TEST D'IPOTESI - confronti

Note Title 30/11/2017

Supponiamo di avere due campioni gaussiani

$X_1 \_\_\_ X_n \qquad \mathbb{P}_{X_i} = N(\mu_X, \sigma^2_X)$ Tra di loro

$Y_1 \_\_\_ Y_k \qquad \mathbb{P}_{Y_j} = N(\mu_Y, \sigma^2_Y)$ indipendenti

$H_0 : \mu_X = \mu_Y$

<u>1° CASO</u> σ^2_X e σ^2_Y sono note

$\overline{X} \qquad \mathbb{P}_{\overline{X}} = N\left(\mu_X, \frac{\sigma^2_X}{n}\right)$

$\overline{X}, \overline{Y}$ sono indipendenti

$\overline{Y} \qquad \mathbb{P}_{\overline{Y}} = N\left(\mu_Y, \frac{\sigma^2_Y}{k}\right)$

$\Rightarrow \mathbb{P}_{\overline{X}-\overline{Y}} = N\left(\mu_X - \mu_Y, \frac{\sigma^2_X}{n} + \frac{\sigma^2_Y}{k}\right)$

$H_0: \mu_X - \mu_Y = d$, d assegnato

$x_1 \_\_ x_n$ Accetto H_0 se $|(\bar{x} - \bar{y}) - d| < \varepsilon$

$y_1 \_\_ y_k$

$$\mathbb{P}\left(|\overline{X} - \overline{Y} - d| > \varepsilon \mid \mu_X - \mu_Y = d\right) =$$

$$Z := \frac{(\overline{X} - \overline{Y}) - d}{\sqrt{\frac{\sigma^2_X}{n} + \frac{\sigma^2_Y}{k}}}$$ ha distribuzione $N(0,1)$

$$= \mathbb{P}\left(\frac{|\overline{X} - \overline{Y} - d|}{\sqrt{\frac{\sigma^2_X}{n} + \frac{\sigma^2_Y}{k}}} > \frac{\varepsilon}{\sqrt{\frac{\sigma^2_X}{n} + \frac{\sigma^2_Y}{k}}} \,\middle|\, \mu_X - \mu_Y = d\right)$$

$$\alpha = \mathbb{P}\left(|Z| > \frac{\varepsilon}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{k}}}\right) = \mathbb{P}\left(Z > \frac{\varepsilon}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{k}}}\right) + \mathbb{P}\left(Z < \frac{-\varepsilon}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{k}}}\right)$$

$$= 2\left(1 - \Phi\left(\frac{\varepsilon}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{k}}}\right)\right)$$

$$\Phi\left(\frac{\varepsilon}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{k}}}\right) = 1 - \frac{\alpha}{2} \qquad \frac{\varepsilon}{\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{k}}} = z_{1-\frac{\alpha}{2}}$$

$$\varepsilon = z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{k}}$$

Accetto H_0 se $\left|\bar{x}-\bar{y}-d\right| < z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{k}}$

Rifiuto H_0 se $\left|\bar{x}-\bar{y}-d\right| \geq z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{k}}$

— o —

2° CASO Le varianze σ_x^2 e σ_y^2 sono uguali ma ignote

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \qquad S_y^2 = \frac{1}{k-1}\sum_{i=1}^{k}(Y_i-\bar{Y})^2$$

$$V_x = \frac{(n-1)S_x^2}{\sigma^2} \quad \mathbb{P}_{V_x} = \chi^2_{n-1} \qquad V_y = \frac{(k-1)S_y^2}{\sigma^2} \quad \mathbb{P}_{V_y} = \chi^2_{k-1}$$

dove σ^2 è il comune valore di σ_x^2 e σ_y^2

V_x e V_y sono indipendenti $\Rightarrow \mathbb{P}_{V_x+V_y} = \chi^2_{n+k-2}$

$$V_x + V_y = \frac{(n-1)S_x^2}{\sigma^2} + \frac{(k-1)S_y^2}{\sigma^2} = \frac{n+k-2}{\sigma^2}\underbrace{\frac{(n-1)S_x^2+(k-1)S_y^2}{n+k-2}}_{S^2_{pool}}$$

