Capitolo 12
Stima per intervallo

Capitolo 12

- Stima per intervallo
- Analogie tra la stima puntuale e per intervallo
- Intervallo di confidenza per la media
- Intervallo di confidenza per la proporzione
- Intervallo di confidenza per la varianza
- Determinazione della numerosità campionaria

Stima per intervallo

- Sia X una v.c. che rappresenta un carattere osservato su una popolazione. Supponiamo che la v.c. sia definita da una funzione di probabilità $f(x;\theta)$ dipendente dal parametro θ incognito.
- Sia $X_1,\dots,X_n$ un campione di dimensione n e $x_1,\dots,x_n$ il corrispondente campione osservato.

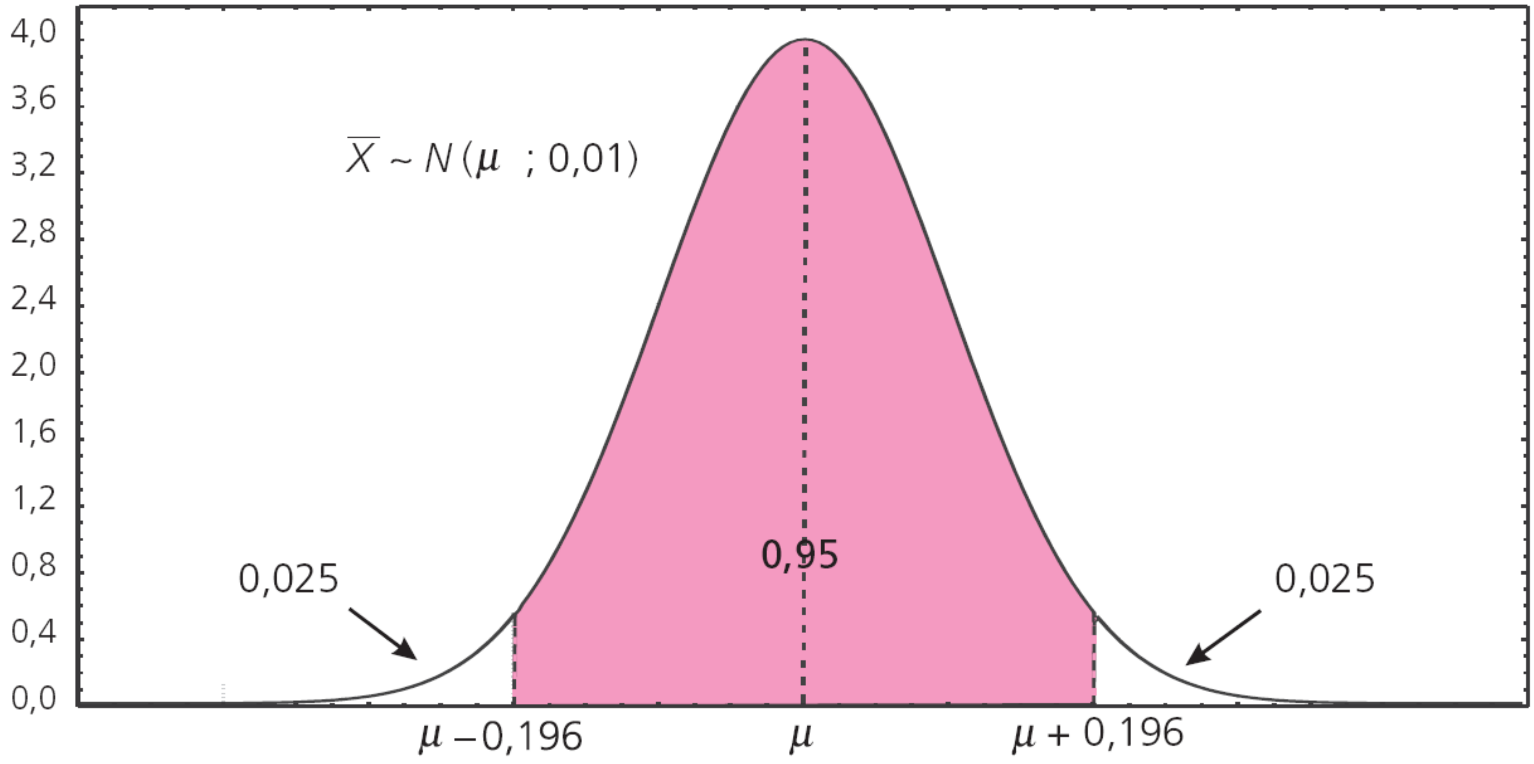
Obiettivo:

