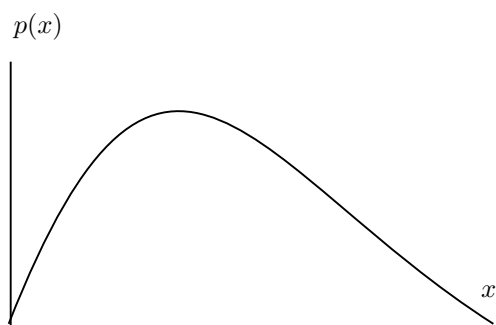


# 1 Занятие 1

## 1.1 Основные распределения в Мат Стат

### 1.1.1 Гамма распределение

$$\xi \sim \Gamma(\lambda, a) \quad \lambda > 0, a > 0$$
$$P(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \quad \{(0, +\infty)\}$$



$$\Gamma(S+1) = S\Gamma(S) \quad \Gamma(S) = \int_0^\infty x^{S-1} e^{-x} dx \quad x > 0$$
$$M[\Gamma] = \int_{-\infty}^\infty \rho(x) dx = \int_0^\infty \frac{\lambda^a}{\Gamma(a)} \left(\frac{t}{\lambda}\right)^a e^{-t} \frac{dt}{\lambda} = \frac{1}{\lambda \Gamma(a)} \int_0^\infty t^a e^{-t} dt = \frac{a}{\lambda}$$
$$D[\xi] = M[\xi^2] - M^2[\xi] = \frac{a^2 + a}{\lambda^2} - \left(\frac{a}{\lambda}\right)^2 = \frac{a}{\lambda^2}$$

**Теорема 1.1** (Свойство суммы).  $\xi_1, \dots, \xi_n$  независимы,  $\xi_i \sim \Gamma(\lambda, a_i)$ ,  
 $\eta = \xi_1 + \dots + \xi_n \sim \Gamma(\lambda, a_1 + \dots + a_n)$

*Доказательство.*  $\xi_1 \sim \Gamma(\lambda, a_1)$ .  $\xi_2 \sim \Gamma(\lambda, a_2)$  - независимые,  $\eta = \xi_1 + \xi_2$

$$\begin{aligned}\Phi(y) &= P(\eta < y) = P(\xi_1 + \xi_2 < y) = \iint_{x_1 + x_2 < y} p(x_1, x_2) dx_1 dx_2 = \\ &= \int_0^y dx_2 \int_0^{y-x_2} \frac{\lambda^{a_1}}{\Gamma(a_1) x_1^{a_1-1} e^{-\lambda x_1}} \frac{\lambda^{a_2}}{\Gamma(a_2) x_2^{a_2-1} e^{-\lambda x_2}} dx_1 \\ \varphi(y) &= \Phi'(y)\end{aligned}$$

□

### 1.1.2 Распределение Парсона $\chi^2$

$\xi_i \sim N(0, 1)$  - независимы,  $\eta = \xi_1^2 + \dots + \xi_n^2 = \chi^2$

$$\begin{aligned}\Phi(y) &= P(\xi^2 < y) = \begin{cases} y \leq 0 & : 0 \\ y > 0 & : P(-\sqrt{y} < \xi < \sqrt{y}) \end{cases} \\ \varphi(x) &= \begin{cases} \frac{1}{2\sqrt{y}} F'(\sqrt{y}) + \frac{1}{2\sqrt{y}} F'(-\sqrt{y}), & y > 0 \\ 0, & y < 0 \end{cases}\end{aligned}$$

$$p(x) = \frac{e^{x^2/2}}{\sqrt{2\pi}}$$

$$\varphi(y) = \begin{cases} \frac{1}{\sqrt{y}} \frac{e^{-y/2}}{\sqrt{2\pi}}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

$$p(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \{(0; +\infty)\} \quad \lambda = \frac{1}{2} \quad a = \frac{1}{2}$$

$$\xi^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right) \quad \xi_1^2 + \dots + \xi_n^2 \sim \Gamma\left(\frac{1}{2}, \frac{n}{2}\right) = \chi^2(n)$$

$n$  - число степеней свободы

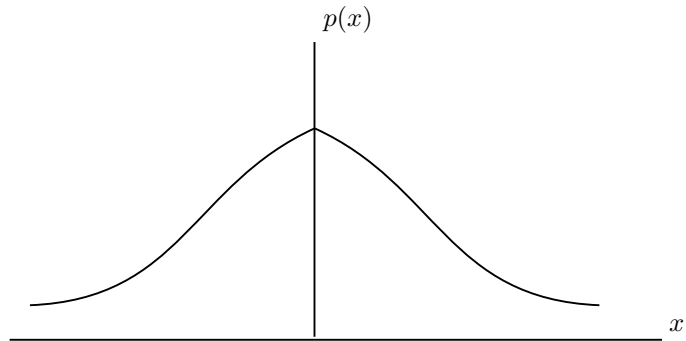
$$M[\eta] = \frac{a}{\lambda} = \frac{n/2}{1/2} = n$$

$$D[\eta] = \frac{a}{\lambda^2} = \frac{n/2}{1/4} = 2n$$

**Теорема 1.2** (Свойство суммы).  $\xi_1, \dots, \xi_m$  - независ,  $\xi_i \sim \chi^2(n_i)$ ,  $\xi_1 + \dots + \xi_n \sim \chi^2(n_1 + \dots + n_m)$

### 1.2 Распределение Стьюдента (Госсет)

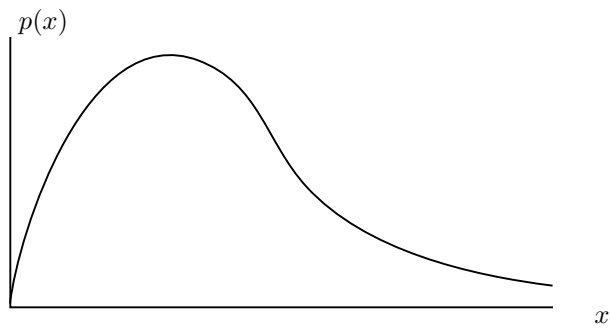
$\xi \sim N(0, 1)$ ,  $\eta \sim \chi^2(m)$  - независимы,  $\frac{\xi}{\sqrt{\eta/m}} \sim t(m)$



$$p(x) = \frac{(m)^{m/2} \Gamma(\frac{m+1}{2})}{\sqrt{\pi} \Gamma(\frac{m}{2}) (x^2 + m)^{\frac{m+1}{2}}}$$

### 1.3 Распределение Фишера

$\xi \sim \chi^2(n)$ ,  $\eta \sim \chi^2(m)$  - независимые,  $\frac{\xi/n}{\eta/m} \sim F(n, m)$



### 1.4 Нормальное распределение

$$p(\vec{x}) = \frac{1}{(\sqrt{2\pi})^n} \frac{1}{\sqrt{\det K}} e^{-\frac{1}{2}(\vec{x}-\vec{a})^T K^{-1}(\vec{x}-\vec{a})}$$

$$\vec{\xi} \sim N(\vec{a}, R)$$

Свойства:

- $\xi \sim N(0, 1), \eta = a\xi + b \sim N(b, a^2)$
- $\xi \sim N(\alpha, \sigma^2), \eta = a\xi + b \sim N(a\alpha + b, \sigma^2 a^2)$
- $\xi \sim N(\vec{0}, E), \vec{\eta} = A\vec{\xi} + \vec{b}, A : n \times n, \det A \neq 0$

$$\begin{aligned}
\Phi(t_1, \dots, t_n) &= P(\eta_1 < t_1, \dots, \eta_n < t_n) = P(\vec{\eta} < \vec{t}) = P(A\vec{\xi} + \vec{b} < \vec{t}) = \\
&= \int \dots \int_{A\vec{x} + \vec{b} < \vec{t}} p(x_1, \dots, x_n) dx_1 \dots dx_n = \\
&\vec{y} = A\vec{x} + \vec{b} \quad J = \left| \frac{\partial \vec{x}}{\partial \vec{y}} \right| \quad \frac{1}{J} = \det A \\
&= \int \dots \int_{\vec{y} < \vec{t}} p(A^{-1}(\vec{y} - \vec{b})) \frac{1}{|\det A|} dy_1 \dots dy_n \\
&\quad \varphi(\vec{t}) = p(A^{-1}(\vec{y} - \vec{b})) \frac{1}{|\det A|} \\
\varphi(\vec{t}) &= \frac{1}{|\det A|} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(A^{-1}(\vec{y} - \vec{b}))^T (A^{-1}(\vec{y} - \vec{b}))} = \\
&= \frac{1}{|\det A|} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(\vec{t} - \vec{b})^T (A^T)^{-1} A^{-1}(\vec{t} - \vec{b})} \\
K &= AA^T \quad \vec{\eta} = A\vec{\xi} + \vec{b} \sim N(\vec{b}, AA^T)
\end{aligned}$$

- $\xi \sim N(\vec{a}, K), \vec{\eta} = A\vec{\xi} + \vec{b} \sim N(A\vec{a} + \vec{b}, AK A^T), A : n \times n, \det A \neq 0$
- Для  $A : m \times n$  два предыдущих свойства так же верны
- $\xi, \eta$  - независ  $\Rightarrow \text{cov}(\xi, \eta) = 0$ , в другую сторону не верно

$$\begin{cases} \xi \sim N(a_1, \sigma_1^2) \\ \eta \sim N(a_2, \sigma_2^2) \\ \text{независимые} \end{cases} \Leftrightarrow (\xi, \eta) \sim N \left( \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right)$$

$$\begin{cases} \xi \sim N(a_1, \sigma_1^2) \\ \eta \sim N(a_2, \sigma_2^2) \\ \text{cov}(\xi, \eta) = 0 \end{cases} \Leftarrow (\xi, \eta) \sim N \left( \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right)$$

**Лемма 1.1** (Лемма Фишера). Пусть  $\vec{\xi} \sim N(\vec{0}, E)$  и  $C$  ортогональная матрица,  $\vec{\eta} = C\vec{\xi}$ , тогда  $\forall k = 1 \dots n-1$  сл. вел.  $\varkappa = \sum_{i=1}^n \xi_i^2 - \eta_1^2 - \eta_2^2 - \dots - \eta_k^2 \sim \chi^2(n-k)$  и вел  $\varkappa, \eta_1, \eta_2, \dots, \eta_k$  независ.

Доказательство.

$$\begin{aligned}\vec{\eta} &\sim N(\vec{0}, \underbrace{CC^T}_E) \\ \eta_1^2 + \dots + \eta_n^2 &= \vec{\eta}^T \vec{\eta} = \vec{\xi}^T C^T C \vec{\xi} = \xi_1^2 + \dots + \xi_n^2 \\ \varkappa &= \eta_{k-1}^2 + \dots + \eta_n^2 \\ \varkappa &= \chi^2(n-k)\end{aligned}$$

□

**Теорема 1.3 (Фишера).** Пусть  $\xi_1, \dots, \xi_n$  независ и  $\xi_i \sim N(a, \sigma^2)$ , тогда:

1.  $\varphi = \sqrt{n} \frac{\bar{\xi} - a}{\sigma} \sim N(0, 1)$ ,  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$
2.  $\psi = \sum_{i=1}^n \frac{(\xi_i - \bar{\xi})^2}{\sigma^2} \sim \chi^2(n-1)$
3.  $\varphi$  и  $\psi$  независ.

Доказательство.

$$\begin{aligned}\varphi &= \frac{1}{\sqrt{n}} \frac{\sum \xi_i - na}{\sigma} = \frac{1}{\sqrt{n}} \sum \left( \frac{\xi_i - a}{\sigma} \right) \\ \frac{\xi_i - a}{\sigma} &= \frac{1}{\sigma} \xi_i - \frac{a}{\sigma} \sim N\left(\frac{a}{\sigma} - \frac{a}{\sigma}, \sigma^2 \frac{1}{\sigma^2}\right) = N(0, 1) \\ \varphi &= \frac{1}{\sqrt{n}} \sum \eta_i = \left( \frac{1}{\sqrt{n}} \dots \frac{1}{\sqrt{n}} \right) \vec{\eta} \sim N(\vec{0}, AA^T) = N(0, 1)\end{aligned}$$

1) Доказан

$$\begin{aligned}\psi &= \sum \left( \underbrace{\frac{\xi_i - a}{\sigma}}_{\eta_i \sim N(0,1)} - \underbrace{\frac{\bar{\xi} - a}{\sigma}}_{\bar{\eta}} \right)^2 = \sum (\eta_i - \bar{\eta})^2 = \sum (\eta_i^2 - 2\eta_i \bar{\eta} + (\bar{\eta})^2) = \\ &= \sum \eta_i^2 - 2\bar{\eta} \sum \eta_i + n(\bar{\eta})^2 = \sum \eta_i^2 - n(\bar{\eta})^2 \\ \eta_i &\sim N(0, 1) \quad \zeta^2 = n\bar{\eta}^2 \quad \zeta = \sqrt{n}\bar{\eta} = \frac{1}{\sqrt{n}} \sum \eta_i = A\vec{\eta} = \varphi\end{aligned}$$

$A = \left( \frac{1}{\sqrt{n}} \dots \frac{1}{\sqrt{n}} \right) \Rightarrow C$  - ортог. матрица (Грамма-Шмидта)

( $A$  получается строчкой матрицы  $C$  и тогда  $\zeta$  - одна из координат в другом базисе и применима Лемма Фишера)

По лемме Фишера  $\psi \sim \chi^2(n-1)$ ,  $\psi$  и  $A\vec{\eta}$  независ

□

**Теорема 1.4** (О проекции). Пусть  $\vec{\xi} \sim N(\vec{0}, \sigma^2 E)$ ,  $L_1 : \dim L_1 = m_1$  и  $L_2 : \dim L_2 = m_2$  два ортогональных подпространства  $\mathbb{R}^n$ ,  $\vec{\eta}_1$  - проекция  $\vec{\xi}$  на  $L_1$ , норм. распр., независ. и  $\frac{|\eta_1|^2}{\sigma^2} \sim \chi^2(\dim L_1)$ ,  $\frac{|\eta_2|^2}{\sigma^2} \sim \chi^2(\dim L_2)$

