28/4/21

Проверка статистических гипотез

ГИПОТЕЗА — любое суждение о неизвестном распределении

ПРОСТАЯ гипотеза однозначно восстанавливает распределение

СЛОЖНАЯ гипотеза неоднозначно восстанавливает распределение

Пример $H_1 = \{X \in N(0,1)\}$ — простая гипотеза

$H_2 = \{X \in N(a,\sigma)\}$ — сложная гипотеза

СТАТИСТИЧЕСКОЕ РЕШАЮЩЕЕ ПРАВИЛО (КРИТЕРИЙ) —

m-точечное отображение выборочного пространства

Обозн $\delta(X) = \begin{cases} 1, & \text{если} \dots \\ \vdots \\ m, & \text{если} \dots \end{cases}$ $1,..,m$ — индексы гипотез $H_1,..,H_m$

$\alpha_i(\delta) = \mathbb{P}_i(\delta \neq i)$ — **ВЕРОЯТНОСТЬ ОШИБКИ i-ГО РОДА**

Пример У врача $H_1 = \{здоров\}$ Ошибка I рода — лечат здорового
$H_2 = \{болен\}$ Ошибка II рода — не лечат больного

ЛУЧШЕ тот критерий, у которого вероятности ошибок меньше
сделать очень маленькими вероятности ошибок всех родов обычно невозм.

Метод минимального расстояние

$H_1 = \{F = F_0\}$ — **ОСНОВНАЯ** гипотеза

$H_2 = \{F \neq F_0\}$ — **АЛЬТЕРНАТИВНАЯ** гипотеза (**АЛЬТЕРНАТИВА**)

В рассматриваемой паре гипотез основная — простая,
альтернативная — сложная

$d(F_1, F_2)$ — **СТАТИСТИЧЕСКОЕ РАССТОЯНИЕ**, если ① $d(F_1,F_2) \geq 0$
② $d(F,F) = 0$

это не метрика! нет нерва треуг.
и точности

Замечание $F_n^*(t) = \dfrac{\#\{x_i \mid x_i < t\}}{n}$

$\underbrace{\sup_t |F_n^*(t) - F_0(t)|}_{} \xrightarrow[n\to\infty]{п.н.} 0$ по т. Глив-Кантели

↳ расстояние Колмогорова $d_K(F_n^*, F_0)$

Статистическое расстояние непр. отн. d_K =>

=> $d(F_n^*, F_0) \xrightarrow[n\to\infty]{} 0$

Вернемся к $H_1 = \{F = F_\theta\}$, $H_2 = \{F \neq F_\theta\}$

Метод мин. расст. : $\delta(X) = \begin{cases} 1, & d(F_n^*, F_0) \le c_\varepsilon \\ 2, & d(F_n^*, F_0) > c_\varepsilon \end{cases}$ → критическое значение

$$\varepsilon = \alpha_1(\delta) = \mathbb{P}_1(\delta = 2) = \underline{\mathbb{P}_1\left(d(F_n^*, F_0) > c_\varepsilon\right) = \varepsilon}$$

↳ отсюда ищем c_ε. По сути — квантиль уровня ε

Критерий Колмогорова

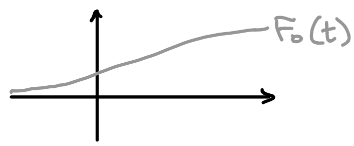
$$d_K(F_1, F_2) = \sup_t |F_1(t) - F_2(t)|$$

Лемма (Свойство непараметричности критерия Колмогорова)

| F_0 (истинное распр) — непрерывная ф-ия

$\Rightarrow \mathbb{P}(d_K(F_n^*, F_0) < t)$ не зависит от F_0 (параметром как раз считается F_0)

Д-во



$$\sup_t |F_n^*(t) - F_0(t)| = \underline{\sup_{0<s<1}} \left| F_n^*(F_0^{-1}(s)) - F_0(F_0^{-1}(s)) \right|$$

↳ квантильное преобраз.

$$= \sup_{0<s<1} |F_n^*(F_0^{-1}(s)) - s| = \Big/ F_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} I\{X_i < t\} \Big/ =$$

$$= \sup_{0<s<1} \left| \frac{1}{n}\sum_{i=1}^{n} I(\underline{F_0(X_i)} < s) - s \right|$$

↳ имеет равномерное на $[0,1]$ распределение

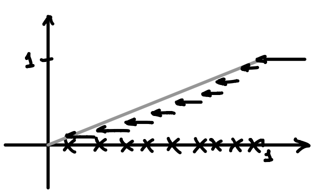
Т.о. $\mathbb{P}(d_K(F_n^*, F_0) < t) = \mathbb{P}(d_K(U_n^*, U_{0,1}) < t)$ — не зависит от F_0 ■

