CAMPIONAMENTO

Il campionamento ci porta alla vera e propria teoria della statistica.
L'obiettivo è generalizzare l'esperimento singolo alla classe di tutti gli esperimenti simili, operando un'estensione dal particolare al generale, chiamata INFERENZA INDUTTIVA.
L'inferenza induttiva è un processo d'azzardo: non si possono fare generalizzazioni assolutamente certe; si possono fare inferenze incerte e misurare il grado di incertezza in termini di probabilità.

DEF.: La totalità degli elementi in esame, circa i quali si vogliono ottenere informazioni è detta POPOLAZIONE OBIETTIVO.

È poco pratico esaminare l'intera popolazione, ma si può esaminare una sua parte (campione) e fare inferenze sulla popolazione obiettivo.
Pb: scelta del campione.

DEF.: Se $X_1, \dots, X_n$ sono n v.a. con densità congiunta espressa nel seguente modo:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f(x_1) \dots f(x_n)$$

dove $f(\cdot)$ è la densità (comune) di ogni X_i, allora $X_1, \dots, X_n$ è detto CAMPIONE CASUALE di ampiezza n estratto da una popolazione con densità $f(\cdot)$.
Il campione casuale è detto POPOLAZIONE CAMPIONATA

N.B: i campionamenti sono pensati $\boxed{\text{indipendenti.}}$

<u>DEF.</u>: Sia $X_1, \dots, X_n$ un campione c. di ampiezza n. La distribuzione congiunta di $X_1, \dots, X_n$ data da

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f(x_1) \cdot \dots \cdot f(x_n)$$

è detta DISTRIBUZIONE CAMPIONARIA di $X_1, \dots, X_n$ estratto da $f(\cdot)$. (Se $X_1, \dots, X_n$ è un c.c. $\Rightarrow X_1, \dots, X_n$ sono indipendenti.)

UN TEMA CENTRALE DELLA STATISTICA è studiare una popolazione con densità $f(\cdot; \theta)$ di forma nota, ma che contiene un parametro incognito θ.

<u>Procedimento</u>: estrarre un campione casuale $X_1, \dots, X_n$ di ampiezza n dalla densità e stimare il parametro θ con il valore di una qualche funzione $t(x_1, \dots, x_n)$, e determinare quale funzione sia la migliore per stimarlo.

<u>DEF.</u>: Una STATISTICA è una funzione di variabili casuali $\underline{t}(x_1, \dots, x_n)$; quindi a suo volta una v.a. che <u>non</u> contiene alcun parametro incognito.

Esempi: Sono statistiche per un campione casuale $X_1, \dots X_n$ estratto da una densità $f(\cdot;\theta)$

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

$$\frac{1}{2}\left\{\min[X_1,\dots,X_n] + \max[X_1,\dots,X_n]\right\}$$

Importanti statistiche sono date da:

DEF.: Sia $X_1, \dots, X_n$ un c.c. dalla densità $f(\cdot)$.
Definiamo MOMENTO CAMPIONARIO r-ESIMO (ASSOLUTO) la quantità

$$M'_r = \frac{1}{n} \sum_{i=1}^{n} X_i^r$$

se $r = 1$ si ottiene la <u>media campionaria</u> $\bar{X}_n$.

Definiamo il MOMENTO CAMPIONARIO r-ESIMO RISPETTO a $\bar{X}_n$ la quantità

$$M_r = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^r$$

N.B.:

I momenti campionari rispecchiano i momenti della popolazione nel senso che il valore atteso di un momento campionario (assoluto) è uguale al corrispondente momento della popolazione.

TEOREMA 1 Sia $X_1, \dots, X_n$ c.c. estratto da una popolazione con densità $f(\cdot)$. Allora

$$E[M'_r] = \mu'_r \ (\text{se } \mu'_r \ \exists) \quad ; \quad \mathrm{var}[M'_r] = \frac{1}{n}\left[\mu'_{2r} - (\mu'_r)^2\right]$$

se $\exists\ \mu'_{2r}$

Infatti

$$E[M'_r] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^r\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i^r\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^r]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mu'_r = \frac{1}{n} n \mu'_r = \mu'_r$$

$$\mathrm{var}[M'_r] = \mathrm{var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^r\right] = \left(\frac{1}{n}\right)^2 \mathrm{var}\left[\sum_{i=1}^{n} X_i^r\right] =$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}[X_i^r] = \frac{1}{n^2}\sum_{i=1}^{n}\left\{E[X_i^{2r}] - (E[X_i^r])^2\right\} =$$

indipendenza

$$= \frac{1}{n}\left\{E[X^{2r}] - (E[X^r])^2\right\} = \frac{1}{n}\left[\mu'_{2r} - (\mu'_r)^2\right]$$

se $r = 1 \Rightarrow M'_1 = \bar{X}_n$ e

$$E[\bar{X}_n] = \mu \qquad \mathrm{var}[\bar{X}_n] = \frac{1}{n}\sigma^2$$

con μ e σ^2 media e varianza di $f(\cdot)$.

Riguardo al momento campionario r-esimo rispetto alla media campionaria consideriamo solo il caso

$$r = 2 \Rightarrow M_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

DEF.: Sia $X_1, \dots, X_n$ c.c. estratto da una popolazione con densità $f(\cdot)$. Definiamo VARIANZA CAMPIONARIA la quantità

$$(S_n^2) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \quad \text{per } n > 1$$

OSSERVAZIONE Si usa S^2 anziché M_2 come definizione di varianza campionaria (entrambe misurano la dispersione del campione) perché il valore atteso di S^2 è uguale alla varianza della popolazione.

TEOREMA 2 Sia $X_1, \ldots, X_n$ c. c. estratto da una popolazione con densità $f(\cdot)$. Si ha che:

$$E[S^2] = \sigma^2 \quad , \quad var[S^2] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right) \quad n>1$$

(no dim.)

N. B:

$$M_1 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\bar{X}_n} - \frac{1}{n} \cdot n\bar{X}_n \equiv 0$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{n}{n-1} \cdot \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}_{M_2} = \frac{n}{n-1} M_2$$

Riassumendo:

M'_r stima μ'_r ; $\bar{X}_n$ stima μ ; S^2 stima σ^2.

OSSERVAZIONE

Abbiamo visto che se $r=1$ $M'_1 = \bar{X}_n$.

In base al Th. 1 (per $r=1$) il valore di $\bar{X}_n$ serve per stimare μ; cioè $E[\bar{X}_n] = \mu$ prova che la media $\bar{X}_n$ è uguale al parametro μ da stimare, cioè la distribuzione di $\bar{X}_n$ è CENTRATA intorno a μ.

Mentre $var[\bar{X}_n] = \frac{1}{n}\sigma^2$ prova che la dispersione dei valori di $\bar{X}_n$ intorno a μ è piccola se l'ampiezza del campione è grande.

LEGGE DEI GRANDI NUMERI

Indicato $E[X]$ con μ nella densità $f(\cdot)$ se si usa solo un numero finito di valori di X (campione casuale di ampiezza n) si possono fare inferenze attendibili circa $E[X]$ che è la media di un numero infinito di valori di X? sì in base alla legge debole dei grandi numeri che si può dimostrare usando la diseguaglianza di Tchebycheff.

Tale legge stabilisce che si può determinare un intero positivo n tale che se si prende un c.c. di ampiezza $\geq n$ da una popolazione di densità $f(\cdot)$ (con $E[X]=\mu$) "la probabilità che la differenza tra la media campionaria $\bar{X}_n$ e μ sia minore di una quantità fissata piccola a piacere, è vicina a 1 quanto si vuole."

Per ogni $\varepsilon>0$ e $0<\delta<1$ $\exists\ n > \dfrac{\sigma^2}{\varepsilon^2\delta}$:

$$P[\,|\bar{X}_n-\mu|<\varepsilon\,]\geq 1-\delta$$

(con μ e σ^2 media e varianza di $f(\cdot)$)

Infatti ricordando Markov:

$$P[g(X) \geq r] \leq \frac{E[g(X)]}{r} \quad , \forall r > 0$$

e $g(x) \geq 0 \ \forall x \in \mathbb{R}$.

Formulazione analoga:

$$P[g(X) < r] \geq 1 - \frac{E[g(X)]}{r}$$

scelta $g(x) = (\bar{X}_n - \mu)^2$ ed $r = \varepsilon^2$

$$P[|\bar{X}_m - \mu| < \varepsilon] = P[(\bar{X}_m - \mu)^2 < \varepsilon^2] \geq 1 - \frac{E[(\bar{X}_n - \mu)^2]}{\varepsilon^2}$$

ma $E[(\bar{X}_m - \mu)^2] = var[\bar{X}_m]$ $\qquad var[Y] = E[(Y - E[Y])^2]$

e $var[\bar{X}_m] = \frac{\sigma^2}{m}$, $E[\bar{X}_m] = \mu$

pertanto

$$P[|\bar{X}_m - \mu| < \varepsilon] \geq 1 - \frac{1}{\varepsilon^2} \cdot \frac{\sigma^2}{m} \geq 1 - \delta$$

per $\delta > \frac{\sigma^2}{m\varepsilon^2}$ oppure $m > \frac{\sigma^2}{\delta \varepsilon^2}$.

Esempi

1) Data una popolazione con media incognita e varianza $\sigma^2 = 1$, calcolare la dimensione del campione estratto affinché sia almeno del 95% la probabilità che la media campionaria disti meno di 0.5 dalla media della popolazione.

$\sigma^2 = 1$

$\varepsilon = 0.5$

$\delta = 0.05$

$$P[|\bar{X}_m - \mu| < \varepsilon] \geq 1 - \delta \implies P[|\bar{X}_n - \mu| < 0.5] \geq 0.95$$

$$0.95 \leq P[(\bar{X}_n - \mu)^2 < (0.5)^2] \geq 1 - \underbrace{\frac{1}{\varepsilon^2}\frac{\sigma^2}{m}} \geq 1 - \delta$$

$$\delta > \frac{\sigma^2}{m\varepsilon^2} \sim m > \frac{\sigma^2}{\delta\varepsilon^2}$$

$$m > \frac{1}{(0.05)\cdot(0.5)^2} = 80$$

2) Quanto deve essere grande un campione per essere sicuri al 99% che $\bar{X}_m$ disti meno di (0.5σ) da μ?

$$0.99 = P[|\bar{X}_m - \mu| < \varepsilon] \geq 1 - \delta$$

$\varepsilon = 0.5\,\sigma$

$\delta = 0.01$

σ incognita

$$m > \frac{\sigma^2}{\delta\varepsilon^2} = \frac{\cancel{\sigma^2}}{(0.01)\cdot(0.5)^2\cancel{\sigma^2}} = 400$$

TEOREMA DEL LIMITE CENTRALE

Il teorema del limite centrale e la teoria della probabilità ci dicono come è approssimativamente distribuita $\bar{X}_n$.

Sia $f(\cdot)$ una densità con media μ e varianza FINITA σ^2. Sia $\bar{X}_n$ la media campionaria di un c.c. di ampiezza n estratto da una popolazione con densità f.

Sia Z_n definita da:

$$Z_n = \frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{var[\bar{X}_n]}} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Allora la distribuzione di Z_n tende alla distribuzione normale standard per $n \to \infty$.

Nulla è detto circa la forma della $f(\cdot)$.

$$Z_n \,\dot{\sim}\, N(0,1) \qquad \dot{\sim} = \text{approssimativamente}$$

$$(\sim \; = \text{esatta})$$

Pb del TLC: Quanto grande deve essere la numerosità del campione n affinché l'approssimazione sia valida?

Regola empirica $\rightarrow$ $n = 30$

Esempio

Sbarre di lunghezza data con f. di densità incognita.

E' nota $\sigma = 0.2 m$.

Prese n sbarre calcolare n in modo che la media campionaria $\bar{X}_n$ disti dal valore atteso μ per meno di 1 cm. con una probabilità > del 97%.

$1 cm = 0.01 m \Rightarrow \varepsilon = 0.01 m$

$\sigma = 0.2 m \Rightarrow \sigma^2 = 0.04 m^2$

con LGN $\quad P[|\bar{X}_n - \mu| < 0.01] > 1 - \dfrac{0.04}{n(0.01)^2} \qquad \delta > \dfrac{\sigma^2}{n\varepsilon^2}$

e $\quad P[|\bar{X}_n - \mu| < 0.01] > 0.97$

$\Rightarrow 1 - \dfrac{0.04}{n\, 10^{-4}} > 0.97 \Rightarrow n \geq \dfrac{4 \cdot 10^2}{0.03} = 1.\bar{3} \cdot 10^4 = \underline{13'333}$

con TLC $\quad |\bar{X}_n - \mu| < 0.01 \iff \dfrac{-0.01}{\frac{0.2}{\sqrt{n}}} < \underbrace{\dfrac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}}_{= Z_n} < \dfrac{0.01}{\frac{0.2}{\sqrt{n}}}$

$\iff |Z_n| < \dfrac{\sqrt{n}}{20}$

$P[|\bar{X}_n - \mu| < 0.01] > 0.97 \Rightarrow P[|Z_n| < \underbrace{\dfrac{\sqrt{n}}{20}}_{= z_\alpha}] > 0.97.$

$P[|Z_n| < z_\alpha] = (P[Z_n < z_\alpha] - 0.5) \cdot 2 = 2P[Z_n < z_\alpha] - 1$

$\Rightarrow P[Z_n < z_\alpha] > \dfrac{1 + 0.97}{2} = 0.985 \Rightarrow z_\alpha = 2.17.$

$\dfrac{\sqrt{n}}{20} > 2.17 \Rightarrow n > 1883,56 \quad \underline{n > 1884}$

CAMPIONAMENTO DA DISTRIBUZIONI NORMALI

Tra tutte le possibili funzioni di un c.c., la più semplice è la media campionaria $\bar{X}_n$.
Se il c.c. (v.i.i.d.) è estratto da una popolazione normale $N(\mu, \sigma^2)$ allora la distribuzione **ESATTA** di $\bar{X}_n$ è **NORMALE** $N\left(\mu, \frac{\sigma^2}{n}\right)$.

