Inferenza Statistica
I prova scritta a.a. 2011-2012 – 12 giugno 2012

I PARTE

ATTENZIONE: I FOGLI A QUADRETTI CONSEGNATI
NON VERRANNO CONSIDERATI

• • •

Cognome e Nome: ________________________ **Canale** SEFA (De Santis) ☐
SES - SG (Perone Pacifico) ☐

Esercizio 1. Si consideri un campione casuale $X_1, \ldots, X_n$ proveniente da una popolazione con funzione di densità di probabilità

$$f_X(x;\theta) = \frac{\theta}{x^{\theta+1}}, \qquad x > 1, \qquad \theta > 2.$$

1. Verificare che $\{f_X(\cdot;\theta),\ \theta \in \Theta\}$ è una famiglia esponenziale, determinare la funzione di verosimiglianza, individuarne il nucleo e determinare una statistica sufficiente per il modello. Dire se la statistica sufficiente determinata è anche completa.

2. Verificare che la stima di massima verosimiglianza e l'informazione osservata risultano rispettivamente uguali a

$$\widehat{\theta}_{mv} = \frac{n}{\ln \prod_{i=1}^n x_i}, \qquad I_n^{oss} = \frac{(\ln \prod_{i=1}^n x_i)^2}{n}.$$

   Determinare il valore di $\widehat{\theta}_{mv}$ e di I_n^{oss} per il campione $\mathbf{x}_3 = (1.5, 1.4, 1.8)$. Calcolare la densità di probabilità del campione $\mathbf{x}_3 = (1.5, 1.4, 1.8)$ assumendo $\theta = 3$.

3. Verificare che valore atteso di X è uguale a

$$\mathrm{E}[X] = \frac{\theta}{\theta - 1}.$$

   Determinare lo stimatore dei momenti di θ. Tale stimatore può essere lo stimatore UMVUE (non distorto di minima varianza) del parametro θ? Calcolare il valore della stima per il campione osservato $\mathbf{x}_3 = (1.5, 1.4, 1.8)$.

Svolgimento:

1. Osservando che la funzione di densità si può scrivere come

$$f_X(x;\theta) = \frac{\theta}{x^{\theta+1}} = \exp\{\ln\theta - (\theta+1)\ln x\}$$

e ponendo $h(x) = 1$, $\eta(\theta) = -(\theta+1)$, $T(x) = \ln x$, $B(\theta) = \ln\theta$, possiamo concludere che $\{f_X(\cdot;\theta),\ \theta \in \Theta\}$ è una famiglia esponenziale.

La funzione di verosimiglianza è

$$L(\theta) = \theta^n \left(\prod_{i=1}^n x_i\right)^{-\theta}.$$

Una statistica sufficiente per il modello è $T(\mathbf{x}) = \prod_{i=1}^n x_i$ che è anche una completa dal momento che $\{f_X(\cdot;\theta),\ \theta \in \Theta\}$ è una famiglia esponenziale.

2. Per trovare la stima di massima verosimiglianza, passiamo alla logverosimiglianza:

$$\ell(\theta) = n\ln\theta - \theta\sum_{i=1}^n \ln x_i$$

calcoliamo la derivata prima

$$\ell'(\theta) = \frac{n}{\theta} - \sum_{i=1}^n \ln x_i$$

e, eguagliandola a 0, otteniamo $\widehat{\theta}_{mv} = \frac{n}{\ln\prod_{i=1}^n x_i}$, che è la stima di massima verosimiglianza dal momento che la derivata seconda

$$\ell''(\theta) = -\frac{n}{\theta^2}$$

è negativa per ogni valore di θ.

L'Informazione osservata è

$$I_n^{oss} = \ell''(\theta)|_{\theta=\widehat{\theta}_{mv}} = \frac{n}{\widehat{\theta}_{mv}^2} = \frac{(\ln\prod_{i=1}^n x_i)^2}{n}.$$

In corrispondenza del campione osservato $\mathbf{x}_3 = (1.5, 1.4, 1.8)$ si ha $\widehat{\theta}_{mv} = 2.26$ e $I_n^{oss} = 0.59$.

La densità congiunta del campione osservato per $\theta = 3$ è

$$f(\mathbf{x};\theta) = \theta^n\left(\prod_{i=1}^n x_i\right)^{-(\theta+1)} = 3^3(1.5\cdot 1.4\cdot 1.8)^{-4} = 0.13$$

3. Il valore atteso è

$$\mathrm{E}(X) = \int_1^{\infty} x \cdot \frac{\theta}{x^{\theta+1}} dx = \theta \int_1^{\infty} x^{-\theta} dx = \theta \left[\frac{x^{-\theta+1}}{-\theta+1} \right]_1^{\infty} = \frac{\theta}{\theta-1}.$$

Lo stimatore dei momenti si ottiene ponendo:

$$\mathrm{E}(X) = \frac{\theta}{\theta-1} = \overline{X}_n \Longrightarrow \widehat{\theta}_{mom} = \frac{\overline{X}_n}{\overline{X}_n - 1}.$$

Senza voler procedere alla verifica della non distorsione di $\widehat{\theta}_{mom}$, in ogni caso possiamo escludere che tale stimatore sia UMVUE perchè non è funzione della statistica sufficiente completa determinata al punto 1. Per il campione osservato $\mathbf{x}_3 = (1.5, 1.4, 1.8)$, la stima è $\widehat{\theta}_{mom} = 2.76$.

