MEDIA EMPIRICA

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

VARIANZA CAMPIONARIA

$$var(x) = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}$$

VARIANZA EMPIRICA

$$var_e(x) = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n}$$

$$= \sum_{i=1}^{n} \frac{(x_i)^2}{n} - \overline{x}^2$$

DEVIAZIONE STANDARD

$$\sigma(x) = \sqrt{var(x)}$$

FUNZIONE DI RIPARTIZIONE EMPIRICA

$$F_e(t) = \frac{\#\{x_i \mid x_i \leq t\}}{n}$$

K-MO PERCENTILE

$K \in (0;100)$

il minimo numero che supera $K\%$ dati

β - QUANTILE

$$\beta = \frac{k}{100} \Rightarrow \beta \in (0;1)$$

COVARIANZA CAMPIONARIA

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

COVARIANZA EMPIRICA

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

$$= \sum_{i=1}^{n} \frac{x_i y_i}{n} - \bar{x}\bar{y}.$$

COEFFICIENTE di CORRELAZIONE

$$r(x,y) = \frac{\mathrm{cov}(x,y)}{\sigma(x)\,\sigma(y)} \quad \text{con } \sigma(x) \neq 0,\ \sigma(y) \neq 0$$

RETTA di REGRESSIONE

$$\min_{a,b \in \mathbb{R}^2} \sum_{i=1}^{n} (y_i - a - b x_i)^2 = \sum_{i=1} (y_i - \bar{y})^2 (1 - r(x,y)^2)$$

La retta $y = a^* + b^* x_i$ è la retta di regressione con

$$b^* = \frac{cov(x,y)}{var(x)}, \quad a^* = -b^*\bar{x} + \bar{y}.$$

NUMERO di APPLICAZIONI

$$\#\{f : fun\ \{1, \dots, k\} \longrightarrow \{1, \dots, n\}\} = n^k$$

PERMUTAZIONI

Ordinare l'insieme $\{1 \dots n\}$
$\Rightarrow n!$

COEFFICIENTE BINOMIALE $k \in [0; n]$

I sottoinsiemi di $\{1, \dots, n\}$ di k elementi sono $\binom{n}{k}$.

FORMULA di BAYES

Hp: A evento, $B_1 \dots B_n$ sist. di altern.

$$P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$$

$$P(B_i|A) = P(A|B_i)\,P(B_i) / P(A)$$

$$= \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j=1}^{n} P(A|B_j) \cdot P(B_j)}$$

INDIPENDENZA STOCASTICA

A, B indipendenti se

$$P(A \cap B) = P(A) \cdot P(B).$$

FUNZIONE DI MASSA

$$p(x_i) = P(x_i)$$

DENSITA' DI PROBABILITA'

$f: \mathbb{R} \to [0; +\infty[$ integrabile t.c.

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

E' associata alla probabilita' sull'insieme A tramite la formula

$$P(A) = \int_A f(x)\,dx.$$

VARIABILE ALEATORIA

Spazio Ω su cui e' definita una probabilità P;

$X : \Omega \to \mathbb{R}$ variabile aleatoria

Legge di probabilità e'

$$P_X(A) = P(X^{-1}(A)).$$

VARIABILE ALEATORIA DISCRETA

$$P_X(A) = P\{X \in A\} = \sum_{x_i \in A} P_X(x_i).$$

VARIABILE ALEATORIA CON DENSITA'

$$P_X(A) = P(X \in A) = \int_A f(x)\,dx.$$

FUNZIONE DI RIPARTIZIONE

$$F_X(x) = P\{X \leq x\}$$

Ha le proprietà:

i) e' debolmente crescente;

ii) $\lim_{x \to 0} F_X(x) = 0$, $\lim_{x \to +\infty} F_X(x) = 1$;

iii) e' continua a destra.

