Inferenza Statistica
I ESONERO A.A. 2011-2012 – 23 aprile 2012

ATTENZIONE: I FOGLI A QUADRETTI CONSEGNATI
NON VERRANNO CONSIDERATI

◇ ◇ ◇

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

Esercizio 1. Si consideri la funzione

$$f_X(x;\theta) = a\frac{x}{\theta^2}\, I_{[0,3\theta]}(x), \qquad \theta > 0.$$

1. Determinare il valore della costante a tale che la funzione $f_X(\cdot;\theta)$ sia la funzione di densità della variabile aleatoria X.

2. Verificare che valore atteso e varianza di X sono rispettivamente uguali a :

$$\mathrm{E}[X] = 2\theta, \qquad \mathrm{V}[X] = \frac{1}{2}\theta^2.$$

3. Considerato un campione casuale di n osservazioni $X_1, \ldots, X_n$ con funzione di densità $f_X(\cdot, \theta)$, determinare la funzione di verosimiglianza di θ, una statistica sufficiente unidimensionale e la stima di massima verosimiglianza ($\widehat{\theta}_{mv}$) di θ. Tracciare il grafico della funzione di verosimiglianza.

4. Determinare lo stimatore dei momenti ($\widehat{\theta}_{mom}$) di θ, e verificare che la sua distribuzione asintotica risulta essere:
$$N\left(\theta, \frac{1}{8n}\theta^2\right).$$

5. Utilizzando l'approssimazione normale del punto precedente, determinare l'espressione della seguente probabilità:
$$\mathrm{P}(|\widehat{\theta}_{mom} - \theta| > \epsilon), \qquad \epsilon > 0.$$

   Verificare che, per n che tende ad infinito, tale probabilità tende a zero, per ogni valore di ϵ e per ogni valore finito di θ. Calcolare quindi il valore della probabilità considerata assumendo $\epsilon = 1$, $\theta = 9$ e $n = 10$.

6. Stabilire se lo stimatore dei momenti ha varianza coincidente con il limite inferiore di Cramer Rao.

Svolgimento:

1. Affinché $f_X(\cdot;\theta)$ sia una densità si deve imporre la condizione

$$\int_{\mathcal{X}} f_X(x;\theta)dx = 1$$

   e determinare la costante a. Si ha quindi

$$\int_0^{3\theta} a\frac{x}{\theta^2}dx = \frac{a}{\theta^2}\frac{x^2}{2}|_0^{3\theta} = \frac{9}{2}a = 1$$

   da cui segue che $a = \frac{2}{9}$.

2. Calcoliamo il valore atteso, il momento secondo e la varianza:

$$\mathrm{E}[X] = \int_0^{3\theta} x\cdot\frac{2}{9}\frac{x}{\theta^2}dx = \frac{2}{9\theta^2}\frac{x^3}{3}|_0^{3\theta} = 2\theta$$

$$\mathrm{E}[X^2] = \int_0^{3\theta} x^2\cdot\frac{2}{9}\frac{x}{\theta^2}dx = \frac{2}{9\theta^2}\frac{x^4}{4}|_0^{3\theta} = \frac{9}{2}\theta^2$$

$$\mathrm{V}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \frac{9}{2}\theta^2 - 4\theta^2 = \frac{1}{2}\theta^2$$

3. La funzione di densità congiunta è:

$$f_n(\mathbf{x}_n;\theta) = \prod_{i=1}^n\left[\frac{2}{9}\frac{x_i}{\theta^2}I_{[0,3\theta]}(x_i)\right] = \left(\frac{2}{9\theta^2}\right)^n\prod_{i=1}^n x_i\prod_{i=1}^n I_{[0,3\theta]}(x_i) = \left(\frac{2}{9\theta^2}\right)^n\prod_{i=1}^n x_i I_{\left[\frac{x_{(n)}}{3},\infty\right)}(\theta).$$

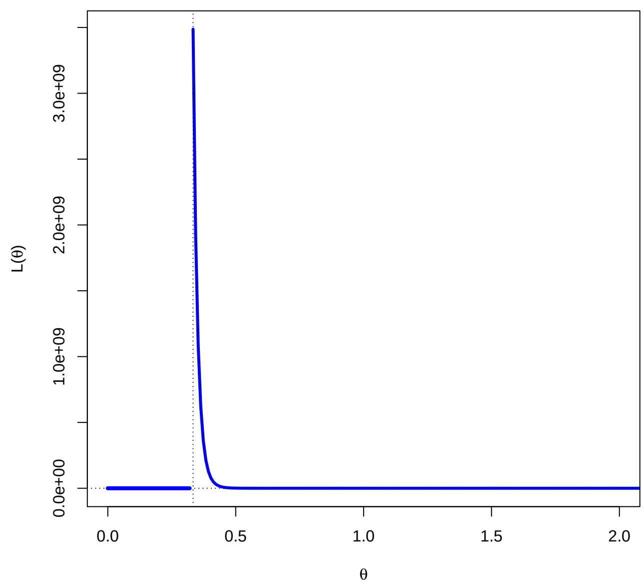
   Pertanto la funzione di verosimiglianza è:

$$L(\theta) = \theta^{-2n}I_{\left[\frac{x_{(n)}}{3},\infty\right)}(\theta).$$

   In base al criterio di fattorizzazione, una statistica sufficiente è: $T(\mathbf{X}) = X_{(n)}$. Per determinare la stima di massima verosimiglianza basta osservare che $L(\theta)$ è funzione decrescente di θ nell'insieme $\left[\frac{x_{(n)}}{3},\infty\right)$ e che quindi assume il suo massimo nel punto $\frac{x_{(n)}}{3}$.

