Stimatori e loro Proprietà

Ogni volta che una statistica campionaria è usata per stimare un parametro, viene chiamata *stimatore*. La sua realizzazione nel campione osservato costituisce la *stima puntuale* del parametro.

Per esempio, la stima MV per la media di una popolazione normale è le media campionaria osservata $\bar{x}$ e lo stimatore è la media campionaria $\bar{X}$.

Lo stimatore MV non è l'unico stimatore possibile di un dato parametro, esistono anche altri metodi di stima (come ad esempio il più famoso metodo dei Minimi Quadrati, o il metodo dei Momenti) e in alcune situazioni essi possono fornire risultati differenti.

Stimatori e loro Proprietà 2

Se per stimare un dato parametron θ esistono diversi stimatori alternativi, le stime fornite da alcuni di essi potrebbero risultare migliori di quelle fornite dagli altri.

Uno dei punti fondamentali dell'approccio frequentista all'inferenza statistica riguarda proprio le

caratteristiche (o proprietà) desiderabili

che rendono uno stimatore migliore di un altro.

Per esempio, supponiamo che il parametro d'interesse sia il centro di simmetria di una distribuzione simmetrica. Sia la media che la mediana campionarie sono stimatori ragionevoli, ma:

Qual è migliore? Secondo quale criterio? Sotto quali condizioni?

Stimatori e loro Proprietà 3

Diventa quindi necessario stabilire dei criteri per poter scegliere in maniera accurata tra stimatori alternativi.

Tali criteri si basano sulle caratteristiche delle loro distribuzioni campionarie, in relazione al vero valore θ^* del parametro.

La prima proprietà, detta ***Non distorsione*** **(*o correttezza*)**, si riferisce alla media della distribuzione:

Uno stimatore *T* del parametro θ è detto *non distorto* se

$$E_\theta(T)=\theta, \text{ per tutti i valori } \theta \in \Theta$$

dove Θ è lo spazio parametrico del modello statistico considerato.

Stimatori e loro Proprietà 4

Uno stimatore *T* la cui distribuzione presenta un valore atteso diverso da θ è detto *distorto* e la differenza

$$B(\theta)=E_\theta(T) - \theta$$

è detta *distorsione* (in inglese *bias*, da cui la lettera B che la indica).

Esempio: La media campionaria è uno stimatore non distorto della media di popolazione μ, a prescindere da quale sia la famiglia $f_\theta(x)$ di distribuzioni.

La varianza campionaria S^2 è uno stimatore non distorto di σ^2. La dimostrazione è immediata se il campione proviene da una popolazione normale, in quanto il valore atteso di S^2 è semplicemente pari al prodotto tra il valore atteso di una v.c. χ^2 con n-1 g.d.l. e il rapporto $\sigma^2/(n-1)$. La non distorsione si può comunque dimostrare anche in casi più generali.

Al contrario, lo stimatore MV di σ^2 è distorto. Infatti il suo valore atteso è $n\sigma^2/(n\text{-}1)$.

L'eventuale distorsione, tuttavia, non è la cosa più importante di cui preoccuparsi nel valutare la validità di uno stimatore!

