hat{\theta}_2) = \sqrt{D\hat{\theta}_1 \cdot D\hat{\theta}_2}$$

*Следовательно, по свойствам ковариации, $\hat{\theta}_1$ и $\hat{\theta}_2$ связаны положительной линейной зависимостью, т. е. ($k > 0$)*

$$\hat{\theta}_1 = k\hat{\theta}_2 + b \tag{17}$$

*4. 17 $\implies$*

$$\underbrace{D\hat{\theta}_1}_{=a^2} = k^2 \underbrace{D\hat{\theta}_2}_{=a^2} \implies k^2 = 1 \implies k = 1$$

*Тогда $\hat{\theta}_1 = \hat{\theta}_2 + b \implies \underbrace{M\hat{\theta}_1}_{=\theta} = \underbrace{M\hat{\theta}_2}_{=\theta} + b \implies b = 0$.*

*Таким образом, 17 $\implies \hat{\theta}_1 = \hat{\theta}_2$.*

# Лекция №22, 18.03.2019

**Вопрос.** *Из предыдущих рассуждений следует, чем меньше дисперсия точечной оценки, тем более эффективной будет её практическое использование. Можно ли сделать дисперсию несмещённой точечной оценки неизвестного параметра некоторого закона распределения сколь угодно малой?*

Оказывается, значения дисперсии всех несмещённых точечных оценок данного параметра (при фиксированном объёме выборки) ограничены снизу (см. неравенство Ра́о-Краме́ра).

Пусть

1. $X$ — непрерывная случайная величина;
2. $f(t, \theta)$ — функция плотности распределения вероятностей случайной величины $X$;
3. $\vec{X} = (X_1,\, \ldots,\, X_n)$ — случайная выборка из генеральной совокупности $X$.

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

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

(поскольку компоненты случайного вектора независимы)

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

**Определение.** *Величина*

$$I(\theta) = M\left\{\left[\frac{\delta \ln f_{\vec{X}}(\vec{T}, \theta)}{\delta\theta}\right]^2\right\}$$

*называется количеством информации по Фишеру (в серии из n наблюдений).*

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

$$\frac{\delta}{\delta\theta}\int \varphi(\vec{T},\, \theta)\, d\vec{T} = \int \frac{\delta\varphi(\vec{T},\, \theta)}{\delta\theta}\, d\vec{T}$$

*Параметрические модели, для которых справедлив данный переход будем называть регулярными.*

**Теорема.** ***Неравнество Ра́о-Краме́ра.*** *Пусть*

1. *Рассматривается регулярная модель;*
2. $\hat{\theta}(\vec{X})$ *— несмещённая точечная оценка параметра $\theta$ закона распределения случайной величины $X$.*

*Тогда*

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

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

**Доказательство.**

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

$$\int\limits_{\mathbb{R}^n} f_{\vec{X}}(\vec{T},\, \theta)\, d\vec{T} = \int\limits_{G^n} f_{\vec{X}}(\vec{T},\, \theta)\, d\vec{T} = 1 \text{ (по условию нормировки)}$$

2. *Продифференцируем подчёркнутое равенство по $\theta$: Правая часть: $\frac{\delta 1}{\delta\theta} = 0$; левая часть*

$$\frac{\delta}{\delta\theta}\int\limits_{G^n} f_{\vec{X}}(\vec{T},\, \theta)\, d\vec{T} = \langle \text{модель явл. рес.?} \rangle =$$

$$= \int\limits_{G^n} \frac{\delta f_{\vec{T}}(\vec{T},\, \theta)}{\delta\theta}\, dT = \langle \frac{\delta \ln y}{\delta\theta} = \frac{1}{y}\cdot\frac{\delta y}{\delta\theta} \implies \frac{\delta y}{\delta\theta} = y\frac{\delta \ln y}{\delta\theta}\rangle =$$

$$= \int\limits_{G^n} \frac{\delta \ln \overbrace{f_{\vec{X}}(\vec{T},\, \theta)}^{y}}{\delta\theta}\, \underbrace{f_X(\vec{T},\, \theta)}_{y}\, d\vec{T} = \langle M[\varphi(\vec{X})] = \int\limits_{R^n} \varphi(\vec{X}) f(\vec{X}) d\vec{X} \rangle =$$

$$= M\left[\frac{\delta \ln f_X(\vec{X},\, \theta)}{\delta\theta}\right]$$

*Таким образом,*

$$M\left[\frac{\delta \ln f_X(\vec{X},\, \theta)}{\delta\theta}\right] = 0 \tag{18}$$

3. *Т. к.* $\hat{\theta}(\vec{X})$ *— несмещённая оценка для* $\theta$*, то*

$$\theta = M[\hat{\theta}(\vec{X})] = \langle M[\varphi(\vec{X})] = \int\limits_{\mathbb{R}^n} \varpi(\vec{X}) f_{\vec{X}}(\vec{x})\, d\vec{x} \rangle = \int\limits_{G^n} \hat{\theta}(\vec{T}) \cdot f_{\vec{X}}(\vec{T},\, \vec{\theta})\, d\vec{T}$$

*Продифференцируем полученное равенство по* $\theta$*: для левой части* $\frac{\delta\theta}{\delta\theta} = 1$*; правая часть:*

$$\frac{\delta}{\delta\theta} \int\limits_{G^n} \hat{\theta}(\vec{T})\, f(\vec{T},\, \theta)\, d\vec{T} = \langle \text{ модель регулярная } \rangle = \int\limits_{G^n} \hat{\theta}(\vec{T}) \frac{\delta f_{\vec{X}}(\vec{T},\, \theta)}{\delta\theta}\, d\vec{T} =$$

$$= \left\langle \frac{\delta y}{\delta\theta} = y \frac{\delta \ln y}{\delta\theta} \right\rangle = \int\limits_{G^n} \hat{\theta}(\vec{T}) \frac{\delta \ln f_{\vec{X}}(\vec{T},\, \theta)}{\delta\theta} \cdot f_{\vec{X}}(\vec{T},\, \theta)\, d\vec{T} = M\left[\hat{\theta}(\vec{X}) \cdot \frac{\delta \ln f_{\vec{X}}(\vec{X},\, \theta)}{\delta\theta}\right]$$

*Таким образом,*

$$M[\hat{\theta}(\vec{X}) \cdot \frac{\delta \ln f_{\vec{X}}(\vec{X},\, \theta)}{\delta\theta}] = 1 \tag{19}$$

4. *Умножим обе части 18 на* $\theta$*:*

$$M[\frac{\delta \ln f_{\vec{X}}(\vec{X},\, \theta)}{\delta\theta} \cdot \theta] = 0 \tag{20}$$

*Вычтем из 19 равенство 20:*

$$M[\frac{\delta \ln f_{\vec{X}}(\vec{X},\, \theta)}{\delta\theta}(\hat{\theta}(\vec{X}) - \theta)] = 1$$

*Возведём обе части этого равенства в квадрат:*

$$1 = \left\{ M[\frac{\delta \ln f_{\vec{X}}(\vec{X},\, \theta)}{\delta\theta}(\hat{\theta}(\vec{X}) - \theta)] \right\}^2 =$$

$$= \left\{ \int\limits_{G^n} \underbrace{\frac{\delta \ln f_{\vec{X}}(\vec{T},\, \theta)}{\delta\theta}}_{a(\vec{T})} \cdot \underbrace{(\hat{\theta}(\vec{T}) - \theta)}_{b(\vec{T})} \cdot \underbrace{f_X(\vec{T},\, \theta)}_{\textit{(вес)}}\, d\vec{T} \right\}^2 =$$

$= \Big\langle$ *Неравенство Коши-Буняковского:* $|(a,\, b)| \leqslant |a| \cdot |b|,$ *где* $|x| = \sqrt{(x,\, x)}$*;*

*В пространстве* $L^2(G^n)$ *скалярное произведение может быть задано в виде:*

$$(a,\, b) = \int\limits_{G^n} a(\vec{T}) b(\vec{T}) p(\vec{T}) d\vec{T}$$

*где* $p(\vec{T}) > 0$ *в* $G^n$*, где* $p(\vec{T})$ *— весовая функция* $\Big\rangle =$

$$= \{(a(\vec{T}),\, b(\vec{T})\}^2 \leqslant \langle \textit{неравенство Коши-Буняковского} \rangle \leqslant$$

$$\leqslant \underbrace{(a(\vec{T}),\, a(\vec{T}))}_{|a|^2} \cdot \underbrace{(b(\vec{T}),\, b(\vec{T}))}_{|b|^2} =$$

$$= \underbrace{\int\limits_{G^n} [\frac{\delta \ln f_{\vec{X}}(\vec{T},\, \theta)}{\delta\theta}]^2 f_{\vec{X}}(\vec{T},\, \theta) d\vec{T}}_{|a|^2} \cdot \underbrace{\int\limits_{G^n} (\hat{\theta}(\vec{T}) - \theta)^2 f_{\vec{X}}(\vec{T},\, \theta) d\vec{T}}_{|b|^2} =$$

$$= M[\frac{\delta \ln f_{\vec{X}}(\vec{X},\, \theta)}{\delta\theta}]^2] \cdot M[(\hat{\theta}(\vec{T}) - \overset{M[\hat{\theta}]}{\theta})^2] = I(\theta) \cdot D[\hat{\theta}]$$

*Таким образом,*

$$1 \leqslant I(\theta) D(\hat{\theta}) \implies D(\hat{\theta}) \geqslant \frac{1}{I(\theta)}$$

**Замечание 1.** *Неравенство Рао-Крамера даёт нижнюю границу для значений дисперсии всевозможных точечных оценок параметра* $\theta$*.*

**Замечание 2.** *Величина*

$$e(\hat{\theta}) = \frac{1}{I(\theta) \cdot D(\hat{\theta})}$$

*называется показателем эффективности по Рао-Крамеру оценки* $\hat{\theta}$*. Из неравенства Рао-Крамера следует, что*

$$0 < e(\hat{\theta}) \leqslant 1$$

*При этом*

$$e(\hat{\theta}) = 1 \iff D\hat{\theta} = \frac{1}{I(\theta)}$$

*т. е. $D\hat{\theta}$ имеет наименьшее возможное значение. В этом случае оценка $\hat{\theta}$ называется эффективной по Рао-Крамеру.*

Вообще говоря, никто не обещал, что для неизвестного параметра $\theta$ каждого из возможных законов распределения генеральной совокупности найдётся несмещённая оценка $\hat{\theta}$, для которой неравенство Рао-Крамера будет выполняться в форме равенства (т. е. никто не обещал, что нижняя графница значения дисперсии всегда достижима). Если для некоторой оценки $\hat{\theta}$ верно $D(\hat{\theta}) = \frac{1}{I(\theta)}$, то эта оценка будет эффективной по Рао-Крамеру и, следовательно, «просто» эффективной.

**Замечание 3.** *В каких случаях неравенство Рао-Крамера выполняется в форме равенства? Из курса линейной алгебры известно, что неравенство К.-Б.? выполняется в форме равенства ТТТ?, когда сомножители $a$, $b$ линейно зависимы, т. е.*

$$e(\hat{\theta}) \iff \langle \textit{см. доказательство Рао-Крамера} \rangle \iff a(\vec{T}) =$$

$$= k \cdot b(\vec{T}) \iff \underbrace{\frac{\delta \ln f_{\vec{X}}(\vec{X}, \theta)}{\delta\theta} = k(\theta) \cdot (\hat{\theta}(\vec{X}) - \theta)}_{\textit{критерий сущ-ния эфф. по Рао-Крамеру}} \tag{21}$$

*При выполнении этого условия оценка $\hat{\theta}$ является эффективной (по Рао-Крамеру) и $k(\theta) = \frac{1}{D(\theta)} = I(\theta)$. Покажем, что $k(\theta) = \frac{1}{D(\theta)}$: Из 21:*

$$[\frac{\delta \ln f_{\vec{X}}(\vec{X}, \theta)}{\delta\theta}]^2 = k^2(\theta) \cdot (\hat{\theta}(\vec{X}) - \theta)^2 \implies$$

$$\implies \underbrace{M[\frac{\delta \ln f_{\vec{X}}(\vec{X}, \theta)}{\delta\theta}]^2}_{I(\theta)} = k^2(\theta) \underbrace{M\{[\hat{\theta} - \theta]^2\}}_{D(\hat{\theta})} \implies$$

$$\implies I(\theta) = k^2(\theta) D(\hat{\theta}) \implies \langle \textit{ т. к. } \hat{\theta} \textit{ эффективная по Р.-К., то } I(\theta) = \frac{1}{D\hat{\theta}} \rangle \implies$$

$$\implies k^2(\theta) = \frac{1}{(D(\hat{\theta}))^2} \implies k(\theta) = \frac{1}{\underbrace{D(\hat{\theta})}_{I(\theta)}}$$

**Замечание 4.** *Если $X$ — дискретная случайная величина, $\theta$ — неизвестный параметр закона распределения случайной величины $X$, то для всевозможных несмещённых оценок $\hat{\theta}$ параметра $\theta$ также будет выполняться неравенство Рао-Крамера, где под $I(\theta)$ понимается*

$$I(\theta) = M\{\left(\frac{\delta \ln(P\{X_1 = t_1\} \cdot \ldots \cdot P\{X_n = t_n\})}{\delta\theta}\right)^2\}$$

# Лекция №23, 25.03.2019

### 3.3.5 Методы построения точечных оценок

Мы изучим два метода:

1. Метод моментов;
2. Метод максимального правдоподобия.

