Inferenza statistica – Esonero 1 – 23 aprile 2025

Cognome, nome e n. di matricola: ______________________________

- Si corregge solo quanto riportato in questi fogli
- Gli studenti che hanno diritto alla riduzione del compito possono saltare i quesiti * e *

Parte A – Quesiti

1. Sia $Z \sim \text{Exp}(1)$, con funzione di densità $f(z) = e^{-z}\mathbb{I}_{[0,+\infty)}(z)$. Determinare l'espressione della funzione di densità $g_{ps}(x;\mu,\sigma)$ per il generico membro della corrispondente famiglia posizione-scala, con parametri $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$.

Risp. Seguendo la definizione (vedi dispense) e tenendo conto del supporto limitato inferiormente, si ha che

$$g_{ps}(x;\mu,\sigma) = \frac{1}{\sigma}e^{-\frac{x-\mu}{\sigma}}\mathbb{I}_{[\mu,\infty)}(x).$$

2. Con riferimento al precedente quesito, sia $X = \sigma Z + \mu$ la v.a. con densità $g_{ps}(x;\mu,\sigma)$. Determinare gli stimatori dei momenti di (μ, σ^2) basati su $X_1, \ldots, X_n$ un campione casuale (iid) con le X_i aventi la stessa distribuzione di X.

Risp.
Si osservi che $\mathbb{E}[Z] = \mathbb{V}[Z] = 1$ (si tratta di v.a. esponenziale di parametro pari a 1). Pertanto $\mathbb{E}[X] = \sigma + \mu$ e $\mathbb{V}[X] = \sigma^2$. Il sistema dei momenti è quindi

$$\begin{cases} \overline{X}_n = \mu + \sigma \\ \widehat{\sigma}^2_n = \sigma^2 \end{cases}$$

da cui si ottiene lo stimatore dei momenti

$$\begin{cases} \widehat{\mu}_m = \overline{X}_n - \widehat{\sigma}_n \\ \widehat{\sigma}^2_m = \widehat{\sigma}^2_n \end{cases}$$

dove $\widehat{\sigma}^2_n$ indica la varianza campionaria.

3. Sia $X_1, \ldots, X_n$ un campione casuale (iid) con $X_i \sim \text{Exp}(\theta)$, , $i = 1, \ldots, n$. Determinare $\widehat{\theta}_{mv}(\mathbf{x}_n)$ e $I_n^{oss}(\mathbf{x}_n)$ (stima di massima verosimiglianza di θ e informazione osservata di Fisher) per un campione osservato $\mathbf{x}_n$.

Risp.
Ricordando che $f_X(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, \quad x \geq 0,\ \theta > 0$ si trova che $L(\theta) = \theta^{-n}e^{-\frac{T_n}{n}}$, dove $T_n = \sum_{i=1}^n x_i$. Con calcoli standard (vedi dispensa) si trova che $\widehat{\theta}_{mv} = \overline{x}_n$ e $I_n^{oss} = \frac{n}{\overline{x}_n^2} = \frac{n^3}{(T_n^2)^2}$.

4. Con riferimento al precedente quesito, determinare l'insieme di verosimiglianza approssimato per θ.

Risp.
$\tilde{L}_q = \widehat{\theta}_{mv} \pm k_q [I_n^{oss}]^{-1/2} = \overline{x}_n \pm k_q \frac{\overline{x}_n}{\sqrt{n}}$, con $k_q = \sqrt{-2 \ln q}$ e $q \in [0,1]$.

5. Sfruttando l'opportuno integrale notevole, calcolare il valore di $\int_0^{+\infty} x^2 e^{-2x} dx$.

Risp.
Integrale gamma: per $a, b > 0$

$$\int_0^{\infty} x^{a-1} e^{-bx} dx = \frac{\Gamma(a)}{b^a}.$$

Ponendo $a = 3$ e $b = 2$ si ottiene che l'integrale in esame risulta uguale a $\frac{\Gamma(3)}{2^3} = \frac{1}{4}$.

6. Sia $X_1, \ldots, X_n$ un campione casuale (iid) con $X_i \sim \mathrm{N}(0, \theta)$, $i = 1, \ldots, n$. Determinare la distribuzione campionaria della statistica $\sum_{i=1}^n X_i^2$.