Determinare due statistiche campionarie:

$$L_1 = L_1(X_1,\dots,X_n) \qquad\qquad L_2 = L_2(X_1,\dots,X_n)$$

tali che $L_1 \leq L_2$ per ogni possibile campione e che l'intervallo $[L_1, L_2]$ contenga il parametro θ con probabilità $1-\alpha$

FIGURA 12.2.1
Distribuzione della media campionaria per una popolazione Normale con varianza nota $\sigma^2 = 0,1$ e dimensione campionaria $n = 10$.



Stima per intervallo

L'intervallo casuale $[L_1(X_1,\dots,X_n), L_2(X_1,\dots,X_n)]$ si definisce **intervallo di confidenza** di livello $1-\alpha$ per il parametro θ se contiene con probabilità $1-\alpha$ il parametro ignoto θ della popolazione, ossia:

$$\boxed{Pr[L_1(X_1,\dots,X_n) \le \theta \le L_2(X_1,\dots,X_n)] = 1-\alpha}$$

In genere si fissano valori di $1-\alpha$ pari a 0,99; 0,95; 0,90 e questo valore viene detto livello di confidenza. Una volta estratto il campione si ottiene l'intervallo di confidenza stimato.

Nota:
Non è possibile sapere se l'intervallo stimato contenga o meno il valore vero del parametro; d'altra parte se si estraesse dalla popolazione un numero sufficientemente elevato di campioni e calcolassimo i corrispondenti intervalli di confidenza, circa il $100(1-\alpha)\%$ di questi conterrebbe il parametro ignoto.

Stima per intervallo - esempio

Esempio:

Sia $X \sim N\left(\mu; \sigma^2 = 0{,}1\right)$

Si consideri un campione di dimensione n=10.

La media campionaria è una v.c. che si distribuisce come

$$\overline{X} \sim N\left(\mu; \sigma^2 / n = 0{,}01\right)$$

$Z = \left(\overline{X} - \mu\right) / \sqrt{0{,}01}$ è una v.c. Normale standardizzata

$$Pr\left(-1{,}96 \le Z \le +1{,}96\right) = 0{,}95$$

da cui possiamo ricavare che

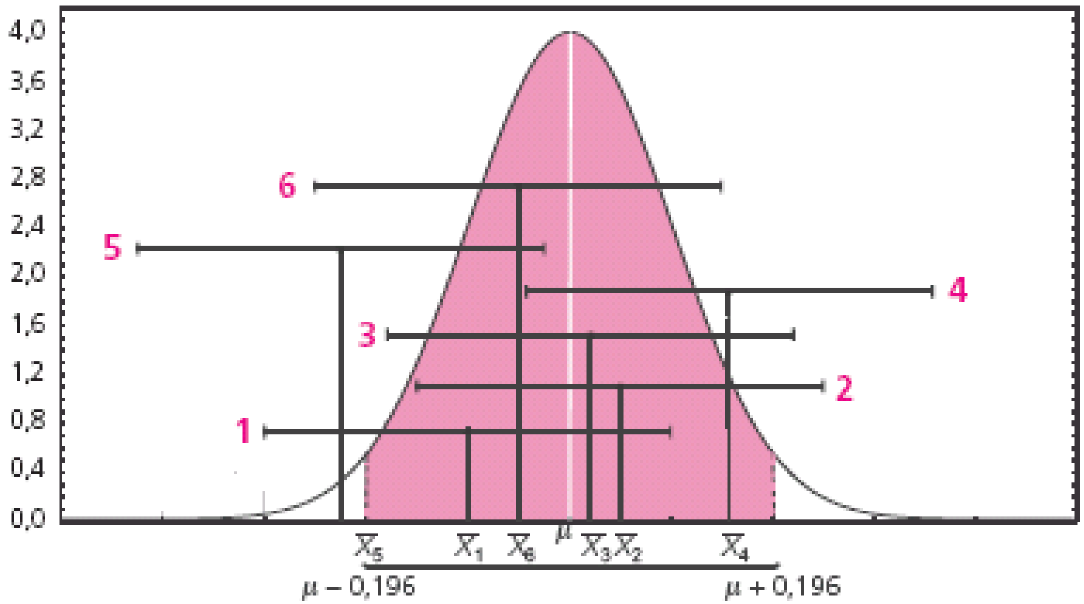
$$Pr\left(\overline{X} - 1{,}96 \cdot 0{,}1 \le \mu \le \overline{X} + 1{,}96 \cdot 0{,}1\right) = 0{,}95$$

