Politecnico di Milano – Facoltà di Ingegneria Informatica

Anno Accademico 2008/2009

Corso di Statistica (2L) per INF e TEL

Docente: Antonio Pievatolo; esercitazioni: Raffaele Argiento

Esercitazione del 9/05/08

Esercizio 1

Tema d'esame del 19/09/05

Abbiamo estratto un campione casuale $X_1, ..., X_n$ dalla densità (di Raleigh) di parametro $\theta > 0$:

$$f(x, \theta) = \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right) \mathbf{1}_{(0,+\infty)} \tag{1}$$

1. Determinate lo stimatore della caratteristica $\kappa(\theta) = \theta^2$ usando il metodo di massima verosimiglianza.
2. Determinate la densità di $Y = X^2$.
3. Chiamiamo $\hat{\kappa}_n$ lo stimatore di massima verosimiglianza di $\kappa(\theta) = \theta^2$: discutete qualche proprietà di $\hat{\kappa}_n$.
4. Determinate la distribuzione della variabile aleatoria $Q_n = \frac{2n\hat{\kappa}_n}{\theta^2}$.
5. Fornite un intervallo di confidenza a due code per $\kappa(\theta)$ di livello $\gamma = 90\%$ per $n = 10$ e $\hat{\kappa}_n = 0.0387$. Quindi deducete un intervallo di confidenza per θ, sempre bilatero e di livello 90%.

Soluzione

1. Sia $\underline{x} = (x_1, \ldots, x_n)$ una realizzazione campionaria, scriviamo la funzione di verosimiglianza basata su $\underline{x}$
$$L(\theta, \underline{x}) = \prod_{i=1}^{n} f(x_i, \theta) = \frac{\prod_{i=1}^{n} x_i}{\theta^{2n}} \exp\left\{-\frac{\sum_{i=1}^{n} x_i^2}{2\theta^2}\right\} \mathbf{1}_{(0,\infty)}(x_{(1)}).$$
Passiamo dunque al logaritmo
$$l(\theta, \underline{x}) = \log(L(\theta, \underline{x})) = -n\log(\theta^2) + \sum_{i=1}^{n} \log(x_i) - \frac{\sum_{i=1}^{n} x_i^2}{2\theta^2}.$$
Differenziando rispetto a θ^2 otteniamo
$$\frac{\partial}{\partial\theta^2} l(\theta, \underline{x}) = -\frac{n}{\theta^2} + \frac{\sum_{i=1}^{n} x_i^2}{2(\theta^2)^2}.$$
Studiando il segno di quest'ultima, rispetto a θ^2 otteniamo
$$\hat{k}_n = \frac{\sum_{i=1}^{n} x_i^2}{2n}.$$
2. Sia $Y = X^2$ e $y > 0$, allora
$$\begin{aligned} \mathbb{P}(Y \le y) &= \mathbb{P}(X^2 \le y) = \mathbb{P}(X \le \sqrt{y}) = \int_0^{\sqrt{y}} \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right) \\ &= \left[-\exp\left\{-\frac{x^2}{2\theta^2}\right\}\right]_0^{\sqrt{y}} = 1 - \exp\left\{-\frac{y}{2\theta^2}\right\}; \end{aligned}$$
Concludiamo che $Y \sim Exp(2\theta^2) \stackrel{\mathrm{d}}{=} \Gamma(1, 2\theta^2)$

3. Sia $X_1, \ldots, X_n$ un campione dalla densità (1).

   (a) Dal punto precedente si deduce che $\sum_{i=1}^n X_i^2 \sim \Gamma(n, 2\theta^2)$, di conseguenza $\hat{k}_n = \frac{\sum_{i=1}^n X_i^2}{2n} \sim \Gamma(n, \frac{\theta^2}{n})$ quindi $\mathbb{E}(\hat{k}_n) = \theta^2$: $\hat{k}_n$ è NON distorto!

   (b) La derivata rispetto a θ^2 della log-verosimiglianza è

   $$\frac{\partial}{\partial\theta^2} l(\theta, \underline{x}) = -\frac{n}{\theta^2} + \frac{\sum^{2n}_{x_i}}{2(\theta^2)^2} = \frac{n}{(\theta^2)^2}\left(\hat{k}_n - \theta^2\right).$$

   Quindi $\hat{k}_n$ è lo stimatore efficiente.

   (c) Dato che il modello (1) è regolare, $\hat{k}_n$ è asintoticamente normale con media θ^2 e varianza $(\theta^2)^2/n$.

4. Dal punto 2. dell'esercizio si deduce che

   $$Q_n = \frac{2n}{\theta^2}\hat{k}_n \sim \Gamma(n, 2) \overset{\mathrm{d}}{=} \Gamma\left(\frac{2n}{2}, 2\right) \overset{\mathrm{d}}{=} \chi^2_{2n}$$

5. La variabile Q_n è una quantità pivotale in quanto la sua distribuzione non dipende da θ. Si fissino i due percentili $q_1 = \chi^2_{2n}\left(\frac{1-\gamma}{2}\right)$ e $q_2 = \chi^2_{2n}\left(\frac{1+\gamma}{2}\right)$, dove γ è compreso fra 0 e 1. Allora;

   $$\mathbb{P}\left(q_1 \leq Q_n \leq q_2\right) = \gamma \Rightarrow \mathbb{P}\left(q_1 \leq \frac{2n}{\theta^2}\hat{k}_n \leq_2\right) = \gamma \Rightarrow \mathbb{P}\left(\frac{2n}{q_2}\hat{k}_n \leq \theta^2 \leq \frac{2n}{q_1}\hat{k}_n\right) = \gamma.$$

   Concludiamo che

   $$\left[\frac{2n}{q_2}\hat{k}_n, \frac{2n}{q_1}\hat{k}_n\right] \tag{2}$$

   è un intervallo di confidenza di livello γ per θ^2. Inoltre dato che $k(\theta) = \theta^2$ per $\theta > 0$ è una funzione monotona crescente, l'intervallo

   $$\left[\sqrt{\frac{2n}{q_2}\hat{k}_n}, \sqrt{\frac{2n}{q_1}\hat{k}_n}\right] \tag{3}$$

   è di confidenza al livello γ per θ.