QUANTILE

β-quantile $:= r \mid P\{X \le r\} \ge \beta,$
$P\{X > r\} \ge 1-\beta.$

VARIABILE ALEATORIA BINOMIALE

n prove di esperimento a due esiti con probabilita' di successo p:
X conta i successi.
$B(n,p)$.
se $n=1$ allora di Bernoulli.

$$p(h) = P\{X=h\} = \binom{n}{h} p^h (1-p)^{n-h}.$$

VARIABILE ALEATORIA GEOMETRICA

Esperimento ripetuto finché non si ha un successo: X conta i tentativi.

$$p(h) = P\{X = n+h \mid X > n\} = P\{X=h\} = (1-p)^{h-1} p$$

VARIABILE ALEATORIA di POISSON

$\lambda > 0$, X assume come valori tutti i naturali e

$$p(h) = P\{X=h\} = e^{-\lambda} \frac{\lambda^h}{h!}$$

VARIABILE ALEATORIA CON DENSITA' ESPONENZIALE

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

con λ positivo

COMPOSIZIONE DI VARIABILI ALEATORIE

X v.a. con densità $f_X \neq 0$ su A con A intervallo aperto;

$h: A \to B$ con B intervallo aperto biunivoca, derivabile e con inversa derivabile

$Y := h \circ X$, allora la densità di Y ha formula:

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y)) \left| \frac{d}{dy} h^{-1}(y) \right| \\ 0 & y \notin B \end{cases}$$

DENSITA' GAUSSIANA STANDARD

$N(0,1)$ ha funzione

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

c.d.f. $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \, dt$

$$\Phi(-x) = 1 - \Phi(x)$$

QUANTILI $q_{1-\alpha} = -q_\alpha$

DENSITA' GAUSSIANA GENERALE

$N(m, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$\downarrow$

$(\sigma X + m)$ con X gaussiana standard

è equivalente

VARIABILI ALEATORIE DOPPIE

disc. $P_{X,Y}(A) = P\{(X,Y) \in A\} = \sum'_{(x_i, y_j) \in A} p(x_i, y_j)$

cont. $P_{X,Y}(A) = P\{(X,Y) \in A\} = \iint_A f(x,y) \, dx \, dy$

SOMMA di BINOMIALI

$$X \sim B(n,p) \quad , \quad Y \sim B(m,p)$$
$$Z = X + Y \sim B(n+m, p)$$

SOMMA di GAUSSIANE

$$X \sim N(m_1, \sigma_1^2) \, , \quad Y \sim N(m_2, \sigma_2^2)$$
$$Z = X + Y \sim N(m_1 + m_2 \, , \sigma_1^2 + \sigma_2^2)$$

e.g. $X \sim N(1,4) \, , \; Y \sim N(2,9)$

$$Z \sim N(0,1)$$

$$X = 2Z + 1$$
$$Y = 3Z + 2$$
$$X + Y \neq 2Z + 1 + 3Z + 2$$
$$X + Y \sim N(1+2, 4+9)$$
$$N(3, 13) = \sqrt{13}\, Z + 3$$

FORMULA della CONVOLUZIONE DISCRETA

X, Y indipendenti $\quad Z = X + Y$

$$p_Z(n) = \sum_{h=0}^{n} p_X(h) \cdot p_Y(n-h)$$

FORMULA della CONVOLUZIONE

X, Y indipendenti $Z = X + Y$

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) \cdot f_Y(z-x)\,dx = \int_{-\infty}^{+\infty} f_Y(y) f_X(z-y)\,dy$$

VALORE ATTESO

$$E[X] = \sum_i x_i \, p(x_i)$$

nel caso di variabili con densità:

$$E[X] = \int_{-\infty}^{+\infty} x f(x)\,dx$$

BINOMIALE

$E[X] = np$,

$Var(X) = np(1-p)$

POISSON

$E[X] = \lambda$, $Var(X) = \lambda$

$E[X^2] = \lambda + \lambda^2$

GAUSSIANA STANDARD

$E[X^n] = 0$

$Var(X) = 1$

GEOMETRICA

$E[X] = 1/p$

$Var(X) = (1-p)/p^2$

ESPONENZIALE

$E[X] = 1/\lambda$

$Var(X) = 1/\lambda^2$

VALORE ATTESO di COMPOSIZIONI di VARIABILI ALEATORIE

i) caso discreto

$Y = g(X)$ con X discreto

$$E[g(X)] = \sum_i g(x_i) \cdot p(x_i)$$

ii) Caso con densita'

$Y = g(X)$ con X con densita'

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) \, dx$$

In generale:

X, Y ammettono valore atteso

$\Rightarrow Z = X + Y$ ha valore atteso e valgono:

- $E[X+Y] = E[X] + E[Y]$
- $|E[X]| \leq E[|X|]$
- $E[aX+b] = a E[X] + b$
- $X \geq Y \Rightarrow E[X] \geq E[Y]$
- X, Y indipendenti $\Rightarrow E[XY] = E[X] E[Y]$

MOMENTO di ORDINE n

$$E[X^n].$$

DISUGUAGLIANZA di MARKOV

Y variabile aleatoria a valori positivi

$a \in \mathbb{R}^+$

$$a\, P\{Y \geq a\} \leq E[Y]$$

DISUGUAGLIANZA di SCHWARTZ

X, Y variabili aleatorie

$$E[|XY|] \leq \sqrt{E[X^2]} \cdot \sqrt{E[Y^2]}$$

VARIANZA in funzione del valore atteso

$$\mathrm{Var}(X) = E[(X - E[X])^2]$$
$$= E[X^2] - E[X]^2$$

Si ricorda che

$\sigma :=$ deviazione standard $= \sqrt{\mathrm{Var}(X)}$

DISUGUAGLIANZA di CHEBYSHEV $d > 0$

$$P\{|X - E[X]| > d\} \leq \mathrm{Var}(X) / d^2$$

COVARIANZA

$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$$

Valgono le proprietà:

- $\text{Cov}(aX+bY+c, Z) = a\text{Cov}(X,Z) + b\text{Cov}(Y,Z)$
- $\text{Cov}(X,Y) = \text{Cov}(Y,X)$
- $\text{Var}(X) = \text{Cov}(X,X)$
- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$

Le seguenti proprietà sono tra loro equivalenti:

- $E[XY] = E[X]E[Y]$
- $\text{Cov}(X,Y) = 0$
- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

In questo caso X, Y sono incorrelate (possono essere indipendenti)

COEFFICIENTE di CORRELAZIONE

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

FUNZIONE GENERATRICE dei MOMENTI

$$G_X(t) = E[e^{tX}] = \begin{cases} \sum_{x_i} e^{tx_i} p(x_i) & X \text{ discreta} \\ \int_{\mathbb{R}} e^{tx} f(x)\,dx & \text{altr.} \end{cases}$$

Valgono le proprietà:

i) $G_{aX+b}(t) = G_X(at)\,e^{tb}$

ii) X, Y indipendenti $\Rightarrow G_{X+Y}(t) = G_X(t) + G_Y(t)$.

$$E[X^n] = \left.\frac{d^n G_X(t)}{dt^n}\right|_{t=0}$$

LEGGE DEBOLE dei GRANDI NUMERI

$X_1, \dots, X_n$ variabili dotate di momento secondo

$\mu = E[X_i]$

$$\lim_{n\to\infty} P\left\{\left|\frac{X_1+\dots+X_n}{n} - \mu\right| > \varepsilon\right\} = 0 \quad \forall \varepsilon > 0$$

TEOREMA LIMITE CENTRALE

$$\lim_{n\to\infty} P\left\{\frac{X_1+\dots+X_n - n\mu}{\sigma\sqrt{n}} \in [a;b]\right\} = \phi(b) - \phi(a)$$

Si nota che

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} = \sqrt{n}\cdot\frac{\overline{X}_n - \mu}{\sigma}$$ è Gaussiana standard per $n \geq 80$ (o 50).

Si nota che per X binomiale

$$\frac{X - np}{\sqrt{np(1-p)}}$$ è approssimativamente gaussiana standard per $np(1-p) \geq 15$ (o 20).

e.g. Stimare $P\{50 \leq X_1 + \dots + X_{60} \quad 0\}$.

