TEST DEL χ^2

Note Title 28/05/2019

$X_1 \dots X_n$ campione statistico relativo ad una distribuzione concentrata su dei valori noti

$$t_1 \text{ — } t_k$$

Conosco la distribuzione se conosco

$$\forall j = 1 \text{ — } k \qquad p_j := \mathbb{P}(X_i = t_j)$$

$x_1 \dots x_n$ dati rilevati

Ci sono n_1 dati uguali a t_1

n_2 dati uguali a t_2

$\vdots$

n_k dati uguali a t_k

$$g(x_1 \text{ — } x_n \mid p_1 \text{ — } p_k) = \mathbb{P}(X_1 = x_1, X_2 = x_2 \dots X_n = x_n)$$
$$= \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) = p_1^{n_1} \cdot p_2^{n_2} \dots p_k^{n_k}$$

$$p_j \geq 0 \quad \forall j = 1 \text{ — } k \qquad \sum_{j=1}^{k} p_j = 1 \quad \longleftarrow$$

$$h(x_1 \text{ — } x_n \mid p_1 \text{ — } p_k) = \log \prod_{j=1}^{k} p_j^{n_j} = \sum_{j=1}^{k} n_j \log p_j$$

$$p_k = 1 - \sum_{j=1}^{k-1} p_j$$

$$H(p_1, \text{ — }, p_{k-1}) = \sum_{j=1}^{k-1} n_j \log(p_j) + n_k \log\left(1 - \sum_{j=1}^{k-1} p_j\right)$$

$$\forall s = 1 \text{ — } k-1 \qquad \frac{\partial H}{\partial p_s} = \frac{n_s}{p_s} + n_k \frac{-1}{\underbrace{1 - \sum_{j=1}^{k-1} p_j}_{= p_k}} = 0$$

$$\frac{n_s}{p_s} = \frac{n_k}{p_k} \qquad \forall s = 1 \_\_ k-1$$

$$p_s = \frac{n_s}{n_k} p_k \qquad \forall s = 1 \_\_ k-1$$

$$\sum_{j=1}^{k} p_j = 1 \qquad \sum_{j=1}^{k-1} \frac{n_j}{n_k} p_k + p_k = 1$$

$$\left(\sum_{j=1}^{k-1} n_j + n_k \right) p_k = n_k \qquad n p_k = n_k$$

$p_k = \frac{n_k}{n}$ = frequenza relativa del valore t_k nel campione $x_1 \_\_ x_n$

$$p_s = \frac{n_s}{n_k} p_k = \frac{n_s}{n_k} \frac{n_k}{n} = \frac{n_s}{n} \qquad \forall s = 1 \_\_ k-1$$

= frequenza relativa del valore t_s nel campione $x_1 \_\_ x_n$

$$p_j = \frac{1}{n} \# \{ i \in \{1 \_\_ n\} : X_i = t_j \} \qquad \forall j = 1 \_\_ k$$

— o —

Sia $p :=$ probabilità di successo in ogni prova di un esperimento binario

$\Rightarrow$ la probabilità di ottenere insuccesso è $1 - p$.

Ripeto l'esperimento n volte: otterrò un numero di successi $k \in \{0, 1, \_\_, n-1, n\}$

$\mathbb{P}$ di ottenere una stringa prescritta $x_1 x_2 \dots x_n$ che contiene k successi e $n-k$ insuccessi

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \dots X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$$
$$= p^k (1-p)^{n-k}$$

$\mathbb{P}$ di ottenere k successi ??

$$\begin{matrix} 1 & 2 & 3 & \dots & n \\ \downarrow & & \downarrow & & \downarrow \\ 1 & & 3 & \dots & n-1 \end{matrix}$$

Ci sono $\binom{n}{k}$ diverse stringhe che contengono esattamente K successi

$\mathbb{P}$ di ottenere esattamente K successi in n prove è

$$\binom{n}{k} p^k (1-p)^{n-k} = B(n,p)(\{k\})$$

— - —

Siano Y_1 — Y_n un campione statistico relativo a una distribuzione concentrata su un insieme finito t_1 — t_k

