12/5/21

$\vec{X} \in N(\alpha_1, \sigma)$, n шт $\qquad \vec{Y} \in N(\alpha_2, \sigma)$, m шт

$H_0: \{\alpha_1 = \alpha_2\}$
$H_1: \overline{H_0}$

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{nS_x^2 + mS_y^2}} \sqrt{\frac{m+n-2}{\frac{1}{m} + \frac{1}{n}}}$$

$$\underbrace{\overline{X} - \overline{Y} \in N\left(0, \sigma\sqrt{\tfrac{1}{n} + \tfrac{1}{m}}\right)}_{\text{т.к. } D\overline{X} = \sigma^2/n,\ D\overline{Y} = \sigma^2/m} \Rightarrow \frac{\overline{X} - \overline{Y}}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \in N(0,1)$$

$$\frac{nS_x^2}{\sigma^2} = \sum_{i=1}^{n} \Big(\underbrace{\frac{X_i - \overline{X}}{\sigma}}_{\in N(0,1)}\Big)^2 \qquad \frac{mS_y^2}{\sigma^2} = \sum_{i=1}^{m} \Big(\underbrace{\frac{Y_i - \overline{Y}}{\sigma}}_{\in N(0,1)}\Big)^2$$

т.о. $\dfrac{nS_x^2}{\sigma^2} + \dfrac{mS_y^2}{\sigma^2} \in \chi^2_{\underbrace{n+m-2}_{(n-1)+(m-1)}}$

$\dfrac{\xi_0}{\sqrt{\frac{1}{N}\chi_N^2}} \in T_N$ – РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА, где $\xi \in N(0,1)$

В нашем случае $N := n+m-2$, т.о.

$$\frac{\overline{X} - \overline{Y}}{\sigma\sqrt{1/n + 1/m}} \cdot \sqrt{\frac{n+m-2}{\frac{nS_x^2 + mS_y^2}{\sigma^2}}} = \frac{\overline{X} - \overline{Y}}{\sqrt{nS_x^2 + mS_y^2}} \cdot \sqrt{\frac{n+m-2}{1/n + 1/m}} =: T$$

$F(n-1, m-1) = \dfrac{nS_x^2}{mS_y^2}$ – РАСПРЕДЕЛЕНИЕ ФИШЕРА

По сути $\dfrac{\chi^2_{n-1}}{\chi^2_{m-1}} \in F_{n-1, m-1}$

Критерий: $\dfrac{X_1^2 + \dots + X_n^2}{Y_1^2 + \dots + Y_m^2} \leq C_\varepsilon$

Критерий Колмогорова – Смирнова

$\vec{X} \in F_1$, n
$\vec{Y} \in F_2$, m

$H_0: F_1 = F_2$
$H_1: \overline{H_0}$

F_1, F_2 – непрерывны
для разрывных неверно

$F^*_{nx}(t)$ по $\vec{X}$, $F^*_{ny}(t)$ по $\vec{Y}$

$d(F^*_{nx}, F^*_{ny}) = \sqrt{\dfrac{mn}{m+n}} \sup\limits_t |F^*_{nx}(t) - F^*_{ny}(t)|$ – РАССТОЯНИЕ КОЛМОГОРОВА

Используется при $m, n \gg 1$

Теорема Колмогорова – Смирнова

$$\mathbb{P}\left(d(F^*_{nx}, F^*_{my}) < t\right) \xrightarrow[n,m\to\infty]{} K(t) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2t^2}$$

→ функция Колмогорова

$$\delta(X) = \begin{cases} 0, & d \le t_\varepsilon \\ 1, & d > t_\varepsilon \end{cases}$$

Задача о сопряженных признаках

$\vec{X}, n$
$\vec{Y}, n$ одного объёма

Получаем $(X_1, Y_1), \ldots, (X_n, Y_n)$ – "двумерие"

при этом X_i, Y_i могут быть векторн.

$H_0 = \{X_i \perp\!\!\!\perp Y_i\}$
$H_1 = \overline{H_0}$

проверяем "сопреженность"

$\mathbb{X}$ разбиваем на $\{\Delta_i \mid i = \overline{1,m}\}$

$\mathbb{Y}$ разбиваем на $\{\delta_j \mid j = \overline{1,k}\}$

выборочные пр-ва

$$\nu_{ij} = \#\{(x_k, y_k) \mid x_k \in \Delta_i,\ y_k \in \delta_j\}$$

$$\nu_{i*} = \sum_{j=1}^{k} \nu_{ij}, \quad \nu_{*j} = \sum_{i=1}^{m} \nu_{ij}$$

$$\sum_{\substack{i=1,\ldots,m \\ j=1,\ldots,k}} \frac{(\nu_{ij} - np_{ij})^2}{np_{ij}} \qquad \{p_{ij}\} \qquad \dim\Theta = mk - 1$$