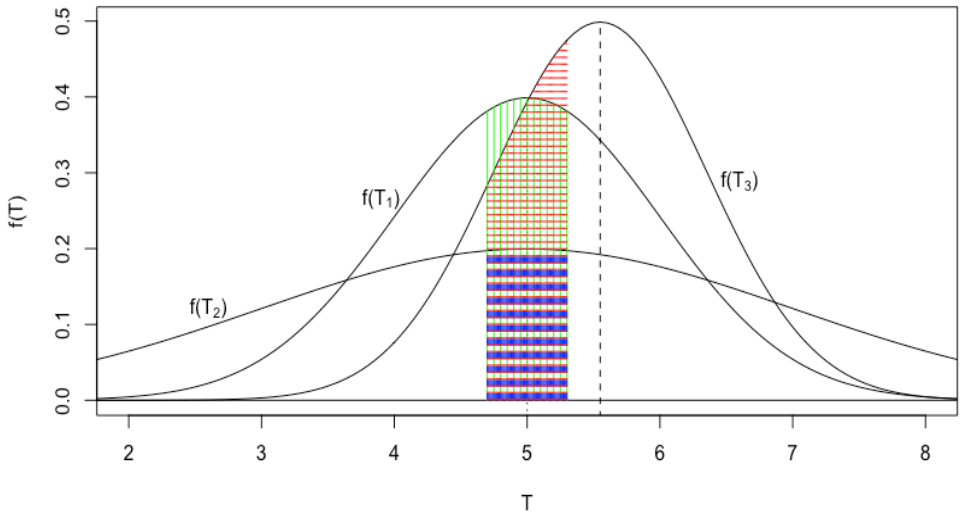
Per esempio, perfino la prima osservazione, cioè X_1, è uno stimatore non distorto della media di una popolazione Normale. Tuttavia sarebbe alquanto scarso!

Una migliore valutazione della 'bontà' di uno stimatore potrebbe basarsi sulla 'distanza' da θ

… ma come si misurano le distanze in statistica?

Consideriamo la situazione rappresentata nella figura seguente, dove si ipotizzano tre stimatori alternativi per un dato parametro di popolazione, il cui vero valore è 5: T_1 e T_2 sono non distorti, mentreT_3 è distorto. Le aree colorate sottese a ciascuna curva rappresentano le probabilità che il corrispondente stimatore assuma valori in un dato intorno di 5.

T_3 è certamente migliore di T_2 (l'area blu è la più piccola di tutte) e forse è quasi altrettanto buono quanto T_1 (l'area rossa è quasi identica a quella verde).



Errore Quadratico Medio (Mean Square Error) e Efficienza

E' definito da:

$$MSE_\theta(T)=E_\theta[(T-\theta)^2]=E_\theta\{[T-E_\theta(T)+E_\theta(T)-\theta]^2\}=$$

$$=E_\theta\{[T-E_\theta(T)]^2\}+E_\theta\{[E_\theta(T)-\theta]^2\}+2E_\theta\{[T-E_\theta(T)]\cdot[E_\theta(T)-\theta]\}=$$

$$=Var_\theta(T)+[B(\theta)]^2+2[E_\theta(T)-\theta]\cdot E_\theta\{[T-E_\theta(T)]\}=$$

$$=Var_\theta(T)+[B(\theta)]^2$$

L'indice MSE tiene conto anche di un possibile trade-off tra distorsione e varianza: in molte situazioni, infatti, è possibile ridurre la distorsione soltanto aumentando la varianza (e viceversa).

Tra stimatori alternativi, si dovrà preferire quello con MSE più piccolo.

Se uno stimatore è non distorto, allora l'MSE coincide con la varianza.

Quindi, stimatori non distorti alternativi, quello migliore è lo stimatore con la varianza più piccola.

Esso è detto stimatore ***più efficiente***.

Se esiste uno stimatore non distorto T^* di θ (o di $\tau(\theta)$), tale che $\text{Var}(T^*) \leq \text{Var}(T)$, per ogni altro stimatore non distorto T di θ (o di $\tau(\theta)$), allora T^* è detto

Uniformly Minimun Variance Unbiased Estimator (**UMVUE**)

Per stabilire se uno stimatore è UMVU non è necessario fare confronti con tutti i possibili stimatori non distorti.

Infatti, il seguente teorema definisce un estremo inferiore per la varianza degli stimatori non distorti:

Disuguaglianza di Cramer-Rao: Dato il modello statistico $\mathcal{M}$ e uno stimatore non distorto T di θ, sotto certe condizioni di regolarità della $f_\theta(x)$ risulta:

$$\mathrm{Var}(T) \geq 1/\mathrm{I}(\theta) = 1/\{nE[\partial \log f_\theta(x)/\partial\theta]^2\}$$

dove $\mathrm{I}(\theta)$ è l'Informazione di Fisher.