$$\mathbb{P}_{\bar{X}} = N\left(\mu_x, \frac{\sigma^2}{n}\right) \qquad \mathbb{P}_{\bar{Y}} = N\left(\mu_y, \frac{\sigma^2}{k}\right)$$

$$\mathbb{P}_{\bar{X}-\bar{Y}} = N\left(\mu_x - \mu_y, \frac{\sigma^2}{n} + \frac{\sigma^2}{k}\right) = N\left(\mu_x - \mu_y, \sigma^2\left(\frac{1}{n} + \frac{1}{k}\right)\right)$$

$\bar{X}, S_x^2$ sono indipendenti

$\bar{Y}, S_y^2$ sono indipendenti

$X_1 \ldots X_n$ e $Y_1 \ldots Y_k$ sono campioni indipendenti

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{k}}}$$ ha distribuzione $N(0,1)$

è indipendente da $V_x + V_y$

$$T = \frac{Z\sqrt{n+k-2}}{\sqrt{V_x + V_y}}$$ ha distribuzione $t(n+k-2)$

H_0 è vera ssE $Z = \dfrac{\bar{X} - \bar{Y} - d}{\sigma\sqrt{\frac{1}{n} + \frac{1}{k}}}$ ha distribuzione $N(0,1)$

H_0 è vera ssE $\dfrac{\bar{X} - \bar{Y} - d}{\sigma\sqrt{\frac{1}{n} + \frac{1}{k}}} \dfrac{\sqrt{n+k-2}}{\sqrt{V_x + V_y}}$ ha distribuzione $t(n+k-2)$

Accetto H_0 ssE $\left|\underbrace{\dfrac{\bar{x} - \bar{y} - d}{\sigma\sqrt{\frac{1}{n} + \frac{1}{k}}} \dfrac{\sqrt{n+k-2}}{\sqrt{V_x + V_y}}}\right| < \varepsilon$

$V_x + V_y = \dfrac{n+k-2}{\sigma^2} S^2_{pool}$ $\quad \left|\dfrac{\bar{x} - \bar{y} - d}{\cancel{\sigma}\sqrt{\frac{1}{n} + \frac{1}{k}}} \dfrac{\cancel{\sqrt{n+k-2}}}{\sqrt{\cancel{n+k-2}}} \dfrac{1}{\cancel{\sigma}} S_{pool}\right| < \varepsilon$

$$\alpha = \mathbb{P}\left(\left|\frac{(\bar{X} - \bar{Y} - d)}{\sqrt{\frac{1}{n} + \frac{1}{k}}} \frac{1}{S_{pool}}\right| > \varepsilon \,\middle|\, \mu_x - \mu_y = d\right) = \mathbb{P}(|T_{n+k-2}| > \varepsilon) =$$

$$= \mathbb{P}(T_{n+k-2} > \varepsilon) + \mathbb{P}(T_{n+k-2} < -\varepsilon) = 2(1 - F_T(\varepsilon))$$

$F_T(\varepsilon) = 1 - \frac{\alpha}{2}$ $\qquad \varepsilon = t_{n+k-2,\,1-\frac{\alpha}{2}}$

DISTRIBUZIONE DI FISHER-SNEDECOR A k ED n GRADI DI LIBERTÀ

E' la distribuzione A.C. associata alla densità

$$f(t) = \begin{cases} \dfrac{\Gamma\left(\frac{k+n}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\dfrac{k}{n}\right)^{k/2} \dfrac{t^{\frac{k}{2}-1}}{\left(1+\frac{k}{n}t\right)^{\frac{k+n}{2}}} & t > 0 \\ 0 & t \le 0 \end{cases}$$