$\chi^2_{N-1-\dim\Theta}$ Теорема Пирсона не прокатит

H_0 – реал. $\longrightarrow$ $p_{ij} = \mathbb{P}(X_1 \in \Delta_i, Y_1 \in \delta_j) = p_{i*} p_{*j}$

размерность $m+k-1$

размерность $k-1$

размерность $m-1$

$$\hat{p}_{i*} = \frac{\nu_{i*}}{n}, \quad \hat{p}_{*j} = \frac{\nu_{*j}}{n} \text{ – ОМП}$$

$$\sum_{\substack{i=1,\ldots,m \\ j=1,\ldots,k}} \left(\frac{\nu_{ij} - n\hat{p}_{i*}\hat{p}_{*j}}{n\hat{p}_{i*}\hat{p}_{*j}}\right) \xRightarrow[n\to\infty]{} \eta \in \chi^2_{mk-1-(m+k-2)} = \chi^2_{(m-1)(k-1)}$$

Пример

# пропуск. \ ср. балл	(2,4)	[4,5]	
[0,30]	11	25	$\nu_{1\cdot} = 36$
$(30,+\infty)$	28	16	$\nu_{2\cdot} = 44$
	$\nu_{\cdot 1} = 39$	$\nu_{\cdot 2} = 41$	

$n = 80$ $\quad m = k = 2$

$$\chi^2 = 8{,}67$$

$c_\varepsilon = 3{,}84$ – 5%-критерий $\quad \varepsilon = 0{,}05$

$3{,}84 \ll 8{,}67 \implies$ сопряжены (связаны)

Байесовский подход к построению критериев

$H_1, \dots, H_m$ – простые гипотезы $\qquad q_1, \dots, q_m$ – априорные вероятности

$$\alpha_i(\delta) = \mathbb{P}_i(\delta \neq i)$$

$$\alpha_q(\delta) = \sum_{i=1}^{m} \alpha_i(\delta)\, q_i \longrightarrow \min$$

Теорема Байеса

$$\exists\, \delta^* : \alpha_q(\delta^*) = \min_{\delta} \alpha_q(\delta)$$

Теорема (критерий Байеса)

$$\delta^*(X) = i, \text{ если } q(i \mid X) = \max_k q(k \mid X) \quad \circledast$$

$$q(k \mid X) = \frac{\varphi_X(k)\, q_k}{\sum_j \varphi_X(j)\, q_j} \qquad \mapsto \ = \prod_{i=1}^{n} f_k(x_i)$$

$$\delta(X) = i, \text{ если } X \in S_i$$

$$\alpha_q(\delta) = \sum_{i=1}^{m} \left(1 - \mathbb{P}_i(\delta = i)\right) q_i = 1 - \sum_{i=1}^{m} \mathbb{P}_i(\delta = i)\, q_i = 1 - \sum_{i=1}^{m} \mathbb{P}(X \in S_i)\, q_i =$$

$$= 1 - \sum_{i=1}^{m} q_i \cdot \int_{S_i} \varphi_{\vec z}(i)\, \mu^n(d\vec z) \geq 1 - \sum_{i=1}^{m} \int_{S_i} \underbrace{\max_k q\ \varphi_{\vec z}(k)}_{\text{от } i \text{ не зависит}} \mu^n(d\vec z) =$$

равенство как раз при $\circledast$

$$= 1 - \int_{\mathbb{X}^n} \max q \cdot \varphi_{\vec z}(k)\, \mu^n(d\vec z) = 1 - \int_{\mathbb{X}^n} \max q(k \mid X)\, f(\vec z)\, \mu^n(d\vec z)$$

Теорема Неймана – Пирсона

H_1, H_2 – простые

$K = \{\delta \mid \alpha_1(\delta) \le \varepsilon\}$

$\exists\, \tilde{\delta} = \delta_{opt} \ : \quad \alpha_2(\tilde{\delta}) = \inf\limits_{\delta} \alpha_2(\delta)$

Критерий Неймана – Пирсона

$$\tilde{\delta}(X) = \begin{cases} 1, & \varphi_X(2)/\varphi_X(1) \le C_\varepsilon \\ 2, & \varphi_X(2)/\varphi_X(1) > C_\varepsilon \end{cases}$$

$$\mathbb{P}\left(\varphi_X(2)/\varphi_X(1) > C_\varepsilon\right) = \varepsilon$$