Inferenza Statistica

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

II PARTE

ATTENZIONE: I FOGLI A QUADRETTI CONSEGNATI
NON VERRANNO CONSIDERATI

•••

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

Esercizio 1. Sia $X_1, \ldots, X_n$ un campione casuale proveniente da una popolazione di Poisson di parametro incognito $\theta/2$, $\theta > 0$.

1. Verificare che $\widehat{\theta}_{mom}$ e $\widehat{\theta}_{mv}$, rispettivamente stimatore dei momenti e di massima verosimiglianza di θ, coincidono e che la loro distribuzione asintotica risulta essere:

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

2. Verificare che la probabilità che la prima osservazione campionaria assuma valore pari a 0 è

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

Determinare la stima di massima verosimiglianza di $g(\theta)$ e un intervallo di confidenza approssimato di livello $1 - \alpha = 0.90$, per un campione osservato di dimensione $n = 20$ per il quale si sa che

$$\frac{1}{2}\overline{x}_n + 2 = 20.$$

3. Utilizzando la distribuzione asintotica di $\widehat{\theta}_{mv}$, verificare che l'intervallo aleatorio

$$\left[2\overline{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\overline{X}_n}{n}}, 2\overline{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\overline{X}_n}{n}}\right]$$

è un intervallo di confidenza approssimato di livello $1 - \alpha$ per θ. Determinare l'espressione della lunghezza (aleatoria) $\mathcal{L}(\mathbf{X}_n)$ dell'intervallo considerato . Determinare il minimo valore di n tale per cui

$$E_\theta[\mathcal{L}(\mathbf{X}_n)^2] < \ell_o$$

assumendo $\alpha = 0.05$, $\theta = 1$ e $\ell_0 = 0.5$.

Svolgimento:

1. Ricordiamo che per una popolazione di Poisson di parametro incognito λ lo stimatore di massima verosimiglianza è $\widehat{\lambda}_{mv} = \overline{X}_n$. Ponendo $\lambda = \frac{\theta}{2} \Longrightarrow \theta = 2\lambda$ e sfruttando la proprietà di equivarianza si ottiene $\widehat{\theta}_{mv} = 2\overline{X}_n$.

   Inoltre, poiché il valore atteso $\mathrm{E}(X)$ coincide con il parametro, per ricavare lo stimatore dei momenti si deve porre

   $$\mathrm{E}(X) = \frac{\theta}{2} = \overline{X}_n \Longrightarrow \widehat{\theta}_{mom} = 2\overline{X}_n.$$

   Quindi $\widehat{\theta}_{mv}$ e $\widehat{\theta}_{mom}$ coincidono. Poiché

   $$\mathrm{E}(2\overline{X}_n) = 2\mathrm{E}(X) = 2 \cdot \frac{\theta}{2} = \theta$$

   e

   $$\mathrm{V}(2\overline{X}_n) = 4\frac{\mathrm{V}(X)}{n} = 4 \cdot \frac{\theta}{2n} = \frac{2\theta}{n},$$

   per il teorema del limite centrale, la loro distribuzione asintotica di $\widehat{\theta}_{mv}$ e $\widehat{\theta}_{mom}$ è $N\left(\theta, \frac{2\theta}{n}\right)$.

2. La probabilità richiesta è

   $$\mathrm{P}\left(X_1 = 0; \frac{\theta}{2}\right) = \frac{e^{-\frac{\theta}{2}}(\frac{\theta}{2})^0}{0!} = e^{-\frac{\theta}{2}} = g(\theta).$$

   Per la proprietà di equivarianza si ha

   $$\widehat{g(\theta)} = g(\widehat{\theta}_{mv}) = e^{-\frac{\widehat{\theta}_{mv}}{2}} = e^{-\overline{X}_n}.$$

   Poiché $g'(\theta) = -\frac{1}{2}e^{-\frac{\theta}{2}}$ e $I(\theta) = \frac{n}{2\theta}$, un intervallo di confidenza approssimato di livello 0.90 per $g(\theta)$ è

   $$g(\widehat{\theta}_{mv}) \pm z_{1-\alpha/2}\sqrt{(g'(\widehat{\theta}_{mv}))^2 I^{-1}(\widehat{\theta}_{mv})} = e^{-\overline{X}_n} \pm z_{0.95}\frac{e^{-\overline{X}_n}}{2}\sqrt{\frac{4\overline{X}_n}{n}}$$

   che in corrispondenza di un campione osservato con $n = 20$ e $\overline{x}_n = 36$ è

   $$e^{-36} \pm 1.65e^{-36}\sqrt{\frac{36}{20}}.$$

3. Sfruttando la distribuzione asintotica dello stimatore di massima verosimiglianza ricavata al punto 1., i.e. $\widehat{\theta}_{mv} \approx N\left(\theta, \frac{2\theta}{n}\right)$, è immediato ricavare l'intervallo approssimato di livello $1-\alpha$ per θ

   $$\left[\widehat{\theta}_{mv} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{2\widehat{\theta}_{mv}}{n}}, \widehat{\theta}_{mv} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{2\widehat{\theta}_{mv}}{n}}\right] = \left[2\overline{X}_n - z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\overline{X}_n}{n}}, 2\overline{X}_n + z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\overline{X}_n}{n}}\right].$$

