Politecnico di Milano – Facoltà di Ingegneria Informatica

Anno Accademico 2009/2010

Corso di Statistica (2L) per INF e TEL

Docente: Antonio Pievatolo; esercitazioni: Raffaele Argiento

Esercitazione del 09/04/10

Esercizio 1

Sia $X_1, \dots, X_n$ un campione iid da una distribuzione Poisson(θ).

1. Calcolare lo stimatore di massima verosimiglianza di θ.
2. Esso è non distorto?
3. È efficiente?

SOLUZIONE

1. Sia $\underline{x} = (x_1, \dots, x_n)$ la realizzazione campionaria relativa al campione $X_1, \dots, X_n$. Allora la funzione di verosimiglianza è:

$$L(\theta; \underline{x}) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \mathrm{e}^{-\theta} \frac{\theta_i^x}{x_i!} = \mathrm{e}^{-n\theta} \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}.$$

La log-verosimiglianza è:

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

Differenziando rispetto a θ

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

Quindi

$$\frac{\partial}{\partial \theta} l(\theta; \underline{x}) \geq 0 \Rightarrow \theta \leq \frac{\sum_{i=1}^n x_i}{n}.$$

Ne consegue che lo stimatore di massima verosimiglianza è $\hat{\theta} = \bar{X}$. quando ha il cappello è uno stimatore

2. Lo stimatore $\hat{\theta}$ trovato al punto precedente è banalmente non distorto.

3. Innazitutto bisogna verificare che le condizioni di regolarità del teorema di Frechét-Cramer-Rao (pg.14 delle dispense Stima Puntuale della Prof.ssa Epifani) siano verificate.

   i. banale.

   ii. banale.

   iii. banale.

   Definiamo le variabili

$$Y_1 = \frac{\partial}{\partial \theta} \ln f_X(X_1; \theta), Y_2 = \frac{\partial}{\partial \theta} \ln f_X(X_2; \theta), \dots, Y_n = \frac{\partial}{\partial \theta} \ln f_X(X_n; \theta)$$

   Si ossevi che $Y_i = \left(\frac{X_i}{\theta} - 1\right)$ per ogni $i = \{1, \dots, n\}$.

prendo la funzione di densità per la var aletatoria
del singolo campione, ci faccio il log(f(X)) e poi
derivo E VIENE Yi

uso l'ipotesi del teorema e applico proprietà della media

iv. $\mathbb{E}\left(\frac{\partial}{\partial\theta}\ln f_X(X_1;\theta)\right)=0 \Leftrightarrow \mathbb{E}(Y_1)=0 \Leftrightarrow \mathbb{E}\left(\frac{X_1}{\theta}-1\right) \Leftrightarrow \frac{\mathbb{E}(X_1)}{\theta}-1=0$. L'ultima uguaglianza è banalmente verificata.

v. $I(\theta)=\mathbb{E}(Y_1^2)=\mathbb{E}\left(\left(\frac{X_i}{\theta}-1\right)^2\right)=\mathbb{E}\left(\frac{X_1^2}{\theta^2}+1-2\frac{X_1}{\theta}\right)=\frac{\mathbb{E}(X_1^2)}{\theta^2}+1-2\frac{\mathbb{E}(X_1)}{\theta}=\frac{1}{\theta}$ Si ha dunque che $0<I(\theta)<\infty$ il che verifica la quinta condizione.

vi. $\mathbb{E}\left(\hat{\theta}\frac{\partial}{\partial\theta}L_\theta(X_1,\dots,X_n)\right)=\mathbb{E}\left(\hat{\theta}\sum_{i=1}^n Y_i\right)=\frac{1}{n}\mathbb{E}\left(\sum_{i=1}^n X_i\left(\frac{\sum_{i=1}^n X_i}{\theta}-1\right)\right)=$
$=\frac{1}{n\theta}\mathbb{E}\left(\left(\sum_{i=1}^n X_i\right)^2\right)-\mathbb{E}\left(\sum_{i=1}^n X_i\right)$. Si osservi che $\sum_{i=1}^n X_i \sim$Poisson$(n\theta)$. Si conclude $\mathbb{E}\left(\hat{\theta}\sum_{i=1}^n Y_i\right)=1$ che dimostra la sesta delle condizioni di regolarità.

Si osservi che

$$\frac{\theta}{n}=\mathrm{Var}(\hat{\theta})=\frac{1}{nI(\theta)}.$$

Concludiamo che lo stimatore ML è efficiente.
Soluzione alternativa: La derivata della log-verosimiglianza si puó fattorizzarecome segue che

$$\frac{\partial}{\partial\theta}L_\theta(X_1,\dots,X_n)=\frac{n}{\theta}\left(\hat{\theta}-\theta\right)$$

dal teorema di FCR (formula (11) pg.14 dispense Stima Puntuale della Prof.ssa Epifani) discende che $\hat{\theta}$ è efficiente.

∎

Esercizio 1

Tema d'Esame del 14/07/2009
Per verificare la Qualità di Servizio di un Sistema Software Aperto di acquisti telematici, abbiamo deciso di registrare le proporzioni giornaliere $X_1,\dots,X_n$ di richieste fallite dei prossimi n giorni. Astraendo, possiamo considerare $X_1,\dots,X_n$ come un campione casuale estratto da una popolazione di densità:

$$f(x,\theta)=\begin{cases}\frac{1}{\theta}x^{\frac{1}{\theta}-1} & 0<x<1\\ 0 & \text{altrove}\end{cases}$$

dove θ è un parametro positivo incognito.