Se dal campione estratto si osserva un valore della media pari a $\overline{x} = 10$, l'intervallo stimato risulta pari a $\left[9{,}804\,; 10{,}196\right]$

Stima per intervallo - esempio

Esempio (continua)

Nella seguente figura si mostrano, in corrispondenza di 6 campioni osservati, gl'intervalli di confidenza stimati per la media della popolazione a un livello di confidenza 0,95.



Osserviamo che dal campione 5 si ottiene un intervallo stimato che non contiene il vero parametro della popolazione.

Analogie con la stima puntuale

Nella seguente tabella sono riportate analogie e differenze tra la stima puntuale e la stima per intervallo.

	Stima puntuale	**Stima intervallare**
Campione casuale	$X_1, \ldots, X_n$	$X_1, \ldots, X_n$
Obiettivo	Stima puntuale del parametro θ	Stima per intervallo del parametro θ
Strumento	Stimatore puntuale: $T = T(X_1, \ldots, X_n)$	Stimatore intervallo di confidenza: $[L_1, L_2] = [L_1(X_1, \ldots, X_n), L_2(X_1, \ldots, X_n)]$
Accuratezza	Errore quadratico medio $MSE(T) = E[(T - \theta)^2]$	Livello di confidenza $P(L_1 \leq \theta \leq L_2) = 1 - \alpha$
Dati campionari	$x_1, \ldots, x_n$	$x_1, \ldots, x_n$
Risultato	Stima puntuale: $t = T(x_1, \ldots, x_n)$	Intervallo di confidenza stimato: $[l_1, l_2] = [L_1(x_1, \ldots, x_n), L_2(x_1, \ldots, x_n)]$

Intervallo di confidenza per la media (varianza nota)

Sia X una v.c. che rappresenta un carattere osservato su una popolazione. Supponiamo che la v.c. sia distribuita come una Normale con varianza nota. Allora sappiamo che:

$$\bar{X} \sim N\left(\mu, \sigma^2/n\right) \quad \Longrightarrow \quad Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