*Доказательство.*  $\vec{\eta}_1 = A_1 \vec{\xi} \sim N(\dots, \dots)$ ,  $\vec{\zeta} = C \vec{\xi}$ ,  $C$  - ортогональная.  $\vec{\zeta} \sim N(\vec{0}, C \sigma^2 E C^T) = N(\vec{0}, \sigma^2 E)$ . Новый ортонормированный базис  $e'_1 \dots e'_m$  в  $L_1$ ,  $e'_{m+1} \dots e'_n$  в  $L_2$ ,  $\vec{\eta}_1 = \zeta_1 e'_1 + \dots + \zeta_m e'_m$ ,  $\vec{\eta}_2 = \zeta_{m+1} e'_{m+1} + \dots + \zeta_n e'_n$

$$\frac{\bar{\xi}}{\sigma} \sim N(\vec{0}, E) \quad \frac{|\eta_1|^2}{\sigma^2} = \sum \frac{\xi_i^2}{\sigma^2} \sim \chi^2(m_1)$$

□

## 2 Порядковые случайные величины

$\xi_1, \xi_2, \dots, \xi_n$  независ.,  $\xi_i \sim F(x)$

$$\eta = \min(\xi_1, \dots, \xi_n) \sim? \quad \zeta = \max(\xi_1, \dots, \xi_n) \sim?$$

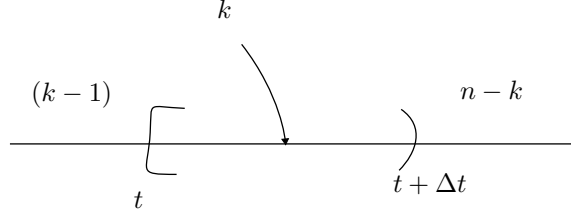
$$\begin{aligned} \Phi(y) &= P(\eta < y) = 1 - P(\eta \geq y) = 1 - P(\min(\xi_1, \dots, \xi_n) \geq y) = \\ &= 1 - P(\xi_1 \geq y, \dots, \xi_n \geq y) = 1 - \prod_{i=1}^n P(\xi_i \geq y) = \\ &= 1 - \prod_{i=1}^n (1 - P(\xi_i < y)) = 1 - (1 - F(y))^n \end{aligned}$$

$$\begin{aligned} \Psi(z) &= P(\zeta < z) = P(\max(\xi_1, \dots, \xi_n) < z) = \\ &= P(\xi_1 < z, \dots, \xi_n < z) = \prod_{i=1}^n P(\xi_i < z) = (F(z))^n \end{aligned}$$

Порядковые величины:

$$\begin{aligned} \xi_{(1)} &= \min(\xi_1, \dots, \xi_n) \\ \xi_{(2)} &= \min(\xi_i : \xi_i \neq \xi_{(1)}) \\ \xi_{(3)} &= \min(\xi_i : \xi_i \neq \xi_{(1)}, \xi_i \neq \xi_{(2)}) \\ &\dots \\ \xi_{(n)} &= \max(\xi_1, \dots, \xi_n) \end{aligned}$$

Положим  $F(x)$  - непрерывна:



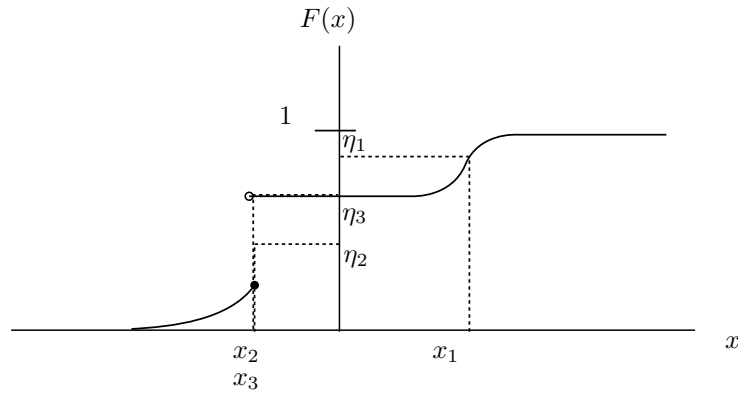
$$\begin{aligned}
P(t \leq \xi_{(k)} < t + \Delta t) &= \\
&= nP(t \leq \xi < t + \Delta t) C_{n-1}^{k-1} (P(\xi < t))^{k-1} (P(\xi \geq t + \Delta t))^{n-k} C_{n-k}^{n-k} = \\
&= \varkappa(t + \Delta t) - \varkappa(t) \\
n \frac{F(t + \Delta t) - F(t)}{\Delta t} C_{n-1}^{k-1} (F(t))^{k-1} (1 - F(t + \Delta t))^{n-k} &= \frac{\varkappa(t + \Delta t) - \varkappa(t)}{\Delta t} \\
\varkappa(t) &= np(t) C_{n-1}^{k-1} (1 - F(t))^{n-k} (F(t))^{k-1}
\end{aligned}$$

$\varkappa(t)$  - плотность распределения  $\xi_{(k)}$ ,  $p(t) = F'(t)$

$\xi_{(1)}$  и  $\xi_{(n)}$  совместное распр

$$\begin{aligned}
G(y, z) &= P(\xi_{(1)} < y, \xi_{(n)} < z) \\
P(\xi_{(n)} < z) &= P(\xi_{(1)} < y, \xi_{(n)} < z) + P(\xi_{(1)} \geq y, \xi_{(n)} < z) \\
\Psi(z) &= (F(z))^n = P(\xi_{(1)} < y, \xi_{(n)} < z) + \prod_{i=1}^n \underbrace{P(y \leq \xi_i < z)}_{F(z) - F(y)} \\
G(y, z) &= P(\xi_{(1)} < y, \xi_{(n)} < z) = \begin{cases} (F(z))^n, & y > z \\ (F(z))^n - (F(z) - F(y))^n, & y \leq z \end{cases}
\end{aligned}$$

### 3 Моделирование случайных величин



$\xi \sim F(x)$ ,  $\eta \sim R(0,1)$ ,  $F(x) = \eta_1 \rightarrow x_1$ , для псевдослучайных чисел вихрь Мерсенна

### 4 Основные задачи статистики

Явление  $\rightarrow$  математическая модель явления  $\rightarrow$  вероятностная модель явления  $\xi_i$ .

Выборка -  $n$  наблюдений над явлением  $\rightarrow$  описательная статистика (непараметрическая), выбор классов (e.g.  $\varepsilon \sim N(a, \sigma^2)$ )  $\rightarrow$  параметрическая статистика (e.g.  $\varepsilon \sim N(a, \sigma^2)$ ,  $a-?$ ,  $b-?$ )

**Пример.** Пытаемся понять как остывает чашка чая.

$$\frac{dT}{dt} = k(T - T_0) + \varepsilon$$

Вероятностная модель явления с двумя случайными величинами (погрешности измерений  $\varepsilon$  и внутри  $k$ ).

Распределения полагаем нормальными и так далее.

1. Определение параметров и оценка их точности
2. Проверка статистических гипотез

Характеристики модели  $\theta$ , по выборке оценка  $\bar{\theta}(\vec{x}_n)$ ,  $n$  - объём выборки.

Статистика -  $\forall$  борелевская функция от  $\vec{x}_n$  (борелевская  $g : \mathbb{R} \rightarrow \mathbb{R} \forall B \in \mathbb{B} \hookrightarrow g^{-1}(B) \in \mathbb{B}$ ).

Воспринимаем  $\vec{x}_n$  с двух сторон:

1. конкретные наблюдения над явлением



2. независимые случайные величины с распределением, одинаковым со случайными величинами в вероятностной модели

#### 4.1 Свойства оценок

$\Theta$  - множество значений  $\theta$ ,  $\theta \in \Theta$ ,  $\tilde{\theta}(\vec{x}_n)$  - оценка  $\theta$  по выборке

1. несмещённость  $\forall \theta \in \Theta \hookrightarrow M[\tilde{\theta}(\vec{x}_n)] = \theta$
2. состоятельность  $\forall \theta \in \Theta \hookrightarrow \tilde{\theta} \xrightarrow{P} \theta$  (i.e.  $\forall \varepsilon > 0 \ P(|\tilde{\theta} - \theta| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0$ )
3. сравнение оценок  $\tilde{\theta}_1$  эффективнее  $\tilde{\theta}_2$ , если  $\forall \theta \in \Theta \hookrightarrow D[\tilde{\theta}_1] \leq D[\tilde{\theta}_2]$  и  $\exists \theta \in \Theta : D[\tilde{\theta}_1] < D[\tilde{\theta}_2]$

**Теорема 4.1** (Достаточное условие состоятельности). Если  $\tilde{\theta}$  является несмещённой оценкой  $\theta$  и  $D[\tilde{\theta}] \xrightarrow{n \rightarrow \infty} 0$ , то  $\tilde{\theta}$  является состоятельной оценкой ( $\forall \theta \in \Theta$ )

*Доказательство.*  $M[\tilde{\theta}] = \theta$ , по неравенству Чебышёва:

$$\forall \varepsilon > 0 \hookrightarrow P(|\tilde{\theta} - \underbrace{M[\tilde{\theta}]}_{\theta}| < \varepsilon) \geq 1 - \frac{D[\tilde{\theta}]}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 1$$

□

---

**Задача (T1).** Пытаемся понять по двум серийным номерам сколько всего танков.

$\xi \sim R(0, \theta)$ ,  $\theta > 0$  вер. модель.,  $\vec{x}_n$  - выборка объёмом  $n$

$$\tilde{\theta}_1 = 2\bar{x} = 2 \frac{1}{n} \sum_{i=1}^n x_i$$

$$\tilde{\theta}_2 = \min x_i$$

$$\tilde{\theta}_3 = \max x_i$$

$$\tilde{\theta}_4 = x_1 + \frac{1}{(n-1)} \sum_{i=2}^n x_i$$

$$M[\xi] = \int_{-\infty}^{\infty} x p(x) dx = \int_0^{\theta} \frac{x}{\theta} dx = \frac{\theta}{2}$$

$$M[\xi^2] = \int_{-\infty}^{\infty} x^2 p(x) dx = \int_0^{\theta} \frac{x^2}{\theta} dx = \frac{\theta^2}{3}$$

Рассматриваем  $\tilde{\theta}_1$ :

Несмещённость  $\forall \theta > 0 M[\tilde{\theta}] = \theta$ :

$$M[\frac{\theta}{n} \sum_{i=1}^n x_i] = \frac{2}{n} \sum_{i=1}^n M[x_i] = 2M[\xi] = \theta \text{ несмещённая}$$

$$D[\tilde{\theta}_1] = D[\frac{2}{n} \sum x_i] = \frac{1}{n^2} \sum D[x_i] = \frac{4}{n} D[\xi] = \frac{\theta^2}{3n} \xrightarrow{n \rightarrow \infty} 0, \text{ по достаточному условию оценка состоятельная}$$

Рассматриваем  $\tilde{\theta}_2$ :

$$\begin{aligned} M[\tilde{\theta}_2] &= \int_{-\infty}^{\infty} y \varphi(y) dy \\ \Phi(y) &= 1 - (1 - F(y))^n \quad \varphi(y) = n(1 - F(y))^{n-1} p(y) \\ M[\tilde{\theta}_2] &= \int_0^{\theta} n(1 - \frac{y}{\theta})^{n-1} \frac{1}{\theta} y dy = \\ &\quad t = 1 - \frac{y}{\theta} \\ &= - \int_1^0 n t^{n-1} (1 - t) \theta dt = \int_0^1 n \theta t^{n-1} dt - \int_0^1 n \theta t^n dt = \\ &= n \theta [1 - \frac{n}{n+1}] = \frac{\theta}{n+1} \quad \text{смещённая} \\ \tilde{\theta}_2' &= (n+1) x_{min} = (n+1) \tilde{\theta}_2 \quad \text{несмещённая} \\ M[\tilde{\theta}_2^2] &= \int_0^{\theta} n(1 - \frac{y}{\theta})^{n-1} \frac{1}{\theta} y^2 dy = \\ &= - \int_1^0 n t^{n-1} (1 - t)^2 \theta^2 dt = \frac{2\theta^2}{(n+1)(n+2)} \\ D[\tilde{\theta}_2] &= \frac{2\theta^2}{(n+1)(n+2)} - \frac{\theta^2}{(n+1)^2} = \theta^2 \left[ \frac{2(n+1) - (n+2)}{(n+1)^2(n+2)} \right] = \\ &= \theta^2 \left[ \frac{n}{(n+1)^2(n+2)} \right] \xrightarrow{n \rightarrow \infty} 0 \quad \text{достаточное не выполняется} \\ D[\tilde{\theta}_2'] &= (n+1)^2 D[\tilde{\theta}_2] = \frac{\theta^2 n}{n+2} \not\rightarrow 0 \end{aligned}$$

Смотрим состоятельность  $\tilde{\theta}_2'$  по определению

$$\forall \theta > 0 \forall \varepsilon > 0 \Leftrightarrow P(|\tilde{\theta}_2' - \theta| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0$$

$$\begin{aligned} P(|\tilde{\theta}_2' - \theta| \geq \varepsilon) &\geq P(\tilde{\theta}_2' > \theta + \varepsilon) = \\ &= P((n+1)x_{min} \geq \theta + \varepsilon) = P(x_{min} \geq \frac{\theta + \varepsilon}{n+1}) = \\ &= 1 - P(x_{min} < \frac{\theta + \varepsilon}{n+1}) = 1 - (1 - (1 - F(\frac{\theta + \varepsilon}{n+1}))^n) = \\ &= (1 - (\frac{\theta + \varepsilon}{\theta(n+1)}))^n \xrightarrow{n \rightarrow \infty} e^{-\frac{\theta + \varepsilon}{\theta}} > 0 \end{aligned}$$

Не является состоятельной!

Смотрим состоятельность  $\tilde{\theta}_2$  по определению:

$$P(\tilde{\theta}_2 < \theta - \varepsilon) + \underbrace{P(\tilde{\theta}_2 > \theta + \varepsilon)}_{=0, \text{ т.к. } \tilde{\theta}_2 = x_{\min}} \\ P(x_{\min} < \theta - \varepsilon = \Phi(\theta - \varepsilon)) = 1 - (1 - \frac{\theta - \varepsilon}{\theta})^n = 1 - \left(\frac{\varepsilon}{\theta}\right)^n \xrightarrow{n \rightarrow \infty} 1$$

Не является состоятельной!

