R - Esercitazione 5

Lorenzo Di Biagio
dibiagio@mat.uniroma3.it

Università Roma Tre

Lunedì 2 Dicembre 2013

Intervalli di confidenza (1)

Sia $X_1, \ldots, X_n$ un campione casuale estratto da un densità $f(x, \theta)$ nota a meno di θ.

Ricordiamo: se L_{INF} e L_{SUP} sono due statistiche che soddisfano $L_{INF} \leq L_{SUP}$ e per le quali $P_\theta(L_{INF} < \theta < L_{SUP}) = \gamma = 1 - \alpha$, dove γ non dipende da θ, allora l'intervallo casuale (L_{INF}, L_{SUP}) si chiama intervallo di confidenza al 100γ percento per θ. γ si dice *livello di confidenza*.

Si "confida" che l'intervallo includa il parametro incognito al 100γ percento, i.e., *prima* dell'estrazione del campione è γ la probabilità che l'intervallo che si sta per costruire contiene il vero valore di θ. In altre parole: prendendo molti campioni, circa il 100γ percento degli intervalli (ad essi relativi) conterrà il vero valore di θ.

Intervalli di confidenza (2)

All'aumentare del livello di confidenza aumenta l'ampiezza dell'intervallo (e ne diminuisce quindi la capacità informativa).

In generale: diminuendo la varianza di L_{INF} e L_{SUP} (a parità di livello di confidenza) si otterranno intervalli più ristretti.

Intervalli di confidenza per la media di una v.c. normale con varianza incognita (1)

Sia $X_1, \ldots, X_n$ un campione casuale estratto da una densità normale di varianza σ^2 e media μ, incognite.

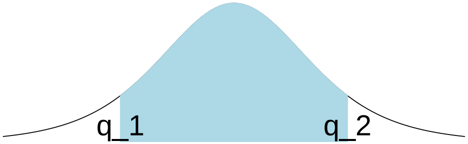
Sappiamo che

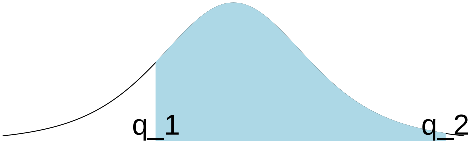
$$T := \frac{\overline{X}_n - \mu}{S/\sqrt{n}}$$

ha una distribuzione t con $n-1$ gradi di libertà (S^2 è la varianza campionaria); dipende solo da μ e non da σ.

Siano q_1, q_2 tali che $P(q_1 < T < q_2) = \gamma = 1 - \alpha$.

Livello di confidenza: $\gamma = 1 - \alpha$





Intervalli di confidenza per la media di una v.c. normale con varianza incognita (2)

Dato che la densità di probabilità di una t di Student è simmetrica rispetto all'asse y, sceglieremo:

- $q_1 = t_{\frac{\alpha}{2}} =$ `qt(0.5*alfa,df=n-1)`
- $q_2 = t_{1-\frac{\alpha}{2}} =$ `qt(1-0.5*alfa,df=n-1)`

Ovviamente $q_1 = -q_2$.

L'intervallo di confidenza della media è quindi:

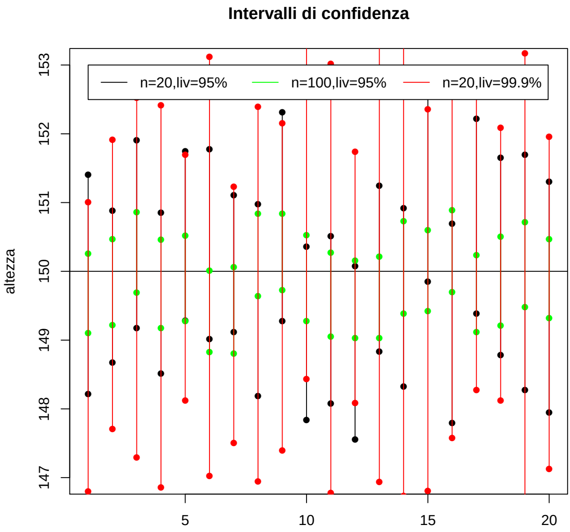
$$\left(\overline{X}_n - t_{1-\frac{\alpha}{2}} \cdot S/\sqrt{n}, \overline{X}_n + t_{1-\frac{\alpha}{2}} \cdot S/\sqrt{n}\right)$$

Intervalli di confidenza per la media di una v.c. normale con varianza incognita (3)

Esercizio 1

1. Generare `m` campioni casuali di ampiezza `n` estratti da una normale di media `media` e varianza `varianza`. Per ognuno di essi calcolare l'intervallo di confidenza per la media (con livello di confidenza = `prob`), assumendo la varianza incognita, e analizzarli graficamente. Ripetere l'esperimento:
   - aumentando l'ampiezza del campione;
   - aumentando il livello di confidenza.

