Математическая статистика

$X_1, \dots, X_n$ - случ. величины. Доступен конеч. набор данных. Как получить $\mathbb{P}$?
Если $X_1, \dots, X_n$ нез., то наз-ся независимой выборкой (аналогично однород.)
Выборка ≡ незав. однород. выборка
$(X_1, \dots, X_n) = \mathbb{X}$ $\quad (\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ $\quad \mathcal{P} = \{\mathbb{P}\}$

мн-во всех мн-в, которые можно постр. из n-мер. параллелепипедов с помощью счётного числа теоретико-множ. операций

Если $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ - параметрич. вид. Задача сводится к поиску „истинного" параметра

Пример: $\mathcal{P} = \{\mathcal{N}_{a,\sigma^2} : a \in \mathbb{R}, \sigma^2 > 0\}$ Распред. отождеств. с точкой в верхней полуплоскости.

1) $T(\mathbb{X}) \approx \theta$ Задача точечного оценивания.
2) $\Theta \subseteq \mathbb{R}$: $T_1(\mathbb{X}) \le \theta \le T_2(\mathbb{X})$ Задача интервального оценивания.
3) $\{\theta = \theta_0\} = H_0$ Задача проверки статистических гипотез.

$(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$ - статистическая структура. Задано много мер. (семейство вероятностных мр-в)

Можем опред-ть: $F(x) = \mathbb{P}(X_1 < x), x \in \mathbb{R}$. $\mathbb{P}$ неизв. => F неизв.
Как опред. $F(x)$ по выборке $\mathbb{X}$?
$(X_{(1)}, \dots, X_{(n)}) = \tilde{\mathbb{X}}$, где $X_{(1)} \le \dots \le X_{(n)}$

$$\mathbb{P}(X_{(1)} < x) = 1 - \mathbb{P}(X_{(1)} \ge x) = 1 - \mathbb{P}(X_1 \ge x, \dots, X_n \ge x) = 1 - \prod_{i=1}^n \mathbb{P}(X_i \ge x) =$$
$$= 1 - (1 - F(x))^n$$
$$\mathbb{P}(X_{(n)} < x) = (F(x))^n$$
=> эл-ты $\tilde{\mathbb{X}}$ имеют разные распределения

$\tilde{\mathbb{X}}$ - вариационный ряд. Эл-ты - порядковые статистики

$1 \le k \le n$
$$\mathbb{P}(X_{(k)} < x) = \sum_{i=k}^n C_n^i (F(x))^i (1 - F(x))^{n-i},$$
$F(x)$ - вер-ть попадания левее x (успех)
$1 - F(x)$ - вер-ть неудачи

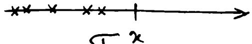
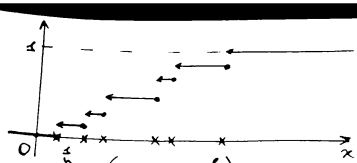


схема Бернулли

(интересно:
(исследование экстремальных порядковых стат.



(скачок $= \frac{1}{n}$)

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i < x)$$

эмпирическая (выборочная) ф-ция распределен.

$$F_n(x) \approx F(x)$$

$$nF_n(x) \sim Bi(n, F(x)) \qquad \mathbb{P}(\mathbb{1}(X_i < x) = 1) = \mathbb{P}(X_i < x) = F(x)$$

$$\mathbb{E}\, nF_n(x) = nF(x), \quad \mathbb{E}F_n(x) = F(x), \quad \mathbb{D}F_n(x) = \frac{F(x)(1-F(x))}{n}$$

, хар-ка точности приближения $F(x)$ с помощью $F_n(x)$

$F_n(x) \xrightarrow{\mathbb{P}} F(x)$ (з-н больших чисел Бернулли), $F_n(x) \xrightarrow{\text{п.н.}} F(x)$ (з-н больших чисел Колмогорова)

Т. Гливенко: $\mathbb{P}\left(\lim\limits_{n\to\infty} \sup\limits_{x} |F_n(x) - F(x)| = 0\right) = 1$

$D_n = \sup\limits_{x} |F_n(x) - F(x)|$ Распределение D_n не зависит от F

$\mathbb{P}(D_n < x) = \mathbb{P}\left(\sup\limits_{y} |F_n(y) - F(y)| < x\right) =$ Для простоты: $F(x)$ – непр. и строго монот.

$$\left\{\exists\, y = F^{-1}(z)\right\} = \mathbb{P}\left(\sup_{0\le z\le 1} |F_n(F^{-1}(z)) - F(F^{-1}(z))| < x\right) =$$

$$= \mathbb{P}\left(\sup_{0\le z\le 1} |F_n(F^{-1}(z)) - z| < x\right) = \mathbb{P}\left(\sup_{0\le z\le 1} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i < F^{-1}(z)) - z\right| < x\right) =$$

$$= \mathbb{P}\left(\sup_{0\le z\le 1} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(F(X_i) < z) - z\right| < x\right) =$$

$$= \mathbb{P}\left(\sup_{0\le z\le 1} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(U_i < z) - z\right| < x\right) \quad \left| \begin{array}{l} U_1, \dots, U_n \\ U_i \sim U[0;1] \end{array} \right|$$

$$\boxed{\begin{array}{l} \xi \sim F_\xi(x) \\ \eta = F_\xi(\xi) \quad \mathbb{P}(\eta < x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases} \\ \text{(повторить)} \end{array}}$$

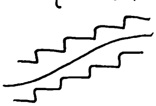
Обозначим: $\mathbb{P}(\sqrt{n}\,D_n < x) = K_n(x)$ ~~Дано $X_1, \dots,$~~

$$\lim_{n\to\infty} \mathbb{P}(\sqrt{n}\,D_n < x) = K(x) \equiv \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2x^2}$$

$\cancel{\mathbb{P}(D_n \le \varepsilon) \ge p}$.

$$\mathbb{P}(T_1(\mathbb{X}) \le F(x) \le T_2(\mathbb{X})) \ge p \qquad \{\sqrt{n}\,D_n \le y\} \equiv \{\sqrt{n}\sup |F_n(x) - F(x)| \le y\} \equiv$$

$$\equiv \left\{F_n(x) - \frac{y}{\sqrt{n}} \le F(x) \le F_n(x) + \frac{y}{\sqrt{n}} \quad \forall x \in \mathbb{R}\right\}$$



$0 < p \le 1$ – задаж. близкое к 1 $\quad K_n(y_{p,n}) = p$ / достоверность вывода

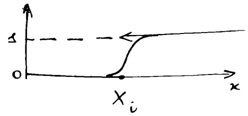
точность $\quad \frac{y_{n,p}}{\sqrt{n}} \gtrsim \varepsilon$

$$\mathbb{P}\left(F_n(x) - \frac{y_{p,n}}{\sqrt{n}} \le F(x) \le F_n(x) + \frac{y_{p,n}}{\sqrt{n}} \quad \forall x \in \mathbb{R}\right) = p$$

Приблизить $F(x)$ непрерывной ф-цией, завис. от выборки?

$\mathbb{1}(X_i<x)$ Эмпирич. ф-ция распред. – ср. арифм. ф-ций распр. $\mathbb{1}(X_i<x)$

$\Rightarrow$ заменим $\mathbb{1}(X_i<x)$ на гладкую.



