Tema di Statistica Matematica

Pre-appello

23 dicembre 2020

1) Enunciare e dimostrare la disuguaglianza di Rao-Cramér riguardante il limite inferiore per la varianza di un qualsiasi stimatore non distorto T_n di una funzione del parametro $\theta \in \Theta$, $\tau(\theta)$.

2) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale da una distribuzione Beta di parametri $\alpha = 1$ e $\beta > 0$. Trovare, se esiste, il test UMP di livello α (vale a dire, individuare la forma della regola di decisione e la regione critica) per il seguente sistema di ipotesi:

$$H_0 : \beta = 1 \ \ vs. \ \ \beta > 1$$

3) Consideriamo un campione casuale $(X_1, X_2, \ldots, X_n)$ di ampiezza n proveniente da una distribuzione di Bernoulli di parametro p. Si stabilisce di rigettare $H_0 : p = \frac{1}{2}$ e accettare $H_1 : p > \frac{1}{2}$ qualora $\sum_{i=1}^{n} X_i \geq c$. Trovare i valori di n e di c che restituiscono una funzione di potenza del test $\eta_{C_\alpha}(p)$ tale per cui, approssimativamente, $\eta_{C_\alpha}(0.5) = 0.10$ e $\eta_{C_\alpha}(0.667) = 0.95$.

4) Tra i dati raccolti nell'ambito di un progetto di monitoraggio della qualità dell'aria dell'Organizzazione Mondiale della Sanità è presente una misura relativa alla concentrazione delle particelle sospese (in $\mu g/m^3$) in aria. Siano X e Y due variabili che misurano rispettivamente la concentrazione di particelle sospese rilevate dalle centraline di rilevamento poste nel centro delle due città di Melbourne e Houston. Utilizzando $n = 13$ osservazioni relative alla variabile X and $m = 16$ osservazioni relative alla variabile Y, vogliamo sottoporre a verifica il seguente sistema di ipotesi:

$$H_0 : \mu_X = \mu_Y \ \ vs. \ \ H_0 : \mu_X < \mu_Y$$

a) Trovare la statistica test e l'associata regione e la regione critica, assumendo che le varianze (sconosciute) siano uguali e $\alpha = 0.05$, giustificando opportunamente i vari passaggi.

b) Avendo osservato $\bar{x}_n = 72.9$, $s_X = 25.6$, $\bar{y}_n = 81.7$ e $s_Y = 28.3$, decidere in merito al sistema di ipotesi in questione.

5) I seguenti dati sono le realizzazioni (arrotondate) di un campione casuale di dimensione $n = 20$ proveniente da una distribuzione Gamma di parametri sconosciuti.

$$131.7, 182.7, 73.3, 10.7, 150.4, 42.3, 22.2, 17.9, 264.0, 154.4,$$
$$4.3, 265.6, 61.9, 10.8, 48.8, 22.5, 8.8, 150.6, 103.0, 85.9$$

Supponiamo di voler fare inferenza sui quartili di questa distribuzione sconosciuta. Sulla base dell'informazione contenuta nella realizzazione campionaria data, scrivere un codice R (commentato) finalizzato a ottenere una stima bootstrap per ciascuno dei suddetti quartili e una misura della sua precisione. Proporre poi una stima della distribuzione campionaria della mediana e calcolare un intervallo di confidenza bootstrap, a livello di confidenza 90%, per la mediana.

6) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale proveniente da una distribuzione cui é associata la seguente funzione di densitá

$$f(x;\theta) = 2\,\theta\,x\,\exp\{-\theta x^2\}\,\mathbb{I}_{\mathbb{R}^+}(x), \;\; \theta > 0$$

a) Verificare se la famiglia di distribuzioni in questione é regolare e, laddove esista, individuare una statistica sufficiente per θ.

b) Trovare lo stimatore di massima verosimiglianza $\hat{\theta}_n$ di θ.

c) Determinare lo stimatore di massima verosimiglianza per $\gamma = 2/\theta$.

d) Calcolare le stime di massima verosimiglianza di θ e di γ ipotizzando di aver estratto dalla popolazione un campione casuale di ampiezza $n = 100$ cosí costituito

x_i	0.01	0.05	0.5	1	2	4	**Totale**
Frequenza	35	25 ù	20	10	5	5	**100**

e) Lo stimatore di massima verosimiglianza di θ é anche stimatore efficiente per θ?

f) Trovare lo stimatore $UMVU$ di θ.

g) Trovare la distribuzione asintotica di $\hat{\theta}_n$ e costruire un intervallo di confidenza di livello 0.95 per θ.

7) Sia $(X_1, X_2, \ldots, X_n, \ldots)$ una successione di variabili casuali i.i.d. aventi ciascuna distribuzione $U(0, 1)$ e sia $M_n = \max(X_1, X_2, \ldots, X_n)$. Dimostrare che

$$n(1 - M_n) \xrightarrow{D} Exp(1)$$

1) Sia X una variabile casusle avente distribuzione continua con densità $f(x)$ purché $0 < x < b < \infty$. Dimostrare che

$$\mathbb{E}(X) = \int_0^b [1 - F(x)]dx$$

con $F(x)$ funzione di distribuzione cumulata di X.

3) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale da una distribuzione di Bernoulli di parametro θ, $\theta \in (0,1)$ e sia $\tau(\theta) = \theta\,(1-\theta)$.

a) Trovare lo stimatore di massima verosimiglianza di $\tau(\theta)$ e verificare se esso è non distorto

b) Trovare la distribuzione asintotica di $\tau(\theta)$ per ogni valore di $\theta \in (0,1)$.

1) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale di ampiezza n proveniente da una distribuzione di Poisson di parametro $\theta > 0$. Trovare il valore atteso condizionato

$$\mathbb{E}\left(X_1 + 2\,X_2 + 3\,X_3 \mid \sum_{i=1}^n X_i \right).$$

2) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale proveniente da una distribuzione Uniforme discreta di parametro θ, la cui funzione di massa é data da

$$\mathbb{P}_\theta(X = x) = \frac{1}{\theta}\,\mathbb{I}_{\{1,2,\ldots,\theta\}}(x)$$

dove θ é un intero positivo non noto.

a) Dimostarare che $Y = \max(X_1, X_2, \ldots, X_n)$ é una statistica sufficiente per θ.

b) Dimostarare che

$$W_n = \frac{Y^{n+1} - (Y-1)^{n+1}}{Y^n - (Y-1)^n}$$

é lo stimatore UMVU per θ e che esso é unico.

3) Siano X_1 e X_2 variabili casuali IID distribuite secondo Poisson di parametro $\theta > 0$.

a) Trovare una statistica sufficiente (e minimale) per θ.

b) Dimostrare che

$$W = \begin{cases} 1 & \text{se } X_1 = 0 \\ 0 & \text{altrimenti} \end{cases}$$

é uno stimatore non distorto per $\tau(\theta) = e^{-\theta}$.

c) Calcolare $\mathbb{E}_\theta[W \mid X_1 + X_2 = y]$.

d) Partendo dallo stimatore non distorto W trovato in b) ricavare lo stimatore UMVU di $\tau(\theta) = e^{-\theta}$

e) Estendere il risultato ottenuto al punto precedente al caso in cui le variabili casuali IID prese in considerazione siano n e non solo due.

4) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale proveniente da una distribuzione Uniforme sull'intervallo $(0, \theta), \theta > 0$.

a) Individuare una statistica sufficiente minimale per θ.

a) Trovare lo stimatore di massima verosimiglianza $\hat{\theta}_n$ e quello del metodo dei momenti $\tilde{\theta}_n$ del parametro θ.

b) Verificare non distorsione e consistenza dei due stimatorio trovati in a).

c) Infine ricavare e confrontare gli errori quadratici medi di $\hat{\theta}_n$ e $\tilde{\theta}_n$ e proporre quello che tra i due ritenerte migliore.

5) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale proveniente da una distribuzione di Bernoulli di parametro $\theta \in (0, 1)$. Trovare il limite inferiore di Rao-Cramér per la varianza

a) dello stimatore plug-in T_n di θ.

b) dello stimatore plug-in della varianza di T_n.

6) La convergenza in distribuzione: definizione, proprietà e principali teoremi a essa collegati.

7) Enunciare e dimostrare il teorema di Rao-Blackwell.

8) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale proveniente da una distribuzione di Poisson di media $\theta > 0$.

a) Trovare una statistica sufficiente e completa per θ.

b) Trovare lo stimatore di massima verosimiglianza di θ^2.

c) Trovare la distribuzione condizionata di X_1 dato $T_n = \sum_{i=1}^n X_i$.

d) Dimostrare che $V_n = X_1^2 - X_1$ è uno stimatore non dstorto per θ^2 e trovare lo stimatore $UMVU$ per θ^2. Coicide con lo stimatore di massima verosimiglianza di θ^2 trovato al punto b)?

e) Trovare lo stimatore $UMVU$ di $P(X_1 = 0) = e^{-\theta}$.

9) Sia $(X_1, X_2, \ldots, X_n)$ un campione casuale estratto da una distribuzione Gamma di parametri (η, θ) con $\eta > 0$ noto mentre $\theta > 0$ é un parametro incognito.

a) Dimostrare che la regione critica del test piú potente di livello α per il sistema di ipotesi

$$H_0 : \theta = \theta_0 \;\; vs. \;\; H_1 : \theta = \theta_1, \;\; \text{con} \;\; \theta_0 < \theta_1$$

ha forma

$$\mathcal{C} = \left\{ (X_1, X_2, \ldots, X_n) \in \mathcal{X} : \; \sum_{i=1}^{n} X_i \geq B \right\}$$

per una qualche costante $B > 0$.

b) Trovare la potenza del test in questione.

c) Il test corrispondente alla regione critica in a) é anche uniformemente piú potente per alternative unilaterali $H_1 : \theta > \theta_0$?