martedì 10 novembre 2020 16:01

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$v=(x_1-\bar{x},\, x_2-\bar{x},\, \_\_\_\_ \, x_n-\bar{x})$$
$$w=(y_1-\bar{y},\, y_2-\bar{y},\, \ldots\ldots,\, y_n-\bar{y})$$

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \langle v, w\rangle$$

$$\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\ \sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2} = \|v\|_2 \cdot \|w\|_2$$

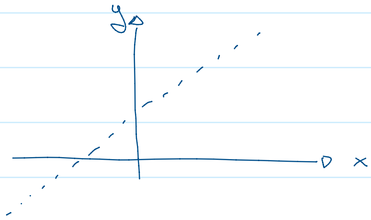
$$|\langle v,w\rangle| \le \|v\|_2 \cdot \|w\|_2$$

$$\frac{|\langle v,w\rangle|}{\|v\|\cdot\|w\|} \le 1 \qquad |\rho(x,y)| \le 1$$

$$\rho(x,y)=1 \iff \langle v,w\rangle = \|v\|\cdot\|w\| \iff v=\lambda w \quad \lambda>0$$

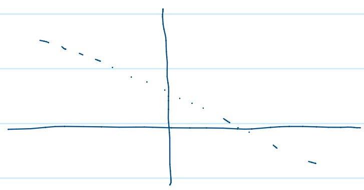
cioè $\exists \lambda>0$ t.c. $(x_1-\bar{x}, x_2-\bar{x} \ldots x_n-\bar{x}) = \lambda(y_1-\bar{y}, \ldots y_n-\bar{y})$

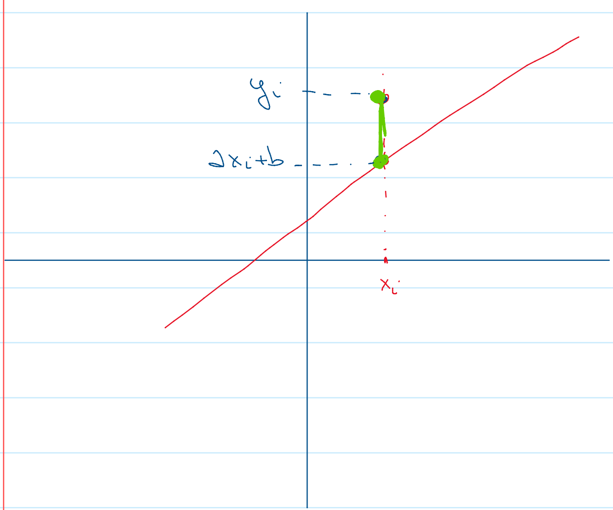
$$x_i-\bar{x} = \lambda(y_i-\bar{y}) \quad \forall i=1 \_\_ n$$
$$y_i = \bar{y} + \frac{1}{\lambda}(x_i-\bar{x}) \quad \forall i=1 \_\_ n$$
$$y = \bar{y} + \frac{1}{\lambda}(x-\bar{x})$$



$$\rho(x,y)=-1 \iff \langle v,w\rangle = -\|v\|\cdot\|w\| \iff$$

$\exists \mu<0$ t.c. $v=\mu w$





$y=ax+b$

(x_i, ax_i+b) $\qquad$ (x_i, y_i)

distanza $= |y_i-(ax_i+b)|$

$$S(a,b) = \sum_{i=1}^{n}(y_i-(ax_i+b))^2$$

$$2\sum_{i=1}^{n}(\bar{y}-a\bar{x}-b)(y_i-\bar{y}) = 2(\bar{y}-a\bar{x}-b)\sum_{i=1}^{n}(y_i-\bar{y}) =$$

$$2\sum_{i=1}^{n}\underbrace{(\bar{y}-a\bar{x}-b)}(y_i-\bar{y}) = 2(\bar{y}-a\bar{x}-b)\sum_{i=1}^{n}(y_i-\bar{y}) =$$

$$= 2(\bar{y}-a\bar{x}-b)\Big(\underbrace{\sum_{i=1}^{n}y_i}_{n\bar{y}} - \underbrace{\sum_{i=1}^{n}\bar{y}}_{n\bar{y}}\Big) \qquad \bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$= 2(\bar{y}-a\bar{x}-b)(n\bar{y}-n\bar{y}) = 0$$

$$2a\sum_{i=1}^{n}(\bar{y}-a\bar{x}-b)(x_i-\bar{x}) = 2a(\bar{y}-a\bar{x}-b)\sum_{i=1}^{n}(x_i-\bar{x}) =$$

$$= 2a(\bar{y}-a\bar{x}-b)\Big(\underbrace{\sum_{i=1}^{n}x_i}_{n\bar{x}} - \underbrace{\sum_{i=1}^{n}\bar{x}}_{n\bar{x}}\Big) = 2a(\bar{y}-a\bar{x}-b)(n\bar{x}-n\bar{x}) = 0$$

