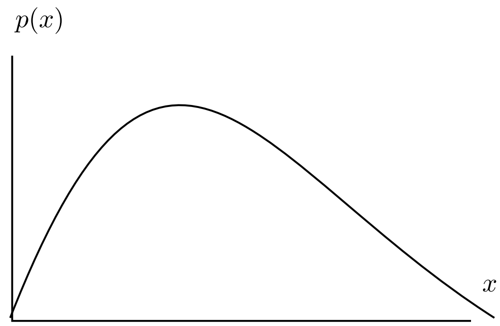
1 Занятие 1

1.1 Основные распределения в Мат Стат

1.1.1 Гамма распределение

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



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

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, $\eta = \xi_1 + \xi_2$

$$\Phi(y) = P(\eta < y) = P(\xi_2 + \xi_2 < y) = \iint_{x_1+x_2<y} p(x_1, x_2) dx_1 dx_2 =$$
$$= \int_0^y dx_2 \int_0^{y-x_1} \frac{\lambda^{a_1}}{\Gamma(a_1) x_1^{a_1-1} e^{-\lambda x_1}} \frac{\lambda^{a_2}}{\Gamma(a_2) x_2^{a_2-2} e^{-\lambda x_2}} dx_2$$
$$\varphi(y) = \Phi'(y)$$

□

1.1.2 Распределение Парсона χ^2

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

$$\Phi(y) = P(\xi^2 < y) = \begin{cases} y \le 0 & : 0 \\ y > 0 & : P(-\sqrt{y} < \xi < \sqrt{y}) \end{cases}$$
$$\varphi(x) = \begin{cases} \frac{1}{2\sqrt{y}} F'(\sqrt{y}) + \frac{1}{2\sqrt{y}} F'(-\sqrt{y}), \ y > 0 \\ 0, \ y < 0 \end{cases}$$
$$p(x) = \frac{e^{x^2/2}}{\sqrt{2\pi}}$$
$$\varphi(y) = \begin{cases} \frac{1}{\sqrt{y}} \frac{e^{-y/2}}{\sqrt{2\pi}}, \ y > 0 \\ 0, \ y \le 0 \end{cases}$$
$$p(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \{(0; +\infty)\} \qquad \lambda = \frac{1}{2} \qquad a = \frac{1}{2}$$
$$\xi^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2}) \qquad \xi_1^2 + \cdots + \xi_n^2 \sim \Gamma(\frac{1}{2}, \frac{n}{2}) = \chi^2(n)$$

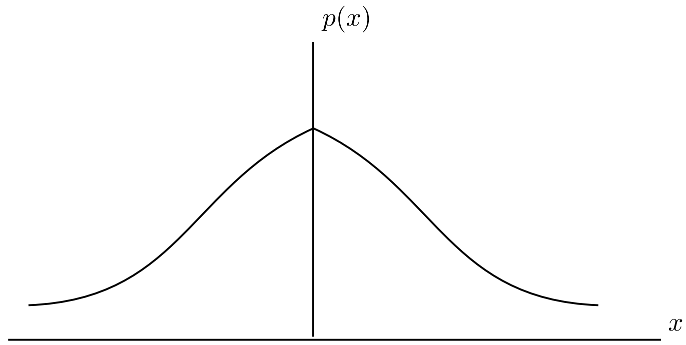
n - число степеней свободы

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

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

1.2 Распределение Стьюдента (Госсет)

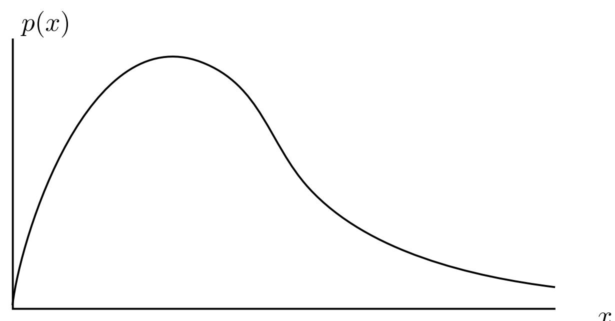
$\xi \sim N(0,1)$, $\eta \sim \chi^2(m)$ - независимы, $\frac{\xi}{\sqrt{\eta/m}} \sim t(m)$



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

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

$$p(\vec{x}) = \frac{1}{(\sqrt{2\pi})^n}\frac{1}{\sqrt{detK}}e^{-\frac{1}{2}(\vec{x}-\vec{a})^T K^{-1}(\vec{x}-\vec{a})}$$
$$\vec{\xi} \sim N(\vec{a}, R)$$

Свойства:

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

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

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

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

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

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.

$$\vec{\eta} \sim N(\vec{0}, \underbrace{CC^T}_{E})$$

$$\eta_1^2 + \dots + \eta_n^2 = \vec{\eta}^T\vec{\eta} = \vec{\xi}^T C^T C \vec{\xi} = \xi_1^2 + \dots + \xi_n^2$$

$$\varkappa = \eta_{k-1}^2 + \dots + \eta_n^2$$

$$\varkappa = \chi^2(n-k)$$

□

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

1. $\varphi = \sqrt{n}\frac{\bar{\xi}-a}{\sigma} \sim N(0,1)$, $\bar{\xi} = \frac{1}{n}\sum_1^n \xi_i$
2. $\psi = \sum_{i=1}^n \frac{(\xi_i-\bar{\xi})^2}{\sigma^2} \sim \chi^2(n-1)$
3. φ и ψ независ.

