Corsi di Inferenza Statistica – II prova scritta - 3 Luglio 2009.

Corsi di Studi A - E - D - G - T

ATTENZIONE: I FOGLI A QUADRETTI CONSEGNATI
NON VERRANNO CONSIDERATI

Cognome e Nome: ______________________ **Corso di Laurea:** __________

Esercizio 1. Si consideri un campione casuale di n osservazioni da una popolazione X con funzione di densità:

$$f_X(x;\theta) = \frac{x^2}{2\,\theta^3}\exp\{-x/\theta\} \qquad \theta > 0, \qquad x > 0.$$

Per questa variabile aleatoria si ha che $\mathbb{E}[X] = 3\,\theta$ e $\mathbb{V}[X] = 3\,\theta^2$.

1. Scrivere il modello statistico probabilistico per il campione casuale $\mathbf{X}_n$, verificare che $f_X(x;\theta)$ appartiene alla famiglia esponenziale e ricavare una statistica sufficiente per θ.
2. Scrivere la funzione di verosimiglianza. Ottenere lo stimatore $\hat{\theta}_{MLE}$ di massima verosimiglianza di θ e lo stimatore di massima verosimiglianza di $\psi = \theta + \theta^2$, funzione del parametro θ.
3. Calcolare il momento primo e secondo della variabile aleatoria $\bar{X}_n$, media campionaria.
4. Verificare che $\hat{\psi}_{MLE}$ è uno stimatore distorto di ψ, calcolarne la distorsione e stabilire se è asintoticamente non distorto.

Svolgimento.
1. Il modello statistico-probabilistico è:

$$\left(\mathcal{X}^n = (\mathbb{R}_+)^n,\;\; f_n(\mathbf{x};\theta) = \frac{\prod x_i}{2^n\theta^3}\exp\{-\sum x_i/\theta\}\;\theta^n,\;\; \Theta = \mathbb{R}_+\right)$$

La distribuzione delle popolazione X appartiene alla famiglia esponenziale e può essere scritta come:

$$f_X(x;\theta) = \frac{x^2}{2}\exp\{-x/\theta - 3\log 2\theta\},$$

con: $h(x) = \frac{x^2}{2}$, $\psi(\theta) = -1/\theta$, $T(x) = x$, $B(\theta) = -3\log 2\theta$. Si ottiene quindi che una statistica sufficiente è data da $T(\mathbf{x}_{oss}) = \sum_{i=1}^n T(x_i) = \sum_{i=1}^n x_i$.
2. Le funzioni di verosimiglianza, di log-verosimiglianza e le derivate prima e seconda della funzione di log-verosimiglianza sono:

$$L_{\mathbf{x}_{oss}}(\theta) = \frac{1}{\theta^{3n}}e^{-\sum x_i/\theta}; \qquad \ell(\theta) = -3n\log\theta - \frac{\sum x_i}{\theta}; \qquad \ell'(\theta) = -\frac{3n}{\theta} + \frac{\sum x_i}{\theta^2}; \qquad \ell''(\theta) = \frac{3n}{\theta^2} - \frac{2\sum x_i}{\theta^3}.$$

Da $\ell'(\theta) = 0$, si ottiene $-3n\theta + \sum x_i = 0$ da cui $\hat{\theta}_{MLE} = \bar{X}/3$ è lo stimatore di massima verosimiglianza. Infatti

$$\ell''(\hat{\theta}_{MLE}) = 9(3n/\bar{x}^2) - 27(2n/\bar{x}^2) = -27n/\bar{x}^2 < 0.$$

Per la proprietà di invarianza, lo stimatore di massima verosimiglianza per ψ è:

$$\hat{\psi}_{MLE} = \hat{\theta}_{MLE} + \hat{\theta}^2_{MLE} = \frac{\bar{X}}{3} + \frac{\bar{X}^2}{9}.$$

3. I momenti primo e secondo della media campionaria sono:

$$\mathbb{E}[\,\bar{X}\,]= \frac{1}{n}\sum_{i=1}^{n} X_i = 3\,\theta;\;\; \mathbb{E}[\,\bar{X}^2\,]= \mathbb{V}[\,\bar{X}\,] + \mathbb{E}\big([\,\bar{X}\,]\big)^2 = \frac{1}{n}\,\mathbb{V}[X] + 9\,\theta^2 = \frac{1}{n}3\,\theta^2 + 9\,\theta^2 = 3\,\theta^2\frac{1+3n}{n}.$$

4. La media dello stimatore $\hat{\psi}_{MLE}$ è:

$$\mathbb{E}[\,\hat{\psi}_{MLE}\,]= \mathbb{E}[\hat{\theta}_{MLE} + \hat{\theta}^2_{MLE}\,]= \mathbb{E}\left[\frac{\bar{X}}{3} + \frac{\bar{X}^2}{9}\right]= \frac{1}{3}\,\mathbb{E}[\,\bar{X}\,] + \frac{1}{9}\,\mathbb{E}[\,\bar{X}^2\,] = \theta + \frac{1+3n}{3n}\,\theta^2 \neq \theta + \theta^2$$

Quindi lo stimatore $\hat{\psi}_{MLE}$ è distorto. La sua distorsione è data da:

$$D(\hat{\psi}_{MLE}) = \mathbb{E}[\,\hat{\psi}\,] - \theta - \theta^2 = \theta^2/3n,$$

e questo implica che lo stimatore è asintoticamente corretto perché la sua distorsione tende a 0 al crescere di n.

Esercizio 2. Dato un campione casuale di n osservazioni estratto da una popolazione X con funzione di densità:

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

lo stimatore di massima verosimiglianza per θ è $\hat{\theta}_{MLE} = -\dfrac{n}{\sum_{i=1}^n \log(1-X_i)}$.