Рассматриваем  $\tilde{\theta}_3 = x_{\max}$ :

$$M[\tilde{\theta}_3] = \int_{-\infty}^{+\infty} z\psi(z)dz = \int_0^\theta n \frac{z^n}{\theta^n} dz = \frac{n}{n+1}\theta \quad \text{смешённая} \\ \Psi(z) = (F(z))^n \quad \psi(z) = n(F(z))^{n-1}p(z) = n\left(\frac{z}{\theta}\right)^{n-1} \frac{1}{\theta} \{(0; \theta)\} \\ D[\tilde{\theta}_3] \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n\theta^2}{(n+2)(n+1)^2} \\ D[\tilde{\theta}_3] \frac{(n+1)^2}{n^2} D[\tilde{\theta}_3] = \frac{\theta^2}{n(n+2)} \xrightarrow{n \rightarrow \infty} 0 \quad \text{состоятельная}$$

Смотрим состоятельность  $\tilde{\theta}_2'$  по определению

$$\forall \theta > 0 \quad \forall \varepsilon > 0 \\ P(|\tilde{\theta}_2' - \theta| \geq \varepsilon) = P(x_{\max} < \theta - \varepsilon) + \underbrace{P(x_{\max} \geq \theta + \varepsilon)}_{=0} = \\ = (F(\theta - \varepsilon))^n = \begin{cases} 0 < \varepsilon < \theta : \left(\frac{\theta - \varepsilon}{\theta}\right)^n \xrightarrow{n \rightarrow \infty} 0 \\ \varepsilon \geq \theta : (0)^n \xrightarrow{n \rightarrow \infty} 0 \end{cases}$$

Является состоятельной!

Рассматриваем  $\tilde{\theta}_4$ :

$$M[\tilde{\theta}_4] = M[x_1 + \sum_{i=2}^n x_i] = M[x_1] + \frac{1}{n-1} \sum_{i=1}^n M[x_i] = \frac{\theta}{2} + \frac{\theta}{2} = \theta \\ D[\tilde{\theta}_4] = D[\xi] + \frac{1}{(n-1)^2} (n-1) D[\xi] = \frac{\theta^2}{12} \frac{n}{n-1} \not\xrightarrow{n \rightarrow \infty} 0$$

Достаточное усл. не работает.

Используем теорему  $\xi_n \xrightarrow{P} \xi$ ,  $\eta_n \xrightarrow{P} \eta$ ,  $\xi_n + \eta_n \xrightarrow{P} \xi + \eta$ .

И ЗБЧ Хинчина:  $\xi_1, \dots, \xi_n$  незав., одинак распр.  $\Rightarrow \frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow{P} M[\xi]$ .

$$\begin{aligned} x_1 &\xrightarrow{P} x_1 & \frac{1}{n-1} \sum_{i=2}^n x_i &\xrightarrow{P} \frac{\theta}{2} \\ \tilde{\theta}_4 &\xrightarrow{P} x_1 + \frac{\theta}{2} \end{aligned}$$

Не состоятельна!

Адекватные остались  $\tilde{\theta}_1 = 2\bar{x}$ ,  $\tilde{\theta}_3' = \frac{n+1}{n} x_{max}$

$$D[\tilde{\theta}_1] = \frac{\theta^2}{3n} > D[\tilde{\theta}_3'] = \frac{\theta^2}{n(n+2)}$$

Лучшая оценка  $\tilde{\theta}_3$ .

## 5 Оптимальность и эффективность оценок

**Определение 5.1.** Несмещённая оценка  $\tilde{\theta}(\vec{x}_n)$  характеристики  $\theta$  называется оптимальной  $\tilde{\theta}_{opt}$  если для  $\forall \theta \in \Theta \Rightarrow D[\tilde{\theta}_{opt}] = \inf D[\tilde{\theta}]$ ,  $\inf$  по всем несмещённым оценкам  $\theta$ .

**Теорема 5.1** (Единственность оптимальной оценки). Если оптимальная оценка существует, то она единственна.

*Доказательство.* Пусть  $\tilde{\theta}_1$  и  $\tilde{\theta}_2$  разные оптимальные оценки

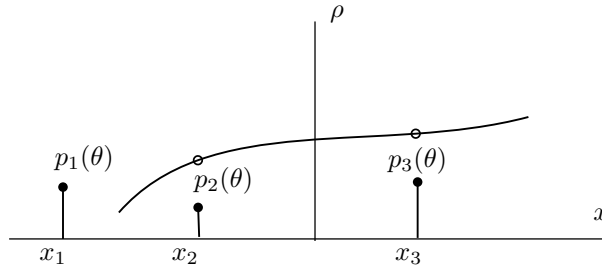
$$\begin{aligned} \tilde{\theta}_3 &= \frac{\tilde{\theta}_1 + \tilde{\theta}_2}{2} & M[\tilde{\theta}_3] &= \theta \\ D[\tilde{\theta}_3] &= \frac{1}{4}D[\tilde{\theta}_1] + D[\tilde{\theta}_2] + \frac{1}{2}cov(\tilde{\theta}_1, \tilde{\theta}_2) = \frac{1}{2}D[\tilde{\theta}_1] + \frac{1}{2}cos(\tilde{\theta}_1, \tilde{\theta}_2) \\ D[a\xi + b\eta] &= a^2D[\xi] + b^2D[\eta] + 2abcov(\xi, \eta) \\ |cov(\tilde{\theta}_1, \tilde{\theta}_2)| &\leq \sqrt{D\tilde{\theta}_1 D\tilde{\theta}_2} = D[\tilde{\theta}_1] \\ D[\tilde{\theta}_3] &\leq D[\tilde{\theta}_1] & D[\tilde{\theta}_3] &= D[\tilde{\theta}_1] \\ cov(\tilde{\theta}_1, \tilde{\theta}_2) &= D[\tilde{\theta}_1] & \Rightarrow & r = 1 \Leftrightarrow \tilde{\theta}_1 = a\tilde{\theta}_2 + b \\ M[\tilde{\theta}_1] &= M[\tilde{\theta}_2] = \theta & D[\tilde{\theta}_1] &= D[\tilde{\theta}_2] \\ \begin{cases} \theta = a\theta + b \\ a^2D[\tilde{\theta}_2] = D[\tilde{\theta}_2] \end{cases} &\Rightarrow \begin{cases} a = 1 \\ b = 0 \end{cases} \\ &\Rightarrow \tilde{\theta}_1 = \tilde{\theta}_2 \end{aligned}$$

Противоречие.

□

Будем рассматривать параметрические вероятностные модели:

$$\begin{aligned}\xi &\sim \rho(x, \theta), \theta \in \Theta \subset \mathbb{R}, x \in A(\theta) \\ \xi &\sim \rho(x, \vec{\theta}), \vec{\theta} \in \Theta \subset \mathbb{R}^m, x \in A(\vec{\theta}) \\ \rho(x, \theta) &= \underbrace{p(x, \theta)\{E\}}_{\text{непр. часть}} + \underbrace{\sum_k p_k(\theta)\{x_k\}}_{\text{дискр. часть}}\end{aligned}$$



## 5.1 Информация Фишера

$$\begin{aligned}I(\theta) &= M \left[ \left( \frac{\partial \ln \rho(x, \theta)}{\partial \theta} \right)^2 \right] = \\ &= \int_E \left( \frac{\partial \ln p(x, \theta)}{\partial \theta} \right)^2 p(x, \theta) dx + \sum_k \left( \frac{\partial \ln p_k(x, \theta)}{\partial \theta} \right)^2 p_k(\theta)\end{aligned}$$

$I(\vec{\theta})$  - информационная матрица Фишера

$$I_{ij}(\vec{\theta}) = M \left[ \frac{\partial \ln p(x, \theta)}{\partial \theta_i} \frac{\partial \ln p(x, \theta)}{\partial \theta_j} \right]$$

**Определение 5.2.** Вероятностная модель  $\xi \sim \zeta(x, \theta)$ ,  $\theta \in \Theta \subset \mathbb{R}$ ,  $x \in A$  называется регулярной, если

1.  $\rho(x, \theta)$  непр дифф по  $\theta$  на  $\Theta$
2.  $\frac{\partial}{\partial \theta} \int_A \rho(x, \theta) dx = \int_A \frac{\partial}{\partial \theta} \rho(x, \theta) dx$  на  $\Theta$
3.  $I(\theta)$  непр на  $\Theta$  и  $I(\theta) > 0$  на  $\Theta$

**Определение 5.3.** Вероятностная модель  $\xi \sim \zeta(x, \vec{\theta})$ ,  $\vec{\theta} \in \Theta \subset \mathbb{R}^m$ ,  $x \in A$  называется регулярной, если

1.  $\rho(x, \vec{\theta})$  непр дифф по  $\vec{\theta}$  на  $\Theta$
2.  $\frac{\partial}{\partial \theta_i} \int_A \rho(x, \vec{\theta}) dx = \int_A \frac{\partial}{\partial \theta_i} \rho(x, \vec{\theta}) dx$  на  $\Theta$ ,  $i = 1, \dots, m$
3.  $I(\vec{\theta})$  положительно определена на  $\Theta$  и  $I_{ij}(\vec{\theta})$  непрер. на  $\Theta$

**Определение 5.4.** Вероятностная модель  $\xi \sim \rho(x, \theta)$ ,  $\theta \in \Theta \subset \mathbb{R}^m$ ,  $x \in A$  называется сильно регулярной, если эта модель регулярна и

1.  $\rho(x, \theta)$   $k$  раз непрерывно дифф по  $\theta$  на  $\Theta$  ( $k \geq 2$ )
2.  $\frac{\partial^l}{\partial \theta^l} \int_A \rho(x, \theta) dx = \int_A \frac{\partial^l}{\partial \theta^l} \rho(x, \theta) dx$ ,  $l = 1, \dots, k$

**Определение 5.5.** Вероятностная модель  $\xi \sim \zeta(x, \vec{\theta})$ ,  $\vec{\theta} \in \Theta \subset \mathbb{R}^m$ ,  $x \in A$  называется сильно регулярной, если эта модель регулярна и

1.  $\rho(x, \vec{\theta})$   $k$  раз непрерывно дифф по  $\theta$  на  $\Theta$  ( $k \geq 2$ )
2. все производные по  $\vec{\theta}$  перестановочные с  $\int$  по  $x$

**Определение 5.6.** Статистика  $\tilde{g}(\vec{x}_n)$  называется регулярной оценкой функции  $g(\theta)$ , если она является несмещённой оценкой и

$$\frac{\partial}{\partial \theta} \int_B \tilde{g}(\vec{x}_n) L(\vec{x}_n, \theta) d\vec{x}_n = \int_B \tilde{g}(\vec{x}_n) \frac{\partial}{\partial \theta} L(\vec{x}_n, \theta) d\vec{x}_n$$

где  $L(\vec{x}_n, \theta)$  - плотность распределения случайного вектора  $\vec{x}_n$   
 $(L(\vec{x}_n, \theta) = \prod_{i=1}^n \rho(x_i, \theta))$ ,  $B = \underbrace{A \times A \times \dots \times A}_n$

**Теорема 5.2** (Достаточное условие регулярности оценки). Если модель регулярна,  $\tilde{g}(\vec{x}_n)$  является несмещ. оценкой  $g(\theta)$  и  $D[\tilde{g}(\vec{x}_n)]$  ограничена на  $\forall$  компакте из  $\Theta$  по  $\theta$ , тогда оценка регулярна.

## 5.2 Неравенство Крамера-Рао

**Теорема 5.3.** Пусть модель является регулярной,  $\tilde{g}(\vec{x}_n)$  является регулярной оценкой дифф функции  $g(\theta)$ . Тогда

$$\forall \theta \in \Theta \hookrightarrow D[\tilde{g}] \geq \frac{(g')^2(\theta)}{nI(\theta)}$$

*Доказательство.*  $\xi \sim \rho(x, \theta)$ ,  $\theta \in \Theta \subset \mathbb{R}$ ,  $x \in A(\theta)$ ,  $\vec{x}_n$  независ. выборка

$L(\vec{x}_n, \theta) = \prod_{i=1}^n \rho(x_i, \theta)$  - распр. выборки  $\vec{x}_n$ ,  $B = A \times \dots \times A$

$$\frac{\partial}{\partial \theta} \int \dots \int_B L(\vec{x}, \theta) d\vec{x} = \frac{\partial}{\partial \theta} 1 = 0$$

в силу регулярности модели

$$\int \dots \int_B \frac{\partial}{\partial \theta} L d\vec{x} = 0$$

Домножаем и делим на  $L$ , там где  $L = 0$  считаем что интеграл 0

$$\int_B \frac{\partial \ln L}{\partial \theta} L d\vec{x} = 0$$

$$M[\tilde{g}] = g(\theta)$$

$$\int_B \tilde{g}(\vec{x}_n) L(\vec{x}_n, \theta) d\vec{x}_n = g(\theta)$$

$$\frac{\partial}{\partial \theta} \int_B \tilde{g}(\vec{x}_n) L(\vec{x}_n, \theta) d\vec{x}_n = \frac{\partial}{\partial \theta} g(\theta)$$

$$\int_B \tilde{g}(\vec{x}_n) \frac{\partial}{\partial \theta} L d\vec{x}_n = g'(\theta)$$

$$\int_B \tilde{g}(\vec{x}_n) \frac{\partial \ln L}{\partial \theta} L d\vec{x}_n = g'(\theta)$$

$$\int_B (\tilde{g}(\vec{x}_n) - g(\theta)) \frac{\partial \ln L}{\partial \theta} L d\vec{x}_n = g'(\theta)$$

$\eta = \tilde{g}(\vec{x}_n) - g(\theta)$  - с.л. вел

$$\zeta = \frac{\partial \ln L(\vec{x}_n, \theta)}{\partial \theta} - \text{с.л. вел.}$$

$$\begin{aligned}
M[\eta] &= 0 & M[\zeta] &= 0 \\
M[\eta\zeta] &= g'(\theta) \\
\text{cov}(\eta, \zeta) &= M[\eta\zeta] - M[\eta]M[\zeta] = g'(\theta) \\
r &= \frac{\text{cov}(\eta, \zeta)}{\sqrt{D\eta D\zeta}} & |r| &\leq 1 \\
\frac{\text{cov}^2(\zeta, \eta)}{D\zeta D\eta} &\leq 1 \\
g'^2(\theta) &\leq D\zeta \underbrace{D[\tilde{g}]}_{D[\tilde{g}]} \\
D\zeta &= M[\zeta^2] - M^2[\zeta] = M[\zeta^2] \\
D\zeta &= D\left[\frac{\partial \ln L}{\partial \theta}\right] = D\left[\sum_{i=1}^n \frac{\partial \ln \rho(x_i, \theta)}{\partial \theta}\right] = \sum_{i=1}^n D\left[\frac{\partial \ln \rho(x_i, \theta)}{\partial \theta}\right] = \\
&= nD\left[\frac{\partial \ln \rho(x_i, \theta)}{\partial \theta}\right] = nM\left[\left(\frac{\partial \ln \rho}{\partial \theta}\right)^2\right] - \underbrace{nM^2\left[\frac{\partial \ln \rho}{\partial \theta}\right]}_0 = nI(\theta)
\end{aligned}$$

□

**Следствие 5.1.**

1. оценка параметра  $\theta$ ,  $g(t) = \theta$ ,

$$D[\tilde{\theta}] \geq \frac{1}{nI(\theta)}$$

2. многомерный аналог нер. Крамера-Рао

$$D[\tilde{g}(\vec{x}_n)] \geq \frac{1}{n} \nabla^T g(\vec{\theta}) I^{-1}(\vec{\theta}) \nabla g(\vec{\theta})$$

**Определение 5.7** (Эффективная оценка). Регулярная оценка  $\tilde{g}(\vec{x}_n)$  функции  $g(\theta)$  называется эффективной ( $\tilde{g}_{eff}$ ), если  $\forall \theta \in \Theta \hookrightarrow D[\tilde{g}_{eff}] = \inf D[\tilde{g}]$ ,  $\inf$  берётся по всем регулярным оценкам.