Доказательство.

$$\varphi = \frac{1}{\sqrt{n}} \frac{\sum^n \xi_i - na}{\sigma} = \frac{1}{\sqrt{n}} \sum^n \left(\frac{\xi_i - a}{\sigma}\right)$$

$$\frac{\xi_i - a}{\sigma} = \frac{1}{\sigma}\xi_i - \frac{a}{\sigma} \sim N(\frac{a}{\sigma} - \frac{a}{\sigma}, \sigma^2\frac{1}{\sigma^2}) = N(0,1)$$

$$\varphi = \frac{1}{\sqrt{n}} \sum^n \eta_i = (\frac{1}{\sqrt{n}} \dots \frac{1}{\sqrt{n}})\vec{\eta} \sim N(\vec{0}, AA^T) = N(0,1)$$

1) Доказан

$$\psi = \sum^n \left(\underbrace{\frac{\xi_i - a}{\sigma}}_{\eta_i \sim N(0,1)} - \underbrace{\frac{\bar{\xi} - a}{\sigma}}_{\bar{\eta}} \right)^2 = \sum^n (\eta_i - \bar{\eta})^2 = \sum^n (\eta_i^2 - 2\eta_i\bar{\eta} + (\bar{\eta})^2) =$$

$$= \sum^n \eta_i^2 - 2\bar{\eta}\sum^n \eta_i + n(\bar{\eta})^2 = \sum^n \eta_i^2 - n(\bar{\eta})^2$$

$$\eta_i \sim N(0,1) \qquad \zeta^2 = n\bar{\eta}^2 \qquad \zeta = \sqrt{n}\bar{\eta} = \frac{1}{\sqrt{n}}\sum^n \eta_i = A\vec{\eta} = \varphi$$

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

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

По лемме Фишера $\psi \sim \chi^2(n-1)$, ψ и $A\bar{\eta}$ 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, $\vec{\eta_1} = \zeta_1 e'_1 + \dots + \zeta_m e'_m$, $\vec{\eta_2} = \zeta_{m+1} e'_{m+1} + \dots + \zeta_n e'_n$

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

□

2 Порядковые случайные величины

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

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

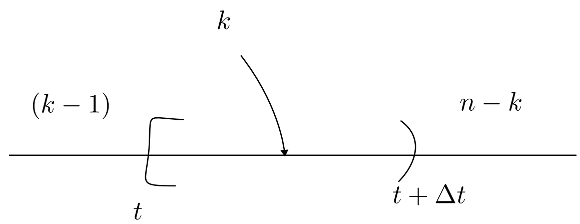
$$\Phi(y) = P(\eta < y) = 1 - P(\eta \geq y) = 1 - P(min(\xi_1, \dots, \xi_n) \geq y) =$$
$$= 1 - P(\xi_1 \geq y, \dots, \xi_n \geq y) = 1 - \prod_{i=1}^{n} P(\xi_i \geq y) =$$
$$= 1 - \prod_{i=1}^{n}(1 - P(\xi_i < y)) = 1 - (1 - F(y))^n$$

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

Порядковые величины:

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

Положим $F(x)$ - непрерывна:



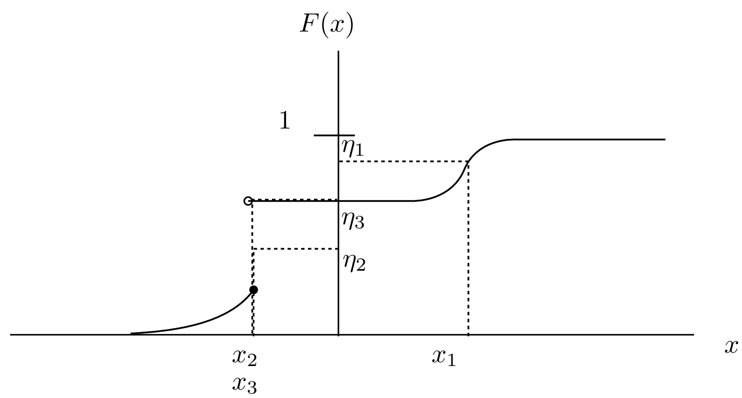
$$P(t \ \le \xi_{(k)} < t+\Delta t) =$$
$$= nP(t \le \xi < t+\Delta t) C_{n-1}^{k-1} (P(\xi < t))^{k-1} (P(\xi \ge t+\Delta t))^{n-k} C_{n-k}^{n-k} =$$
$$= \varkappa(t+\Delta t) - \varkappa(t)$$
$$n \frac{F(t+\Delta t) - F(t)}{\Delta t} C_{n-1}^{k-1} (F(t))^{k-1} (1 - F(t+\Delta t))^{n-k} = \frac{\varkappa(t+\Delta t) - \varkappa(t)}{\Delta t}$$
$$\varkappa(t) = np(t) C_{n-1}^{k-1} (1-F(t))^{n-k} (F(t))^{k-1}$$

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

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

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

3 Моделирование случайных величин



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

4 Основные задачи статистики

Явление $\to$ математическая модель явления $\to$ вероятностная модель явления ξ_i.