STATISTICA INFERENZIALE

$x = (x_1 \ldots x_n)$

$\exists (\Omega, \mathcal{E}, \mathbb{P})$ $\quad X_1, X_2, \ldots X_n : \Omega \to \mathbb{R}$ v.a. t.c.

- le $X_1 \ldots X_n$ sono i.i.d.
- $\exists \omega \in \Omega$: $\quad x_1 = X_1(\omega)$

$$x_2 = X_2(\omega)$$
$$\vdots$$
$$x_n = X_n(\omega)$$

cioè $\exists \omega \in \Omega$: $x_i = X_i(\omega)$ $\forall i = 1 \ldots n$

DEF CAMPIONE STATISTICO

Una famiglia finita di v.a. $X_1 \ldots X_n$ si dice un CAMPIONE STATISTICO se le v.a. $X_1 \ldots X_n$, definite sullo stesso spazio probabilizzato $(\Omega, \mathcal{E}, \mathbb{P})$ sono indipendenti e identicamente distribuite.
La loro comune distribuzione $\mathbb{P}_{X_i}$, $i = 1 \ldots n$, si dice DISTRIBUZIONE CAMPIONARIA.

Sia $X_1 \ldots X_n$ un campione statistico e sia $f : \mathbb{R}^n \to \mathbb{R}$ Borel-misurabile. Allora la v.a. $Y = f(X_1 \ldots X_n)$ si dice una STATISTICA DEL CAMPIONE

Sia $(X_i)_{i \in \mathbb{N}}$ una successione di v.a. i.i.d.
$\forall n \in \mathbb{N}$ $\quad X_1, X_2 \ldots X_n$ è un campione statistico di cardinalità n
Sia $f_n : \mathbb{R}^n \to \mathbb{R}$ Borel misurabile $\quad \forall n \in \mathbb{N}$
$Y_n := f_n(X_1 \ldots X_n)$

Sia $f_n : \mathbb{R}^n \to \mathbb{R}$ Borel misurabile $\forall n \in \mathbb{N}$
$Y_n := f_n(X_1 \ldots X_n)$
Se Y_n CONVERGE IN PROBABILITÀ AD UNA COSTANTE λ
dico che Y_n è uno STIMATORE CONSISTENTE IN SENSO DEBOLE DI λ

$$\lim_{n \to \infty} \mathbb{P}\left(|Y_n - \lambda| > \delta\right) = 0 \qquad \forall \delta > 0$$

Se Y_n CONVERGE $\mathbb{P}$-QUASI CERTAMENTE AD UNA COSTANTE λ
dico che Y_n è uno STIMATORE CONSISTENTE IN SENSO FORTE DI λ

Se Y_n è uno stimatore consistente di λ e se $\mathbb{E}[Y_n] = \lambda$ $\forall n \in \mathbb{N}$,
dico che Y_n è uno STIMATORE CORRETTO (O NON DISTORTO) di λ

DEF MEDIA CAMPIONARIA E VARIANZA CAMPIONARIA
Sia $X_1 \ldots X_n$ campione statistico
Chiamo media campionaria la statistica $\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$

$$\overline{X}(\omega) := \frac{1}{n} \sum_{i=1}^{n} X_i(\omega) \qquad \forall \omega \in \Omega$$

Chiamo varianza campionaria la statistica $S^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})$

PROPRIETÀ DI MEDIA CAMPIONARIA E VARIANZA CAMPIONARIA
Sia $X_1 \ldots X_n$ campione statistico di cardinalità n, t.c. il valore atteso e la varianza della distribuzione campionaria esistono e sono finiti. Indico μ il valore atteso e σ^2 la varianza.
Allora

$$\mathbb{E}[\overline{X}] = \mu \qquad \mathrm{Var}[\overline{X}] = \frac{\sigma^2}{n} \qquad \mathbb{E}[S^2] = \sigma^2$$