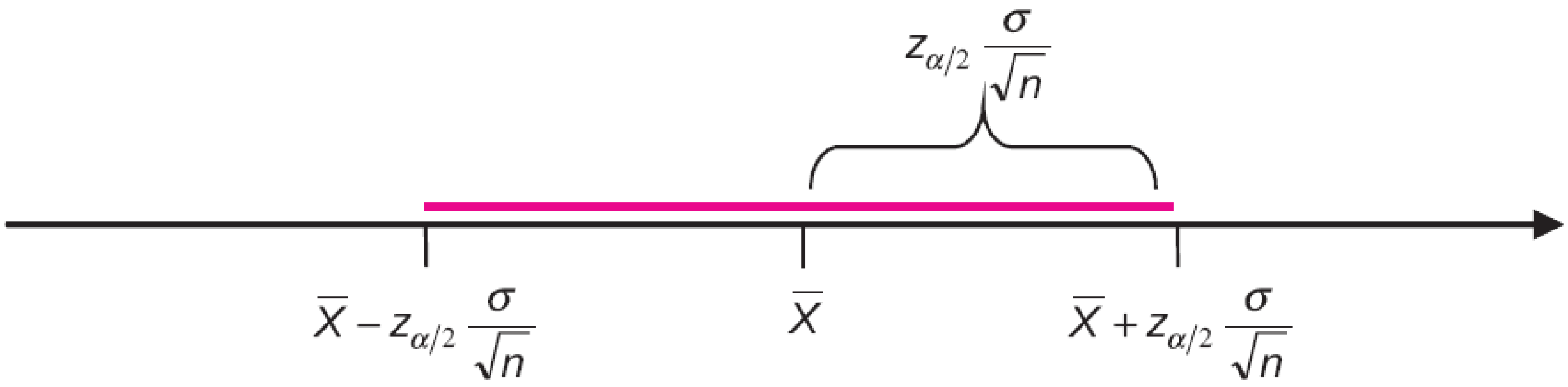
$$P\left(-z_{\alpha/2} \leq Z \leq +z_{\alpha/2}\right) = 1-\alpha$$

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

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

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

FIGURA 12.3.1
Ampiezza di un intervallo di confidenza.



Intervallo di confidenza per la media (varianza nota)

Dato un campione casuale estratto da una popolazione Normale con media ignota e varianza nota, l'intervallo di confidenza per la media della popolazione al livello di confidenza $1-\alpha$ è:

$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Esempio

Siano $n = 10$ $\sigma^2 = 9$ $1-\alpha = 0{,}99$

Dalle tavole della Normale standardizzata si ottiene

$$z_{\alpha/2} = z_{0{,}005} = 2{,}576$$

Se $\overline{x} = 4{,}924$ si ottiene:

$$\left[4{,}924 \pm 2{,}576\frac{\sqrt{9}}{\sqrt{10}}\right] \rightarrow [2{,}4802 , 7{,}3678]$$

Intervallo di confidenza per la media (varianza nota)

La lunghezza dell'intervallo di confidenza si ricava dalla differenza tra estremo superiore e estremo inferiore:

$$\text{Lunghezza} = 2z_{\alpha/2}\left(\sigma/\sqrt{n}\right)$$

Dipende da:

1. **la dimensione del campione**
2. **il livello di confidenza**
3. **la varianza della popolazione**

Intervenendo sulla dimensione del campione o sul livello di confidenza si può aumentare o diminuire la lunghezza dell'intervallo. Una volta fissati questi due elementi, al variare dei campioni estratti, la lunghezza degli intervalli corrispondenti rimane costante.

Intervallo di confidenza per la media (varianza nota)

La lunghezza dell'intervallo di confidenza si ricava dalla differenza tra estremo superiore e estremo inferiore:

$$\text{Lunghezza} = 2z_{\alpha/2}\left(\sigma/\sqrt{n}\right)$$

Esempio:

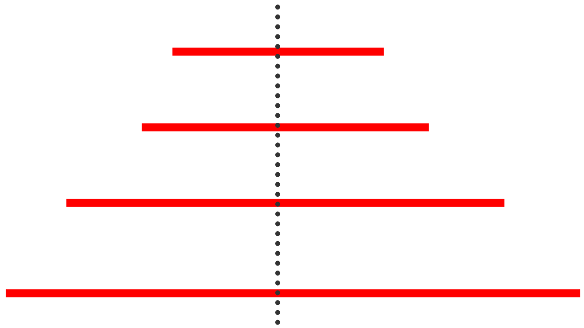
Fissato $1-\alpha$

$n = 100$

$n = 70$

$n = 50$

$n = 10$



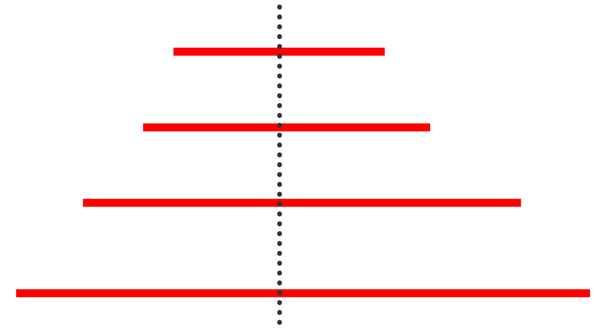
Fissato n

$1-\alpha = 0{,}85$

$1-\alpha = 0{,}90$

$1-\alpha = 0{,}95$

$1-\alpha = 0{,}99$



Intervallo di confidenza per la media (varianza ignota)

Sia X una v.c. che rappresenta un carattere osservato su una popolazione. Supponiamo che la v.c. sia distribuita come una Normale con media e varianza ignota.

Per stimare la varianza della popolazione si utilizza lo stimatore varianza campionaria corretta:

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

Pertanto la v.c. $T = (\bar{X} - \mu)/(S/\sqrt{n})$ si distribuisce come una v.c. *t-Student* con $n-1$ gradi di libertà.

Intervallo di confidenza per la media (varianza ignota)

Dato un campione casuale di dimensione n estratto da una popolazione Normale con media e varianza entrambe ignote, l'intervallo di confidenza per la media a livello $1-\alpha$ è dato da:

$$\left[\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

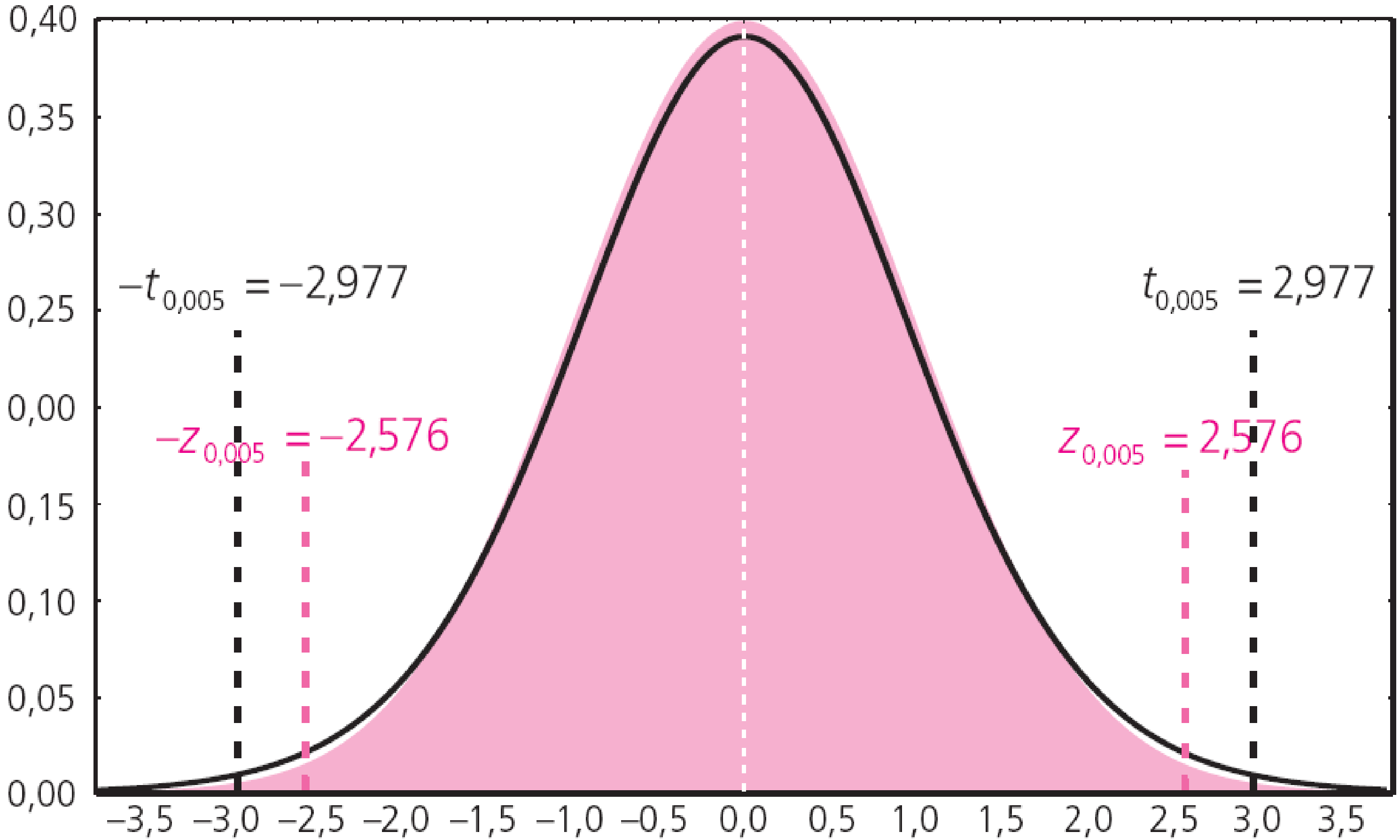
La lunghezza dell'intervallo di confidenza è data in questo caso da:

$$\text{Lunghezza} = 2t_{\alpha/2}\left(S/\sqrt{n}\right)$$

Nota

Al variare dei campioni estratti, la lunghezza degli intervalli corrispondenti non rimane costante poiché varia il valore di S

FIGURA 12.4.1
Centili della t-Student con 14 g.l. (nero) e della Normale standardizzata (blu).



Intervallo di confidenza per la media (popolazioni non Normali)

Quando non è nota la popolazione, ma il campione ha una dimensione sufficientemente grande, possiamo considerare un'approssimazione dell'intervallo di confidenza per la media ottenuta attraverso il teorema del limite centrale.

Per n sufficientemente grande possiamo utilizzare il seguente intervallo di confidenza a livello $1-\alpha$:

$$\left[\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right]$$

Intervallo di confidenza per una proporzione (campioni di dimensione elevata)

Quando la popolazione è riferita a un carattere che può assumere solo due modalità (popolazione Bernoulliana), siamo interessati all'intervallo di confidenza per una proporzione π, ad esempio, la proporzione di maschi nella popolazione. Come sappiamo un buon stimatore per π è la media campionaria $\bar{X}$.

Si ha: $E(\bar{X}) = \pi \quad V(\bar{X}) = \pi(1-\pi)/n$
inoltre, dal teorema del limite centrale sappiamo che al crescere della dimensione campionaria la distribuzione della $\bar{X}$ tende alla Normale, pertanto

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

Intervallo di confidenza per una proporzione (campioni di dimensione elevata)

$$1-\alpha \cong P\left(-z_{\alpha/2} \le \frac{\overline{X}-\pi}{\sqrt{\pi(1-\pi)/n}} \le +z_{\alpha/2}\right) =$$

$$= P\left(\overline{X} - z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}} \le \pi \le \overline{X} + z_{\alpha/2}\sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

Tuttavia gli estremi dell’intervallo dipendono ancora dal parametro incognito e dunque devono essere sostituiti con degli stimatori, ottenendo il seguente intervallo di confidenza al livello $1-\alpha$:

$$\left[\overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\,,\overline{X} + z_{\alpha/2}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}\right]$$

Una regola pratica: se $n\overline{x} \ge 5$ **e** $n(1-\overline{x}) \ge 5$ **, allora l’aprossimazione Normale è adeguata.**