Выборка - n наблюдений над явлением $\to$описательная статистика (непараметрическая), выбор классов (e.g. $\varepsilon \sim N(a, \sigma^2)$) $\to$ 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 ε и внутри k).

Распределения полагаем нормальными и так далее.

1. Определение параметров и оценка их точности

2. Проверка статистических гипотиз

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

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

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

1. конкретные наблюдения над явлением

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

4.1 Свойства оценок

Θ - множество значений θ, $\theta \in \Theta$, $\tilde{\theta}(\vec{x}_n)$ - оценка θ 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. $\forall\varepsilon > 0\ P(|\tilde{\theta} - \theta| \geq \varepsilon) \underset{n\to\infty}{\to} 0$)
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 θ и $D[\tilde{\theta}] \underset{n\to\infty}{\to} 0$, то $\tilde{\theta}$ является состоятельной оценкой ($\forall\theta \in \Theta$)

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

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

$\square$

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

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

$$\tilde{\theta}_1 = 2\bar{x} = 2\frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\tilde{\theta}_2 = \min x_i$$
$$\tilde{\theta}_3 = \max x_i$$
$$\tilde{\theta}_4 = x_1 + \frac{1}{(n-1)}\sum_{i=2}^{n} x_i$$

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

Рассматриваем $\tilde{\theta}_1$:

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

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

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

Рассматриваем $\tilde{\theta}_2$:

$$M[\tilde{\theta}_2] = \int_{-\infty}^{\infty} y\varphi(y)dy$$

$$\Phi(y) = 1 - (1 - F(y))^n \qquad \varphi(y) = n(1 - F(y))^{n-1}p(y)$$

$$M[\tilde{\theta}_2] = \int_0^{\theta} n(1 - \frac{y}{\theta})^{n-1}\frac{1}{\theta}ydy =$$

$$t = 1 - \frac{y}{\theta}$$

$$= -\int_1^0 nt^{n-1}(1-t)\theta dt = \int_0^1 n\theta t^{n-1}dt - \int_0^1 n\theta t^n dt =$$

$$= n\theta[1 - \frac{n}{n+1}] = \frac{\theta}{n+1} \quad \text{смещённая}$$

$$\tilde{\theta}'_2 = (n+1)x_{min} = (n+1)\tilde{\theta}_2 \quad \text{несмещённая}$$

$$M[\tilde{\theta}_2^2] = \int_0^{\theta} n(1 - \frac{y}{\theta})^{n-1}\frac{1}{\theta}y^2 dy =$$

$$= -\int_1^0 nt^{n-1}(1-t)^2\theta^2 dt = \frac{2\theta^2}{(n+1)(n+2)}$$

$$D[\tilde{\theta}_2] = \frac{2\theta^2}{(n+1)(n+2)} - \frac{\theta^2}{(n+1)^2} = \theta^2\left[\frac{2(n+1)-(n+2)}{(n+1)^2(n+2)}\right] =$$

$$= \theta^2\left[\frac{n}{(n+1)^2(n+2)}\right] \underset{n\to\infty}{\to} 0 \quad \text{достаточное не выполнятеся}$$

$$D[\tilde{\theta}'_2] = (n+1)^2 D[\tilde{\theta}_2] = \frac{\theta^2 n}{n+2} \not\to 0$$

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

$$\forall\theta > 0\, \forall\varepsilon > 0 \hookrightarrow P(|\tilde{\theta}'_2 - \theta| \geq \varepsilon) \underset{n\to\infty}{\to} 0$$

$$P(|\tilde{\theta}'_2 - \theta| \geq \varepsilon) \geq P(\tilde{\theta}'_2 > \theta + \varepsilon) =$$

$$= P((n+1)x_{min} \geq \theta + \varepsilon) = P(x_{min} \geq \frac{\theta+\varepsilon}{n+1}) =$$

$$= 1 - P(x_{min} < \frac{\theta+\varepsilon}{n+1}) = 1 - (1 - (1 - F(\frac{\theta+\varepsilon}{n+1}))^n) =$$

$$= (1 - (\frac{\theta+\varepsilon}{\theta(n+1)}))^n \underset{n\to\infty}{\to} e^{-\frac{\theta+\varepsilon}{\theta}} > 0$$

Не является состоятельной!

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

$$P(\tilde{\theta}_2 < \theta - \varepsilon) + \underbrace{P(\tilde{\theta}_2 > \theta + \varepsilon)}_{=0,\text{т.к.}\tilde{\theta}_2 = x_{min}}$$

$$P(x_{min} < \theta - \varepsilon = \Phi(\theta - \varepsilon)) = 1 - (1 - \frac{\theta - \varepsilon}{\theta})^n = 1 - \left(\frac{\varepsilon}{\theta}\right)^n \underset{n\to\infty}{\rightarrow} 1$$

Не является состоятельной!