1. Si utilizzi il test di Neyman-Pearson per il confronto tra le ipotesi

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta = \theta_1 \qquad (\theta_0 < \theta_1)$$

e si mostri che la regione di accettazione è

$$RA = \{\mathbf{X} \in \mathcal{X}^n : \hat{\theta}_{\text{MLE}} < K\}. \tag{1}$$

2. Verificare che asintoticamente si ha $\hat{\theta}_{MLE} \approx N(\theta, \frac{\theta^2}{n})$.

3. Ottenere l'espressione dell'intervallo di confidenza approssimato per θ di livello 0.80 e calcolarne gli estremi supponendo di avere osservato un campione di $n = 225$ unità, per il quale $\hat{\theta}_{MLE} = 1.5$.

4. Si consideri il campione osservato di cui al punto precedente. Si vogliono confrontare le ipotesi:

$$H_0 : \theta_0 = 1.3 \qquad H_1 : \theta_1 = 1.6.$$

Calcolare il valore approssimato della potenza del test (1), ottenuto ponendo $K = 1.464$.

Svolgimento.

1. Per il test di Neyman-Pearson, la regione di accettazione è: $RA = \{\mathbf{x} \in \mathcal{X}^n : \lambda(\mathbf{x}) > k\}$, dove $\lambda(\mathbf{x}) = f_n(\mathbf{x}; \theta = \theta_0)/f_n(\mathbf{x}; \theta = \theta_1) = L_{\mathbf{x}}(\theta_0)/L_{\mathbf{x}}(\theta_1)$, e dove k è tale che $\mathbb{P}(\mathbf{X} \in RR|H_o) = \alpha$, con α fissato. Dato che

$$\lambda(\mathbf{x}) = \frac{\theta_0^n\ \left[\prod(1-x_i)\right]^{\theta_0-1}}{\theta_1^n\ \left[\prod(1-x_i)\right]^{\theta_1-1}} = \left(\frac{\theta_0}{\theta_1}\right)^n \left[\prod_{i=1}^n (1-x_i)\right]^{\theta_0-\theta_1}$$

si ha che $\lambda(\mathbf{x}) > k$ quando:

$n(\log\theta_0 - \log\theta_1) + (\theta_0-\theta_1)\sum_{i=1}^n \log(1-x_i) > \log k$ e, dato che $\theta_0 - \theta_1 < 0$, si ottiene

$\sum_{i=1}^n \log(1-x_i) < \dfrac{\log k - n(\log\theta_0 - \log\theta_1)}{\theta_0-\theta_1}$, da cui: $\dfrac{n}{\sum_{i=1}^n \log(1-x_i)} > \dfrac{n(\theta_0-\theta_1)}{\log k - n(\log\theta_0 - \log\theta_1)}$.

Quindi $RA = \{\mathbf{X} \in \mathcal{X}^n : \hat{\theta}_{\text{MLE}} < K\}$, con $K = -\dfrac{n(\theta_0-\theta_1)}{\log k - n(\log\theta_0-\log\theta_1)}$.

2. $\hat{\theta}_{MLE} \approx N(\theta, I^{-1}(\theta))$, dove $I(\theta) = n\mathbb{E}\big[-\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta)\big]$ è l'informazione attesa di Fisher. Dato che $\log f_X(x;\theta) = \log\theta + (\theta-1)\log(1-x)$ e $\frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta) = -\theta^{-2}$, si ottiene che $I^{-1}(\theta) = \frac{\theta^2}{n}$ e quindi $\hat{\theta}_{MLE} \approx N(\theta, \frac{\theta^2}{n})$.

3. L'intervallo di confidenza approssimato di livello $1-\alpha$ si ottiene utilizzando la distribuzione asintotica di $\hat{\theta}_{MLE}$ dove il valore di θ nell'espressione della varianza è sostituito da $\hat{\theta}_{MLE}$. Nel caso di distribuzione campionaria $N(\theta, V^2)$, per uno stimatore $T(\mathbf{X})$, l'intervallo di confidenza è dato da:

$$IC = (T(\mathbf{x}_{oss}) - z_{1-\alpha/2}\, V\ ,\ T(\mathbf{x}_{oss}) + z_{1-\alpha/2}\, V).$$

In questo caso $T(\mathbf{X}) = \hat{\theta}_{MLE}$ con distribuzione campionaria asintotica è $N(\theta, \hat{\theta}^2_{MLE}/n)$ ed il valore di $z_{\alpha/2} = z_{.9} = 1.28$. Quindi l'intervallo di confidenza approssimato è:

$$\widetilde{IC} \approx \big(\hat{\theta}_{MLE} - z_{.9}\, (\hat{\theta}_{MLE}/\sqrt{n})\, , \hat{\theta}_{MLE} - z_{.9}\, (\hat{\theta}_{MLE}/\sqrt{n})\big) = \big(1.5 - 1.28(0.1)\ ,\ 1.5 + 1.28(0.1)\big) = (1.372\ ,\ 1.628).$$

4. Il valore approssimato della potenza del test è:

$$\begin{aligned}
1-\beta &= \mathbb{P}(RR; \theta_1) = \mathbb{P}(\hat{\theta}_{MLE} > K; \theta_1) = \mathbb{P}\left(\frac{\hat{\theta}_{MLE} - \theta_1}{\hat{\theta}_{MLE}/\sqrt{n}} > \frac{K - \theta_1}{\hat{\theta}_{MLE}/\sqrt{n}}\right) \\
&\approx \mathbb{P}\left(Z > \frac{1.464 - 1.6}{0.1}\right) = \mathbb{P}(Z > -1.36) = 1 - \Phi(-1.36) = 0.913.
\end{aligned}$$