Esempio: Consideriamo la popolazione $X \sim \text{Ber}(\theta)$ e stimiamo la proporzione θ mediante la media campionaria.

Sappiamo che tale stimatore è non distorto, quindi $E(\bar{X})=\theta$, mentre $\text{Var}(\bar{X})=\theta(1-\theta)/n$.

Secondo la disuguaglianza di Cramer-Rao, abbiamo:

$\log f_\theta(x)=x\log\theta + (1-x)\log(1-\theta)$; $\quad \partial \log f_\theta(x)/\partial\theta = x/\theta - (1-x)/(1-\theta)$

$\partial^2 \log f_\theta(x)/\partial\theta^2 = -x/\theta^2 - (1-x)/(1-\theta)^2$

$I_1(\theta)=E[\partial \log f_\theta(x)/\partial\theta]^2 = -E[\partial^2 \log f_\theta(x)/\partial\theta^2] = -E[-x/\theta^2 - (1-x)/(1-\theta)^2]=$

$=1/\theta + 1/(1-\theta) = 1/[\theta(1-\theta)]$

$\Rightarrow I(\theta) = n/[\theta(1-\theta)]$ e $1/I(\theta)= \theta(1-\theta)/n$

Quindi $\bar{X}$ è uno stimatore UMVU.

Proprietà Asintotiche

L'idea chiave è che, aumentando la dimensione campionaria, possiamo migliorare il comportamento dello stimatore.

Per esempio, uno stimatore distorto può diventare non distorto quando *n* diventa sufficientemente grande, cioè si allarga la base informativa:

E' il caso della varianza campionaria non corretta (stimatore MV per la varianza di una popolazione Normale), il cui valore atteso $(n-1)\sigma^2/n \to \sigma^2$ per $n \to \infty$.

- ***Non distorsione (correttezza) asintotica:*** come nel precedente esempio, uno stimatore *T* di θ è asintoticamente non distorto se
$$E_\theta\left(T_n(\boldsymbol{X})\right)\underset{n\to\infty}{\longrightarrow}\theta$$
- ***Consistenza:*** Una statistica T_n è uno stimatore consistente di θ, se $T_n \xrightarrow{p} \theta$ per $n \to \infty$, cioè:
$\lim_{n\to\infty} \Pr(|T_n - \theta| < \varepsilon) = 1$, oppure: $\lim_{n\to\infty} \Pr(|T_n - \theta| \geq \varepsilon) = 0$
- ***Normalita' Asintotica***: T_n è asintoticamente Normale se:
$\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, V)$,
dove V è la varianza (o matrice di covarianza) asintotica.

  Se V è più piccola della varianza di tutti gli altri stimatori, allora T_n è detto ***Best Asymptotically Normal*** (**BAN**)

Proprietà dello stimatore di Massima Verosimiglianza

Sia T lo stimatore MV di un parametro θ. Allora:

- T è *invariante*: se T è lo stimatore MV del parametro θ, lo stimatore MV di una funzione $\tau(\theta)$ invertibile (cioè biunivoca, o uno a uno) è $\tau(T)$
- Se S è una statistica sufficiente per θ, allora lo stimatore MV di θ (se esiste) è funzione di S

- Se esiste uno stimatore non distorto e efficiente per θ, esso è lo stimatore MV
- *T* è asintoticamente non distorto ed efficiente (la sua varianza asintotica raggiunge il limite di Cramer-Rao), quindi è consistente
- *T* è asintoticamente distribuito secondo una Normale

In una parola, *T* è *Consistente e Asintoticamente Normale* (CAN)

Stima intervallare e Intervalli di Confidenza

Supponiamo di disporre di uno stimatore *T* per θ.

La sua realizzazione empirica nel campione osservato ci fornisce una stima puntuale per il parametro, cioè $t = \hat{\theta}$

… tuttavia non è una buona idea!