**Теорема 5.4.** Если эффективная оценка  $\exists$ , то она единственна.

*Доказательство.* Так же как и оптимальная только нужно доказать что  $\tilde{\theta}_3 = \frac{\tilde{\theta}_1 + \tilde{\theta}_2}{2}$  - регулярная. □



**Теорема 5.5** (Достаточное условие эффективности). Пусть выполнены условия нер. Крамера-Рао и  $D[\tilde{g}] = \frac{g'^2}{nI(\theta)}$ , тогда  $\tilde{g}$  эффективная оценка  $g(\theta)$ .

**Теорема 5.6** (Теорема о частоте). Частота появления события  $A$  в  $n$  независимых опытах является несмещённой, состоятельной и эффективной оценкой вероятности появления этого события.

*Доказательство.* (на примере)

$$\begin{aligned}\xi &\sim \rho(x, \theta) = \theta\{1\} + (1 - \theta)\{0\} & \theta &\in (0, 1) \\ \xi &\sim Bi(1, \theta) & \nu &= \frac{m}{n} \\ \vec{x}_n &= (0, 0, 1, \dots) & \nu &= \frac{1}{n} \sum x_i = \bar{x} \quad \tilde{\theta} = \bar{x}\end{aligned}$$

1. несмещённость ( $\xi \sim Bi(l, \theta)$ ,  $M[\xi] = l\theta$ ,  $D[\xi] = l\theta(1 - \theta)$ )

$$M[\bar{x}] = \frac{1}{n} M[\sum x_i] = M[\xi]$$

2. состоятельность

$$D[\tilde{\theta}] = D[\frac{1}{n} \sum x_i] = \frac{1}{n^2} n D[\xi] = \frac{1}{n} \theta(1 - \theta) \xrightarrow{n \rightarrow \infty} 0$$

состоятельна по достаточному условию

3. эффективность, модель регулярна

$$I(\theta) = \frac{l}{\theta(1 - \theta)} \Big|_{l=1} = \frac{1}{\theta(1 - \theta)}$$

$\tilde{\theta} = \bar{x}$  - регулярная оценка?

$D[\tilde{\theta}] = \frac{1}{n} \theta(1 - \theta)$  огран на  $\forall$  компакте из  $(0, 1)$

Является регулярной ✓

Неравенство Крамера-Рао:

$$D[\tilde{\theta}] = \frac{1}{n} \theta(1 - \theta) \geq \frac{1}{nI(\theta)} = \frac{\theta(1 - \theta)}{n}$$

достигает нижней грани  $\Rightarrow$  эффективная (в данном случае ещё и оптимальная)

□

## 6 Описательная стат. (непараметр. стат.)

$\vec{x}_n$  - выборка

Вероятностная модель - все распределения, кроме сингулярных и вырожденных.

1. Вариационный ряд - упорядоченная выборка

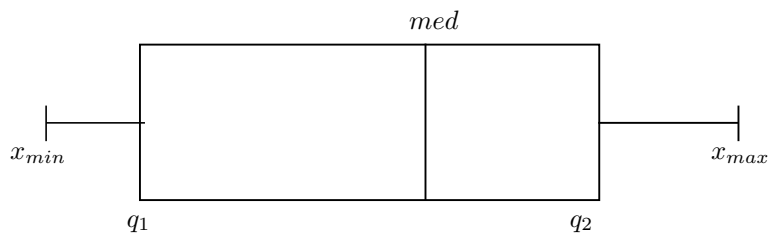
$$x_{min} = x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)} = x_{max}$$

$x_{(k)}$  -  $k$ -ая порядковая сл. вел

2. Размах выборки  $l = x_{max} - x_{min}$
3. Медиана выборки  $med$

$$mes = \begin{cases} x_{(k+1)}, n = 2k + 1 \\ \frac{x_{(k+1)} + x_{(k)}}{2}, n = 2k \end{cases}$$

4. Мода - эл. выборки, который встречается чаще всего
5. Квартили  $q_1, q_2$  (медианы половинок)
6. Boxplot



$\varepsilon = q_2 - q_1$ , если  $x_{min} < q_1 - 1.5\varepsilon$  или  $x_{max} > q_2 + 1.5\varepsilon$  рисуем усики до  $q_1 - 1.5\varepsilon$  или  $q_2 + 1.5\varepsilon$  соответственно, а дальше выбросы обозначаем точками для каждого значения

7. эмпирическая функция распределения

$$\tilde{F}(x) = \frac{m(x)}{n}$$

где  $m(x)$  число элементов выборки, которые  $< x$ .

$$F(x) = P(\underbrace{\xi < x}_A)$$

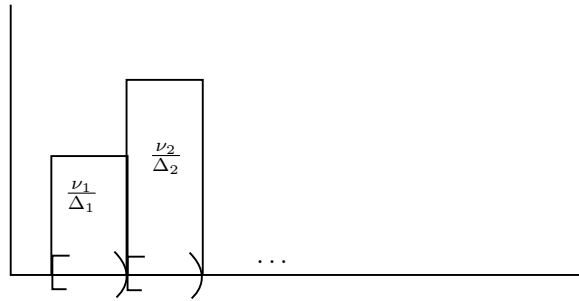
$\tilde{F}$  является несмещённой, состоятельной и эффективной оценкой  $F(x)$  (по Т. о частоте).

## 8. Гистограмма

Статистический ряд

$$\underbrace{[y_1, y_2)}_{\nu_1 = \frac{m_1}{n}}, \underbrace{[y_2, y_3)}_{\nu_2} \dots \underbrace{[y_k, y_{k+1})}_{\nu_k}$$

Эмперически  $k = 1 + \log_2 n$



$$\nu_m = P(y_m \leq \xi < y_{m+1}) = \int_{y_m}^{y_{m+1}} p(x) dx = p(\bar{x}) \Delta_m$$

## 9. числовые характеристики

$\alpha_k = M[\xi^k]$  момент  $k$ -го порядка

$$\tilde{\alpha}_k = \frac{1}{n} \sum_{i=1}^n x_i^k \quad \tilde{\alpha}_1 = \bar{x}$$

- 
- несмещ:  $M[\alpha_k] = \frac{1}{n} M[\sum x_i^k] = M[\xi^k] = \alpha_k$
  - состоятельность: ЗБЧ Хинчина  $\tilde{\alpha}_k \xrightarrow{P} \alpha_k = M[\xi^k]$
- 

$\mu_k = M[(\xi - M\xi)^K]$  - центральный момент  $k$ -го порядка

$$\tilde{\mu}_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$$

- СОСТОЯТЕЛЬНОСТЬ

$$\begin{aligned}\mu_k &= M\left[\sum_{m=0}^k C_k^m \xi^m (-1)^{k-m} (M\xi)^{k-m}\right] = \\ &= \sum_{m=0}^k C_k^m (-1)^{k-m} \alpha_m (\alpha_1)^{k-m} \\ \tilde{\mu}_k &= \sum_{m=0}^k C_k^m (-1)^{k-m} \tilde{\alpha}_m (\tilde{\alpha}_1)^{k-m}\end{aligned}$$

Теорема наследования сходимости:

$$\xi_n \xrightarrow{p} \xi, f(x) \text{ непр на } \mathbb{R} \Rightarrow f(\xi_n) \xrightarrow{p} f(\xi)$$

$$\xi_n \xrightarrow{p} C, f(x) \text{ непр в точке } x = C \Rightarrow f(\xi_n) \xrightarrow{p} f(C)$$

$$\tilde{\alpha}_k \xrightarrow{p} \alpha_k, f(x_1, \dots, x_n) \text{ непр} \Rightarrow \tilde{\mu}_k \xrightarrow{p} \mu_k$$

- несмещённость

$$\begin{aligned}\mu_2 &= D\xi \quad \tilde{\mu}_2 = \tilde{\alpha}_2 - (\tilde{\alpha})^2 \\ M[\tilde{\mu}_2] &= M\left[\frac{1}{n} \sum x_i^2\right] - M\left[\left(\frac{1}{n} \sum x_i\right)^2\right] = \\ &= M\xi^2 - (D[\bar{x}] + (M[\bar{x}])^2) = M\xi^2 - \frac{1}{n^2} n D\xi - (M\xi)^2 = \\ &= D\xi \left[1 - \frac{1}{n}\right] = \mu_2 \frac{n-1}{n}\end{aligned}$$

$$S^2 = \frac{n}{n-1} \tilde{\mu}_2, M[S^2] = \mu_2 \text{ несмещ оценка дисперсии}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

---

коэффициент асимметрии

$$\begin{aligned}\gamma &= \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}} \\ \tilde{\gamma} &= \frac{\tilde{\mu}_3}{\tilde{\mu}_2^{3/2}} \xrightarrow{p} \gamma\end{aligned}$$

10. оценка распределения статистики

$$\vec{x}_n \quad \tilde{g}(\vec{x}_n)$$

$\vec{x}_n$  становится вероятностной моделью, вытаскиваем 1000 подвыбоок с повторением элементов того же объёма  $\vec{x}_n^*$ ,  $\vec{x}_n^* \rightarrow \tilde{g}_i^*(\vec{x}_n^*)$ , строим гистограмму из  $\tilde{g}_1^* \dots \tilde{g}_{1000}^*$

## 7 Методы находж. параметров модели

$$\xi \sim \rho(x, \vec{\theta}), \quad \vec{\theta} \in \Theta \subset \mathbb{R}^m, \quad x \in A$$

Парметрическая модель,  $\vec{x}_n$  выборка

1. Метод моментов (Пирсон)

$$\alpha_k(\vec{\theta}) = M[\xi^k] \rightarrow \tilde{\alpha}_k = \alpha_k(\vec{\theta})$$

Если можем решить систему получаем  $\tilde{\vec{\theta}}$  оценку методом моментов (ОММ)

**Замечание.** ОММГ (ОММ Группированная) для статистического ряда:

$$\underbrace{[y_1, y_2)}_{\nu_1} \dots \underbrace{[y_k, y_{k+1})}_{\nu_k} \quad \nu_i = \frac{m_i}{n}$$

$$z_1 = \frac{y_1 + y_2}{2}, \dots, z_k = \frac{y_k + y_{k+1}}{2}$$

$$\tilde{\alpha}_l = \sum_{i=1}^k z_i^l \nu_i$$

2. Оценка Методом Правдоподобия (Фишер):

---

монета 10 раз, 6 орлов, 4 решки

(a)  $p = \frac{1}{2}$

(b)  $p = \frac{1}{3}$  - орёл,  $p = \frac{2}{3}$  - решка

$$\left(\frac{1}{2}\right)^{10} = 0.00098 \quad \left(\frac{1}{3}\right)^6 \left(\frac{2}{3}\right)^4 = 0.00027$$


---

$$L(\vec{x}_n, \theta) = \prod_{i=1}^n \rho(x_i, \theta)$$

$fix \theta, L(\vec{x}_n)$  - плотность распределения выборки

$fix \vec{x}_n, L(\theta) = \prod_{i=1}^n \rho(x_i, \theta)$  - функция правдоподобия (здесь  $x_i$  - элемент выборки), пытаемся максимизировать  $L(\theta) \rightarrow \max$

**Замечание.** ОМПГ для статистического ряда

$$\begin{aligned}
& \underbrace{[y_1, y_2)}_{\nu_1} \cdots \underbrace{[y_k, y_{k+1})}_{\nu_k} \quad \nu_i = \frac{m_i}{n} \\
& P_1(\theta) = \int_{-\infty}^{y_2} \rho(x, \theta) dx \\
& P_2(\theta) = \int_{y_2}^{y_3} \rho(x, \theta) dx \\
& \quad \dots \\
& P_k(\theta) = \int_{y_k}^{+\infty} \rho(x, \theta) dx \\
& L(\theta) = P_1^{m_1} \dots P_k^{m_k} \quad L(\theta) \rightarrow \max
\end{aligned}$$

## 8 Асимптотические свойства оценок

$n \rightarrow \infty$

1. состоятельность  $\forall \theta \in \Theta \hookrightarrow \tilde{g}(\vec{x}_n) \xrightarrow{P} g(\theta)$
2. асимптотическая несмещённость  $\forall \theta \in \Theta \hookrightarrow M[\tilde{g}(\vec{x}_n)] \xrightarrow{n \rightarrow \infty} g(\theta)$
3. асимптотическая эффективность  $\forall \theta \in \Theta \hookrightarrow nD[\tilde{g}(\vec{x}_n)] \xrightarrow{n \rightarrow \infty} \frac{g'^2(\theta)}{I(\theta)}$
4. асимптотическая нормальность  $\forall \theta \in \Theta \hookrightarrow \sqrt{n}(\tilde{g}(\vec{x}_n) - g(\theta)) \xrightarrow{F} \eta \sim N(0, \sigma^2)$

### 8.1 Асимптотические свойства ОММ

$$\begin{aligned}
\alpha_k(\vec{\theta}) = \tilde{\alpha}_k \rightarrow \tilde{\theta} &= f(\tilde{\alpha}_1, \dots, \tilde{\alpha}_k) \\
\tilde{\alpha}_k &\xrightarrow{P} \alpha_k
\end{aligned}$$

$f$  - непр в  $\alpha_1, \dots, \alpha_k$ , тогда по теореме наследования  $\tilde{\theta} \xrightarrow{P} f(\alpha_1, \dots, \alpha_k) = \vec{\theta}$   
состоятельная

Теорема о наследовании нормальности:

$\sqrt{n}(\xi_n - a) \rightsquigarrow N(0, \sigma^2)$  и  $f(x) \in C'(r)$  и  $f'(a) \neq 0$ , тогда

$$\sqrt{n}(g(\xi_n) - g(a)) \rightsquigarrow N(0, g'^2(a)\sigma^2)$$

ЦПТ:

$$\begin{aligned}\frac{\sum x_i^k - nM[\xi^k]}{\sqrt{nD[\xi^k]}} &\rightsquigarrow N(0, 1) \\ \frac{\tilde{\alpha}_k - a_k}{\sqrt{\alpha_{2k} - a_k^2}} \sqrt{m} &\rightsquigarrow N(0, \underbrace{\alpha_{2k} - a_k^2}_{\sigma_k^2}) \\ (f(\tilde{\alpha}_k) - f(\alpha_j))\sqrt{n} &\rightsquigarrow N(0, f'^2(\alpha_k)\sigma_k^2) \\ (\tilde{\theta} - \theta)\sqrt{n} &\rightsquigarrow N(0, f'^2(\alpha_k)\sigma_k^2)\end{aligned}$$

### 8.1.1 Многомерное ЦПТ

$$\begin{aligned}\alpha &= \begin{pmatrix} \alpha_{s_1} \\ \vdots \\ \alpha_{s_k} \end{pmatrix} \quad \tilde{\alpha} = \begin{pmatrix} \tilde{\alpha}_{s_1} \\ \vdots \\ \tilde{\alpha}_{s_k} \end{pmatrix} \quad \tilde{\alpha}_{s_1} = \frac{1}{n} \sum_{i=1}^n x_i^{s_1} \\ (\tilde{\alpha} - \alpha)\sqrt{n} &\rightsquigarrow N(\vec{0}, K) \quad K_{ij} = \alpha_{s_i + s_j} - \alpha_{s_i} \alpha_{s_j}\end{aligned}$$

$f(x) \in C'(R^k)$  и  $\nabla f(\alpha) \neq 0$

$$(f(\tilde{\alpha}) - f(\alpha))\sqrt{n} \rightsquigarrow N(0, \nabla^T f(\alpha) K \nabla f(\alpha))$$

## 8.2 Асимптотические свойства ОМП

**Теорема 8.1.** Пусть вероятностная модель сильно регулярна и множество  $\Theta$  открыто.

Тогда ОМП является состоятельной, асимп. несмещ., асимп. эффект. и асимп. нормальной.

