STATISTICA DESCRITTIVA

Titolo nota 16/11/2017

Si occupa di dati raccolti su popolazioni molto ampie.

- INDIVIDUI: singolo oggetto su cui raccolgo l'informazione
- POPOLAZIONE: insieme degli individui oggetto dell'indagine
- CARATTERE OSSERVATO: è o una quantità misurata o una qualità rilevata su ogni individuo della popolazione

- Altezza dei residenti nel Comune di Firenze
  Anagrafe: ogni persona registrata è un individuo e ad ogni persona chiedo quanto è alto
- Reddito familiare dei residenti nel Comune di Firenze
  Anagrafe: ogni nucleo familiare è un individuo e su ogni nucleo familiare rilevo il reddito.
- Colore degli occhi dei residenti del Comune
  Anagrafe: ogni persona registrata è un individuo e a cui chiedo il colore degli occhi.
- La numerosità dei nuclei familiari
  Anagrafe: ogni nucleo familiare è un individuo e vedo quanti sono i componenti

CARATTERI
- CARATTERI QUALITATIVI
- CARATTERI NUMERICI
  - DISCRETI
  - CONTINUI

Supponiamo di avere una popolazione di n individui
Su ogni individuo rilevo (o misuro) lo stesso carattere

$x_1, x_2 \ldots x_n$

Chiamo $x = (x_1 \_\_ x_n)$ il vettore dei dati rilevati (o misurati)

Si chiama CAMPIONE STATISTICO (EMPIRICO)

n si dice CARDINALITÀ DEL CAMPIONE

Se il campione è relativo ad un carattere qualitativo o numerico discreto chiamo MODALITÀ DEL CAMPIONE l'insieme dei valori che vi appaiono

Le modalità si indicano $z_1 \_\_ z_k$, $k \leq n$

Se il campione è relativo ad un carattere numerico continuo $\Rightarrow \exists [a,b) \supset \{x_1 \_\_ x_n\}$

Scelgo $N \in \mathbb{N}$ e suddivido $[a,b)$ in N parti uguali

$$I_j = \left[a + \frac{b-a}{N}(j-1), \; a + \frac{b-a}{N} j\right) \quad j = 1 \_\_ N$$

Se almeno un valore tra $x_1 \_\_ x_n$ appartiene a I_j, dico che I_j è una CLASSE MODALE DEL CAMPIONE

Sia $x = (x_1 \_\_ x_n)$ campione relativo ad un carattere qualitativo o numerico discreto e siano $z_1 \_\_ z_k$ le sue modalità

$j = 1 \_\_ k \qquad n_j := \# \{ i \in \{1 \_\_ n\}, \; x_i = z_j \}$

n_j si dice EFFETTIVO o FREQUENZA ASSOLUTA della modalità z_j

Nel caso numerico continuo, chiamo EFFETTIVO in I_j

$j = 1 \_\_ k \qquad n_j := \# \{ i \in \{1, \_\_ n\}, \; x_i \in I_j \}$

In entrambi i casi $p_j := \frac{n_j}{n}$ si dice FREQUENZA RELATIVA (della modalità z_j o della classe modale I_j)

$n_1 \; n_2 \; \ldots \; n_k$

Se $\exists! \, \bar{j} \in \{1 \ldots k\}$ t.c. $n_{\bar{j}} = \max\{n_1 \ldots n_k\}$, dico che $z_{\bar{j}}$ (o $I_{\bar{j}}$) è la MODA DEL CAMPIONE

Se $\exists \, s \geq 2$ $\bar{j}_1, \ldots \bar{j}_s$ t.c. $n_{\bar{j}_1} = \ldots = n_{\bar{j}_s} = \max\{n_1 \ldots n_k\}$, allora $z_{\bar{j}_1} \ldots z_{\bar{j}_k}$ (o $I_{\bar{j}_1} \ldots I_{\bar{j}_k}$) si dicono VALORI MODALI

— o —

Consideriamo solo CARATTERI NUMERICI

$x = (x_1 \; \ldots \; x_n)$

$\Rightarrow$ Riordino $x_1 \ldots x_n$ in ordine crescente

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n-1)} \leq x_{(n)}$$

1° CASO $\quad n = 2m+1$

$$\underbrace{x_{(1)} \leq \ldots \; x_{(m)}}_{m \text{ dati}} \leq x_{(m+1)} \leq \underbrace{x_{(m+2)} \leq \ldots \leq x_{(2m+1)}}_{m \text{ dati}}$$