(vedi applicazioni del capitolo sulle v.c. congiunte).

N.B: se la popolazione ha f. di densità qlq vale il TLC.

ma se $f \sim N(\mu, \sigma^2) \Rightarrow \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

CONSEGUENZA

$$Z_n = \frac{\bar{X}_n - E[\bar{X}_n]}{\sqrt{var[\bar{X}_n]}} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Infatti

$$E[Z_n] = \frac{\sqrt{n}}{\sigma} E[\bar{X}_n - \mu] = \frac{\sqrt{n}}{\sigma}(\mu - \mu) \equiv 0$$

$$var[Z_n] = \frac{n}{\sigma^2} var[\bar{X}_n - \mu] = \frac{n}{\sigma^2} var[\bar{X}_n] = \frac{n}{\sigma^2} \cdot \frac{\sigma^2}{n} \equiv 1$$

Ricorda che se

$$X_i \sim N(\mu, \sigma^2) \Rightarrow Z_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$$

Posso considerare

$$Z_i^2 = \frac{(X_i - \mu)^2}{\sigma^2}$$ cioè il quadrato di normale standard

Definisco la funzione

$$U := \sum_{i=1}^{n} Z_i^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$ SOMMA DI QUADRATI DI NORMALI STANDARD

Si può provare

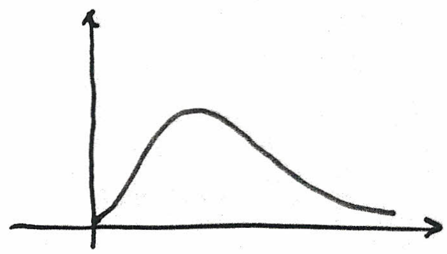
TEOREMA 3 : $U \sim \chi^2_n$ CHI-QUADRO con n gradi di libertà

La somma di quadrati di normali standard indip. ha una distribuzione χ^2 con gradi di libertà uguali al numero dei termini dell'addizione.

N.B. Ricorda che χ^2 è una funzione GAMMA dove $\lambda = \frac{1}{2}$, $r = \frac{n}{2}$ (n) chiamato grado di libertà.

$$f(x, r, \lambda) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}$$

$$\chi^2_n = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}}{\Gamma\left(\frac{n}{2}\right)}$$



Posso definire (in analogia a $\bar{X}_n$)

$$\bar{Z}_n = \frac{1}{n} \sum_{i=1}^{n} Z_i = \frac{1}{n} \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sigma} = \frac{1}{\sigma}\left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)\right]$$

$$= \frac{1}{\sigma}\left[\frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \sum_{i=1}^{n} \mu\right] = \frac{1}{\sigma}\left(\bar{X}_n - \frac{1}{n} \cdot n\mu\right)$$

$$= \frac{\bar{X}_n - \mu}{\sigma}$$

Si può provare che

$$\bar{Z}_n \sim N\left(0, \frac{1}{n}\right)$$

$$E[\bar{Z}_n] = E\left[\frac{1}{\sigma}(\bar{X}_n - \mu)\right] = \frac{1}{\sigma} E[\bar{X}_n - \mu] = \frac{1}{\sigma}(\mu - \mu) \equiv 0$$

$$var[\bar{Z}_n] = \frac{1}{\sigma^2} var[\bar{X}_n - \mu] = \frac{1}{\sigma^2} var[\bar{X}_n] = \frac{1}{\sigma^2} \cdot \frac{\sigma^2}{n} = \frac{1}{n}$$

Riprendiamo la funzione $Z_n \sim N(0,1)$ e si consideri il quadrato $(Z_n)^2$. In base al Teorema 3 si ha che
$(Z_n)^2 \sim \chi^2_1$

ma anche

$$(Z_n)^2 = \frac{(\bar{X}_n - \mu)^2}{\frac{\sigma^2}{n}} = \frac{n}{\sigma^2}(\bar{X}_n - \mu)^2 = n\left[\frac{\bar{X}_n - \mu}{\sigma}\right]^2 = n(\bar{Z}_n)^2$$

Consideriamo ora la funzione

$$V = U - Z_n^2$$

Si può provare

<u>TEOREMA 4</u> : $V \sim \chi^2_{n-1}$

$U \sim \chi^2_n$ e $Z_n^2 \sim \chi^2_1$ (per differenza).

Inoltre

$$V = \sum_{i=1}^n Z_i^2 - Z_n^2 = \sum_{i=1}^n Z_i^2 - n\bar{Z}_n^2 \overset{\text{CONTI}}{=} \sum_{i=1}^n (Z_i - \bar{Z}_n)^2$$

$$\underset{\text{CONTI }(\circ)}{=} \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{\sigma^2}(n-1) S^2$$

poichè

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

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

$$= \frac{1}{n-1} \Big[\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n \underbrace{\sum_{i=1}^{n} X_i}_{= n\overline{X}_n} + n\overline{X}_n^2 \Big]$$

$$= \frac{1}{n-1} \Big[\sum_{i=1}^{n} X_i^2 - n\overline{X}_n^2 \Big]$$

CONTI (•) $\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2 = \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}_n}{\sigma}\right)^2 = \sum_{i=1}^{n} \left[\frac{(X_i - \mu)}{\sigma} - \frac{(\overline{X}_n - \mu)}{\sigma}\right]^2$

$$= \sum_{i=1}^{n} (Z_i - \overline{Z}_n)^2$$

OSSERVAZIONE

Col termine "grado di libertà" si indica il n° dei quadrati indipendenti nella sommatoria.

Nel thm 3 sono n

Nel thm. 4 sono $n-1$ perchè vale la relazione di vincolo

$$\sum_{i=1}^{n} (Z_i - \overline{Z}_n) \equiv 0.$$

(cioè abbassa il grado di libertà)

Infine date $\overline{Z}_n \sim N(0,1)$ e $V \sim \chi^2_{n-1}$ con $\overline{Z}_n$ e V indipendenti si definisce la funzione.

$$T_{n-1} := \frac{\overline{Z}_n}{\sqrt{\frac{V}{n-1}}}$$

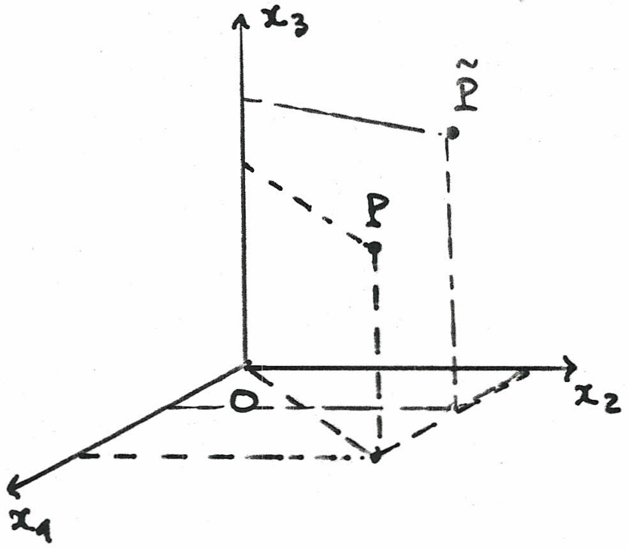
$$T_{n-1} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \cdot \sqrt{\frac{n-1}{\frac{(n-1)S^2}{\sigma^2}}} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \cdot \frac{\sigma}{S} = \frac{\bar{X}_n - \mu}{\frac{S}{\sqrt{n}}}$$

Si può provare

<u>TEOREMA 5</u> : $T_{n-1} \sim t_{n-1}$ t di STUDENT con $(n-1)$ gradi di libertà

Esempio

Si vuole localizzare un oggetto nello spazio, ma la misurazione porta un errore, in ognuna delle 3 direzioni, che è una v.a. normale $N(\mu=0, \sigma=2\,m.)$ Supponendo i 3 errori indipendenti, determinare la probabilità che la distanza tra posizione misurata e quella reale sia maggiore di 3 metri



$P(x_1, x_2, x_3)$ REALE

$\tilde{P}(y_1, y_2, y_3)$ MISURATA

$y_1 = x_1 + X_1$

$y_2 = x_2 + X_2$ $\quad X_i \quad i=1,2,3$

$y_3 = x_3 + X_3$ $\quad$ ERRORI

D = distanza tra P e $\tilde{P}$.

$$D^2 = \overline{P\tilde{P}}^2 = (y_1-x_1)^2+(y_2-x_2)^2+(y_3-x_3)^2 = X_1^2+X_2^2+X_3^2$$

$$X_i \sim N(0,2)$$

$$Z_i = \frac{X_i-\mu}{\sigma} = \frac{X_i}{2} \sim N(0,1)$$

$$Y = \sum_{i=1}^{3} Z_i^2 \quad \text{somma di quadrati di normali standard}$$

PER IL TH. 3 $\quad Y \sim \chi^2_{n=3}$

$$P(D>3) = P(D^2>9) = P\left(\frac{X_1^2}{4}+\frac{X_2^2}{4}+\frac{X_3^2}{4} > \frac{9}{4}\right) = P\left(Z_1^2+Z_2^2+Z_3^2 > \frac{9}{4}\right)$$

$$= P\left(Y>\frac{9}{4}\right) = 1-\underbrace{P\left(Y\le\frac{9}{4}\right)}_{\substack{\Downarrow \\ \text{da programma} \\ \simeq 0.4778}} \simeq 0.5222$$

Se la localizzazione avviene nel piano

$\Rightarrow D^2 = X_1^2 + X_2^2$

$\Rightarrow Y \sim \chi^2_{m=2}$

Ma $\chi^2_{m=2} = \Gamma\left(\alpha = \frac{m}{2} = 1, \lambda = \frac{1}{2}\right) = \Gamma\left(1, \frac{1}{2}\right) \equiv \exp\left(\lambda = \frac{1}{2}\right)$

Perciò Y è exp.

$$f(y) = \lambda e^{-\lambda y}$$

$$P(Y \leq y) = F(y) = 1 - e^{-\lambda y}$$

$$P(D > 3) = P(D^2 > 9) = P\left(Y > \frac{9}{4}\right) = 1 - P\left(Y \leq \frac{9}{4}\right)$$

$$= e^{-\lambda y}\Big|_{\lambda = \frac{1}{2}, y = \frac{9}{4}} = e^{-9/8} \approx 0.3247.$$

STIMA DI PARAMETRI

Quando si fa della probabilità si suppone che le distribuzioni siano completamente note, in statistica invece si fa dell'inferenza su parametri sconosciuti utilizzando i dati osservati.

Nell'inferenza statistica sono presenti due problemi importanti: la STIMA e la VERIFICA DI IPOTESI.

Un tipo di stima è la STIMA PUNTUALE che consiste nel trovare una statistica $t(x_1, \dots, x_n)$ detta STIMATORE PUNTUALE che permetta di stimare il parametro incognito "ϑ" o una sua funzione "$\tau(\vartheta)$" quando l'incognita ϑ della funzione di densità $f(\cdot, \vartheta)$ appartiene ad un insieme di parametri.

Il secondo tipo di stima è la STIMA PER INTERVALLI che consiste nel definire due statistiche $t_1(X_1, \dots, X_n)$, $t_2(X_1, \dots, X_n)$ con $t_1 < t_2$ cosicché

(t_1, t_2) costituisca un intervallo ~~per~~ di valori plausibili per θ o $\tau(\theta)$ per il quale si può determinare la probabilità che $\tau(\theta)$ o θ vi appartenga.

STIMA PUNTUALE DI PARAMETRI

Gli stimatori sono v.a. Il valore deterministico assunto da uno stimatore si chiama STIMA.

PROBLEMA: individuare la forma opportuna dello stimatore e calcolare la sua distribuzione.

- trovare una statistica da usare come stimatore
- scegliere criteri per definire e ottenere uno stimatore "ottimale" fra i molti possibili

Le proprietà che può possedere uno stimatore che verranno discusse sono:

- CORRETTEZZA O NON DISTORSIONE
- CONSISTENZA

Per determinare degli stimatori puntuali ci sono vari metodi.

METODO DEI MOMENTI

Dato $f_X(x; \theta_1, .., \theta_k)$ una f. di densità con

k parametri incogniti.

$$\mu'_r = E[X^r] = \mu'_r(\theta_1, .., \theta_k)$$

Dato un c.c. $X_1, .., X_n$

$$M'_J = \frac{1}{n} \sum_{i=1}^{n} X_i^J \qquad J = 1, .., k$$

Costruiamo il sistema di k equazioni

$$(\bullet) \quad M'_J = \mu'_J(\theta_1, .., \theta_k) \quad J = 1, .., k$$

nelle k incognite $\theta_1, .., \theta_k$.

La soluzione unica di $(\bullet)$ $(\bar{\theta}_1, .., \bar{\theta}_k)$ sarà lo stimatore cercato, ottenuto sostituendo ai momenti della popolazione i momenti campionari.

<u>Esempi</u>

1) $X_1, .., X_n$ c.c. estratto da una popolazione con densità normale (μ, σ^2).

$$(\theta_1, \theta_2) = (\mu, \sigma)$$

Poichè

$$\mu = \mu'_1$$

$$\sigma^2 = \mu'_2 - (\mu'_1)^2$$

$$\begin{cases} \mu_1' = \mu \\ \mu_2' = \sigma^2 + (\mu_1')^2 = \sigma^2 + \mu^2 \end{cases}$$

Le equazioni dei momenti:

$$\begin{cases} M_1' = \mu_1' = \mu_1'(\mu, \sigma) = \mu \\ M_2' = \mu_2' = \mu_2'(\mu, \sigma) = \sigma^2 + \mu^2 \end{cases} \qquad \begin{matrix} \theta_1 = \mu \\ \theta_2 = \sigma \end{matrix}$$

ma

$$M_1' \doteq \bar{X}_n$$

$$M_2' \doteq \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

perciò

$$\begin{cases} \bar{\Theta}_1 = \bar{X}_n \quad \text{è lo stimatore di } \mu \\ \bar{\Theta}_2 = \sqrt{M_2' - \bar{X}_n^2} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}_n^2} \end{cases}$$

$$= \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2} \quad \text{è lo stimatore di } \sigma$$

$$= \sqrt{M_2}$$

2) $X_1, \dots, X_n$ c.c. estratto da una popolazione con densità esponenziale $f(x; \theta) = \theta e^{-\theta x}$, $x \geq 0$.