*Доказательство.*  $L(\theta) \rightarrow \max, \ln L \rightarrow \max, \frac{d \ln L(\tilde{\theta})}{d\theta} = 0$  Ряд Тейлора с остаточным членом в форме Лагранжа:

$$\begin{aligned}\underbrace{\frac{d \ln L(\theta)}{d\theta}}_0 &= \frac{d \ln L}{d\theta}(\theta) + \frac{d^2 \ln L}{d\theta^2}(\theta^*)(\tilde{\theta} - \theta) \\ \tilde{\theta} - \theta &= -\frac{(\ln L)'(\theta)}{(\ln L)''(\theta^*)} \\ \frac{d \ln L}{d\theta} &= \frac{d}{d\theta}(\ln \prod p(x_i, \theta)) = \sum_{i=1}^n \frac{d \ln p(x_i, \theta)}{d\theta}\end{aligned}$$

$$\text{ЗБЧ Хинчина: } \frac{1}{n} \sum \frac{d \ln p}{d\theta} \xrightarrow{p} M \left[ \frac{d \ln p(\xi, \theta)}{d\theta} \right]$$

$$\int_{-\infty}^{+\infty} p(x, \theta) dx = 1$$

$$\int_{-\infty}^{+\infty} \frac{d}{d\theta} p(x, \theta) dx = 1$$

$$\int_{-\infty}^{+\infty} \frac{d \ln p}{d\theta} p dx = 0$$

$$M \left[ \frac{d \ln p}{d\theta} \right] = 0$$

$$\frac{d}{d\theta} \int_{-\infty}^{+\infty} \frac{d \ln p}{d\theta} p dx = 0$$

$$\int_{-\infty}^{+\infty} \frac{d^2 \ln p}{d\theta^2} p dx + \int_{-\infty}^{+\infty} \frac{d \ln p}{d\theta} \frac{dp}{d\theta} dx = 0$$

$$M \left[ \frac{d^2 \ln p}{d\theta^2} \right] + \underbrace{M \left[ \left( \frac{d \ln p}{d\theta} \right)^2 \right]}_{I(\theta) > 0} = 0$$

$$\frac{d^2 \ln L}{d\theta^2} = \sum_{i=1}^n \frac{d^2 \ln p(x_i, \theta^*)}{d\theta^2}$$

$$\frac{1}{n} \sum \frac{d \ln p(x_i, \theta^*)}{d\theta^2} \xrightarrow{p} -I(\theta^*) \neq 0$$

$$\text{Состоятельность доказана: } \theta^* \xrightarrow{p} \theta$$

$$\frac{\sum \frac{d \ln p}{d\theta} - \overbrace{n M \left[ \frac{d \ln p}{d\theta} \right]}^{=0}}{\sqrt{n D \left[ \frac{d \ln p}{d\theta} \right]}} \rightsquigarrow N(0, 1)$$



Лемма Слущкого:  $\xi_n \xrightarrow{F} \xi$ ,  $\eta_n \xrightarrow{P} C$ ,  $x_n \eta_n \xrightarrow{F} \xi C$

$$\begin{aligned}\tilde{\theta} - \theta &= -\frac{\frac{\sum \frac{d \ln p}{d\theta}}{\sqrt{nI(\theta)}} \sqrt{nI(\theta)}}{\frac{1}{-nI(\theta)} \sum \frac{d^2 \ln p}{d\theta^2} (-nI(\theta))} \\ a &= \frac{\sum \frac{d \ln p}{d\theta}}{\sqrt{nI(\theta)}} \rightsquigarrow N(0, 1) \quad b = \frac{1}{-nI(\theta)} \sum \frac{d^2 \ln p}{d\theta^2} \rightsquigarrow 1 \\ (\tilde{\theta} - \theta) \sqrt{nI(\theta)} &= \frac{a}{b} \rightsquigarrow N(0, 1) \\ (\tilde{\theta} - \theta) \sqrt{n} &\rightsquigarrow N(0, \frac{1}{I(\theta)}) \\ D[(\tilde{\theta} - \theta) \sqrt{n}] &= nD[\tilde{\theta}] \xrightarrow{n \rightarrow \infty} \frac{1}{I(\theta)}\end{aligned}$$

Асимптотическая эффективность. □

**Следствие 8.1.**  $\tilde{\theta} \xrightarrow{P} \theta$  - сост.  $\tilde{\theta}$  ОМП,  $\theta \in \Theta \subset \mathbb{R}$

$$\sqrt{n}(\tilde{\theta} - \theta) \rightsquigarrow N(0, \frac{1}{I(\theta)})$$

- $g(\theta) \in C'(\Theta)$  и  $g'(\theta) \neq 0$  на  $\Theta$

$$\sqrt{n}(g(\tilde{\theta}) - g(\theta)) \rightsquigarrow N(0, g'^2(\theta) \frac{1}{I(\theta)})$$

$$g(\tilde{\theta}) \xrightarrow{P} g(\theta)$$

$g(\tilde{\theta})$  сост. оценка  $g(\theta)$ , асим. несмещ., асим. эффект., асим. норм.

- многомерный аналог

$$\sqrt{n}(\vec{\tilde{\theta}} - \vec{\theta}) \rightsquigarrow N(\vec{0}, I^{-1}(\vec{\theta}))$$

$$g(\vec{\theta}) \in C'(\mathbb{R}^m) \quad \nabla g(\vec{\theta}) \neq 0 \quad \theta \in \Theta \subset \mathbb{R}^m$$

$$\sqrt{n}(g(\vec{\tilde{\theta}}) - g(\vec{\theta})) \rightsquigarrow N(\vec{0}, \nabla^T g(\vec{\theta}) I^{-1}(\vec{\theta}) \nabla g(\vec{\theta}))$$

**Пример.**  $\xi \sim R(0, \theta)$ ,  $\theta > 0$ , ОМП:  $\tilde{\theta} = x_{max}$

$$\tilde{\theta}' = \frac{n+1}{n} x_{max} \text{ несмещ}$$

$$M[x_{max}] = \frac{n}{n+1} \theta \xrightarrow{n \rightarrow \infty} \theta \text{ асим. несм.}$$

$$x_{max} \xrightarrow{P} \theta \text{ сост.}$$

$D[x_{max}] = \frac{n\theta^2}{(n+1)^2(n+2)}$ ,  $I(\theta) = \frac{1}{\theta^2}$ ,  $nD[x_{max}] \not\xrightarrow{n \rightarrow \infty} \frac{1}{I(\theta)}$ , значит не является эффективной (на самом деле оценка сверхэффективная, модель нерегулярна)

условия теоремы не выполнены)

Пусть асим. норм.

$$\begin{aligned}\sqrt{n}(\underbrace{\tilde{\theta}}_{x_{max}} - \theta) &\rightsquigarrow N(0, \sigma^2) \\ P(\sqrt{n}(x_{max} - \theta) < x) &\xrightarrow{n \rightarrow \infty} \Phi(x) \quad \forall x \in \mathbb{R} \\ P(\sqrt{n}(x_{max} - \theta) < 0) &= 1 \rightarrow 1 \quad \Phi(0) = \frac{1}{2}\end{aligned}$$

Противоречие  $\Rightarrow$  не является асим. нормальной.

## 9 Доверительный интервал

**Определение 9.1.** Доверительным интервалом величины  $h$  вероятностной модели называется случайный интервал, который покрывает значение  $h$  с вероятностью, не меньшей  $\beta$ .

$$\begin{aligned}I &= (g_1(\vec{x}_n), g_2(\vec{x}_n)) \\ P((g_1, g_2) \ni \theta) &\geq \beta\end{aligned}$$

$\beta$  - доверительная вероятность, чаще всего 0.9, 0.95, 0.99.

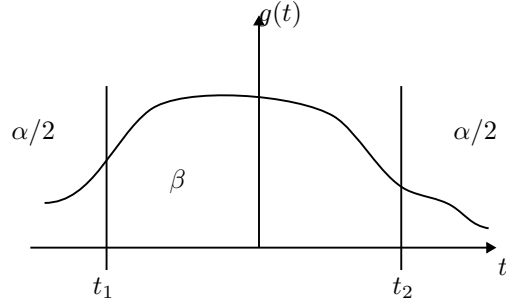
### 9.1 Методы построения доверит. инт.

- точный
- асимптотический
  - по ОММ
  - по ОМП
- численные
  - параметрический бутстрап
  - непараметрический бутстрап

#### 9.1.1 Точный метод

$h, \vec{x}_n \rightarrow f(g, \vec{x}_n) \sim g(t)$  (созерцание)

1.  $g(t)$  - плотность распр. (сл. вел. непр.)



$\alpha + \beta = 1$ , квантиль  $F(x_p) = p$ ,  $x_p$  - квантиль порядка  $p$ ,  $F(x_p) = \int_{-\infty}^{x_p} g(t)dt$

$$t_1 = g_{\alpha/2} = g_{\frac{1-\beta}{2}} \quad t_2 = g_{\beta+\frac{\alpha}{2}} = g_{\frac{1+\beta}{2}}$$

$$P(t_1 < f(h, \vec{x}_n) < t_2) = \beta$$

$$g_1(\vec{x}_n) < h < g_2(\vec{x}_n)$$

2.  $g(t)$  содержит дискр. части, сдвигаем  $\beta$  так чтобы получить точное равенство

**Пример.**  $\xi \sim N(\theta_1, \theta_2^2)$ ,  $\theta_1 \in \mathbb{R}$ ,  $\theta_2 > 0$

$$\tilde{\theta}_1 = \bar{x} \quad \tilde{\theta}_2^2 = S^2$$

Теорема Фишера:

$$\sqrt{n} \frac{\bar{x} - \theta_1}{\theta_2} \sim N(0, 1) \quad \frac{S^2(n-1)}{\theta_2^2} \sim \chi^2(n-1)$$

$$t_1 = \chi_{\frac{1-\beta}{2}}^2(n-1) \quad t_2 = \chi_{\frac{1+\beta}{2}}^2(n-1)$$

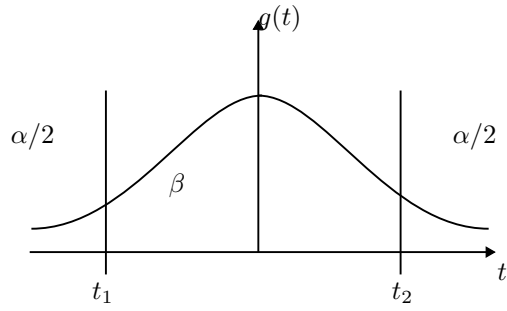
$$P(t_1 < \frac{S^2(n-1)}{\theta_2^2} < t_2) = B$$

$$\frac{S^2(n-1)}{t_2} < \theta_2^2 < \frac{S^2(n-1)}{t_1}$$

$$\sqrt{\frac{S^2(n-1)}{t_2}} < \theta_2 < \sqrt{\frac{S^2(n-1)}{t_1}}$$

$$\frac{\sqrt{n} \frac{\bar{x} - \theta_1}{\theta_2}}{\sqrt{\frac{S^2(n-1)}{\theta_2^2(n-1)}}} \sim t(n-1)$$

$$\sqrt{n} \frac{\bar{x} - \theta_1}{S} \sim t(n-1)$$



$$t_1 = t_{\frac{1-\beta}{2}}(n-1) \quad t_2 = t_{\frac{1+\beta}{2}}(n-1)$$

$$P(t_1 < \sqrt{n} \frac{\bar{x} - \theta_1}{S} < t_2) = \beta$$

$$\bar{x} - \frac{St_2}{\sqrt{n}} < \theta_1 < \bar{x} - \frac{St_1}{\sqrt{n}}$$

## 9.2 Асимптотический метод

$$h, \vec{x}_n \rightarrow f(h, \vec{x}_n) \rightsquigarrow g(t)$$

По ОММ

$$\begin{aligned}
\sqrt{n}(g(\vec{\theta}) - g(\vec{\theta})) &\rightsquigarrow N(\vec{0}, \nabla^T g(\vec{\theta}) I^{-1}(\vec{\theta}) \nabla g(\vec{\theta})) \\
\alpha = \begin{pmatrix} \alpha_{s_1} \\ \vdots \\ \alpha_{s_k} \end{pmatrix} \quad \tilde{\alpha} = \begin{pmatrix} \tilde{\alpha}_{s_1} \\ \vdots \\ \tilde{\alpha}_{s_k} \end{pmatrix} \quad K_{ij} = \alpha_{s_i + s_j} - \alpha_{s_i} \alpha_{s_j} \\
(\tilde{\alpha}_j - \alpha_j) \sqrt{n} &\rightsquigarrow N(0, \alpha_{2k} - \alpha_k^2) \\
\frac{\tilde{\alpha}_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}} \sqrt{n} &\rightsquigarrow N(0, 1) \\
\tilde{\alpha}_k \xrightarrow{p} \alpha_k \quad \sqrt{\tilde{\alpha}_{2k} - \tilde{\alpha}_k^2} &\xrightarrow{p} \sqrt{\alpha_{2k} - \alpha_k^2} \\
\frac{\sqrt{\alpha_{2k} - \alpha_k^2}}{\sqrt{\tilde{\alpha}_{2k} - \tilde{\alpha}_k^2}} &\xrightarrow{p} 1
\end{aligned}$$

