Inferenza statistica – 3 luglio 2023 – Parte A

Cognome, nome e n. di matricola: – SOLUZIONI –

[SI CORREGGE SOLO QUANTO RIPORTATO IN QUESTI FOGLI] Gli studenti che hanno diritto alla riduzione del compito [DSA] possono saltare 3 quesiti e rispondere solo ai rimanenti 7

A – Quesiti [dove necessario giustificare adeguatamente le risposte]

1. Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale da $\mathrm{N}(\mu, \sigma^2)$. Si consideri come parametro di interesse la varianza σ^2. Fornire le ditribuzioni esatte dei due stimatori S_n^2 e $\widehat{\sigma}_n^2$ ed elencarne tutte le proprietà inferenziali note.

   Risp. $S_n^2 \sim Ga^R\left(\frac{n-1}{2}, \frac{n-1}{2\sigma^2}\right)$ $\quad \hat{\sigma}_n^2 \sim Ga^R\left(\frac{n-1}{2}, \frac{n}{2\sigma^2}\right)$

   S_n^2 è: ND per θ, consistente, fun SSAT suff., asint Norm UMVUE di σ^2 MA inammissibile

   $\hat{\sigma}_n^2$ è, distorto per σ^2, consist, fun stat suff, asint Norm e asint effic Ammissibile.

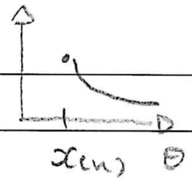
2. Con riferimento al precedente esercizio, argomentare perché lo stimatore S_n^2 è inammissibile.

   Risp. Perché $MSE_\theta[\hat{\sigma}_n^2] < MSE_\theta[S_n^2]$ $\forall \sigma^2 \in \mathbb{R}^+$

   [Vedi Es. 4.69 DISPENSE)

3. Sia $X_1, \ldots, X_n$ un campione casuale da popolazione con funzione di densità $f_X(x;\theta) = \frac{2x}{\theta^2} \cdot \mathbb{I}_{[0,\theta]}(x)$ con parametro $\theta > 0$. Determinare l'espressione della funzione di verosimiglianza e verificare che lo stimatore di massima verosimiglianza di θ corrisponde al massimo campionario $X_{(n)}$.

   Risp. $L(\theta) = c \, \frac{1}{\theta^{2n}} \, \mathbb{I}_{[x_{(n)}, +\infty)}(\theta)$

   $\Rightarrow \hat{\theta}_{MV} = x_{(n)}$

   $F_X(x) = \begin{cases} 0 & x \leq 0 \\ x^2/\theta^2 & 0 \leq x < \theta \\ 1 & x \geq \theta \end{cases}$

4. Con riferimento al precedente esercizio, determinare la distribuzione dello stimatore di massima verosimiglianza, calcolare la sua distorsione e studiare l'asintotica correttezza.

   $Y = X_{(n)}$ **Risp.** $F_Y(y) = [F_X(y)]^n \Rightarrow f_Y(y) = n[F_X(y)]^{n-1} f_X(y)$

   $= n \frac{x^{2n}}{\theta^{2n}} \frac{2x}{\theta^2} \mathbb{I}_{[0,\theta]}(x) = 2n \frac{x^{2n+1}}{\theta^{2n+2}} \mathbb{I}_{[0,\theta)}(x)$

   $\Rightarrow \mathbb{E}(Y) = \frac{2n}{\theta^{2n+2}} \int_0^\theta x^{2n+1} \cdot x \, dx$

   $= \frac{2n}{\theta^{2n+2}} \frac{\theta^{2n+3}}{2n+3} = \frac{2n}{2n+3}\theta \xrightarrow[n \to \infty]{} \theta$

   $x_{(n)}$ è DIST ma asint ND

5. Sia $\mathbf{X}_n = (X_1, \ldots, X_n)$ un campione casuale da un modello di Bernoulli $X_i \sim Ber(\theta)$. Determinare lo stimatore di massima verosimiglianza del parametro $\psi = \log \frac{\theta}{1-\theta}$ e la sua distribuzione asintotica.