#### 3.3.5.1 Метод моментов

Пусть

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

   (требуется оценить значения параметров $\theta_1, \ldots, \theta_r$).

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

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

1. Найти выражения для $r$ первых теоретических моментов случайной величины $X$ (т. к. функция распределения случайной величины $X$ зависит от параметров $\theta_1$, $\ldots$, $\theta_r$, то и теоретические моменты также будут зависеть от этих параметров):

   $m_1(\theta_1, \ldots, \theta_r) = M[X]$

   $\ldots$

   $m_r(\theta_1, \ldots, \theta_r) = M[X^r]$

2. Нужно приравнять выражения для теоретических моментов и их выборочных аналогов:

$$\begin{cases} m_1(\theta_1, \ldots, \theta_r) = \hat{m}_1(\vec{X}) \\ \ldots \\ m_r(\theta_1, \ldots, \theta_r) = \hat{m}_r(\vec{X}) \end{cases} \tag{22}$$

   Это — система из $r$ нелинейных уравнений относительно $r$ неизвестных $\theta_1$, $\ldots$, $\theta_r$.

3. Решаем получившуюся систему относительно $\theta_1, \ldots, \theta_r$:

$$\begin{cases} \theta_1 = \hat{\theta}_1(\vec{X}) \\ \ldots \\ \theta_r = \hat{\theta}_r(\vec{X}) \end{cases}$$

**Замечание 1.** *Некоторые уравнения системы 22 удобно записывать не относительно начальных моментов, а относительно центральных:*

$$\mathring{m}_j(\theta_1, \ldots, \theta_r) = \hat{\nu}_j(\vec{X})$$

*где* $\mathring{m}_j = M[(X - MX)^j]$, $\hat{\nu}_j(\vec{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^j$. *Выбор между начальными моментами для записи уравнения делают исходя из того, что система 22 должна решаться как можно проще.*

**Замечание 2.** *Поэтому выборочные моменты* $\hat{m}_k$ *и* $\hat{\nu}_k$ *являются состоятельными оценками своих теоретических аналогов, то и оценки параметров, полученные с использованием метода моментов, также будут состоятельными (при условии, что решение системы 22 непрерывно зависит от правых частей).*

**Замечание 3.** *Поскольку оценки* $\hat{m}_k$, $\hat{\nu}_k$ *при* $k \geqslant 2$ *являются смещёнными оценками соответствующих моментов, то и оценки параметров, полученных с использованием метода моментов, также могут быть смещёнными.*

**Пример.** $X \sim R[a, b]$ *С использованием метода моментов построить точечные оценки для* $a$ *и* $b$.

**Решение.**

1.

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0 & \text{иначе} \end{cases}$$

$a, b$ — *неизвестные параметры* $\implies r = 2$. *Нужно составить систему из двух уравнения для первого и второго моментов:*

$m_1(a, b) = MX = \frac{a+b}{2}$ — *первый начальный момент;*

$\mathring{m}_2(a, b) = DX = \frac{(b-a)^2}{12}$ — *второй центральный момент;*

2. *Составим систему*

$$\begin{cases} \frac{a+b}{2} = \hat{m}_1(\vec{X}) \\ \frac{(b-a)^2}{12} = \hat{\nu}_2(\vec{X}) \end{cases}$$

$$(\hat{m}_1(\vec{X}) = \overline{X},\ \hat{\nu}_2(\vec{X}) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}_{\hat{\sigma}^2(\vec{X})})$$

3. *Решаем получившуюся систему относительно $a$ и $b$:*

$$\begin{cases} a + b = 2\overline{X} \\ (b-a)^2 = 12\sigma^2 \end{cases}$$

*Приходим к*

$$\begin{cases} a = \overline{X} \mp \sqrt{3}\hat{\sigma} \\ b = \overline{X} \pm \sqrt{3}\hat{\sigma} \end{cases}$$

*Т. к. $a < b$, то*

$$\begin{cases} \hat{a}(\vec{X}) = \overline{X} - \sqrt{3}\hat{\sigma}(\vec{X}) \\ \hat{b}(\vec{X}) = \overline{X} + \sqrt{3}\hat{\sigma}(\vec{X}) \end{cases}$$

#### 3.3.5.2 Метод максимального правдоподобия

Пусть

1. $X$ — случайная величина, закон распределения которой зависит от неизвестных параметров $\hat{\theta} = (\theta_1, \ldots, \theta_r)$;
2. $\vec{X} = (X_1, \ldots, X_n)$ — случайная выборка из генеральной совокупности $X$.

**Определение.** *Функцией правдоподобия случайной выборки $\vec{X}$ называется функция*

$$L(\vec{X}, \theta_1, \ldots, \theta_r) = p(X_1, \vec{\theta}) \cdot \ldots \cdot p(X_n, \vec{\theta})$$

*где $p(x, \vec{\theta}) = P\{X = x\}$, если $X$ — дискретная случайная величина; $p(x, \vec{\theta}) = f(x, \vec{\theta})$, если $X$ — непрерывная случайная величина ($f$ — функция плотности распределения случайной величины $X$).*

**Замечание.** *Очевидно, что если в функцию $L$ подставить удачное значение вектора $\vec{\theta}$ (т. е. подставить значения параметров $\theta_1, \ldots, \theta_n$, которые близки к их теоретическим значениям), то значение функции $L$ будет относительно большим. Если же использовать «неудачные» значения вектора $\vec{\theta}$, то соответствующие значение $L$ будет сравнительно малым.*

В методе максимального правдоподобия в качестве точечной оценки вектора $\vec{\theta}$ неизвестных параметров используют такое значение, которое максимизирует функцию правдоподобия $L$, т. е.

$$\forall \vec{x} \in X_n \; L(\vec{X}, \hat{\vec{\theta}}) \geqslant L(\vec{X}, \vec{\theta}), \; \vec{\theta} \in \Theta$$

Для построения такой оценки необходимо решить задачу

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

Эта задача эквивалентна задаче

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

которая часто решается проще.

В некоторых случаях для нахождения оптимального значения вектора $\vec{\theta}$ можно использовать необходимые условия экстремума ФНП?:

$$\frac{\delta L(\vec{X},\,\vec{\theta})}{\delta\theta} = 0,\ j = \overline{1;\,r} \tag{23}$$

или (для эквивалентной задачи)

$$\frac{\delta \ln L(\vec{X},\,\vec{\theta})}{\delta\theta_j} = 0,\ j = \overline{1;\,r} \tag{24}$$

**Замечание.** *Уравнения 23 (или уравнения 24) называют уравнения правдоподобия.*

**Пример.** $X \sim R[a,\,b]$. *Построить для $a$, $b$ оценки максимального правдоподобия.*

**Решение.**

*1.*

$$f(x,\,a,\,b) = \begin{cases} \frac{1}{b-a}, & x \in [a,\,b] \\ 0 & \textit{иначе} \end{cases}$$

*2. Запишем функцию правдоподобия*

$$L(\vec{X},\vec{\theta}) = f(X_1,\,a,\,b)\cdot\ldots\cdot f(X_n,a,b) = \begin{cases} \frac{1}{(b-a)^n}, & a \leqslant X_{(1)},\,X_{(n)} \leqslant b \\ 0 & \textit{иначе} \end{cases}$$

$$\ln L = \begin{cases} -n\ln(b-a), & a \leqslant X_{(1)},\,X_{(n)} \leqslant B \\ -\infty & \textit{иначе} \end{cases}$$

*Составим уравнения правдоподобия:*

$$\begin{cases} \frac{\delta \ln L}{\delta a} = \frac{n}{n-a} = 0 \\ \frac{\delta \ln L}{\delta b} = \frac{-n}{b-a} = 0 \end{cases}$$

$\Longrightarrow$ $b - a = \infty$? *Поскольку от неизвестных параметров $a$, $b$ зависят границы области, в которой $f(x,a,b) > 0$, то дифференцирование по этим параметрам затруднительно. По этой причине будем максимизировать функцию $L$ «в лоб», не используя условия экстремума.*

*3.*

$$L(\vec{X}, a, b) = \begin{cases} \frac{1}{(b-a)^n}, & x \in [a;\, b] \\ 0 & \textit{иначе} \end{cases}$$

(a) *Если $X_{(1)} < a$ или $X_{(n)} > b$, то $L = 0$. По этой причине $a \leqslant X_{(1)}$, $X_{(n)} \leqslant b$;*

(b) *Чем меньше размер отрезка $[a,\, b]$, т. е. чем меньше $b - a$, тем больше $\frac{1}{b-a}$, следовательно, тем больше $L(x,\, a,\, b) = \frac{1}{(b-a)^n}$. Таким образом, для максимизации функции $L$ нужно брать $a$, $b$ так, чтобы $b - a \to \min$;*

(c) *Из а, б $\implies \hat{a}(\vec{X}) = X_{(1)}$, $\hat{b}(\vec{X}) = X_{(n)}$.*

## 3.4 Интервальные оценки

### 3.4.1 Основные понятия

**Задача.** ***Вторая задача математической статистики.***

*Пусть $X$ — случайная величина, закон которой известен с точностью до вектора $\vec{\theta} = (\theta_1,\, \ldots,\, \theta_r)$ неизвестных параметров (т. е. известен общий вид закона распределения случайной величины $X$, но этот закон создаёт неизвестные параметры $\theta_1,\, \ldots,\, \theta_r$). Требуется: оценить значение $\vec{\theta}$.*

Ранее для решения этой задачи использовались точечные оценки, недостаток которых состоит в том, что они не дают вероятностных характеристик точности оценивания параметров. Помимо точечных оценок для решения второй задачи математической статистики используется другой подход, занимающийся в построении доверительных интервалов.

Для простоты будем считать, что $r = 1$, т. е. $\vec{\theta} = (\theta)$.

Пусть $\gamma \in (0; 1)$. ([45])

**Определение.** *$\gamma$-доверительным интервалом (доверительным интервалом уровня $\gamma$) для параметра $\theta$ называется пара статистик*

$$\underline{\theta}(\vec{X}),\, \overline{\theta}(\vec{X})$$

*таких, что*

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

**Замечание 1.** *Таким образом, $\gamma$-доверительный интервал — это интервал, который покрывает теоретическое значение параметра $\theta$ с вероятностью $\gamma$.*

**Замечание 2.** *Границы $\underline{\theta}(\vec{X})$, $\overline{\theta}(\vec{X})$ $\gamma$-доверительного интервала являются функциями случайной выборки, т. е. сами являются случайными величинами.*

[45] $\gamma$ — строчная буква «гамма» греческого алфавита. — Прим. ред.

**Замечание 3.** *Вероятностной характеристикой $\gamma$-доверительного интервала является случайная величина $l(\vec{X}) = \overline{\theta}(\vec{X}) - \underline{\theta}(\vec{X})$, которая называется размахом этого интервала.*

**Замечание 4.** *Вероятность совершить ошибку при $\gamma$-доверительном интервале:*

$$P\{\theta \notin (\underline{\theta}(\vec{X}),\, \overline{\theta}(\vec{X}))\} = 1 - \gamma$$

**Замечание 5.** *Иногда удобно строить так называемые односторонние доверительные интервалы.*

**Определение.** *Односторонним нижним (верхним) $\gamma$-доверительным интервалом для параметра $\theta$ называется статистика $\underline{\theta}(\vec{X})$ (соответственно $\overline{\theta}(\vec{X})$) такая, что $P\{\theta \geqslant \underline{\theta}(\vec{X})\} = \gamma$ (соответственно $P\{\theta \leqslant \overline{\theta}(\vec{X})\} = \gamma$).*

### 3.4.2 Построение доверительных интервалов

Пусть

1. $\theta$ — неизвестный параметр закона распределения случайной величины $X$;

2. $g(\vec{X},\, \theta)$ — некоторая статистика.

