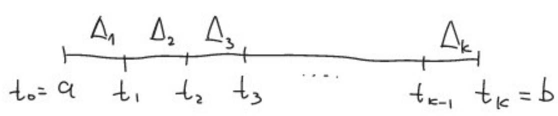
Критерий хи-квадрат.

Критерий χ^2 Пирсона

Основывается на группированных данных. Область значений F_0 делят на некоторое число интервалов, затем строят функцию отклонения Ψ по разностям теоретических вероятностей попадания в интервалы и эмпирических частот.



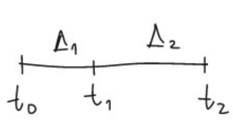
$$\nu_i = \sum_{i=1}^{n} I(x \in \Delta_i) \qquad p_i = P(x \in \Delta_i) = F_0(t_i) - F_0(t_{i-1})$$

$$\nu_1 + \dots + \nu_n = n \qquad p_1 + \dots + p_n = 1$$

$$d(F_n^*, F_0) = (\chi^2)^* = \sum_{i=1}^{n} \frac{(\nu_i - np_i)^2}{np_i} \Rightarrow \eta \in \chi^2_{k-1} \text{ (при верной } H_0\text{)}$$

$$\delta = \begin{cases} 0, & (\chi^2)^* < c \\ 1, & (\chi^2)^* \geq c \end{cases}, \text{ где } \chi^2_{k-1}(c) = 1 - \varepsilon$$

Докажем теорему Пирсона для k = 2



$$\Psi = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(\nu_2 - np_2)^2}{np_2} = \frac{(\nu_1 - np_1)^2}{np_1} + \frac{(\overbrace{(n-\nu_1) - n(1-p_1)}^{n-\nu_1-n+np_1})^2}{n(1-p_1)} =$$

$\nu_1 + \nu_2 = n$

$p_1 + p_2 = 1$

$$= \frac{(\nu_1 - np_1)^2(1 - \cancel{p_1}) + \cancel{(np_1 - \nu_1)^2} \cdot p_1}{np_1(1-p_1)} = \left(\frac{\nu_1 - \overbrace{np_1}^{\text{EX Бернулли}}}{\sqrt{\underbrace{np(1-p)}_{\text{DX Бернулли}}}}\right)^2 \Rightarrow \xi^2, \quad \xi \in N_{0,1}$$

Сх. Бернулли: попадаем в Δ_1 – успех (p_1)

в Δ_2 – неудача ($1-p_1$) $\qquad \nu_1$ – число успехов

Состоятельность χ^2

$H_0 = \{\vec{X} \in F_0\}$

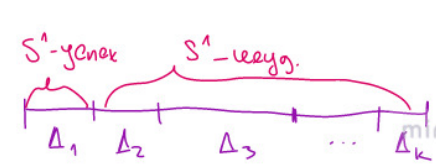
$H_a = \{\vec{X} \notin F_0\}$

– Для этих гипотез не всегда будет выполнятся k_2.
Если распределение $\vec{X}$ $\mathcal{F}_1 \neq \mathcal{F}_0$ имеет такие же как у $\mathcal{F}_0$ вероятности p_i попадания в каждый из Δ_i то по данной ф-ции $(\chi^2)^*$ эти распределения различить невозможно.

Поэтому введем более корректные гипотезы:

$$H_0' = \{\mathcal{F} : \forall i \quad P(X_1 \in \Delta_i) = p_i\}$$

$$H_a' = \{\mathcal{F} : \exists i : P(X_1 \in \Delta_i) \neq p_i\}$$



Введем случайную величину $\delta^i(x) = \begin{cases} 1, & x \in \Delta_i \\ 0, & x \notin \Delta_i \end{cases}$ – Сх. Бернулли $\begin{pmatrix} \text{успех} - p_i \\ \text{неуд.} - (1-p_i) \end{pmatrix}$

Тогда по ЗБЧ, $\dfrac{\delta_1^i + \ldots + \delta_n^i}{n} \xrightarrow{P} \mathbb{E}\,\delta_1^i = p_i$

$$\alpha_2 = P_{H_a}\left(\sum_{j=1}^{k} \frac{(\nu_j - np_j)^2}{np_j} < C \right)$$

Если верна $H_a \Rightarrow \exists i: \; P(x \in \Delta_i) \neq p_i$

По ЗБЧ

$\dfrac{\delta_1^i + \ldots + \delta_n^i}{n} \xrightarrow{P} \mathbb{E}\,\delta_n^i = P(x \in \Delta_i) = p_k \neq p_i$

Тогда, рассмотрим i-ое слагаемое: $\dfrac{(\nu_i - np_i)^2}{np_i} = \dfrac{n}{p_i}\left(\dfrac{\nu_i}{n} - p_i\right)^2 \xrightarrow{P} \dfrac{n}{p_i}\underbrace{\left(p_k - p_i\right)^2}_{>0} \xrightarrow[n\to\infty]{} \infty$

Поэтому

$$\alpha_2(\delta) = P_{H_a}\left(\sum_{j=1}^{n} \frac{(\nu_j - np_j)^2}{np_j} < C \right) \to 0$$

$\searrow \infty$