DIM $$\mathbb{E}[\overline{X}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\mathbb{E}[X_i]}_{\mu \ \forall i} = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{\cancel{n}}\cancel{n}\mu = \mu$$

$$\mathrm{Var}[\overline{X}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] =$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} \mathrm{Var}[X_i] + 2\sum_{1 \le i < j \le n} \underbrace{\mathrm{Cov}(X_i, X_j)}_{=0 \quad i \neq j}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\mathrm{Var}[X_i]}_{\sigma^2}$$

$$= \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$$

$$\mathbb{E}[S^2] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = \frac{1}{n-1}\mathbb{E}\left[\underbrace{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right] \leftarrow$$

$$\mathbb{E}[S^2] = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right] = \frac{1}{n-1}\mathbb{E}\Big[\underbrace{\sum_{i=1}^{n}(X_i-\overline{X})^2}\Big] \leftarrow$$

$$\rightarrow \sum_{i=1}^{n}(X_i-\overline{X})^2 = \sum_{i=1}^{n}\left(X_i^2-2\overline{X}X_i+\overline{X}^2\right) = \sum_{i=1}^{n}X_i^2 - 2\overline{X}\underbrace{\sum_{i=1}^{n}X_i}_{n\overline{X}} + \sum_{i=1}^{n}\overline{X}^2$$

$$= \sum_{i=1}^{n}X_i^2 - 2\overline{X}\cdot n\overline{X} + n\overline{X}^2 = \sum_{i=1}^{n}X_i^2 - n\overline{X}^2 \leftarrow$$

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\right)$$

$$\mathbb{E}[S^2] = \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n}X_i^2 - n\overline{X}^2\right] =$$

$$= \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n}(X_i-\mu+\mu)^2 - n(\overline{X}-\mu+\mu)^2\right] =$$

$$= \frac{1}{n-1}\left(\mathbb{E}\left[\sum_{i=1}^{n}(X_i-\mu+\mu)^2\right] - n\,\mathbb{E}\left[(\overline{X}-\mu+\mu)^2\right]\right) =$$

$$= \frac{1}{n-1}\left(\mathbb{E}\left[\sum_{i=1}^{n}\left((X_i-\mu)^2+2\mu(X_i-\mu)+\mu^2\right)\right] - n\,\mathbb{E}\left[(\overline{X}-\mu)^2+2\mu(\overline{X}-\mu)+\mu^2\right]\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}\left[(X_i-\mu)^2+2\mu(X_i-\mu)+\mu^2\right] - n\,\mathbb{E}\left[(\overline{X}-\mu)^2+2\mu(\overline{X}-\mu)+\mu^2\right]\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}\left(\underbrace{\mathbb{E}[(X_i-\mu)^2]}_{Var[X_i]=\sigma^2} + \underbrace{\mathbb{E}[2\mu(X_i-\mu)]}_{=0} + \mathbb{E}[\mu^2]\right) - n\left(\underbrace{\mathbb{E}[(\overline{X}-\mu)^2]}_{\sigma^2/n} + \underbrace{\mathbb{E}[2\mu(\overline{X}-\mu)]}_{} + \underbrace{\mathbb{E}[\mu^2]}_{}\right)\right)$$

$$\mathbb{E}[(X_i-\underset{\substack{\| \\ \mathbb{E}[X_i]}}{\mu})^2] = Var[X_i]$$

$$\mathbb{E}[2\mu(X_i-\mu)] = 2\mu\,\mathbb{E}[X_i-\mu] = 2\mu(\underbrace{\mathbb{E}[X_i]}_{\substack{\| \\ \mu}}-\mu) = 0$$

$$\mathbb{E}[(\overline{X}-\underset{\substack{\| \\ \mathbb{E}[\overline{X}]}}{\mu})^2] = Var[\overline{X}] = \frac{\sigma^2}{n}$$

$$\mathbb{E}[2\mu(\overline{X}-\mu)] = 2\mu\,\mathbb{E}[\overline{X}-\mu] = 2\mu(\mathbb{E}[\overline{X}]-\mu) = 2\mu(\mu-\mu) = 0$$

$$\mathbb{E}[S^2] = \frac{1}{n-1}\left(\underbrace{\sum_{i=1}^{n}(\sigma^2+\mu^2)} - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right) =$$

$$= \frac{1}{n-1}\left(n(\sigma^2+\cancel{\mu^2}) - \sigma^2 - \cancel{n\mu^2}\right) = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