   Risp. $\hat{\theta}_{MV} = \bar{X}_n \quad \hat{\psi}_{MV} = \ln \frac{\bar{X}_n}{1-\bar{X}_n}$

   $\hat{\psi} \,\dot\sim\, N\left(\ln \frac{\theta}{1-\theta}, \; g'(\theta)^2 \cdot I_n^{-1}(\theta)\right) = N\left(\ln \frac{\theta}{1-\theta}, \frac{1}{n\theta(1-\theta)}\right)$

   $g'(\theta)^2 = \frac{1}{\theta^2(1-\theta)^2}$ $\quad I_n^{-1}(\theta) \to \frac{\theta(1-\theta)}{n}$

$\hat{\theta}_{MV} = 0.4$

6. Con riferimento al precedente esercizio, si è osservato un campione di $n = 100$ osservazioni con 40 successi. Scrivere la regione di rifiuto per verificare il sistema di ipotesi $H_0 : \psi = 0$ contro $H_1 : \psi \neq 0$ mediante un opportuno test asintotico di dimensione $\alpha = 0.01$ e stabilire la conclusione del test

Risp. $\tilde{W} = \dfrac{\hat{\varphi} - \varphi_0}{\sqrt{v(\hat{\varphi})|_{\psi = \hat{\psi}_{MV}}}} = \dfrac{\ln 0.4/0.6 - 0}{\sqrt{100 \cdot (0.4)(0.6)}} =$

Rif se

$|\tilde{W}| > z_{1-\alpha/2} = z_{0.995} = 2.57 \Rightarrow$

7. Sia $X_1, \ldots, X_n$ un campione casuale da popolazione con la seguente discreta $f_X(x;\theta) = Pr\{X = x;\theta\} = \theta(1-\theta)^x$ con valori $x \in \mathcal{X} = \{0, 1, 2, \ldots\}$ e parametro $\theta \in \Theta = (0, 1)$. Verificare che è possibile costruire un test UMP per testare il sistema $H_0 : \theta \leq 0.5$ contro $H_1 : \theta \geq 0.5$ e scrivere la generica forma della regione di rifiuto.

Risp. $f_X(x;\theta) \in$ a famiglia espon $\Rightarrow$ posso usare teo di KR poiché le ip sono UNILATERALI; Rapp. veros è $\downarrow$ in $T = \sum x_i \Rightarrow R = \{x_n : \sum x_i \leq K\}$

Infatti se $\theta_2 > \theta_1$ $\lambda_{21}(x_n) = \left(\dfrac{\theta_2}{\theta_1}\right)^n \left(\dfrac{1-\theta_2}{1-\theta_1}\right)^{\sum x_i}$ $\downarrow$ con $\sum x_i$

8. Si consideri la seguente densità di probabilità $h(w) = e^{-w}\mathbb{I}_{(0,\infty)}(w)$ e la variabile casuale $W \sim h(w)$. Verificare che la mediana della distribuzione vale $\log 2 = 0.6931472$.

Risp. m = mediana di W è quel valore t.c.:

$\int_{-\infty}^{m} h(w)\,dw = 1/2$. Poiché $\int_{-\infty}^{m} h(w)\,dw = \int_{0}^{m} e^{-w} = -e^{-w}\Big|_0^m = 1 - e^{-m}$

$\Rightarrow m$ deve soddisfare

$1 - e^{-m} = 1/2 \leftrightarrow e^{-m} = 1/2 \leftrightarrow -m = \ln 1/2 \leftrightarrow \boxed{m = \ln 2}$

9. Con riferimento al precedente esercizio, costruire in base ad essa una famiglia parametrica di distribuzioni con parametro di posizione-scala (μ, σ) determinando la generica espressione della funzione di densità della v.c. $X \sim f_X(\cdot; \mu, \sigma)$.