In effetti, se *T* è una variabile aleatoria continua, siamo quasi sicuri che la nostra stima sarà sbagliata, essendo

$$\Pr(T = \theta) \approx 0!$$

Tuttavia, potremmo valutare altre probabilità …

Intervalli di Confidenza 2

Per esempio, nel caso in cui si conosca la distribuzione del cosiddetto

$$\textit{errore campionario } E=(T-\theta),$$

si potrebbe valutare quanto siano probabili affermazioni come la seguente:

"in valore assoluto l'errore campionario è inferior a ε", per un dato ε piccolo a piacere.

Ma la distribuzione dell'errore campionario è la stessa di T.

Per questo motive è importante specificare la distribuzione campionaria di uno stimatore!

Supponiamo di voler fare inferenza sulla media incognita di una data popolazione.

Se la popolazione è Normale, o la dimensione campionaria *n* è sufficientemente grande (in modo da poter applicare il TCL), possiamo porre:

$$\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right) \Longrightarrow E = (\bar{X} - \mu) \sim N(0; \frac{\sigma^2}{n})$$

e quindi possiamo calcolare la probabilità Pr($|E| < \varepsilon$).

Esempio: Supponiamo di considerare n=16, σ^2=25 e ε=0.05;

avremmo:

$$\Pr(|E| \leq \varepsilon) = \Pr\left(|Z| \leq \frac{0.05}{\sigma/\sqrt{n}} = \frac{0.05}{5/4} = 0.04\right)$$

$$= \Phi(0.04) - \Phi(-0.04) = 2 * \Phi(0.04) - 1 = 0.032$$

In via più generale, per un dato livello α, risulta:

$$\Pr\left(|E| \leq \varepsilon_{\alpha/2}\right) = \Pr\left(|Z| \leq z_{\alpha/2}\right) = 1 - \alpha,$$

con $z_{\alpha/2}$: $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$

La precedente espressione può essere riscritta in forma equivalente:

$$1-\alpha = \Pr\left(\frac{|\bar{X}-\mu|}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) =$$

$$= \Pr\left(-z_{\alpha/2}\sigma/\sqrt{n} \leq \bar{X}-\mu \leq z_{\alpha/2}\sigma/\sqrt{n}\right) =$$

$$= \Pr\left(\bar{X}-z_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \bar{X}+z_{\alpha/2}\sigma/\sqrt{n}\right)$$

In altre parole, la probabilità che un intervallo casuale $\Im$,

con estremi casuali $\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}$ e $\bar{X} + z_{\alpha/2}\sigma/\sqrt{n}$,

possa includere la media di popolazione è $1 - \alpha$.

L'intervallo casuale $\Im$ è un *intervallo di probabilità*, in quanto i suoi estremi dipendono dalla probabilità $1 - \alpha$ di includere μ.

Analogamente alla distribuzione campionaria di $\bar{X}$, questa probabilità si riferisce all'intero spazio campionario.

Pertanto, così come la media campionaria $\bar{X}$ è una variabile casuale con infinite realizzazioni possibili,

una delle quali è il valore $\bar{x}$ osservato nel campione estratto,

l'intervallo casuale $\Im$ è un oggetto aleatorio con infinite realizzazioni possibili,

una delle quali è l'intervallo osservato

$$\lambda = [\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}; \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}]$$

Allo stesso tempo, mentre sappiamo che $\Pr(\bar{X} = \mu) \approx 0$,

non ha alcun senso considerare $\Pr(\bar{x} = \mu)$,

poiché la media osservata non è un oggetto casuale (è infatti una costante, essendo una delle possibili realizzazioni della variabile casuale $\bar{X}$).