**Определение.** *Статистику $g(\vec{X},\, \theta)$ будем называть центральной, если закон её распределения не зависит от $\theta$, т. е. $F_g(x,\, \theta) \equiv F_g(x)$, где $F_g$ — функция распределения случайной величины $g$.*

**Пример.** *Пусть $X \sim N(m,\, \sigma^2)$, где $m$ — неизвестно, $\sigma^2$ — известно. Рассмотрим статистику $g(\vec{X},\, m) = \frac{m-\overline{X}}{\sigma} \cdot \sqrt{n}$, где $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Найдём закон распределения случайной величины $g$.*

*1.* $g(\vec{X},\, m) = \frac{m\sqrt{n}}{\sigma} - \frac{1}{\sigma\sqrt{n}}(X_1 + \ldots + X_n)$, *т. е. $g$ является линейным преобразованием нормальной случайных величин $X_1,\, \ldots,\, X_n$, т. е., следовательно, $g$ сама является нормальной случайной величиной ($g \sim N(m_g,\, \sigma_g^2)$).*

*2.* $m_g = M[g] = M[\frac{m-\overline{X}}{\sigma}\sqrt{n}] = \frac{m-\cancel{M[\overline{X}]}^{\,m}}{\sigma}\sqrt{n} = 0$

$\sigma_g^2 = D[g] = D[\frac{m-\overline{X}}{\sigma}\sqrt{n}] = (\frac{\sqrt{n}}{\sigma})^2 D[\cancel{m} - \overline{X}] = \frac{n}{\sigma^2}\underbrace{D[\overline{X}]}_{\frac{\sigma^2}{n}} = 1$

*3. Таким образом, $g(\vec{X},\, m) \sim N(0,\, 1)$, не зависит от неизвестного параметра $m$.*

*Таким образом, $g$ — центральная статистика.*

# Лекция №24, 01.04.2019

Пусть

1. $g(\vec{X},\,\theta)$ является центральной;

2. Случайная величина $g$ является непрерывной случайной величиной, т. е. $F_g(x)$ — непрерывная (из этого следует, что уравнение $F_g(x) = \varpi$ ([46]) имеет единственное решение $\forall\varpi \in (0;\,1)$; $q_\varpi$ — решение этого уравнение — квантиль уровня $\varpi$);

3. $q(\vec{X},\,\theta)$ как функция параметра $\theta$ является монотонно возрастающей;

4. Выбраны значения $\alpha_1,\,\alpha_2 \geqslant 0$ такие, что $\alpha_1 + \alpha_2 = 1 - \gamma$.

Тогда из свойств непрерывных случайных величин

$$\gamma = P\{q_{\alpha_1} < g(\vec{X},\,\theta) < q_{1-\alpha_2}\} = \langle q \text{ может возрастать с ростом } \theta\rangle =$$
$$= P\{\underbrace{g^{-1}(\vec{X},\,q_{\alpha_1})}_{\underline{\theta}(\vec{X})} < \theta < \underbrace{g^{-1}(\vec{X},\,q_{1-\alpha_2})}_{\overline{\theta}(\vec{X})}\}$$

Можно заметить, что согласно определению $\gamma$-доверительного интервала в качестве его границ можно взять отрезки

$$\underline{\theta}(\vec{X}) = g^{-1}(\vec{X},\,q_{\alpha_1}),\ \overline{\theta}(\vec{X}) = g^{-1}(\vec{X},\,q_{1-\alpha_2})$$

**Замечание 1.** *Можно заметить, что построенный $\gamma$-доверительный интервал зависит от выбора $\alpha_1$ и $\alpha_2$. Как правило, используют значения $\alpha_1 = \alpha_2$; тогда $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, где $\alpha = 1 - \gamma$.*

**Замечание 2.** *Если взять $\alpha_2 = 0$, то $\alpha_1 = 1 - \gamma$. В этом случае $q_{1-\alpha_2} = +\infty$ и доверительный интервал имеет вид $(\underline{\theta}(\vec{X}),\,+\infty)$, где $\underline{\theta}(\vec{X}) = g^{-1}(\vec{X},\,q_{1-\gamma})$ — нижняя граница $\gamma$-доверительного множества?.*

**Замечание 3.** *Если взять $\alpha_1 = 0$, то $\alpha_2 = 1 - \gamma$. В этом случае $q_{\alpha_1} = -\infty$ и доверительный интервал имеет вид $(-\infty,\,\overline{\theta}(\vec{X}))$, где $\overline{\theta}(\vec{X}) = g^{-1}(\vec{X},\,q_\gamma)$ — верхняя граница $\gamma$-доверительного множества?.*

**Пример.** *Пусть $X \sim N(m,\,\sigma^2)$, где $m$ — неизвестно, $\sigma^2$ — известно. Выше было показано, что статистика*

$$g(\vec{X},\,m) = \frac{m - \vec{X}}{\sigma}\sqrt{n} \sim N(0,\,1)$$

*Построим с её помощью $\gamma$-доверительный интервал для $m$.*

$$\gamma = P\{u_{\alpha_1} < g(\vec{X},\,m) < u_{1-\alpha_2}\} = P\{u_{\alpha_1} < \frac{m - \overline{X}}{\sigma}\sqrt{n} < u_{1-\alpha_2}\} =$$
$$= P\{\underbrace{\overline{X} + \frac{\delta u_{\alpha_1}}{\sqrt{n}}}_{\underline{m}(\vec{X})} < m < \underbrace{\overline{X} + \frac{\delta u_{1-\alpha_2}}{\sqrt{n}}}_{\overline{m}(\vec{X})}\}$$

[46] $\varpi$ — буква «пи» греческого алфавита. — Прим. ред.

*Таким образом,*

$$\underline{m}(\vec{X}) = \overline{X} + \frac{\delta u_{\alpha_1}}{\sqrt{n}},\ \overline{m}(\vec{X}) = \overline{X} + \frac{\delta u_{1-\alpha_2}}{\sqrt{n}}$$

**Замечание 1.** *Размах $\gamma$-доверительного интервала, построенного в примере:*

$$l(\vec{X}) = \overline{m} - \underline{m} = \frac{\sigma}{\sqrt{n}}(u_{1-\alpha_2} - u_{\alpha_1})$$

*Можно заметить, что в этом примере размах является детерминизированной величиной (т. е. величиной, которая не является случайной). При этом очевидно, что размах будет минимальным, если $\alpha_1 = \alpha_2$. В этом случае*

$$\alpha_1 = \alpha_2 = \frac{\alpha}{2} = \frac{1-\gamma}{2}$$

$$u_{\alpha_1} = u_{\frac{\alpha}{2}} = \langle \text{в силу симметрии графика } f_g(x) \rangle = -u_{1-\frac{\alpha}{2}}$$

$$u_{\alpha_2} = u_{\frac{1+\gamma}{2}}$$

*Тогда*

$$l = \frac{2\sigma}{\sqrt{n}} \cdot u_{\frac{1+\gamma}{2}}$$

*При этом*

$$\underline{m}(\vec{X}) = \overline{X} - \frac{\sigma u_{\frac{1+\gamma}{2}}}{\sqrt{n}},\ \overline{m}(\vec{X}) = \overline{X} + \frac{\sigma u_{\frac{1+\gamma}{2}}}{\sqrt{n}}$$

**Замечание 2.** *Зависимость $l$ от параметров. При $\gamma \to 1 - 0$ $l \to +\infty$ (видно из функции, т. к. $u_{\frac{1+\gamma}{2}} \to +\infty$). Это означает, что чем более достоверную оценку мы хотим получить, тем меньше определённой она будет — ситуация, типичная для математической статистики. Для повышения надёжности результатов при сохранении их детерминизированности следует увеличивать объём наблюдений $n$.*

**Пример.** $X \sim N(m, \sigma^2)$, *где $m$ — неизвестно, $\sigma^2$ — неизвестна. Построить доверительный интервал для $m$.*

*Использовать статистику*

$$g(\vec{X}, m) = \frac{m - \overline{X}}{\sigma}\sqrt{n}$$

*не получится, т. к. $\sigma$ — неизвестна. Используем вместо $\sigma$ исправленную оценку для среднеквадратичного отклонения:*

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

$$g(\vec{X}, m) = \frac{m - \overline{X}}{S(\vec{X})}\sqrt{n}$$

*Запишем $g$ в виде*

$$g(\vec{X},\, m) = \frac{\overbrace{\frac{m-\overline{X}}{\sigma}\sqrt{n}}^{=\xi}}{\sqrt{\underbrace{\frac{(n-1)S^2(\vec{X})}{\sigma^2}}_{=\eta}}}\sqrt{n-1} = \left\langle \xi = \frac{m-\overline{X}}{\sigma}\sqrt{n},\ \eta = \frac{(n-1)S^2(\vec{X})}{\sigma^2}\right\rangle =$$

$$= \frac{\xi}{\sqrt{\eta}}\sqrt{n-1}$$

1. $\xi \sim N(0,\,1)$ *(см. предыдущие примеры);*
2. *Можно показать, что* $\eta \sim \xi^2(n-1)$, *причём случайные величины* $\xi$ *и* $\eta$ *— независимые;*
3. *Таким образом, случайная величина $g$ имеет распределение $St(n-1)$. Построим $\gamma$-доверительный интервал для $m$: $\alpha_1 = \alpha_2 = \frac{1-\gamma}{2}$. Из свойств непрерывных случайных величин:*

   $$\gamma = P\{-t_{\frac{1+\gamma}{2}} < g(\vec{X},\, m) < t_{\frac{1+\gamma}{2}}\} = P\{-t_{\frac{1+\gamma}{2}} < \frac{m-\overline{X}}{S(\vec{X})}\sqrt{n} < t_{\frac{1+\gamma}{2}}\} =$$

   $$P\{\underbrace{\overline{X} - \frac{S(\vec{X}\cdot t_{\frac{1+\gamma}{2}})}{\sqrt{n}}}_{\underline{m}(\vec{X})} < m < \underbrace{\overline{X} + \frac{S(\vec{X}\cdot t_{\frac{1+\gamma}{2}})}{\sqrt{n}}}_{\overline{m}(\vec{X})}\}$$

   *Таким образом,*

   $$\underline{m}(\vec{X}) = \overline{X} - \frac{S(\vec{X}\cdot t_{\frac{1+\gamma}{2}})}{\sqrt{n}},\ \overline{m}(\vec{X}) = \overline{X} + \frac{S(\vec{X}\cdot t_{\frac{1+\gamma}{2}})}{\sqrt{n}}$$

**Пример.** $X \sim (m,\, \sigma^2)$, *где $m$ — неизвестно, $\sigma^2$ — неизвестно. Оценить $\sigma^2$. Рассмотрим статистику*

$$g(\vec{X},\, \sigma^2) = \frac{S^2(\vec{X})}{\sigma^2}\cdot(n-1) \sim \xi^2(n-1)$$

*(см. предыдущий пример). Покажем $\alpha_1 = \alpha_2 = \frac{\alpha}{2} \implies \alpha_1 = \frac{1-\gamma}{2}$.*

$$\gamma = P\{h_{\frac{1-\gamma}{2}} < g(\vec{X},\, \sigma^2) < h_{\frac{1+\gamma}{2}}\} = P\{h_{\frac{1-\gamma}{2}} < \frac{S^2(n-1)}{\sigma^2} < h_{\frac{1+\gamma}{2}}\} =$$

$$= P\{\frac{1}{h_{\frac{1+\gamma}{2}}} < \frac{\sigma^2}{S^2(n-1)} < \frac{1}{h_{\frac{1+\gamma}{2}}}\} = P\{\underbrace{\frac{S^2(n-1)}{h_{\frac{1+\gamma}{2}}}}_{\underline{\sigma^2}} < \sigma^2 < \underbrace{\frac{S^2(n-1)}{h_{\frac{1+\gamma}{2}}}}_{\overline{\sigma^2}}\}$$

*Таким образом,*

$$\underline{\sigma^2} = \frac{S^2(\vec{X})(n-1)}{h_{\frac{1+\gamma}{2}}},\ \overline{\sigma^2} = \frac{S^2(\vec{X})(n-1)}{h_{\frac{1-\gamma}{2}}}$$

## 3.5 Проверка параметрических гипотез

### 3.5.1 Основные понятия

Пусть $X$ — случайная величина, закон распределения которой неизвестен (известен не полностью).

**Определение.** *Статистической гипотезой называется любое утверждение относительно закона распределения случайной величины $X$.*

**Определение.** *Статистическая гипотеза называется простой, если она однозначно определяет закон распределения случайной величины $X$ (однозначно задаёт функцию распределения случайной величины $X$ как функцию своего аргумента). В противном случае статистическая гипотеза называется сложной.*

**Определение.** *Статистическая гипотеза называется параметрической, если она является утверждением относительно значений неизвестных параметров известного закона распределения.*

**Пример 1.** *Пусть $X \sim N(m, \sigma^2)$, где $m$, $\sigma^2$ — неизвестны. Рассмотрим гипотезы:*

$H_1 = \{m = 1, \sigma^2 = 5\}$ *(параметрическая простая гипотеза)*

$H_2 = \{m > 1, \sigma^2 = 5\}$ *(параметрическая сложная гипотеза)*

$H_3 = \{m = 1\}$ *(параметрическая сложная гипотеза)*

**Пример 2.** $X$ *— случайная величина.*

$H_1 = \{X \sim Exp(5)\}$ *(непараметрическая простая гипотеза)*

$H_1 = \{X$ *— нормальная сл-ная величина*$\}$ *(непараметрическая сложная гипотеза)*

$H_1 = \{X \sim \Pi(\lambda),\ \lambda > 5\}$ *(непараметрическая сложная гипотеза)*

Проверку статистических гипотез проводят следующим образом:

1. Формулируют основную гипотезу $H_0$;
2. Формулируют конкурирующую (альтернативную) гипотезу $H_1$; $H_0 \cap H_1 = \emptyset$, но, возможно, $H_0 + H_1$ не исчерпывают все возможные случаи;
3. На основании имеющейся выборки $\vec{X} \in X_n$ принимают решение об истинности $H_0$ или $H_1$.