$$E[X] = 1$$
$$Var(X) = 3/5$$

$$P\{50 \leq X_1 + \dots + X_{60} \leq 70\} =$$

$$P\left\{\frac{50 - 60\cdot 1}{\sqrt{\frac{3}{5}\cdot 60}} \leq \frac{X_1 + \dots + X_{60} - 60\cdot 1}{\sqrt{\frac{3}{5}\cdot 60}} \leq \frac{70 - 60\cdot 1}{\sqrt{\frac{3}{5}\cdot 60}}\right\}$$

$$= P\left\{-\frac{10}{6} \leq Z \leq \frac{10}{6}\right\}$$ con $Z \sim N(0,1)$

$$= \phi\left(\frac{5}{3}\right) - \phi\left(-\frac{5}{3}\right) \approx 0.902$$

DENSITÀ Γ

$r > 0, \lambda > 0$

$$\Gamma(r, \lambda) : f(x) = \begin{cases} \frac{1}{\Gamma(r)} \lambda^r x^{r-1} e^{-\lambda x}, & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$E[X^\beta] = \frac{\Gamma(r+\beta)}{\Gamma(r)\lambda^\beta}, \quad Var(X) = \frac{r}{\lambda^2}$$

$$E[X] = \frac{r}{\lambda}, \quad E[X^2] = \frac{(r+1)r}{\lambda^2}$$

$$G_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r$$

DENSITÀ χ^2

$X_1, \dots, X_n$ gaussiane standard

$(X_1^2 + \dots + X_n^2)$ ha densità $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$

si chiama chi-quadro a n gradi di libertà

DENSITA' di STUDENT

$$T_n = \frac{X}{\sqrt{\frac{C_n}{n}}} = \sqrt{n}\,\frac{X}{\sqrt{C_n}}$$

con $X \sim N(0,1)$

$C_n \sim \chi^2(n)$

X, C_n indipendenti

E' una funzione pari e valgono

- $F_n(-x) = 1 - F_n(x)$
- $\tau_{(\alpha, n)} = -\tau_{(1-\alpha, n)}$

con $\tau(\alpha, n)$ alfa-quantile di T_n.

STIMA di MASSIMA VEROSIMIGLIANZA

i) caso discreto

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} p_\theta(x_i)$$

ii) caso con densita'

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

Si cerca $\hat{\theta}(x_1, \dots, x_n)$

$= \max L(\theta; x_1, \dots, x_n)$

STIMA COL METODO dei MOMENTI

$$E_\theta[X] = \sum_{i=1}^{n} \frac{x_i}{n}$$ media empirica

se non funziona si passa a

$$E_\theta[X^2] = \sum_{i=1}^{n} \frac{x_i^2}{n}$$

Il risultato e' $\tilde{\theta}$.

DENSITA' ESPONENZIALE

$$\tilde{\theta} = \frac{1}{\overline{x}} = \hat{\theta}.$$

DENSITA' BINOMIALE $B(m, \theta)$

$$\hat{\theta} = \frac{1}{m} \cdot \sum_{i=1}^{m} \frac{x_i}{m} = \tilde{\theta}$$

DENSITA' di POISSON

$$\hat{\theta} = \overline{x} = \tilde{\theta}$$

INTERVALLO di FIDUCIA AL LIVELLO $(1-\alpha)$

$$P_\theta\{\theta \in I\} \geq (1-\alpha) \quad \text{con } \alpha \in (0;1)$$

INTERVALLO di FIDUCIA per la MEDIA di un CAMPIONE GAUSSIANO $(1-\alpha)$

$X_1, \dots, X_n$ con varianza σ^2

E' un intervallo bilatero di fiducia al $(1-\alpha)\cdot 100\%$ di formula

$$I = \left[\bar{x} - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} ; \bar{x} + \underbrace{\frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}}_{\text{precisione}}\right]$$

PRECISIONE RELATIVA $:= \dfrac{\text{precisione}}{\bar{x}}$

Nel caso di varianza sconosciuta:

$$I = \left[\bar{x} - t_{(1-\frac{\alpha}{2}, n-1)} \frac{\bar{\sigma}}{\sqrt{n}} ; \bar{x} + t_{(1-\frac{\alpha}{2}, n-1)} \bar{\sigma}/\sqrt{n}\right]$$

con $\bar{\sigma} = \sqrt{\dfrac{\sum_i (x_i - \bar{x})^2}{n-1}}$. se unilatero α anziché $\frac{\alpha}{2}$.

