Teoria e Laboratorio di Inferenza Statistica - Statistica Inferenziale - Statistica 2

Corsi di Laurea D E e SAF. Anno Accademico 2006-2007
Esonero n. 1 - 21 aprile 2007. Ore 10:30 - 12:30

Soluzioni

Esercizio 1 Sia $X_1, X_2, \dots, X_n$ un campione casuale proveniente da una popolazione di Poisson di parametro $\theta = 0.1$.

a) Determinare la distribuzione esatta e l'approssimazione normale per la statistica campionaria $U_n = 2Y_n - 3$, dove $Y_n = \sum_{i=1}^n X_i$.

b) Determinare la probabilità che U_n assuma valori positivi, considerando $n = 20$.

Svolgimento.

a) La distribuzione esatta di $Y_n = \sum_{i=1}^n X_i$ è la distribuzione di Poisson di parametro $n\theta$. Quindi Y_n può assumere i valori $y = 0, 1, \dots, y \dots$ con probabilità

$$f_{Y_n}(y;\theta) = \mathbb{P}(Y_n = y;\theta) = \frac{e^{-n\theta}(n\theta)^y}{y!}.$$

La variabile aleatoria $U_n = 2Y_n - 3$ assume di conseguenza i valori $u = 0, -3, -1, 1, 3, 5, \dots 2y - 3, \dots$ con probabilità

$$f_{U_n}(u;\theta) = \mathbb{P}(U_n = u;\theta) = \mathbb{P}(2Y_n - 3 = u;\theta) = \mathbb{P}(Y_n = \frac{u+3}{2};\theta) = f_{Y_n}((u+3)/2;\theta) = \frac{e^{-n\theta}(n\theta)^{(u+3)/2}}{[(u+3)/2]!}.$$

b) Sono verificate le ipotesi del teorema del limite centrale: le v.a. $X_1, \dots, X_n$ sono i.i.d., e valore atteso e varianza di X_i sono finiti. Pertanto,

$$Y_n = \sum_{i=1}^n X_i \mathrel{\dot\sim} N(n\theta, n\theta) = N(2, 2).$$

Per le proprietà delle v.a. normali, una combinazione lineare di una v.a. normale ha ancora distribuzione normale. Poichè

$$\mathbb{E}_\theta[U_n] = 2\mathbb{E}_\theta[Y_n] - 3 = 2n\theta - 3, \qquad \mathbb{V}_\theta[U_n] = 4\mathbb{V}_\theta[Y_n] = 4n\theta,$$

si ha che

$$U_n = 2Y_n - 3 \mathrel{\dot\sim} N(2n\theta - 3, 4n\theta) = N(1, 8),$$

e quindi che

$$\mathbb{P}(U_n > 0) \approx \mathbb{P}\left(\frac{U_n - 1}{\sqrt{8}} > \frac{-1}{\sqrt{8}}\right) = \mathbb{P}\left(Z > -0.36\right) = 1 - \Phi(-0.36) = \Phi(0.36) = 0.64.$$

NOTA BENE: NON CONFONDERE L' APPROSSIMAZIONE NORMALE DI UNA DISTRIBUZIONE CAMPIONARIA CON L'APPROSSIMAZIONE NORMALE DELLA F.NE DI VEROSIMIGLIANZA.

Esercizio 2. Si consideri un campione casuale di n osservazioni da una popolazione X con distribuzione di probabilità:

$$f_X(x;\theta) = \frac{1}{\theta^2}\, x\, \exp\left\{-\frac{x}{\theta}\right\}, \qquad \theta > 0, \qquad x > 0.$$

a) Scrivere il modello statistico probabilistico e verificare che $f_X(x;\theta)$ appartiene alla famiglia esponenziale.

b) Determinare la funzione di verosimiglianza $L_{\mathbf{x}_{oss}}(\theta)$.

c) Ottenere, se esiste, una statistica sufficiente per θ.

d) Determinare lo stimatore di massima verosimiglianza per θ.

e) Determinare l'espressione dell'informazione osservata di Fisher.

Svolgimento.

a) Modello statistico:

$$\left(\mathcal{X}^n = (\mathbb{R}^+)^n,\ f_n(\mathbf{x};\theta) = \theta^{-2n}\,[\prod_{i=1}^{n} x_i I_{\mathbb{R}^+}(x_i)]\ \exp\left\{-\frac{\sum_{i=1}^n x_i}{\theta}\right\},\ \Theta = \mathbb{R}^+\right).$$

La distribuzione di X appartiene alla famiglia esponenziale in quanto può essere scritta come:

$$f_X(x;\theta) = x\ \exp\left\{-2\log\theta - \frac{x}{\theta}\right\} I_{\mathbb{R}^+}(x) = h(x)\exp\{\eta(\theta)T(x) - B(\theta)\},$$

con:

$$h(x) = xI_{\mathbb{R}^+}(x) \qquad \eta(\theta) = -\frac{1}{\theta} \qquad T(x) = x \qquad B(\theta) = 2\log\theta.$$

b) La funzione di verosimiglianza è:

$$L(\theta;\mathbf{x}_n) = \theta^{-2n}\, e^{-\frac{\sum_{i=1}^n x_i}{\theta}}$$

c) Una statistica sufficiente è data da $T(\mathbf{x}) = \sum_{i=1}^n x_i$.