Risp. Ricordando le proprietà delle v.a. normali e chi quadrato (iid, in particolare l'additività) si ha che: $X_i \sim \mathrm{N}(0,\theta) \Longrightarrow \frac{X_i - 0}{\sqrt{\theta}} \sim \mathrm{N}(0,1) \Longrightarrow \frac{X_i^2}{\theta} \sim \chi_1^2 \Longrightarrow \sum_{i=1}^n \frac{X_i^2}{\theta} \sim \chi_n^2 = \mathrm{Ga}^R(\frac{n}{2}, \frac{1}{2}) \Longrightarrow \sum_{i=1}^n X_i^2 = \theta \cdot \sum_{i=1}^n \frac{X_i^2}{\theta} \sim \mathrm{Ga}^R(\frac{n}{2}, \frac{1}{2\theta})$. Ci si arriva anche ricordando che $\frac{nS_0^2}{\theta} \sim \chi_n^2$, dove $S_0^2 = \sum_{i=1}^n \frac{(X_i - \mu_0)^2}{n}$, qui con $\mu_0 = 0$.

7. Sia $X_1, \ldots, X_n$ un campione casuale (iid) da una popolazione con distribuzione Unif$[0, \frac{\theta}{2}]$, $\theta > 0$. Determinare: (a) $L(\theta; \mathbf{x}_n)$, funzione di verosimiglianza di θ; (b) una statistica sufficiente minimale; (c) $\widehat{\theta}_{mv}$, stima di massima verosimiglianza di θ.

Risp.
(a) Con calcoli standard si mostra che $L(\theta) = \left(\frac{2}{\theta}\right)^n \mathbb{I}_{[2x_{(n)}, \infty)}(\theta)$, dove $x_{(n)}$ indica la statistica massimo campionario.
(b) Per il criterio di fattorizzazione $x_{(n)}$ è la statistica sufficiente minimale per il modello.
(c) Dal grafico di $L(\theta)$ o da considerazioni analitiche elementari si evince che $L(\theta)$ ha un solo punto di massimo in $\widehat{\theta}_{mv} = 2x_{(n)}$.

8. Con riferimento al precedente esercizio, determinare l'espressione di $\overline{L}(\theta; \mathbf{x}_n)$, funzione di verosimiglianza relativa di θ e l'espressione di $L_q(\mathbf{x}_n)$, insieme di verosimiglianza di livello $q \in (0,1)$ per θ. Fornire risultati analitici e rappresentazione grafica.

Risp.
Si verifica facilmente che

$$\overline{L}(\theta) = \frac{L(\theta)}{L(\widehat{\theta}_{mv})} = \ldots = \left[\frac{2x_{(n)}}{\theta}\right]^n \mathbb{I}_{[2x_{(n)},\infty)}(\theta).$$

L'insieme L_q è quindi dato dai valori di θ tali che

$$\begin{cases} \theta \geq 2x_{(n)} \\ \left[\frac{2x_{(n)}}{\theta}\right]^n \geq q \end{cases} \Longleftrightarrow \theta \in [2x_{(n)}, q^{-1/n} 2x_{(n)}].$$

La rappresentazione grafica si omette per brevità.

9. Sia X una v.a. con funzione di densità $f_X(x;\theta) = \theta x^{\theta-1}\ \mathbb{I}_{[0,1]}(x)$, $\theta > 0$. Determinare $\mu_2(\theta) = \mathbb{E}_\theta[X^2]$ e il valore di θ per il quale $\mu_2(\theta) = \frac{1}{2}$.

Risp.
$\mathbb{E}[X^2] = \int_0^1 x^2 \cdot \theta x^{\theta-1} dx = \theta \int_0^1 x^{\theta+1} dx = \frac{\theta}{\theta+2}$. Tale quantità è pari a 1 $\Longleftrightarrow \theta = 2$.

10. Con riferimento al precedente esercizio, supponendo di avere osservato un campione di dimensione $n = 1$ con $x = \frac{1}{2}$, stabilire quale tra i valori $\theta_a = 2$ o $\theta_b = 1$ è più verosimile.

Risp.
$\lambda_{ab}(x) = \frac{\theta_a x^{\theta_a - 1}}{\theta_b x^{\theta_b - 1}} = \frac{2 \cdot \frac{1}{2}}{1} = 1$. I due valori di θ considerati sono quindi verosimili allo stesso livello.

