Inferenza statistica – Esonero 1, Parte B1 – 6 aprile 2024

[SEFA] [SES] [SG] [altro] + **Cognome, nome e n. di matricola:** ______________________

[**SI CORREGGE SOLO QUANTO RIPORTATO IN QUESTI FOGLI**] La prova personalizzata per gli <u>studenti che ne hanno diritto</u> prevede l'esclusione dello svolgimento dei punti 5 e 7

Problema.
Sia $\mathbf{X}_n = (X_1, \ldots X_n)$ un campione i.i.d. da una popolazione X con funzione di densità

$$f_X(x;\theta) = \theta e^{-\theta} x^{-(\theta+1)}, \qquad x > \frac{1}{e}, \quad \theta > 0$$

1. Verificare se il modello per la singola osservazione X_i appartiene alla classe delle famiglie esponenziali.
   Risp.

   $f_X(x;\theta) = \frac{1}{x}\exp\{-\theta \ln x - (\theta - \ln\theta)\}$. Si verifica l'appartenenza alla famiglia esponenziale ponendo $h(x) = \frac{1}{x}$, $T(x) = \ln x$, $\eta(\theta) = -\theta$, $B(\theta) = \theta - \ln\theta$.

2. Determinare il modello statistico per l'n-pla campionaria $\mathbf{X}_n = (X_1, \ldots X_n)$.
   Risp.

   $\left\{\mathcal{X}^n = \left(\frac{1}{e}, +\infty\right)^n, f_n(\mathbf{x}_n;\theta) = \theta^n e^{-n\theta} T_n(\mathbf{x}_n)^{-(\theta+1)}, \Theta = (1, +\infty)\right\}$, con $T_n(\mathbf{x}_n) = \prod_{i=1}^n x_i$. NB: posso trovare la statistica sufficiente anche dal punto precedente, ottenendo $\sum_{i=1}^n \ln x_i = \sum_{i=1}^n T(x_i)$ (funzione monotona di $\prod_{i=1}^n x_i$).

3. Individuare una statistica sufficiente e minimale per il parametro θ fornendo l'opportuna argomentazione.
   Risp.

   Per il criterio di fattorizzazione, $T_n(\mathbf{x}_n) = \prod_{i=1}^n x_i$ è statistica sufficiente minimale. Lo è anche $\sum_{i=1}^n \ln x_i = \sum_{i=1}^n T(x_i)$ in quanto funzione monotona di $\prod_{i=1}^n x_i$.

4. Determinare la stima di massima verosimiglianza $\widehat{\theta}_{mv}(\mathbf{x}_n)$ in corrispondenza del seguente campione osservato
   $$(x_1 = 0.9, x_2 = 1.5, x_3 = 1.0)$$
   Risp.

   $L(\theta) \propto \theta^n e^{-n\theta} T_n^{-\theta}$ e quindi $\ell(\theta) = n\ln\theta - n\theta - \theta \ln T_n$. Pertanto $\ell'(\theta) = 0 \iff \frac{n}{\theta} - n - \ln T_n = 0$ da cui si ottiene $\widehat{\theta}_{mv} = \frac{n}{n + \ln T_n}$. Inoltre $\ell''(\theta) = -\frac{n}{\theta^2}$, che è minore di zero per qualsiasi valore di θ e quindi $\widehat{\theta}_{mv}$ è effettivamente un massimo per $\ell(\theta)$ e $L(\theta)$.

5. Verificare che l'informazione osservata di Fisher corrisponde alla seguente espressione:

$$I_{oss} = \frac{n}{\widehat{\theta}_{mv}(\mathbf{x}_n)^2}$$

Risp.

Dall'espressione di $\ell''(\theta)$ ottenuta nel punto precedente si ricava facilmente che $I_n^{oss} = \frac{n}{\widehat{\theta}_{mv}^2}$.

6. Determinare l'espressione dell'insieme approssimato di verosimiglianza di livello $q = 0.5$ per θ (basato su approssimazione normale di $\overline{L}(\theta; \mathbf{x}_n)$) e calcolare numericamente l'intervallo corrispondente ad un campione di 100 osservazioni dal quale è risultato $\widehat{\theta}_{mv}(\mathbf{x}_n) = 1.43$. Sostituendo i valori numerici si ottiene $\widehat{\theta}_{mv} = 0.91$.

Risp.

$\tilde{L}(\theta) = \exp\{-\frac{1}{2}(\theta - \widehat{\theta}_{mv})^2 I_n^{oss}\}$ e $\tilde{L}_q = \widehat{\theta}_{mv} \pm k_q [I_n^{oss}]^{-1/2}$ con $k_q = \sqrt{-2\ln q}$, e con $\widehat{\theta}_{mv}$ e I_n^{oss} ottenuti nei punti che precedono. Con i valori campionari proposti si ha che $\tilde{L}_q = 1.43 \pm k_q \frac{1.43}{10}$.

7. Determinare l'espressione del valore atteso della variabile $X \sim f(x;\theta)$, mostrare che $\psi = E[X]$ è una funzione biunivoca di θ e calcolare la stima di massima verosimiglianza $\widehat{\psi}_{mv}(\mathbf{x}_n)$ in corrispondenza del campione di ampiezza $n = 100$ di cui al punto precedente.

Risp.

$\mathbb{E}(X) = \theta e^{-\theta} \int_{1/e}^{\infty} x \, \frac{1}{x^{\theta+1}} = \theta e^{-\theta} \int_{1/e}^{\infty} \frac{1}{x^{\theta}} = \theta e^{-\theta} \frac{x^{-\theta+1}}{-\theta+1}\Big|_{1/e}^{\infty} = \ldots = \frac{\theta e^{-1}}{\theta - 1}$.