Statistica Computazionale
Esempio di compito (2)

È richiesta una relazione contenente il codice R*, debitamente commentato, necessario per risolvere i problemi posti. Le procedure usate devono essere giustificate a parole. Anche i grafici, se presenti, utilizzati nella discussione devono essere inseriti nell'elaborato.*

Dove richiesto, devono essere usate 10000 replicazioni (Monte Carlo o bootstrap).

Non è possibile consultare alcun materiale, tranne l'"help" di R*.*

Il tempo a disposizione è di 135 minuti.

Esercizio 1

Si consideri la distribuzione EE, caratterizzata dalla funzione di densità

$$f(x) = \alpha \, \lambda \, (1 - e^{-\lambda x})^{\alpha-1} \, e^{-\lambda x} \qquad x > 0 \, ,$$

dove $\alpha > 0$ e $\lambda > 0$.

a. Si scrivano due funzioni in R che servono per calcolare la funzione di ripartizione e la funzione quantile per la distribuzione EE. [Ossia funzioni di tipo "p" e tipo "q", rispettivamente]

b. Si scriva una funzione in R che permette, tramite generazione per inversione, di ottenere valori generati dalla distribuzione EE. [Ossia una funzione di tipo "r"]

c. Si verifichi graficamente che la funzione di tipo "r" costruita al punto precedente, quando $\alpha = 2$ e $\lambda = 3$. [È utile il comando `hist(x, prob=T, add=T, nclass=100)`, dove `x` contiene i valori generati]

d. Si usi la distribuzione EE, con $\alpha = 2$ e $\lambda = 3$ per proporre, via campionamento per importanza, un'approssimazione della media di una variabile casuale $\cos(X)$, se X ha distribuzione Gamma$(2, 3)$. Si fornisca anche un intervallo di confidenza per l'approssimazione. [Si ricordi che dgamma$(x, 2, 3)$ restituisce il valore in x della densità di una distribuzione Gamma$(2, 3)$]

Esercizio 2

Il file `esercizio_2.dat` contiene osservazioni relative a due variabili X e Y, tra le quali si assume che sia valida la relazione

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon \ ,$$

con $E(\epsilon) = 0$.

a. Si scriva il codice R necessario per ottenere delle replicazioni bootstrap (bootstrap semiparametrico) degli stimatori dei minimi quadrati di β_0, β_1 e β_2. [Si ricordi che il comando R per ottenere gli stimatori dei minimi quadrati per il modello descritto è `lm(y ~ x + I(x^2))`]

b. Si valutino, usando il bootstrap semiparametrico, la distorsione e la deviazione standard degli stimatori dei minimi quadrati di β_0, β_1 e β_2.

c. Si ottenga, sempre usando il bootstrap semiparametrico, un intervallo di confidenza con grado di fiducia 0.95 per $E(Y|X = 0)$ e lo si usi per verificare l'ipotesi $H_0\colon\ E(Y|X = 0) = 0$, rispetto all'alternativa bidirezionale.

Esercizio 3

Il file `esercizio_3.dat` contiene un campione $x_1, \ldots, x_n$ che si ritiene generato da variabili casuali i.i.d. con distribuzione normale troncata, avente funzione di densità

$$f(x) = \frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \cdot \frac{1}{\sigma} \qquad a \leq x \leq b \ ,$$

dove $\phi(\cdot)$ e $\Phi(\cdot)$ sono rispettivamente la funzione di densità e la funzione di ripartizione di una distribuzione $\mathcal{N}(0, 1)$, mentre $\mu \in \mathcal{R}$ e $\sigma > 0$ sono i parametri di interesse. I valori di a e b sono noti e pari a -2 e 4, rispettivamente.

a. Si scriva una funzione R che calcola la funzione di log-verosimiglianza in corrispondenza di due generici valori di μ e σ.

b. Si disegni un grafico della funzione di log-verosimiglianza e se ne commenti l'andamento: quali valori suggerisce il grafico per la stima di massima verosimiglianza? [È utile la funzione contour]

c. Si ottenga la stima di massima verosimiglianza di (μ, σ) impiegando la funzione nlm e adottando una opportuna riparametrizzazione. Per la scelta dei valori di partenza, ci si aiuti con la rappresentazione grafica ottenuta al punto b.

d. Si ottenga un intervallo di confidenza con grado di fiducia 0.95 per σ utilizzando la distribuzione asintotica dello stimatore di massima verosimiglianza. [Può essere utile la funzione fdHess nella biblioteca nlme]