$x_{(m+1)}$ si dice MEDIANA DEL CAMPIONE

2° CASO $\quad n = 2m$

$$\underbrace{x_{(1)} \leq \ldots \leq x_{(m)}}_{m \text{ dati}} \leq \underbrace{x_{(m+1)} \leq \ldots \leq x_{(2m)}}_{m \text{ dati}}$$

$\frac{x_{(m)} + x_{(m+1)}}{2}$ si dice MEDIANA DEL CAMPIONE

MEDIA e VARIANZA CAMPIONARIA

$x = (x_1 \; — \; x_n)$ chiamo media del campione x

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad z_1 - - z_k$$

$$\bar{x} = \frac{1}{n} \left(x_1 + x_2 \; - - + x_n \right) = \frac{1}{n} \left(n_1 z_1 + n_2 z_2 + - - + n_k z_k \right)$$

$$= p_1 z_1 + p_2 z_2 + — + p_k z_k = \sum_{j=1}^{k} p_j z_j$$

Chiamo VARIANZA CAMPIONARIA DI x il numero

$$Var[x] = \sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$\sqrt{\sigma_x^2}$ si dice DEVIAZIONE STANDARD o SCARTO QUADRATICO MEDIO si indica $Std[x]$

$$\sigma_x^2 = \frac{1}{n-1} \left((x_1 - \bar{x})^2 + - - + (x_n - \bar{x})^2 \right)$$

$$= \frac{1}{n-1} \left(n_1 (z_1 - \bar{x})^2 + — n_k (z_k - \bar{x})^2 \right) =$$

$$= \frac{n}{n-1} \left(p_1 (z_1 - \bar{x})^2 + — + p_k (z_k - \bar{x})^2 \right.$$

$$= \frac{n}{n-1} \sum_{i=1}^{k} p_i (z_i - \bar{x})^2$$

$$x = (x_1 \text{ —— } y_n)$$
$$y = (y_1 \text{ —— } y_n)$$

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

COVARIANZA DI x E y

$$Corr(x,y) = \frac{Cov(x,y)}{Std[x]\, Std[y]} =$$

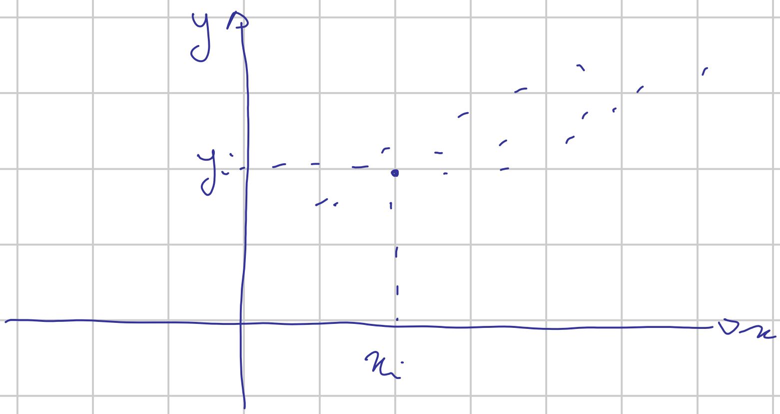
$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right)\left(\sum_{i=1}^{n} (y_i - \bar{y})^2\right)}}$$

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

$$-1 \leq \frac{\underline{v} \cdot \underline{w}}{\|\underline{v}\| \cdot \|\underline{w}\|} \leq 1$$

Vale 1 sse $\underline{v} = a\underline{w}$ $a > 0$ $y = ax + b$

-1 sse $\underline{v} = -a\underline{w}$ $a < 0$ $y = ax + b$



$$y = ax + b \qquad x = x_i \qquad (x_i, ax_i + b)$$

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

$$S(a,b) = \sum_{i=1}^{n} \left(y_i - \bar{y} + \bar{y} - a(x_i - \bar{x} + \bar{x}) - b \right)^2$$

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

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

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

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

$$= (n-1)\sigma_y^2 + a^2(n-1)\sigma_x^2 + n(\bar{y} - a\bar{x} - b)^2$$
$$- 2a(n-1)\,\mathrm{Cov}(x,y)$$

$$\sum_{i=1}^{n} (y_i - \bar{y}) = n\bar{y} - n\bar{y} = 0$$

$$= \underbrace{a^2(n-1)\sigma_x^2 - 2a(n-1)\,\mathrm{Cov}(x,y) + (n-1)\sigma_y^2}_{f(a)} + n(\bar{y} - a\bar{x} - b)^2$$

$$f'(a) = 2a(n-1)\sigma_x^2 - 2(n-1)\,\mathrm{Cov}(x,y) = 0$$

$$\Rightarrow \qquad a = \frac{\mathrm{Cov}(x,y)}{\sigma_x^2}$$

$$b = \bar{y} - a\bar{x} = \bar{y} - \frac{\mathrm{Cov}(x,y)}{\sigma_x^2}\bar{x}$$