1. Determinate lo stimatore dei momenti di θ.
2. Determinate lo stimatore di massima verosimiglianza di θ.
3. Verificate se lo stimatore di massima verosimiglianza di θ è efficiente.
4. Determinate lo stimatore di massima verosimiglianza della proporzione attesa $\mathbb{E}_\theta(X)$ di richieste giornaliere fallite. Quindi dimostrate che uno stimatore efficiente per $\mathbb{E}_\theta(X)$ NON esiste. (*Giustificate rigorosamente la risposta.*)
5. Monitorando il sistema software per $n=4$ giorni registriamo i seguenti risultati: il primo giorno falliscono 4 su 480 richieste arrivate, il secondo 5 su 450, il terzo 3 su 300 e il quarto 6 su 500. Qual è la stima di θ basata sul metodo dei momenti? E quella basata sul metodo di massima verosimiglianza?

SOLUZIONE

1. $\mathbb{E}_\theta(X_1)=\int_0^1 x\frac{1}{\theta}x^{\frac{1}{\theta}-1}\,dx=\frac{1}{\theta}\left[\frac{x^{1/\theta+1}}{1/\theta+1}\right]_0^1=\frac{1/\theta}{1/\theta+1}=\frac{1}{\theta+1}$

ed $\mathbb{E}_\theta(X_1)=\bar{X}$ se e solo se $\theta=1/\bar{X}-1$; segue che $\widehat{\theta}_{mom}=1/\bar{X}-1$ è lo stimatore dei momenti di θ.

importante da sapere.

metodo del momento: trovo momento primo e lo eguaglio alla media campionaria e la uso direttamnete poiché il suo valore "noto" a priori

2. Studiamo la funzione di verosimiglianza del campione $X_1, \dots, X_n$:

$$L_\theta(x_1, \dots, x_n) = \prod_{i=1}^n \left(\frac{1}{\theta} x_i^{\frac{1}{\theta}-1} \right) = \theta^{-n} \left(\prod_{i=1}^n x_i \right)^{\frac{1}{\theta}-1} \qquad \theta > 0$$

$$\log L_\theta(x_1, \dots, x_n) = -n \log \theta + \left(\frac{1}{\theta} - 1 \right) \sum_{j=1}^n \log x_j$$

$$\frac{\partial \log L_\theta(x_1, \dots, x_n)}{\partial \theta} = \frac{n}{\theta^2} \left(- \frac{\sum_{j=1}^n \log x_j}{n} - \theta \right) \geq 0 \tag{1}$$

se e solo se $\theta \leq -(1/n) \sum_{j=1}^n \log x_j$; infine osserviamo che $-\sum_{j=1}^n \log X_j > 0$ poiché $P(0 < X_j < 1) = 1$. Segue che $\widehat{\theta}_{ML} = -\dfrac{\sum_{j=1}^n \log X_j}{n}$ è stimatore ML per θ.

3. La densità $f(x, \theta)$ è "regolare" e leggiamo nell'equazione (1) che la derivata del logaritmo della funzione di verosimiglianza *"essenzialmente"* è funzione lineare della differenza fra $\widehat{\theta}_{ML}$ e θ. Ma (1) è CNS affinchè la varianza di $\hat{\theta}$ raggiunga il confine di Fréchet-Cramér-Rao, cioè affinché $\widehat{\theta}_{ML}$ sia stimatore efficiente.

4. Poiché $\kappa = \mathbb{E}_\theta(X_1) = 1/(1+\theta)$, allora, $\widehat{\kappa}_{ML} = 1/(1 + \widehat{\theta}_{ML})$. Inoltre, deriviamo da (1) che

$$\frac{\partial \log L_\theta(x_1, \dots, x_n)}{\partial \theta} = \frac{n}{\theta^2} ((1 + \widehat{\theta}_{ML}) - (1 + \theta)) = \frac{n}{\theta^2} \left(\frac{1}{\widehat{\kappa}_{ML}} - \frac{1}{\kappa} \right) :$$

cioè $\partial \log L_\theta / \partial \theta$ è funzione lineare di $\widehat{\kappa}_{ML}^{-1}$, quindi non può esserlo di $\widehat{\kappa}_{ML}$: non essendo soddisfatta una CNS per l'efficienza, allora $\widehat{\kappa}_{ML}$ non è stimatore efficiente di κ. D'altro canto se uno stimatore efficiente per κ esiste, allora esso è necessariamente stimatore ML: rimane così stabilito che NON esiste nessun stimatore efficiente di κ.

5. Deriviamo dai dati forniti il seguente campione: $x_1 = 4/480,\ x_2 = 5/450$, $x_3 = 3/300,\ x_4 = 6/500$; quindi $\bar{x} = 373/36000 \simeq 0.0104$ e $\widehat{\theta}_{mom} = 35627/373 \simeq 95.515$; $\widehat{\theta}_{ML} \simeq 4.579$. ∎