Se F è una v.a. con distribuzione di Fisher-Snedecor a k ed n gradi di libertà, allora

$$E[F] = \begin{cases} \dfrac{n}{n-2} & n > 2 \\ +\infty & n = 1,2 \end{cases}$$

$$Var[F] = \begin{cases} \dfrac{2n^2(k+n-2)}{k(n-2)^2(n-4)} & n > 4 \\ +\infty & n = 3,4 \\ \nexists & n = 1,2 \end{cases}$$

$f_{k,n,\alpha}$ è il quantile di F di livello α

Siano U e V v.a. indipendenti

$\mathbb{P}_U = \chi^2_k \qquad \mathbb{P}_V = \chi^2_n$

Definisco $F = \dfrac{U/k}{V/n}$ $\Rightarrow$ allora F ha distribuzione di Fisher-Snedecor con k ed n gradi di libertà

$\mathbb{P}_U = f(u)\,du \qquad \mathbb{P}_V = g(v)\,dv$

$F = \varphi \circ (U,V) \qquad \varphi : (u,v) \in \mathbb{R}^2 \longmapsto \frac{u}{k}\,\frac{n}{v} \in \mathbb{R}$

$\mathbb{P}_{U,V} = f(u)\, g(v)\, du\, dv$

$$f(u) = \begin{cases} \dfrac{1}{\Gamma(k/2)} \left(\dfrac{1}{2}\right)^{k/2} u^{\frac{k}{2}-1} \exp\left(-\dfrac{u}{2}\right) & u > 0 \\ 0 & u \leq 0 \end{cases}$$

$\psi : \mathbb{R} \to \mathbb{R}$ di Borel non negativa

$$\int_{\mathbb{R}} \psi(t)\, \mathbb{P}_{\underset{= \varphi \circ (U,V)}{F}}(dt) = \int_{\mathbb{R}^2} \psi(\varphi(u,v))\, f(u)\, g(v)\, du\, dv =$$

$$= \int_{(0,+\infty)^2} \psi\Big(\underbrace{\frac{u}{k}\,\frac{n}{v}}_{t}\Big) \frac{1}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{2}\right)^{k/2} \left(\frac{1}{2}\right)^{n/2} u^{k/2-1}\, v^{n/2-1} \cdot \exp\left(-\frac{u}{2}\right) \exp\left(-\frac{v}{2}\right) du\, dv$$

$t = \frac{u}{k}\,\frac{n}{v} \qquad u = \frac{kv}{n}\, t$

$$\int_0^{+\infty} \psi(t) \frac{t^{k/2-1}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{2}\right)^{\frac{k+n}{2}} \left(\frac{k}{n}\right)^{k/2} \cdot \left(\int_0^{+\infty} v^{\frac{n+k}{2}-1} \exp\left(-\frac{1}{2} v \left(1 + \frac{kt}{n}\right)\right) dv \right) dt$$

$y = \frac{v}{2}\left(1 + \frac{kt}{n}\right) \qquad v = \frac{2ny}{n+kt}$

$$\int_0^{+\infty} \psi(t) \frac{\Gamma\left(\frac{n+k}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{n}{2}\right)}\, t^{k/2-1} \left(\frac{n}{n+kt}\right)^{\frac{n+k}{2}} dt$$

TEST D'IPOTESI SULL'UGUAGLIANZA DELLE VARIANZE

$X_1 \text{ — } X_k \qquad \mathbb{P}_{X_i} = N(\mu_x, \sigma_x^2)$ campioni

$Y_1 \text{ — } Y_n \qquad \mathbb{P}_{Y_j} = N(\mu_y, \sigma_y^2)$ indipendenti.