   Sia ora $\gamma = 0.9$, $n = 10$ e $\hat{\kappa}_n = 0.0387$, allora $q_1 = \chi^2_{20}(0.05)$ e $q_2 = \chi^2_{20}(0.95) = 31.4$. Sostituendo in tali valori in (2) si ottiene la realizzazione $(0.0246, 0.07118)$, mentre $(0.1570, 0.2665)$ é la realizzazione dell'intervallo in (3).

$\blacksquare$

Esercizio 2

Una moneta viene lanciata $n = 1000$ volte ottenendo 432 volte testa.

1. Determinare lo stimatore di massima verosimiglianza (ML) per θ, la probabilità di ottenere testa in un lancio.

2. Discutere qualche proprietà esatta e asintotica dello stmimatore $\hat{\theta}_n$ ricavato al punto 1.

3. Determinate l'informazione di Fisher del modello $I(\theta)$ ed un suo stimatore ML.

4. Costruire un intervallo di confidenza bilatero asintotico per θ di livello $\gamma = 0.95$.

5. Quanto deve essere grande n affinché l'intervallo calcolato al punto 4. abbia ampiezza al più pari a 0.02.

Soluzione

Per $i = 1, \dots, n$ si definiscano le variabili

$$X_i = \begin{cases} 1 & \text{se si ottiene testa all'i-esimo lancio} \\ 0 & \text{altrimenti} \end{cases}.$$

Costruendo così il campione $X_1, \dots, X_n$ da una popolazione Bernoulli(θ). La cui densità discreta è

$$p_X(x;\theta) = \theta^x (1-\theta)^{1-x}; \quad x \in \{0,1\}, \quad \theta \in [0,1].$$

1. Data la realizzazione $\underline{x} = (x_1, \dots, x_n)$ la verosimiglianza è

   $$L(\theta; \underline{x}) = \prod_{i=1}^{n} p_X(x_i \theta) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}.$$

   La log-verosimiglianza

   $$l(\theta; \underline{x}) = \log L(\theta; \underline{x}) = \left(\sum_{i=1}^n x_i \right) \log(\theta) + \left(n - \sum_{i=1}^n x_i \right) \log(1-\theta),$$

   Derivando rispetto a θ si ottiene

   $$\frac{\partial}{\partial \theta} l(\theta; \underline{x}) = \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum_{i=1}^n x_i}{1-\theta} = \frac{1}{\theta(1-\theta)} \left(\frac{\sum_{i=1}^n x_i}{n} - \theta \right).$$

   Dato che $\frac{\partial}{\partial \theta} l(\theta; \underline{x}) > 0 \;\Rightarrow\; \theta < \bar{x}_n$,, Si ottiene che

   $$\hat{\theta}_n = \bar{X}_n$$

   dove $\bar{X}_n = \sum_{i=1}^n X_i$ è la statistica media campionaria

2. Dato che $X_i \sim \mathit{Binomiale}(\theta)$, si ha che $\hat{\theta}$ é uno stimatore non distorto, esso ha inoltre varianza $\frac{\theta(1-\theta)}{n}$, ed è quindi consistente in media quadratica. Da come abbiamo scritto $\frac{\partial}{\partial \theta} \log L(\theta; x_1, \dots, x_n)$ è semplice rendersi conto che $\hat{\theta}$ è efficiente.

3. Dato che $\hat{\theta}$ è efficiente $I(\theta) = \frac{1}{n \mathrm{Var}(\hat{\theta})} = \frac{1}{\theta(1-\theta)}$. Per la proprietà di invarianza dello stimatore di massima verosimiglianza $\hat{I} = I(\hat{\theta})$.

4. Dato che il modello è regolare, lo stimatore di massima verosimiglianza $\hat{\theta}_n$ è asintoticamente gaussiano con media $\mathbb{E}(\hat{\theta}_n) = \theta$, e $\mathrm{Var}\hat{\theta}_n = \frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}$ (Si osservi come tali considerazioni si possono dedurre anche dal teorema de limite centrale). In conclusione, la quantità $\sqrt{nI(\hat{\theta})}(\hat{\theta} - \theta)$ ha distribuzione asintotica approssimativamente normale standard. Essa è dunque una quantità (asintoticamente) pivotale, dalla quale ricaviamo l'intervallo di confidenza asintotico:

   $$\left[\hat{\theta} - z_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}; \hat{\theta} + z_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \right],$$

   la cui realizzazione sulla base del campione fornito nel testo è $[0.4013; 04627]$.

5. La lunhezza dell'intervallo definito al punto 4. è

   $$\mathcal{L}_\gamma(n) = 2 z_{\frac{1+\gamma}{2}} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}.$$

   Questa è una quantità aleatoria, tuttavia osserviamo che se θ appartiene all'intervallo $[0,1]$, allora

   $$0 < \theta(1-\theta) < \frac{1}{4}.$$

Quindi con probabilità 1 si ha che

$$\mathcal{L}_\gamma(n) < z_{\frac{1+\gamma}{2}} \frac{1}{\sqrt{n}}.$$

Affinche la lunghezza dell'intervallo sia minore di 0.02 si scegliere n che soddisfa la disequazione

$$z_{\frac{1+\gamma}{2}} \frac{1}{\sqrt{n}} < 0.02$$

ovvero $n > 9506.25$.

∎