Esempio

Si vuole ottenere una stima intervallare della proporzione di fumatori presenti in una certa regione. A tal fine viene osservato un campione casuale di 120 persone, di cui 78 sono fumatori.
Quindi la stima puntuale della proporzione è data da:

$$\bar{x} = 78/120 = 0{,}65$$

e l'intervallo di confidenza al livello $1-\alpha = 0{,}95$ **è:**

$$\left[0{,}65 - z_{0{,}025}\sqrt{\frac{0{,}65(0{,}35)}{120}}\,,\, 0{,}65 + z_{0{,}025}\sqrt{\frac{0{,}65(0{,}35)}{120}}\right] = [0{,}56\,,\,0{,}74]$$

In questo caso: $n\bar{x} = 120 \cdot 0{,}65 = 78 \geq 5$

$$n(1-\bar{x}) = 120 \cdot 0{,}35 = 42 \geq 5$$

Intervallo di confidenza per la varianza

Si consideri una popolazione Normale con media e varianza entrambe ignote.
Come stimatori puntuali dei due parametri si possono utilizzare: $\overline{X} \to \mu \qquad S^2 \to \sigma^2$

Si può dimostrare che la v.c. $(n-1)S^2 / \sigma^2 \sim \chi^2_{n-1}$. Pertanto:

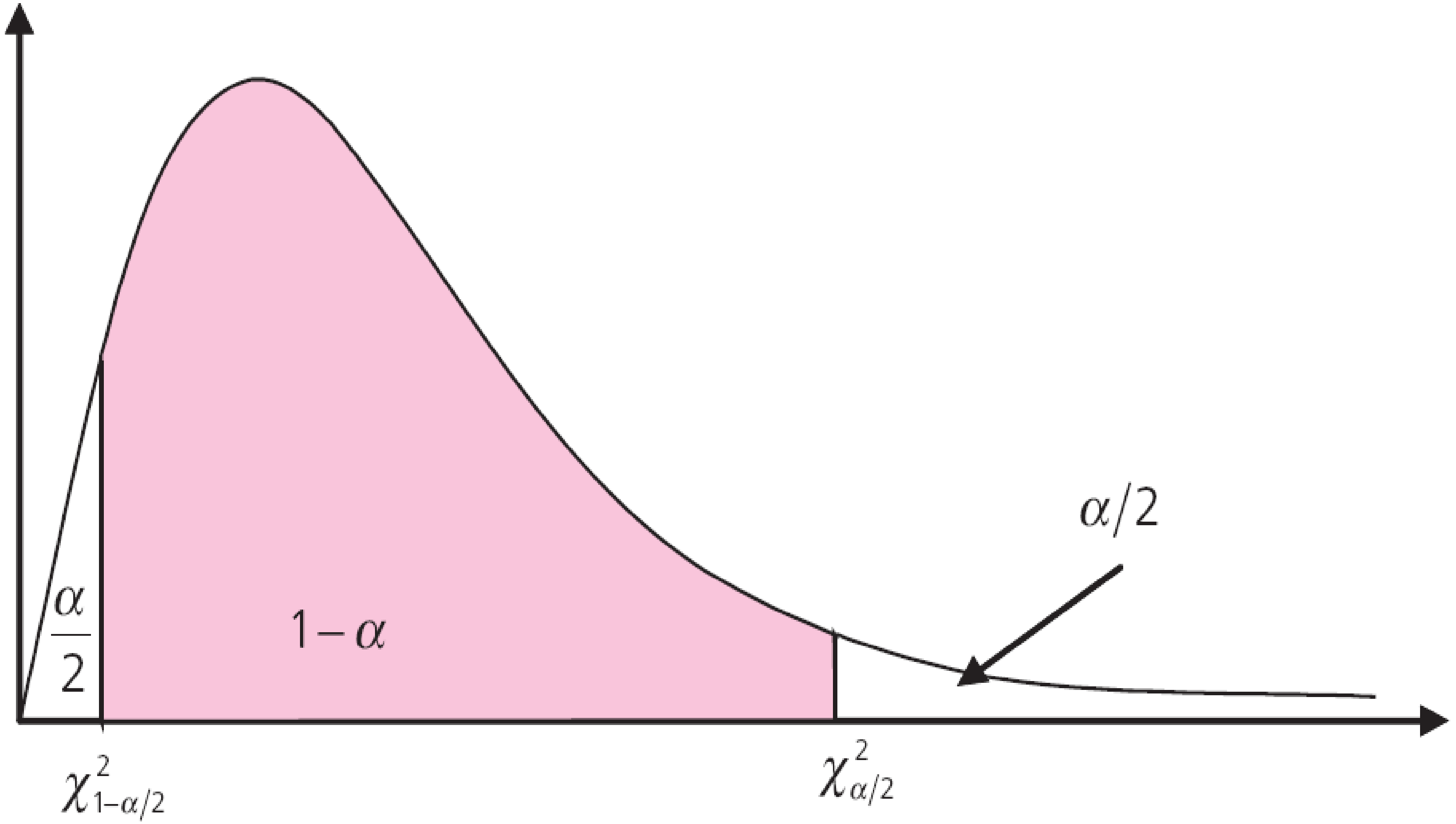
$$1-\alpha = P\left(\chi^2_{1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{\alpha/2} \right) =$$