Parte B – Problema. Sia $\mathbf{X}_n = (X_1, \dots X_n)$ un campione casuale (iid) da una popolazione X con funzione di densità $f_X(x;\theta) = \frac{\theta^3}{2x^4}\, e^{-\frac{\theta}{x}}\, \mathbb{I}_{\mathbb{R}^+}(x),\ \theta > 0$. Per la v.a. X sappiamo che $\mathbb{E}_\theta[X] = \frac{\theta}{2}$ e $\mathbb{V}_\theta[X] = \frac{\theta^2}{4}$.

•••

1. Determinare l'espressione dello stimatore dei momenti $\widehat{\theta}_m(\mathbf{X}_n)$ del parametro θ e calcolarne valore atteso, varianza e MSE. Discutere quindi distorsione e consistenza (forte e debole) dello stimatore.

Risp.
$\mathbb{E}_\theta[X] = \frac{\theta}{2} \Longrightarrow \widehat{\theta}_m = 2\overline{X}_n$.
$\mathbb{E}_\theta[\overline{X}_n] = \theta\ \forall\theta \Longrightarrow \widehat{\theta}_m$ stimatore non distorto di θ.
$\mathbb{V}_\theta(2\overline{X}_n) = \frac{\theta^2}{n}$.
$\mathrm{MSE}(\theta, \widehat{\theta}_m) = \frac{\theta^2}{n} \longrightarrow 0$ per $n \longrightarrow \infty\ \forall\theta$ e quindi $\widehat{\theta}_m$ è stimatore consistente (in senso forte e debole) di θ.

2. Determinare la distribuzione asintotica di $\widehat{\theta}_m(\mathbf{X}_n)$ basata sul teorema del limite centrale e la probabilità che lo stimatore dei momenti assuma valori inferiori a $\delta > 0$ [in termini della f.ne di ripartizione della v.a. $\mathrm{N}(0,1)$].

Risp.
Sono soddisfatte le ipotesi del TLC e quindi
$\overline{X}_n \dot{\sim} \mathrm{N}\left(\frac{\theta}{2}, \frac{\theta^2}{4n}\right)$
$\Longrightarrow$
$\widehat{\theta}_m = 2\overline{X}_n \dot{\sim} \mathrm{N}\left(\theta, \frac{\theta^2}{n}\right)$
$\Longrightarrow$
$\mathbb{P}_\theta[2\overline{X}_n < \delta] \approx \Phi\left(\frac{\sqrt{n}(\delta-\theta)}{\theta}\right)$.

3. Verificare che il modello per la singola osservazione X_i appartiene alla classe delle famiglie esponenziali.

Risp.
$f_X(x;\theta) = h(x)\exp\{\eta(\theta)T(x) - B(\theta)\}$ con
$h(x) = \frac{1}{2x^4}$,
$\eta(\theta) = -\theta$,
$T(x) = \frac{1}{x}$,
$B(\theta) = -3\ln\theta$.

4. Verificare che $T_n(\mathbf{X}_n) = \sum_{i=1}^n \frac{1}{X_i}$ è statistica sufficiente minimale per modello.

Risp.
$T(x) = \frac{1}{x} \Longrightarrow T_n(\mathbf{x}_n) = \sum_{i=1}^n \frac{1}{x_i}$
statistica sufficiente minimale (per il criterio di fattorizzazione) e completa (in quanto famiglia esponenziale).

5. Determinare $\widehat{\theta}_{mv}(\mathbf{X}_n)$, stimatore di massima verosimiglianza di θ.

Risp.
$L(\theta) = c \cdot \theta^{3n} e^{-\theta t_n}$ con $t_n = \sum_{i=1}^n \frac{1}{x_i}$.
$\ell'(\theta) = 0 \Longleftrightarrow \theta = \frac{3n}{t_n}$.
$\ell''(\theta) = -\frac{3n}{\theta^2} < 0 \ \forall \theta$ e quindi $\widehat{\theta}_{mv} = \frac{3n}{t_n}$.

6. Determinare l'espressione di $I_n(\theta)$, informazione attesa di Fisher, e quella di $\mathrm{cr}(\theta)$, limite inferiore di Cramer-Rao per gli stimatori non distorti di θ.

Risp.
$\ell''(\theta) = -\frac{3n}{\theta^2}$
$\Longrightarrow$
$I_n(\theta) = \frac{3n}{\theta^2}$
$\Longrightarrow$
$\mathtt{cr}(\theta) = \frac{\theta^2}{3n}$.
Osservare che, in questo caso, per il calcolo dell'informazione attesa:
- possiamo usare l'espressione di $\ell''(\theta)$ determinata nel punto precedente, ovvero la formula di $I_n(\theta)$ basata su $f_n(\mathbf{X}_n; \theta)$;
- non serve effettuare il valore atteso in quanto in tale espressione non compaiono variabili aleatorie;
- si può anche usare la formula $-nI_1(\theta)$, ovvero la formula basata su $f_X(X; \theta)$, e si ottiene lo stesso risultato.