$$y = \frac{\mathrm{Cov}(x,y)}{\sigma_x^2} x + \bar{y} - \frac{\mathrm{Cov}(x,y)}{\sigma_x^2}\bar{x}$$

$$y - \bar{y} = \frac{\mathrm{Cov}(x,y)}{\sigma_x^2}(x - \bar{x}) \qquad \text{passa per } (\bar{x}, \bar{y})$$

$$\min S(a,b) = (n-1)\frac{\mathrm{Cov}^2(x,y)}{\sigma_x^2} - 2(n-1)\,\mathrm{Cov}(x,y)\frac{\mathrm{Cov}(x,y)}{\sigma_x^2}$$
$$+ (n-1)\sigma_y^2 = (n-1)\left(\sigma_y^2 - \frac{\mathrm{Cov}^2(x,y)}{\sigma_x^2}\right)$$
$$= (n-1)\sigma_y^2\left(1 - \frac{\mathrm{Cov}^2(x,y)}{\sigma_x^2\sigma_y^2}\right) =$$

$$= (n-1)\,\sigma_y^2 \left(1 - \mathrm{Corr}^2(x,y)\right)$$

—— o ——

STATISTICA INFERENZIALE

$X_1 \ldots X_n$ famiglia finita di v.a.

Dico che $X_1 \ldots X_n$ è un CAMPIONE STATISTICO se le v.a. $X_1 \ldots X_n$ sono i.i.d.

n si dice CARDINALITÀ DEL CAMPIONE

N.B. Se $\mathbb{P}_{X_i} = f(x)\,dx \quad \forall i = 1 \ldots n$, allora

$$\mathbb{P}_{X_1 \ldots X_n} = f(x_1) f(x_2) \ldots f(x_n)\, dx_1 \ldots dx_n$$

La comune distribuzione $\mathbb{P}_{X_i}$ si dice DISTRIBUZIONE CAMPIONARIA.

Definisco MEDIA CAMPIONARIA la v.a.

$$\overline{X}_n = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$f(x_1 \ldots x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$

$f \circ (X_1 \ldots X_n)$

Definisco VARIANZA CAMPIONARIA la v.a.

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

PROP Sia $X_1 \ldots X_n$ un campione statistico t.c.

$E[X_i] = \mu \in \mathbb{R} \qquad Var[X_i] = \sigma^2 \qquad \forall i = 1 \ldots n$

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

$E[S^2] = \sigma^2$

N.B. Se Y è una statistica di $X_1 \ldots X_n$ cioè $Y = f \circ (X_1 \ldots X_n)$ $\quad f : \mathbb{R}^n \to \mathbb{R}$ e

$E[Y] = \alpha(X_1 \ldots X_n)$ (quantità legata al campione)

dico che Y è uno STIMATORE ESATTO DELLA QUANTITÀ

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

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

$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

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

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

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

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

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

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

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

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

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

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

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

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\sigma^2 + \mu^2 - 0\right) - \frac{n}{n-1}\left(\frac{\sigma^2}{n} + \mu^2 + 0\right)$$

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

— o —

T = successo = 1 $\qquad$ C = insuccesso = 0

$x_1 \; - - - - \; x_n$

$X_1 \; - - - - \; X_n$

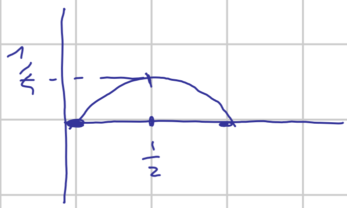
Esce Testa con probabilità p ad ogni lancio

$$\mathbb{P}_{X_i} = \mathcal{B}(p)$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \mathbb{E}[\bar{X}] = p \qquad \mathrm{Var}[\bar{X}] = \frac{p(1-p)}{n}$$

$$\mathbb{P}(|\bar{X} - p| > t) \le \frac{p(1-p)}{nt^2}$$

$$p \in [0,1] \qquad f(p) = p(1-p)$$



$$= p - p^2 =$$

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

$$\mathbb{P}(|\bar{X} - p| > t) \le \frac{1}{4nt^2}$$

$$x_1 = X_1(\omega) \quad x_2 = X_2(\omega) \quad - - - . \quad x_n = X_n(\omega)$$

$$\bar{X}(\omega) = \frac{1}{n}\sum_{i=1}^{n} X_i(\omega) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Voglio che $\mathbb{P}\left(|\bar{X} - p| < \frac{1}{20}\right) \ge 90\%$ $\Leftarrow$

$$\mathbb{P}\left(|\bar{X} - p| > \frac{1}{20}\right) \le \frac{1}{10} \qquad t = \frac{1}{20} \qquad \frac{1}{4nt^2} \le \frac{1}{10}$$

$$\frac{1}{4nt^2} \le \frac{1}{10} \qquad \frac{400}{4n} \le \frac{1}{10} \qquad \frac{100}{n} \le \frac{1}{10}$$

$$n \ge 1000$$

DISTRIBUZIONI $\Gamma(\alpha, \lambda)$

$$f(x) = \begin{cases} \dfrac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

