Inferenza Statistica

II ESONERO A.A. 2011-2012 – 31 maggio 2012

ATTENZIONE: I FOGLI A QUADRETTI CONSEGNATI
NON VERRANNO CONSIDERATI

• • •

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

Esercizio 1. Sia X una variabile aleatoria assolutamente continua con funzione di densità di probabilità

$$f_X(x;\theta) = x(\theta+2)x^{\theta}, \qquad x \in (0,1), \qquad \theta > -2.$$

1. Verificare che lo stimatore di massima verosimiglianza di θ è

$$\widehat{\theta}_{mv}(\mathbf{x}_n) = -\frac{n}{\ln \prod_{i=1}^{n} x_i} - 2.$$

   Determinare il limite inferiore di Cramer-Rao per gli stimatori non distorti di θ.

2. Verificare che la distribuzione asintotica dello stimatore di massima verosimiglianza è

$$N\left(\theta, \frac{(\theta+2)^2}{n}\right).$$

   Considerato un campione casuale di $n = 20$ osservazioni in corrispondenza delle quali si ha $\ln \prod_{i=1}^{n} x_i = -19$, determinare la stima di massima verosimiglianza di θ e l'intervallo di confidenza approssimato di livello $1-\alpha = 0.80$.

3. Verificare che

$$E[X] = \frac{\theta+2}{\theta+3} = 1 - \frac{1}{\theta+3}.$$

   Determinare per questa quantità la stima di massima verosimiglianza e l'intervallo di confidenza approssimato di livello 0.95, utilizzando i dati del punto precedente.

Svolgimento:

1. La funzione di verosimiglianza è:

$$L(\theta;\mathbf{x}_n) = \left(\prod_{i=1}^{n} x_i\right)^{\theta} (\theta+2)^n$$

La log-verosimiglianza è:

$$l(\theta) = \theta \ln \prod_{i=1}^{n} x_i + n \ln(\theta + 2)$$

Annullando la derivata prima di $l(\theta)$

$$\frac{d}{d\theta} l(\theta) = \ln \prod_{i=1}^{n} x_i + \frac{n}{\theta + 2} = 0$$

si ottiene

$$\widehat{\theta}_{mv} = -\frac{n}{\ln \prod_{i=1}^{n} x_i} - 2$$

che è un punto di massimo in quanto la derivata seconda di $l(\theta)$ è negativa per ogni θ

$$\frac{d^2}{d\theta^2} l(\theta) = -\frac{n}{(\theta + 2)^2}.$$

Il limite inferiore di Cramer-Rao per gli stimatori non distorti di θ è:

$$\frac{1}{I(\theta)} = \frac{1}{-\mathbb{E}[\frac{d^2}{d\theta^2} l(\theta)]} = \frac{(\theta + 2)^2}{n}$$

2. La distribuzione asintotica dello stimatore di massima verosimiglianza è Normale con media θ e varianza $I^{-1}(\theta)$, quindi

$$\widehat{\theta}_{mv} \sim N\left(\theta; \frac{(\theta + 2)^2}{n}\right).$$

Per il campione in esame, con $n = 20$, $\ln \prod_{i=1}^{n} x_i = -19$, la stima di massima verosimiglianza è

$$\widehat{\theta}_{mv} = -\frac{n}{\ln \prod_{i=1}^{n} x_i} - 2 = -\frac{20}{-19} - 2 = -\frac{18}{19} = -0.95$$

e l'intervallo approssimato di livello $1 - \alpha = 0.80$

$$\widehat{\theta}_{mv} \pm z_{1-\alpha/2} \sqrt{\frac{(\widehat{\theta}_{mv} + 2)^2}{n}} = -0.95 \pm 1.28 \frac{-0.95 + 2}{\sqrt{20}} = (-1.25; -0.65).$$

3. Il valore atteso è

$$\mathbb{E}(X) = \int_0^1 x \cdot f_X(x; \theta) dx = \int_0^1 x \cdot x^{\theta+1} (\theta + 2) dx = (\theta + 2) \frac{x^{\theta+3}}{\theta + 3}|_0^1 = \frac{\theta + 2}{\theta + 3}.$$

Per la proprietà di invarianza, la stima di massima verosimiglianza di $g(\theta) = \mathbb{E}(X) = \frac{\theta+2}{\theta+3}$ è

$$\widehat{g(\theta)} = g(\widehat{\theta}_{mv}) = \frac{\widehat{\theta}_{mv} + 2}{\widehat{\theta}_{mv} + 3}.$$

Per ottenere l'intervallo di confidenza approssimato a livello 0.95, si ricorre al metodo delta, osservando che $g'(\theta) = \frac{1}{(\theta+3)^2}$ e ricordando l'espressione trovata al punto 1. per $I_n(\theta)$

$$g(\widehat{\theta}_{mv}) \pm z_{1-\alpha/2}\sqrt{(g'(\widehat{\theta}_{mv}))^2 I_n(\widehat{\theta}_{mv})^{-1}}$$

$$\frac{\widehat{\theta}_{mv}+2}{\widehat{\theta}_{mv}+3} \pm z_{1-\alpha/2}\sqrt{\left(\frac{1}{(\widehat{\theta}_{mv}+3)^2}\right)^2 \cdot \frac{(\widehat{\theta}_{mv}+2)^2}{n}}$$

$$\frac{\widehat{\theta}_{mv}+2}{\widehat{\theta}_{mv}+3} \pm z_{1-\alpha/2}\frac{(\widehat{\theta}_{mv}+2)}{\sqrt{n}(\widehat{\theta}_{mv}+3)^2}$$

$$\frac{-0.95+2}{-0.95+3} \pm 1.96\frac{(-0.95+2)}{\sqrt{20}(-0.95+3)^2}$$

$$0.51 \pm 1.96 \cdot 0.056 = (0.4; 0.62)$$

Esercizio 2. Con riferimento al parametro del modello statistico dell'esercizio n.1, si consideri il sistema di ipotesi

$$H_0 : \theta = \theta_0 = 1, \qquad H_1 : \theta = \theta_1 = 0.$$

1. Verificare che, per $n = 1$, il test di Neyman-Pearson ha la seguente regione di accettazione:

$$A = \{x \in (0,1) : \; x > k\},$$

dove k un generico valore positivo. Assumendo $k = 4/5$, determinare le probabilità di errore di I e II specie e la potenza del test.