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

$$M[\tilde{\theta}_3] = \int_{-\infty}^{+\infty} z\psi(z)dz = \int_0^\theta n\frac{z^n}{\theta^n}dz = \frac{n}{n+1}\theta \qquad \text{смешённая}$$

$$\Psi(z) = (F(z))^n \qquad \psi(z) = n(F(z))^{n-1}p(z) = n\left(\frac{z}{\theta}\right)^{n-1}\frac{1}{\theta}\{(0;\theta)\}$$

$$D[\tilde{\theta}_3]\frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n\theta^2}{(n+2)(n+1)^2}$$

$$D[\tilde{\theta}'_3]\frac{(n+1)^2}{n^2}D[\tilde{\theta}_3] = \frac{\theta^2}{n(n+2)} \underset{n\to\infty}{\rightarrow} 0 \qquad \text{состоятельная}$$

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

$$\forall\theta > 0 \; \forall\varepsilon > 0$$

$$P(|\tilde{\theta}'_2 - \theta| \geq \varepsilon) = P(x_{max} < \theta - \varepsilon) + \underbrace{P(x_{max} \geq \theta + \varepsilon)}_{=0} =$$

$$= (F(\theta - \varepsilon))^n = \begin{cases} 0 < \varepsilon < \theta: \; \left(\dfrac{\theta - \varepsilon}{\theta}\right)^n \underset{n\to\infty}{to} 0 \\ \varepsilon \geq \theta: \; (0)^n \underset{n\to\infty}{\rightarrow} 0 \end{cases}$$

Является состоятельной!

Рассматриваем $\tilde{\Theta}_4$:

$$M[\tilde{\theta}_4] = M[x_1 + \sum_{i=2}^n x_i] = M[x_1] + \frac{1}{n-1}\sum_{i=1}^n M[x_i] = \frac{\theta}{2} + \frac{\theta}{2} = \theta$$

$$D[\tilde{\theta}_4] = D[\xi] + \frac{1}{(n-1)^2}(n-1)D[\xi] = \frac{\theta^2}{12}\frac{n}{n-1} \underset{n\to\infty}{\not\rightarrow} 0$$

Достаточое 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: $\xi_1, \ldots, \xi_n$ незав., одинак распр. $\Rightarrow \frac{1}{n}\sum_{i=1}^n \xi_i \xrightarrow{p} M[\xi]$.

$$x_1 \xrightarrow{p} x_1 \qquad \frac{1}{n-1}\sum_{i=2}^n x_i \xrightarrow{p} \frac{\theta}{2}$$

$$\tilde{\theta}_4 \xrightarrow{p} x_1 + \frac{\theta}{2}$$

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 θ 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 θ.

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

$$\tilde{\theta}_3 = \frac{\tilde{\theta}_1 + \tilde{\theta}_2}{2} \qquad M[\tilde{\theta}_3] = \theta$$

$$D[\tilde{\theta}_3] = \frac{1}{4}D[\tilde{\theta}_1] + D[\tilde{\theta}_2] + \frac{1}{2}cov(\tilde{\theta}_1, \tilde{\theta}_2) = \frac{1}{2}D[\tilde{\theta}_1] + \frac{1}{2}cos(\tilde{\theta}_1, \tilde{\theta}_2)$$

$$D[a\xi + b\eta] = a^2D[\xi] + b^2D[\eta] + 2abcov(\xi, \eta)$$

$$|cov(\tilde{\theta}_1, \tilde{\theta}_2)| \le \sqrt{D\tilde{\theta}_1 D\tilde{\theta}_2} = D[\tilde{\theta}_1]$$

$$D[\tilde{\theta}_3] \le D[\tilde{\theta}_1] \qquad D[\tilde{\theta}_3] = D[\tilde{\theta}_1]$$

$$cov(\tilde{\theta}_1, \tilde{\theta}_2) = D[\tilde{\theta}_1] \quad \Rightarrow \quad r = 1 \Leftrightarrow \tilde{\theta}_1 = a\tilde{\theta}_2 + b$$

$$M[\tilde{\theta}_1] = M[\tilde{\theta}_2] = \theta \qquad D[\tilde{\theta}_1] = D[\tilde{\theta}_2]$$

$$\begin{cases} \theta = a\theta + b \\ a^2D[\tilde{\theta}_2] = D[\tilde{\theta}_2] \end{cases} \Rightarrow \begin{cases} a = 1 \\ b = 0 \end{cases}$$

$$\Rightarrow \tilde{\theta}_1 = \tilde{\theta}_2$$

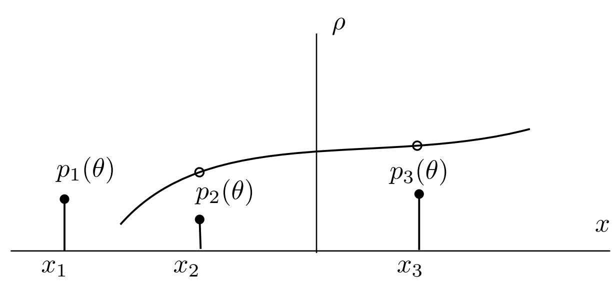
Противоречие. $\square$

Будем рассматриватть параметрические вероятностные модели:

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

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

$$\rho(x,\theta) = \underbrace{p(x,\theta)\{E\}}_{\text{непр. часть}} + \underbrace{\sum_k p_k(\theta)\{x_k\}}_{\text{дискр. часть}}$$



5.1 Информация Фишера

$$I(\theta) = M\left[(\frac{\partial \ln \rho(x,\theta)}{\partial\theta})^2\right] =$$

$$= \int_E \left(\frac{\partial \ln p(x,\theta)}{\partial\theta}\right)^2 p(x,\theta)dx + \sum_k \left(\frac{\partial \ln p_k(x,\theta)}{\partial\theta}\right)^2 p_k(\theta)$$

$I(\vec{\theta})$ - 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. Вероятностьная 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 θ на Θ
2. $\frac{\partial}{\partial\theta}\int_A \rho(x,\theta)dx = \int_A \frac{\partial}{\partial\theta}\rho(x,\theta)dx$ на Θ
3. $I(\theta)$ непр на Θ и $I(\theta) > 0$ на Θ

Определение 5.3. Вероятностьная 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 Θ
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 Θ, $i = 1, \dots, m$
3. $I(\vec{\theta})$ положительно определена на Θ и $I_{ij}(\vec{\theta})$ непрер. на Θ

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

1. $\rho(x, \theta)$ k раз непрерывно дифф по θ на Θ ($k \geq 2$)
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 θ на Θ ($k \geq 2$)
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 Θ по θ, тогда оценка регулярна.

5.2 Неравенство Крамера-Рао

Теорема 5.3. Пусть модель является регулярной, $\tilde{g}(\vec{x}_n)$ является регулярной оценкой 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 $\vec{x}_n$, $B = A \times \cdots \times A$

$$\frac{\partial}{\partial \theta} \underset{B}{\int \cdots \int} L(\vec{x}, \theta) d\vec{x} = \frac{\partial}{\partial \theta} 1 = 0$$

в силу регулярности модели

$$\underset{B}{\int \cdots \int} \frac{\partial}{\partial \theta} L d\vec{x} = 0$$

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

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

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

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

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

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

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

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

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

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

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

□

Следствие 5.1.

1. оценка параметра θ, $g(t) = \theta$,
$$D[\tilde{\theta}] \ge \frac{1}{nI(\theta)}$$
2. многомерный аналог нер. Крамера-Рао
$$D[\tilde{g}(\vec{x}_n)] \ge \frac{1}{n}\nabla^T g(\vec{\theta}) I^{-1}(\vec{\theta}) \nabla g(\vec{\theta})$$

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)

$$\xi \sim \rho(x, \theta) = \theta\{1\} + (1-\theta)\{0\} \qquad \theta \in (0, 1)$$

$$\xi \sim Bi(1, \theta) \qquad \nu = \frac{m}{n}$$

$$\vec{x}_n = (0, 0, 1, \dots) \qquad \nu = \frac{1}{n}\sum x_i = \bar{x} \qquad \tilde{\theta} = \bar{x}$$

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

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

   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 ваборка

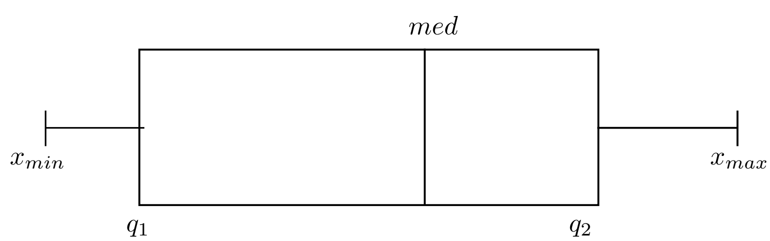
$$x_{min} = x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)} = x_{max}$$

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

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

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

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 $< x$.

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

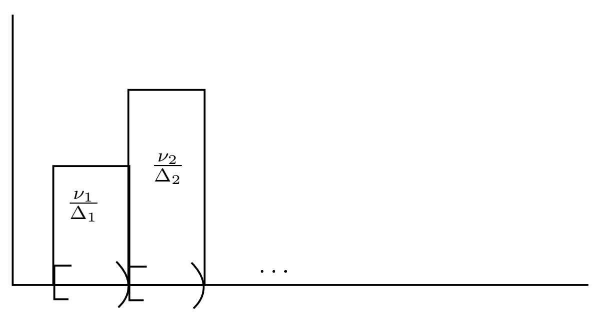
$\tilde{F}$ является несмещённой, состоятельной и эффективной оценкой $F(x)$ (по Т. о частоте).