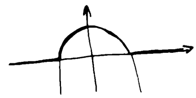
$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}\Phi}_{\text{ядро}}\Big(\frac{x-X_i}{\underbrace{h_n}_{\text{ширина окна}}}\Big) \text{ – ядерная оценка}$$

(коэф-т гладкости)

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

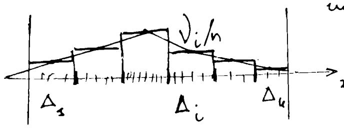
$$\left(\frac{1}{n}\sum_{i=1}^{n}\Phi\left(\frac{x-X_i}{h_n}\right)\right)'_x=\frac{1}{nh_n}\sum_{i=1}^{n}\varphi\left(\frac{x-X_i}{h_n}\right)\approx f(x)\equiv F'(x)$$ – ядерная оценка плотности

$$\varphi(x)\geq 0;\ \int_{-\infty}^{\infty}\varphi(x)\,dx=1,\quad \int_{-\infty}^{\infty}x\varphi(x)\,dx=0;\quad \int_{-\infty}^{\infty}x^2\varphi(x)\,dx=1$$ – св-ва ядра



k

гистограмма



~~Пусть~~ вместо ф.р. найдём хар-ки:

$$a=\int_{-\infty}^{\infty}x\,dF(x),\quad \sigma^2=\int_{-\infty}^{\infty}x^2\,dF(x)-a^2\qquad a\approx\overline{X}\equiv\frac{1}{n}\sum_{i=1}^{n}X_i\qquad \mathbb{E}\overline{X}=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}X_i=\frac{1}{n}\cdot na=a$$

$$\overline{X}\xrightarrow[n\to\infty]{\text{п.н.}}a\qquad \sigma^2\approx\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\equiv S^2\qquad \mathbb{E}S^2=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i-\overline{X})^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i-a+a-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i-a)^2+\frac{2}{n}\sum_{i=1}^{n}\mathbb{E}(X_i-a)(a-\overline{X})+\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(a-\overline{X})^2=(*)$$

$$\begin{cases}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i-a)^2=\sigma^2,\quad \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\overline{X}-a)^2=\frac{\sigma^2}{n},\quad \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}(X_i-a)(a-\overline{X})=\\ =2\mathbb{E}(\overline{X}-a)(a-\overline{X})=-2\mathbb{E}(\overline{X}-a)^2=-\frac{2\sigma^2}{n}\end{cases}$$

(по закону больших чисел сход. к дисперсии)

$$(*)=\sigma^2+\frac{\sigma^2}{n}-\frac{2\sigma^2}{n}=\sigma^2-\frac{\sigma^2}{n}=\sigma^2\left(\frac{n-1}{n}\right)\neq\sigma^2\ \Rightarrow \text{рассм. } \overline{S}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

$$\mu_k=\mathbb{E}X_i^k\approx\frac{1}{n}\sum_{i=1}^{n}X_i^k=m_k\qquad \frac{\sqrt{n}(m_k-\mu_k)}{\sqrt{\mathbb{D}X_i^k}}=\frac{\sum_{i=1}^{n}X_i^k-n\mu_k}{\sqrt{n(\mu_{2k}-\mu_k^2)}}\xrightarrow{d}\xi\sim\mathcal{N}(0,1)$$

$$\mathbb{D}X_i^k=\mu_{2k}-(\mu_k)^2\qquad \mu_{2k}<\infty$$

$$\mathbb{P}(m_k<x)\approx\cancel{\Phi}\,?\quad \mathbb{P}\Big(\sum_{i=1}^{n}X_i<x\Big)\approx\Phi\left(\frac{x-n\mu_k}{\sqrt{n(\mu_{2k}-\mu_k^2)}}\right)\Rightarrow\mathbb{P}(m_k<x)\approx\Phi\left(\frac{\sqrt{n}(x-\mu_k)}{\sqrt{\mu_{2k}-\mu_k^2}}\right)$$

$$\mathbb{P}(m_k<x)=\mathbb{P}\Big(\sum_{i=1}^{n}X_i<nx\Big)$$

Улюков. Параметрическая статистика

$X_1, \dots, X_n$ - выборка (н.о.р.с.в.)

Предполагаем, что семейство (~~вероятностей~~ распределений) параметризовано

$F(x, \theta), \theta \in \Theta \subset \mathbb{R}^m$ ∀ прообраз борелевского мн-ва явл. борелевским

Опр. Статистикой наз. ∀ измерим. ф-ция $T(x)$ $\forall x \in \mathbb{R}^n : \mathbb{R}^n \to \mathbb{R}^m$

$(T_1(x), T_2(x), \dots, T_m(x))$

Ф-ция правдоподобия.

∃ наблюд. с.в. - абс. непр. ($F(x,\theta)$ абс. непр.)

$$F(x,\theta) = \int_{-\infty}^{x} \underbrace{p(u,\theta)}_{\substack{\geq 0 \\ \text{пл-ть распр.}}} du$$

$L(x,\theta) = p(X_1,\theta)p(X_2,\theta)\dots p(X_n,\theta)$ - ф-ция правдоподобия (совмест. пл-ть)

Пример: ∃ $X_1, \dots, X_n$ $\mathcal{N}(\theta_1, \theta_2^2)$

$$\frac{1}{\sqrt{2\pi}\theta_2} e^{-\frac{1}{2\theta_2^2}(X_1-\theta_1)^2} \dots \frac{1}{\sqrt{2\pi}\theta_2} e^{-\frac{1}{2\theta_2^2}(X_n-\theta_1)^2} = \frac{1}{(\sqrt{2\pi}\theta_2)^n} e^{-\frac{1}{2\theta_2^2}\sum_i (X_i-\theta_1)^2}$$

∃ $X_1, \dots, X_n$ р.р. $[\theta_1, \theta_2]$

$$p(x,\theta_1,\theta_2) = \begin{cases} \frac{1}{\theta_2-\theta_1}, & x \in [\theta_1, \theta_2] \\ 0, & \text{в др. случаях} \end{cases}$$

$X_{(1)} \leq \dots \leq X_{(n)}$ - порядковые статистики

$$L(X,\theta_1,\theta_2) = \frac{1}{(\theta_2-\theta_1)^n} \underbrace{H(\theta_2 - X_{(n)}) H(X_{(1)} - \theta_1)}_{\mathbb{1}(\{X_{(n)} \leq \theta_2\} \cap \{X_{(1)} \geq \theta_1\})}$$

$$H(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$(\Omega, \mathcal{A})$ μ и ν - меры

Опр. ν абс. непр. отн. μ, если $\forall A \in \mathcal{A}$ $\mu(A) = 0 \Rightarrow \nu(A) = 0$.

Теорема Радона - Никодима

Если ν абс. непр. отн. μ, то ∃ и почти всюду единств. ф-ция X:

$$\nu(A) = \int_A X(\omega)\mu(d\omega)$$

Опр. Статистическая структура допускает ф-цию правдоподобия, если ∃ μ, опр. на σ-алгебре борелевских мн-в относительно которой абс. непр. сл. в. $X_1, \dots, X_n$ $\forall \theta$

Из Т. Р-Н: $F(x,\theta)=\int\limits_{-\infty}^{x} p(u,\theta)\mu(du)$ — если с.в. не абс.непр. → обоб. пл-ть → считающая мера →

$L(X,\theta)=p(X_1,\theta),\dots,p(X_n,\theta)$ – обобщ. пл-ть относ. меры μ.

Пусть знач. сл.в. $X_1,\dots,X_n$ сосредоточ. в $(a_1,\dots,a_n,\dots)$

Назовём μ считающей, если $\forall B\in\mathcal{B}$ $\mu(B)=\#$ точек a_i, лежащих в B

$a_i\in B,\ a_j\notin B\ i\neq j \Rightarrow \mu(B)=1 \quad \mathbb{P}(X_1\in B)=\mathbb{P}(X_1=a_i)$

$\mathbb{P}(X_1\in B)=\int\limits_B p(u,\theta)\mu(du)=p(a_i,\theta)\underset{1}{\underset{\parallel}{\mu(B)}}$

$X_1 - \Pi(\theta)$

$\mathbb{P}(X_1=k)=e^{-\theta}\frac{\theta^k}{k!}$, $p(x,\theta)=e^{-\theta}\frac{\theta^x}{x!}$, $L(x,\theta)=e^{-n\theta}\frac{\theta^{\sum x_i}}{x_1!\dots x_n!}$

$X_1=\begin{cases}1, & \theta\\ 0, & 1-\theta\end{cases}$ $p(x,\theta)=\theta^x(1-\theta)^{1-x}$ (приблизим такой ф-цией) $L(x,\theta)=\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}$

Повторение (параметрич. стат.)

$\mathbb{X}=(X_1,\dots,X_n)\in\mathbb{R}^n \quad (\mathbb{R}^n,\mathcal{B}_n,\mathcal{P})$, $\mathcal{P}=\{\mathbb{P}_\theta:\theta\in\Theta\}$, $\Theta\subseteq\mathbb{R}^m$

$\mathbb{P}_\theta(X_i<x)=F(x;\theta) \quad \hat\theta=\hat\theta(\mathbb{X})\approx\theta$

1. $\mathbb{E}_\theta\hat\theta(\mathbb{X})=\theta \quad \forall\theta\in\Theta$ — несмещённая оценка

2. $\hat\theta(\mathbb{X})=\hat\theta_n(\mathbb{X})\xrightarrow[n\to\infty]{\mathbb{P}_\theta}\theta\ \forall\theta\in\Theta$ (если сх-ть п.н. по $\mathbb{P}_\theta$, то сильно сост. оценка) — состоятельная оценка

Опр. Измеримая ф-ция выборки $T(\mathbb{X})$, не завис. от θ наз-ся статистикой.