Лемма Слущкого  $\xi_n \xrightarrow{F} \xi, \eta_n \xrightarrow{p} C, \xi_n \eta_n \xrightarrow{F} C\xi$

$$\frac{\tilde{\alpha}_k - \alpha_k}{\sqrt{\alpha_{2k} - \alpha_k^2}} \sqrt{n} \frac{\sqrt{\alpha_{2k} - \alpha_k^2}}{\sqrt{\tilde{\alpha}_{2k} - \tilde{\alpha}_k^2}} \rightsquigarrow N(0, 1)$$

|   |
|---|
| $ \frac{\sqrt{n}(g(\tilde{\alpha}) - g(\alpha))}{\sqrt{\nabla^T g(\tilde{\alpha}) K(\tilde{\alpha}) \nabla g(\tilde{\alpha})}} \rightsquigarrow N(0, 1) \quad \text{ОММ} $    |
| $ \frac{\sqrt{n}(g(\vec{\theta}) - g(\vec{\theta}))}{\sqrt{\nabla^T g(\vec{\theta}) I^{-1}(\vec{\theta}) \nabla g(\vec{\theta})}} \rightsquigarrow N(0, 1) \quad \text{ОМП} $ |

**Пример.**  $\xi \sim \rho(x) = x\theta\{(0, 1)\} + (1 - \frac{\theta}{2})\{2\}, \theta \in (0, 2)$

$\vec{x}_n = \{0.53, 0.84, 0.1, 0.83, \underbrace{2, \dots, 2}_{16}\}, n = 20$

1. ОММ  $\tilde{\theta} = \frac{3}{2}(2 - \bar{x}), \tilde{\theta} = 0.4275, S = 0.6, \beta = 0.95$

$$\tilde{\theta} = \frac{3}{2}(2 - \tilde{\alpha}_1) = g(\tilde{\alpha}_1) \quad \theta = g(\alpha_1)$$

$$\frac{\sqrt{n}(g(\tilde{\alpha}_1) - g(\alpha_1))}{\sqrt{\nabla^T g(\tilde{\alpha}) K(\tilde{\alpha}) \nabla g(\tilde{\alpha})}} \rightsquigarrow N(0, 1)$$

$$\nabla g = 0 \frac{3}{2} \quad K_{11} = \alpha_2 - \alpha_1^2$$

$$\frac{\sqrt{n}(\tilde{\theta} - \theta)}{\sqrt{-\frac{3}{2}(-\frac{3}{2})(\tilde{\alpha}_2 - \tilde{\alpha}_1^2)}} \rightsquigarrow N(0, 1)$$

$$\tilde{\mu}_2 = \tilde{\alpha}_2 - \tilde{\alpha}_1^2 = 0.342 \quad \tilde{\mu}_2 = S^2 \frac{n-1}{n}$$

$$t_1 = u_{\frac{1-0.95}{2}} = u_{0.025} = -1.96 \quad t_2 = u_{\frac{1+0.95}{2}} = u_{0.025} = 1.96$$

$$-1.96 < \frac{\sqrt{20}(0.4275 - \theta)}{\sqrt{\frac{9}{4} \cdot 0.342}} < 1.96$$

$$0.0435 < \theta < 0.811 \quad l = 0.768$$

2. ОМП  $\tilde{\theta} = 2(1 - \nu) = 2(1 - \frac{16}{20}) = 0.4$

$$\frac{\sqrt{n}(\tilde{\theta} - \theta)}{\sqrt{I^{-1}(\theta)}} \rightsquigarrow N(0, 1)$$

$$I(\theta) = \frac{1}{\theta(2 - \theta)}$$

$$\frac{\sqrt{n}(\tilde{\theta} - \theta)}{\sqrt{\tilde{\theta}(2 - \tilde{\theta})}} \rightsquigarrow N(0, 1)$$

$$-1.96 < \frac{\sqrt{20}(0.4 - \theta)}{\sqrt{0.4 \cdot 1.6}} < 1.96$$

$$0.049 < \theta < 0.751 \quad l = 0.702$$

ОМП довер. инт. дисперсии

$$D\xi = \frac{11}{12}\theta - \frac{4}{9}\theta^2$$

$$\tilde{D}\xi = \frac{11}{12} \cdot 0.4 - \frac{4}{9} \cdot 0.4^2 = 0.296$$

$$\frac{\sqrt{n}(g(\tilde{\theta}) - g(\theta))}{\sqrt{\nabla^T g(\tilde{\theta}) I^{-1}(\tilde{\theta}) \nabla g(\tilde{\theta})}} \rightsquigarrow N(0, 1)$$

$$\nabla g(\theta) = \frac{11}{12} - \frac{8}{9}\theta$$

$$-1.96 < \frac{\sqrt{20}(\tilde{D}\xi - D\xi)}{\sqrt{(\frac{11}{12} - \frac{8}{9} \cdot 0.4)^2 \cdot 0.4 \cdot 1.6}} < 1.96$$

$$0.1 < D\xi < 0.492$$

### 9.3 Численные методы

#### 1. Непараметрические методы

$h, \vec{x}_n \rightarrow \tilde{h}$ , берём выборку за вероятностную модель, из выборки формируем подвыборку  $N = 1000$ ,  $\tilde{h} - h = \tilde{\Delta}$

(а)  $\vec{x}_n^*$  с повторение элем.  $\Delta_1^* = \tilde{h}^* - \tilde{h}$

(б) ...

(с) ???

(d) ...

(e)  $\vec{x}_n^*$  с повторение элем.  $\Delta_{1000}^* = \tilde{h}^* - \tilde{h}$

(f) Profit!

Вариационный ряд  $\Delta_{(1)}^*, \dots, \Delta_{(1000)}^*$

$$\begin{aligned} K_1 &= \left[ \frac{1-\beta}{2} \cdot 1000 \right] & K_2 &= \left[ \frac{1+\beta}{2} \cdot 1000 \right] \\ t_1 &= \Delta_{(k_1)} & t_2 &= \Delta_{(k_2)} \\ P(t_1 < \tilde{h}^* - \tilde{h} < t_2) &\approx \beta \\ P(t_1 < h^* - h < t_2) &\approx \beta \\ \tilde{h} - t_2 < h < \tilde{h} - t_1 \end{aligned}$$

## 2. Параметрический бутстрап

$h, \tilde{h}, \Delta = \tilde{h} - h, \xi \sim \rho(x, h), \vec{x}_n \rightarrow \tilde{h}$  - сост. и несм. оценка

$\xi \sim \rho(x, \tilde{h})$  моделируем выборки  $\vec{x}_n^*, N = 50000$

$\Delta_i^* = \tilde{h}^* - \tilde{h}, \Delta_{(1)}^* \dots \Delta_{(N)}^*$

$$\begin{aligned} K_1 &= \left[ \frac{1-\beta}{2} \cdot N \right] & K_2 &= \left[ \frac{1+\beta}{2} \cdot N \right] \\ t_1 &= \Delta_{(k_1)}^* & t_2 &= \Delta_{(k_2)}^* \\ P(t_1 < \tilde{h}^* - \tilde{h} < t_2) &\approx \beta \\ P(t_1 < h^* - h < t_2) &\approx \beta \end{aligned}$$

**Пример.**  $\vec{x}_n = \{0.53, 0.84, 0.1, 0.83, \underbrace{2, \dots, 2}_{16}\}, n = 20$

### 1. Непараметрический бутстрап $\tilde{\theta} = 0.4$ ОМП $\tilde{\theta} = 2(1 - \nu)$

- $\vec{x}_n^* = 0.53, \dots, m = 14, \Delta_1^* = \tilde{\theta}^* - \tilde{\theta} = 0.2$
- ...
- $\Delta_{1000}^* = 0$

$$\begin{aligned} \beta &= 0.95 & k_1 &= 25 & k_2 &= 975 \\ t_1 &= \Delta_{(25)}^* = -0.3 & t_2 &= \Delta_{(975)}^* = 0.4 \\ P(-0.3 < \tilde{\theta}^* - \tilde{\theta} < 0.4) &\approx 0.95 \\ P(-0.3 < \tilde{\theta} - \theta < 0.4) &\approx 0.95 \\ 0 < \theta < 0.7 & l &= 0.7 \end{aligned}$$

2. Параметрический бутстрап  $\tilde{\theta} = 0.4$

$$\xi \sim \rho(x, \theta) = x\theta\{(0, 1)\} + (1 - \frac{\theta}{2})\{2\} = x \cdot 0.4\{(0, 1)\} + 0.8\{2\}$$

$N = 10000$ , делаем выборки из модели и дальше так же как и непарам.

#### 9.4 Доверительный интервал для частоты

$\nu = \frac{m}{n}$  хорошая оценка  $P(A)$

Интегральная теорема Муавра-Лапласа:

$$\begin{aligned}\frac{m - np}{\sqrt{np(1-p)}} &\rightsquigarrow N(0, 1) \\ \frac{\nu - p}{\sqrt{p(1-p)}} \sqrt{n} &\rightsquigarrow N(0, 1) \\ \frac{\nu - p}{\sqrt{\nu(1-\nu)}} \sqrt{n} &\rightsquigarrow N(0, 1)\end{aligned}$$

#### 9.5 Доверительный интервал функции распределения

$F, \tilde{F}(x) = \frac{m}{n}$ , где  $m$  - кол-во элементов меньше  $x$

$$\frac{\tilde{F}(x) - F(x)}{\sqrt{\tilde{F}(x)(1 - \tilde{F}(x))}} \sqrt{n} \rightsquigarrow N(0, 1)$$

### 10 Проверка статистических гипотез

**Определение 10.1.** Гипотеза - любое высказывание о вероятностной модели.

**Определение 10.2.** Простая гипотеза - однозначное определение вероятностной модели.

**Определение 10.3.** Сложная гипотеза - неоднозначное определение вер. модели.

$\rho \sim N(0, a), H : a = 2$  - простая,  $H : a > 2$  - сложная

**Определение 10.4.**  $H_0$  - основная гипотеза,  $H_1$  - альтернативная (отклонение от основной)

$H_0 : a = 2, H_1 : a > 2$



## 10.1 Принцип от маловероятного

Пусть  $H_0$  - верна. Событие  $A$ .  $P(A|H_0)$  - мала. Событие  $A$  наблюдаемо. Отвергаем  $H_0$ , иначе нет оснований отвергнуть  $H_0$ .

## 11 Критерии согласия

### 11.1 Критерий Пирсона

Вероятностная модель,  $\vec{x}_n$

$H_0 : \xi \sim F(x)$  - простая гипотеза

$H_1 : \bar{H}_0$  - сложная

Полная группа событий:  $A_1 \dots A_k$  - конечная группа событий,  $A_1 + \dots + A_k = \Omega$ ,  $A_i A_j = \emptyset$   $i \neq j$ ,  $P(A_i) > 0$

$H_0 : P_i = P(A_i), \tilde{P}_i = \nu_i = \frac{m_i}{n}$

$$\Delta = n \sum_{i=1}^k \frac{(p_i - \nu)^2}{p_i}$$
$$\Delta = n \sum_{i=1}^k \frac{(p_i - \frac{m_i}{n})^2}{p_i} = \sum_{i=1}^k \frac{(np_i - m_i)^2}{np_i}$$

**Теорема 11.1.** Если  $H_0$  верна, то  $\Delta \rightsquigarrow \chi^2(k-1)$

**Замечание.** Нормальная аппроксимация при  $n \geq 50$ ,  $np_i \geq 5$  (можно мухлевать:  $np_i \geq 1$  и в 20% случаев можем разрешить  $np_i < 5$ )

**Пример.**  $H_0$  : красные автомобили штрафуют в 2 раза чаще остальных

$H_1 : \bar{H}_0$

Выборка: 150 штрафов, 90 красные,  $A_{red}$ ,  $A_{other}$

|        |       |       |
|--------|-------|-------|
|        | $A_r$ | $A_o$ |
| $p_i$  | 2/3   | 1/3   |
| $np_i$ | 100   | 50    |
| $np_i$ | 90    | 60    |

$$\hat{\Delta} = \frac{(100 - 90)^2}{100} + \frac{(50 - 60)^2}{50} = 3$$

$\alpha$  - уровень значимости,  $\alpha = .1; \boxed{.05}; .01$

$$H_0 : \Delta \rightsquigarrow \chi^2(1), k = 2$$

$$\text{p-value} = P(\Delta \geq \tilde{\Delta} | H_0) = \int_3^\infty q(t) dt = 0.083$$

Нет веских оснований отвергнуть  $H_0$ .

**Замечание.** Правила использования p-value:

1. Если  $\text{p-value} \leq \alpha$ , то  $H_0$  отвергается, результаты значимы, p-value - мера значимости
2. Если  $\text{p-value} > \alpha$ , то нет оснований отвергать  $H_0$ , результаты незначимы.

Либо гипотеза верна (отклонения объясняются случайными факторами), либо критерий недостаточно мощный.

p-value не является вероятностью  $H_0$  и не является вероятностью случайных факторов!

**Пример** (Закон Бенфорда). В больших массивах данных, полученных естественным путём, следуют определённому распределению.

$$P_i = \log_{10}(1 + \frac{1}{d_i}) \quad d_i = 1, \dots, 9$$

Рассмотрим числа Фибоначи (первые 100 штук)

|        | 1   | 2    | 3    | 4   | 5    | 6    | 7    | 8    | 9    |
|--------|-----|------|------|-----|------|------|------|------|------|
| $m_i$  | 29  | 18   | 13   | 9   | 8    | 6    | 5    | 7    | 5    |
| $p_i$  | 0.3 | 0.18 | 0.12 | 0.1 | 0.08 | 0.07 | 0.06 | 0.05 | 0.04 |
| $np_i$ | 3   | 18   | 12   | 10  | 8    | 7    | 6    | 5    | 4    |

8 и 9 объединяем чтобы попадать под условия применимости

$$\tilde{\Delta} = \frac{(30 - 29)^2}{30} + \dots + \frac{(9 - 12)^2}{9} = 1.53 \quad \Delta \rightsquigarrow \chi^2(7)$$

$$\text{p-value} = P(\Delta \geq \tilde{\Delta} | H_0) = \int_{1.53}^\infty q(t) dt = 0.981$$

Нет оснований отвергать  $H_0$ .

