Corsi di Inferenza Statistica – Prova scritta - 11 Settembre 2009.

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

ATTENZIONE: I FOGLI A QUADRETTI CONSEGNATI <u>NON</u> VERRANNO CONSIDERATI

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

Esercizio 1 Sia $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ un campione casuale di n osservazioni estratte da una popolazione X con distribuzione di probabilità:

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

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

1. Determinare la funzione di verosimiglianza $L_{\mathbf{x}_{oss}}(\theta)$ ed ottenere lo stimatore di massima verosimiglianza $\hat{\theta}_{MLE}$
2. Verificare che l'espressione dell'informazione osservata di Fisher è $\dfrac{n^3}{(\sum_{i=1}^n \log x_i)^2}$. Scrivere l'approssimazione normale della funzione di verosimiglianza e spiegare perché utilizzando questa approssimazione si può ricavare un intervallo di verosimiglianza approssimato ad un livello fissato q.

   Parte facoltativa. Scrivere gli estremi dell'intervallo approssimato.
3. Studiare la correttezza e la consistenza dello stimatore $\hat{\theta}_{MLE}$.
4. Controllare se $\hat{\theta}_{MLE}$ è uno stimatore UMVUE.

Svolgimento.
1. La funzione di verosimiglianza è:

$$L_{\mathbf{x}_{oss}}(\theta) = \prod_{i=1}^n \theta^{-1} x_i^{-\frac{1}{\theta}-1} = \theta^{-n}\left(\prod_{i=1}^n x_i\right)^{-\frac{1}{\theta}-1}.$$

La funzione di log-verosimiglianza e le sue derivate sono:
$\ell(\theta) = -n\log\theta + (-\frac{1}{\theta}-1)\sum_{i=1}^n \log x_i \qquad \ell'(\theta) = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n \log x_i \qquad \ell''(\theta) = \frac{n}{\theta^2} - \frac{2\sum_{i=1}^n \log x_i}{\theta^3}.$
Si ottiene quindi: $\ell'(\theta) = 0 \Rightarrow \hat{\theta}_{MLE} = \frac{\sum_{i=1}^n \log x_i}{n}$, infatti $\ell''(\hat{\theta}_{MLE}) = \frac{n}{\hat{\theta}^2_{MLE}} - \frac{2\sum_{i=1}^n \log x_i}{\hat{\theta}^3_{MLE}} = -\frac{n^3}{(\sum_{i=1}^n \log x_i)^2} < 0.$

2. L'informazione osservata di Fisher è $I(\hat{\theta}_{MLE}) = -\ell''(\hat{\theta}_{MLE}) = \frac{n^3}{(\sum_{i=1}^n \log x_i)^2}$. Nell'intorno di $\hat{\theta}$ si può approssimare la funzione di verosimiglianza di un qualsiasi modello statistico con una distribuzione normale di media $\hat{\theta}$ e di varianza $I^{-1}(\hat{\theta})$ e l'intervallo di verosiniglianza approssimato di livello q si ottiene come l'intervallo di verosimiglianza di livello q di una distribuzione normale.

Parte facoltativa. L'intervallo di verosimiglianza di livello q della distribuzione normale $L_{\mathbf{x}_{oss}}(\theta) \approx N(\hat{\theta}_{MLE}, I^{-1}(\hat{\theta}_{MLE}))$ che approssima $L_{\mathbf{x}_{oss}}(\theta)$ è dato da: $\left(\hat{\theta}_{MLE} \pm \sqrt{I^{-1}(\hat{\theta}_{MLE})} \ \sqrt{-2\ \log q}\right)$ quindi, sostituendo i valori trovati per $\hat{\theta}_{MLE}$ e $I(\hat{\theta}_{MLE})$, si ottiene che: $I_q \approx \left(\frac{\sum_{i=1}^n \log x_i}{n} \pm \frac{(\sum_{i=1}^n \log x_i)^2}{n^3}\sqrt{-2\ \log q}\right)$.

3. La media dello stimatore di massima verosimiglianza è

$$\mathbb{E}[\hat{\theta}_{MLE}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \log X_i\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\log X_i\right] = \frac{1}{n}\sum_{i=1}^n \theta = \theta.$$

E quindi $\hat{\theta}_{MLE}$ è uno stimatore corretto. Si ha che

$$MSE(\hat{\theta}_{MLE}) = \mathbb{V}[\hat{\theta}_{MLE}] = \mathbb{V}[\frac{1}{n}\sum_{i=1}^n \log X_i] = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\left[\log X_i\right] = \frac{1}{n^2}\sum_{i=1}^n \theta^2 = \frac{\theta^2}{n}.$$

Ciò significa che $\hat{\theta}_{MLE}$ è uno stimatore consistente perché per $n \to \infty$, si ha $MSE(\hat{\theta}_{MLE}) \to 0$.