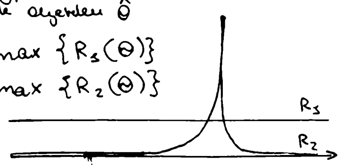
Опр. Статистика со знач. в параметрич. мн-ве ($T(\mathbb{X})\in\Theta$) наз-ся оценкой.

$R(\theta)=\mathbb{E}_\theta(\hat\theta-\theta)^2 \underset{\text{если несмещ.}}{=} \mathbb{D}_\theta\hat\theta(\mathbb{X})$

квадратичный риск оценки $\hat\theta$

$\max\{R_1(\theta)\}$

$\max\{R_2(\theta)\}$



$X_1,\dots,X_n: X_i\sim U[0,\theta]$

$\mathbb{E}_\theta X_1=\frac{\theta}{2}$

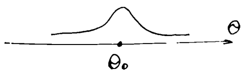
$\theta\approx 2\mathbb{E}_\theta X_1=\frac{2}{n}\sum\limits_{i=1}^n X_i=\hat\theta_1(\mathbb{X})$

$\theta\approx\hat\theta_2(\mathbb{X})=X_{(n)}\cdot\frac{n+1}{n}$

Д/з: убедиться $\mathbb{D}_\theta\hat\theta_2(\mathbb{X})=\underset{n\to\infty}{O}\left(\frac{1}{n^2}\right)$

$\varphi(\theta) \geq 0$ $\quad \int\limits_{\Theta} \varphi(\theta)\,d\theta = 1$; $B_1 = \int\limits_{\Theta} R_1(\theta)\varphi(\theta)\,d\theta$, $B_2 = \int\limits_{\Theta} R_2(\theta)\varphi(\theta)\,d\theta$

Если $B_1 < B_2$, то R_1 лучше. (Байесовский критерий)



R_1, R_2 хорошо различимы в окрест. θ_0

арг. параметр

Если распр. $X_1,\dots,X_n$ дискретно, то $L(\theta;\vec{x}) = \prod\limits_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i) = \mathbb{P}_\theta(X = \vec{x})$

абс. непр.: $L(\theta;\vec{x}) = \prod\limits_{i=1}^{n} p(x_i;\theta)$, $p(x,\theta) = \frac{d}{dx}F(x,\theta)$

вер-ть наблюдать то, что мы набл. как ф-цию от ~~параметра~~ θ.

$U(X;\theta) = \frac{\partial}{\partial\theta}\ln L(\theta,X) = \frac{1}{L(\theta;X)}\cdot\frac{\partial}{\partial\theta}L(\theta;X)$

($L(\theta;X) \neq 0$ и дост. гладкая)

ф-ция вклада

$$E_\theta U(X;\theta) = \int\limits_{\mathbb{R}^n} U(\vec{x};\theta)L(\theta;\vec{x})\,d\vec{x} = \int\limits_{\mathbb{R}^n} \frac{1}{\cancel{L(\theta;\vec{x})}}\frac{\partial}{\partial\theta}L(\theta;\vec{x})\cancel{L(\theta;\vec{x})}\,d\vec{x} =$$

$$= \frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}^n} L(\theta;\vec{x})\,d\vec{x} = \frac{\partial}{\partial\theta}1 = 0$$

условия регулярности

$$E_\theta U^2(X;\theta) = D_\theta U(X;\theta) \equiv i_n(\theta)$$

Опр. Дисперсия ф-ции вклада наз-ся Фишеровской инф-ей, содержащейся в выборке X.

$$i_n(\theta) = D_\theta U(X;\theta) = D_\theta \frac{\partial}{\partial\theta}\ln L(\theta;X) = D_\theta \frac{\partial}{\partial\theta}\sum_{i=1}^{n}\ln p(X_i;\theta) =$$

$$= D_\theta \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln p(X_i;\theta) = \sum_{i=1}^{n} D_\theta \frac{\partial}{\partial\theta}\ln p(X_i;\theta) = \sum_{i=1}^{n} D_\theta U(X_i;\theta) = n D_\theta U(X_1;\theta)$$

$$\equiv n\cdot i(\theta) \Rightarrow \underline{i_n(\theta) = n\cdot i(\theta)}$$

$$E_\theta T(X) \equiv \tau(\theta), \quad \mathcal{T}_\tau = \{T(X) : E_\theta T(X) = \tau(\theta)\}$$

Теорема. (предп., что вып. усл. регулярности) $\forall T \in \mathcal{T}_\tau$ $D_\theta T(X) \geq \frac{(\tau'(\theta))^2}{n\cdot i(\theta)}$

(чем больше инф., тем меньше дисперсия)

нер-во Рао Крамера

Док-во: $E_\theta T(\mathbb{X}) = \tau(\theta) = \int T(\vec{x}) L(\theta; \vec{x}) d\vec{x}$

$$\tau'(\theta) = \int T(\vec{x}) \frac{\partial}{\partial\theta} \ln L(\theta; \vec{x}) \cdot L(\theta; \vec{x})\, d\vec{x} = E_\theta [T(\mathbb{X}) U(\mathbb{X}; \theta)] =$$

$$= \mathrm{cov}_\theta (T(\mathbb{X}), U(\mathbb{X}, \theta))$$

$$(\tau'(\theta))^2 \le D_\theta T(\mathbb{X}) \cdot \underbrace{\mathbb{D} U(\mathbb{X}; \theta)}_{n i(\theta)} \quad \blacksquare$$

Зам. Обращ. в рав-во, когда $T(\mathbb{X}) = a(\theta) + b(\theta) U(\mathbb{X}; \theta)$

Опр. Оценка $T(\mathbb{X})$, обращ. нер-во Рао-Крамера в рав-во, наз-ся эффективной.

Условное математическое ожидание

Определение для дискретных: $\xi: x_1, x_2$; $\eta: y_1, y_2$

$$\mathbb{P}(\xi = x_i | \eta = y_j) = \frac{\mathbb{P}(\xi = x_i, \eta = y_j)}{\mathbb{P}(\eta = y_j)}; \ \mathbb{E}(\xi | \eta = y_j) = \sum_i x_i \mathbb{P}(\xi = x_i | \eta = y_j) = f(y_j)$$

Опр. Условным мат. ожид. дискр. вел. ξ отн. дискр. вел η наз-ся случ. вел. $\mathbb{E}(\xi|\eta) = f(\eta)$.

$$\forall B \in \mathcal{B} \quad \mathbb{E}[\mathbb{E}(\xi|\eta) \mathbb{1}(\eta \in B)] = \sum_{j: y_j \in B} f(y_j) \mathbb{P}(\eta = y_j) = \sum_{j: y_j \in B} \left[\sum_k x_k \frac{\mathbb{P}(\xi = x_k, \eta = y_j)}{\mathbb{P}(\eta = y_j)} \right] \mathbb{P}(\eta = y_j)$$

$$= \sum_{j: y_j \in B} \sum_k x_k \mathbb{P}(\xi = x_k, \eta = y_j) = \mathbb{E}\, \xi \mathbb{1}(\eta \in B) \Rightarrow \mathbb{E}[\mathbb{E}(\xi|\eta) \mathbb{1}(\eta \in B)] = \mathbb{E}\, \xi \mathbb{1}(\eta \in B)$$

Для абс. непр: $\xi, \eta \sim p_{\xi,\eta}(x, y)$

$$\forall B \in \mathcal{B}_2 \quad \mathbb{P}((\xi, \eta) \in B) = \iint_B p_{\xi,\eta}(x, y)\, dx\, dy \ ; \ p_\eta(y) = \int_{-\infty}^{\infty} p_{\xi,\eta}(x, y)\, dx$$

$$\mathbb{P}(\xi < x | \eta = y) = \frac{\mathbb{P}(\xi < x, \eta = y)}{\mathbb{P}(\eta = y)} = \frac{0}{0} \ ??$$

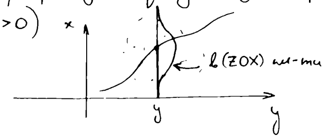
предп., что совместн. пл-ть удовл. усл. Т о сред.