$H_0 : \sigma_x^2 = \sigma_y^2 \qquad H_A : \sigma_x^2 \neq \sigma_y^2$

$\dfrac{(k-1) S_x^2}{\sigma_x^2}$ ha distribuzione χ^2_{k-1}

$\dfrac{(n-1) S_y^2}{\sigma_y^2}$ ha distribuzione χ^2_{n-1}

$$F = \frac{\frac{\cancel{(k-1)} S_x^2}{\sigma_x^2} \cdot \cancel{\frac{1}{k-1}}}{\frac{\cancel{(n-1)} S_y^2}{\sigma_y^2} \cdot \cancel{\frac{1}{n-1}}}$$

ha distribuzione di Fisher-Snedecor con $k-1$ e $n-1$ gradi di libertà

$x_1 \text{ — } x_k$ accetto H_0 se $\left| \dfrac{s_x^2}{s_y^2} - 1 \right| < \varepsilon$

$y_1 \text{ — } y_n$

$$\alpha = \mathbb{P}\left(\left| \frac{S_x^2}{S_y^2} - 1 \right| > \varepsilon \;\middle|\; \sigma_x^2 = \sigma_y^2 \right) =$$

$$= \mathbb{P}\left(\frac{S_x^2}{S_y^2} > 1 + \varepsilon \;\middle|\; \sigma_x^2 = \sigma_y^2 \right) + \mathbb{P}\left(\frac{S_x^2}{S_y^2} < 1 - \varepsilon \;\middle|\; \sigma_x^2 = \sigma_y^2 \right)$$

$$= \mathbb{P}\left(F_{k-1, n-1} > 1 + \varepsilon \right) + \mathbb{P}\left(F_{k-1,n-1} < 1 - \varepsilon \right)$$

Accetto H_0 se $1 - \varepsilon_1 < \dfrac{s_x^2}{s_y^2} < 1 + \varepsilon_2$ con ε_1 e ε_2 da scegliere

$$\alpha = \mathbb{P}\left(\frac{S_X^2}{S_Y^2} > 1 + \varepsilon_2\right)\Bigg|_{\sigma_X^2 = \sigma_Y^2} + \mathbb{P}\left(\frac{S_X^2}{S_Y^2} < 1 - \varepsilon_1 \,\Big|\, \sigma_X^2 = \sigma_Y^2\right)$$

$$= \mathbb{P}\left(F_{k-1,n-1} > 1 + \varepsilon_2\right) + \mathbb{P}\left(F_{k-1,n-1} < 1 - \varepsilon_1\right)$$

$$\mathbb{P}\left(F_{k-1,n-1} > 1 + \varepsilon_2\right) = \frac{\alpha}{2} \qquad \underbrace{\mathbb{P}\left(F_{k-1,n-1} < 1 + \varepsilon_2\right)} = 1 - \frac{\alpha}{2}$$

$$\mathbb{P}\left(F_{k-1,n-1} < 1 - \varepsilon_1\right) = \frac{\alpha}{2}$$

$$1 + \varepsilon_2 = f_{k-1,n-1,1-\frac{\alpha}{2}}$$

$$1 - \varepsilon_1 = f_{k-1,n-1,\frac{\alpha}{2}}$$

$$\underbrace{f_{k-1,n-1,\frac{\alpha}{2}}}_{\frac{1}{f_{n-1,k-1,1-\frac{\alpha}{2}}}} < \frac{S_X^2}{S_Y^2} < f_{k-1,n-1,1-\frac{\alpha}{2}}$$ accetto H_0

Altrimenti rifiuto H_0

$U \qquad \mathbb{P}_U = \chi^2_{k-1}$

$V \qquad \mathbb{P}_V = \chi^2_{n-1}$

$$F = \frac{U/_{k-1}}{V/_{n-1}}$$ ha distribuzione di Fisher con $k-1$ ed $n-1$ gradi di libertà

Sia $f = f_{k-1,n-1,\frac{\alpha}{2}}$

$$\frac{\alpha}{2} = \mathbb{P}\left(F_{k-1,n-1} \leq f\right) = \mathbb{P}\left(\frac{U/_{k-1}}{V/_{n-1}} \leq f\right) =$$

$$= \mathbb{P}\left(\frac{V/_{n-1}}{U/_{k-1}} \geq \frac{1}{f}\right) = \mathbb{P}\left(F_{n-1,k-1} \geq \frac{1}{f}\right)$$

$$\mathbb{P}\left(F_{n-1,k-1} \leq \frac{1}{f}\right) = 1 - \frac{\alpha}{2}$$

$$\frac{1}{F_{k-1,\,n-1,\,\frac{\alpha}{2}}} = F_{n-1,\,k-1,\,1-\frac{\alpha}{2}}$$