OSSERVAZIONE Sia X v.a. con distribuzione A.C. e densità $f(x)$, $\alpha, \beta \in \mathbb{R}$ $\alpha \neq 0$. Considero $Y := \alpha X + \beta$

$F_Y(t) = \mathbb{P}(Y \le t) = \mathbb{P}(\alpha X + \beta \le t)$ ⇐

① $\alpha > 0$ $\quad F_Y(t) = \mathbb{P}\left(X \le \frac{t-\beta}{\alpha}\right) = \int_{-\infty}^{\frac{t-\beta}{\alpha}} f(x)dx =$ $\qquad x = \frac{y-\beta}{\alpha}$, $\quad dx = \frac{1}{\alpha}dy$

$= \int_{-\infty}^{t} f\left(\frac{y-\beta}{\alpha}\right)\frac{1}{\alpha}dy$ $\qquad y = \beta + \alpha x$, $\quad x \to -\infty \quad y \to -\infty$, $\quad x = \frac{t-\beta}{\alpha}$, $\quad y = \beta + \cancel{\alpha}\frac{t-\beta}{\cancel{\alpha}} = t$

$F_Y(t) = \int_{-\infty}^{t} \boxed{\frac{1}{\alpha} f\left(\frac{y-\beta}{\alpha}\right)} dy$

è la densità di Y

$\mathbb{P}_Y = g(y)dy \qquad g(y) = \frac{1}{|\alpha|} f\left(\frac{y-\beta}{\alpha}\right) \qquad \forall y \in \mathbb{R}$

② $\alpha < 0$ $\quad F_Y(t) = \mathbb{P}\left(X \ge \frac{t-\beta}{\alpha}\right) = \int_{\frac{t-\beta}{\alpha}}^{+\infty} f(x)dx$ $\qquad x = \frac{y-\beta}{\alpha}$, $\quad dx = \frac{1}{\alpha}dy$

$= \int_{t}^{-\infty} f\left(\frac{y-\beta}{\alpha}\right)\frac{1}{\alpha}dy =$ $\qquad y = \beta + \alpha x$, $\quad x = \frac{t-\beta}{\alpha} \quad y = \beta + \cancel{\alpha}\frac{t-\beta}{\cancel{\alpha}} = t$

$\alpha = -|\alpha| \qquad \int_{t}^{-\infty} \frac{-1}{|\alpha|} f\left(\frac{y-\beta}{\alpha}\right) dy =$ $\qquad x \to +\infty \quad y \to -\infty$

$F_Y(t) = \int_{-\infty}^{t} \boxed{\frac{1}{|\alpha|} f\left(\frac{y-\beta}{\alpha}\right)} dy \quad \Rightarrow \mathbb{P}_Y = g(y)dy$ con $g(y) = \frac{1}{|\alpha|} f\left(\frac{y-\beta}{\alpha}\right)$

⇒ In entrambi i casi $\mathbb{P}_Y = g(y)dy$

e $g(y) = \frac{1}{|\alpha|} f\left(\frac{y-\beta}{\alpha}\right)$

— o —

Sia X v.a. con distribuzione A.C. e densità $f(x)$

Consideriamo $Y := X^2$

$$F_Y(t) = \mathbb{P}(X^2 \le t) = \begin{cases} 0 & t < 0 \\ \mathbb{P}(X = 0) = 0 & t = 0 \\ \mathbb{P}(-\sqrt{t} \le X \le \sqrt{t}) & t > 0 \end{cases} = \begin{cases} 0 & t \le 0 \\ \mathbb{P}(-\sqrt{t} \le X \le \sqrt{t}) & t > 0 \end{cases}$$

$t > 0$

$$F_Y(t) = \mathbb{P}(-\sqrt{t} \le X \le \sqrt{t}) = \int_{-\sqrt{t}}^{\sqrt{t}} f(x)dx = \int_{-\sqrt{t}}^{0} f(x)dx + \int_{0}^{\sqrt{t}} f(x)dx$$

$\int_{-\sqrt{t}}^{0} f(x)dx$ $\qquad -\sqrt{t} \quad 0 \quad \mathbb{R}$ $\qquad x = -\sqrt{y} = -y^{1/2} = -\sqrt{y}$, $\quad dx = -\frac{1}{2}y^{-1/2}dy = \frac{-1}{2\sqrt{y}}dy$