d) Per determinare lo stimatore di massima verosimiglianza si considera:

$$\ell(\theta) = \log L(\theta;\mathbf{x}_n) = -2n\log\theta - \frac{\sum_{i=1}^n x_i}{\theta}$$

che ha derivata prima data da:

$$\ell'(\theta) = -\frac{2n}{\theta} + \frac{\sum_{i=1}^n x_i}{\theta^2} \qquad \text{e} \qquad \text{quindi} \qquad \widehat{\theta}_{mv} = \frac{\sum_{i=1}^n x_i}{2n}.$$

Inoltre $\ell''(\theta) = 2n/\theta^2 - 2\sum_{i=1}^n x_i/\theta^3$ che risulta essere < 0 in $\theta = \widehat{\theta}_{mv}$.

e) L'informazione osservata di Fisher è data da:

$$I(\widehat{\theta}) = -\ell''(\theta)\big|_{\theta=\widehat{\theta}_{mv}} = \frac{2n}{\widehat{\theta}^2_{mv}} = \frac{8n^3}{(\sum_{i=1}^n x_i)^2}.$$

Esercizio 3. Sia

$$\mathbf{x}_{oss} = (0.56, 0.47, 0.30, 0.60, 0.22, 0.41, 0.76, 0.38, 0.08, 0.29, 0.57, 0.97, 0.81, 0.87, 0.36, \\ 0.20, 1.27, 0.20, 1.38, 1.12, 0.46, 0.52, 1.17, 0.32, 0.21, 0.61, 0.61, 1.47, 0.64, 0.08)$$

Un campione di numerosità $n = 30$ estratto da una popolazione distribuita come la variabile aleatoria X dell'esercizio 2. Si ha che $\sum_{i=1}^{n} x_i = 17.91$, $(\sum_{i=1}^{n} x_i)^2 = 320.77$, $(\sum_{i=1}^{n} x_i)^3 = 5744.96$. Si determini

a) Il valore della stima di massima verosimiglianza $\widehat{\theta}_{mv}$.

b) Il valore dell'informazione osservata di Fisher.

c) L'intervallo di verosimiglianza approssimato di livello $q = 0.146$.

d) Il valore approssimato della seguente probabilità, utilizzando il valore calcolato di $\widehat{\theta}_{mv}$ e tenendo conto che la variabile aleatoria X ha media $\mathbb{E}[X] = 2\theta$ e varianza $\mathbb{V}[X] = 2\theta^2$,

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i < 10\right).$$

Svolgimento.

a) Il valore della stima di massima verosimiglianza è:

$$\widehat{\theta}_{mv} = \frac{\sum_{i=1}^{n} x_i}{2n} = \frac{17.91}{60} = 0.2985.$$

b) Il valore dell'informazione osservata di Fisher:

$$I_n^{oss}(\mathbf{x}_n) = \frac{8n^3}{(\sum_{i=1}^{n} x_i)^2} = \frac{216.000}{320.77} = 673.38.$$

c) L'intervallo di verosimiglianza approssimato si ottiene sfruttando l'approssimazione normale della funzione di verosimiglianza (DA NON CONFONDERE CON L'APPROSSIMAZIONE NORMALE DI UNA DISTRIBUZIONE CAMPIONARIA) nell'intorno del punto $\theta = \widehat{\theta}_{mv}$, da cui si ottiene che, ponendo $k = \sqrt{-2 \log q} = 1.96$:

$$L_q = \widehat{\theta}_{mv} \pm k \times [I_n^{oss}(\mathbf{x}_n)]^{-1/2} = 0.2985 \pm \frac{1.96}{\sqrt{673.38}} = 0.2985 \pm 0.0755.$$

e l'intervallo è $(0.22, 0.37)$.

d) Le ipotesi del TLC sono verificate: variabili aleatorie i.i.d. con valore atteso e varianza finiti. Dato che $\sum_{i=1}^{n} X_i \,\dot\sim\, N(2n\theta, 2n\theta^2) =$ (per $\theta = \widehat{\theta}_{mv}$)$= N(17.91, 5.34)$, si ottiene:

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i < 10\right) \approx \mathbb{P}\left(Z < \frac{10 - 17.91}{2.32}\right) = \mathbb{P}\left(Z < -3.41\right) = \Phi(-3.41) = \texttt{pnorm}(-3.41) = 0.00032.$$

NOTA BENE: NON CONFONDERE L' APPROSSIMAZIONE NORMALE DI UNA DISTRIBUZIONE CAMPIONARIA CON L'APPROSSIMAZIONE NORMALE DELLA F.NE DI VEROSIMIGLIANZA.