**Определение.** *Правило, посредством которого принимается решение об истинности $H_0$ или $H_1$, называется статистическим критерием проверки гипотез.*

Задают критерий проверки гипотез обычно с использованием так называемого критического множества $W \subseteq X_n$. При этом решающее правило имеет вид:

$$\vec{x} \in W \implies \begin{cases} \text{отклоняют } H_0 \\ \text{принимают } H_1 \end{cases}, \ \vec{x} \notin W \iff \vec{x} \in \underbrace{\overline{W}}_{X_n \setminus W} \implies \begin{cases} \text{принимают } H_0 \\ \text{отклоняют } H_1 \end{cases}$$

**Замечание 1.** *Задать критерий проверки гипотезы и задать критическое множество — одно и то же.*

**Замечание 2.** *При использовании любого критерия возможны ошибки двух видов:*

1. *Принять конкурирующую гипотезу при использовании основной гипотезы — ошибка I рода. Вероятность совершения этой ошибки обозначим $P\{\vec{x} \in W | H_0\} = \alpha$;*

2. *Принять $H_0$ при истинности $H_1$ — ошибка II рода. Вероятность её совершения: $P\{\vec{x} \notin W | H_1\} = P\{\vec{x} \in X_n \setminus W | H_1\} = \beta$.*

**Замечание 1.** *Множество $X_n \setminus W$ называется доверительным множеством;*

**Замечание 2.** *Конечно, при построении критерия хотелось мы сделать так, чтобы $\alpha$ и $\beta$ минимизировались, однако это невозможно. По этой причине при построении критерия исходят из принципа*

$$\begin{cases} \beta \to \min \\ \alpha = const \end{cases}$$

*или (что эквивалентно)*

$$\begin{cases} \beta = const \\ \alpha \to \min \end{cases}$$

*При этом $\alpha$ называют уровнем значительного критерия, а $1 - \beta$ называют мощностью критерия (таким образом, при построении критерия максимизируют его мощность при фиксированном уровне значимости).*

# Лекция №25, 08.04.2019

Разрыва материала между 24 и 26 лекциями не видно. Эта лекция вообще была?

# Лекция №26, 15.04.2019

### 3.5.2 Критерий Неймана-Пирсона проверки двух простых гипотез

Пусть

1. $X$ — случайная величина;
2. $F(x, \theta)$ — функция распределения случайной величины $X$ (известен общий вид функции $F$, но она зависит от неизвестного параметра $\theta$).

Рассмотрим две простые параметрические гипотезы:

$$H_0 = \{\theta = \theta_0\} \text{ и } H_1 = \{\theta = \theta_1\}$$

где $\theta_0 \neq \theta_1$. Введём в рассмотрение статистику

$$\varphi(\vec{X}) = \frac{L(\vec{X}, \theta_1)}{L(\vec{X}, \theta_0)}$$

где $L(\vec{X}, \theta)$ — функция правдоподобия.

**Определение.** *Статистика $\varphi$ называется отношением правдоподобия.*

Если значения функции достаточно большие, то это ассоциируется с истинностью $H_1$. Очевидно, что «большие» значения статистики $\varphi$ ассоциируются с истинностью конкурирующей гипотезы $H_1$, поэтому критическое множество должно иметь вид:

$$W = \{\vec{x} \in X_n \colon \varphi(\vec{x}) \geqslant C_\varphi\}$$

где константа $C_\varphi$ выбирается из условия

$$\alpha = P\{\varphi(\vec{X}) \geqslant C_\varphi | \theta = \theta_0\} \tag{25}$$

**Замечание.** *При построении критерия зафиксировано некоторое $\alpha = const$ — вероятность совершения ошибки I рода. Чтобы построенный критерий имел уровень значимости $\alpha$, необходимо, чтобы $P\{\vec{X} \in W | H_0\} = \alpha$ — в общем случае. В рассматриваемом случае $\vec{X} \in W \iff \varphi(\vec{X}) \geqslant C_\varphi$.*

$$H_0 \text{ — истина} \iff \theta = \theta_0$$

*поэтому*

$$P\{\vec{X} \in W | H_0\} = P\{\varphi(\vec{X}) \geqslant C_\varphi | \theta = \theta_0\}$$

**Замечание 1.** *Построенный критерий называется критерием Неймана-Пирсона.*

**Замечание 2.** *Если $X$ — непрерывная случайная величина, то плотность распределения случайной выборки $\vec{X}$ имеет вид*

$$f_{\vec{X}}(x_1, \ldots, x_n, \theta) = \overbrace{f(x_1, \theta) \cdot \ldots \cdot f(x_n, \theta)}^{=L(x_1, \ldots, x_n, \theta)}$$

*где $f(x, \theta)$ —- функция плотности распределения случайной величины $X$.*

По этой причине условие 25 может быть записано в следующем виде:

$$\int\limits_{\varphi(t_1,\,\dots,\,t_n)\geqslant C_\varphi}\dots\int L(t_1,\,\dots,\,t_n,\,\theta)\,dt_1\dots dt_n = 1$$

**Пример.** *Пусть* $X \sim N(m,\,\sigma^2)$, *где* $m$ *— неизвестно,* $\sigma^2$ *— известно. Рассмотрим задачу проверки двух простых гипотез*

$$H_0 = \{m = m_0\} \text{ и } H_1 = \{m = m_1\}$$

*где* $m_0 < m_1$. *В этом примере функция правдоподобия имеет вид:*

$$L(x_1,\,\dots,\,x_n,\,m) = [\frac{1}{\sqrt{2\pi}\sigma}]^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-m)^2}$$

*Тогда отношение правдоподобия:*

$$\varphi(\vec{X}) = \frac{L(\vec{X},\,m_1)}{L(\vec{X},\,m_0)} = \frac{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-m_1)^2}}{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-m_0)^2}} = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [(X_i-m_1)^2-(X_i-m_0)^2]} =$$
$$= e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [(X_i-m_1)^2-(X_i-m_0)^2]} = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [X_i^2-2X_im_1+m_1^2-X_i^2+2X_im_0-m_0^2]} =$$
$$= e^{\frac{m_1-m_0}{\sigma^2}(\sum_{i=1}^n X_i)-\frac{n}{2\sigma^2}[m_1^2-m_0^2]} \tag{26}$$

*Выше было показано, что критическое множество должно иметь вид*

$$W = \{\vec{x}\colon \varphi(\vec{x}) \geqslant C_\varphi\}$$

*где* $C_\varphi = const$ *выбирается из условия*

$$P\{\varphi\vec{X} \geqslant C_\varphi | H_0\} = \alpha$$

*Условие*

$$\varphi(\vec{x}) \geqslant C_\varphi \iff \ln\varphi(\vec{X}) \geqslant \ln C_\varphi \iff \langle \textit{см. 26} \rangle \iff$$
$$\iff \ln[e^{\frac{m_1-m_0}{\sigma^2}(\sum_{i=1}^n X_i)-\frac{n}{2\sigma^2}[m_1^2-m_0^2]}] \geqslant \ln C_\varphi \iff$$
$$\iff \frac{m_1-m_0}{\sigma^2}\left(\sum_{i=1}^n X_i\right) - \frac{n}{2\sigma^2}[m_1^2-m_0^2] \geqslant \ln C_\varphi \iff$$
$$\iff \frac{m_1-m_0}{\sigma^2}\left(\sum_{i=1}^n X_i\right) \geqslant \ln C_\varphi + \frac{n}{2\sigma^2}[m_1^2-m_0^2] \iff$$
$$\iff \langle m_1 > m_0 \implies m1 - m_0 > 0\rangle \iff \sum_{i=1}^n X_i \geqslant \underbrace{\frac{\sigma^2}{m_1-m_0}[\ln C_\varphi - \frac{n}{2\sigma^2}[m_1^2-m_0^2]]}_{c=const}$$

*Таким образом,*

$$W = \{\vec{x} \in X_n\colon \sum_{i=1}^{n} x_i \geqslant c\}$$

*где c выбирается из условия*

$$\alpha = P\{\varphi(\vec{X}) \geqslant C_\varphi | H_0\} = P\{\sum_{i=1}^{n} X_i \geqslant C | m = m_0\}$$

*(эти события эквивалентны). Если истина $H_0$, т. е. $m = m_0$, то случайная величина*

$$\sum_{i=1}^{n} X_i \sim (nm_0, n\sigma^2)$$

*Таким образом,*

$$\alpha = P\{\sum_{i=1}^{n} X_i \geqslant C | m = m_0\} = 1 - P\{\sum_{i=1}^{n} X_i < C | m = m_0\} = 1 - \Phi(\frac{c - nm_0}{\sqrt{n\sigma^2}})$$

*то есть*

$$\Phi(\frac{c - nm_0}{\sigma\sqrt{n}}) = 1 - \alpha$$

*Таким образом*

$$\frac{c - nm_0}{\sigma\sqrt{n}} = u_{1-\alpha}$$

*($u_{1-\alpha}$ — квантиль уровня $1 - \alpha$ стандартного нормального распределения).* $c = \sigma u_{1-\alpha}\sqrt{n} + nm_0$

*Таким образом, критерий имеет вид:*

$$\sum_{i=1}^{n} X_i \geqslant \delta u_{1-\alpha}\sqrt{n} + nm_0 \implies \begin{cases} \text{принять } H_1, \\ \text{отклонить } H_0 \end{cases}$$

$$\sum_{i=1}^{n} X_i < \delta u_{1-\alpha}\sqrt{n} + nm_0 \implies \begin{cases} \text{принять } H_0, \\ \text{отклонить } H_1 \end{cases}$$

*При этом вероятность совершения ошибки второго рода:*

$$\beta = P\{\vec{X} \notin W | H_1\} = P\{\sum_{i=1}^{n} X_i < C | m = m_1\} =$$

$$= \langle c = \delta u_{1-\alpha}\sqrt{n} + nm_0; \text{при } m = m_1\colon \sum_{i=1}^{n} X_i \sim N(nm_1, n\sigma^2)\rangle =$$

$$= \Phi(\frac{(\sigma u_{1-\alpha}\sqrt{n} + nm_0) - nm_1}{\sigma\sqrt{n}}) = \Phi(\frac{\sigma u_{1-\alpha}\sqrt{n} - n(m_1 - m_0)}{\sigma\sqrt{n}}) = \Phi(u_{1-\alpha} - \sqrt{n}\frac{m_1 - m_0}{\sigma})$$

**Замечание.** *Если в условиях предыдущего примера $m_1 < m_0$, то критическое множество должно задаваться условием $\sum_{i=1}^n x_i \leqslant C$, где $C = const$ выбирается из условия $P\{\sum_{i=1}^n X_i > C|H_0\} = \alpha$, т. е. $P\{\sum_{i=1}^n X_i < C|m = m_0\} = \alpha$, т. е. $\Phi(c - \frac{nm_0}{\sigma\sqrt{n}}) = \alpha$, откуда*

$$\frac{C - nm_0}{\sigma\sqrt{n}} = u_\alpha \implies C = \underbrace{\delta u_\alpha + \sqrt{n}}_{-\delta u_{1-\alpha} - \sqrt{n}} + nm_0$$

*(т. к. $u_\alpha = u_{1-\alpha}$ в случае ст. норм? распределения).*

### 3.5.3 Проверка сложный параметрических гипотез

Пусть

1. $X$ — случайная величина;
2. $F(x, \theta)$ — функция распределения случайной величины $X$ (общий вид функции $F$ известен, но $F$ зависит от неизвестного параметра $\theta$).

Рассмотрим задачу проверки двух сложных гипотез:

$$H_0 = \{\theta \in \Theta_0\} \text{ и } H_1 = \{\theta \in \Theta_1\}$$

где $\Theta_0 \cap \Theta_1 = \emptyset$.