Di fatti risulta:

$$\Pr(\bar{x} = \mu) = \begin{cases} 0, \text{se } \bar{x} \neq \mu \text{ e conseguentemente l'evento } \bar{x} = \mu \text{ è impossibile} \\ 1, \text{se effettivamente } \bar{x} \equiv \mu \text{ e quindi l'evento } \bar{x} = \mu \text{ è certo} \end{cases}$$

Analogamente, non ha senso dire che l'intervallo osservato (la realizzazione osservata di un intervallo casuale) ha probabilità $1-\alpha$ di includere la media di popolazione.

Di fatti risulta:

$$\Pr\left(\bar{x}-\frac{z_{\alpha/2}\sigma}{\sqrt{n}}\leq\mu\leq\bar{x}+\frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right)=\begin{cases}0,\text{ se }\mu\text{ non è incluso (evento impossibile)}\\1,\text{ se }\mu\text{ è effettivamente incluso (evento certo)}\end{cases}$$

In altre parole, parliamo di probabilità quando ci riferiamo all'intervallo casuale, ma non possiamo più farlo quando ci riferiamo all'intervallo osservato i.

Tuttavia, se l'intervallo aleatorio ha probabilità $1 - \alpha$ di includere la media di popolazione,

sappiamo anche che, tra tutti gli intervalli osservabili che appartengono allo spazio campionario, una frazione pari a $1 - \alpha$ include la media di popolazione, mentre la parte rimanente no.

Noi non sappiamo se il nostro campione osservato appartenga al primo gruppo (nel quale gli intervalli osservati contengono la media di popolazione) o al secondo.

Tuttavia, poiché il primo gruppo è molto più ampio (ammesso che $1 - \alpha$ sia elevato),

possiamo ***confidare*** (avere fiducia) che il nostro campione vi appartenga e quindi che il nostro intervallo osservato i includa la media di popolazione.

Il nostro ***livello di confidenza*** (o di fiducia) è proprio $1 - \alpha$!

Per questo motivo, la stima intervallare per μ fornita da

$$\lambda = [\bar{x} - z_{\alpha/2}\sigma/\sqrt{n}; \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}]$$

è detta

Intervallo di confidenza

e $1 - \alpha$ è il *livello di confidenza*.