Risp. $f_X(x;\mu,\sigma) = \dfrac{1}{\sigma} h\left(\dfrac{x-\mu}{\sigma}\right) = \dfrac{1}{\sigma} e^{-(x-\mu)/\sigma} \mathbb{I}_{(0,\infty)}\left(\dfrac{x-\mu}{\sigma}\right)$

$= \dfrac{1}{\sigma} e^{-(x-\mu)/\sigma} \mathbb{I}_{(\mu, +\infty)}(x)$

10. Con riferimento al precedente esercizio, determinare la mediana, il valore atteso e la varianza della v.a. X

$\boxed{\text{V.A}}$ **Risp.** $X = \sigma W + \mu \Rightarrow \mathbb{E}(X) = \sigma\,\mathbb{E}(W) + \mu$

$= \sigma * 1 + \mu = \sigma + \mu$

$[W \sim \exp(1) \Rightarrow \mathbb{E}[W] = 1]$

$\boxed{\text{median}}$ (invariante rispetto a trasf lineari)

$\text{Med}(X) = \sigma\,\text{Med}(W) + \mu = \boxed{\sigma \ln 2 + \mu}$

Inferenza statistica – 3 luglio 2023 – Parte B

Cognome, nome e n. di matricola: SOLUZIONI

[SI CORREGGE SOLO QUANTO RIPORTATO IN QUESTI FOGLI]
Gli studenti che hanno diritto alla riduzione del compito [DSA] possono saltare possono saltare 3 quesiti e rispondere solo ai rimanenti 7

B – Problema. Sia $X_1, \ldots, X_n$ un campione casuale da una popolazione X con funzione di densità di probabilità $f_X(x;\theta) = x(\theta+2)x^\theta$, $\quad x \in (0,1), \quad \theta > -2$.

1. Verificare se la il modello $(f_X(x;\theta),\ \theta \in \Theta)$ costituisce una famiglia esponenziale.

Risp.

$$\underbrace{f_X(x;\theta)}_{} = \underbrace{x}_{h(x)} \exp\{ \underbrace{\theta}_{\eta(\theta)} \underbrace{\ln x}_{T(x)} + \underbrace{\ln(\theta+2)}_{-B(\theta)} \} \Rightarrow \text{OK}$$

2. Determinare le espressioni di $L(\theta;\mathbf{x}_n)$ (funzione di verosimiglianza di θ), di $T(\mathbf{x}_n)$ (una statistica sufficiente) e di θ, $\widehat{\theta}_{mv}(\mathbf{x}_n)$ (stima di massima verosimiglianza di θ).

Risp.

$$L(\theta) = c\,(\theta+2)^n \left(\prod_i^n x_i\right)^\theta \Rightarrow T(\underline{x}_n) = \prod_i^n x_i$$

$$\ell(\theta) = \ln c + n \ln(\theta+2) + \theta \ln T \Rightarrow \ell'(\theta) = \frac{n}{\theta+2} + \ln T$$

$\Rightarrow$ (★)

3. Sapendo che $\overline{X}_n$ è uno stimatore non distorto di $\psi = \frac{\theta+2}{\theta+3}$, determinare $\widehat{\theta}_m(\mathbf{X}_n)$, stimatore dei momenti di θ.

Risp. $\mathbb{E}[\overline{X}_n] = \dfrac{\theta+2}{\theta+3} \Rightarrow \overline{X}_n = \dfrac{\theta+2}{\theta+3} \Rightarrow \overline{X}_n\theta + 3\overline{X}_n = \theta + 2$

$\Rightarrow \theta(\overline{X}_n - 1) = 2 - 3\overline{X}_n \Rightarrow \hat{\theta}_m = \dfrac{2 - 3\overline{X}_n}{\overline{X}_n - 1} = \dfrac{3\overline{X}_n - 2}{1 - \overline{X}_n}$

4. Stabilire se $\widehat{\theta}_m(\mathbf{X}_n)$ è lo stimatore UMVUE di θ [motivare/argomentare la risposta].

Risp. $\hat{\theta}_m$ non può essere UMVUE in quanto non è f.ne di $T(\underline{x}_n) = \prod_i^n x_i$

(★) $\ell'(\theta) = 0 \iff n + (\theta+2)\ln T = 0 \iff \theta \ln T = -2\ln T - n$

$$\hat{\theta} = -\frac{2\ln T + n}{\ln T}$$

$= -\,n \quad - 2$

5. Verificare che, asintoticamente, $\widehat{\theta}_{mv}(\mathbf{X}_n) \dot{\sim} \mathrm{N}\left(\theta, \frac{(\theta+2)^2}{n}\right)$.