— o —

$X_1 \,\_\_\_\, X_n \qquad t_1 \,\_\_\_\, t_k$

$\forall j = 1 \,\_\_\, k \qquad \mathbb{P}(X_j = t_j) = p_j \qquad \sum_{j=1}^{k} p_j = 1 \qquad p_j \geq 0$

$x_1 \,\_\_\_\, x_n$

n_1 volte il valore t_1

n_2 volte il valore t_2

⋮

n_k volte il valore t_k

$(X_1 \,\_\_\_\, X_n)$ valutate su $x_1 \,\_\_\, x_n$

$$\mathbb{P}(X_1 = x_1)\,\mathbb{P}(X_2 = x_2) \cdots \mathbb{P}(X_n = x_n) = p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

$$f(x_1 \,\_\, x_n \mid p_1 \,\_\, p_k)$$

$$g(x_1 \,\_\, x_n \mid p_1 \,\_\, p_k) = \log f(x_1 \,\_\, x_n \mid p_1 \,\_\, p_k) = \log\left(p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}\right) = \sum_{j=1}^{k} n_j \log p_j$$

$$\max\left\{ g(x_1 \,\_\, x_n \mid p_1 \,\_\, p_k) \;\middle|\; \sum_{j=1}^{k} p_j = 1 \right\}$$

$$F(p_1 \,\_\, p_k, \lambda) = \sum_{j=1}^{k} n_j \log p_j - \lambda\left(\sum_{j=1}^{k} p_j - 1\right)$$

$$\frac{\partial F}{\partial p_j} = \frac{n_j}{p_j} - \lambda = 0 \qquad \frac{n_j}{p_j} = \lambda \qquad p_j = \frac{n_j}{\lambda}$$

$$\frac{\partial F}{\partial \lambda} = -\left(\sum_{j=1}^{k} p_j - 1\right) = 0 \qquad 1 = \sum_{j=1}^{k} p_j = \sum_{j=1}^{k} \frac{n_j}{\lambda} = \frac{1}{\lambda}\, n$$

$\Rightarrow \lambda = n \quad \Rightarrow \quad p_j = \dfrac{n_j}{n}$ frequenza relativa

$Y_1 \ldots Y_n$ campione statistico distribuito sui valori:

$t_1 \ldots t_k$

$p_1 \ldots p_k \qquad p_j > 0 \qquad \sum_{j=1}^{k} p_j = 1$

$H_0 : \mathbb{P}(Y_i = t_j) = p_j \qquad \forall j = 1 \ldots k$

$H_A : \exists\, j \in \{1 \ldots k\} : \mathbb{P}(Y_1 = t_j) \neq p_j$

Per $j = 1 \ldots k \qquad X_j := \#\{ i \in \{1 \ldots n\} : Y_i = t_j \}$

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

$$\Rightarrow \mathbb{E}[X_j] = np_j$$

$\sum_{j=1}^{k} a_j (x_j - np_j)^2 \qquad a_j > 0 \qquad a_j = \frac{1}{np_j}$

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

per $n \to \infty$, distribuzione di T tende χ^2_{k-1}

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

$\alpha = \mathbb{P}(T \geq \varepsilon \mid H_0) = 1 - \mathbb{P}(T \leq \varepsilon \mid H_0) \Leftarrow ?$

$\mathbb{P}(T \leq \varepsilon \mid H_0) = 1 - \alpha \qquad \varepsilon \sim \chi^2_{k-1,1-\alpha}$

Accetto H_0 se $\sum_{j=1}^{k} \frac{(x_j - np_j)^2}{np_j} < \chi^2_{k-1,1-\alpha}$