4. Il limite inferiore di Cramer Rao per la varianza di uno stimatore corretto è dato dall'inverso dell'informazione attesa di Fisher. L'informazione attesa di Fisher risulta essere:

$$\begin{aligned} I(\theta) &= -n\,\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\left[\log f_X(x;\theta)\right]\right] = -n\,\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2}\left[-\log\theta - \left(\frac{1}{\theta}+1)\right)\log x\right]\right\} = \\ & -n\,\mathbb{E}\left[\frac{1}{\theta^2} - \frac{2\log x}{\theta^3}\right] = -n\left[\frac{1}{\theta^2} - \frac{2\theta}{\theta^3}\right] = \frac{n}{\theta^2}. \end{aligned}$$

L'informazione attesa di Fisher è pari all'inverso della varianza di $\hat{\theta}_{MLE}$ che risulta quindi essere uno stimatore UMVUE.

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

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

1. Verificare che la distribuzione campionaria asintotica di $\hat{\theta}_{MLE}$ è $\hat{\theta}_{MLE} \approx N\left(\theta, \dfrac{(\theta+2)^2}{n}\right)$.

   Parte facoltativa. È necessario conoscere esplicitamente l'espressione di $\hat{\theta}_{MLE}$? Spiegare perchè.

2. Ottenere un intervallo di confidenza approssimato di livello 0.95 per θ assumendo che in un campione osservato di numerosità $n = 144$, il valore della stima di massima verosimiglianza ottenuto sia $\hat{\theta}_{MLE} = 1.6$.

3. Per il confronto tra le ipotesi

$$H_0 : \theta = \theta_0 \text{ vs } H_1 : \theta < \theta_0,$$

   si consideri il test con la seguente regione di rifiuto

$$RR = \{\mathbf{x} \in \mathcal{X}^n : \hat{\theta}_{MLE} < k\}$$

   (test di Wald). Utilizzando l'approssimazione normale per $\hat{\theta}_{\text{MLE}}$, si verifichi che la regione di rifiuto per un test con probabilità di errore di prima specie pari ad α si ottiene ponendo $k = z_\alpha \dfrac{\hat{\theta}_{MLE}+2}{\sqrt{n}} + \theta_0$, dove z_α indica il percentile di livello α della distribuzione normale standardizzata.

4. Per il campione osservato della domanda **2**, si esegua il test per $\theta_0 = 1.5$ a livello $\alpha = 0.05$.

Svolgimento.

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

$$\log f_X(x;\theta) = \log(\theta+2) + (\theta+1)\log x\ , \qquad \text{e che} \qquad \frac{\partial^2}{\partial\theta^2}\log f_X(x;\theta) = -\frac{1}{(\theta+2)^2},$$

si ha $I(\theta) = \dfrac{n}{(\theta+2)^2}$, quindi $\hat{\theta}_{MLE} \approx N\left(\theta, \dfrac{(\theta+2)^2}{n}\right)$.

Parte facoltativa. No, non è necessario. Perchè l'approssimazione campionaria asintotica dello stimatore di massima verosimiglianza non dipende dalla sua espressione matematica. Dipende solo dall'espressione matematica dell'informazione attesa di Fisher, il cui inverso è la varianza asintotica.

2. Per costruire un intervallo di confidenza approssimato per $\hat{\theta}_{MLE}$ si deve usare la distribuzione campionaria asintotica ottenuta al punto precedente, dove il valore di θ nell'espressione della varianza è sostituito da $\hat{\theta}_{MLE}$.

L'intervallo è dato da

$$\widetilde{IC}_{(1-\alpha)} = \left(\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \frac{\hat{\theta}+2}{\sqrt{n}}\right) = \left(1.6 \pm 1.96\ \frac{3.6}{12}\right) = (1.6 \pm 0.588) = (1.012\ ,\ 2.188)$$

3. Dato che, come al punto **2**, la distribuzione campionaria asintotica di $\hat{\theta}$ è: $\hat{\theta} \approx N\left(\theta, \frac{(\hat{\theta}+2)^2}{n}\right)$, si ha:

$$\mathbb{P}\left\{RR|\theta=\theta_0\right\} = \mathbb{P}\{\hat{\theta} < k|\theta=\theta_0\} = \mathbb{P}\left\{\frac{\hat{\theta}-\theta_0}{(\hat{\theta}+2)/\sqrt{n}} < \frac{k-\theta_0}{(\hat{\theta}+2)/\sqrt{n}}\right\} \approx \Phi\left(\frac{k-\theta_0}{(\hat{\theta}+2)/\sqrt{n}}\right) = \alpha.$$

Da cui si ottiene che $\frac{k-\theta_0}{(\hat{\theta}+2)/\sqrt{n}} = z_\alpha$ e $k = z_\alpha \frac{\hat{\theta}+2}{\sqrt{n}} + \theta_0$

4. Dato che $\hat{\theta} = 1.6$ e che $z_\alpha = z_{0.05} = -1.64$, si ottiene che $k = -1.64\frac{3.6}{12} + 1.5 = -1.64(.3) + 1.5 = -0.49 + 1.5 = 1.01$ si ha dunque che $\hat{\theta} = 1.6 > k = 1.01$, il valore di $\hat{\theta}$ non appartiene alla regione di rifiuto e si accetta H_0.