Se $\lambda = \frac{1}{2}$ $\alpha = \frac{n}{2}$ $n \in \mathbb{N}$ la distribuzione $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ si dice DISTRIBUZIONE DI PEARSON (o DISTRIBUZIONE χ^2) CON n GRADI DI LIBERTÀ e si dice χ^2_n

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

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} \, dx \qquad \alpha > 0$$

$$\Gamma(1) = 1 \qquad \Gamma(\alpha+1) = \alpha\Gamma(\alpha) \implies \Gamma(n) = (n-1)!$$

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{+\infty} x^{-1/2} e^{-x} \, dx \qquad x = y^2 \qquad dx = 2y\,dy$$

$$= \int_0^{+\infty} \cancel{y^{-1}} e^{-y^2} 2\cancel{y} \, dy = \int_{\mathbb{R}} e^{-y^2} dy = \sqrt{\pi}$$

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}$$

$$\Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3 \cdot 1}{2 \cdot 2}\sqrt{\pi}$$

$$\Gamma\left(\frac{7}{2}\right) = \frac{5}{2}\Gamma\left(\frac{5}{2}\right) = \frac{5 \cdot 3 \cdot 1}{2 \cdot 2 \cdot 2}\sqrt{\pi}$$

$$\Gamma\left(\frac{2n+1}{2}\right) = \frac{(2n-1)!!\sqrt{\pi}}{2^n}$$

Se X è una v.a. con $\mathbb{P}_X = \chi^2_n$

$$\mathbb{E}[X] = \frac{\alpha}{\lambda} = \frac{n/2}{1/2} = n \qquad Var[X] = \frac{\alpha}{\lambda^2} = \frac{(n/2)}{(1/2)^2} = 2n$$

PROPRIETÀ Se X e Y sono v.a. indipendenti con $\mathbb{P}_X = \chi^2_n$ e $\mathbb{P}_Y = \chi^2_k$

Allora $\mathbb{P}_{X+Y} = \chi^2_{n+k}$

DIM

$$\mathbb{P}_X = f(x)\,dx \qquad f(x) = \begin{cases} \frac{1}{\Gamma(n/2)}\left(\frac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$\mathbb{P}_Y = g(x)\,dx \qquad g(x) = \begin{cases} \frac{1}{\Gamma(\frac{k}{2})}\left(\frac{1}{2}\right)^{k/2} x^{k/2-1} e^{-x/2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$\mathbb{P}_{X+Y} = h(x)\,dx \qquad h(x) = \int_{\mathbb{R}} f(y)\,g(x-y)\,dy$$

$$h(x) = \int_0^x f(y)\,g(x-y)\,dy \qquad \begin{aligned} x-y &> 0 \\ y &< x \end{aligned}$$

$x \leq 0 \quad h(x) = 0$

$$x > 0 \quad h(x) = \int_0^x \frac{1}{\Gamma(\frac{n}{2})\Gamma(\frac{k}{2})}\left(\frac{1}{2}\right)^{\frac{n}{2}+\frac{k}{2}} y^{n/2-1} \cancel{e^{-\frac{y}{2}}} (x-y)^{\frac{k}{2}-1} e^{-\frac{(x-\cancel{y})}{2}}\,dy$$

$$= \frac{e^{-x/2}}{\Gamma(\frac{n}{2})\Gamma(\frac{k}{2})}\left(\frac{1}{2}\right)^{\frac{n+k}{2}} \int_0^x y^{n/2-1} (x-y)^{k/2-1}\,dy$$

$$y = xt \qquad dy = x\,dt \qquad t = \frac{y}{x} \qquad x-y = x(1-t)$$

$$= \frac{e^{-x/2}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{2}\right)^{\frac{n+k}{2}} \int_0^1 x^{n/2 - 1}\, t^{n/2-1}\, x^{\frac{k}{2} \cancel{-1}} (1-t)^{k/2-1} \cancel{x}\, dt$$