Rifiuto H_0 altrimenti

$t_1 \quad t_2 \qquad \mathbb{P}(Y_i = t_1) = p_1 \qquad p_1, p_2 \geq 0$

$\mathbb{P}(Y_i = t_2) = p_2 \qquad p_1 + p_2 = 1 \qquad p_2 = 1 - p_1$

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

$$X_2 = n - X_1$$

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

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

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

$\{X_i\}_{i=1}^{\infty}$ v.a. i.i.d. e sia F_0 la loro comune legge

$$t \in \mathbb{R} \qquad Y_i(\omega, t) = \begin{cases} 1 & X_i(\omega) \leq t \\ 0 & X_i(\omega) > t \end{cases}$$

$$Y_i(\omega, t) = \mathbb{1}_{(-\infty, t]}(X_i(\omega))$$

$$\mathbb{E}[Y_i(\cdot, t)] = \mathbb{P}(X_i \leq t) = F_0(t)$$

$$\mathbb{E}[Y_i^2(\cdot, t)] = \mathbb{P}(X_i \leq t) = F_0(t)$$

$$Var[Y_i(\cdot, t)] = F_0(t) - F_0^2(t) \leq F_0(t) \leq 1$$

$$G_n(\omega, t) = \frac{1}{n} \sum_{i=1}^{n} Y_i(\omega, t) \qquad S_n = \sum X_i$$

$$\frac{S_n}{n}$$

$$\mathbb{P}\left(\left|G_n(\cdot,t) - F_0(t)\right| > \varepsilon\right) \le \frac{1}{n\varepsilon^2}$$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|G_n(\cdot,t) - F_0(t)\right| > \varepsilon\right) = 0$$ uniformemente rispetto a t

$G_n(\cdot,t)$ funzione di ripartizione empirica.

$$G_n(\omega,t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,t]}(X_i(\omega)) =$$

$$= \frac{1}{n} \# \{ i \in \{1 \_ n\} : X_i(\omega) \le t \}$$

$G_n(\omega,t)$ prende i valori $0, \frac{1}{n}, \frac{2}{n}, \_\_ \frac{n-1}{n}, 1$

$G_n(\omega,\cdot)$ è una funzione monotona crescente e costante a tratti

$X_1 \_\_\_ X_n$ campione statistico

$x_1 \_\_\_ x_n$ dati rilevati - Sia F_0 una legge assegnata

H_0 : $F_0(t)$ è legge del campione

$$\mathbb{P}(X_i \le t) = F_0(t) \qquad \forall t \in \mathbb{R}$$

H_A : $\exists t \in \mathbb{R}$: $\mathbb{P}(X_i \le t) \ne F_0(t)$

$$G_n(\omega,t) = \frac{1}{n} \# \{ i \in \{1 \_ n\} : X_i(\omega) \le t \}$$

$$g_n(x_1 \_ x_n, t) = \frac{1}{n} \# \{ i \in \{1 \_ n\} : x_i \le t \}$$

$$\sup_{t\in\mathbb{R}} \left| g_n(x_1 \_ x_n, t) - F_0(t) \right|$$

$$D_n^{(\omega)} = \sup_{t\in\mathbb{R}} \left| G_n(\omega,t) - F_0(t) \right| \qquad G_n^{(\cdot,t)} = g_n(X_1 \_ X_n)(\cdot,t)$$

Accetto H_0 se $\sup_{t\in\mathbb{R}} |g_n(x_1 - x_n, t) - F_0(t)| < \varepsilon$

Rifiuto H_0 se $\sup_{t\in\mathbb{R}} |g_n(x_1 - x_n, t) - F_0(t)| \geq \varepsilon$

e devo usare D_n per scegliere ε in base al livello di significatività α prescelto.