$$\varepsilon > 0 \quad \mathbb{P}(\xi < x | y \le \eta < y + \varepsilon) = \frac{\mathbb{P}(\xi < x; y \le \eta < y + \varepsilon)}{\mathbb{P}(y \le \eta < y + \varepsilon)} = \frac{\int_{-\infty}^{x} \int_{y}^{y+\varepsilon} p_{\xi,\eta}(x, y)\, dx\, dy}{\int_{y}^{y+\varepsilon} p_\eta(y)\, dy} =$$

$$= \frac{\varepsilon \int_{-\infty}^{x} p_{\xi,\eta}(x, y')\, dx}{\varepsilon \cdot p_\eta(y'')} = \int_{-\infty}^{x} \frac{p_{\xi,\eta}(x, y')}{p_\eta(y'')}\, dx \xrightarrow[\varepsilon \to 0]{} \int_{-\infty}^{x} \frac{p_{\xi,\eta}(x, y)}{p_\eta(y)}\, dx$$

Опр. Условной плотностью случ. вел. ξ при усл. $\eta = y$ наз-ся ф-ция:

$p_{\xi|\eta=y}(x) = \dfrac{p_{\xi,\eta}(x,y)}{p_\eta(y)}$ (для опред-ти $p_\eta > 0$)



$$\mathbb{E}(\xi|\eta=y) = \int_{-\infty}^{\infty} x\, p_{\xi|\eta=y}(x)\,dx = f(y)$$

Опр. Условным мат. ожиданием сл. в. ξ отн. η наз-ся: $\mathbb{E}(\xi|\eta) = f(\eta)$

$$\mathbb{P}(\mathbb{E}(\xi|\eta) \in B) = \mathbb{P}(f(\eta) \in B) = \mathbb{P}(\eta \in f^{-1}(B))$$

$$\mathbb{E}[\mathbb{E}(\xi|\eta)\mathbb{1}(\eta\in B)] = \mathbb{E} f(\eta)\mathbb{1}(\eta \in B) = \int_B f(y)p_\eta(y)dy = \int_B\int_{-\infty}^{\infty} \frac{x p_{\xi,\eta}(x,y)}{p_\eta(y)}dx \cdot p_\eta(y)dy =$$

$$= \int_B\int_{-\infty}^{\infty} x p_{\xi,\eta}(x,y)dx\,dy = \mathbb{E}\,\xi\,\mathbb{1}(\eta\in B)$$

$(\Omega, \mathcal{A}, \mathbb{P})$, $\mathcal{F} \subseteq \mathcal{A}$ $\quad$ $(\{\omega : \eta(\omega) \in B\} \in \mathcal{F})$

Опр. Случ. вел. η наз. $\mathcal{F}$-измеримой, если $\forall B \in \mathfrak{B}$ $\eta^{-1}(B) \in \mathcal{F}$
(меньше вариантов поведения, чем $\mathcal{A}$)

Опр. У.М.О. с.в. ξ от σ-алг. $\mathcal{F}$ наз. с.в. $\mathbb{E}(\xi|\mathcal{F})$:

1) $\mathbb{E}(\xi|\mathcal{F})$ — $\mathcal{F}$-измерима $\quad$ (У.М.О. измеримо отн. $\mathcal{F}$)
2) $\forall A \in \mathcal{F}$ $\mathbb{E}[\mathbb{E}(\xi|\mathcal{F})\mathbb{1}(A)] = \mathbb{E}\,\xi\,\mathbb{1}(A)$

Опр. σ-алг. порождена с.в. η, если она состоит из мн-в $\eta^{-1}(B), \forall B \in \mathfrak{B}$ и обозн. $\sigma(\eta)$

Опр. $\mathbb{E}(\xi|\eta) = \mathbb{E}(\xi|\sigma(\eta))$

$\mathbb{E}\,\xi\,\mathbb{1}(A) = \nu(A) \ll \mathbb{P}(A)$ $\quad$ $\exists$ ф-ция ≥ 0 т....

мера $\quad$ $\mathbb{P}(A) = 0 \Rightarrow \nu(A) = 0$ $\quad$ Т.Р-Н.

$\nu(A)$ абс. непр. отн. $\mathbb{P}(A)$

Св-ва У.М.О.: (проверить для дискр. случая)

1) $\forall B \in \mathcal{B}$ $\mathbb{P}(\xi \in B | \mathcal{F}) = \mathbb{E}(\mathbb{1}(\xi \in B) | \mathcal{F})$

2) $\mathbb{E}(a\xi_1 + b\xi_2 | \mathcal{F}) \overset{п.н.}{=} a\,\mathbb{E}(\xi_1 | \mathcal{F}) + b\,\mathbb{E}(\xi_2 | \mathcal{F})$

3) η - $\mathcal{F}$ измерима. $\mathbb{E}(\xi \cdot \eta | \mathcal{F}) \overset{п.н.}{=} \eta\, \mathbb{E}(\xi | \mathcal{F})$

4) ξ, η - независ. $\mathbb{E}(\xi | \eta) \overset{п.н.}{=} \mathbb{E}\xi$

5) $\mathbb{E}(\mathbb{E}(\xi | \mathcal{F})) = \mathbb{E}\xi$

Матстат.

$(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P})$, $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$, $\mathbb{X} = (X_1, \dots, X_n)$, $T(\mathbb{X})$

$X_i \sim U[0;\theta]$

$T_1(\mathbb{X}) = \frac{2}{n}\sum_{i=1}^n X_i = 2\overline{X}$, $\mathbb{D}T_1(\mathbb{X}) \sim \frac{1}{n}$ $(n \to \infty)$

$T_2(\mathbb{X}) = \frac{n+1}{n} X_{(n)} = \frac{n+1}{n} \max_{1 \le i \le n} X_i$, $\mathbb{D}T_2(\mathbb{X}) \sim \frac{1}{n^2}$ $(n \to \infty)$

$\mathbb{E}_\theta \frac{n+1}{n} X_{(n)} = (*)$

$\mathbb{P}_\theta(X_{(n)} < x) = \mathbb{P}_\theta(X_1 < x, \dots, X_n < x) = \prod_{i=1}^n \mathbb{P}_\theta(X_i < x) = \frac{x^n}{\theta^n}$

$p_{X_{(n)}}(x) = \frac{n x^{n-1}}{\theta^n}$

$(*) = \frac{n+1}{n}\int_0^\theta \frac{x \cdot n \cdot x^{n-1}}{\theta^n} dx = \theta$, $\mathbb{D}_\theta T_2(\mathbb{X}) = \mathbb{E}_\theta T_2^2(\mathbb{X}) - \theta^2$

$\mathbb{E}_\theta T_2^2(\mathbb{X}) = \mathbb{E}_\theta \left(\frac{n+1}{n}\right)^2 X_{(n)}^2 = \frac{(n+1)^2}{n^2}\int_0^\theta \frac{x^2 n x^{n-1}}{\theta^n} dx = \frac{(n+1)^2}{n\theta^n} \cdot \frac{\theta^{n+2}}{n+2} = \theta^2 \cdot \frac{(n+1)^2}{n(n+2)}$

$\mathbb{D}_\theta T_2(\mathbb{X}) = \theta^2\left(\frac{(n+1)^2}{n^2+2n} - 1\right) = \theta^2\left(\frac{1}{n^2+2n}\right) = \frac{\theta^2}{n^2+2n} \sim \frac{1}{n^2}$ $(n \to \infty)$

Достаточные статистики. Критерий факторизации

Оп. $T(\mathbb{X})$ - достаточная, если $\mathbb{E}_\theta(\mathbb{X} | T(\mathbb{X}))$ не зависит от θ.
($\forall B \in \mathcal{B}_n$ $\mathbb{P}_\theta(\mathbb{X} \in B | T(\mathbb{X}))$ не зав. от θ)

Критерий факторизации: $T(\mathbb{X})$ дост. (для θ), если и только если $L(\theta; \mathbb{X}) = g(T(\mathbb{X}), \theta) \cdot h(\mathbb{X})$. Рассм. дискр. случай:

1) $\exists T(\mathbb{X})$ - достат. $\Rightarrow \forall t$ $\mathbb{P}_\theta(\mathbb{X} = x | T(\mathbb{X}) = t) = h(x, t)$

$\vec{x}, t : T(\vec{x}) = t \Rightarrow L(\theta; \vec{x}) = \mathbb{P}_\theta(\mathbb{X} = \vec{x}) = \mathbb{P}_\theta(\mathbb{X} = \vec{x}, T(\vec{x}) = t) = \mathbb{P}_\theta(T(\vec{x}) = t) \cdot$
$\mathbb{P}_\theta(\mathbb{X} = \vec{x} | T(\vec{x}) = t) = g(t, \theta) h(\vec{x})$