**Пример 1.** $\Theta_0 = \{\theta > \theta_0\}$, $\Theta_1 = \{\theta \leqslant \theta_0\}$.

**Пример 2.** $\Theta_0 = \{\theta < \theta_0\}$, $\Theta_1 = \{\theta > \theta_1\}$, *где* $\theta_0 < \theta_1$.

В этом случае, критерий, как и раньше, задаётся с использованием критического множества $W$, а решающее правило имеет вид:

$$\vec{x} \in W \implies \begin{cases} \text{отклоняют } H_0 \\ \text{принимают } H_1 \end{cases}, \ \vec{x} \notin W \iff \vec{x} \in \underbrace{\overline{W}}_{X_n \backslash W} \implies \begin{cases} \text{принимают } H_0 \\ \text{отклоняют } H_1 \end{cases}$$

При этом ошибки первого и второго родов определяются так же, как и раньше, но теперь их вероятности зависят от $\theta$:

$$\alpha(\theta) = P\{\vec{X} \in W|\theta \in \Theta_0\}$$
$$\beta(\theta) = P\{\vec{X} \in X_n \setminus W|\theta \in \Theta_1\}$$

**Определение.** *Величина*

$$\alpha = \sup_{\theta \in \Theta_0} \alpha(\theta)$$

*называется размером критерия.*

**Определение.** *Функция*

$$M(\theta) = P\{\vec{X} \in W|\theta\} \tag{27}$$

*называется функцией мощности критерия.*

**Замечание 1.** *Условие 27, принятое в математической статистике, было бы удачнее записать в виде*

$$M(t) = P\{\vec{X} \in W|\theta = t\}$$

*т. е. $M(t)$ — вероятность события $\{\vec{X} \in W\}$ при условии, что неизвестный параметр имеет значение $\theta$.*

**Замечание 2.** *Через функцию мощности можно выразить вероятности совершения ошибок первого и второго рода:*

$$\alpha(\theta) = M(\theta),\ \theta \in \Theta_0$$
$$\beta(\theta) = 1 - M(\theta),\ \theta \in \Theta_1$$

# Лекция №27, 22.04.2019

**Определение.** *Критерий, который при заданном размере $\alpha$ максимизирует функцию мощности одновременно по всем возможным критериям при всех $\theta \in \Theta_1$, называется равномерно наиболее мощным.*

**Замечание 1.** *Таким образом, равномерно наиболее мощный критерий, если он существует, при заданном размере $\alpha$ минимизирует вероятность совершения ошибки второго рода по всем возможным критериям при всех значениях $\theta \in \Theta_2$.*

**Замечание 2.** *Равномерно наиболее мощный критерий существуют лишь в некоторых частных случаях при проверке гипотез о значениях одномерных параметров.*

**Пример.** *Пусть $X \sim N(m, \sigma^2)$, где $m$ — неизвестно, $\sigma^2$ — известно. Рассмотрим задачу проверки гипотез*

$$\underbrace{H_0 = \{m = m_0\}}_{\text{простая гипотеза}} \text{ и } \underbrace{H_1 = \{m > m_0\}}_{\text{сложная гипотеза}}$$

1. *Ранее была решена задача проверки двух простых гипотез*

$$H_0 = \{m = m_0\} \text{ и } H_1 = \{m = m_1\}$$

*где $m1 > m_0$. При этом критическое множество имело вид*

$$W = \{\vec{x} \in X_n\colon \sum_{i=1}^{n} x_i \geqslant nm_0 + u_{1-\alpha}\sigma\sqrt{n}\} \tag{28}$$

2. *Т. к. построенное выше критическое множество не зависит от $m_1$, то фактически этот критерий является равномерно наиболее мощным для проверки гипотез $H_0 = \{m = m_0\}$ и $H_1 = \{m > m_0\}$. Таким образом, для рассматриваемой задачи критическое множество имеет вид 28.*

**Замечание.** *Если в условиях предыдущего примера рассматривается задача проверки гипотез*

$$H_0 = \{m = m_0\} \text{ и } H_1 = \{m < m_0\}$$

*то аналогичным образом приходим к выводу, что критическое множество будет иметь вид*

$$W = \{\vec{x} \in X_n \colon \sum_{i=1}^{n} x_i \leqslant nm_0 - u_{1-\alpha}\sigma\sqrt{n}\}$$

*(см. аналогичное замечание после примера о проверке $H_0 = \{m = m_0\}$ против $H_1 = \{m = m_1\}$ при $m_0 > m_1$).*

**Пример.** *Пусть $X \sim N(m, \sigma^2)$, $m$ — неизвестно, $\sigma^2$ — известно. Рассмотрим задачу проверки $H_0 = \{m = m_0\}$ против $H_1 = \{m \neq m_0\}$. В предыдущих примерах рассмотрение аналогичной задачи привело к критическим множествам вида:*

1. $H_1 = \{m > m_0\}$, $W = \{\vec{x} \in X_n \colon T(\vec{x}) \geqslant u_{1-\alpha}\}$, *где $u_{1-\alpha}$ — квантиль стандартного нормального распределения.*

$$T(\vec{X}) = \frac{\overline{X} - m_0}{\sigma}\sqrt{n}$$

*(при истинности гипотезы $H_0$ $T(\vec{X}) \sim N(0, 1)$).*

2. $H_1 = \{m < m_0\}$, $W = \{\vec{x} \in X_n \colon T(\vec{X}) \leqslant u_{1-\alpha}\}$, *где $T$ и $u_{1-\alpha}$ имеют тот же смысл, что и в пункте а.*

*Воспользуемся статистикой $T$ и в рассматриваемом примере, когда $H_1 = \{m \neq m_0\}$. ?*

*Тогда при истинности конкурирующей гипотезы $H_0$ статистика $T$ будет принимать большие по абсолютной величине значения, поэтому критическое множество можно записать в виде*

$$W = \{\vec{x} \in X_n \colon |T(\vec{x})| \geqslant u_{1-\frac{\alpha}{2}}\}$$

*В этом случае размер критерия (т. е. вероятность совершить ошибку первого рода) равен $\alpha$.*

**Пример.** *Пусть $X \sim N(m, \sigma^2)$, где $m$ — неизвестно, $\sigma^2$ — известно. Рассмотрим задачу проверки*

$$\underbrace{H_0 = \{m = m_0\}}_{\text{сложная (т. к. дисперсия неизвестна)}} \quad \text{против} \quad \underbrace{H_1 = \{m > m_0\}}_{\text{сложная}}$$

*В этом примере целесообразно воспользоваться статистикой*

$$T(\vec{X}) = \frac{\overline{X} - m_0}{S(\vec{X})}\sqrt{n} \sim St(n-1) \text{ (при истинности } H_0\text{)}$$

*Аналогично предыдущим примерам критическое множество можно задать в виде*

$$W = \{\vec{x} \in X_n \colon T(\vec{X}) \geqslant t_{1-\alpha}^{(n-1)}\}$$

*где* $t_{1-\alpha}^{(n-1)}$ *— квантиль уровня* $1-\alpha$ *распределения* $St(n-1)$.

**Замечание.** *Пусть в условиях предыдущего примера требуется проверить следующие гипотезы:*

1. $H_0 = \{m = m_0\}$ *против* $H_1 = \{m < m_0\}$ *и*
2. $H_0 = \{m = m_0\}$ *против* $H_1 = \{m \neq m_0\}$.

*Рассуждая аналогично, с использованием статистики*

$$T(\vec{X}) = \frac{\overline{X} - m_0}{S(\vec{X})}\sqrt{n}$$

*зададим критические множества в виде*

1. $W = \{\vec{x} \in X_n \colon T(\vec{X}) \leqslant -t_{1-\alpha}^{(n-1)}\}$
2. $W = \{\vec{x} \in X_n \colon |T(\vec{X})| \geqslant t_{1-\frac{\alpha}{2}}^{(n-1)}\}$

**Пример.** *Пусть*

1. $X \sim N(m_1, \sigma_1^2)$;
2. $Y \sim N(m_2, \sigma_2^2)$;
3. $m_1$, $m_2$ *— неизвестно,* $\sigma_1$, $\sigma_2$ *— известны. Рассмотрим задачи проверки гипотез:*
   (a) $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 > m2\}$;
   (b) $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 < m_2\}$;
   (c) $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 \neq m_2\}$.

   *Рассмотрим случайную величину* $Z = X - Y$, $MZ = MX - MY$, *поэтому сформулированные задачи эквивалентны задачам:*

   (a) $H_0 = \{m = 0\}$ *против* $H_1 = \{m > 0\}$;
   (b) $H_0 = \{m = 0\}$ *против* $H_1 = \{m < 0\}$;

(c) $H_0 = \{m = 0\}$ *против* $H_1 > \{n \neq 0\}$,

*где* $m = M[Z]$.

*Рассмотрим статистику*

$$T(\vec{X}, \vec{Y}) = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}}}$$

*где* $n_1$ *— объём выборки* $\vec{X}$, $n_2$ *— объём выборки* $\vec{Y}$. *Каков закон распределения случайной величины* $G$ *при истинности* $H_0$? $T$ *является линейной комбинацией нормальных случайных величин, следовательно,* $T$ *сама имеет нормальное распределение.*

$$M[T] = \frac{1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}(M[\overline{X}] - M[\overline{Y}]) = \langle \text{при истинности } H_0\ m_1 = m_2 \rangle = 0$$

$$D[T] = \frac{1}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot [D\overline{X} + D\overline{Y}] = \frac{1}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}] = 1$$

*Таким образом, при истинности* $H_0$ *статистика* $T \sim N(0, 1)$. *По этой причине критические множества в каждой из рассматриваемых задач имеют вид:*

(a) $W = \{(\vec{x}, \vec{y}) \in X_n : T(\vec{x}, \vec{y}) \geqslant u_{1-\alpha}\}$, $m > 0$;

(b) $W = \{(\vec{x}, \vec{y}) \in X_n : T(\vec{x}, \vec{y}) \leqslant -u_{1-\alpha}\}$, $m < 0$;

(c) $W = \{(\vec{x}, \vec{y}) \in X_n : |T(\vec{x}, \vec{y})| \geqslant u_{1-\frac{\alpha}{2}}\}$, $m \neq 0$.

# Лекция №28, 29.04.2019

**Пример.** *Пусть*

1. $X_1 \sim N(m_1, \sigma_1^2)$, *где* $m_1$, $\sigma_1^2$ *— неизвестные;*
2. $X^2 \sim N(m_2, \sigma_2^2)$, *где* $m_2$, $\sigma_2^2$ *— неизвестные.*

*Рассмотрим задачи проверки гипотез*

1. $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 > m2\}$;
2. $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 < m_2\}$;
3. $H_0 = \{m_1 = m_2\}$ *против* $H_1 = \{m_1 \neq m_2\}$.

*Рассуждая аналогично предыдущему примеру, заключаем, что статистика*

$$\tilde{T}(\vec{X},\,\vec{Y}) = \underbrace{\frac{(n_1-1)S^2(\vec{X})}{\sigma_1^2}}_{\sim\chi^2(n_1-1)} + \underbrace{\frac{(n_2-1)S^2(\vec{Y})}{\sigma_2^2}}_{\sim\chi^2(n_2-1)} \sim \chi^2(n_1+n_2-2)$$

*При истинности $H_0$ статистика*

$$\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,\,1)$$

*поэтому статистика*

$$T(\vec{X},\,\vec{Y}) = \frac{(\overline{X}-\overline{Y})\sqrt{n_1+n_2-2}}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{(n_1-1)S^2(\vec{X})+(n_2-1)S^2(\vec{Y})}} \sim St(n_1+n_2-2)$$

*Таким образом, критические множества для каждой из рассматриваемых задач выглядят так же, как и в предыдущем примере при условии $u \longleftrightarrow t^{(n_1+n_2-2)}$.*

**Пример.** *Пусть*

1. $X \sim N(m_1,\,\sigma_1^2)$; $Y \sim N(m_2,\,\sigma_2^2)$;
2. $m_1,\,m_2,\,\sigma_1^2,\,\sigma_2^2$ — *неизвестны.*

*Рассмотрим задачи проверки гипотез*

1. $H_0 = \{\sigma_1 = \sigma_2\}$ *против* $H_1 = \{\sigma_1 > \sigma 2\}$;
2. $H_0 = \{\sigma_1 = \sigma_2\}$ *против* $H_1 = \{\sigma_1 < \sigma_2\}$;
3. $H_0 = \{\sigma_1 = \sigma_2\}$ *против* $H_1 = \{\sigma_1 \neq \sigma_2\}$.

*Рассмотрим статистику*

$$T(\vec{X},\,\vec{Y}) = \frac{S^2(\vec{X})}{S^2(\vec{Y})} \sim F(n_1-1,\,n_2-2) \text{ (распределение Фишера)}$$

*Рассмотрим задачу а. Рассуждая аналогично предыдущим примерам, приходим к выводу, что критическое множество имеет вид*

$$W = \{(\vec{x},\,\vec{y})\colon T(\vec{x},\,\vec{y}) \geqslant f_{1-\alpha}^{(n_1-1,\,n_2-1)}\}$$

*В задаче б:*

$$W = \{(\vec{x},\,\vec{y})\colon T(\vec{x},\,\vec{y}) \leqslant f_{\alpha}^{(n_1-1,\,n_2-1)}\}$$

*В задаче в:*

$$W = \{(\vec{x},\,\vec{y})\colon (T(\vec{x},\,\vec{y}) \leqslant f_{\frac{\alpha}{2}}^{(n_1-1,\,n_2-1)}) \vee (T(\vec{x},\,\vec{y}) \geqslant f_{1-\frac{\alpha}{2}}^{(n_1-1,\,n_2-1)})\}$$

## 3.6 Проверка непараметрических гипотез: критерии согласия

### 3.6.1 Основные понятия

До сих пор рассматривалась вторая задача математической статистики:

Дано $X$ — случайная величина, закон распределения которой известен с точностью до вектора $\vec{\theta}$ неизвестных параметров. Требуется:оценить значение $\vec{\theta}$.