Ad esempio: si può porre `m`=20, `n`=20, `media`=150, `varianza`=9, `prob`=0.95; aumentare poi `n` a 100 e `prob` a 0.999.



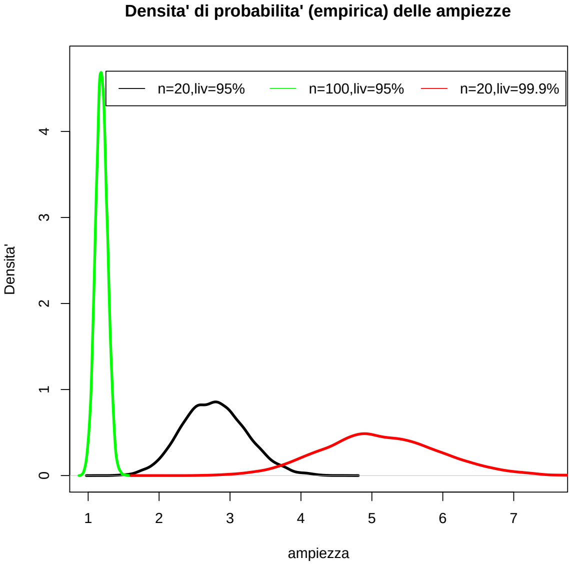
Intervalli di confidenza
n=20,liv=95%
n=100,liv=95%
n=20,liv=99.9%
altezza
153
152
151
150
149
148
147
5
10
15
20

Intervalli di confidenza per la media di una v.c. normale con varianza incognita (4)

Esercizio 2

1. Generare `m` campioni casuali di ampiezza `n` estratti da una normale di media `media` e varianza `varianza`. Per ognuno di essi calcolare l'intervallo di confidenza per la media (con livello di confidenza = `prob`), assumendo la varianza incognita. Determinare la percentuale degli intervalli che contengono il vero valore della media. Determinare le ampiezze degli intervalli e analizzarne graficamente la distribuzione. Ripetere l'esperimento:
    - aumentando l'ampiezza del campione;
    - aumentando il livello di confidenza.

Ad esempio: porre `m`=10000 e `n`, `media`, `varianza`, `prob` come nell'Esercizio 1.



Densita' di probabilita' (empirica) delle ampiezze
n=20,liv=95%
n=100,liv=95%
n=20,liv=99.9%
Densita'
ampiezza
0
1
2
3
4
1
2
3
4
5
6
7

Intervalli di confidenza per la media di una v.c. normale con varianza nota

Sia $X_1, \ldots, X_n$ un campione casuale estratto da una densità normale di varianza σ^2 nota e media μ incognita.
Sappiamo che

$$Z := \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

ha una distribuzione normale standard.
$P(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}) = 1 - \alpha$, da cui l'intervallo di confidenza per la media μ è
$\left(\overline{X}_n - z_{1-\frac{\alpha}{2}} \cdot \sigma/\sqrt{n}, \overline{X}_n + z_{1-\frac{\alpha}{2}} \cdot \sigma/\sqrt{n}\right)$.

Con R: l'intervallo di confidenza si calcola con la funzione `z.test` del pacchetto `BSDA` indicando la varianza della popolazione con il parametro `sigma.x` e il livello di confidenza con il parametro `conf.level`

Esercizio 3

Aprire il dataset "statura.csv". Supporre che le stature, sia per gli uomini che per le donne, ubbidiscano ad una legge normale.

Calcolare un intervallo di confidenza per la media delle stature degli uomini e delle donne (separatamente). Supponendo che la varianza per la statura degli uomini sia 121 e per le donne 49, determinare un intervallo di confidenza per la media delle stature degli uomini e delle donne.