La distribuzione è nota se per ogni $j = 1$ — k conosco $p_j :=$ densità della distribuzione in t_j
$$= \mathbb{P}(Y_i = t_j)$$

Siano p_1^0 — p_k^0 assegnati $\quad p_j^0 \geq 0 \quad \forall j = 1 - k$
$$\sum_{j=1}^{k} p_j^0 = 1$$

$H_0 : p_j = p_j^0 \ \forall j = 1 - k \qquad H_1 : \exists j$ t.c. $p_j \neq p_j^0$

Per ogni $j = 1$ — k considero
$$X_j = \#\{ i \in \{1 - n\} : Y_i = t_j \}$$

$$\mathbb{P}_{X_j} = B(n, p_j) \qquad \forall j = 1 \ldots k$$

$$\mathbb{E}[X_j] = np_j \qquad Var[X_j] = np_j(1-p_j)$$

$$|X_j - np_j^\circ|^2 \qquad a_j > 0 \qquad \sum_{j=1}^{k} a_j (X_j - np_j^\circ)^2 < \varepsilon$$

TEOREMA DI PEARSON (no dim)

Se $\mathbb{P}_{X_j} = Bin(n, p_j^\circ)$ $\forall j = 1 - k$, allora le v.a.
e $X_1 + X_2 \ldots + X_k = n$

$$\sum_{j=1}^{k} \frac{(X_j - np_j^\circ)^2}{np_j^\circ}$$

, per $n \to \infty$, converge in legge a una v.a. con distribuzione χ^2_{k-1}

Accetto H_0 se $\sum_{j=1}^{k} \frac{(x_j - np_j^\circ)^2}{np_j^\circ} < \varepsilon$

Probabilità di commettere errore di 1ª specie è

$$\mathbb{P}\left(\sum_{j=1}^{k} \frac{(X_j - np_j^\circ)^2}{np_j^\circ} \geq \varepsilon \;\middle|\; \mathbb{P}_{X_j} = B(n, p_j^\circ)\ \forall j \right)$$

$$\simeq \mathbb{P}(\chi_{k-1} \geq \varepsilon) = 1 - F_{\chi^2_{k-1}}(\varepsilon) = \alpha$$

$$F_{\chi^2_{k-1}}(\varepsilon) = 1 - \alpha \qquad \varepsilon = \chi^2_{k-1, 1-\alpha}$$

Effettuo l'esperimento: ottengo $y_1 \ldots y_n \in \{t_1 \ldots t_k\}$
Per ogni $j = 1 \ldots k$ pongo $x_j = \#\{i : y_i = t_j\}$
Accetto H_0 se e solo

$$\sum_{j=1}^{k} \frac{(x_j - np_j^\circ)^2}{np_j^\circ} < \chi^2_{k-1, 1-\alpha}$$

$K=2 \qquad t_1 \qquad t_2 \qquad X_1 \qquad X_2 \qquad X_1+X_2=n$

$p_1^\circ + p_2^\circ = 1$

$X_2 = n - X_1$

$p_2^\circ = 1 - p_1^\circ$

$$\frac{(X_1-np_1^\circ)^2}{np_1^\circ} + \frac{(X_2-np_2^\circ)^2}{np_2^\circ} =$$

$$= \frac{(X_1-np_1^\circ)^2}{np_1^\circ} + \frac{(\cancel{n}-X_1-n(\cancel{1}-p_1^\circ))^2}{n(1-p_1^\circ)}$$

$$(X_1-np_1^\circ)^2\left(\frac{1}{np_1^\circ}+\frac{1}{n(1-p_1^\circ)}\right) = (X_1-np_1^\circ)^2\,\frac{1-\cancel{p_1^\circ}+\cancel{p_1^\circ}}{np_1^\circ(1-p_1^\circ)}$$

$$= \left(\frac{X_1-np_1^\circ}{\sqrt{np_1^\circ(1-p_1^\circ)}}\right)^2 = \left(\frac{X_1-E[X_1]}{\sqrt{Var[X_1]}}\right)^2$$