— o —

$$G_n(\omega, t) = \frac{1}{n} \sum_{i=1}^{n} 1\!\!1_{(-\infty, t]}(X_i(\omega))$$

$$D_n(\omega) = \sup_{t\in\mathbb{R}} |G_n(\omega, t) - F_0(t)|$$

$$\mathbb{P}(D_n \geq d) = \mathbb{P}\left(\sup_{t\in\mathbb{R}} \left| \frac{\#\{i : X_i \leq t\}}{n} - F_0(t) \right| \geq d \right) = (*)$$

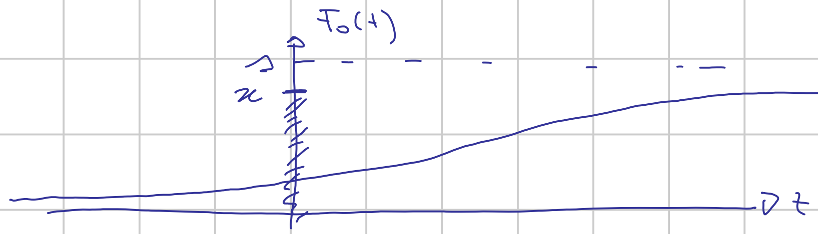
Se F_0 fosse strettamente monotona crescente

$$\#\{i : X_i \leq t\} = \#\{i : F(X_i) \leq F_0(t)\}$$

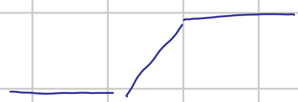
$$(*)\ \mathbb{P}\left(\sup_{t\in\mathbb{R}} \left| \frac{\#\{i : F_0(X_i) \leq F_0(t)\}}{n} - F_0(t) \right| \geq d \right)$$

Supponiamo che F_0 sia continua

$Z_i = F_0 \circ X_i$ è distribuito su $[0,1)$



$\mathbb{P}(Z_i \leq x) = x$ $\qquad \mathbb{P}_{Z_i} = U([0,1])$



$$\mathbb{P}(D_n \geq d) = \mathbb{P}\left(\sup_{t \in \mathbb{R}} \left| \frac{\#\{i : U_i \leq F_0(t)\}}{n} - F_0(t) \right| \geq d \right)$$

$U_i := F_0 \circ X_i$ campione uniformemente distribuito su $(0,1)$

$$= \mathbb{P}\left(\sup_{y \in (0,1)} \left| \frac{\#\{i : U_i \leq y\}}{n} - y \right| \geq d \right)$$

$$\sup_{t \in \mathbb{R}} \left| \frac{\#\{i : x_i \leq F_0(t)\}}{n} - F_0(t) \right|$$

$$x_1 \leq x_2 \leq \dots \leq x_n$$

$$\frac{\#\{i : x_i \leq F_0(t)\}}{n} - F_0(t)$$

$t < x_1$ $\qquad |-F_0(t)| = F_0(t) \leq F_0(x_1)$

$x_1 \leq t < x_2$ $\qquad \frac{1}{n} - F_0(t) \qquad F_0(x_1) \leq F_0(t) \leq F_0(x_2)$

$$\frac{1}{n} - F_0(\underset{2}{x_2}) \leq \frac{1}{n} - F_0(t) \leq \frac{1}{n} - F_0(\underset{1}{x_1})$$

$$\sup_{t < x_1} \left| \frac{\#\{i : x_i \leq F_0(t)\}}{n} - F_0(t) \right| = F_0(x_1)$$

$$\sup_{x_1 \leq t < x_2} \left| \frac{\#\{i : x_i \leq F_0(t)\}}{n} - F_0(t) \right| = \max\left\{ \left| \frac{1}{n} - F_0(x_2) \right|, \left| \frac{1}{n} - F_0(x_1) \right| \right\}$$

$$\vdots$$

$$\sup_{x_{n-1} \leq t < x_n} \left| \frac{\#\{i : x_i \leq F_0(t)\}}{n} - F_0(t) \right| = \max\left\{ \left| \frac{n-1}{n} - F_0(x_n) \right|, \left| \frac{n-1}{n} - F_0(x_{n-1}) \right| \right\}$$

$$\sup_{t \geq x_n} \left| \frac{\#\{i : x_i \leq F_0(t)\}}{n} - F_0(t) \right| = 1 - F_0(x_n)$$

$$\mathbb{P}\left(\sqrt{n}\, D_n \leq t\right)$$