Рассмотрим первую задачу математической статистики:

Дано: $X$ — случайная величина, закон распределения которой неизвестен. Требуется: найти закон распределения случайной величины $X$.

Решение этой задачи сводится к проверке основной гипотезы

$$H_0 = \{F(t) \equiv F_0(t)\} = \{(\forall t \in \mathbb{R})(F(t) = F_0(t))\}$$

где $F(t)$ — функция распределения случайной величины $X$, $F_0(t)$ — некоторая известная функция распределения, против конкурирующей гипотезы

$$H_1 = \neg H_0 = \{(\exists t \in \mathbb{R})(F(t) \neq F_0(t))\}$$

**Замечание.** *В таком виде задача формулирована для проверки простой гипотезы $H_0$. Однако на практике гораздо чаще возникает задача проверки гипотезы об общем виде функции распределения случайной величины $X$. Например,*

$$H_0 = \{X \text{ имеет нормальное распределение (с некоторыми значениями параметров)}\}$$

*Или*

$$H_0 = \{X \text{ имеет распределение Пуассона (для некоторого значения параметра)}\}$$

*В этом случае гипотеза $H_0$ является сложной и (если $\vec{\theta}$ — вектора неизвестных параметров) имеет вид:*

$$H_0 = \{(\exists \vec{\theta}_0)(\forall t \in \mathbb{R})(F(t) = F_0(t, \vec{\theta}_0)\}$$

*где $F(t)$ — функция распределения случайной величины $X$, $F_0(t, \vec{\theta})$ — некоторая функция распределения, известная с точностью до вектора $\vec{\theta}_0$. При этом конкурирующая гипотеза:*

$$H_1 = \neg H_0 = \{(\forall \vec{\theta})(\exists t \in \mathbb{R})(F(t) \neq F_0(t, \vec{\theta})\}$$

Проверка основной гипотезы $H_0$ сводится к оценке величины

$$\triangle(F_n, F_0)$$

рассогласования эмпирической функции распределения и предполагаемой функции распределения $F_0$.

**Определение.** *Криетрием согласия называется статистический критерий, предназначенный для проверки корректности гипотезы о том, что предполагаемый закон распределения $F_0(t, \vec{\theta})$ случайной величины $X$ соответствует экспериментальным данным, представленным эмпирической функцией распределения $F_n(t)$.*

**Замечание.** *При выдвижении основной гипотезы $H_0$ могут быть использованы следующие соображения:*

1. *Анализ априорной информации об изученном процессе или явлении и её сопоставление и исходными предпосылками построения конкретных моделей (нормальной, биномиальной, экспоненциальной);*

2. *Построение эмпирической функции распределения по имеющейся выборке $\vec{x}$. На основании вида этой функции может быть выдвинута гипотеза $H_0$;*

3. *Построение гистограммы и использование её для выдвижения гипотезы $H_0$.*

### 3.6.2 Критерий Колмогорова для простой гипотезы

Пусть

1. $X$ — непрерывная случайная величина;

2. $\vec{X}$ — случайная выборка из генеральной совокупности $\vec{X}$.

Рассмотрим задачу проверки гипотезы $H_0 = \{F(t) \equiv F_0(t)\}$ против $H_1 = \neg H_0$.

**Замечание.** *Здесь $F_0(t)$ — полностью известная функция распределения, которая не зависит ни от каких-то ни было неизвестных параметров. По этой причине $H_0$ — простая гипотеза.*

Для решения этой задачи рассмотрим статистику $\triangle(\vec{X})$, реализации которой определяются соотношением

$$\triangle(\vec{x}) = \sup |F_n(t) - F_0(t)|,\ t \in \mathbb{R}$$

где $F_n(t)$ — эмпирическая функция распределения, построенная по выборке $\vec{x}$. Очевидно, что «малые» значения статистики $\triangle$ свидетельствуют об истинности $H_0$, а «большие» — об истинности $H_1$. По этой причине критическое множество имеет вид

$$W = \{\vec{x} \in X_n \colon \triangle(\vec{x}) \geqslant \sigma_{1-\alpha}\}$$

где $\sigma_{1-\alpha}$ — квантиль уровня $1 - \alpha$ закона распределения случайной величины $\triangle(\vec{X})$. При этом собственно решающее правило имеет стандартный вид:

$$\begin{aligned} \vec{x} \in W &\implies \text{принять } H_1,\ \text{отклонить} H_0 \\ \vec{x} \in \overline{W} &\implies \text{принять } H_0,\ \text{отклонить} H_1 \end{aligned}$$

**Замечание.** *О законе распределения случайной величины* $\triangle(\vec{X})$.

1. *Построенный критерий предполагает использование квантилей закона распределения случайной величины* $\triangle(\vec{X})$. *Вообще говоря, этот закон распределения зависит от:*
   (a) *От выбранной функции* $F_0(t)$;
   (b) *От теоретического закона распределения случайной величины* $X$.
2. *В математической статистике доказано утверждение:*

   **Теорема.** *Пусть*
   (a) $Y \sim R[0;\, 1]$;
   (b) $\hat{R}_n(t,\, \vec{Y})$ — *выборочная функция распределения случайной величины* $Y$, *отвечающая выборке* $\vec{Y}$.

   *Тогда при исполнении* $H_0$ *функция распределения случайной величины* $\triangle(\vec{X})$ *совпадает с функцией распределения случайной величины* $Z(n) = \sup\limits_{t\in[0;\,1]} |\hat{R}_n(t,\, \vec{Y}) - t|$.

   *Таким образом, для всевозможных законов распределения случайной величины* $X$ *и всевозможных функций распределения* $F_0$, *участвующих в построении основной гипотезы* $H_0$, *закон распределения случайной величины* $\triangle(\vec{X})$ *совпадает с законом распределения случайной величины* $Z(n)$.
3. *Для каждого фиксированного* $n \in \mathbb{N}$ *закон распределения случайной величины* $Z(n)$ *известен. Для* $n \leqslant 100$ *составлены таблицы значений соответствующей функции распределения.*
4. *Что делать, если* $n > 100$?
   (a) *Колмогоров доказал, что*

   $$\lim_{n\to\infty} P\{\sqrt{n}Z(n) < t\} = K(t),\ t > 0$$

   *где* $K(t) = \sum_{n=-\infty}^{+\infty}(-1)^k e^{-2k^2t^2}$. *Другими словами, при* $n \to \infty$ *последовательность случайных величин* $\sqrt{n}Z(n)$ *слабо сходится к случайной величине* $A$, *функцией распределения которой является*

   $$F_a(t) = \begin{cases} 0, & t \leqslant 0 \\ k(t), & t > 0 \end{cases}$$

(b) *Таким образом, при достаточно больших $n$ квантили случайной величины $\sqrt{n}Z(n)$ совпадают с квантилями соответствующих уровней случайной величины $A$. Если а м— квантиль случайной величины $A$, б — квантиль случайной величины $Z(n)$, то*

$$\sigma_{1-\alpha} = \frac{a_{1-\alpha}}{\sqrt{n}} \tag{29}$$

*Таблица значений функции $K(t)$ составлены, их можно найти в литературе (интернете).*

5. *Как показывает практика, соотношением 29 можно пользоваться при $n \geqslant 20$.*

**Замечание.** *Построенный критерий называется критерием Колмогорова для простой гипотезы.*

# Лекция №29, 06.05.2019

### 3.6.3 Критерий хи-квадрат для простой гипотезы

Пусть

1. $X$ — *дискретная* случайная величина;
2. $X$ может принимать конечное множество значений $a_1, \ldots, a_l$ с неизвестными вероятностями $p_1, \ldots, p_l$.

Требуется проверить основную гипотезу

$$H_0 = \{p_1 = p_1^0, \ldots, p_l = p_l^0\}$$

(где $p_1^0, \ldots, p_l^0$ — некоторые известные значения) против конкурирующей гипотезы

$$H_1 = \neg H_0 = \{\exists k \in \{1, \ldots, l\} \colon p_k \neq p_k^0\}$$

Для решения этой задачи рассмотрим статистики $n_1(\vec{X}), \ldots, n_l(\vec{X})$, где выборочные значения $n_k(\vec{X})$

$$n_k(\vec{x}) = \{\text{кол-во элементов выборки } \vec{x} \text{ которые имеют значение } n_k\}$$

**Замечание.** *Очевидно, что $n_1(\vec{X}) + ... + n_l(\vec{X}) = n$, поэтому случайные величины $n_1(\vec{X}), \ldots, n_l(\vec{X})$ зависимы.*

**Теорема.** ***Теорема Пирсона.***

*Пусть выполнены сделанные выше предположения. Тогда при истинности $H_0$ последовательность случайных величин*

$$\sum_{k=1}^{l} \frac{(n_k(\vec{X}) - np_k)^2}{np_k}$$

*слабо сходится к случайной величине, имеющей распределение $\chi^2(l-1)$.*

Согласно этой теореме, при $n \to \infty$ случайная величина

$$\triangle(\vec{X}) = \sum_{k=1}^{l} \frac{(n_k(\vec{X}) - np_k^0)^2}{np_k^0} = n \sum_{k=1}^{l} \frac{(\frac{n_k(\vec{X})}{n} - p_k^0)^2}{p_k^0}$$

сходится к случайной величине, распределённой по закону $\chi^2(l-1)$.

Очевидно, что истинность гипотезы $H_0$ ассоциируется с «малыми» значениями статистики $\triangle(\vec{X})$, а истинность конкурирующей гипотезы $H_1$ — с «большими» положительными значениями. По этой причине критическое множество можно задать в виде

$$W = \{\vec{x} \in X_n \colon \triangle(\vec{x}) \geqslant h_{1-\alpha}^{(l-1)}\}$$

где $h_{1-\alpha}^{(l-1)}$ — квантиль уровня $1-\alpha$ распределения $\chi^2(l-1)$.

**Замечание 1.** *Построенный критерий называется критерием хи-квадрат ($\chi^2$) или критерием Пирсона.*

**Замечание 2.** *В отличие от критерия Колмогорова, критерий хи-квадрат носит асимптотический характер, т. к. статистика $\triangle(\vec{X})$ имеет распределение, близкое к $\chi^2(l-1)$ лишь при достаточном больших $n$.*

*Эмпирические рекомендации таковы:*

1. *При $l \lesssim 20$: должно быть $np_k \geqslant 10$, $k = \overline{1;\, l}$ для использования критерия;*

2. *При $l > 20$, достаточно выполнения условия $np_k \geqslant 5$, $k = \overline{1;\, l}$.*

### 3.6.4 Критерий Колмогорова для сложной гипотезы

На практике задачи проверки простой гипотезы

$$H_0 = \{F(t) \equiv F_0(t)\}$$

встречается крайне редко; как правило, требуется проверить гипотезу о принадлежности закона распределения случайной величины $X$ заданному классу (нормальному, экспоненциальному и т. д.) По этой причине основная гипотеза $H_0$ будет сложной:

$$H_0 = \{(\exists\vec{\theta})(\forall t \in \mathbb{R})(F(t) = F_0(t, \vec{\theta}))\}$$

где $F(t)$ — теоретический ("истинный") закон распределения случайной величины $X$, $F_0$ — предполагаемый закон распределения случайной величины, $\vec{\theta}$ — вектор параметров закона $F_0$.

Как правило, конкурирующую гипотезу формулируют так:

$$H_1 = \neg H_0$$

Для решения задачи проверки гипотезы $H_0$ против $H_1$ естественно сделать следующее:

1. Построить точечную оценку $\hat{\vec{\theta}}(\vec{X})$ для значения вектора параметров $\vec{\theta}$;

2. Использовать критерий Колмогорова для проверки простой гипотезы.

$$H_0 = \{F(t) \equiv F_0(t, \hat{\vec{\theta}}(\vec{x}))\}$$

   где $\hat{\vec{\theta}}(\vec{x})$ — выборочное значение построенной оценки.

Недостатком данного подхода является то, что в этом случае критерий перестаёт быть параметрическим, т. к. распределение модифицированной статистики.

$$\triangle(\vec{X}) = \sup_{t\in\mathbb{R}} |F(t) - F_0(t, \hat{\vec{\theta}}(\vec{x}))|$$

зависит от выбранной точечной оценки, т. е. от закона распределения случайной величины $\hat{\vec{\theta}}(\vec{x})$. Однако можно показать, что если

1. $\hat{\vec{\theta}}(\vec{X})$ — оценка максимально правдоподобна для вектора $\vec{\theta}$,

2. Элементы $F_0(t, \vec{\theta})$ параметрического семейства получаются из какого-нибудь одного представителя с использованием преобразований сдвига и масштаба (вдоль оси $Ot$), т. е.
$$F_0(t, \vec{\theta}) = \tilde{F}_0(\tfrac{t-a}{b})$$
   где $\tilde{F}_0(t)$ — какая-то фиксированная функция из рассматриваемого семейства $F_0(t, \vec{\theta})$; $a$, $b$ — компоненты, значения которых зависят от значения $\vec{\theta}$ в левой части,

то для исполнения критерия Колмогорова достаточно иметь только одну таблицу квантилей для каждого семейства (т. е. одну таблицу для нормального семейства, одну для экспоненциального и т. д.) Также таблицы составлены для наиболее важных семейств.