8. Гистограмма

   Статистический ряд

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

   Эмпирически $k = 1 + \log_2 n$



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

9. числовые характеристики

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

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

   - 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

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

Теорема наследования сходимости:

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

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

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

- несмещённость

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

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

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

коэффициент асимметрии

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

10. оценка распределения статистики

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

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

7 Методы нахожд. параметров модели

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

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

1. Метод моментов (Пирсон)

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

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

   Замечание. ОММГ (ОММ Групированная) для статистического ряда:

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

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

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

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

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

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

$fix\,\theta$, $L(\vec{x}_n)$ - 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 $L(\theta) \to \max$

Замечание. ОМПГ для статистического ряда

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

$$P_1(\theta) = \int_{-\infty}^{y_2} \rho(x, \theta) dx$$

$$P_2(\theta) = \int_{y_2}^{y_3} \rho(x, \theta) dx$$

$$\dots$$

$$P_k(\theta) = \int_{y_k}^{+\infty} \rho(x, \theta) dx$$

$$L(\theta) = P_1^{m_1} \dots P_k^{m_k} \qquad L(\theta) \to \max$$

8 Асимптотические свойства оценок

$n \to \infty$

1. состоятельность $\forall \theta \in \Theta \hookrightarrow \tilde{g}(\vec{x}_n) \xrightarrow{p} g(\theta)$
2. асимптотическая несмещённость $\forall \theta \in \Theta \hookrightarrow M[\tilde{g}(\vec{x}_n)] \underset{n \to \infty}{\to} g(\theta)$
3. асимптотическая эффективность $\forall \theta \in \Theta \hookrightarrow nD[\tilde{g}(\vec{x}_n)] \underset{n \to \infty}{\to} \frac{g'^2(\theta)}{I(\theta)}$
4. асимпотическая 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

$$\alpha_k(\vec{\theta}) = \tilde{\alpha}_k \to \tilde{\vec{\theta}} = f(\tilde{\alpha}_1, \dots, \tilde{\alpha}_k)$$

$$\tilde{\alpha}_k \xrightarrow{p} \alpha_k$$

f - непр в $\alpha_1, \dots, \alpha_k$, тогла по теореме наследования $\tilde{\vec{\theta}} \xrightarrow{p} f(\alpha_1, \dots, \alpha_k) = \vec{\theta}$ 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:

$$\frac{\sum x_i^k - nM[\xi^k]}{\sqrt{nD[\xi^k]}} \rightsquigarrow N(0,1)$$

$$\frac{\tilde{\alpha}_k - a_k}{\sqrt{\alpha_{2k} - a_k^2}}\sqrt{m} \rightsquigarrow N(0, \underbrace{\alpha_{2k} - a_k^2}_{\sigma_k^2})$$

$$(f(\tilde{\alpha}_k) - f(\alpha_j))\sqrt{n} \rightsquigarrow N(0, f'^2(\alpha_k)\sigma_k^2)$$

$$(\tilde{\theta} - \theta)\sqrt{n} \rightsquigarrow N(0, f'^2(\alpha_k)\sigma_k^2)$$

8.1.1 Многомерное ЦПТ

$$\alpha = \begin{pmatrix} \alpha_{s_1} \\ \vdots \\ \alpha_{s_k} \end{pmatrix} \qquad \tilde{\alpha} = \begin{pmatrix} \tilde{\alpha}_{s_1} \\ \vdots \\ \tilde{\alpha}_{s_k} \end{pmatrix} \qquad \tilde{\alpha}_{s_1} = \frac{1}{n}\sum_{i=1}^{n} x_i^{s_1}$$

$$(\tilde{\alpha} - \alpha)\sqrt{n} \rightsquigarrow N(\vec{0}, K) \qquad K_{ij} = \alpha_{s_i+s_j} - \alpha_{s_i}\alpha_{s_j}$$

$f(x) \in C'(R^k)$ 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 Θ открыто.

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

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

$$\underbrace{\frac{d \ln L(\theta)}{d\theta}}_{0} = \frac{d \ln L}{d\theta}(\theta) + \frac{d^2 \ln L}{d\theta^2}(\theta^*)(\tilde{\theta} - \theta)$$

$$\tilde{\theta} - \theta = -\frac{(\ln L)'(\theta)}{(\ln L)''(\theta^*)}$$

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

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

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

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

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

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

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

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

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

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

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

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

$$\frac{\sum\frac{d\ln p}{d\theta} - \overbrace{nM[\frac{d\ln p}{d\theta}]}^{=0}}{\sqrt{nD[\frac{d\ln p}{d\theta}]}}\rightsquigarrow N(0,1)$$

Лемма Слуцкого: $\xi_n \xrightarrow{F} \xi$, $\eta_n \xrightarrow{p} C$, $x_n\eta_n \xrightarrow{F} \xi C$

$$\tilde\theta - \theta = -\frac{\frac{\sum \frac{d\ln p}{d\theta}}{\sqrt{nI(\theta)}}\sqrt{nI(\theta)}}{\frac{1}{-nI(\theta)}\sum \frac{d^2\ln p}{d\theta^2}(-nI(\theta))}$$

$$a = \frac{\sum \frac{d\ln p}{d\theta}}{\sqrt{nI(\theta)}} \rightsquigarrow N(0,1) \qquad b = \frac{1}{-nI(\theta)}\sum \frac{d^2\ln p}{d\theta^2} \rightsquigarrow 1$$

$$(\tilde\theta - \theta)\sqrt{nI(\theta)} = \frac{a}{b} \rightsquigarrow N(0,1)$$

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

$$D[(\tilde\theta - \theta)\sqrt{n}] = nD[\tilde\theta] \underset{n\to\infty}{\to} \frac{1}{I(\theta)}$$

Асимптотическая эффективность. $\square$

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 Θ

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

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

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

- многомерный аналог

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

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

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

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

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

$M[x_{max}] = \frac{n}{n+1}\theta \underset{n\to\infty}{\to} \theta$ асим. несм.