Si ha

$$M_1' = \mu_1' = \mu_1'(\theta) = \frac{1}{\theta} \qquad \mu_1' = \mu = E[X]$$

e quindi

$$\bar{\Theta} = \frac{1}{M_1'} = \frac{1}{\bar{X}_n}$$

METODO DELLA MASSIMA VEROSIMIGLIANZA

Siano $X_1, \dots, X_n$ n. v.a. indipendenti
Sia $f(x_1, \dots, x_n; \theta)$ la f. di densità congiunta nota a meno di un parametro incognito θ.

obiettivo: stimare θ partendo dai valori osservati $x_1, \dots, x_n$.

Interpretiamo $f(x_1, \dots, x_n; \theta)$ come la verosimiglianza (o plausibilità, o credibilità) che si realizzi la n-pla di dati $x_1, \dots, x_n$ quando θ è il vero valore assunto dal parametro.

Pare ragionevole adottare come stima di θ quel valore $\hat{\theta}$ che rende massima la verosimiglianza, cioè $f(x_1, \dots, x_n; \theta)$ quando i dati osservati sono $x_1, \dots, x_n$.

La funzione $f(x_1, \dots, x_n; \theta)$ è detta FUNZIONE DI VERO=SIMIGLIANZA.
Per l'indipendenza si ha che:

$$f(x_1, \dots, x_n; \theta) = f_{X_1}(x_1; \theta) \cdot \dots \cdot f_{X_n}(x_n; \theta)$$

Lo stimatore di max. ver. è pertanto la soluzione dell'equazione

$$\frac{d}{d\theta} f(x_1, \dots, x_n; \theta) = 0$$

Nel calcolare il valore $\hat{\theta}$ di θ che massimizza f si usa spesso la funzione $\log f(x_1, \dots, x_n; \theta)$, poichè

f e $\log f$ assumono il massimo in corrispondenza dello stesso valore di θ.

Infatti

$$\frac{d}{d\theta}\log f(\theta) = \frac{1}{f(\theta)}\, f'(\theta) \quad = 0 \iff f'(\hat{\theta}) = 0$$

$$\frac{d^2}{d\theta^2}\log f(\theta) = \frac{f''(\theta)\cdot f(\theta) - f'(\theta)^2}{f^2(\theta)}$$

$$\left.\frac{d^2}{d\theta^2}\log f(\theta)\right|_{\theta=\hat{\theta}} = \frac{f''(\hat{\theta})}{f(\hat{\theta})} < 0 \iff f''(\hat{\theta}) < 0$$

A volte è più facile operare con $\log f$ che con f.

<u>Esempi</u>

1) Dato un c.c. di ampiezza n estratto da una popolazione caratterizzata da una f. di densità di Bernoulli, cioè

$$f(x,p) = p^{x} q^{(1-x)}$$

$0 \leq p \leq 1$ e $q = 1-p$.

La funzione di verosimiglianza è:

$$f(x_1,\dots,x_n;p) = \underbrace{(p^{x_1} q^{1-x_1})}_{f_{x_1}(x_1;p)} \cdot \dots \cdot \underbrace{(p^{x_n} q^{1-x_n})}_{f_{x_n}(x_n;p)}$$

$$f(x_1, \dots, x_m; p) = p^{x_1 + \dots + x_n} \, q^{n-(x_1+\dots+x_n)}$$

Posto $y = \sum_{i=1}^{n} x_i$

$$f(y; p) = p^y \, q^{n-y}$$

$$\log f(p) = y \log p + (n-y) \log (1-p)$$

$$\frac{d}{dp} \log f(p) = y \cdot \frac{1}{p} + (n-y) \cdot \frac{1}{1-p} \cdot (-1) = \frac{y - np}{p(1-p)}$$

$$\frac{d}{dp} \log f(p) = 0 \quad \text{se} \quad \underline{\hat{p}} = \frac{y}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i = \underline{\underline{\bar{x}_n}}$$

$$\frac{d^2}{dp^2} \log f(p) = -y \cdot \frac{1}{p^2} - (n-y) \cdot \frac{1}{(1-p)^2}$$

$$\left. \frac{d^2}{dp^2} \log f(p) \right|_{\hat{p} = \frac{y}{n}} = -\frac{n^3}{y(n-y)} < 0 \quad \text{sempre vero}$$

$$0 \leq y = x_1 + \dots + x_n \leq n$$

$\Rightarrow$ $\hat{p} = \bar{x}_n$ è la stima di massima verosimiglianza per p.

PROPRIETÀ DEGLI STIMATORI PUNTUALI

Esistono stimatori che siano in qualche modo migliori di altri?

Definiamo ora alcune proprietà, che uno stimatore può possedere o no, utili per decidere se uno stimatore è da preferirsi ad un altro.

CORRETTEZZA

Una misura utile, anche se rozza, della bontà di uno stimatore di $\tau(\theta)$ è l'ERRORE QUADRATICO MEDIO (= MSE)

$$MSE[T](\theta) = E\left[[T-\tau(\theta)]^2\right]$$

dove $T = t(X_1, .., X_n)$ è uno stimatore di $\tau(\theta)$.

Esso misura la dispersione dei valori di T rispetto a $\tau(\theta)$ (come la varianza di una v.a. X misura la sua dispersione attorno alla media)

<u>DEF.</u> $T = t(X_1, .., X_n)$ è uno stimatore CORRETTO o NON DISTORTO di $\tau(\theta)$ se e solo se, $\forall \theta \in \Theta$:

$$E[T] = \tau(\theta)$$

<u>N.B</u>: Uno stimatore con MSE minimo è raro.

Speranza: restringendoci alla classe degli stimatori non distorti trovare quello con MSE minimo.

DEF.: Si definisce DISTORSIONE di T la quantità

$$D[T](\theta) = \tau(\theta) - E[T] \quad (\gtreqless 0)$$

Se T è non distorto $\Rightarrow D[T]=0$.

Vale la seguente relazione:

$$MSE[T](\theta) = \mathrm{var}[T] + (D[T])^2$$

Infatti:

$$MSE[T] = E[(T-\tau)^2] = E\Big[\big[\overbrace{T - E[T]}^{a} + \overbrace{E[T] - \tau}^{b}\big]^2\Big]$$

$$= E[\overbrace{(T-E[T])^2}^{a^2}] + 2E[\underbrace{(T-E[T])}_{a}\underbrace{(E[T]-\tau)}_{b}] +$$

$$+ E[\overbrace{(\tau - E[T])^2}^{b^2}]$$

ma $(E[T]-\tau)$ non dipende dalle x_i e quindi

$$E[(T-E[T])\underbrace{(E[T]-\tau)}_{\text{costante}}] = (E[T]-\tau)\,\underbrace{E[T-E[T]]}_{=\,0} = 0$$

$$E[\underbrace{(\tau-E[T])^2}_{\text{costante}}] = (\tau-E[T])^2 = D^2[T]$$

e quindi

$$MSE[T](\theta) = \underline{E[(T-E[T])^2]} + D^2[T](\theta)$$

$$= \underline{\mathrm{var}[T]} + D^2[T](\theta).$$

- se $\underline{T}$ è <u>non distorto</u> $\underline{MSE[T](\theta) = \mathrm{var}[T]}$.

Esempio Dato un c.c. estratto da una popolazione con f. di densità normale (μ, σ^2)

Abbiamo già ricavato che gli stimatori per μ e per σ^2 sono:

$$\mu \longrightarrow \bar{X}_n$$

$$\sigma^2 \longrightarrow M_2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

Poichè $E[\bar{X}_n] = \mu \Rightarrow \bar{X}_n$ è non distorto

Inoltre

$$MSE[\bar{X}_n] = E[(\bar{X}_n - \mu)^2] = var[\bar{X}_n] = \frac{\sigma^2}{n}$$

Invece

$$E[M_2] = \frac{n-1}{n} E[S^2] = \frac{n-1}{n} \sigma^2 \neq \sigma^2 \Rightarrow M_2 \text{ è distorto}$$

$E[S^2] = \sigma^2$ (th. 2)

Cercare uno stimatore con MSE minimo tra quelli non distorti equivale a cercare uno stimatore a varianza minima nella stessa classe.

Un limite inferiore della varianza di stimatori non distorti è dato dal seguente teorema

DISUGUAGLIANZA DI RAO-CRAMÈR

Dato un c.c. $X_1, \dots, X_n$ estratto da $f(\cdot;\theta)$ con $\theta \in \Theta \subseteq \mathbb{R}$ (Θ intervallo) (θ unid.) e T uno stimatore non distorto di $\tau(\theta)$, si ha:

$$\operatorname{var}[T] \geq \frac{[\tau'(\theta)]^2}{n E\left[\left(\frac{1}{f}\frac{\partial}{\partial\theta} f\right)^2\right]}$$

$$\frac{1}{f}\frac{\partial f}{\partial\theta} = \frac{\partial}{\partial\theta}(\log f)$$

Esempio. Dato $X_1, \dots, X_n$ estratto da

$$f(x;\theta) = \theta e^{-\theta x} \qquad x \geq 0$$

preso $\tau(\theta) \equiv \theta$

$\tau'(\theta) = 1$

$$\operatorname{var}[T] \geq \frac{1}{n E\left[\left(\frac{1}{f}\frac{\partial}{\partial\theta} f\right)^2\right]}$$

$$\frac{\partial f}{\partial\theta} = e^{-\theta x} - \theta x e^{-\theta x} = (1-\theta x) e^{-\theta x}$$

$$\left(\frac{1}{f}\frac{\partial f}{\partial\theta}\right) = \frac{1}{\theta} - x$$

$$E\left[\left(\frac{1}{\theta}-X\right)^2\right] = E\left[\left(X-\frac{1}{\theta}\right)^2\right] \underset{\substack{\uparrow \\ E[X]=\frac{1}{\theta}}}{=} var[X] = \frac{1}{\theta^2}$$

$$\Rightarrow var[T] \geq \frac{\theta^2}{n}$$

CONSISTENZA

Poichè uno stimatore di $\tau(\theta)$ dipende dal numero di campionamenti enunciamo una proprietà definita in termini di ampiezza crescente del campione.

<u>DEF.</u>: Uno stimatore T_n di $\tau(\theta)$ è detto CONSISTENTE (in media quadratica) se e solo se $\forall \theta \in \Theta$:

$$\lim_{n \to \infty} E\left[\left[T_n - \tau(\theta)\right]^2\right] = 0$$

Poichè $E\left[\left[T_n - \tau(\theta)\right]^2\right] = MSE[T_n]$ e $MSE[T_n]$ si scrive

$$MSE[T_n] = D^2[T_n] + var[T_n]$$

si ha che

$$var[T_n] \underset{n \to \infty}{\longrightarrow} 0$$

$$D[T_n] \underset{n \to \infty}{\longrightarrow} 0$$

Esempio

Abbiamo già visto che per un c.c. estratto da una popolazione con f. di densità normale (μ, σ^2)

$\bar{X}_n$ è uno stimatore non distorto per μ.

$$\lim_{n\to\infty} E[(\bar{X}_n-\mu)^2] = \lim_{n\to\infty} var[\bar{X}_n] = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

$\Rightarrow$ $\bar{X}_n$ è uno stimatore consistente.

Esempio 1

Campione casuale estratto da una popolazione con funzione di densità rettangolare su $[0, \theta]$.

1) stimatore di θ col metodo dei momenti

2) correttezza

3) consistenza.

$$f(x,\theta) = \begin{cases} \frac{1}{\theta} & 0 \le x \le \theta \\ 0 & \text{altrove} \end{cases}$$

$$E[X] = \frac{\theta}{2} \; ; \; var[X] = \frac{\theta^2}{12}$$

1) $\mu'_r = E[X^r]$

$$M'_r = \frac{1}{n} \sum_{i=1}^{n} X_i^r$$

$$\mu'_1 = \mu'_1(\theta)$$

$$M'_1 = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}_n \quad \Rightarrow \quad \underline{E[X] = \bar{X}_n}$$

$$\Downarrow$$

$$\mu'_1 = E[X] \qquad \frac{\theta}{2} = \bar{X}_n \quad \Rightarrow \quad \boxed{\hat{\theta} = 2\bar{X}_n}$$

2) $E[\hat{\theta}] \stackrel{?}{=} \theta$ $\qquad E[\bar{X}_n] = \mu$

$$E[\hat{\theta}] = E[2\bar{X}_n] = 2E[\bar{X}_n] = 2\frac{\theta}{2} = \theta \Rightarrow \hat{\theta} \text{ CORRETTO}$$

3) $\hat{\theta}$ è consistente se $MSE(T_n) \to 0$

$$MSE(T_n) = var(T_n) + \underset{\overset{|||}{0}}{D^2(T_n)} = var[2\bar{X}_n] = 4\,var[\bar{X}_n]$$

$$= 4\frac{\sigma^2}{n} = \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n} \xrightarrow[n\to\infty]{} 0 \Rightarrow \hat{\theta} \text{ CONSISTENTE}$$

Esempio 2

C.C. estratto da una popolazione con f. di densità
$N(\mu, \sigma^2)$.

Dato $T_n = \frac{X_1 + X_n}{2}$, dire se è corretto e consistente.
per μ.