### 3.6.5 Критерий хи-квадрат для сложной гипотезы

Пусть

1. $X$ — дискретная случайная величина;

2. $X$ — может принимать значения из множества $a_1, \ldots, a_l$ с неизвестными вероятностями $p_1, \ldots, p_l$;

3. Эти вероятности $p_k$, $k = \overline{1; l}$ зависят от неизвестных параметров $\vec{\theta}$, где $\vec{\theta} \in \Theta$, т. е. в отличие от критерия $\chi^2$ для простой гипотезы теперь $p_n = p_k(\vec{\theta})$, $\vec{\theta} \in \Theta$, $k = \overline{1; l}$.

По этой причине основную гипотезу можно записать в виде

$$H_0 = \{P\{X = a_k\} = p_{k_0}(\vec{\theta}),\, k = \overline{1;\, l}\}$$

Т. е. $p_{k_0}(\theta)$ — известные функции, предполагаемые зависимости вероятностей $p_k$ от параметров $\vec{\theta}$; $P\{X = a_k\} = p_k(\vec{\theta})$ — теоретические ("истинные") зависимости этих вероятностей от параметров, эти зависимости нам неизвестны.

Конкурирующую гипотезу обычно выбирают как $H_1 = \neg H_0$; для решения задачи проверки $H_0$ против $H_1$ используется модифицированный критерий хи-квадрат.

1. Сначала строят оценку максимального правдоподобия для вектора параметров $\vec{\theta}$: $\hat{\vec{\theta}}(\vec{X})$;

2. Вычисляют выборочное значение $\hat{\vec{\theta}}(\vec{x})$, $n_k(\vec{x})$ — (см. критерий $\chi^2$ для простой гипотезы);

3. Рассматривают статистику

$$\chi^2(\vec{X}) = \sum_{k=1}^{l} \frac{[n_k(\vec{X}) - np_{k_0}(\hat{\vec{\theta}}(\vec{X}))]^2}{np_{k_0}(\hat{\vec{\theta}}(\vec{X}))} = n\sum_{k=1}^{l} \frac{[\frac{n_k(\vec{X})}{n} - p_{k_0}(\hat{\vec{\theta}}(\vec{X}))]^2}{p_{k_0}(\vec{\theta}(\vec{X}))}$$

   которая, в случае выполнения определённых условий гладкости функции $p_{k_0}(\vec{\theta})$ при $n \to \infty$ слабо сходится к случайной величине, имеющей распределения $\chi^2(l-r-1)$, где $r$ — размерность вектора $\vec{\theta}$.

4. Поскольку при истинности основной гипотезы $H_0$ статистика $\chi^2(\vec{X})$ принимает «малые» значения, а при истинности $H_1$ — «большие» значения, критическое множество можно записать в следующем виде:

$$W = \{\vec{x}\colon \chi^2(x^2) \geqslant h_{1-\alpha}^{(l-r-1)}\}$$

   где $h_{1-\alpha}^{(l-r-1)}$ — квантиль уровня $1-\alpha$ распределения $\chi^2(l-r-1)$.

**Замечание.** ***О построении оценки максимального правдоподобия в рассматриваемом случае.***

*При истинности основной гипотезы $H_0$ функция правдоподобия имеет следующий вид:*

$$L(\vec{X},\, \vec{\theta}) = \prod_{j=1}^{n} P\{X = X_j\} =$$

$$= \langle \textit{т. к. среди компонент вектора } \vec{X} \textit{ значение } a_k \textit{ встречается } n_k(\vec{X}) \textit{ раз,}$$
$$\textit{а элементы вектора } \vec{X} \textit{ могут быть различными способами раскиданы по позициям}$$
$$\textit{в этом векторе, то это выражение можно записать в следующем виде}\rangle =$$

$$= \frac{n!}{n_1! \cdot \ldots \cdot n_k!} \prod_{k=1}^{l} \left[p_{k_0}(\vec{\theta})\right]^{h_k(\vec{X})}$$

*где* $\sum_{k=1}^{l} n_k(\vec{X}) = n$*. Тогда уравнения правдоподобия*

$$\frac{\delta \ln L}{\delta \theta_j} = 0,\ j = \overline{1;\, r}$$

*примут вид*

$$\sum_{k=1}^{l} \frac{n_k(\vec{X})}{p_{k_0}(\vec{\theta})} \cdot \frac{\delta p_{k_0}(\vec{\theta})}{\delta \theta_j} = 0,\ j = \overline{1;\, r}$$

# Лекция №30, 13.05.2019

### 3.6.6 Двухвыборочная задача

Пусть

1. $X$, $Y$ — случайные величины;
2. $F(t)$ — функция распределения случайной величины $X$, $G(t)$ — функция распределения случайной величины $Y$.
3. $\overline{X}$ — случайная выборка из генеральной совокупности $X$ (объёма $n_1$), $\overline{Y}$ — случайная выборка из генеральной совокупности $Y$ (объёма $n_2$).

Требуется проверить гипотезу

$$H_{)} = \{X \text{ и } Y \text{ одинаково распределены}\} = \{(\forall t \in \mathbb{R})(F(t) = G(t))\}$$

против $H_1 = \neg H_0$. Если случайные величины $X$ и $Y$ непрерывны, то для решения этой задачи можно использовать статистику $\triangle(\vec{X},\, \vec{Y})$, выборочные значения которой выражаются формулой

$$\triangle(\vec{x},\, \vec{y}) = \sup_{t \in \mathbb{R}} |F_{n_1}(t) - G_{n_2}(t)|$$

где $F_{n_1}(t)$, $G_{n_2}(t)$ — эмпирические функции распределения, отвечающие выборкам $\vec{x}$ и $\vec{y}$. Если истина $H_0$, то в соответствии с теоремой о сходимости функции распределения заключаем, что при достаточно больших $n_1$ и $n_2$ значения статистики должны быть «малыми», а при истинности $H_1$ — «большими». По этой причине критическое множество можно задать в виде:

$$W = \{(\vec{x}, \vec{y})\colon \triangle(\vec{x}, \vec{y}) \geqslant S_{1-\alpha}\}$$

где $\alpha \in (0; 1)$ — заданный уровень значимости критерия; $S_{1-\alpha}$ — квантиль уровня $1-\alpha$ закона распределения статистики $\triangle$ при истинности $H_0$.

**Замечание 1.** *Построенный критерий называется критерием Смирнова.*

**Замечание 2.** *О законе распределения статистки $\triangle(\vec{X}, \vec{Y})$.*

1. *Доказано, что* при истинности $H_0$ *закон распределения статистики $\triangle$ не зависит от $F(t)$ — теоретического (т. е. «истинного» закона) закона, распределения случайной величины $X$ (поскольку $H_0$ предполагается истинной, $X$ и $Y$ имеет одинаковые распределения);*

2. *Для небольших $n_1$ и $n_2$ соответствующие распределения табулированы (т. е. составлены таблица их квантилей);*

3. *Смирнов доказал, что для $t > 0$*

$$P\{\sqrt{\frac{n_1 n_2}{n_1+n_2}}\triangle(\vec{X}, \vec{Y}) < t\} \xrightarrow[n_1\to\infty,\, n_2\to\infty]{} k(t)$$

   *где $k(t) = \sum_{k=-\infty}^{+\infty}(-1)^k e^{-2k^2t^2}$. Другими словами, при достаточно больших $n_1$ и $n_2$ можно считать, что случайная величина*

$$A = \sqrt{\frac{n_1 n_2}{n_1+n_2}} \cdot \triangle(\vec{X}, \vec{Y})$$

   *имеет своей функцией распределения $k(t)$, $t > 0$ (т. к. из определения статистики $\triangle$ вытекает, что $\triangle \geqslant 0$, то $A \geqslant 0 \implies F_A(t) \equiv 0$, $t \leqslant 0$).*

**Замечание 3.** *О вычислении значений статистики $\triangle$.*

1. *Рассмотрим вариационный ряд $z_{(1)}, \ldots, z_{n_1+n_2}$ объединённой выборки*

$$(X_1, \ldots, X_{n_1}, Y_1, \ldots, \ldots, Y_{n_2})$$

   *Можно показать, что*

$$\triangle(\vec{X}, \vec{Y}) = \max_{i=\overline{1;\, n}} |F_{n_1}(z_{(i)}) - G_{n_2}(z_{(i)})|$$

   *где $N = n_1 + n_2$.*

2. *Обозначим*

$$d_i = \begin{cases} 1, & z_{(i)} \text{ — одно из наблюдений случайной величины} X \\ 0, & z_{(i)} \text{ — -=-} \end{cases}$$

*(здесь $z_{(i)}$ — реализация $Z_{(i)}$).*

*Обозначим также*

$$S_j = \frac{j \cdot n_1}{N} - \sum_{k=1}^{j} d_k, \; j = \overline{1;\, N}$$

*Тогда*

$$\triangle(\vec{x},\, \vec{y}) = \frac{N}{n_1 n_2} \max\{S_1,\, \ldots,\, S_n\}$$

**Замечание.** *В этом разделе мы изучили некоторые критерии согласия, которые используются для проверки гипотезы о том, что закон распределения случайной величины $X$ имеет заданный вид.*

*Другую большую группу критериев для проверки непараметрических гипотез составляет так называемые критерии независимости. Они предназначены для решения следующей задачи:*

*Дано:*

1. $(X,\, Y)$ — *случайный вектор;*

2. $(\vec{X},\, \vec{Y})$ — *случайная выборка.*

*Требуется проверить гипотезу $H_0 = \{X \text{ и } Y \text{ независимые}\}$ против $H_1 = \neg H_0$.*

## 3.7 Элементы регрессионного анализа

### 3.7.1 Основные определения

**Пример.** *Пусть $Y_1$ — объём выпуска некоторым предприятием готовой продукции (млн. изделий в годы), а $Y_2$ — процент брака в этой продукции. Руководство предприятия считает, что величины $Y_1$ и $Y_2$ в основном зависят от следующих «параметров»:*

1. $X_1$ — *объём заработной платы (млн. руб.);*

2. $X_2$ — *объём вложений в социальный пакет (млн. руб.);*

3. $X_3$ — *объём вложений в подготовку/переподготовку (повышение квалификации кадров) (млн. руб.).*

*Для оптимизации производственного процесса руководство предприятия хочет знать зависимость величин $Y_1$, $Y_2$ от величин $X_1$, $X_2$, $X_3$, т. е. хочет знать зависимости $Y_j = Y_j(X_1,\, X_2,\, X_3)$, $j = \overline{1;\, 2}$.*

Особенностями данного примера являются следующие:

1. Величины $X_1$, $X_2$, $X_3$ являются детерминированными (т. е. не случайными);

2. Величины $Y_1$, $Y_2$ являются случайными.

Последнее объясняется тем, что на значения $Y_1$, $Y_2$ могут влиять неучтённые факторы (например, качество сырья, кол-во аварий, вложения в технику безопасности). Кроме этого, случайный характер величин $Y_1$ и $Y_2$ может быть связан с ошибками измерения (например, используется упрощённая схема контроля за качеством готовой продукции).

В этом случае связь между величинами $Y_1$, $Y_2$ и величинами $X_1$, $X_2$, $X_3$ является не функциональной, а стохастической (т. е. случайной).

**Определение.** *Говорят, что переменная $Y$ стохастически зависит от переменных $X_1, \ldots, X_p$, если на изменения значений переменных $X_1, \ldots, X_p$ величина $Y$ реагирует изменением своего закона распределения.*

**Пример.** *Пусть $Y$ — случайная величина, функция плотности распределения которой имеет вид*

$$f_Y(y) = \frac{1}{\sqrt{2\pi}X_2} e^{-\frac{(y-X_1)^2}{2X_2^2}}, \; y \in \mathbb{R}$$

*В этом случае $Y$ стохастически зависит от $X_1$, $X_2$.*

Задачи, связанные с установкой стохастических зависимостей между случайной величиной $Y$ и детерминизированными величинами $X_1, \ldots, X_p$, носящими количественный характер, составляет предмет изучения регрессионного анализа.

В регрессионном анализе используют модель чёрного ящика как наиболее общую модель отображения.

*(картинка с чёрным ящиком, в которую идут стрелки от $X_*$ и $\varepsilon_*$ и из которого выходят стрелки в $Y_*$)*

При этом используется терминология Ф-отображений, осуществляемых чёрным ящиком. Здесь $X_1, \ldots, X_p$ — факторы, $Y_1, \ldots, Y_s$ — отклики, $\varepsilon_1, \ldots, \varepsilon_m$ — случайные возмущения.