2. Verificare che, per un campione casuale di dimensione n, la generica regione di accettazione del test di Neyman-Pearson è così definita:

$$A = \{\mathbf{x}_n : \; \widehat{\theta}_{mv}(\mathbf{x}_n) > k\},$$

dove k è un generico valore positivo e dove l'espressione di $\widehat{\theta}_{mv}(\mathbf{x}_n)$ (stimatore di massima verosimiglianza) è fornita nell'esercizio 1.

3. Utilizzando la distribuzione asintotica di $\widehat{\theta}_{mv}$, verificare che la regione di accettazione del test con probabilità di errore di I specie pari a $\alpha = 0.05$ è:

$$A = \left\{\mathbf{x}_n : \; \widehat{\theta}_{mv}(\mathbf{x}_n) > \theta_0 - 1.65\frac{\theta_0 + 2}{\sqrt{n}}\right\}.$$

Stabilire se, con i dati a disposizione (vedi Punto 2 dell'esercizio precedente), l'ipotesi nulla viene o meno accettata.

[**Suggerimento**: utilizzare $[I_n(\theta_0)]^{-1}$ come varianza asintotica per l'approssimazione normale].

Svolgimento:

1. Per $n = 1$ il rapporto delle verosimiglianze è:

$$\frac{L(\theta_0; x)}{L(\theta_1; x)} = \frac{x^{\theta_0}(\theta_0 + 2)}{x^{\theta_1}(\theta_1 + 2)} = x^{\theta_0 - \theta_1}\frac{\theta_0 + 2}{\theta_1 + 2} = \frac{3}{2}x$$

che è funzione crescente di x.
In base al Lemma di Neyman-Pearson la regione di accettazione è dunque:

$$A = \{x \in (0,1) : \; x > k\}.$$

Fissando $k = 4/5$, possiamo calcolare:

$$\alpha = P(R|H_0) = P(X < k|\theta = \theta_0) = \int_0^{4/5} x^{\theta_0 + 1}(\theta_0 + 2)dx = 3\int_0^{4/5} x^2 dx = 3\frac{x^3}{3}|_0^{4/5} = \frac{64}{125} = 0.512$$

$$\beta = P(A|H_1) = P(X > k|\theta = \theta_1) = \int_{4/5}^{1} x^{\theta_1 + 1}(\theta_1 + 2)dx = 2\int_{4/5}^{1} x dx = 2\frac{x^2}{2}|_{4/5}^{1} = 1 - \frac{16}{25} = 0.36$$

$$1-\beta = 1-0.36 = 0.64$$

2. Per un campione casuale di dimensione n il rapporto delle verosimiglianze è:

$$\frac{L(\theta_0;\mathbf{x}_n)}{L(\theta_1;\mathbf{x}_n)} = \frac{\left(\prod_{i=1}^n x_i\right)^{\theta_0}(\theta_0+2)^n}{\left(\prod_{i=1}^n x_i\right)^{\theta_1}(\theta_1+2)^n} = \left(\prod_{i=1}^n x_i\right)^{\theta_0-\theta_1}\left(\frac{\theta_0+2}{\theta_1+2}\right)^n = \left(\prod_{i=1}^n x_i\right)\left(\frac{3}{2}\right)^n =$$

che è funzione crescente di $\prod_{i=1}^n x_i$ e quindi è funzione crescente anche di $-\frac{n}{\ln\prod_{i=1}^n x_i}-2=\widehat{\theta}_{mv}$. In base al Lemma di Neyman-Pearson la regione di accettazione è dunque:

$$A = \{\mathbf{x}_n : \ \widehat{\theta}_{mv}(\mathbf{x}_n) > k\}.$$

3. Utilizzando la distribuzione asintotica di $\widehat{\theta}_{mv}$ e imponendo $\alpha = 0.05$ si ha:

$$\alpha = P(R|H_0) = P(\widehat{\theta}_{mv} < k|\theta_0) = P\left(\frac{\widehat{\theta}_{mv}-\theta_0}{\sqrt{\frac{(\theta_0+2)^2}{n}}} < \frac{k-\theta_0}{\sqrt{\frac{(\theta_0+2)^2}{n}}}\right) = \Phi\left(\frac{k-\theta_0}{\sqrt{\frac{(\theta_0+2)^2}{n}}}\right)$$

$$\Longrightarrow z_\alpha = \frac{k-\theta_0}{\sqrt{\frac{(\theta_0+2)^2}{n}}} \Longrightarrow k = z_\alpha\frac{(\theta_0+2)}{\sqrt{n}}+\theta_0 = -1.65\frac{3}{\sqrt{20}}+1 = -0.107$$

Poiché la stima di massima verosimiglianza è $\widehat{\theta}_{mv} = -0.95 < k = -0.107$, si rifiuta l'ipotesi nulla.