## 11.2 Критерий Колмогорова (непр. распр.)

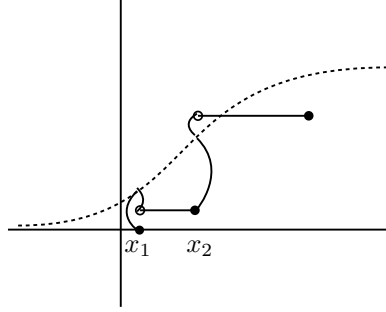
$$H_0 : \xi \sim F(x), H_1 : \bar{H}_0, \vec{x}_n$$

$$\Delta = \sqrt{n} \sup_{x \in \mathbb{R}} |\tilde{F}(x) - F(x)|$$

- где  $\tilde{F}$  - империческая функция распределения.

**Теорема 11.2.** Если  $H_0$  верна, то  $\Delta \rightsquigarrow K(x)$   $K(x)$  - функ. распределения Колмогорова

$$K(x) = P(\Delta < x) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2 x^2} \{(0, \infty)\}$$



$$\sup_{\mathbb{R}} |\tilde{F}(x) - F(x)| = \max_{i=1, \dots, n} \max(|\tilde{F}(x_i - 0) - F(x_i)|, |\tilde{F}(x_i + 0) - F(x_i)|)$$

**Пример.**  $H_0 : \xi \sim R(0, 1)$ ,  $H_1 : \bar{H}_0$ ,  $\vec{x}_n = (0.2; 0.5; 0.8; 0.9)$

$$\tilde{\Delta} = \sqrt{4} \max(0.2; 0, 25; 0.3; 0.15) = 2 * 0.3 = 0.6$$

$$H_0 : \Delta \rightsquigarrow K(x)$$

$$\text{p-value} = P(\Delta \geq \tilde{\Delta} | H_0) = 1 - K(\tilde{\Delta}) = -2 \sum_{k=1}^{\infty} (-1)^j e^{-2k^2 \tilde{\Delta}^2} = 0.8642$$

4 элемента - вообще беда, используем бутстрап, так как считаем в вероятности, что верна  $H_0$  по хорошему бутстрап параметрический

- $x_4^* \rightarrow \Delta_1^* = \sqrt{n} \sup |\tilde{F}^*(x) - F(x)|$
- ...
- $x_4^* \rightarrow \Delta_N^* = \sqrt{n} \sup |\tilde{F}^*(x) - F(x)|$

$N = 10000 - 50000$ , вариационный ряд  $\Delta_{(1)}^* \dots \Delta_{(N)}^*$

$$\text{p-value} = \frac{K}{N}, K - \text{число } \Delta_{(i)}^* \geq \tilde{\Delta}$$

$$\text{p-value} = 0.47$$

### 11.3 Критерий Пирсона для слож. гипотезы

$H_0 : \xi \sim F(x, \vec{\theta}), \vec{\theta} \in \Theta \subset \mathbb{R}^m, H_1 : \bar{H}_0, \vec{x}_n, A_1 \dots A_k$  - полная группа событий

$$P_i(\vec{\theta}) = P(A_i) \quad P_i = \nu_i$$

$$\Delta = \sum_{i=1}^n \frac{(np_i(\vec{\theta}) - m_i)^2}{np_i(\vec{\theta})}$$

**Теорема 11.3.** Если  $H_0$  верная и  $\tilde{\theta}$  есть ОМПГ, то  $\Delta \rightsquigarrow \chi^2(k-1-m)$

**Пример.**  $H_0 : \xi \sim R(0, \theta), \theta > 0, H_1 : \bar{H}_0$

|        | $[0, 1)$ | $[1, 2)$ | $[2, 3)$ | $[3, 4)$ | $[4, \infty)$ |
|--------|----------|----------|----------|----------|---------------|
| $m_i$  | 25       | 10       | 15       | 30       | 20            |
| $p_i$  | 1/5      | 1/5      | 1/5      | 1/5      | 1/5           |
| $np_i$ | 20       | 20       | 20       | 20       | 20            |

$$P_i = \int_{a_i}^{b_i} \frac{1}{\theta} dx \quad P_1 = P_2 = P_3 = P_4 = \frac{1}{\theta}$$

$$P_5 = \int_4^{\infty} p(x, \theta) dx = \int_4^{\infty} \frac{1}{\theta} dx = \frac{\theta - 4}{\theta}$$

$$L(\theta) = \left(\frac{1}{\theta}\right)^{25+10+15+30} \left(\frac{\theta-4}{\theta}\right)^{20} \rightarrow \max$$

$$\ln L = -80 \ln \theta + 20 \ln(\theta - 4) - 20 \ln \theta = -100 \ln \theta + 20 \ln(\theta - 4) \rightarrow \max$$

$$(\ln L)' = -\frac{100}{\theta} + \frac{20}{\theta - 4} = 0 \quad \tilde{\theta} = 5$$

$$\Delta \rightsquigarrow \chi^2(5-1-1) = \chi^2(3)$$

$$\tilde{\Delta} = \frac{(25-20)^2}{20} + \dots + \frac{(20-20)^2}{20} = 12.5$$

$$\text{p-value} = P(\Delta \geq \tilde{\Delta} | H_0) = \int_{12.5}^{\infty} q(t) dt = 0.00585$$

Отвергаем  $H_0$ .

**Пример.**  $H_0$  : распределение Эрланга  $\xi \sim p(x) = \frac{1}{\lambda^2} x e^{-x/\lambda} \{(0, +\infty)\}$ ,  $\lambda > 0, H_1 : \bar{H}_0$

|        | $[0; 2.5)$ | $[2.5; 5)$ | $[5; 7.5)$ | $[7.5; 10)$ | $[10; 12.5)$ | $[12.5; 15)$ |
|--------|------------|------------|------------|-------------|--------------|--------------|
| $m_i$  | 12         | 17         | 12         | 4           | 3            | 2            |
| $np_i$ | 12.9       | 16.4       | 10.4       | 5.5         | 2.6          | 2.2          |

$$\int_a^b p(x)dx = \frac{a}{\lambda}e^{-a/\lambda} - \frac{b}{\lambda}e^{-b/\lambda} + e^{-a/\lambda} - e^{-b/\lambda}$$

$$P_1 = \int_0^{2.5} p(x)dx \quad \dots \quad P_6 = \int_{12.5}^{\infty} p(x)dx$$

$$L(\lambda) = P_1^{12} \dots P_6^2 \rightarrow \max$$

Считаем экстремум численно.

$$\tilde{\lambda} = 2.531$$

После объединения и пересчёта:

$$\tilde{\lambda} = 2.542 \quad \tilde{\Delta} = 0.756$$

$$\text{p-value} = P(\Delta \geq \tilde{\Delta} | H_0) = 0.86$$

#### 11.4 Критерий Колмогорова для сл. гип. (непр. распр.)

$H_0 : \xi \sim F(x, \vec{\theta}), \vec{\theta} \in \Theta \subset \mathbb{R}^m, H_1 : \bar{H}_0, \vec{x}_n, A_1 \dots A_k$  - полная группа событий

$$\Delta = \sqrt{n} \sup_{x \in \mathbb{R}} |\tilde{F}(x) - F(x, \vec{\theta})|$$

Параметрический бутстрап:  $\vec{x}_n \rightarrow \tilde{\vec{\theta}}$  - сост. оценка (любой метод)

$$\tilde{\Delta} = \sqrt{n} \sup_{x \in \mathbb{R}} |\tilde{F}(x) - F(x, \tilde{\vec{\theta}})|$$

$\xi \sim F(x, \tilde{\vec{\theta}}), N = 10000 - 50000$

- $\vec{x}_n^* \rightarrow \tilde{\vec{\theta}}^* \rightarrow \Delta_1^* = \sqrt{n} \sup_{x \in \mathbb{R}} |\tilde{F}^*(x) - F(x, \tilde{\vec{\theta}})|$
- ...

Вариационный ряд  $\Delta_{(1)}^* \dots \Delta_{(N)}^*$ ,  $\text{p-value} = \frac{K}{N}$ ,  $K$  - число элементов  $\Delta_{(i)}^* \geq \tilde{\Delta}$

**Пример** (Эрланг).

$$\tilde{\Delta} = 0.3982 \quad \text{p-value} = 0.9284$$

#### 11.5 Проверка гипотезы однородности

$A_1 \dots A_k$  - полная группа

- 1 выборка  $m_{11} \dots m_{1k}$
- ...
- 1 выборка  $m_{l1} \dots m_{lk}$

$H_0$  : все выборки из одного распределения,  $H_1 : \bar{H}_0$

$$n = n_1 + \dots + n_l \quad \nu_j = \frac{\sum_{i=1}^l m_{ij}}{n}$$

$$\Delta_1 = \sum_{j=1}^k \frac{(m_{1j} - n_1 \nu_j)^2}{n_1 \nu_j} \quad \Delta_s = \sum_{j=s}^k \frac{(m_{sj} - n_s \nu_j)^2}{n_s \nu_j}$$

$$\Delta = \Delta_1 + \dots + \Delta_l$$

**Теорема 11.4.** Если  $H_0$  верна, то  $\Delta \rightsquigarrow \chi^2((k-1)(l-1))$

**Пример** (Коллоквиум 2023).

|     |   |   |    |    |    |
|-----|---|---|----|----|----|
|     | 2 | 3 | 4  | 5  |    |
| 1гр | 0 | 1 | 10 | 5  | 16 |
| 2гр | 1 | 0 | 9  | 1  | 11 |
| 3гр | 2 | 4 | 6  | 4  | 16 |
| 4гр | 1 | 0 | 8  | 6  | 15 |
|     | 4 | 5 | 33 | 16 |    |

|                 |                 |
|-----------------|-----------------|
| 2 и 5           | 3 и 4           |
| 5               | 11              |
| 2               | 9               |
| 6               | 10              |
| 7               | 8               |
| $\frac{20}{58}$ | $\frac{38}{58}$ |

$$\Delta_1 = 0.074 \quad \Delta_2 = 1.294 \quad \Delta_3 = 0.064 \quad \Delta_4 = 0.986$$

$$\tilde{\Delta} = 2.418 \quad \chi^2(1 \cdot 3)$$

$$\text{p-value} = P(\Delta \geq \tilde{\Delta} | H_0) = 0.49$$

## 11.6 Проверка гипотезы независимости

$(\xi, \eta)$ ,  $H_0 : \xi$  и  $\eta$  - независимы,  $H_1 : \bar{H}_0$ ,  $A_i$  и  $B_j$  - полные группы событий

|          |          |          |          |
|----------|----------|----------|----------|
|          | $A_1$    | $\dots$  | $A_k$    |
| $B_1$    | $m_{11}$ | $\dots$  | $m_{1k}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $B_l$    | $m_{l1}$ | $\dots$  | $m_{lk}$ |

$$P_1 = \frac{m_{11} + \dots + m_{1k}}{n} \quad q_1 = \frac{m_{11} + \dots + m_{l1}}{n}$$

$$\Delta = \sum_{i,j} \frac{(m_{ij} - np_i q_j)^2}{np_i q_j}$$

**Теорема 11.5.** Если  $H_0$  верна, то  $\Delta \rightsquigarrow \chi^2((k-1)(l-1))$ .

**Пример.**  $n = 246$ , зависимость успеваемости от родного города

|   | Москва           | Подмоск          | Остальное        |                  |
|---|------------------|------------------|------------------|------------------|
| 2 | 13               | 10               | 17               | $\frac{40}{246}$ |
| 3 | 35               | 38               | 19               | $\frac{92}{246}$ |
| 4 | 21               | 21               | 26               | $\frac{68}{246}$ |
| 5 | 16               | 11               | 19               | $\frac{46}{246}$ |
|   | $\frac{85}{246}$ | $\frac{80}{246}$ | $\frac{81}{246}$ |                  |

$$\tilde{\Delta} = 11.05 \quad \text{p-value} = \int_{11.05}^{\infty} q(t)dt = 0.087$$

Нет оснований отвергать  $H_0$ .

## 11.7 Проверка гипотезы случайности

$H_0$  : элементы выборки независимы и одинакового распределены,  $H_1 : \bar{H}_0$ ,  $I_n$  - число инверсий в выборке

**Теорема 11.6.** Если  $H_0$  верно, то

$$\Delta = \frac{I_n - \frac{n(n-1)}{4}}{\sqrt{\frac{n^3}{36}}} \rightsquigarrow N(0, 1)$$

**Пример.**  $n = 10$ ,  $\vec{x}_n = (4, 10, 8, 8, 6, 7, 7, 9, 8, 9)$

$$I_n = 0 + 8 + 3 + 3 + \dots + 0 = 15$$

$$\tilde{\Delta} = \frac{15 - \frac{10 \cdot 9}{4}}{\sqrt{\frac{10^3}{36}}} = -1.42$$

$$\text{p-value} = P(|\Delta| \geq |\tilde{\Delta}| | H_0) = 2P(\Delta \geq |\tilde{\Delta}| | H_0) = 2 \int_{1.42}^{\infty} q(t)dt = 0.155$$

Нет оснований отвергать  $H_0$ .

## 11.8 Критерий Смирнова для проверки гипотезы однородности (непр. расп.)

$H_0 : F(x) = G(x)$ ,  $H_1 : \bar{H}_0$  Выборки  $\vec{x}_n, \vec{y}_m$

$$\Delta = \sqrt{\frac{nm}{n+m}} \sup_R |\tilde{F}(x) - \tilde{G}(x)|$$

где  $\tilde{F}$  и  $\tilde{G}$  - эмпирические функции распределения

**Теорема 11.7.** Если  $H_0$  верна, то  $\Delta \rightsquigarrow K(x)$

## 12 Использование доверительных интервалов для проверки гипотез

Вероятностная модель,  $h$  - характеристика вероятностной модели,  $\tilde{h}$  - оценка

$$H_0 : h \in T_0, H_1 : h \in T_1, T_0 \cap T_1 = \emptyset$$

$\tilde{h} \rightarrow$  доверительный интервал (точный, асимптотический или непараметрический бутстраповский)

Если  $I \subset T_1$ , то  $H_0$  отвергаем

**Пример.**  $H_0 : h \leq 3, H_1 : h > 3$

- Если  $\tilde{h} \leq 3$  сразу говорим что нет оснований отвергать гипотезу
- Если  $\tilde{h} > 3$ , считаем доверительный интервал (правосторонний):  $\tilde{h} \rightarrow I = [h_0, +\infty)$

**Пример.**  $H_0 : h = 3, H_1 : h \neq 3$ , в таком случае интервал двухсторонний:  $I = (\tilde{h} - \delta, \tilde{h} + \delta)$

## 13 Методы построения критериев

$G_{кр} (G_{cr})$  - критическая область

$$H_0 \text{ отвергаем} \Leftrightarrow \vec{x}_n \in G_{cr}$$

$$P(\vec{x}_n \in G | H_0) \leq \alpha$$

Ошибка I рода  $\alpha_1 = P(H_1 | H_0)$ .