7. Trovare la funzione di densità di $Y = \frac{1}{X}$ e verificare che si tratta della densità di una v.a. $\mathrm{Ga}(3, \mathtt{rate} = \theta)$[1].

Risp.
Applicando la formula in nota si trova
$f_Y(y; \theta) = \frac{\theta^3 y^4}{2} e^{-\theta y} \cdot \frac{1}{y^2} = \frac{\theta^3}{2} y^2 e^{-\theta y}, \quad y \geq 0$
che è la funzione di densità di una v.a. $\mathrm{Ga}(3, \mathtt{rate} = \theta)$.

[1] Ricorda che se X è una v.a. assolutamente continua con funzione di densità $f_X(\cdot)$ e se $Y = g(X)$ con g invertibile, allora $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$, dove $g^{-1}(\cdot)$ indica la funzione inversa di $g(\cdot)$.

8. Sfruttando il risultato precedente e le proprietà delle v.a. Gamma, trovare la distribuzione della statistica $T_n = \sum_{i=1}^n \frac{1}{X_i}$ e quella di $\frac{1}{T_n}$.

Risp.
Per le proprietà delle v.a. gamma,
$Y_i = \frac{1}{X_i} \sim \text{Ga}(3, \texttt{rate} = \theta)$
$\Longrightarrow$
$T_n = \sum_{i=1}^n \frac{1}{X_i} \sim \text{Ga}(3n, \texttt{rate} = \theta)$
$\Longrightarrow$ (per la relazione tra v.a. gamma e v.a. gamma inversa)
$\frac{1}{T_n} \sim \text{InvGa}(3n, \texttt{rate} = \theta)$.

9. Sfruttando il risultato precedente, dimostrare e argomentare attraverso la procedura Rao-Blackwell e Lehmann-Scheffe che lo stimatore $\widehat{\theta}_u^\star(\mathbf{X}_n) = \frac{3n-1}{T_n} = \frac{3n-1}{\sum_{i=1}^n \frac{1}{X_i}}$ è l'UMVUE di θ.

Risp.
Per le proprietà della v.a. gamma inversa (vedi anche nota sotto)
$\mathbb{E}_\theta\left[\widehat{\theta}_u^\star\right] = (3n-1)\mathbb{E}_\theta\left[\frac{1}{T_n}\right] = (3n-1)\frac{\theta}{3n-1} = \theta \; \forall\theta.$
Pertanto $\widehat{\theta}_u^\star$ è:
- non distorto per θ;
- funzione di T_n, statistica sufficiente e completa per il modello.
Per i teoremi di RB-LS si tratta quindi di UMVUE per θ.

10. Verificare che[2], in questo caso, $\mathbb{V}_\theta[\widehat{\theta}_u^\star(\mathbf{X}_n)] > \text{cr}(\theta)$.

(Opzionale). Determinare l'espressione della funzione score $S_n(\mathbf{x}_n;\theta)$ e giustificare il risultato del punto 10.

Risp.
Per quanto riportato in nota e quanto trovato al punto 6, si ha:
$\mathbb{V}_\theta(\widehat{\theta}_u^\star) = (3n-1)^2 \cdot \frac{\theta^2}{(3n-1)^2(3n-2)} = \frac{\theta^2}{3n-2} > \frac{\theta^2}{3n} = \texttt{cr}(\theta), \forall\theta.$
- - -
Risposta alla domanda opzionale
- - -
Non si ha raggiungiumento del limite inferiore di CR in quanto $\widehat{\theta}_u^\star$ e la funzione score $S_n(\mathbf{X}_n,\theta)$ non sono in relazione lineare tra loro.
Infatti,
$S_n(\mathbf{X}_n,\theta) = \frac{3n}{\theta} - \frac{3n}{\widehat{\theta}_u^\star}$ (relazione inversa).

[2] Ricorda che se $X \sim \text{InvGa}(a, \texttt{rate} = b)$, allora $\mathbb{E}(X) = \frac{b}{a-1}$ e $\mathbb{V}(X) = \frac{b^2}{(a-1)^2(a-2)}$.