2) $\exists L(\theta;\vec{x}) = g(T(\vec{x}),\theta)\cdot h(\vec{x})$

$\vec{x}: T(\vec{x}) = t$ (фикс.) $\Rightarrow \mathbb{P}_\theta(\mathbb{X}=\vec{x} \mid T(\mathbb{X})=t) = \dfrac{\mathbb{P}_\theta(\mathbb{X}=\vec{x}, T(\mathbb{X})=t)}{\mathbb{P}_\theta(T(\mathbb{X})=t)} = \dfrac{\mathbb{P}_\theta(\mathbb{X}=\vec{x})}{\mathbb{P}_\theta(T(\mathbb{X})=t)} =$

$$= \frac{L(\theta;\vec{x})}{\sum\limits_{\vec{x}:T(\vec{x})=t} L(\theta;\vec{x})} = \frac{g(T(\vec{x}),\theta)h(\vec{x})}{\sum\limits_{\vec{x}:T(\vec{x})=t} g(T(\vec{x}),\theta)h(\vec{x})} = \frac{g(T(\vec{x}),\theta)h(\vec{x})}{\sum\limits_{\vec{x}:T(\vec{x})=t} g(t,\theta)h(\vec{x})} = \frac{h(\vec{x})}{\sum\limits_{\vec{x}:T(\vec{x})=t} h(\vec{x})}$$ – не зав. от θ

Пример 1) $X_i \sim Poiss(\theta)$: $\mathbb{P}(X_i=k) = e^{-\theta}\cdot\dfrac{\theta^k}{k!}$, $k=0,1,2,\dots$

$$L(\theta,\mathbb{X}) = \prod_{i=1}^{n} e^{-\theta}\cdot\frac{\theta^{x_i}}{x_i!} = \underbrace{e^{-n\theta}\cdot\theta^{\sum_{i=1}^n x_i}}_{g}\cdot\underbrace{\prod_{i=1}^{n}\frac{1}{x_i!}}_{h} \Rightarrow T(\mathbb{X}) = \sum_{i=1}^{n} X_i$$ – достат.

2) $X_i \sim 1-e^{-\theta x}$, $x \geq 0$

$$p(x,\theta) = \theta e^{-\theta x},\quad L(\theta,x) = \underbrace{\theta^n e^{-\theta\sum_{i=1}^n X_i}}_{g}\cdot\underbrace{1}_{h} \Rightarrow T(\mathbb{X}) = \sum_{i=1}^{n} X_i$$ – достат.

3) $X_i \sim U[0;\theta]$

$$p(x,\theta) = \frac{1}{\theta}\mathbb{1}(0\leq x\leq\theta),\quad L(\theta,x) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbb{1}(0\leq X_i\leq\theta) = \frac{1}{\theta^n}\mathbb{1}(\theta\geq X_{(n)})\mathbb{1}(X_{(1)}\geq 0)$$

Задачи: 1) $X_i \sim U[-\theta,\theta]$, 2) $X_i \sim U[\theta,\theta+1]$ $\Rightarrow T(\mathbb{X}) = X_{(n)}$ – достат.

3) $X_i \sim U[\theta, 2\theta]$.

4) $X_i \sim p(x,\theta) = \dfrac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\dfrac{(x-a)^2}{2\sigma^2}\right\}$ $\quad\theta = (a,\sigma^2)$

$$L(a,\sigma^2;x) = \frac{1}{(\sqrt{2\pi})^n\sigma^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-a)^2\right\} = (2\pi)^{-\frac{n}{2}}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\left[na^2+\sum_{i=1}^{n}X_i^2-2a\sum_{i=1}^{n}X_i\right]\right\}$$

$$T(\mathbb{X}) = \left(\sum_{i=1}^{n}X_i^2, \sum_{i=1}^{n}X_i\right) \div (\overline{X}, S^2)$$

Доверительные интервалы

Точеч. оценивание: $X_1,\dots,X_n$ - выборка н.о.р.с.в. $X_i \sim F(x,\theta)$ $\theta\in\Theta$

$T(\mathbb{X})$ - стат. (оценка) при $\forall x$-реализации X $t = T(x)$, t - приближ. знач. $\tau(\theta)$

Интервальное оценивание: Δ - интервал подбирается т.о.: $\mathbb{P}_\theta(\theta\in\Delta)\geq\gamma$, $\gamma\in(0,1)$

γ - уровень доверия (Δ_γ - γ - дов. интервал)

Δ: $\underbrace{T_1=T_1(\mathbb{X})}_{\text{нижн. дов. граница}}$, $\underbrace{T_2=T_2(\mathbb{X})}_{\text{верх. дов. граница}}$, $T_1<T_2$ $\quad \Delta=(T_1,T_2)$ $\quad \mathbb{P}_\theta(T_1<\theta<T_2)\geq\gamma$ $\forall\theta\in\Theta$

$l=\mathbb{E}(T_2-T_1)$ - длина дов. интервала

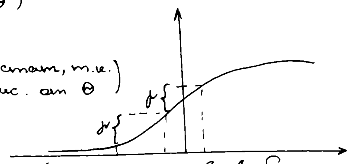
(Односторонние дов. области $\theta<T_2$ или $T_1<\theta$)

Метод центральной статистики

$\exists\, G(\mathbb{X},\theta)$ - центральная статистика (не стат, т.к. завис. от θ)

а) $G(X,\theta)$ - непр. и монотон. ф-ция θ

б) $F_G(t)$ - непр. и не зависит от θ



• (много вариантов выбора g_1, g_2)

$\mathbb{P}(g_1 < G(\mathbb{X},\theta) < g_2) = \gamma$

не зав. от θ по п.б)

$G(\mathbb{X},\theta)=g_1$, T_1 - реш. отн. θ

$G(\mathbb{X},\theta)=g_2$, T_2 - реш. отн. θ

$\Rightarrow \quad g_1<G(\mathbb{X},\theta)<g_2 \Leftrightarrow T_1<\theta<T_2$

$\Rightarrow \mathbb{P}(T_1<\theta<T_2)=\gamma \Rightarrow (T_1,T_2)=\Delta_\gamma$

<u>Пример</u>: $X_1,\dots,X_n\sim N(\theta,1)$

$G(X,\theta)=(\overline{X}-\theta)\sqrt{n}$, $\overline{X}=\frac{\sum X_i}{n}$

$G(X,\theta)\sim N(0,1)$

$\mathbb{P}(g_1<G(X,\theta)<g_2)=\gamma \Rightarrow \Phi(g_2)-\Phi(g_1)=\gamma$, $\Phi(x)=\int\limits_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$

$\mathbb{P}(g_1<\sqrt{n}(\overline{X}-\theta)<g_2)=\mathbb{P}\left(\overline{X}-\frac{g_2}{\sqrt{n}}<\theta<\overline{X}-\frac{g_1}{\sqrt{n}}\right)$

$\Delta_\gamma=\left(\overline{X}-\frac{g_2}{\sqrt{n}}, \overline{X}-\frac{g_1}{\sqrt{n}}\right)$, $\Phi(g_2)-\Phi(g_1)=\gamma$

$l=\frac{g_2-g_1}{\sqrt{n}} \Rightarrow g_2-g_1\to\min$ и $\Phi(g_2)-\Phi(g_1)=\gamma$

$L(g_1,g_2,\lambda)=g_2-g_1+\lambda(\Phi(g_2)-\Phi(g_1)-\gamma)$ $\quad \begin{cases}-1-\lambda\varphi(g_1)=0\\ 1+\lambda\varphi(g_2)=0\\ \Phi(g_2)-\Phi(g_1)=\gamma\end{cases}$

$\varphi(g_1) = \varphi(g_2) \Rightarrow g_1 = -g_2$

табл.

$\Phi(g_2) - \Phi(-g_2) = p \Rightarrow 2\Phi(g_2) - 1 = p \Rightarrow \Phi(g_2) = \frac{p+1}{2} \Rightarrow g_2 = C_p$

$\Rightarrow \Delta_p^* = \left(\overline{X} - \frac{C_p}{\sqrt{n}}, \overline{X} + \frac{C_p}{\sqrt{n}}\right)$

$l = \frac{2C_p}{\sqrt{n}}$.

$X_1, \dots, X_n \sim F(x, \theta)$ монот. и непр. по θ (монот. нужна, чтобы обе-ть именно вид интервала)

$G(X,\theta) = -\sum_{i=1}^{n} \ln F(X_i, \theta)$

$F(X_i, \theta) \sim R[0,1]$, $-\ln F(X_i,\theta) \sim P(1) \sim \Gamma(1,1) \Rightarrow G(X,\theta) \sim \Gamma(1, n)$

$G(X,\theta) = g_i \leftarrow$ сложно.