Вектор $\varepsilon_x = \vec{Y} - \Phi(\vec{X})$ называется вектором случайных ошибок.

### 3.7.2 Метод наименьших квадратов

Рассмотрим частный случай отображения Φ, в котором $p = m = s = 1$. Пусть у нас имеются результаты $n$ наблюдений. $t_1, \ldots, t_n$ — $n$ значений фактора $X$, а $y_1, \ldots, y_n$ — $n$ отвечающих им откликов. Тогда

$$\begin{cases} y_1 = \Phi(t_1) + \varepsilon_1 \\ \ldots \\ y_n = \Phi(t_n) + \varepsilon_n \end{cases}$$

где $\varepsilon_1$, ..., $\varepsilon_n$ — реализация случайной ошибки $\varepsilon$.

**Замечание 1.** *Здесь $\varepsilon$ — случайная величина, а $\varepsilon_1$, ..., $\varepsilon_n$ — её реализации.*

**Замечание 2.** *Рассматриваемая модель имеет вид $Y = \Phi(X) + \varepsilon$. В функции $\Phi$ нет ничего случайного, т. е. если известно значение $X$, то ему отвечает вполне определённое значение $\Phi(x)$.*

*При этом сама функция $\Phi$ неизвестна неизвестна. Требуется по имеющимся результатам наблюдений подобрать некоторую функцию $\hat{\Phi}$ так, чтобы она наилучшим (в некотором смысле) образом аппроксимировала функцию $\Phi(x)$.*

# Лекция №31, 20.05.2019

1. Часто функцию $\hat{\Phi}$ ищут в виде

$$\hat{\Phi}(t) = \theta_1\psi_1(t) + \ldots + \theta_p\psi_p(t) \tag{30}$$

где $\psi_j(t)$ — известные функции, которые часто называют базовыми; $\theta_j$, $j = \overline{1;\, p}$ — неизвестные параметры.

В этом случае задача тестирования построения отображения $\hat{\Phi}$ сводится к задаче подбора значений коэффициентов $\theta_1$, ..., $\theta_p$.

**Замечание 1.** *Модель 30 называют линейной по параметрам, т. к. каждый параметр $\theta_j$ входит в правую часть линейно.*

**Замечание 2.** *Задача выбора базисных функций $\psi_j(t)$, $j = \overline{1;\, n}$, является плохо формулированной и успех её рассмотрения часто зависит от интуиции и опыта исследователей.*

2. С учётом 30 результаты наблюдений можно записать в виде:

$$\begin{cases} y_1 = \theta_1\psi_1(t_1) + \ldots + \theta_p\psi_p(t_1) + \epsilon_1 \\ \ldots \\ y_n = \theta_1\psi_1(t_n) + \ldots + \theta_p\psi_p(t_n) + \epsilon_n \end{cases} \tag{31}$$

Используем обозначения:

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},\ \vec{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix},\ \vec{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix},\ \Psi = \begin{bmatrix} \psi 1(t_1) & \ldots & \psi_p(t_1) \\ \vdots & \ddots & \vdots \\ \psi_1(t_n) & \ldots & \psi_p(t_n) \end{bmatrix}$$

([47])

Чем более удачно подобран вектор $\vec{\theta}$, тем меньше будет значения функционала? $S(\vec{\theta})$.

[47] $\Psi$ — прописная буква «пси» греческого алфавита. — Прим. ред.

**Замечание.**

$$S(\vec{\theta}) = \begin{bmatrix} y_1 - \hat{\Phi}(t_1) \\ \vdots y_n - \hat{\Phi}(t_n) \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} y_1 - \hat{\Phi}(t_1) \\ \vdots y_n - \hat{\Phi}(t_n) \end{bmatrix} = (\vec{y} - \Psi\vec{\theta})^{\mathsf{T}}(\vec{y} - \Psi\vec{\theta}) = \|\vec{y} - \Psi\vec{\theta}\|^2$$

*где $\|\cdot\|$ — евклидова норма.*

**Определение.** *Оценкой, полученной по методу наименьших квадратов (МНК-оценкой) вектора $\vec{\theta}$ называют такое его значение $\hat{\vec{\theta}}$, которое доставляет наименьшее значение функционалу $S(\vec{\theta})$, т. е.*

$$S(\hat{\vec{\theta}}) = \min_{\vec{\theta}\in\mathbb{R}^p} S(\vec{\theta})$$

3. Как найти эту МНК-оценку?

**Теорема.** *Пусть* $\operatorname{rg}\Psi = p$. *([48]) Тогда* $\hat{\vec{\theta}} = (\psi^{\mathsf{T}}\psi)^{-1}\psi^{\mathsf{T}}\vec{y}$.

**Теорема.** ***О свойствах МНК-оценки.***

*Пусть*

(a) $\varepsilon \sim N(0, \sigma^2)$, $i = \overline{1;\, n}$;

(b) *Реализация случайной величины $\epsilon$ в серии из $n$ наблюдений независима.*

(c) $\operatorname{rg}\Psi = p$;

(d) $\hat{\theta} = (\psi^{\mathsf{T}}\psi)^{-1}\psi^{\mathsf{T}}\vec{y}$ — *МНК-оценка для $\vec{\theta}$.*

*Тогда*

(a) *$\hat{\theta}$ является несмещённой и состоятельной оценкой для $\vec{\theta}$ (из несмещённости следует $M[\hat{\vec{\theta}}] = \vec{\theta}$);*

(b) *$\vec{\theta} \sim N(\vec{\theta}, \Sigma)$ — нормальный случайный вектор, где $\Sigma = \sigma^2(\psi^{\mathsf{T}}\psi)^{-1}$;*

(c) *Интервальная оценка уровня $1-\alpha$ для параметра $\theta_j$ имеет вид $(\hat{\theta}_j - \triangle_j,\, \hat{\theta}_j + \triangle_j)$, где $\triangle_j = t_{1-\alpha}^{(n-p)}\sqrt{\frac{d_j}{n-p}S(\hat{\vec{\theta}})}$, $j = \overline{1;\, p}$; $t_{1-\alpha}^{(n-p)}$ — квантиль уровня $1-\alpha$ распределения $St(n-p)$, $d_j$ — $j$-ый элемент главной диагонали матрицы $(\psi^{\mathsf{T}}\psi)^{-1}$.*

[48] $\operatorname{rg} X$ — ранг матрицы $X$. —Прим. ред.

# A Комбинаторика

Пусть $X$ — некое множество. Для примеров определим $X = \{1, 2, 3, 4, 5, 6\}$.

## A.1 Сочетания без повторений

**Определение.** *Сочетанием без повторений из $n$ ($n = |X|$) элементов по $m$ называется любое неупорядоченное подмножество множества $X$, содержащее $m$ различных элементов.*

Кол-во таких подмножеств обозначается как $C_n^m$ и равно

$$C_n^m = \binom{n}{m} = \frac{n!}{m!(n-m)!}$$

## A.2 Размещения без повторений

**Определение.** *Размещением без повторений из $n$ элементов (исходного множества, $n = |X|$) по $m$ (длина кортежа) называется кортеж, состоящий из $m$ различных элементов множества $X$.*

**Примеры.** $(1, 2, 4)$, *но не* $(5, 5, 4)$.

Кол-во возможных размещений без повторений:

$$A_n^m = \frac{n!}{(n-m)!}$$

## A.3 Перестановки без повторений

Перестановки без повторений — крайний случай размещений без повторений.

**Определение.** *Перестановкой без повторений называют называют кортеж, состоящий из $n = |X|$ различных элементов множества $X$.*

Кол-во возможных перестановок без повторений:

$$P_n = A_n^n = n!$$

**Замечание.** *Три предыдущих понятия связаны как*

$$A_n^m = P_m C_n^m$$

## A.4 Размещения с повторениями

**Определение.** *Размещением с повторениями из* $n$ *(*$n = |X|$*) по* $m$ *элементов называется любой элемент из* $X^m = X \times X \times \ldots \times X$.

**Примеры.** $(1,\, 2,\, 3,\, 4,\, 5)$, $(1,\, 4,\, 4,\, 4,\, 2)$.

Кол-во возможных размещений с повторениями:

$$\widetilde{A}_n^m = n^m$$

## A.5 Перестановки с повторениями

**Определение.** *Перестановкой с повторениями называют кортеж длины* $n$ *из элементов множества* $X$*, в котором каждый элемент* $x_i \in X$ *повторяется* $n_i$, $i = \overline{1, k}$ *раз (*$\sum_i^k n_i = n$*).*

Кол-во возможных перестановок с повторениями:

$$P_n(n_1,\, n_2,\, \ldots,\, n_k) = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!},\ n = n_1 + n_2 + \ldots + n_k$$

# B Заметки редактора

## B.1 Рекомендуемая литература

Для изучения Теории Вероятностей, параллельно с этими лекциями, можно читать:

- Очевидно, *«Теория Вероятностей»* из серии *«Математика в Техническом Университете»* (номер книги $XVI$). Данные лекции являются де-факто сокращённой версией этой книги;

- Саркисян П. С. (семинарист) рекомендует: *An Introduction to Probability Theory and its applications by William Feller*. Очень много хороших примеров... и опечаток;

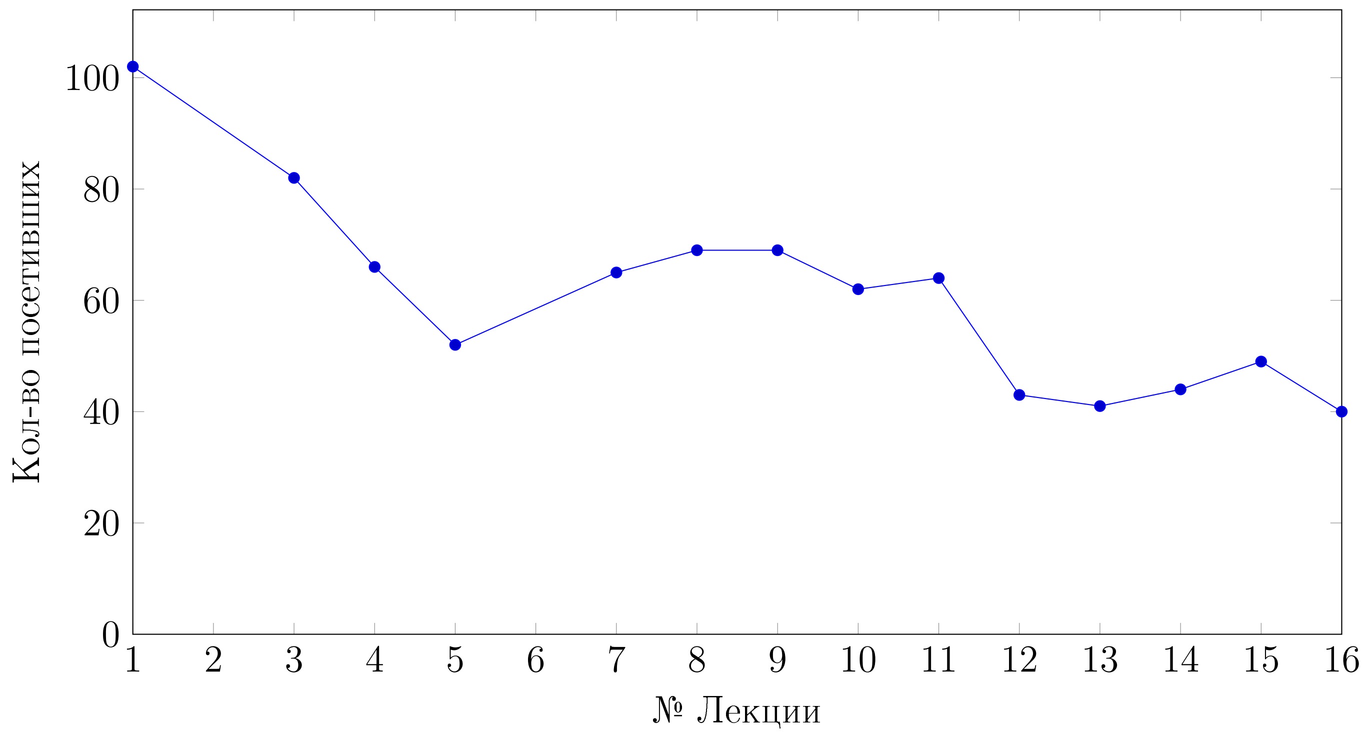
- *A First Course in Probability by Sheldon M. Ross*, части 1-8. Курс примерно совпадает;

- Уровень «Хардкор», или книга для тех, кто хочет изучать *современную* теорию вероятностей: *Probability Theory: The Logic of Science by E. T. Jaynes.* Книга весьма слабо совпадает с описанным курсом: например, понятие вероятности в ней не строится на аксиомах Колмогорова, а выводится как численная функция степени правдоподобия утверждения.

И, да: восславь ~~солнце~~ Library Genesis (и ещё одна полезная ссылка).

# C Статистические данные о средней посещаемости

## C.1 Осенний семестр




## C.2 Весенний семестр