$= \int f(-\sqrt{y})\frac{-1}{2\sqrt{y}}dy = \int \frac{1}{2\sqrt{y}} f(-\sqrt{y})dy$ $\qquad x = -\sqrt{y} \qquad x^2 = y$

$\qquad\quad {}^{0} \qquad\qquad\qquad\qquad\qquad {}^{t}$

$$= \int_t^{} f(-\sqrt{y}) \cdot \frac{-1}{2\sqrt{y}} dy = \int_0^{} \frac{1}{2\sqrt{y}} f(-\sqrt{y}) dy$$

$x = -\sqrt{y}$ $\quad x^2 = y$

$x = -\sqrt{t} \to y = t$

$x = 0 \to y = 0$

$$\int_0^{\sqrt{t}} f(x) dx$$

$x = \sqrt{y} = y^{1/2}$

$dx = \frac{1}{2} y^{-1/2} dy$

$$\int_0^{t} f(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} dy$$

$y = x^2$

$x = 0 \quad y = 0$

$x = \sqrt{t} \quad y = t$

$$t > 0 \Rightarrow F_Y(t) = \int_0^{\sqrt{t}} \frac{1}{2\sqrt{y}} f(-\sqrt{y}) dy + \int_0^{\sqrt{t}} f(\sqrt{y}) \frac{1}{2\sqrt{y}} dy =$$

$$= \int_0^{\sqrt{t}} \frac{1}{2\sqrt{y}} \left(f(\sqrt{y}) + f(-\sqrt{y}) \right) dy = \int_{-\infty}^{t} g(y) dy \qquad t > 0$$

$$\boxed{g(y) = \begin{cases} 0 & y \leq 0 \\ \frac{1}{2\sqrt{y}} \left(f(\sqrt{y}) + f(-\sqrt{y}) \right) & y > 0 \end{cases}}$$

è la densità di $Y = X^2$

$$t \leq 0 \quad F_Y(t) = 0 = \int_{-\infty}^{t} 0 \, dy = \int_{-\infty}^{t} g(y) dy \qquad t \leq 0$$

$$F_Y(t) = \int_{-\infty}^{t} g(y) dy$$

X si dice una v.a. con distribuzione gaussiana di parametri $\mu \in \mathbb{R}$ e $\sigma^2 > 0$ se X ha distribuzione A.C. con densità

o distribuzione normale

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2} \right) \qquad \forall x \in \mathbb{R}$$

Si scrive $\mathbb{P}_X = N(\mu, \sigma^2)$

N.B. $f(x) > 0 \quad \forall x \in \mathbb{R}$

$$\int_{\mathbb{R}} f(x) dx = 1$$

$$\int_{\mathbb{R}} f(x) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \underbrace{\exp\left(\frac{-(x-\mu)^2}{2\sigma^2} \right)}_{\exp(-y^2)} dx$$

$y = \frac{x-\mu}{\sigma\sqrt{2}}$

$x = \mu + \sigma y \sqrt{2}$

$dx = \sigma\sqrt{2}\, dy$

$$= \frac{1}{\sqrt{2\pi\cancel{\sigma^2}}} \int_{\mathbb{R}} e^{-y^2} \cdot \cancel{\sigma\sqrt{2}} \, dy = \frac{1}{\sqrt{\pi}} \underbrace{\int_{\mathbb{R}} e^{-y^2} dy}$$

$x \to -\infty \quad y \to -\infty$

$x \to +\infty \quad y \to +\infty$

$$\frac{1}{\sqrt{\cancel{2}\pi\cancel{\sigma^2}}}\int_{\mathbb{R}} e^{-y^2}\cdot \cancel{\sqrt{2}}\,\cancel{\sigma}\,dy = \frac{1}{\sqrt{\pi}}\underbrace{\int_{\mathbb{R}} e^{-y^2}dy}_{=\sqrt{\pi}}$$

$x \to -\infty \quad y \to -\infty$

$x \to +\infty \quad y \to +\infty$

$$= \frac{1}{\sqrt{\pi}}\sqrt{\pi} = 1$$

N.B. $\mu = 0$ $\sigma = 1$, la distribuzione $N(0,1)$ si dice distribuzione normale standard e la densità è

$$f_0(x) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-x^2}{2}\right) \qquad \forall x \in \mathbb{R}$$