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$E[X] = \mu$; $var[X] = \sigma^2$

1) T_n è corretto?

$$E[T_n] \overset{?}{=} \mu$$

$$E[T_n] = \frac{E[X_1 + X_n]}{2} = 2\frac{E[X]}{2} = \mu \Rightarrow T_n \text{ è corretto}$$

2) T_n è consistente?

$$\lim_{n \to \infty} E[(T_n - \mu)^2] \overset{?}{=} 0$$

$$MSE(T_n) = E[(T_n-\mu)^2] = var[T_n] + \underbrace{D^2(T_n)}_{= 0} = var[T_n]$$

$$= var\left[\frac{X_1 + X_n}{2}\right] = \frac{1}{4} var[X_1 + X_n]$$

$$= \frac{1}{4}\{\underbrace{var[X_1]}_{= var[X]} + \underbrace{var[X_n]}_{= var[X]} + 2\underbrace{cov[X_1, X_n]}_{= 0}\}$$

perchè X_i IID perchè X_i indip.

$$= \frac{1}{4} \cdot 2\, var[X] = \frac{var[X]}{2} = \frac{\sigma^2}{2} \neq 0 \Rightarrow T_n \text{ NON}$$

È consistente

Esempio 5

C.C. estratto da una popolazione con funzione di densità di Poisson: $P(\lambda)$

1) det. lo stimatore di max. verosimigliante per λ.

2) correttezza

3) consistenza

$$f(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$E[X] = \lambda$; $var[X] = \lambda$.

1) $$f(x_1, \dots, x_m, \lambda) = \frac{\lambda^{x_1} e^{-\lambda}}{x_1!} \cdots \frac{\lambda^{x_m} e^{-\lambda}}{x_m!} = \frac{e^{-m\lambda} \lambda^{x_1 + \dots + x_m}}{x_1! \dots x_m!}$$

$$= \frac{e^{-m\lambda} \lambda^{\sum_{i=1}^m x_i}}{x_1! \dots x_m!}$$

$$\log f = \log e^{-m\lambda} + \log \lambda^{\sum x_i} - \log x_1! - \dots - \log x_m!$$

$$= -m\lambda + \sum_{i=1}^m x_i \log\lambda - \log x_1! - \dots - \log x_m!$$

$$\begin{cases} \frac{d}{d\lambda} \log f = -m + \frac{1}{\lambda} \sum_{i=1}^m x_i = 0 \Rightarrow \hat{\lambda} = \bar{x}_m \\ \frac{d^2}{d\lambda^2} \log f \Big|_{\hat\lambda} = -\frac{1}{\lambda^2} \sum x_i \Big|_{\hat\lambda} = -\frac{n}{\bar{x}_n} < 0 \end{cases}$$ stima di max verosimigliante

2) $E[\hat{\lambda}] = E[\bar{X}_m] = \mu = E[X] = \lambda \Rightarrow \hat{\lambda}$ corretto

3) $$MSE[\hat{\lambda}] = var[\hat{\lambda}] + \underbrace{D^2[\hat{\lambda}]}_{=0} = var[\bar{X}_m] = \frac{\sigma^2}{m} = \frac{\lambda}{m} \xrightarrow[n\to\infty]{} 0$$

$\Rightarrow \hat{\lambda}$ consistente

Esempio 4

Dato

$$f(x,\vartheta) = \begin{cases} c\, x\, e^{-\frac{x^2}{\vartheta}} & x \geq 0 \\ 0 & \text{altrove} \end{cases}$$

1. det. c affinchè f sia una f. di densità
2. det. la f. di max verosimiglianza
3. stimare ϑ col metodo della max. ver.
4. stimare ϑ col metodo dei momenti
5. verificare la correttezza o meno degli stimatori ottenuti al punto 3) e al punto 4).
6. verificare l' efficienza (= varianza minima).

1) $c \geq 0$

$$1 = \int_0^\infty c\, x\, e^{-\frac{x^2}{\vartheta}}\, dx = -c\frac{\vartheta}{2}\int_0^\infty -2\frac{x}{\vartheta}\, e^{-\frac{x^2}{\vartheta}}\, dx$$

$$= -c\frac{\vartheta}{2}\left[e^{-\frac{x^2}{\vartheta}}\right]_0^\infty = c\frac{\vartheta}{2}$$

$$\Rightarrow \boxed{c = \frac{2}{\vartheta}}$$

2) c.c. estratto da una popolazione con $f = \frac{2}{\vartheta}x\, e^{-\frac{x^2}{\vartheta}}$, $x \geq 0$

$$f(x_1,\dots,x_n,\vartheta) = f(x_1,\vartheta)\cdot\dots\cdot f(x_n,\vartheta)$$

$$= \frac{2}{\vartheta}x_1\, e^{-\frac{x_1^2}{\vartheta}} \cdots \frac{2}{\vartheta}x_n\, e^{-\frac{x_n^2}{\vartheta}}$$

$$= \frac{2^n}{\vartheta^n}x_1\cdots x_n\, e^{-\frac{1}{\vartheta}(x_1^2+\dots+x_n^2)}$$

$$= 2^n\vartheta^{-n}x_1\cdots x_n\, e^{-\frac{1}{\vartheta}\sum_{i=1}^n x_i^2}$$

3)

$$\log f = \log 2^n + \log \theta^{-n} + \log x_1 + \dots + \log x_n - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2$$

$$= n \log 2 - n \log \theta + \log x_1 + \dots + \log x_n - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2$$

$$\frac{d}{d\theta}\log f = -n\cdot\frac{1}{\theta} - \left(-\frac{1}{\theta^2}\right)\sum_{i=1}^{n} x_i^2 = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2$$

$$\frac{d}{d\theta}\log f = 0 \Rightarrow \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \Rightarrow \boxed{\hat{\theta} = M_2'}$$

$$\frac{d^2}{d\theta^2}\log f = -n\cdot\left(-\frac{1}{\theta^2}\right) - \frac{2}{\theta^3}\sum_{i=1}^{n} x_i^2 = \frac{1}{\theta^2}\left[n - \frac{2}{\theta}\sum_{i=1}^{n} x_i^2\right]$$

$$\left.\frac{d^2}{d\theta^2}\log f\right|_{\theta=\hat{\theta}} = \frac{1}{\hat{\theta}^2}\left[n - 2\sum_{i=1}^{n} x_i^2 \cdot \frac{n}{\sum_{i=1}^{n} x_i^2}\right] = -\frac{n}{\hat{\theta}^2} < 0$$

$\Rightarrow \hat{\theta}$ pto di max (ok)

4) $\mu_r' = M_r'$ dove $\mu_r' = E[X^r]$ $M_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r$

Essendo una sola l'incognita (θ)

$$\mu_1' = \mu_1'(\theta)$$

$$M_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

ma

$$\mu_1' = E[X] \quad \Rightarrow \quad E[X] = \bar{X}_n$$

Ora $E[X]$ deve essere calcolato

$$E[X] = \int_0^\infty (-x)\left(\frac{-2}{\theta}x\right) e^{-\frac{x^2}{\theta}}\, dx$$

Integro per parti:

$$= \Big[\underbrace{(-x)\, e^{-\frac{x^2}{\theta}}}_{=0}\Big]_0^\infty - \int_0^\infty (-1)\, e^{-\frac{x^2}{\theta}}\, dx = \int_0^\infty e^{-\frac{x^2}{\theta}}\, dx \underset{(\bullet)}{=}$$

applicando Hospital

Risultato di Analisi

$$\int_{-\infty}^{+\infty} e^{-x^2}\,dx = \sqrt{\pi}$$

ovvero

$$\frac{1}{\sqrt{\pi}}\int_{-\infty}^{+\infty} e^{-x^2}\,dx = 1 \Rightarrow \frac{1}{\sqrt{\pi}}\int_{0}^{\infty} e^{-x^2}\,dx = \frac{1}{2} \quad (*)$$

$$\Downarrow$$

$$\int_{0}^{\infty} e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}$$

Nel nostro caso poniamo:

$$\frac{x}{\sqrt{\theta}} = y \Rightarrow dx = \sqrt{\theta}\,dy$$

$$x = 0 \Rightarrow y = 0$$

$$x = \infty \Rightarrow y = \infty$$

$$\overset{(\bullet)}{=} \int_{0}^{\infty} e^{-y^2}\sqrt{\theta}\,dy \overset{(*)}{=} \sqrt{\theta}\cdot\frac{\sqrt{\pi}}{2}$$

Perciò $E[X] = \frac{\sqrt{\pi\theta}}{2}$

da cui

$$\bar{X}_n = \frac{\sqrt{\pi\theta}}{2} \Rightarrow \boxed{\hat{\theta} = \frac{4}{\pi}\bar{X}_n^2}$$

5) Uno stimatore T è corretto se $E[T] = \theta$

Nel nostro caso $E[\hat{\theta}] = \theta$.

Per lo stimatore di max. ver.

$$E[\hat{\theta}] = E[M_2'] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] = \frac{1}{n}\sum_{i=1}^{n} E[X_i^2]$$

ma X_i sono i.i.d.

$$= \frac{1}{n}\cdot n\,E[X^2] = E[X^2]$$

Devo pertanto calcolare $E[X^2]$.

$$E[X^2] = \int_0^\infty (-x^2)\cdot\left(-\frac{2}{\theta}x\right) e^{-\frac{x^2}{\theta}}\,dx$$

per parti

$$= \Big[\underbrace{-x^2 e^{-\frac{x^2}{\theta}}}_{=0}\Big]_0^\infty - \int_0^\infty e^{-\frac{x^2}{\theta}}\cdot(-2x)\,dx = -\theta\int_0^\infty -\frac{2x}{\theta} e^{-\frac{x^2}{\theta}}\,dx$$

applicando Hospital due volte

$$= (-\theta)\left[e^{-\frac{x^2}{\theta}}\right]_0^\infty = \theta \quad (\text{ok})$$

$\hat{\theta} = M_2'$ è corretto.

Per lo stimatore col metodo dei momenti

$$E[\hat{\theta}] = E\left[\frac{4}{\pi}\overline{X}_n^2\right] = \frac{4}{\pi}E[\overline{X}_n^2] \overset{(\cdot)}{=}$$

$$\overline{X}_m^2 = \overline{X}_m\cdot\overline{X}_m = \left(\frac{1}{m}\sum_{i=1}^m X_i\right)\left(\frac{1}{m}\sum_{J=1}^m X_J\right) =$$

$$= \frac{1}{m^2}\left[\sum_{\substack{i=1\\ i=J}}^m X_i^2 + 2\sum_{i<J} X_i X_J\right]$$

$$\overset{(\cdot)}{=} \frac{4}{\pi}\left\{E\left[\frac{1}{m^2}\sum_{i=1}^m X_i^2 + \frac{2}{m^2}\sum_{i<J} X_i X_J\right]\right\}$$

$$= \frac{4}{\pi}\left\{\frac{1}{m^2}\sum_{i=1}^m E[X_i^2] + \frac{2}{m^2}\sum_{i<J} E[X_i X_J]\right\} \overset{(\cdot\cdot)}{=}$$

Poichè X_i sono i.i.d.

$$E[X_i X_J] = E[X_i]\,E[X_J] \quad (\text{teorema})$$

$$\sum_{i=1}^m E[X_i^2] = m\,E[X^2]$$

$$\overset{(\bullet\bullet)}{=} \frac{4}{\pi}\Big[\frac{1}{m^2} n \underbrace{E[X^2]}_{\substack{\shortparallel \\ \theta \\ \text{(calcolato)}}} + \frac{2}{m^2} \sum_{i<j} E[X_i]\, E[X_j]\Big]$$

N.B. : $E[X_i] = E[X_j] = E[X]$

$$= \frac{4}{\pi}\Big[\frac{1}{m}\theta + \frac{2}{m^2}\, \frac{m(m-1)}{2} \underbrace{(E[X])^2}_{\substack{\shortparallel \\ \frac{\sqrt{\pi\theta}}{2}}}\Big] \quad \text{(calcolato)}$$

$$= \frac{4}{\pi}\Big[\frac{\theta}{m} + \frac{m-1}{m}\,\frac{\pi\theta}{4}\Big]$$

$$= \frac{4\theta}{m\pi} + \frac{m-1}{m}\theta \neq \theta$$

$\Rightarrow \hat{\theta} = \frac{4}{\pi}\bar{X}_m^2$ non è corretto

Tuttavia

$$\lim_{m\to\infty} E[\hat{\theta}] = \underbrace{\lim_{m\to\infty} \frac{4\theta}{m\pi}}_{\to 0} + \underbrace{\lim_{m\to\infty} \frac{m-1}{m}}_{\to 1}\theta = \theta$$

$\Rightarrow$ <u>$\hat{\theta}$ è asintoticamente corretto.</u>

6) Consideriamo lo stimatore di max. ver.