$\forall i=1 \longrightarrow n \qquad Z_i(\omega) = \mathbb{1}_{\{Y_i=t_1\}}(\omega) \qquad \mathbb{P}_{Z_i} = B(p_1^\circ)$ e indipendenti

e $X_1 = \sum_{i=1}^{n} Z_i$

— o —

TEST DI KOLMOGOROV-SMIRNOV

LEMMA Se X è una v.a. con legge F_X, allora la v.a. $Y := F_X \circ X$ ha distribuzione $U([0,1])$.

dim solo nel caso $\mathbb{P}_X$ v.a. distribuzione A.C.

$\mathbb{P}_X = f(x)\,dx$

— o —

X v.a. a valori in $\mathbb{R}$

$\varphi : \mathbb{R} \to \mathbb{R}$ di Borel

$\psi : \mathbb{R} \to [0,+\infty]$ di Borel

$$\mathbb{E}[\psi\circ(\varphi\circ X)] = \int_{\mathbb{R}} \psi(t)\,\mathbb{P}_{\varphi\circ X}(dt) = \int_{\mathbb{R}} (\psi\circ\varphi)(x)\,\mathbb{P}_X(dx) = \mathbb{E}[(\psi\circ\varphi)\circ X]$$

v.a. Y, funzione di Borel non negativa θ

$$\mathbb{E}[\theta \circ Y] = \int_{\mathbb{R}} \theta(y) \mathbb{P}_Y(dy)$$

Scelgo $f = F_X$ e ottengo che $\forall \varphi$ di Borel non negativa

$$\int_{\mathbb{R}} \varphi(t) \mathbb{P}_{F_X \circ X}(dt) = \int_{\mathbb{R}} \varphi(F_X(x)) \mathbb{P}_X(dx) =$$

$$= \int_{\mathbb{R}} \varphi(F_X(x)) f(x) dx = (*)$$

$$y = F_X(x) \qquad dy = F'_X(x) dx = f(x) dx$$

$x \to -\infty \quad y \to 0, \quad x \to +\infty, \quad y \to 1$

$$(*) = \int_0^1 \varphi(y) dy = \int_{\mathbb{R}} \varphi(y) \mathbb{1}_{[0,1]}(y) dy$$

$\varphi(t) = \mathbb{1}_A(t) \qquad A \subset \mathbb{R}$ Boreliano

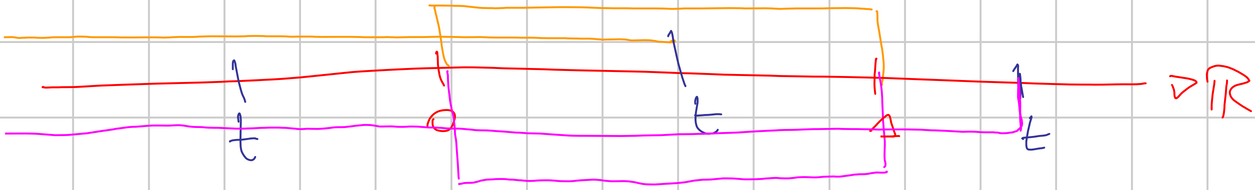
$$\int_{\mathbb{R}} \mathbb{1}_A(t) \mathbb{P}_{F_X \circ X}(dt) = \mathbb{P}(F_X \circ X \in A)$$

$$\forall A \in \mathcal{B}(\mathbb{R}) \qquad \mathbb{P}(F_X \circ X \in A) = \int_{\mathbb{R}} \mathbb{1}_A(y) \mathbb{1}_{[0,1]}(y) dy$$