$$= e^{-x/2}\, x^{\frac{n+k}{2}-1} \frac{1}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{2}\right)^{\frac{n+k}{2}} \underbrace{\int_0^1 t^{n/2-1} (1-t)^{\frac{k}{2}-1} dt}_{\text{converge}}$$

$$= C\, e^{-x/2}\, x^{\frac{n+k}{2}-1}$$

$\int_{\mathbb{R}} f(x)dx = 1 \Rightarrow C =$ coeff della Pearson a $n+k$ gradi di libertà

$$C = \frac{1}{\Gamma\left(\frac{n+k}{2}\right)} \left(\frac{1}{2}\right)^{\frac{n+k}{2}}$$

$$\Rightarrow \mathbb{P}_{X+Y} = \chi^2_{n+k}$$

$\mathbb{P}_X = \Gamma(\alpha, \lambda)$ $\quad \mathbb{P}_Y = \Gamma(\beta, \lambda)$

X e Y indipendent. $\Rightarrow \mathbb{P}_{X+Y} = \Gamma(\alpha+\beta, \lambda)$

— o —

$\mathbb{P}_X = N(0,1)$ $\quad \mathbb{P}_X = f(x)dx$ $\quad f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ $\quad x \in \mathbb{R}$

$Y = X^2$ $\quad \mathbb{P}_Y = g(x)\,dx$ $\quad g(x) = \begin{cases} \frac{1}{2\sqrt{x}}\left(f(\sqrt{x}) + f(-\sqrt{x})\right) & x > 0 \\ 0 & x \leq 0 \end{cases}$

$x > 0$ $\quad g(x) = \frac{1}{2\sqrt{x}}\left(f(\sqrt{x}) + f(\sqrt{x})\right) = \frac{f(\sqrt{x})}{\sqrt{x}} =$

$$= \frac{1}{\sqrt{2\pi}}\, x^{-1/2} \exp\left(-\frac{x}{2}\right) \Rightarrow \mathbb{P}_Y = \chi^2_1$$

$$= \frac{1}{\sqrt{2\pi}}\, x^{1/2 - 1} \exp\left(-\frac{x}{2}\right)$$

TEOREMA Siano $X_1 \ldots X_n$ v.a. indipendenti t.c.

$\mathbb{P}_{X_i} = N(\mu_i, \sigma_i^2)$, allora la v.a.

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i} \right)^2$$ ha distribuzione χ^2_n

DIM $\quad \mathbb{P}_{X_i} = N(\mu_i, \sigma_i^2) \qquad Z_i = \frac{X_i - \mu_i}{\sigma_i} = N(0,1)$

$n = 1 \qquad Z_1^2 \quad \chi^2_1$

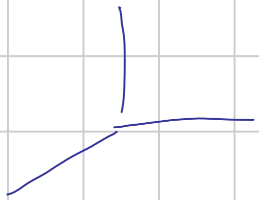
$n = 2 \qquad Z_1^2 + Z_2^2 \qquad \mathbb{P}_{Z_1^2} = \mathbb{P}_{Z_2^2} = \chi^2_1$

X_1 e X_2 indipendenti $\Rightarrow$ Z_1^2 e Z_2^2 indipendenti

$\mathbb{P}_{Z_1^2 + Z_2^2} = \chi^2_2$

$n = 3 \qquad \underbrace{(Z_1^2 + Z_2^2)}_{\chi^2_2} + \underbrace{Z_3^2}_{\chi^2_1} \qquad \mathbb{P}_{Z_1^2 + Z_2^2 + Z_3^2} = \chi^2_{2+1} = \chi^2_3$

— o —



• $\mathbb{P}_{X_i} = N(0, \sigma^2) \qquad i = 1, 2, 3$

$\mathbb{P}(X_1^2 + X_2^2 + X_3^2 < \varepsilon)$

$Z_i = \frac{X_i}{\sigma} \Rightarrow \mathbb{P}_{Z_i} = N(0,1) \qquad X_i = \sigma Z_i$

$X_1^2 + X_2^2 + X_3^2 = \sigma^2 (Z_1^2 + Z_2^2 + Z_3^2)$

$\mathbb{P}\left(Z_1^2 + Z_2^2 + Z_3^2 < \frac{\varepsilon}{\sigma^2} \right) > 1 - \alpha$

$F_{\chi^2_3}(t) = 1 - \alpha \qquad \bar{t} \qquad \frac{\varepsilon}{\sigma^2} < \bar{t}$

COROLLARIO Sia $X_1 \ldots X_n$ un campione statistico di distribuzione campionaria $N(\mu, \sigma^2)$

Allora la v.a. $\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$ ha distribuzione χ^2_n -