Esercizio 2. Con riferimento al modello statistico dell'esercizio 1, si consideri la quantità

$$g(\theta) = \frac{2}{\theta - 1}.$$

1. Verificare che lo stimatore
$$T(\mathbf{X}_n) = 2(\overline{X}_n - 1)$$
uno stimatore non distorto per $g(\theta)$. Sapendo che
$$\mathrm{V}[X] = \frac{\theta}{(\theta-2)(\theta-1)^2},$$
determinare la varianza e la distribuzione asintotica di T.

2. Determinare l'errore quadratico medio (MSE) di T e stabilire se lo stimatore è consistente. Calcolare il valore dell'errore quadratico medio di T, assumendo $\theta = 3$ e $n = 16$.

3. Utilizzando l'approssimazione normale determinata punto 1 si ottiene che, $\forall \epsilon > 0$:
$$\mathrm{P}(|T(\mathbf{X}_n) - g(\theta)| > \epsilon) \approx 2\Phi\left(-\frac{\epsilon}{v_n(\theta)}\right) \tag{1}$$
dove
$$v_n(\theta) = \left[\frac{4\theta}{n(\theta-2)(\theta-1)^2}\right]^{1/2}$$
e dove $\Phi()$ è la funzione di ripartizione della v.a. Normale standard.

   a) Calcolare il valore di (1) assumendo $\epsilon = 1/8$, $\theta = 3$ e $n = 16$.

   b) Verificare che, per n che tende ad infinito, la probabilità definita in (1) tende a 0, per ogni valore di $\epsilon > 0$ e per ogni valore di $\theta > 0$.

   c) Dimostrare la relazione (1) utilizzando l'approssimazione normale per la distribuzione di T.

Svolgimento:

1. Lo stimatore $T(\mathbf{X}_n) = 2(\overline{X}_n - 1) = 2\overline{X}_n - 2$ è non distorto per $g(\theta)$, infatti:
$$\mathrm{E}(T(\mathbf{X}_n)) = \mathrm{E}(2\overline{X}_n - 2) = 2\mathrm{E}(X) - 2 = 2\frac{\theta}{\theta-1} - 2 = \frac{2\theta - 2\theta + 2}{\theta - 1} = \frac{2}{\theta-1}.$$
La varianza di $T(\mathbf{X}_n)$ è
$$\mathrm{V}(T(\mathbf{X}_n)) = \mathrm{V}(2\overline{X}_n - 2) = 4\frac{\mathrm{V}(X)}{n} = \frac{4\theta}{n(\theta-2)(\theta-1)^2}.$$
La distribuzione asintotica di T è dunque
$$N\left(\frac{2}{\theta-1}, \frac{4\theta}{n(\theta-2)(\theta-1)^2}\right).$$

2. Poiché T è non distorto, il suo errore quadratico medio coincide con la varianza

$$MSE(T(\mathbf{X}_n) = \mathrm{V}(T(\mathbf{X}_n)) = \frac{4\theta}{n(\theta-2)(\theta-1)^2}.$$

Poiché $MSE(T(\mathbf{X}_n) \to \infty$ per $n \to \infty$, T è consistente. Per $\theta = 3$ e $n = 16$ si ha $MSE(T(\mathbf{X}_n) = 3/16 = 0.1875$.

3. a) La probabilità in (1) vale

$$\mathrm{P}(|T(\mathbf{X}_n) - g(\theta)| > \epsilon) \approx 2\Phi\left(-\frac{\epsilon}{v_n(\theta)}\right) = 2\Phi\left(-\frac{1/8}{\sqrt{3/16}}\right) = 2\Phi(-0.289) = 2 \cdot 0.386 = 0.772.$$

b) Per $n \to \infty$, si ha che

$$v_n(\theta) = \left[\frac{4\theta}{n(\theta-2)(\theta-1)^2}\right]^{1/2} \to 0 \Longrightarrow -\frac{\epsilon}{v_n(\theta)} \to -\infty \Longrightarrow 2\Phi\left(-\frac{\epsilon}{v_n(\theta)}\right) \to 0$$

e quindi la probabilità in (1) tende a 0 per ogni ϵ e per ogni θ.

c) Sapendo che

$$T \approx N\left(\frac{2}{\theta-1}, \frac{4\theta}{n(\theta-2)(\theta-1)^2}\right)$$

si verifica facilmente che

$$\begin{aligned}\mathrm{P}(|T(\mathbf{X}_n) - g(\theta)| > \epsilon) &= 2P(T(\mathbf{X}_n) - g(\theta) > \epsilon) = 2P\left(\frac{T(\mathbf{X}_n) - g(\theta)}{v_n(\theta)} > \frac{\epsilon}{v_n(\theta)}\right) \\ &\approx 2\Phi\left(-\frac{\epsilon}{v_n(\theta)}\right).\end{aligned}$$