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

E quindi l'intervallo per la varianza al livello $1-\alpha$:

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

FIGURA 12.6.1
Centili di una distribuzione Chi-quadrato.



Determinazione numerosità campionaria

Si consideri una popolazione Normale con media ignota e varianza nota. Ci si può chiedere quale debba essere la dimensione campionaria necessaria ad assicurare che la semi-lunghezza dell'intervallo non superi un certo valore δ.
Dalla formula della lunghezza dell'intervallo di confidenza si ottiene:

$$n = \left(z_{\alpha/2} \frac{\sigma}{\delta} \right)^2$$

Quando la popolazione non è Normale o la varianza è ignota si utilizza:

$$n = \left(z_{\alpha/2} \frac{S}{\delta} \right)^2$$

tuttavia in questo caso è necessario che la numerosità risultante sia sufficientemente grande (>120)

Determinazione numerosità campionaria

Nel caso di popolazione Bernoulliana si ha:

$$n = z_{\alpha/2}^{2} \frac{\hat{\pi}(1-\hat{\pi})}{\delta^{2}}$$

dove $\hat{\pi}$ indica il valore della stima preliminare di π.

Se non si hanno informazioni a priori sul parametro incognito si usa fissare valore prudenziale pari a

$$\hat{\pi} = 0{,}5$$

Esempio

Si vuole stimare la numerosità necessaria per ottenere un intervallo di confidenza per π (ad es. la proporzione di persone propense a dare la preferenza a un certo candidato) in modo tale che la semi-lunghezza dell'intervallo di confidenza al livello 0,95 non sia superiore a 0,05.

$$n = 1{,}96^2 \frac{0{,}5(0{,}5)}{0{,}05^2} = 384{,}16 \approx 385$$

TABELLA 12.7.1 Intervalli di confidenza: schema riassuntivo

Popolazione	Parametri noti	Parametri ignoti	Dimensione del campione	Stima dell'intervallo di confidenza
		Intervallo di confidenza per la media		
Normale	σ^2	μ	qualsiasi	$\bar{x} \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
Normale		μ, σ^2	piccolo	$\bar{x} \pm t_{\alpha/2,(n-1)} \dfrac{s}{\sqrt{n}}$
Qualsiasi		μ, σ^2	grande	$\bar{x} \pm z_{\alpha/2} \dfrac{s}{\sqrt{n}}$
		Intervallo di confidenza per una proporzione		
Bernoulli		π	grande	$\bar{x} \pm z_{\alpha/2} \sqrt{\dfrac{\bar{x}(1-\bar{x})}{n}}$
Bernoulli		π	piccolo	*si veda la* (12.5.2)
		Intervallo di confidenza per la varianza		
Normale		μ, σ^2	qualsiasi	$\left[\dfrac{(n-1)s^2}{\chi^2_{\alpha/2,(n-1)}} ; \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2,(n-1)}} \right]$