$A = (-\infty, t]$

$\mathbb{P}(F_X \circ X \leq t) =$ legge di $F_X \circ X$ in t

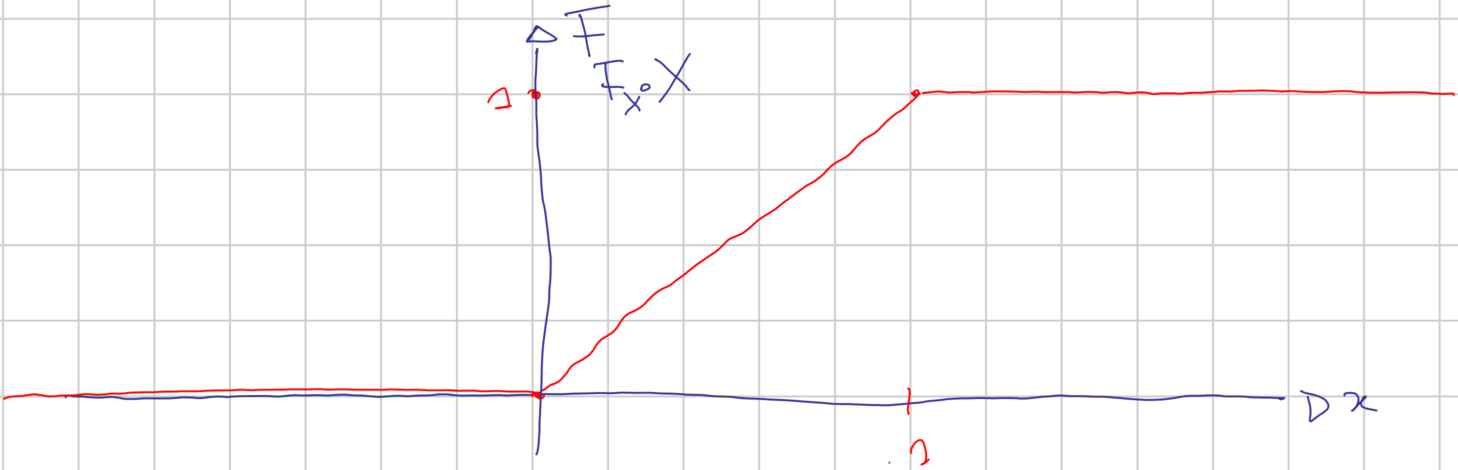
$$\mathbb{P}(F_X \circ X \leq t) = \int_{\mathbb{R}} \mathbb{1}_{(-\infty, t]}(y) \mathbb{1}_{[0,1]}(y) dy$$



Se $t<0$ $\Rightarrow \mathbb{P}(F_X\circ X\le t) = 0$

Se $t\in[0,1)$ $\Rightarrow \mathbb{P}(F_X\circ X\le t) = t$

Se $t\geq 1$ $\Rightarrow \mathbb{P}(F_X\circ X\le t) = 1$



— · —

Sia $X_1 \dots X_n$ campione statistico con legge F

Per $i=1\dots n$ e per ogni $t\in\mathbb{R}$

$$Y_i(\omega,t) = \mathbb{1}_{(-\infty,t]}(X_i(\omega)) = \begin{cases} 1 & X_i(\omega)\le t \\ 0 & X_i(\omega)>t \end{cases}$$

$$p = \mathbb{P}(Y_i(\cdot,t)=1) = \mathbb{P}(X_i\le t) = F(t)$$

$\forall t\in\mathbb{R}$

$Y_1(\cdot,t)\dots Y_n(\cdot,t)$ campione statistico $\mathbb{P}_{Y_i(\cdot,t)} = B(F(t))$

$$\mathbb{E}[Y_i(\cdot,t)] = F(t) \qquad \mathrm{Var}[Y_i(\cdot,t)] = F(t)(1-F(t)) \le \frac{1}{4}$$

$$x(1-x) = -(x^2+x) = -\left(x^2+x+\frac{1}{4}\right)+\frac{1}{4} = \frac{1}{4}-\left(x-\frac{1}{2}\right)^2$$