Ошибка II рода  $\alpha_0 = P(H_0 | H_1)$ .

Если  $H_0$  и  $H_1$  сложные, то  $\alpha_1 = \sup_{H_0} P(\vec{x}_n \in G | H_0)$ ,  $\alpha_2 = \sup_{H_1} P(\vec{x}_n \notin G | H_1)$

**Определение 13.1** (Мощность критерия).

$$W(H_1) = P(\vec{x}_n \in G | H_1)$$

Для простой гипотезы  $W = 1 - \alpha_2$



**Определение 13.2** (Состоятельность критерия). Критерий называется состоятельным, если

$$\forall H_1 \hookrightarrow W(H_1) \xrightarrow{n \rightarrow \infty} 1$$

**Замечание.**  $\alpha_1$  - наиболее опасная ошибка

$\alpha_1 \leq \alpha$  обязательно,  $\alpha_2 \rightarrow \min$ , решаем оптимизационную задачу  $\rightarrow G_{opt}$

**Пример.**  $H_0$  : болен,  $H_1$  : здоров

$\alpha_1 = P(H_1|H_0)$  - признали больного здоровым - беда

$\alpha_2 = P(H_0|H_1)$  - признали здорового больным - не так критично

### 13.1 Теорема Неймана-Пирсона для проверки простой $H_0$ против простой $H_1$

$\alpha_2 \rightarrow \min, \alpha_1 \leq \alpha$

$$\begin{aligned} W = 1 - \alpha_2 = P(\vec{x}_n \in G|H_1) &\rightarrow \max \\ \int_G L_1(\vec{x}_n) d\vec{x}_n = \int_G \underbrace{\frac{L_1}{L_0}}_l L_0 d\vec{x}_n &\rightarrow \max \\ G : l \geq C \quad \int_G L_0 d\vec{x}_n &\leq L \end{aligned}$$

$l$  - отношение правдоподобия

**Пример.**  $n = 25, \bar{x} = 9.3, \alpha = 0.02$

$H_0 : \xi \sim N(10, 4), H_1 : \xi \sim N(9, 4)$

$$\begin{aligned} l = \frac{L_1}{L_0} = \frac{\prod_{i=1}^n p_1(x_i)}{\prod_{i=1}^n p_0(x_i)} &= \frac{\left(\frac{1}{\sqrt{2\pi \cdot 2}}\right)^n e^{-\frac{1}{8} \sum_{i=1}^n (x_i - 9)^2}}{\left(\frac{1}{\sqrt{2\pi \cdot 2}}\right)^n e^{-\frac{1}{8} \sum_{i=1}^n (x_i - 10)^2}} = \\ &= e^{-\frac{1}{8} \sum_{i=1}^n (2x_i - 10)} = e^{-\frac{1}{4} n \bar{x} + \frac{19}{8} n} \geq C \\ -\frac{1}{4} n \bar{x} + \frac{19}{8} n &\geq \ln C \quad G : \bar{x} \leq A \\ P(\bar{x} \leq A|H_0) &\leq \alpha \end{aligned}$$

По Т. Фишера:

$$\begin{aligned} \frac{\bar{x} - a}{\sigma} \sqrt{n} &\sim N(0, 1) \\ P\left(\frac{\bar{x} - 10}{2} \sqrt{25} \leq \frac{A - 10}{2} \sqrt{25}\right) &\leq \alpha \end{aligned}$$

$\frac{5}{2}(A - 10) = U_\alpha$  - квантиль порядка  $\alpha$  для нормального распределения

$$\frac{A - 10}{2}5 = U_\alpha = -2.054$$

$$G : \bar{x} \leq 9.18$$

$\bar{x} \notin G$  нет оснований отвергнуть  $H_0$

$$W = P(\bar{x} \in G | H_1) = P\left(\frac{\bar{x} - 9}{2}5 \leq \frac{9.18 - 9}{2}5\right) = \int_{-\infty}^{0.45} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0.67$$

$$\alpha_1 = 0.02 \quad \alpha_2 = 1 - W = 0.33$$

Сколько нужно  $n$  чтобы имело смысл

$$\alpha_1 = 0.01 \quad \alpha_2 = 0.1$$

$$G : \bar{x} \leq B$$

$$\alpha_1 = P(\bar{x} \leq B | H_0) = P\left(\frac{\bar{x} - 10}{2}\sqrt{n} \leq \frac{B - 10}{2}\sqrt{n}\right) = 0.01$$

$$\alpha_2 = P(\bar{x} > B | H_0) = P\left(\frac{\bar{x} - 9}{2}\sqrt{n} \leq \frac{B - 10}{2}\sqrt{n}\right) = 0.1$$

$$\begin{cases} \frac{B - 10}{2}\sqrt{n} = U_{0.01} = -2.33 \\ \frac{B - 9}{2}\sqrt{n} = U_{0.9} = 1.28 \end{cases}$$

$$B = 9.36 \quad n = 53$$

**Пример.**  $\xi \sim \xi(x) = (2 - 2\theta)x\{(0, 1)\} + \theta\{0\}$

$$H_0 : \theta = \frac{1}{3}, H_1 : \theta = \frac{1}{2}, n = 1, \alpha = 0.35$$

$$l = \frac{L_1}{L_0} = \frac{\rho_1(x)}{\rho_0(x)}$$

| x        | (0, 1)                         | 2                               |
|----------|--------------------------------|---------------------------------|
| 1        | $\frac{x}{4x/3} = \frac{3}{4}$ | $\frac{1/2}{1/3} = \frac{3}{2}$ |
| $H_0, p$ | $\frac{2}{3}$                  | $\frac{1}{3}$                   |
| $H_1, p$ | $\frac{1}{2}$                  | $\frac{1}{2}$                   |

$$G : l \geq C \quad P(l \geq C | H_0) \leq 0.35$$

$$G : l \geq \frac{3}{2}, \alpha_1 = \frac{1}{3}, \alpha_2 = P(l < C | H_1) = \frac{1}{2}$$

**Пример.** Те же условия что и в прошлом но  $n = 2$  !:

|        | (0, 1)   | 2   |
|--------|--|---|
| (0, 1) | $\frac{1 \cdot 1 \cdot x_1 \cdot x_2}{\frac{4}{3}x_1 \frac{4}{3}x_2} = \frac{9}{16}$ | $\frac{x_2 \cdot \frac{1}{2}}{\frac{4}{3}x_2 \frac{1}{3}} = \frac{9}{8}$            |
| 2      | $\frac{9}{8}$  | $\frac{\frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{3} \cdot \frac{1}{3}} = \frac{9}{4}$ |

$H_0$ :

|        |               |               |
|--------|---------------|---------------|
|        | (0, 1)        | 2             |
| (0, 1) | $\frac{4}{9}$ | $\frac{2}{9}$ |
| 2      | $\frac{2}{9}$ | $\frac{1}{9}$ |

$H_1$ :

|        |               |               |
|--------|---------------|---------------|
|        | (0, 1)        | 2             |
| (0, 1) | $\frac{1}{4}$ | $\frac{1}{4}$ |
| 2      | $\frac{1}{4}$ | $\frac{1}{4}$ |

$$G : l \geq C \quad P(l \geq C | H_0) \leq 0.35$$

$$G : l \geq \frac{9}{4}, \alpha_1 = \frac{1}{9}, \alpha_2 = P(l < C | H_1) = \frac{3}{4}, W = \frac{1}{4}$$

Асимптотический критерий  $n \rightarrow \infty$

$$l = \frac{L_1}{L_0} = \prod_{i=1}^n \frac{\rho_1(x_i)}{\rho_0(x_i)} \geq C$$

$$\ln l = \sum_{i=1}^n \ln \underbrace{\frac{\rho_1(x_i)}{\rho_0(x_i)}}_{\eta_i} \geq \ln C$$

ЦПТ

Параметрический бутстрап

$m$  - число появлений 2

$$l = \frac{L_1}{L_0} = \frac{\prod x_i \cdot \left(\frac{1}{2}\right)^2}{\left(\frac{4}{3}\right)^{n-m} \prod x_i \left(\frac{1}{3}\right)^m}$$

$$l = \left(\frac{3}{4}\right)^{n-m} \left(\frac{3}{2}\right)^n \geq C \quad P(l \geq C | H_0) \leq \alpha$$

$$\rho_0(x) = \frac{4}{3}x\{(0, 1)\} + \frac{1}{3}\{2\}$$

- $\vec{x}_n^* \rightarrow m_1^* \rightarrow l_1^*$
- ...
- $\vec{x}_n^* \rightarrow m_N^* \rightarrow l_N^*$

Вариационный ряд  $l_{(1)}^* \dots l_{(N)}^*$ ,  $k = [(1 - \alpha)N]$ ,  $C = l_{(k)}^*$

$n = 1$ ,  $\alpha = 0.02$ ,  $N = 50000$ ,  $C = 4$

$W = P(l \geq C | H_1)$ , те же самые процедуры с другим распределением  $W = \frac{k}{N}$ ,  $W = 0.17$

**Пример** (Т11 из задания).  $H_0 : \xi \sim p_0(x) = 1\{(0, 1)\}$   $H_1 : \xi \sim p_0(x) = \frac{e}{e-1}e^{-x}\{(0, 1)\}$

---

a)  $n = 1, \alpha$

$$l = \frac{L_1}{L_0} = \frac{\frac{e}{e-1}e^{-x}}{1} \geq C \quad e^{-x} \geq B \quad x \leq A$$

$$P(x \leq A|H_0) = \alpha \quad \int_0^A 1dx = A = \alpha$$

$$G : x \leq \alpha \quad \alpha_1 = \alpha$$

$$W = P(x \leq A|H_1) = \int_0^\alpha \frac{e}{e-1}e^{-x}dx = \frac{e}{e-1}(1 - e^{-\alpha}) \quad \alpha_2 = 1 - W$$


---

b)  $n = 2, \alpha$

$$l = \frac{L_1}{L_0} = \frac{\left(\frac{e}{e-1}\right)^2 e^{-x_1}e^{-x_2}}{1 \cdot 1} \geq C \quad x_1 + x_2 \leq A$$

$$P(x_1 + x_2 \leq A|H_0) = \alpha \quad \iint_{x_1+x_2 \leq A} 1dx_1dx_2 = \frac{A^2}{2} = \alpha \quad A = \sqrt{2\alpha}$$

$$G : x_1 + x_2 \leq \sqrt{2\alpha} \quad \alpha_1 = \alpha$$

$$W = P(x_1 + x_2 \leq A|H_1) = \iint_{x_1+x_2 \leq A} \left(\frac{e}{e-1}\right)^2 e^{-x_1-x_2}dx_1dx_2 =$$

$$= \left(\frac{e}{e-1}\right)^2 \int_0^A dx_1 \int_0^{A-x_1} e^{-x_1}e^{-x_2}dx_2 =$$

$$= \left(\frac{e}{e-1}\right)^2 \int_0^A e^{-x_1}(1 - e^{-(A-x_1)})dx_1 =$$

$$= \left(\frac{e}{e-1}\right)^2 \int_0^A (e^{-x_1} - e^{-A})dx_1 = \left(\frac{e}{e-1}\right)^2 (1 - e^{-A} - Ae^{-A}) \quad A = \sqrt{2\alpha}$$

$$\alpha = 1 - W$$


---

c)

$$\begin{aligned}
l &= \frac{L_1}{L_0} = \prod \frac{p_1(x_i)}{p_0(x_i)} \geq C \\
\ln l &= \sum \ln \underbrace{\frac{p_1(x_i)}{p_0(x_i)}}_{\eta_i} \geq \ln C \\
\frac{\sum \eta_i - nM\eta_i}{\sqrt{nD\eta_i}} &\rightsquigarrow N(0, 1) \\
P(\ln l \geq \ln C | H_0) &= \alpha \quad \eta = \ln\left(\frac{e}{e-1}e^{-x}\right) = \ln \frac{e}{e-1} - x \\
\ln l &= \sum \ln \frac{e}{e-1} - \sum x_i \geq \ln C \quad G: \sum x_i \leq A \\
P(\sum x_i \leq A | H_0) &= \alpha \quad P\left(\frac{\sum x_i - nMx}{\sqrt{nDx}} \leq \frac{A - nMx}{\sqrt{nDx}} | H_0\right) = \alpha \\
Mx &= \frac{1}{2} \quad Dx = \frac{1}{2}(b-a)^2 = \frac{1}{12} \\
\frac{A - n\frac{1}{2}}{\sqrt{n\frac{1}{2}}} &= u_\alpha \quad A = n\frac{1}{2} + \sqrt{n\frac{1}{12}}u_\alpha \\
G: \sum x_i &\leq n\frac{1}{2} + u_\alpha\sqrt{\frac{n}{12}} \quad \alpha_1 = \alpha \\
W &= P(\sum x_i \leq A | H_1) = P\left(\frac{\sum x_i - nMx}{\sqrt{nDx}} \leq \frac{A - nMx}{\sqrt{nDx}} | H_1\right) \\
Mx &= \int_0^1 x \frac{e}{e-1} e^{-x} dx = \frac{e-2}{e-1} \quad Mx^2 = \frac{2e-5}{e-1} \quad Dx = \frac{e^2-3e+1}{(e-1)^2} \\
W &= \int_{-\infty}^B \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad B = \frac{\frac{n}{2} + U_\alpha \sqrt{\frac{n}{12}} - n\frac{e-2}{e-1}}{\sqrt{n\frac{e^2-3e+1}{(e-1)^2}}} \quad \alpha_2 = 1 - W \\
B &= \frac{\sqrt{n} \cdot 0.082 + U_\alpha \sqrt{\frac{1}{12}}}{\sqrt{\frac{e^2-3e+1}{(e-1)^2}}} \xrightarrow{n \rightarrow \infty} \infty
\end{aligned}$$

Критерий состоятелен.

---


$$\alpha = 0.05, n = 1, x \leq 0.05, W = 0.077, \alpha_2 = 0.923$$

$$\alpha = 0.05, n = 2, x_1 + x_2 \leq 0.316, W = 0.102, \alpha_2 = 0.898$$

$$\alpha = 0.05, n = 10, \sum x_i \leq 3.5, B = -0.78, W = 0.22, \alpha_2 = 0.78$$