Calcoliamo $var[\hat{\theta}]$

$$var[\hat{\theta}] = var[M_2'] = var\Big[\frac{1}{m}\sum_{i=1}^{m} X_i^2\Big] = \frac{1}{m^2} var\Big[\sum_{i=1}^{m} X_i^2\Big]$$

$$\underset{\substack{\uparrow \\ X_i \text{ i.i.d.}}}{=} \frac{1}{m^2}\sum_{i=1}^{m} var[X_i^2] = \frac{1}{m^2}\, m\, var[X^2] = \frac{var[X^2]}{m}$$

Dalla regola $var[Y] = E[Y^2] - (E[Y])^2$

$$var[X^2] = E[X^4] - (E[X^2])^2$$

Devo calcolare $E[X^4]$.

$$E[X^4] = \int_0^{\infty} (-x^4)\left(-\frac{2}{\theta}x\right) e^{-\frac{x^2}{\theta}} dx = \Big[\underbrace{-x^4 e^{-\frac{x^2}{\theta}}}_{=0}\Big]_0^{\infty} -$$

per l'Hospital

$$-\int_0^{\infty} e^{-\frac{x^2}{\theta}} \cdot (-4x^3)\, dx$$

$$= \int_0^{\infty} 4x^3 e^{-\frac{x^2}{\theta}} dx = 2\theta \int_0^{\infty} (-x^2)\left(-2\frac{x}{\theta}\right) e^{-\frac{x^2}{\theta}} dx$$

$$= 2\theta \left[\Big[\underbrace{-x^2 e^{-\frac{x^2}{\theta}}}_{=0}\Big]_0^{\infty} - \int_0^{\infty} e^{-\frac{x^2}{\theta}} (-2x)\, dx \right]$$

per l'Hospital

$$= 2\theta(-\theta) \int_0^{\infty} \left(-\frac{2x}{\theta}\right) e^{-\frac{x^2}{\theta}} dx = -2\theta^2 \left[e^{-\frac{x^2}{\theta}} \right]_0^{\infty} = 2\theta^2$$

Riassumendo

$$\text{var}[\hat{\theta}] = \frac{1}{n}\left[2\theta^2 - (\theta)^2\right] = \frac{\theta^2}{n}$$

$$\boxed{\text{var}[\hat{\theta}] = \frac{\theta^2}{n}}$$

Applico ora la disuguaglianza di Rao-Cramér

$$\text{var}[T] \geq \frac{1}{nE\left[\left(\frac{1}{f}\frac{\partial}{\partial\theta}f\right)^2\right]}$$

$$\frac{1}{f}\frac{\partial}{\partial\theta}f = \frac{\partial}{\partial\theta}(\log f)$$

$$f(x,\theta) = \frac{2}{\theta} x e^{-\frac{x^2}{\theta}}$$

$$\log f = \log\frac{2}{\theta} + \log x - \frac{x^2}{\theta} = \log 2 - \log\theta + \log x - \frac{x^2}{\theta}$$

$$\frac{\partial}{\partial\theta}\log f = -\frac{1}{\theta} - x^2 \cdot \left(-\frac{1}{\theta^2}\right) = -\frac{1}{\theta} + \frac{x^2}{\theta^2}$$

$$\left(\frac{\partial}{\partial\theta}\log f\right)^2 = \left(\frac{x^2}{\theta^2} - \frac{1}{\theta}\right)^2 = \frac{x^4}{\theta^4} - 2\frac{x^2}{\theta^3} + \frac{1}{\theta^2}$$

$$E\left[\left(\frac{1}{f}\frac{\partial f}{\partial \theta}\right)^2\right] = E\left[\frac{X^4}{\theta^4} - \frac{2x^2}{\theta^3} + \frac{1}{\theta^2}\right]$$

$$= \frac{1}{\theta^4} \underbrace{E[X^4]}_{=2\theta^2} - \frac{2}{\theta^3} \underbrace{E[X^2]}_{=\theta} + \frac{1}{\theta^2}$$

$$= \frac{1}{\theta^4} \cdot 2\theta^2 - \frac{2}{\theta^3} \cdot \theta + \frac{1}{\theta^2} = \frac{1}{\theta^2}$$

$\mathrm{var}[\hat{\theta}] \geq \dfrac{1}{n \cdot \frac{1}{\theta^2}} = \dfrac{\theta^2}{n}$ limite inferiore della varianza

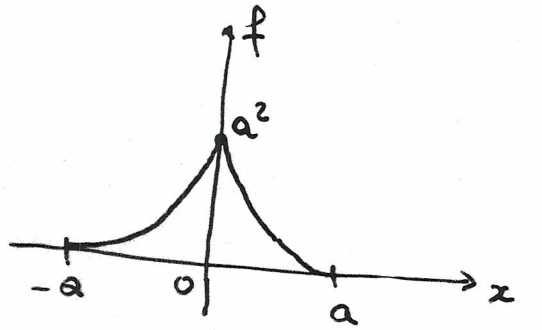
ma $\mathrm{var}[\hat{\theta}] = \dfrac{\theta^2}{n}$ perciò tra gli stimatori corretti è quello con varianza minima $\Rightarrow$ con MSE minimo $\Rightarrow$ più EFFICIENTE!

Esempio

Una v.a è distribuita con la legge:

$$f(x) = C_a \begin{cases} (x+a)^2 & -a \leq x \leq 0 \\ (x-a)^2 & 0 \leq x \leq a \\ 0 & \text{altrove} \end{cases}$$

1. Calcolare C_a
2. Calcolare $E[X]$, var$[X]$.
3. Calcolare uno stimatore di a col metodo dei momenti e il valore atteso del suo quadrato.



$$1 = C_a \cdot 2\int_0^a (x-a)^2\,dx = 2C_a \int_0^a (x^2 - 2ax + a^2)\,dx$$

$$= 2C_a \left(\frac{1}{3}x^3 - 2a\cdot\frac{1}{2}x^2 + a^2\cdot x\right)_0^a = 2C_a \cdot \frac{1}{3}a^3$$

$$\Rightarrow C_a = \frac{3}{2a^3}$$

$E[X] \equiv 0$

var$[X] = E[X^2]$

$$E[X^2] = C_a\, 2\int_0^a (x-a)^2 x^2\,dx = \frac{3}{a^3}\int_0^a (x^4 - 2ax^3 + a^2x^2)\,dx$$

$$= \frac{3}{a^3}\left(\frac{1}{5}a^5 - 2a\cdot\frac{1}{4}a^4 + a^2\cdot\frac{1}{3}a^3\right) = \frac{3}{a^3}\cdot\frac{a^5}{30} = \frac{a^2}{10}$$

$M_1' = \mu_1'$

ma $\mu_1' = E[X] = 0$

e $M_1' = \bar{X}_n$ $\quad \Rightarrow \quad \boxed{\bar{X}_n = 0 \ \lightning}$

allora calcolo

$$M_2' = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

$M_2' = \mu_2'$

$\mu_2' = E[X^2] = \dfrac{a^2}{10}$ $\quad \Rightarrow \quad \dfrac{a^2}{10} = M_2' \quad \Rightarrow \quad \underline{\underline{\bar{a} = \sqrt{10\, M_2'}}}$

$$E[\bar{a}^2] = E[10\, M_2'] = 10\, E[M_2'] = 10\, E\left[\frac{X_1^2 + \dots + X_n^2}{n}\right]$$

$$= \frac{10}{n} \sum_{i=1}^{n} E[X_i^2] = \frac{10}{n} \cdot n\, E[X^2] = \frac{10}{n} \cdot n \cdot \frac{a^2}{10} = a^2$$

il quadrato dello stimatore è corretto.

STIMA PER INTERVALLI

Le stime puntuali sono utili anche se non del tutto soddisfacenti.

Dato c.c. $X_1, .., X_n$ estratto da $N(\mu = \theta, \sigma^2 \text{ nota})$ si ha che $\bar{X}_n$ è lo stimatore di massima verosimiglianza per μ.

$\not\Rightarrow \bar{X}_n = \mu$, ma che $\bar{X}_n$ è vicina a μ.

A volte è preferibile produrre un intervallo per il quale si ha un certo livello di fiducia o confidenza che μ vi appartenga.

IMPORTANTE: Per ottenere un intervallo di confidenza dobbiamo utilizzare la distribuzione di probabilità dello stimatore puntuale.

INTERVALLI DI CONFIDENZA PER LA MEDIA

per un c.c. estratto da una normale

a) σ^2 nota

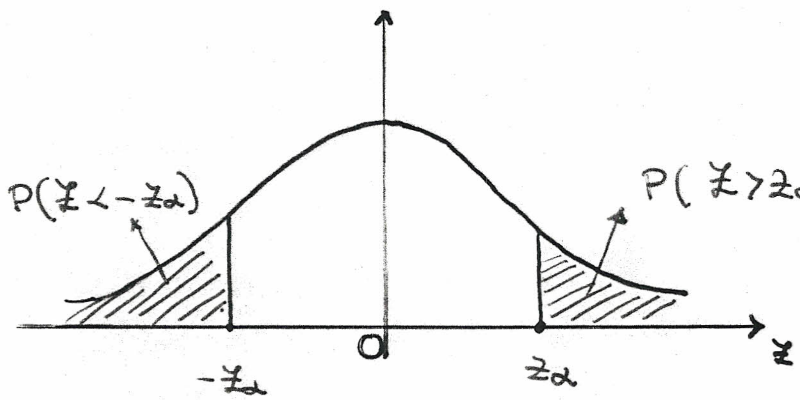
b) σ^2 non nota

a)

$X_i \sim N(\mu, \sigma^2)$ $i = 1, \dots, n$ $\Rightarrow \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\Rightarrow \boxed{Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)}$$

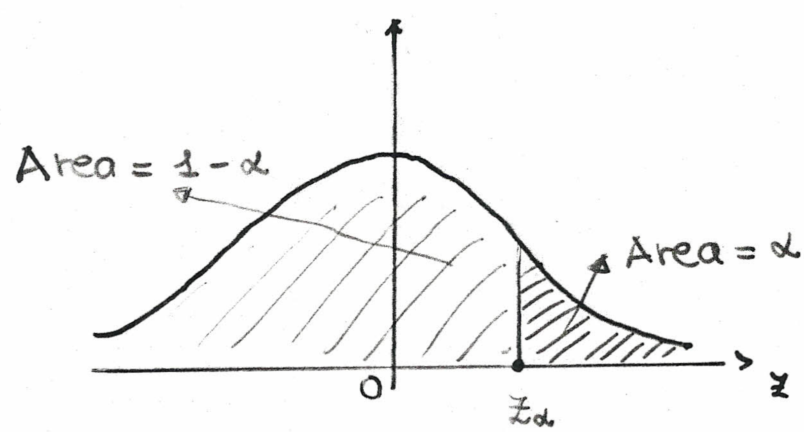
Z_n è indipendente da μ (incognita) $\Rightarrow$ Z_n è una statistica.



Ricordiamo che

$$\Phi(-z_\alpha) = P(Z_n < -z_\alpha) = P(Z_n > z_\alpha) = 1 - P(Z_n < z_\alpha)$$
$$= 1 - \Phi(z_\alpha)$$

In generale si definiscono intervalli di confidenza al livello $(1-\alpha)$, $\forall \alpha \in (0,1)$.



$\forall \alpha \in (0,1)$

$z_\alpha : P(Z_n > z_\alpha) = \alpha$

$P(Z_n > z_\alpha) = 1 - \phi(z_\alpha)$

$\Rightarrow \; 1 - \phi(z_\alpha) = \alpha \; \Rightarrow \; \phi(z_\alpha) = 1 - \alpha$

$\Rightarrow \; z_\alpha = \Phi^{-1}(1-\alpha)$

Si ha che $\boxed{-z_\alpha = z_{1-\alpha}}$

Infatti

$P(Z_n < -z_\alpha) = P(Z_n > z_\alpha) = \alpha$

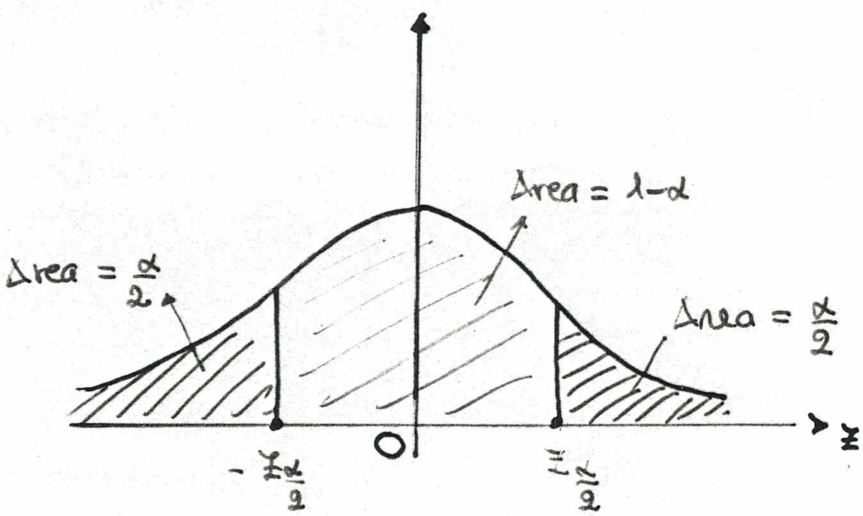
ma $P(Z_n < -z_\alpha) = 1 - P(Z_n > -z_\alpha)$

$P(Z_n > -z_\alpha) = 1 - P(Z_n < -z_\alpha) = 1 - \alpha$

Definisco

$z_{1-\alpha} : \; P(Z_n > z_{1-\alpha}) = 1 - \alpha$

da cui $-z_\alpha = z_{1-\alpha}$.



$\forall \, \alpha \in (0,1) \quad P\left(-z_{\frac{\alpha}{2}} < Z_n < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$

Poiché $Z_n = \dfrac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$

$$1-\alpha = P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}\right)$$

$$= P\left(-z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} < \bar{X}_n-\mu < z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right)$$

$$= P\left(\bar{X}_n - z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right)$$

Pertanto l'intervallo di confidenza bilaterale al livello $(1-\alpha)$ per μ è:

$$\underline{\left(\bar{x}_n - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \bar{x}_n + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)}$$

Mentre da

$P(Z_n > z_\alpha) = \alpha$ e $P(Z_n < -z_\alpha) = \alpha$

si ottengono gli intervalli unilaterali al livello $(1-\alpha)$ per μ.

- $\alpha = P(Z_n > z_\alpha) = P\left(\frac{\bar{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} > z_\alpha\right) = P\left(\mu < \bar{X}_n - z_\alpha\frac{\sigma}{\sqrt{n}}\right)$

$$1-\alpha = P(Z_n < z_\alpha) = P\left(\mu > \bar{X}_n - z_\alpha\frac{\sigma}{\sqrt{n}}\right)$$

si ottiene l'int. di conf. unilat. destro

$$\underline{\left(\bar{x}_n - z_\alpha\frac{\sigma}{\sqrt{n}},\ +\infty\right)}$$

- $\alpha = P(Z_n < -z_\alpha) = P\left(\frac{\bar{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} < -z_\alpha\right) = P\left(\mu > \bar{X}_n + z_\alpha\frac{\sigma}{\sqrt{n}}\right)$

$$1-\alpha = P\left(\mu < \bar{X}_n + z_\alpha\frac{\sigma}{\sqrt{n}}\right)$$

Si ottiene l'int. di conf. unilat. sinistro

$$\left(-\infty,\ \bar{x}_n + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

Esempi

Voglio calcolare

- $P[-1.96 < Z_n < 1.96] = \int_{-1.96}^{1.96} \varphi(z)\,dz = 2\int_{0}^{1.96} \varphi(z)\,dz$

$$= 2\,[0.975 - 0.5] = 2 \cdot 0,475 = 0.95$$

↑
TAB

$\approx$

$P\left[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right] = 0.95$ è int. di conf. bilat.

cioè il 95% delle volte μ starà ad una distanza non superiore a $\left(1.96\frac{\sigma}{\sqrt{n}}\right)$ dalla media dei dati.
Dopo aver calcolato $\bar{x}_n$ (media dei dati), concludo che con il 95% di confidenza la media vera μ della distribuzione appartiene a:

$$I = \left(\bar{x}_n - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x}_n + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

0.95 è il LIVELLO DI CONFIDENZA

$\Rightarrow \alpha = 0.05$.