Risp. $\ell''(\theta) = -\dfrac{n}{(\theta+2)^2} \Rightarrow I_n^{-1}(\theta) = \dfrac{(\theta+2)^2}{n}$

$\Rightarrow \hat{\theta}_{MV} \dot{\sim} N\left(\theta, \dfrac{(\theta+2)^2}{n}\right)$

6. Determinare l'intervallo di confidenza approssimato per θ di livello $1-\alpha$.

Risp. $\tilde{C} = \hat{\theta}_{MV} \pm z_{1-\frac{\alpha}{2}} \dfrac{\hat{\theta}_{MV}+2}{\sqrt{n}}$

7. Si consideri un campione casuale con $n = 20$ e $\ln \prod_{i=1}^n x_i = \sum_{i=1}^n \ln x_i = -19$. Calcolare il valore di $\widehat{\theta}_{mv}(\mathbf{x}_n)$ e degli estremi dell'intervallo di confidenza osservato che si ottengono ponendo $1-\alpha = 0.95$.

Risp. $\hat{\theta}_{MV} = -\dfrac{n}{\ln T} - 2 = \dfrac{20}{19} - 2 \simeq -1$

$\Rightarrow \tilde{C} \simeq -1 \pm 1.96 \; 20/19 / \sqrt{20} = [-1.41, -0.48]$

8. Considerare il sistema di ipotesi semplici $H_0: \theta = \theta_0 = 1$ vs. $H_1: \theta = \theta_1 = 0$. Determinare la regione di accettazione di H_0 per test di Neyman-Pearson in funzione di $\widehat{\theta}_{mv}(\mathbf{x})$ e la regola di rifiuto di H_0 del test asintotico di ampiezza α.

Risp. $\theta_0 > \theta_1 \Rightarrow$ Rif H_0 se $\hat{\theta}_{MV} < k$

$\Rightarrow \dfrac{\hat{\theta}_{MV} - \theta_0}{\sqrt{I_n^{-1}(\theta_0)}} < z_\alpha \iff \hat{\theta}_{MV} < \theta_0 + z_\alpha I_n^{-1}(\theta_0)$

9. Verificare che, per $n = 1$ la regione di accettazione di H_0 del test di Neyman-Pearson è l'insieme $A = \{x \in (0,1) : x \geq k\}$, $\quad k \in \mathbb{R}^+$.

Risp. $\lambda_{01} > k \iff \dfrac{x(\theta_0+2)x^{\theta_0}}{x(\theta_1+2)x^{\theta_1}} = c\, x^{\theta_0-\theta_1}$ ↑ con x

$\Rightarrow \lambda_{01}(x) > k \iff x \geq k$

10. Con riferimento al precedente quesito, calcolare le probabilità di errore di I e II tipo e la potenza del test, assumendo $k = 4/5$.

Risp. $\alpha = \mathbb{P}_{\theta_0}[X < 4/5] = \int_0^{4/5} x(\theta_0+2)x^{\theta_0}\,dx = (*)$

$1-\beta = \mathbb{P}_{\theta_1}[X < 4/5] = \int_0^{4/5} x(\theta_1+2)x^{\theta_1}\,dx =$

$(*)\quad \int_{4/5}^{1} 3x^2\,dx = x^3 \Big|_0^{4/5} = (4/5)^3$

$(**)\quad \int_{4/5}^{1} 2x\,dx = x^2 \Big|_0^{4/5} = \left(\frac{4}{5}\right)^2$

$\Rightarrow \beta = 1 - \left(\frac{4}{5}\right)^2$