Теорема Колмогорова

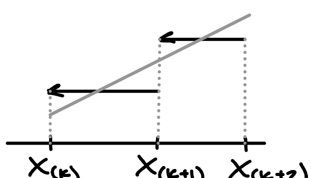
$$\mathbb{P}(\sqrt{n}\, d_K(F_n^*, F_0) < t) \xrightarrow[n\to\infty]{} \underline{K(t)} := \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2t^2}$$

↳ функция Колмогорова

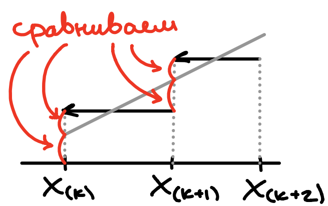
Как считать расчетное задание



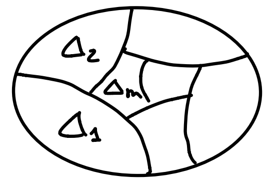
если приблизить:



$\sup |F_n^* - F_0|$ достигается в узлах



Расстояние χ^2



Делаем конечное разбиение $\mathbb{X}$

$$\{\Delta_1, \dots, \Delta_m\} \qquad \nu_i = n\mathbb{P}_n^*(\Delta_i)$$

$$\mathcal{P}_1(\Delta) = \mathbb{P}_1(X_i \in \Delta)$$

$$X_i \in \mathbb{X}$$

$$\mathcal{P}_2(\Delta) = \mathbb{P}_2(X_i \in \Delta)$$

$$d_{\chi^2}(\mathcal{P}_1, \mathcal{P}_2) = \sum_{i=1}^{m} \frac{(\mathcal{P}_1(\Delta_i) - \mathcal{P}_2(\Delta_i))^2}{\mathcal{P}_2(\Delta_i)}$$ — РАССТОЯНИЕ χ^2

$$d(\mathbb{P}_n^*, \mathcal{P}_0) = \sum_{i=1}^{m} \frac{\left(\frac{\nu_i}{n} - p_i\right)}{p_i} = \sum_{i=1}^{n} \frac{(\nu_i - np_i)^2}{n^2 p}$$

p_i → истинное распр

Теорема Пирсона

$$n d_{\chi^2}(\mathbb{P}_n^*, \mathcal{P}_0) \Rightarrow \eta \in \chi^2_{m-1} \iff \eta \overset{d}{=} \sum_{i=1}^{m+1} \xi_i^2, \quad \vec{\xi} \text{ - нор СВ } N_{0,1}$$

Теорема Пирсона (обобщенная)

$$\begin{cases} H_0 = \{X \in \{F_\theta\}\} \\ H_1 = \overline{H_0} \end{cases}$$ — основная гипотеза сложная

$$\Rightarrow n\tilde{d}_{\chi^2}(\mathbb{P}_n^*, \mathcal{P}_0) \Rightarrow \eta \in \chi^2_{m-1-\dim(\theta)}$$

<u>Д-во</u>
схема

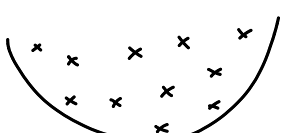
$$\tilde{d}(\mathbb{P}_n^*, \mathcal{P}_0) = \sum \frac{(\nu_i - np_i)^2}{n^2 p_i}, \text{ где } p_i = p_i(\theta)$$

вместо θ возьмем ОМП $\hat{\theta}_n$

<u>Пример</u> $X \in \{N(\alpha, \sigma) \mid \alpha \in \mathbb{R}, \sigma > 0\}$ $\qquad \theta = (\alpha, \sigma) \leftarrow (\overline{X}, S)$

<u>Замечание</u> При небольших объемах выборки разбиваем так, что $np_i \approx \nu_i$ и $\nu_i \geq 10$

<u>Пример</u> Лондон, II Мировая Война



Южный Лондон

$H_0 = \{$ракеты попадают наугад$\}$
$H_1 = \{$ракеты точны$\}$
535 ракет

576 квадратов

i – Кол-во попаданий в квадрат
ν_i – кол-во таких квадратов

i	0	1	2	3	4+
ν_i	229	211	93	35	8

$p = \frac{1}{576}$ – вероятность попадания ракеты в отдельный квадрат, если H_0 верна

По сути проверяем Пуассоновость распределения $(np = \lambda)$

$np = \frac{535}{576} \approx 0,9$

$d_{\chi^2}(\mathbb{P}^*, \mathcal{P}_0) \approx 2,5$

$\hookrightarrow \Pi_{np}$