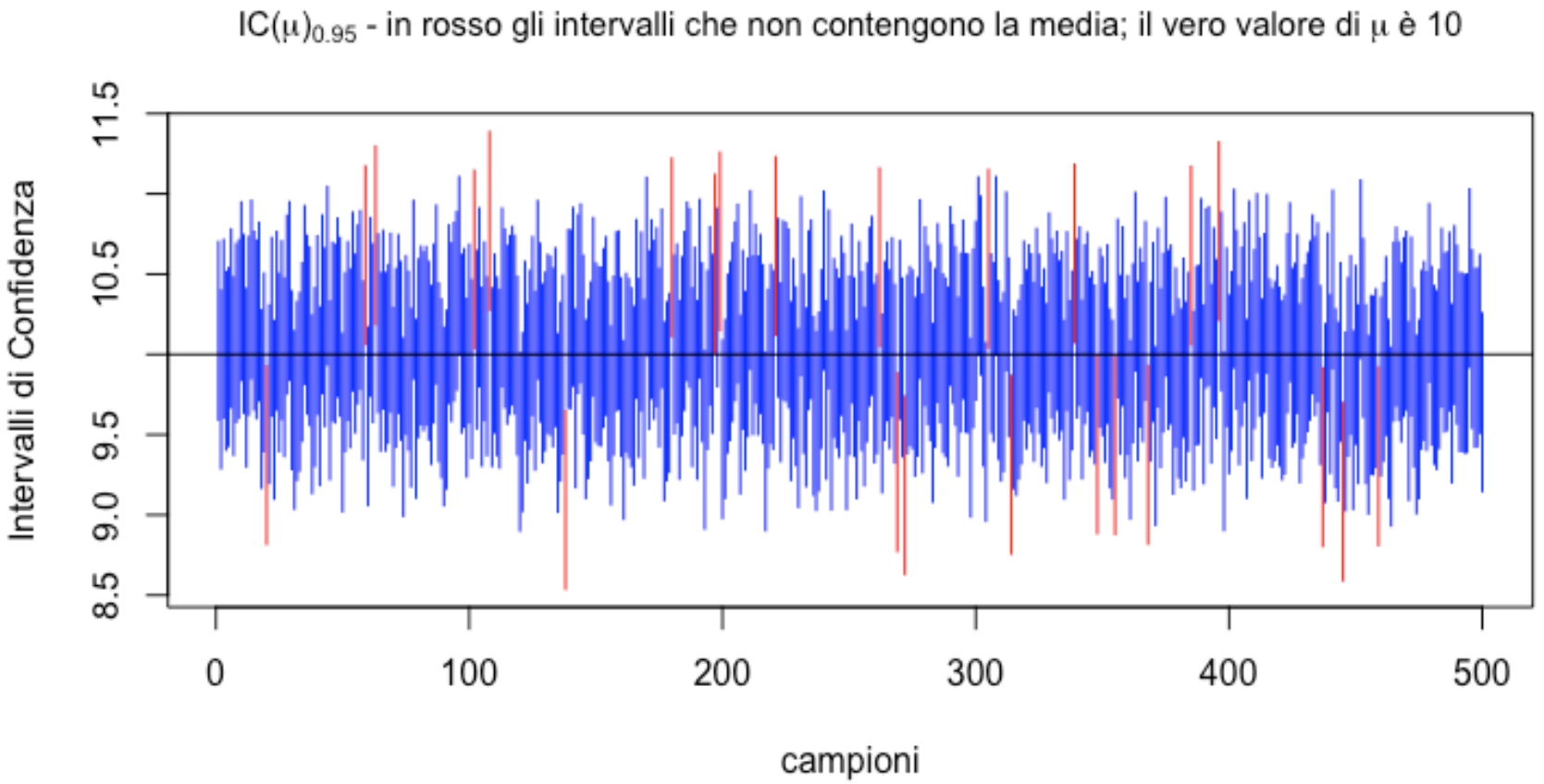
Esempio: Con R possiamo costruire lo spazio campionario generato estraendo 500 diversi campioni di 50 unità da una popolazione N(10, 16).

Per ognuno di essi costruiamo un intervallo di confidenza al livello di confidenza 0.95.

Possiamo osservare che la maggior parte di essi include la media di popolazione, mentre solo una piccola parte no.

Graficamente abbiamo:

Intervalli di Confidenza 14



Più in generale, siano $TL=tl(X_1, X_2, \ldots, X_n)$ e $TU=tu(X_1, X_2, \ldots, X_n)$ due statistiche campionarie tali che

a) $tl(x_1, x_2, \ldots, x_n) \leq tu(x_1, x_2, \ldots, x_n)$ per ogni possibile realizzazione del campione casuale $(X_1, X_2, \ldots, X_n)$

b) $Pr(TL \leq \theta \leq TU) \geq 1 - \alpha, \forall \theta \in \Theta$

Quindi TL e TU sono gli estremi di un intervallo casuale

$\Im$ =[TL, TU], con probabilità almeno pari a $1 - \alpha$.

La realizzazione osservata di $\Im$, $\imath = [tl, tu]$, è detta

intervallo di confidenza per θ, al livello di confidenza $1 - \alpha$.

A seconda del parametro d'interesse e del suo stimatore, la distribuzione campionaria di riferimento cambia, e di conseguenza i quantili da utilizzare nella determinazione degli estremi dell'intervallo. Tuttavia, il ragionamento sottostante è sempre lo stesso!

- IC per la media di una popolazione Normale con varianza incognita

  Al posto della condizione di probabilità

$$1-\alpha \leq \Pr\left(\frac{|\bar{X}-\mu|}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right),$$

  dovremmo considerare

$$1-\alpha = \Pr\left(\frac{|\bar{X}-\mu|}{\sqrt{S^2/n}} \leq t_{n-1;\alpha/2}\right)$$

  dove $t_{n-1;\alpha/2}$: $\mathrm{F}\left(t_{n-1;\alpha/2}\right) = 1-\alpha/2$, essendo F(.) la funzione di ripartizione.