Метод построения доверительного интервала на основе точечной оценки

$\exists\, T(X)$ – оценка θ

$F_T(t, \theta)$ непр. и монотонна по θ.

$\mathbb{P}_\theta(t_1 < T(X) < t_2) = F_T(t_2, \theta) - F_T(t_1, \theta) = p$, $t_i = t_i(\theta)$

$\mathbb{P}_\theta(T(X) \le t_1) = \frac{\alpha}{2} = \frac{1-p}{2} \Rightarrow F_T(t_1, \theta) = \frac{1-p}{2}$ (t_1, t_2) – центр. дов. интервал

$\mathbb{P}_\theta(T(X) \ge t_2) = \frac{\alpha}{2} = \frac{1-p}{2} \Rightarrow 1 - F_T(t_2, \theta) = \frac{1-p}{2}$

$T(X) = t_1$

$T(X) = t_2$

подст. $T(X)$ в t_1, t_2 и реш. отн. θ

Асимптотический доверительный интервал

Ц.П.Т.: $\frac{\sum_{i=1}^{n} X_i - na(\theta)}{\sqrt{n\sigma^2(\theta)}} \Rightarrow \mathcal{N}(0,1)$, где $a(\theta) = \mathbb{E}_\theta X_i$, $\sigma^2(\theta) = \mathbb{D}_\theta X_i$

$G'(X,\theta) = \frac{\sum X_i - na(\theta)}{\sqrt{n\sigma^2(\theta)}}$ (нет гарантии монотонности)

$\mathbb{P}(g_1 < G'(X,\theta) < g_2) \approx \Phi(g_2) - \Phi(g_1) = p$

погрешность

$X_i \sim \begin{cases} 1, & \theta \\ 0, & 1-\theta \end{cases}$ $\sum_{i=1}^{n} X_i \sim Bi(n, \theta)$, $G'(X,\theta) = \frac{\sum_{i=1}^{n} X_i - n\theta}{\sqrt{n\theta(1-\theta)}} = \frac{\sqrt{n}(\overline{X} - \theta)}{\sqrt{\theta(1-\theta)}}$

$g_1 < \frac{\sqrt{n}(\overline{X} - \theta)}{\sqrt{\theta(1-\theta)}} < g_2 \Rightarrow \overline{X} - \frac{g_2\sqrt{\theta(1-\theta)}}{\sqrt{n}} < \theta < \overline{X} - \frac{g_1\sqrt{\theta(1-\theta)}}{\sqrt{n}}$, $\theta(1-\theta) \le \frac{1}{4}$

(чтобы изб-ся от θ в концах – расширим инт)

$\sigma'(X,\theta) = \frac{\sum X - na(\theta)}{\sqrt{n\sigma^2(\theta)}}$

$\tilde{\sigma}(\theta) \to \tilde{\sigma}^2$ - сост. оценка θ $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

Проверка статистических гипотез

$(\mathbb{R}^n, \mathcal{B}_n, \mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\})$

Опр. Статистической гипотезой наз-ся $\forall$ подмножество мн-ва $\mathcal{P}$ ($\mathcal{P}_0 \subseteq \mathcal{P}$)

$\mathcal{P}_0 \subseteq \mathcal{P} \iff \Theta_0 \subseteq \Theta$, $\mathcal{P}_0 = \{\mathbb{P}_\theta : \theta \in \Theta_0\}$ Фактически нужно опр. наход. ли θ в Θ_0 или нет?

$\mathcal{N}(a, \sigma^2)$

$H_0 : a > 0$

$H_0 : a = a_0$

- Если Θ_0 сост. из одного эл-та, то гипотеза простая, иначе - сложная

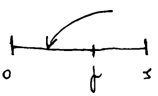
Опр. Критерий - измерим. ф-ция выборки (стат-ка) со значениями в отрезке $[0;1]$

$\varphi(\mathbb{X}) = p \in [0;1]$ - вер-ть того, что при заданном X гипотеза должна быть отвергнута.

- Если $\varphi(\mathbb{X}) = \begin{cases} 0 \\ 1 \end{cases}$, то крит. наз-ся нерандомизированным
- Если есть знач. из $(0,1)$, то рандомизированным

Нерандомиз.: $\mathbb{R}^n = \underset{\uparrow}{K} \cup \overline{K}$, $\mathbb{X} \in K \Rightarrow \varphi(\mathbb{X}) = 1$, $\mathbb{X} \in \overline{K} \Rightarrow \varphi(\mathbb{X}) = 0$

критическое мн-во ; $\varphi(\mathbb{X}) = \mathbb{1}(\mathbb{X} \in K)$



Можно постр. φ так, чтобы вер-ть (α) принять неверную гипотезу была крайне мала.

Критерий согласия

выборка $\Rightarrow$ гистограмма $\Rightarrow$ похоже на норм. распред. $\Rightarrow$ нужно проверить согласие с норм. распр.

1. Крит. согласия Колмогорова

$H_0 : F = F_0$

$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i < x) \approx F(x)$

$D_n = \sup_x |F_n(x) - F_0(x)|$, $K_n(y) = \mathbb{P}_0(\sqrt{n} D_n < y) \xrightarrow[n\to\infty]{} K(y)$

$D_n \ge y_0 \Rightarrow \overline{H_0}$

$D_n < y_0 \Rightarrow H_0$

$P_0(D_n \ge y_0) = \alpha$, вер-сть невозмож. события

вер-ть отвергнуть верную гипотезу

$K_n(y) = P_0(\sqrt{n} D_n < y) = P_0\left(D_n < \frac{y}{\sqrt{n}}\right) = 1 - \alpha = K_n(k_n(1-\alpha))$

$\Rightarrow y_0 = \frac{k_n(1-\alpha)}{\sqrt{n}}$

$k_n(1-\alpha) : K_n(k_n(1-\alpha)) = 1-\alpha$

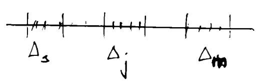
$K = \left\{ \vec{x} : D_n(\vec{x}) \ge \frac{k_n(1-\alpha)}{\sqrt{n}} \right\}$

Можно пользоваться только тогда, когда гипотеза простая (нельзя заменить F_0 на оценку $F(x, \hat{\theta})$, т.к. неизв. распр. D_n.

В $D_n = \sup_x |F_n(x) - F_0(x)|$ ← только случайность здесь (зав-ть от выборки))

Можно провести два исследования: $(\underbrace{X_1, \dots}_{\theta_0}, \underbrace{X_n}_{H_0})$ 1) находим θ_0 2) проверяем гипотезу

2. Критерий χ^2



$\nu_j = \sum_{i=1}^{n} \mathbb{1}(X_i \in \Delta_j)$, $\nu_1 + \dots + \nu_m = n$

$p_j = P(X_i \in \Delta_j)$

1. H_0 - простая

$p_j = P_0(X_i \in \Delta_j)$ $\quad \nu_j \sim Bi(n, p_j)$, $E_0 \nu_j = n p_j$

$X^2 = \sum_{j=1}^{m} \frac{(\nu_j - n p_j)^2}{n p_j}$

Т. верна только когда ~~крайние~~ разбиение ~~границы~~ не зависит от выборки

Если X^2 очень мала, то "что-то не так" с независимостью выборки

$X^2 \ge y_0 \Rightarrow \overline{H_0}$

$X^2 < y_0 \Rightarrow H_0$

Пусть $\xi_1, \dots, \xi_m$ - н.о.р.с.в., $\xi_i \sim \mathcal{N}(0,1)$

$\eta_m = \xi_1^2 + \dots + \xi_m^2$, $\eta_m \sim \chi^2_m$, степени свободы

$p_{\eta_m}(x) = \frac{1}{2^{m/2}\Gamma\left(\frac{m}{2}\right)} x^{m/2 - 1} \cdot e^{-x/2}$, $x \ge 0$

Т (Пирсона). H_0 верна $\Rightarrow P_0(X^2 < y) \xrightarrow[n \to \infty]{} \chi^2_{m-1}(y)$

$\chi^2_{m-1}(\varkappa_{m-1}(\alpha)) = \alpha$, $y_0 = \varkappa_{m-1}(1-\alpha)$

Отличие от 1-го крит. только в асимпт.; погрешность неизвестна

На закаде: $X^2 = \text{æ}$, $\mathbb{P}_0(X^2 > \text{æ}) = 1 - p$ (большое p - хорошо) 0,2 - 0,3 уже достаточно
наблюдали выборку новая выборка p-value