$x_{max} \xrightarrow{p} \theta$ сост.

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

Пусть асим. норм.

$$\sqrt{n}(\underbrace{\tilde{\theta}}_{x_{max}} - \theta) \rightsquigarrow N(0, \sigma^2)$$

$$P(\sqrt{n}(x_{max} - \theta) < x) \underset{n\to\infty}{\to} \Phi(x)\ \forall x \in \mathbb{R}$$

$$P(\sqrt{n}(x_{max} - \theta) < 0) = 1 \to 1 \qquad \Phi(0) = \frac{1}{2}$$

Противоречие $\Rightarrow$ не является асим. нормальной.

9 Доверительный интервал

Определение 9.1. Доверительным интервалом величины h вероятностной модели называется случайный интервал, который накрывает значение h с вероятностью, не меньшей β.

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

β - доверительная вероятность, чаще всего $0.9, 0.95, 0.99$.

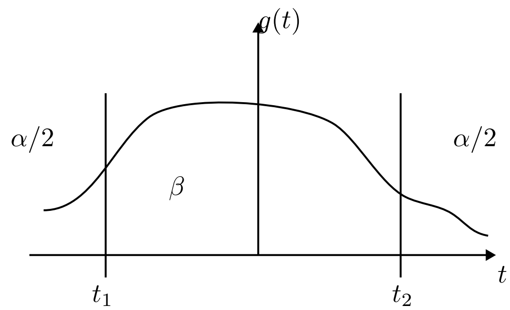
9.1 Методы построения доверит. инт.

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

9.1.1 Точный метод

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

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



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

$$t_1 = g_{\alpha/2} = g_{\frac{1-\beta}{2}} \qquad t_2 = g_{\beta+\frac{\alpha}{2}} = g_{\frac{1+\beta}{2}}$$
$$P(t_1 < f(h, \vec{x}_n) < t_2) = \beta$$
$$g_1(\vec{x}_n) < h < g_2(\vec{x}_n)$$

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

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

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

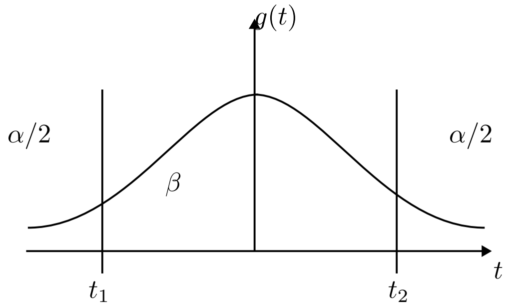
Теорема Фишера:

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

$$t_1 = \chi^2_{\frac{1-\beta}{2}}(n-1) \qquad t_2 = \chi^2_{\frac{1+\beta}{2}}(n-1)$$
$$P(t_1 < \frac{S^2(n-1)}{\theta_2^2} < t_2) =; B$$
$$\frac{S^2(n-1)}{t_2} < \theta_2^2 < \frac{S^2(n-1)}{t_1}$$
$$\sqrt{\frac{S^2(n-1)}{t_2}} < \theta_2 < \sqrt{\frac{S^2(n-1)}{t_1}}$$

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

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



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

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

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

9.2 Асимптотический метод

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

По ОММ

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

$$\alpha = \begin{pmatrix} \alpha_{s_1} \\ \vdots \\ \alpha_{s_k} \end{pmatrix} \qquad \tilde{\alpha} = \begin{pmatrix} \tilde{\alpha}_{s_1} \\ \vdots \\ \tilde{\alpha}_{s_k} \end{pmatrix} \qquad K_{ij} = \alpha_{s_i+s_j} - \alpha_{s_i}\alpha_{s_j}$$

$$(\tilde{\alpha}_j - \alpha_j)\sqrt{n} \rightsquigarrow N(0, \alpha_{2k} - \alpha_k^2)$$

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

$$\tilde{\alpha}_k \xrightarrow{p} \alpha_k \qquad \sqrt{\tilde{\alpha}_{2k} - \tilde{\alpha}_k^2} \xrightarrow{p} \sqrt{\alpha_{2k} - \alpha_k^2}$$

$$\frac{\sqrt{\alpha_{2k} - \alpha_k^2}}{\sqrt{\tilde{\alpha}_{2k} - \tilde{\alpha}_k^2}} \xrightarrow{p} 1$$