   La lunghezza aleatoria dell'intervallo è:

   $$\mathcal{L}(\mathbf{X}_n) = 2z_{1-\frac{\alpha}{2}}\sqrt{\frac{4\overline{X}_n}{n}},$$

da cui

$$\mathcal{L}(\mathbf{X}_n)^2 = 4z^2_{1-\frac{\alpha}{2}} \frac{4\overline{X}_n}{n}.$$

Quindi si ha

$$\mathbb{E}(\mathcal{L}(\mathbf{X}_n)^2) = 4z^2_{1-\frac{\alpha}{2}} \frac{4\mathbb{E}(\overline{X}_n)}{n} = 4z^2_{1-\frac{\alpha}{2}} \frac{4\theta}{2n} = 8z^2_{1-\frac{\alpha}{2}} \frac{\theta}{n} < \ell_0 \Longleftrightarrow$$

$$\Longleftrightarrow n > \frac{8\theta z^2_{1-\alpha/2}}{\ell_0} = \frac{8(1.96)^2}{0.5} = 61.4 \Longrightarrow n = 62$$

Esercizio 2. Sia $X_1, \ldots, X_n$ un campione casuale proveniente da una popolazione con funzione di densità

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

1. Si consideri il sistema di ipotesi

$$H_0 : \theta = \theta_0 = 4, \qquad H_1 : \theta = \theta_1 = 3.$$

   Determinare, per $n = 1$, il rapporto delle verosimiglianze e verificare che il test di Neyman-Pearson ha la seguente regione di accettazione:

$$A = \{x_1 \in (0,1) : \ x_1 < k\},$$

   dove k è un generico valore positivo. Determinare la probabilità di errore di I specie assumendo $k = 3$.

2. Verificare che, per $n > 1$, 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 un generico valore positivo e dove $\widehat{\theta}_{mv}(\mathbf{x}_n) = \frac{n}{\ln \prod_{i=1}^n x_i}$.

3. Verificare che la distribuzione asintotica di $\widehat{\theta}_{mv}$

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

   e quindi che la regione di accettazione del test con probabilità di errore di I specie pari a $\alpha = 0.05$ è:

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

   Stabilire se, con per un campione osservato di dimensione $n = 30$ in cui $(\ln \prod_{i=1}^{30} x_i) = 11$ l'ipotesi nulla viene o meno accettata.

Svolgimento:

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

$$\frac{L(\theta_0)}{L(\theta_1)} = \frac{\theta_0 x^{-\theta_0}}{\theta_1 x^{-\theta_1}} = \frac{\theta_0}{\theta_1} x^{-(\theta_0 - \theta_1)} = \frac{4}{3} x^{-1}.$$

   Poiché tale rapporto è funzione decrescente di x, in base al Lemma di Neyman-Pearson la regione di accettazione è

$$A = \{x_1 \in (0,1) : \ x_1 < k\},$$

   dove k è un generico valore positivo. La probabilità dell'errore di I specie è

$$\alpha = P(RR; \theta = \theta_0) = P(X_1 > 3; \theta_0) = 1 - \int_1^3 \theta_0 x^{-(\theta_0+1)} = 1 - \Big[-x^{\theta_0}\Big]_1^3 = 1 + \frac{1}{3^{\theta_0}} - 1 = \frac{1}{3^4} = 0.012$$

2. Per $n > 1$ il rapporto delle verosimiglianze è

$$\frac{L(\theta_0)}{L(\theta_1)} = \frac{\theta_0^n \left[\prod_{i=1}^n x_i\right]^{-\theta_0}}{\theta_1^n \left[\prod_{i=1}^n x_i\right]^{-\theta_1}} = \left(\frac{\theta_0}{\theta_1}\right)^n \left[\prod_{i=1}^n x_i\right]^{-(\theta_0-\theta_1)} = \left(\frac{4}{3}\right)^n \left[\prod_{i=1}^n x_i\right]^{-1}.$$

Poiché tale rapporto è funzione decrescente di $\prod_{i=1}^n x_i$ e quindi funzione crescente di $\widehat{\theta}_{mv} = \frac{n}{\ln \prod_{i=1}^n x_i}$, in base al Lemma di Neyman-Pearson la regione di accettazione è

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

3. Poiché $I(\theta) = \frac{n}{\theta^2}$, la distribuzione asintotica di $\widehat{\theta}_{mv}$ è

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

Per determinare k_α poniamo

$$\alpha = P(RR; \theta = \theta_0) = P(\widehat{\theta}_{mv} < k_\alpha; \theta_0) = P\left(\frac{\widehat{\theta}_{mv} - \theta_0}{\sqrt{\theta_0^2/n}} < \frac{k_\alpha - \theta_0}{\sqrt{\theta_0^2/n}}\right) = \Phi\left(\frac{k_\alpha - \theta_0}{\theta_0/\sqrt{n}}\right)$$

$$\Longrightarrow z_\alpha = \frac{k_\alpha - \theta_0}{\theta_0/\sqrt{n}} \Longrightarrow k_\alpha = \theta_0 + z_\alpha \frac{\theta_0}{\sqrt{n}} = 4 - 1.65 \cdot \frac{4}{\sqrt{n}}.$$

Per un campione osservato di dimensione $n = 30$ in cui $(\ln \prod_{i=1}^{30} x_i) = 11$, si ha

$$\widehat{\theta}_{mv} = \frac{30}{11} = 2.727 < k_\alpha = 4 - 1.65 \cdot \frac{4}{\sqrt{30}} = 5.20,$$

pertanto si rifiuta l'ipotesi nulla.