Предположим, что $X_i \sim F(x, \theta)$, $\theta = (\theta_1, \dots, \theta_r)$

$X^2 = \sum_{j=1}^{m} \frac{(\nu_j - np_j)^2}{np_j}$, $p_j = p_j(\hat{\theta})$, $\hat{\theta}^*$ - ОМП (полином.) В Т. Пирсона: χ^2_{m-1-r}
Т. Фишера для сложных гипотез

$(\nu_1, \dots, \nu_k)$ Обозначения $\Delta_1 \dots \Delta_k$ (было m)

$\mathbb{P}(\nu_1 = n_1, \dots, \nu_k = n_k) = \frac{n!}{n_1! \cdot \dots \cdot n_k!} p_1^{n_1} \dots p_k^{n_k}$

$n_1 + \dots + n_k = n$

$p_i = \mathbb{P}(X_1 \in \Delta_1)$ $p_i = p_i(\theta)$ $L^*(\theta; \nu_1, \dots, \nu_k) = \frac{n!}{\nu_1! \cdot \dots \cdot \nu_k!} \prod_{i=1}^{k} (p_i(\theta))^{\nu_i}$

$p_1 + \dots + p_k = 1$

Т. Фишера верна только для оценок макс. правд. полином. ф-ции: (если очень хочется, то можно и другие)

$\hat{\theta}^* = \underset{\theta}{\operatorname{argmax}}\, L^*(\theta; \nu_1, \dots, \nu_k)$

$\mathbb{X} = (X_1, \dots, X_n)$, $X_i \sim p(x)$ $\mathcal{P} = \{\mathbb{P}_0, \mathbb{P}_1\}$, строго говоря, нужно ещё ввести меру

$H_0: p = p_0$ - выбираем главной гипотезой (например: минимизируем ущерб при ошибке)

$H_1: p = p_1$

($\varphi(\mathbb{X})$ - вер-ть не принять нулевую гипотезу)

$\mathbb{E}_0 \varphi(\mathbb{X})$ - вер. ошибки I го рода

$\mathbb{E}_1(1 - \varphi(\mathbb{X}))$ - вер. ошибки II го рода (гипотеза верна, но мы её отвергаем)

$Y = \begin{cases} 0 \\ 1 \end{cases}$ — номер гип., кот. верна на самом деле (неизв.) $Z = \begin{cases} 0 \\ 1 \end{cases}$ — наш „диагноз" (изв.)

c_{ij}, если $Y = j$, $Z = i$ — потери (ущерб)

$c_{01} > c_{11}$
$c_{10} > c_{00}$

$\omega = \mathbb{P}(Y = 0)$

Задача: построить критерий $\varphi(\mathbb{X})$: $\mathbb{E} C_{ZY} \to \min$

$$\mathbb{E} C_{ZY} = c_{00} \mathbb{P}(Z=0, Y=0) + c_{01} \mathbb{P}(Z=0, Y=1) + c_{10} \mathbb{P}(Z=1, Y=0) + c_{11} \mathbb{P}(Z=1, Y=1) =$$

$$= c_{00} \mathbb{P}(Z=0|Y=0) \overbrace{\mathbb{P}(Y=0)}^{\omega} + c_{01} \mathbb{P}(Z=0|Y=1) \overbrace{\mathbb{P}(Y=1)}^{1-\omega} + c_{10} \mathbb{P}(Z=1|Y=0) \overbrace{\mathbb{P}(Y=0)}^{\omega} +$$

$$+ c_{11} \mathbb{P}(Z=1|Y=1) \underbrace{\mathbb{P}(Y=1)}_{1-\omega} = c_{00} \omega \mathbb{E}_0(1 - \varphi(\mathbb{X})) + c_{01}(1-\omega) \mathbb{E}_1(1 - \varphi(\mathbb{X})) +$$

$$+ c_{10} \omega \mathbb{E}_0 \varphi(\mathbb{X}) + c_{11}(1-\omega) \mathbb{E}_1 \varphi(\mathbb{X}) = c_{00} \omega \int_{\mathbb{R}^n} (1 - \varphi(\vec{x})) p_0(\vec{x}) d\vec{x} +$$

совместная пл-ть!

$$+ c_{01}(1-\omega) \int_{\mathbb{R}^n} (1 - \varphi(\vec{x})) p_1(\vec{x}) d\vec{x} + c_{10} \omega \int_{\mathbb{R}^n} \varphi(\vec{x}) p_0(\vec{x}) d\vec{x} + c_{11}(1-\omega) \int_{\mathbb{R}^n} \varphi(\vec{x}) p_1(\vec{x}) d\vec{x} =$$

$$= c_{00}\omega + c_{01}(1-\omega) + \int_{\mathbb{R}^n} \varphi(\vec{x}) \left[\omega(c_{10} - c_{00}) p_0(\vec{x}) - (1-\omega)(c_{01} - c_{11}) p_1(\vec{x}) \right] d\vec{x}$$

$\cancel{\Rightarrow \varphi(\vec{x})}$

$$\Rightarrow \varphi(\vec{x}) = \begin{cases} 1, & \frac{p_0(\vec{x})}{p_1(\vec{x})} < \frac{(1-\omega)(c_{01}-c_{11})}{\omega(c_{10}-c_{00})} \\ 0, & \text{—}''\text{—} \geq \text{—}''\text{—} \end{cases}$$

Если $c_{01} = c_{10} = 1$
$c_{00} = c_{11} = 0$

$\mathbb{E}\, C_{ZY} = \ldots = (1-\omega)\int\limits_{\mathbb{R}^n}(1-\varphi(\vec{x}))p_1(\vec{x})d\vec{x} + \omega\int\limits_{\mathbb{R}^n}\varphi(\vec{x})p_0(\vec{x})d\vec{x}$ $\Rightarrow$ минимизируется вер-ть ошибки

полная вер-ть ошибки

(если ω неизв.)
Применяя принцип максимума энтропии $\omega = \frac{1}{2}$
(или Лапласовского индетерминизма)

число

$\Rightarrow l = \frac{p_0(\vec{x})}{p_1(\vec{x})} \lessgtr 1$

монотон. по $\vec{x}$
(байесовское правило)

$\frac{p_0(X)}{p_1(X)}$ ← ср-ция правдоподобия

Рассм. задачу с знач. неравноправием гипотез

Зафикс. малое α, и потребуем $E_0\varphi(X) \leq \alpha$

И при этом условии констр. крит.: $E_1(1-\varphi(X)) \to \min$

Т. (Лемма Неймана-Пирсона)

$\varphi, \varphi^* : \mathbb{E}_0\varphi(X) = \mathbb{E}_0\varphi^*(X) = \alpha \quad \Rightarrow E_1(1-\varphi(X)) \leq \mathbb{E}_1(1-\varphi^*(X))$

$$\varphi(X) = \begin{cases} 0, & cp_0(\vec{x}) > p_1(\vec{x}) \\ 1, & cp_0(\vec{x}) < p_1(\vec{x}) \\ \lambda, & \text{—}''\text{—} = \text{—}''\text{—} \end{cases}$$

байесовский крит.