$1.96 = z_{\frac{\alpha}{2}} = z_{0.025}$

$\bullet$ $P[Z_n < 1.645] = \int_{-\infty}^{1.645} \varphi(z)\,dz = 0.95$ ← TAB

$\approx P(\bar{X}_n - 1.645 \frac{\sigma}{\sqrt{n}} < \mu)$

è l'int. di conf. unilat. dx, cioè

$$\mu \in \left(\bar{x}_n - 1.645 \frac{\sigma}{\sqrt{n}}, +\infty\right)$$

$\alpha = 0.05$

$z_\alpha = z_{0.05} = 1.645$

mentre l'int. di conf. unilat. sx è:

$$\mu \in \left(-\infty, \bar{x}_n + 1.645 \frac{\sigma}{\sqrt{n}}\right)$$

VALORI PIÙ UTILIZZATI

$90\% = 1-\alpha$

$1.28 = z_\alpha = z_{0.10}$ $1.28 \to 0.899$

$1.645 = z_{\frac{\alpha}{2}} = z_{0.05}$ $2(0.95 - 0.5) = 0.90$

$95\% = 1-\alpha$

$1.645 = z_\alpha = z_{0.05}$ $1.64 \to 0.949$, $1.65 \to 0.951$ $\Rightarrow 1.645 \to 0.95$

$1.96 = z_{\frac{\alpha}{2}} = z_{0.025}$ $2(0.975 - 0.5) = 0.95$

$99\% = 1-\alpha$

$2.33 = z_\alpha = z_{0.01}$ $2.32 \to 0.989$

2.326 $2.33 \to 0.9901$

$2.58 = z_{\frac{\alpha}{2}} = z$ $2(0.995 - 0.5) = 0.99$

N.B. Ci sono molti possibili intervalli di confidenza bilat. con lo stesso livello di confidenza.

$P[-1.68 < Z_n < 2.70] = 0.95$

Pb. Scelta dell' intervallo migliore.

$\Rightarrow$ si sceglie quello più corto.

OSSERVAZIONE: L' intervallo simmetrico rispetto a $\bar{X}_n$ è quello che rende minima l'ampiezza dell' intervallo.

Esempio: c.c. $n = 4$ $N(\mu, \sigma = 3)$

$1.2 ; 3.4 ; 0.6 ; 5.6.$

$\bar{x}_n = 2.7$ è la stima di max ver. per μ.

Int. bilat. simm. al 95% è:

$$\left(2.7 - 1.96 \cdot \frac{3}{2}, 2.7 + 1.96 \cdot \frac{3}{2}\right) = (-0.24, 5.64)$$

e $\ell(I) = 5.88$

quello non simm. è:

$$\left(2.7 - 2.7 \cdot \frac{3}{2}, 2.7 + 1.68 \cdot \frac{3}{2}\right) = (-1.35, 5.22)$$

e $\ell(I) = 6.57$.

(unilat. dx $(0.23, +\infty)$; unilat. sx $(-\infty, 5.16)$)

per esercizio.

In alcune situazioni si richiede che l'intervallo di confidenza al livello $(1-\alpha)$ abbia una lunghezza fissata.

Pb: determinare la numerosità del campione che garantisce il risultato.

Esempio

$n = ?$ di un c.c. estratto da $N(\mu, \sigma = 0.2)$ affinché $I_{bil.}$ al 99% per μ abbia lunghezza $\ell(I) = 0.1$.

$$\Big(\bar{x}_n \pm \underbrace{2.58}_{= z_{\frac{\alpha}{2}}} \frac{\sigma}{\sqrt{n}}\Big) \Rightarrow \ell(I) = 2 \cdot (2.58) \frac{\sigma}{\sqrt{n}}$$

$$5.16 \frac{\sigma}{\sqrt{n}} = 0.1 \Rightarrow n = (51.6 \underbrace{\sigma}_{= 0.2})^2 \approx 106.5$$

Poiché n deve essere intero $\Rightarrow n = 107$.

b) $Z_n = \dfrac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ contiene σ incognita

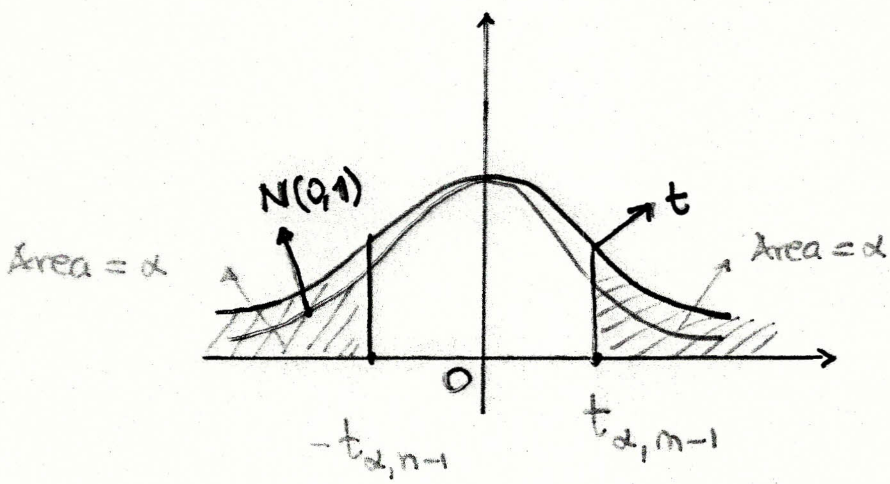
$\Rightarrow$ non è una statistica

mentre

$T_{n-1} = \dfrac{\bar{X}_n - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$ è indipendente da μ e da σ.

(simmetrica rispetto ad 0 come la normale).

Ricordiamo che la distribuzione t di Student dipende da 1 parametro = g. di libertà.



$\forall \alpha \in (0,1)$

$$t_{\alpha,n-1} : P(T_{n-1} > t_{\alpha,n-1}) = \alpha$$

$$P(T_{n-1} < -t_{\alpha,n-1}) = P(T_{n-1} > t_{\alpha,n-1}) = \alpha$$

$$\| $$

$$1 - P(T_{n-1} > -t_{\alpha,n-1})$$

$$\Rightarrow P(T_{n-1} > -t_{\alpha,n-1}) = 1-\alpha$$

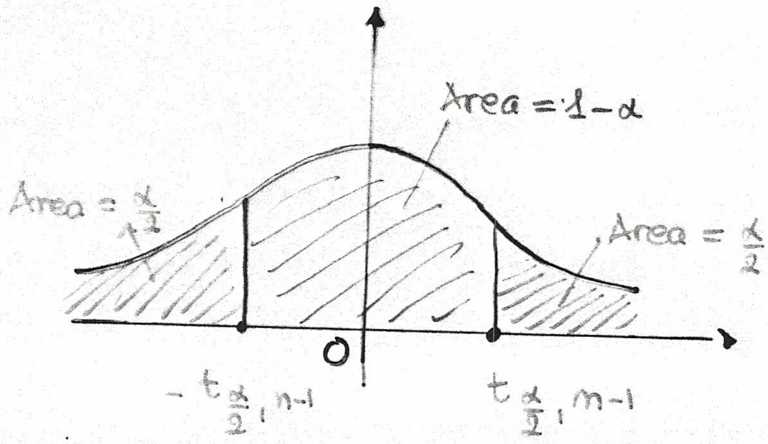
allora definisco

$$t_{1-\alpha,n-1} : P(T_{n-1} > -t_{1-\alpha,n-1}) = 1-\alpha$$

è pertanto verificato

$$-t_{\alpha,n-1} = t_{1-\alpha,n-1}$$

In analogia a quanto fatto per il caso a) si conoscono gli intervalli di conf. bilat. e unilat. dx e sx al livello $(1-\alpha)$, $\forall \alpha \in (0,1)$.



$\forall \alpha \in (0,1)$

$$P\left(-t_{\frac{\alpha}{2},n-1} < T_{n-1} < t_{\frac{\alpha}{2},n-1}\right) = 1-\alpha$$

$$T_{n-1} = \frac{\overline{X}_n - \mu}{\frac{S}{\sqrt{n}}}$$

$$1-\alpha = P\left(\overline{X}_n - t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}} < \mu < \overline{X}_n + t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}\right)$$

l'int. di conf. bilat. al livello $(1-\alpha)$ per μ è:

$$\left(\overline{x}_n - t_{\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}}\ ,\ \overline{x}_n + t_{\frac{\alpha}{2},n-1} \cdot \frac{s}{\sqrt{n}}\right)$$

Gli intervalli unilaterali si ricavano:

- $1-\alpha = P\left(T_{n-1} < t_{\alpha,n-1}\right) = P\left(\mu > \overline{X}_n - t_{\alpha,n-1} \cdot \frac{S}{\sqrt{n}}\right)$

$\Rightarrow \left(\overline{x}_n - t_{\alpha,n-1} \cdot \frac{s}{\sqrt{n}}, +\infty\right)$ int. unil. dx.

• $1-\alpha = P(T_{n-1} > -t_{\alpha,n-1}) = P\left(\mu < \overline{X}_n + t_{\alpha,n-1} \cdot \frac{S}{\sqrt{n}}\right)$

$\Rightarrow (-\infty, \overline{x}_n + t_{\alpha,n-1} \, s/\sqrt{n})$ int. unil. sx

Confronto con l'esempio $n=4$, $\overline{x}_n = 2.7$.

$1-\alpha = 0.95$

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$t_{0.025,3} = 3.182 \leftarrow$ TABELLA ora σ^2 non è nota

$$\overline{x}_n \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}$$

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2\right]$$

$$s^2 = \frac{1}{3}\left[(1.2)^2 + (3.4)^2 + (0.6)^2 + (5.6)^2 - 4\cdot(2.7)^2\right]$$

$s \approx 2.277$

ottengo un intervallo + largo: $\ell(I) = 7.10$
(le code sono più pesanti).

■

N.B. Se non è più valida l'ipotesi che la distribuzione della popolazione sia normale, a condizione che n sia sufficientemente grande per il TLC

$$\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \quad \text{e} \quad \frac{\overline{X}_n - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

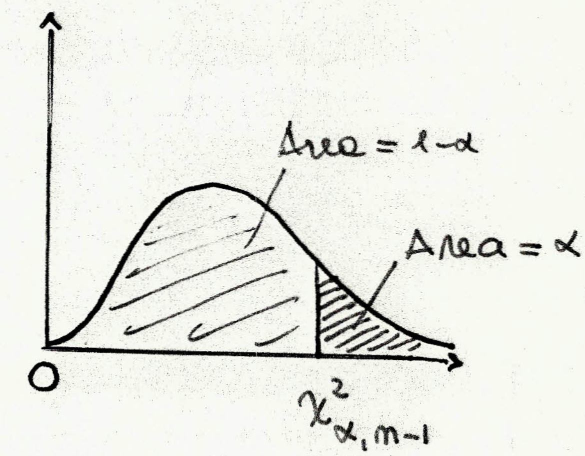
INTERVALLI DI CONFIDENZA PER LA VARIANZA per un c.c. estratto da una normale

a) μ non nota

Si utilizza la statistica

$$V = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Ricordiamo che:



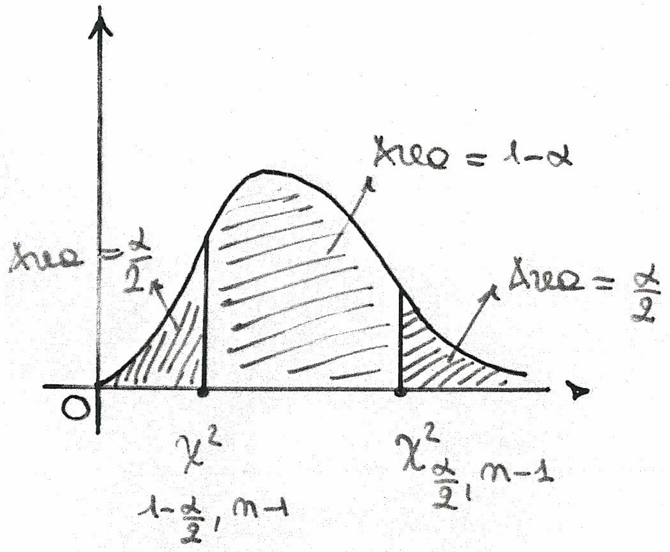
$\forall \alpha \in (0,1)$

$$\chi^2_{\alpha,n-1} : P(V > \chi^2_{\alpha,n-1}) = \alpha$$

e quindi

$$P(0 < V < \chi^2_{\alpha,n-1}) = 1-\alpha$$

Si costruiscono pertanto gli intervalli di confidenza bilat. e unilat. dx e sx al livello $(1-\alpha)$, $\forall \alpha \in (0,1)$.