L'intervallo casuale diventa quindi

$$\Im = \left[\bar{X} - t_{n-1;\frac{\alpha}{2}}\sqrt{\frac{S^2}{n}};\ \bar{X} + t_{n-1;\alpha/2}\sqrt{\frac{S^2}{n}}\right]$$

mentre la sua realizzazione empirica è:

$$\lambda = \mathrm{CI}(\mu)_{1-\alpha} = \left[\underbrace{\bar{x} - t_{n-1;\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}}_{tl};\ \underbrace{\bar{x} + t_{n-1;\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}}_{tu}\right]$$

- IC per la varianza di una popolazione Normale

Ricordando che

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

possiamo considerare la probabilità

$$1-\alpha = \Pr\left(\chi^2_{n-1;1-\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1;\alpha/2}\right)$$

dove $\chi^2_{n-1;\alpha/2}: \mathrm{F}\left(\chi^2_{n-1;\alpha/2}\right) = 1-\alpha/2$

La precedente condizione probabilistica può essere riscritta nel seguente modo:

$$1-\alpha = \Pr\left(\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}\right)$$

cioé, $\mathrm{TL} = \frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}}$ e $\mathrm{TU} = \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$, quindi

$$\lambda = \mathrm{CI}(\sigma^2)_{1-\alpha} = \left[\underbrace{\frac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}}_{tl}; \underbrace{\frac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}}_{tu}\right]$$

- IC per la proporzione di una popolazione Bernoulliana

  Ricordando che, sef $X \sim \text{Ber}(\pi)$ e n è abbastanza grande da poter applicare il TCL, si ha:

$$\frac{\bar{X}-\pi}{\sqrt{\pi(1-\pi)/n}} \approx N(0,1),$$

  possiamo partire dalla condizione probabilistica

$$1-\alpha = \Pr\left(\frac{|\bar{X}-\pi|}{\sqrt{\pi(1-\pi)/n}} \leq z_{\alpha/2}\right)$$

Come nel caso della Student-*t* (popolazione normale con varianza incognita), la varianza è incognita e possiamo stimarla mediante $\bar{X}(1-\bar{X})/n$. Ora,tuttavia, poiché *n* è sufficientemente grande l'approssimazione alla distribuzione Normale continua a valere, perciò:

$$\Im = \left[\bar{X} - z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}};\ \bar{X} + z_{\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right]$$

con realizzazione osservata

$$\lambda = \mathrm{CI}(\pi)_{1-\alpha} = \left[\bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}};\ \bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

Intervalli di Wald Approssimati

In molti casi è impossibile costruire intervalli di confidenza 'esatti' (cioè utilizzando le distribuzioni esatte).

Consideriamo un campione scasuale ($X_1, X_2, \ldots, X_n$) estratto da una popolazione con probabilità o funzione di densità $f(x;\theta)$.

Sia T_n uno stimatore consistente di θ con standard error se(T_n).

Allora la statistica

$$\mathrm{Z}(\theta) = \frac{T_n - \theta}{se(T_n)},$$

detta statististica-zeta, sotto condizioni di regolarità si distribuisce asintoticamente secondo una N(0; 1).

Approximate Wald intervals 2

L'intervallo

$$[T_n - z_{\frac{\alpha}{2}}\, se(T_n);\ [T_n - z_{\frac{\alpha}{2}}\, se(T_n)]$$

rappresenta un intervallo di confidenza approssimato per θ all' $(1-\alpha)100\%$ ed è detto ***intervallo di confidence approssimato di Wald***.

Per esempio, l'intervallo di confidenza asintotico definito per il parametro di una popolazione Bernoulliana è intervallo di confidence approssimato di Wald.