, в абс. непр. случае вер-ть рав-ва $= 0$
$\Rightarrow \lambda$ не важна

Док-во:

$S_0 = \{\vec{x} : cp_0(\vec{x}) > p_1(\vec{x})\}$

$S_1 = \{\vec{x} : cp_0(\vec{x}) < p_1(\vec{x})\}$

$S_2 = \{\vec{x} : cp_0(\vec{x}) = p_1(\vec{x})\}$

$x \in S_0 \Rightarrow (\varphi^*(\vec{x}) - \varphi(\vec{x}))p_1(\vec{x}) \leq (\varphi^*(\vec{x}) - \varphi(x))p_0(\vec{x})c$

$x \in S_1 \Rightarrow$ —"—

$x \in S_2 \Rightarrow (\varphi^*(\vec{x}) - \varphi(\vec{x}))(p_1(\vec{x}) - cp_0(\vec{x})) = 0$

$$\int\limits_{\mathbb{R}^n}(\varphi^*(x) - \varphi(x))(p_1(\vec{x}) - cp_0(\vec{x}))dx = \int\limits_{\mathbb{R}^n}(\varphi^*(\vec{x}) - \varphi(\vec{x}))p_1(\vec{x})d\vec{x} - c\int\limits_{\mathbb{R}^n}(\varphi^*(\vec{x}) - \varphi(\vec{x}))p_0(\vec{x})d\vec{x} \leq 0$$

$$\Rightarrow \int\limits_{\mathbb{R}^n}(\varphi^*(\vec{x}) - \varphi(\vec{x}))p_1(\vec{x})d\vec{x} \leq c\int\limits_{\mathbb{R}^n}(\varphi^*(\vec{x}) - \varphi(\vec{x}))p_0(\vec{x})d\vec{x} = c\int\limits_{\mathbb{R}^n}\varphi^*(\vec{x})p_0(\vec{x})d\vec{x} - c\int\limits_{\mathbb{R}^n}\varphi(\vec{x})p_0(\vec{x})d\vec{x}$$

$= c\cdot\mathbb{E}_0\varphi^*(\vec{x}) - c\,\mathbb{E}_0\varphi(\vec{x}) = 0 \Rightarrow \mathbb{E}_1(1-\varphi(X)) \leq \mathbb{E}_1(1-\varphi^*(X))$ ∎

$\mathbb{X} = (X_1, \dots, X_n)$ - дискр.

$$T = \frac{p_1(\mathbb{X})}{p_0(\mathbb{X})}, \quad l_T(\alpha) \leftarrow \begin{cases} \mathbb{P}_0(T \le l_T(\alpha)) \ge \alpha \\ \mathbb{P}_0(T \ge l_T(\alpha)) \ge 1-\alpha \end{cases}, \quad c \equiv l_T(1-\alpha)$$

$$\varphi(\mathbb{X}) = \varphi(T) = \begin{cases} 1, & T > c \\ 0, & T < c \\ \lambda, & T = c \end{cases}, \quad \mathbb{E}_0 \varphi(T) = \sum_k \varphi(t_k) \mathbb{P}_0(T = t_k) =$$

$$= \sum_{k:\, t_k < c} \underbrace{\varphi(t_k)}_{=0} \mathbb{P}_0(T = t_k) + \sum_{k:\, t_k > c} \underbrace{\varphi(t_k)}_{=1} \mathbb{P}_0(T = t_k) +$$

$$+ \underbrace{\varphi(c)}_{=\lambda} \mathbb{P}_0(T = c) = \mathbb{P}_0(T > c) + \lambda \mathbb{P}_0(T = c)^{0} = \alpha$$

$$\Rightarrow \lambda = \frac{\alpha - \mathbb{P}_0(T > c)}{\mathbb{P}_0(T = c)},$$

(вычислять в терминах ф.р.)

$\mathbb{X} = (X_1, \dots, X_n)$

$$H_0: p_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a_0)^2}{2}}$$

$$H_1: p_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a_1)^2}{2}}$$

$a_0 > a_1$

$$T(\mathbb{X}) = \frac{p_0(\mathbb{X})}{p_1(\mathbb{X})}; \quad p_0(\mathbb{X}) = \prod_{i=1}^n p_0(X_i) = (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^n (X_i - a_0)^2} =$$

$$= (2\pi)^{-n/2} e^{-\frac{1}{2}\sum_{i=1}^n X_i^2 - \frac{n}{2} a_0^2 + a_0 \sum_{i=1}^n X_i}$$

$$T(\mathbb{X}) = \underbrace{e^{-\frac{n}{2}(a_0^2 - a_1^2)}}_{>0} e^{\overbrace{(a_0 - a_1)}^{>0} \sum_{i=1}^n X_i}$$

$$\Rightarrow T \gtrless c \Leftrightarrow \sum_{i=1}^n X_i \gtrless c^*$$

$$\sum_{i=1}^n X_i \sim \mathcal{N}(na, n)$$

$$\alpha = \mathbb{P}_0(T < c) = \mathbb{P}_0\left(\sum_{i=1}^n X_i < c^*\right) = \Phi\left(\frac{c^* - na_0}{\sqrt{n}}\right); \quad \underbrace{\frac{c^* - na_0}{\sqrt{n}}}_{<0} = u_\alpha$$

$\Rightarrow c^* = na_0 + u_\alpha \sqrt{n}$ не зав. от. a_1 => равномерно наиболее мощный критерий

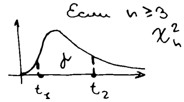
$Y = (x_1, \dots, x_n)$ из $\ell(X) \in \{P_\theta, \theta \in \Theta\} = \{N(0, \theta^2)\}$

$H_0 : \theta = \theta_0$

$H_1 : \theta \neq \theta_0$

Построить дов. инт-л для θ^2 с коэф. доверия γ

Если $n \geq 3$



Рассм. $\left(\frac{x_1}{\theta}\right)^2 + \dots + \left(\frac{x_n}{\theta}\right)^2 = T(Y, \theta^2)$ - ц.с. $\sim \chi^2_n$

$P(t_1 < T(Y, \theta^2) < t_2) = \gamma \quad (1)$

<u>НМК $\Leftrightarrow$ кратч. ДИ</u>

Центр. инт-л: t_1, t_2 найдены по χ^2_n

$(1) \Leftrightarrow \mathbb{P}\left(\frac{\sum x_i^2}{t_2} < \theta^2 < \frac{\sum x_i^2}{t_1}\right) = \gamma$

Проверим гипотезу.

По задан. α строится S-крит. (Если $y = (x_1, \dots, x_n) \in S$, то H_0 отверг.
—"— $\in \overline{S}$, то H_0 приним.)

$\alpha = \mathbb{P}(H_1 | H_0) = \mathbb{P}(Y \in S | H_0)$

Для $\forall \theta \in \Theta$ опред. мн-во $\overline{S}(\theta) \subset \mathcal{X}$ выбороч. пр-во

$P_\theta(Y \in \overline{S}(\theta)) = 1 - \alpha$

Определим для $\forall y \in \mathcal{X}$ $T(y) \subset \Theta : \theta \in T(y) \Leftrightarrow y \in \overline{S}(\theta)$

$\Rightarrow P_\theta(\theta \in T(y)) = P_\theta(y \in \overline{S}(\theta)) = 1 - \alpha$

$\Rightarrow T(y)$ можно считать доверит. мн-во для θ с коэф. доверия $1 - \alpha$.

<u>Опр.</u> Критерий наз-ся состоятельным, если его мощность $\xrightarrow[n \to \infty]{} 1$

$H_0 : p_i = p_{i0} \quad i = \overline{1, k} \qquad \hat{\psi}^2 = \sum_{i=1}^{k} \frac{(\nu_i - np_{i0})^2}{np_{i0}}$

$H_1 : p_i = p_{i1}$

<u>Теор.</u> При $\forall$ простой альтернативе H_1

$\underset{\text{мощность}}{\mathbb{P}(H_1 | H_1)} = \mathbb{P}(\hat{\psi}^2 > \psi_{кр.} | H_1) \longrightarrow 1$, т.е. критерий явл. состоятельным

Док-во: $\exists\, c = \psi_{кр.}$. $\mathbb{P}(\hat{\psi}^2 < c | H_1) = (*)$

Если H_1, то $\nu_i \sim Bi(n, p_{i1})$

$$\hat{\psi}^2 = \sum_{i=1}^{k} \frac{(\nu_i - np_{i1})^2}{np_{i0}} + 2 \sum_{i=1}^{k} (\nu_i - np_{i1}) \frac{p_{i1} - p_{i0}}{p_{i0}} + \sum_{i=1}^{k} \frac{n(p_{i1} - p_{i0})^2}{p_{i0}} =$$

$$= T_1^2 + \underbrace{2 \sum_{i=1}^{k} (\nu_i - np_{i1}) \frac{p_{i1} - p_{i0}}{p_{i0}}}_{T_2} + nA$$

$(*) = \mathbb{P}(T_1^2 + T_2 + nA < c | H_1) \leq \mathbb{P}(T_2 + nA < c | H_1) = \mathbb{P}(T_2 < -nB | H_1) \leq$

$$\leq \mathbb{P}(|T_2| > nB \mid H_1) \leq \frac{\mathbb{E} T_2^2}{n^2 B^2} \longrightarrow 0$$

нер-во Маркова

Коши-Буняковского

$$\mathbb{E} T_2^2 \leq 4n \sum_{i=1}^{k} p_{i1}(1-p_{i1}) \sum_{i=1}^{k} \frac{(p_{i1}-p_{i0})^2}{p_{i0}^2}$$

т.к. $\mathbb{E}_1(\nu_i - np_{i1})^2 = np_{i1}(1-p_{i1})$ ∎