Il valore $\chi^2_{1-\frac{\alpha}{2}, n-1}$ si ottiene ovviamente da

$$P\left(V > \chi^2_{1-\frac{\alpha}{2}, n-1}\right) = 1 - \frac{\alpha}{2}$$

Pertanto

$$1-\alpha = P\left(\chi^2_{1-\frac{\alpha}{2}, n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}, n-1}\right)$$

$$= P\left[\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right]$$

l'intervallo di conf. bil. al livello $(1-\alpha)$ per σ^2 sarà:

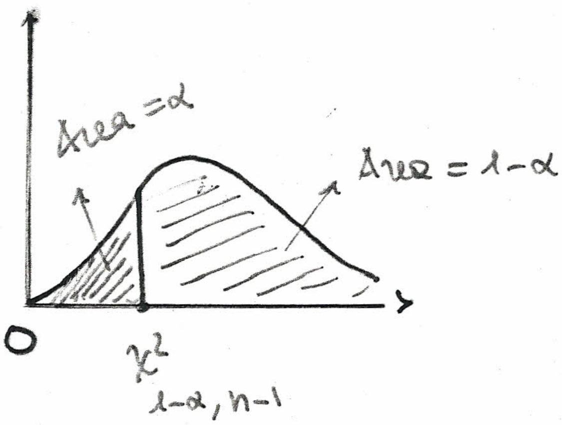
$$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\right)$$

Gli intervalli di conf. unilat. si ricavano:

- $1-\alpha = P\left(V < \chi^2_{\alpha, n-1}\right) = P\left(\frac{S^2(n-1)}{\sigma^2} < \chi^2_{\alpha, n-1}\right)$

$$= P\left(\sigma^2 > \frac{(n-1)S^2}{\chi^2_{\alpha, n-1}}\right)$$

$$\Rightarrow \left(\frac{(n-1)S^2}{\chi^2_{\alpha, n-1}}, \infty \right) \quad \underline{\text{int. unilat. dx}}$$



$$1-\alpha = P\left(V > \chi^2_{1-\alpha, n-1}\right) = P\left(\frac{(m-1)}{\sigma^2} S^2 > \chi^2_{1-\alpha, m-1}\right)$$

$$= P\left(\sigma^2 < \frac{(m-1) S^2}{\chi^2_{1-\alpha, n-1}}\right)$$

$$\Rightarrow \left(0, \frac{(m-1) S^2}{\chi^2_{1-\alpha, n-1}} \right) \quad \underline{\text{int. unilat. sx}}$$

Esempio : Produzione automatica di rondelle di spessore ridotto in mm.

c.c. $m = 10$

spessore :

0.123	0.128
0.133	0.120
0.124	0.124
0.125	0.130
0.126	0.126

$\overline{x}_m = 0.125$

$\overline{x}^2_m = 0.0158$

$$s^2 = \frac{1}{m-1}\left[\sum_{i=1}^{m} x_i^2 - m\,\overline{x}_m^2\right]$$

$$= \frac{1}{9}\left[(0.123)^2 + \dots + (0.126)^2 - 10 \cdot 0.0158\right]$$

$$\simeq 1.366 \cdot 10^{-5}$$

$$\sigma^2 \in \left(\frac{(m-1)s^2}{\chi^2_{\frac{\alpha}{2}, m-1}} , \frac{(m-1)s^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \right)$$

voglio un livello pari al 90%.

$1 - \alpha = 0.90 \Rightarrow \alpha = 0.1$ e $\frac{\alpha}{2} = 0.05$, $1 - \frac{\alpha}{2} = 0.95$

$$\chi^2_{\frac{\alpha}{2}, n-1} = \chi^2_{0.05, 9} \approx 16.919$$

$$\chi^2_{1-\frac{\alpha}{2}, n-1} = \chi^2_{0.95, 9} \approx 3.325$$

$$\Rightarrow \sigma^2 \in (7.26 \cdot 10^{-6}, 36.97 \cdot 10^{-6})$$

OSSERVAZIONE

Si possono costruire int. di conf. bilat. e unil. per la varianza anche quando μ è nota.

caso a)

Si può utilizzare ancora la statistica

$$Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

e calcolare p. es.

$$P[a < Z_n < b] = P\left[\frac{\bar{X}_n - \mu}{\frac{b}{\sqrt{n}}} < \sigma < \frac{\bar{X}_n - \mu}{\frac{a}{\sqrt{n}}}\right] = 1-\alpha$$

$a = -z_{\frac{\alpha}{2}}$, $b = z_{\frac{\alpha}{2}}$

ma in questo caso trovo un int. di conf. bilat. per σ e non per σ^2. (••)

caso b)

Si utilizza la statistica

$$U := \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$$

$$1-\alpha = P[a < U < b] = P\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{b} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{a}\right]$$

dove $b = \chi^2_{\frac{\alpha}{2}, n}$, $a = \chi^2_{1-\frac{\alpha}{2}, n}$.

(••) int. unil. dx : $\left(\frac{\bar{X}_n - \mu}{\frac{z_\alpha}{\sqrt{n}}}, +\infty\right)$

" " sx : $\left(-\infty, (\bar{X}_n - \mu)/_{-z_\alpha/\sqrt{n}}\right)$

OSSERVAZIONI

La lunghezza di un intervallo di confidenza per la media μ, nota la varianza σ^2 è dato da:

$$l = 2 z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

dove $(1-\alpha)$ è il grado di fiducia.

- se n aumenta la $l(I)$ cala $\Rightarrow$ stima più precisa
- se il grado di fiducia $(1-\alpha)$ aumenta la $l(I)$ aumenta perchè $z_{\frac{\alpha}{2}}$ aumenta $\Rightarrow$ stima meno precisa
- se σ aumenta cioè aumenta la variabilità del campione, la $l(I)$ aumenta

Ora n, α possono essere controllati, σ dipende dal tipo di dati.

n, α sono obiettivi in conflitto.

Se voglio aumentare la precisione della stima senza diminuire il grado di fiducia, devo aumentare n.

SCHEMA

- stima per μ se σ è nota $\Rightarrow N(0,1)$

  se σ non è nota $\Rightarrow T_{n-1}$ solo per piccoli campioni ($n \leq 30$)

  $\Rightarrow N(0,1)$ sostituendo σ con s.

- stima per σ^2 se μ è nota $\Rightarrow \chi^2_n$

  se μ non è nota $\Rightarrow \chi^2_{n-1}$

e non conta la numerosità del campione.

Esempi

1) Stimare il numero medio di battiti cardiaci al minuto di una popolazione con $f = N(\mu, \sigma = 10)$ considerando un c.c. di 49 soggetti : $\bar{X}_n = 90$.
Determinare intervalli di confidenza bilat. per μ al 90% (95%; 99%)

- $1-\alpha = 0.90 \Rightarrow \alpha = 0.1 \Rightarrow \frac{\alpha}{2} = 0.05$

$z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$

$$90 - 1.645 \cdot \frac{10}{7} < \mu < 90 + 1.645 \cdot \frac{10}{7}$$

$$\Downarrow$$

$$\mu \in (87,65 ; 92,35)$$

$\ell(I) = 4,70$.

- $1-\alpha = 0.95 \quad \alpha = 0.05 \quad \frac{\alpha}{2} = 0.025 \quad z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

$\ell(I) = 5,60$

- $(1-\alpha) = 0.99 \quad \alpha = 0.01 \quad \frac{\alpha}{2} = 0.005 \quad z_{\frac{\alpha}{2}} = z_{0.005} \cong 2.58$
$= 2.576$

$\ell(I) = 7,36$

se $(1-\alpha)$ aumenta $\ell(I)$ aumenta.

2) Da una popolazione si estrae un campione di 16 oggetti di cui si misura il peso.

Stimare μ sapendo che $\bar{x}_m = 3.42$ gr , $S = 0.68$ gr con un grado di fiducia del 99%.

Poichè faccio misure $\Rightarrow$ la f. di densità adatta è la normale

$n = 16$

$\nu = n - 1 = 15$ grado di libertà della t di student.

$1 - \alpha = 0.99 \qquad \alpha = 0.01 \qquad \frac{\alpha}{2} = 0.005$

$$t_{\frac{\alpha}{2}, n-1} = t_{0.005, 15} = 2.947$$

$$3,42 - 2,947 \cdot \frac{0,68}{\sqrt{16}} < \mu < 3.42 + 2,947 \cdot \frac{0,68}{\sqrt{16}}$$

$$\Downarrow$$

$$\mu \in (2,91 ; 3,93)$$

Uso la t di student perchè non conosco σ^2 ed il campione è piccolo.

3) In una scuola si sceglie un campione di 16 studenti da una classe 5^a e se ne misura l'altezza.

La varianza campionaria $s^2 = 37,09$ cm.

Determinare un int. di conf. (bil.) al 95% (99%) per la σ^2 della popolazione costituita da tutti gli studenti di 5^a.

Faccio misure $\Rightarrow$ la f. di densità adatta è la normale.

$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \qquad \frac{\alpha}{2} = 0.025$

$$1-\frac{\alpha}{2} = 0.975$$

$\nu = n-1 = 15$ g. di libertà

$$\frac{(n-1)\,s^2}{\chi^2_{\frac{\alpha}{2},\,n-1}} < \sigma^2 < \frac{(n-1)\,s^2}{\chi^2_{1-\frac{\alpha}{2},\,n-1}}$$

$$\chi^2_{\frac{\alpha}{2},\,n-1} = \chi^2_{0.025,\,15} = 27.488$$

$$\chi^2_{1-\frac{\alpha}{2},\,n-1} = \chi^2_{0.975,\,15} = 6.262$$

$$\frac{15 \cdot 37{,}09}{27{,}488} < \sigma^2 < \frac{15 \cdot 37{,}09}{6.262}$$

$$\Downarrow$$

$$\sigma^2 \in (20{,}23\ ;\ 88{,}84)$$

4) Le misure dei diametri di un c.c. di 200 sfere da cuscinetto prodotte da una macchina, in una settimana forniscono i seguenti dati:

$\bar{x}_n = 0.824$ cm.

- $s = 0.042$ cm.

Det. un int. di conf. (bil.) per μ al 95% (99%).

Faccio misure $\Rightarrow$ la f. di densità è normale,

a) Non è nota σ^2 $\Rightarrow$ dovrei applicare T_{n-1}.

ma la tabella della t di student arriva fino a $\nu = 50$ e

nel nostro caso $\nu = n-1 = 199$!!

Soluzione

Poichè i grafici della t di Student e della normale standard sono simili, posso usare ancora la stima con $N(0,1)$ sostituendo σ con s.

$1-\alpha = 0.95 \qquad \alpha = 0.05 \qquad \frac{\alpha}{2} = 0.025$

$z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

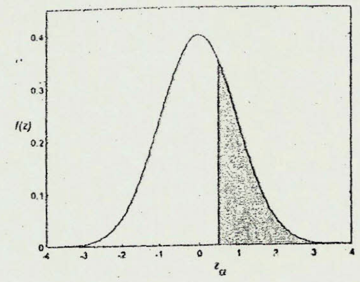
$$0.824 - 1.96 \cdot \frac{0{,}042}{\sqrt{200}} < \mu < 0.824 + 1.96 \cdot \frac{0{,}042}{\sqrt{200}}$$

$$\Downarrow$$

$$\mu \in (0.818 \,;\, 0.83)$$

Tavola 4 – Percentili per la distribuzione normale standardizzata

La tavola fornisce i valori di z_α per i quali $P(z > z_\alpha) = \alpha \cdot 100\% = q\%$, per alcuni valori notevoli di q.



$q\%$	z_α	$q\%$	z_α	$q\%$	z_α	$q\%$	z_α
50	0.000	9	1.341	2.9	1.896	0.4	2.652
45	0.126	8	1.405	2.8	1.911	0.3	2.748
40	0.253	7	1.476	2.7	1.927	0.2	2.878
35	0.385	6	1.555	2.6	1.943	0.1	3.090
30	0.524	5	1.645	2.5	1.960		
29	0.553	4.9	1.655	2.4	1.977	0.09	3.121
28	0.583	4.8	1.665	2.3	1.995	0.08	3.156
27	0.613	4.7	1.675	2.2	2.014	0.07	3.195
26	0.643	4.6	1.685	2.1	2.034	0.06	3.239
25	0.674	4.5	1.695	2.0	2.054	0.05	3.291
24	0.706	4.4	1.706	1.9	2.075	0.04	3.353
23	0.739	4.3	1.717	1.8	2.097	0.03	3.432
22	0.772	4.2	1.728	1.7	2.120	0.02	3.540
21	0.806	4.1	1.739	1.6	2.144	0.01	3.719
20	0.842	4.0	1.751	1.5	2.170	0.005	3.891
19	0.878	3.9	1.762	1.4	2.197	0.001	4.265
18	0.915	3.8	1.774	1.3	2.226	0.0005	4.417
17	0.954	3.7	1.787	1.2	2.257	0.0001	4.753
16	0.994	3.6	1.799	1.1	2.290	0.00005	4.892
15	1.036	3.5	1.812	1.0	2.326	0.00001	5.199
14	1.080	3.4	1.825	0.9	2.366	0.000005	5.327
13	1.126	3.3	1.838	0.8	2.409	0.000001	5.612
12	1.175	3.2	1.852	0.7	2.457	0.0000005	5.731
11	1.227	3.1	1.866	0.6	2.512	0.0000001	5.998
10	1.282	3.0	1.881	0.5	2.576	0.00000005	6.109

Valori della funzione di ripartizione della normale standardizzata $\Phi(z) = \alpha$										
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4.0	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

	Percentili della normale standardizzata $z = \Phi^{-1}(\alpha)$											
$1-\alpha$	$\Phi(z)$	0.7000	0.7500	0.8000	0.9000	0.9500	0.9750	0.9900	0.9950	0.9990	0.9995	0.9999	$1-\frac{\alpha}{2}$
	z	0.5244	0.6745	0.8416	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902	3.2905	3.7195	

$\nu = n \quad \alpha = n$

Percentili della variabile Gamma con $\lambda = 1/2$ ed $\alpha = \nu/2$ (χ^2 con ν gradi di libertà)

