

Analisi Statistica dei Dati: foglio di lavoro sul metodo bootstrap

Luca La Rocca

30 ottobre 2018

1. Leggete i dati presentati nella sezione 1.1.8 del testo di riferimento dal file `dataSurvival.csv` e selezionate, usando la funzione `subset`, le sole osservazioni non censurate, salvandole nel vettore `exactTimes`.
2. Rappresentate graficamente, avvalendovi della funzione `ecdf`, la funzione di ripartizione empirica F_n^\bullet determinata dalle osservazioni $x_1^\bullet, x_2^\bullet, \dots, x_n^\bullet$ in `exactTimes` ($n = 47$), quindi calcolate la corrispondente media empirica \bar{x}_n^\bullet e visualizzatela nella console di R come stima di $\mu = \mathbb{E}(X)$, cioè come stima della sopravvivenza media per la popolazione X da cui supponiamo provengano $x_1^\bullet, x_2^\bullet, \dots, x_n^\bullet$.
3. Create una matrice `bootData` di dimensioni $n \times B$ (es. $B = 10000$) le cui colonne contengano ciascuna un campione casuale di osservazioni generato da F_n^\bullet mediante la funzione `sample`.
4. Calcolate la media di ciascuna colonna della matrice `bootData` e salvate i risultati nel vettore `bootMeans`.
5. Rappresentate graficamente, applicando la funzione `hist` a `bootMeans`, la distribuzione bootstrap della media empirica \bar{x}_n .
6. Calcolate mediante la distribuzione bootstrap e visualizzate nella console di R una stima della distorsione e un errore standard per la media empirica \bar{X}_n come stimatore per μ .
7. Calcolate, usando la funzione `quantile`, un intervallo di stima bootstrap per μ al livello $\gamma = 0.99$.
8. Calcolate un intervallo di stima bootstrap per il coefficiente di variazione $\xi = \sigma/\mu$, dove $\sigma = \text{Sd}(X)$, quindi verificate se l'intervallo trovato contiene il valore $\xi_0 = 1$ (corrispondente all'ipotesi di popolazione esponenziale).

Testo di riferimento

L. Held & D. Sabanés Bové. *Applied Statistical Inference*. Springer, 2014.