INTERVALLO di FIDUCIA per la MEDIA di un CAMPIONE di BERNOULLI $(1-\alpha)$

$$I = \left[\bar{x} - q_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \; ; \; \bar{x} + q_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

$$\hat{p} = \bar{x}$$

INTERVALLO di FIDUCIA per la VARIANZA di un CAMPIONE GAUSSIANO $(1-\alpha)$

$$sx: \; I = \left[0 \; ; \; \frac{(n-1)\bar{\sigma}^2}{\chi^2_{(\alpha, n-1)}} \right]$$

$$dx: \; I = \left[\frac{(n-1)\bar{\sigma}^2}{\chi_{(1-\alpha, n-1)}} \; ; \; \infty \right[$$

TEST sulla MEDIA di un CAMPIONE GAUSSIANO con VARIANZA NOTA (Z-TEST)

$$H_0) \; m = m_0 \quad \text{contro} \quad H_1) \; m \neq m_0$$

REGIONE CRITICA $C = \left\{ |\bar{X} - m_0| > \sigma/\sqrt{n} \cdot q_{1-\frac{\alpha}{2}} \right\}$

$$\bar{\alpha} = 2\left(1 - \phi\left(\frac{\sqrt{n}}{\sigma}|\bar{x} - m_0|\right)\right).$$

CURVA OPERATIVA

$$\beta(m) = \phi\left(\frac{\sqrt{n}(m_0 - m)}{\sigma} + q_{1-\frac{\alpha}{2}}\right)$$

$$- \phi\left(\frac{\sqrt{n}(m_0 - m)}{\sigma} - q_{1-\frac{\alpha}{2}}\right)$$

TEST UNILATERO

$$\mathcal{H}_0)\ m \leq m_0 \,, \quad \mathcal{H}_1)\ m \geq m_0 :$$

$$C = \left\{ (\bar{x} - m_0) > \frac{\sigma}{\sqrt{n}} q_{1-\alpha} \right\}$$

$$\bar{\alpha} = 1 - \phi\left(\frac{\sqrt{n}}{\sigma}(\bar{x} - m_0)\right)$$

$$\beta(m) = \phi\left(\frac{\sqrt{n}(m_0 - m)}{\sigma} + q_{1-\alpha}\right)$$

Se $\mathcal{H}_0$ e' $m \geq m_0$ si ribaltano le disuguaglianze e si sostituisce $q_{1-\alpha}$ con q_α.

TEST sulla MEDIA di un CAMPIONE GAUSSIANO con VARIANZA SCONOSCIUTA (T-TEST)

Test bilatero $\mathcal{H}_0)\ m = m_0$, $\mathcal{H}_1)\ m \neq m_0$

$$C = \left\{ |\bar{x} - m_0| > \frac{\bar{\sigma}}{\sqrt{n}} T_{(1-\frac{\alpha}{2}, n-1)} \right\}$$

$$\bar{\alpha} = 2\left(1 - F_{n-1}\left(\frac{\sqrt{n}}{\bar{\sigma}} |\bar{x} - m_0|\right)\right)$$

con F_{n-1} c.d.f. di $T(n)$.

TEST su un CAMPIONE di BERNOULLI

Test bilatero $\mathcal{H}_0)\ p = p_0$, $\mathcal{H}_1)\ p \neq p_0$

$$C = \left\{ \frac{\sqrt{n}\,|\bar{x} - p_0|}{\sqrt{p_0(1-p_0)}} > q_{1-\frac{\alpha}{2}} \right\}$$

$$\bar{\alpha} = 2\left(1 - \phi\left(\frac{\sqrt{n}\,|\hat{p} - p_0|}{\sqrt{p_0(1-p_0)}}\right)\right).$$

TEST sulla VARIANZA di un CAMPIONE GAUSSIANO

Test unilatero $\mathcal{H}_0)\ \sigma^2 \leq \sigma_0^2$, $\mathcal{H}_1)\ \sigma^2 > \sigma_0^2$

$$C = \left\{ \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma_0^2} > \chi^2_{(1-\alpha, n-1)} \right\}$$

$$\bar{\alpha} = 1 - G_{n-1}\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma_0^2} \right)$$

con G_{n-1} c.d.f. di chi quadro

$$\beta(\sigma) = G_{n-1}\left(\frac{\sigma_0^2}{\sigma^2} \cdot \chi^2_{(1-\alpha, n-1)} \right).$$