$(1-\alpha)$

ν	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0.00004	0.00016	0.00098	0.00393	0.01579	2.70554	3.84146	5.02390	6.63489	7.8794
2	0.01002	0.02010	0.05064	0.10259	0.21072	4.60518	5.99148	7.37778	9.21035	10.5965
3	0.07172	0.11483	0.21579	0.35185	0.58438	6.25139	7.81472	9.34840	11.34488	12.8380
4	0.20698	0.29711	0.48442	0.71072	1.06362	7.77943	9.48773	11.14326	13.27670	14.8601
5	0.41175	0.55430	0.83121	1.14548	1.61031	9.23635	11.07048	12.83249	15.08632	16.7496
6	0.67573	0.87208	1.23734	1.63538	2.20413	10.64464	12.59158	14.44935	16.81187	18.5475
7	0.98925	1.23903	1.68986	2.16735	2.83311	12.01703	14.06713	16.01277	18.47532	20.2777
8	1.34440	1.64651	2.17972	2.73263	3.48954	13.36156	15.50731	17.53454	20.09016	21.9548
9	1.73491	2.08789	2.70039	3.32512	4.16816	14.68366	16.91896	19.02278	21.66605	23.5892
10	2.15585	2.55820	3.24696	3.94030	4.86518	15.98717	18.30703	20.48320	23.20929	25.1880
11	2.60320	3.05350	3.81574	4.57481	5.57779	17.27501	19.67515	21.92002	24.72502	26.7568
12	3.07379	3.57055	4.40378	5.22603	6.30380	18.54934	21.02606	23.33666	26.21696	28.2996
13	3.56504	4.10690	5.00874	5.89186	7.04150	19.81193	22.36203	24.73558	27.68818	29.8193
14	4.07466	4.66042	5.62872	6.57063	7.78954	21.06414	23.68478	26.11893	29.14116	31.3194
15	4.60087	5.22936	6.26212	7.26093	8.54675	22.30712	24.99580	27.48836	30.57795	32.8014
16	5.14216	5.81220	6.90766	7.96164	9.31224	23.54182	26.29622	28.84532	31.99986	34.2670
17	5.69727	6.40774	7.56418	8.67175	10.08518	24.76903	27.58710	30.19098	33.40872	35.7183
18	6.26477	7.01490	8.23074	9.39045	10.86494	25.98942	28.86932	31.52641	34.80524	37.1563
19	6.84392	7.63270	8.90651	10.11701	11.65091	27.20356	30.14351	32.85234	36.19077	38.5821
20	7.43381	8.26037	9.59077	10.85080	12.44260	28.41197	31.41042	34.16958	37.56627	39.9968
21	8.03360	8.89717	10.28291	11.59132	13.23960	29.61509	32.67056	35.47886	38.93223	41.4009
22	8.64268	9.54249	10.98233	12.33801	14.04149	30.81329	33.92446	36.78068	40.28945	42.7956
23	9.26038	10.19569	11.68853	13.09051	14.84795	32.00689	35.17246	38.07561	41.63833	44.1813
24	9.88620	10.85635	12.40115	13.84842	15.65868	33.19624	36.41503	39.36406	42.97978	45.5583
25	10.51965	11.52395	13.11971	14.61140	16.47341	34.38158	37.65249	40.64650	44.31401	46.9279
26	11.16022	12.19818	13.84388	15.37916	17.29188	35.56316	38.88513	41.92314	45.64164	48.2897
27	11.80765	12.87847	14.57337	16.15139	18.11389	36.74123	40.11327	43.19452	46.96284	49.6450
28	12.46128	13.56467	15.30785	16.92788	18.93924	37.91591	41.33715	44.46079	48.27817	50.9935
29	13.12107	14.25641	16.04705	17.70838	19.76774	39.08748	42.55695	45.72228	49.58783	52.3355
30	13.78668	14.95346	16.79076	18.49267	20.59924	40.25602	43.77295	46.97922	50.89218	53.6718
31	14.45774	15.65547	17.53872	19.28056	21.43357	41.42175	44.98534	48.23192	52.19135	55.0024
32	15.13402	16.36220	18.29079	20.07191	22.27059	42.58473	46.19424	49.48044	53.48566	56.3279
33	15.81518	17.07348	19.04666	20.86652	23.11019	43.74518	47.39990	50.72510	54.77545	57.6483
34	16.50130	17.78910	19.80624	21.66428	23.95225	44.90316	48.60236	51.96602	56.06085	58.9637
35	17.19173	18.50887	20.56938	22.46501	24.79665	46.05877	49.80183	53.20331	57.34199	60.2745
36	17.88675	19.23263	21.33587	23.26862	25.64329	47.21217	50.99848	54.43726	58.61915	61.5810
37	18.58588	19.96027	22.10562	24.07494	26.49209	48.36339	52.19229	55.66798	59.89256	62.8831
38	19.28882	20.69141	22.87849	24.88389	27.34296	49.51258	53.38351	56.89549	61.16202	64.1812
39	19.99583	21.42614	23.65430	25.69538	28.19579	50.65978	54.57224	58.12005	62.42809	65.4753
40	20.70658	22.16420	24.43306	26.50930	29.05052	51.80504	55.75849	59.34168	63.69077	66.7660
41	21.42075	22.90556	25.21452	27.32556	29.90708	52.94850	56.94240	60.56055	64.94998	68.0526
42	22.13838	23.65014	25.99866	28.14405	30.76542	54.09019	58.12403	61.77672	66.20629	69.3360
43	22.85957	24.39757	26.78537	28.96471	31.62546	55.23018	59.30352	62.99031	67.45929	70.6157
44	23.58362	25.14801	27.57454	29.78750	32.48713	56.36852	60.48090	64.20141	68.70964	71.8923
45	24.31098	25.90120	28.36618	30.61226	33.35038	57.50529	61.65622	65.41013	69.95690	73.1660
46	25.04130	26.65719	29.16002	31.43900	34.21517	58.64053	62.82961	66.61647	71.20150	74.4367
47	25.77450	27.41582	29.95616	32.26761	35.08142	59.77429	64.00113	67.82064	72.44317	75.7038
48	26.51067	28.17697	30.75450	33.09807	35.94914	60.90661	65.17076	69.02257	73.68256	76.9689
49	27.24937	28.94059	31.55493	33.93029	36.81823	62.03753	66.33865	70.22236	74.91939	78.2305
50	27.99082	29.70673	32.35738	34.76424	37.68864	63.16711	67.50481	71.42019	76.15[illegible]	79.4898

grado di libertà

ν	0.75	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995	0.9999
	Percentili della variabile casuale *t* di Student									
1	1.00000	1.37638	3.07768	6.31375	12.70615	31.82096	63.65590	318.2888	636.5776	3185.272
2	0.81650	1.06066	1.88562	2.91999	4.30266	6.96455	9.92499	22.32846	31.59977	70.70601
3	0.76489	0.97847	1.63775	2.35336	3.18245	4.54071	5.84085	10.21428	12.92443	22.20273
4	0.74070	0.94096	1.53321	2.13185	2.77645	3.74694	4.60408	7.17293	8.61008	13.03852
5	0.72669	0.91954	1.47588	2.01505	2.57058	3.36493	4.03212	5.89353	6.86850	9.67644
6	0.71756	0.90570	1.43976	1.94318	2.44691	3.14267	3.70743	5.20755	5.95872	8.02334
7	0.71114	0.89603	1.41492	1.89458	2.36462	2.99795	3.49948	4.78525	5.40807	7.06408
8	0.70639	0.88889	1.39682	1.85955	2.30601	2.89647	3.35538	4.50076	5.04137	6.44242
9	0.70272	0.88340	1.38303	1.83311	2.26216	2.82143	3.24984	4.29689	4.78089	6.00936
10	0.69981	0.87906	1.37218	1.81246	2.22814	2.76377	3.16926	4.14366	4.58676	5.69387
11	0.69744	0.87553	1.36343	1.79588	2.20099	2.71808	3.10582	4.02477	4.43688	5.45289
12	0.69548	0.87261	1.35622	1.78229	2.17881	2.68099	3.05454	3.92960	4.31784	5.26314
13	0.69383	0.87015	1.35017	1.77093	2.16037	2.65030	3.01228	3.85204	4.22093	5.11063
14	0.69242	0.86805	1.34503	1.76131	2.14479	2.62449	2.97685	3.78743	4.14031	4.98490
15	0.69120	0.86624	1.34061	1.75305	2.13145	2.60248	2.94673	3.73286	4.07279	4.88013
16	0.69013	0.86467	1.33676	1.74588	2.11990	2.58349	2.92079	3.68615	4.01487	4.79049
17	0.68919	0.86328	1.33338	1.73961	2.10982	2.56694	2.89823	3.64576	3.96511	4.71482
18	0.68836	0.86205	1.33039	1.73406	2.10092	2.55238	2.87844	3.61048	3.92174	4.64846
19	0.68762	0.86095	1.32773	1.72913	2.09302	2.53948	2.86094	3.57933	3.88332	4.59026
20	0.68695	0.85996	1.32534	1.72472	2.08596	2.52798	2.84534	3.55183	3.84956	4.53903
21	0.68635	0.85907	1.32319	1.72074	2.07961	2.51765	2.83137	3.52709	3.81930	4.49247
22	0.68581	0.85827	1.32124	1.71714	2.07388	2.50832	2.81876	3.50497	3.79223	4.45172
23	0.68531	0.85753	1.31946	1.71387	2.06865	2.49987	2.80734	3.48497	3.76764	4.41563
24	0.68485	0.85686	1.31784	1.71088	2.06390	2.49216	2.79695	3.46678	3.74537	4.38187
25	0.68443	0.85624	1.31635	1.70814	2.05954	2.48510	2.78744	3.45019	3.72514	4.35160
26	0.68404	0.85567	1.31497	1.70562	2.05553	2.47863	2.77872	3.43498	3.70666	4.32367
27	0.68369	0.85514	1.31370	1.70329	2.05183	2.47266	2.77068	3.42101	3.68949	4.29922
28	0.68335	0.85465	1.31253	1.70113	2.04841	2.46714	2.76326	3.40820	3.67392	4.27593
29	0.68304	0.85419	1.31143	1.69913	2.04523	2.46202	2.75639	3.39627	3.65952	4.25382
30	0.68276	0.85377	1.31042	1.69726	2.04227	2.45726	2.74998	3.38521	3.64598	4.23403
31	0.68249	0.85337	1.30946	1.69552	2.03951	2.45283	2.74404	3.37488	3.63347	4.21540
32	0.68223	0.85300	1.30857	1.69389	2.03693	2.44868	2.73849	3.36528	3.62183	4.19794
33	0.68200	0.85265	1.30774	1.69236	2.03452	2.44479	2.73329	3.35633	3.61091	4.18222
34	0.68177	0.85232	1.30695	1.69092	2.03224	2.44115	2.72839	3.34796	3.60073	4.16767
35	0.68156	0.85201	1.30621	1.68957	2.03011	2.43772	2.72381	3.34003	3.59112	4.15312
36	0.68137	0.85172	1.30551	1.68830	2.02809	2.43450	2.71948	3.33261	3.58210	4.13973
37	0.68118	0.85144	1.30485	1.68709	2.02619	2.43144	2.71541	3.32562	3.57366	4.12751
38	0.68100	0.85118	1.30423	1.68595	2.02439	2.42857	2.71157	3.31900	3.56566	4.11528
39	0.68083	0.85093	1.30364	1.68488	2.02269	2.42584	2.70791	3.31274	3.55809	4.10480
40	0.68067	0.85070	1.30308	1.68385	2.02107	2.42326	2.70446	3.30692	3.55096	4.09433
41	0.68052	0.85048	1.30254	1.68288	2.01954	2.42080	2.70118	3.30125	3.54426	4.08385
42	0.68038	0.85026	1.30203	1.68195	2.01808	2.41847	2.69807	3.29594	3.53772	4.07454
43	0.68024	0.85006	1.30155	1.68107	2.01669	2.41625	2.69511	3.29092	3.53160	4.06639
44	0.68011	0.84987	1.30109	1.68023	2.01537	2.41414	2.69229	3.28611	3.52578	4.05707
45	0.67998	0.84968	1.30065	1.67943	2.01410	2.41212	2.68959	3.28146	3.52025	4.04892
46	0.67986	0.84951	1.30023	1.67866	2.01289	2.41019	2.68701	3.27709	3.51487	4.04194
47	0.67975	0.84934	1.29982	1.67793	2.01174	2.40834	2.68456	3.27287	3.50992	4.03379
48	0.67964	0.84917	1.29944	1.67722	2.01063	2.40658	2.68221	3.26894	3.50497	4.02681
49	0.67953	0.84902	1.29907	1.67655	2.00957	2.40489	2.67995	3.26509	3.50046	4.02040
50	0.67943	0.84887	1.29871	1.67591	2.00856	2.40327	2.67779	3.26138	3.49595	4.01400

$(1-\alpha)$

Intervalli con livello di confidenza $1-\alpha$ per campioni normali.

$$X_1, X_2, \ldots, X_n \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

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

Ipotesi	θ	Intervallo bilaterale	Intervallo sinistro	Intervallo destro
σ^2 nota	μ	$\overline{X}_n \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$\left(-\infty, \overline{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$	$\left(\overline{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right)$
σ^2 non nota	μ	$\overline{X}_n \pm t_{\frac{\alpha}{2},n-1} \frac{S}{\sqrt{n}}$	$\left(-\infty, \overline{X}_n + t_{\alpha,n-1} \frac{S}{\sqrt{n}}\right)$	$\left(\overline{X}_n - t_{\alpha,n-1} \frac{S}{\sqrt{n}}, \infty\right)$
μ non nota	σ^2	$\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)$	$\left(0, \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}\right)$	$\left(\frac{(n-1)S^2}{\chi^2_{\alpha,n-1}}, \infty\right)$
μ nota	σ^2	$\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2},n}}, \frac{\cdot}{\chi^2_{1-\frac{\alpha}{2},n}}\right)$	$\left(0, \frac{\cdot}{\chi^2_{1-\alpha,n}}\right)$	$\left(\frac{\cdot}{\chi^2_{\alpha,n}}, \infty\right)$