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

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

$$\boxed{\frac{\sqrt{n}(g(\tilde{\alpha}) - g(\alpha))}{\sqrt{\nabla^T g(\tilde{\alpha}) K(\tilde{\alpha}) \nabla g(\tilde{\alpha})}} \rightsquigarrow N(0,1) \qquad \text{ОММ}}$$

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

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

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

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

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

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

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

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

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

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

$$-1.96 < \frac{\sqrt{20}(0.4275-\theta)}{\sqrt{\frac{9}{4}\cdot 0.342}} < 1.96$$
$$0.0435 < \theta < 0.811 \qquad l = 0.768$$

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

$$\frac{\sqrt{n}(\tilde{\theta}-\theta)}{\sqrt{I^{-1}(\theta)}} \rightsquigarrow N(0,1)$$
$$I(\theta) = \frac{1}{\theta(2-\theta)}$$
$$\frac{\sqrt{n}(\tilde{\theta}-\theta)}{\sqrt{\tilde{\theta}(2-\tilde{\theta})}} \rightsquigarrow N(0,1)$$
$$-1.96 < \frac{\sqrt{20}(0.4-\theta)}{\sqrt{0.4\cdot 1.6}} < 1.96$$
$$0.049 < \theta < 0.751 \qquad l = 0.702$$

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

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

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

   (b) ...

   (c) ???

(d) ...

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

(f) Profit!

Вариационный ряд $\Delta^*_{(1)}, \ldots, \Delta^*_{(1000)}$

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

2. Параметричекий бутстрап

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

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

$\Delta^*_i = \tilde{h}^* - \tilde{h}$, $\Delta^*_{(1)} \ldots \Delta^*_{(N)}$

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

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

1. Непараметрический бутсрап $\tilde{\theta} = 0.4$ ОМП $\tilde{\theta} = 2(1 - \nu)$
   - $\vec{x}_n^* = 0.53, \ldots$, $m = 14$, $\Delta^*_1 = \tilde{\theta}^* - \tilde{\theta} = 0.2$
   - ...
   - $\Delta^*_{1000} = 0$

$$\beta = 0.95 \qquad k_1 = 25 \qquad k_2 = 975$$
$$t_1 = \Delta^*_{(25)} = -0.3 \qquad t_2 = \Delta^*_{(975)} = 0.4$$
$$P(-0.3 < \tilde{\theta}^* - \tilde{\theta} < 0.4) \approx 0.95$$
$$P(-0.3 < \tilde{\theta} - \theta < 0.4) \approx 0.95$$
$$0 < \theta < 0.7 \qquad l = 0.7$$

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:

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

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 маловеороятного

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, $A_1 + \dots + A_n = \Omega$, $A_i A_j = \emptyset$ $i \neq j$, $P(A_i) > 0$

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

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

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

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

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

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

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

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 p-value $\leq \alpha$, то H_0 отвергается, результаты значимы, p-value - мера значимости

2. Если 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.

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

Рассмотрим числа Фибоначи (первые 100 штпук)

	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 приминимости

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

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

Нет оснований отвергать H_0.

11.2 Критерий Колмогорова (непр. распр.)

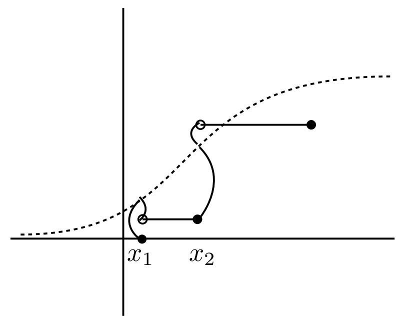
$H_0 : \xi \sim F(x)$, $H_1 : \bar{H}_0$, $\vec{x}_n$

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

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

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

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



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

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

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

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

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

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

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

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

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

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

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

Теорема 11.3. Если H_0 верная и $\tilde{\vec{\theta}}$ 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_1	25	10	15	30	20
p_i	1/5	1/5	1/5	1/5	1/5
np_i	20	20	20	20	20

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

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

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

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

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

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

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

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

Отвергаем H_0.

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

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

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

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

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

Считаем экстремум численно.

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

После объединения и пересчёта:

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

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

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

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

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

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

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

Пример (Эрланг)**.**

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

11.5 Проверка гипотезы однородности

$A_1 \dots A_k$ - полная группа

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

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

$$n = n_1 + \dots + n_l \qquad \nu_j = \frac{\sum_{i=1}^{l} m_i j}{n}$$

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

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

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

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
$\frac{20}{58}$	$\frac{38}{58}$

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

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

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

11.6 Проверка гипотезы независимости

(ξ, η), $H_0 : \xi$ и η - 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}

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

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

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	$\frac{46}{246}$
	$\frac{85}{246}$	$\frac{80}{246}$	$\frac{81}{246}$	

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

Нет оснований отвергать H_0.

11.7 Проверка гипотезы случайности

H_0 : элементы выборки независимы и одинакого 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. $n = 10$, $\vec{x}_n = (4, 10, 8, 8, 6, 7, 7, 9, 8, 9)$

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

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

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

Нет оснований отвергать H_0.