   Lo stimatore di massima verosimiglianza è $\widehat{\theta}_{mv} = \frac{X_{(n)}}{3}$.

La funzione di verosimiglianza può essere rappresentata dal seguente grafico (ad esempio fissando $n = 10$ e $x_{(n)} = 1$):



4. Per determinare lo stimatore dei momenti poniamo:
$$\mathrm{E}[X] = 2\theta = \overline{X}_n \Longrightarrow \widehat{\theta}_{mom} = \frac{1}{2}\overline{X}_n.$$
Essendo funzione della media campionaria lo stimatore dei momenti si distribuisce asintoticamente come una normale, ovvero $\widehat{\theta}_{mom} \approx N\left(\mathrm{E}(\widehat{\theta}_{mom}), \mathrm{V}(\widehat{\theta}_{mom})\right)$. Poiché
$$\mathrm{E}(\widehat{\theta}_{mom}) = \mathrm{E}\left(\frac{1}{2}\overline{X}_n\right) = \frac{1}{2}2\theta = \theta$$
e
$$\mathrm{V}(\widehat{\theta}_{mom}) = \mathrm{V}\left(\frac{1}{2}\overline{X}_n\right) = \frac{1}{4}\frac{1}{n}\mathrm{V}(X) = \frac{1}{4}\frac{1}{2n}\theta^2 = \frac{\theta^2}{8n}$$
si ha che
$$\widehat{\theta}_{mom} \approx N\left(\theta, \frac{\theta^2}{8n}\right).$$

5. Utilizzando l'approssimazione normale del punto precedente, determiniamo la seguente probabilità:
$$\begin{aligned}
\mathrm{P}(|\widehat{\theta}_{mom} - \theta| > \epsilon) &= 1 - \mathrm{P}(\theta - \epsilon < \widehat{\theta}_{mom} < \theta + \epsilon) = \\
&= 1 - \mathrm{P}\left(\frac{\theta - \epsilon - \theta}{\sqrt{\theta^2/8n}} < \frac{\widehat{\theta}_{mom} - \theta}{\sqrt{\theta^2/8n}} < \frac{\theta + \epsilon - \theta}{\sqrt{\theta^2/8n}}\right) = \\
&= 1 - \Phi\left(\frac{\epsilon\sqrt{8n}}{\theta}\right) + \Phi\left(\frac{-\epsilon\sqrt{8n}}{\theta}\right)
\end{aligned}$$

Questa probabilità tende a 0 per n che tende ad infinito per ogni valore di ϵ e per ogni valore finito di θ, infatti $\Phi\left(\frac{\epsilon\sqrt{8n}}{\theta}\right) \to 1$ e $\Phi\left(\frac{-\epsilon\sqrt{15n}}{\theta}\right) \to 0$.

Assumendo $\epsilon = 1$, $\theta = 9$ e $n = 10$ si ha

$$\mathrm{P}(|\widehat{\theta}_{mom} - \theta| > \epsilon) = 1 - \Phi\left(\frac{\epsilon\sqrt{8n}}{\theta}\right) + \Phi\left(\frac{-\epsilon\sqrt{8n}}{\theta}\right) = 1 - \Phi(0.99) + \Phi(-0.99) = 1 - 0.84 + 0.16 = 0.32$$

6. Il modello non è regolare (il supporto dipende da θ), quindi non si può l limite inferiore di Cramer Rao.

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

$$f_X(x;\theta) = \left(\frac{2-\theta}{2}\right)^x \left(\frac{\theta}{2}\right)^{1-x}, \qquad x = \{0,1\}, \qquad \theta \in [0,2].$$

1. Determinare il modello statistico e stabilire se $\{f_X(\cdot;\theta),\ \theta \in \Theta\}$ una famiglia esponenziale.
2. Determinare la funzione di verosimiglianza di θ e una statistica sufficiente per il modello. Calcolare la stima di massima verosimiglianza di θ e l'informazione osservata di Fisher per il campione osservato $\mathbf{x}_3 = (1,0,1)$.
3. Per un campione di dimensione $n = 100$, determinare gli estremi dell'insieme di verosimiglianza approssimato di livello $q = 0.146$ assumendo che la stima di verosimiglianza del parametro sia pari a $1/2$ e l'informazione osservata pari a 25.
4. Determinare l'errore quadratico medio dello stimatore di massima verosimiglianza e studiarne la consistenza.

Svolgimento:

1. Il modello statistico è $\left\{\mathcal{X}^n = \{0,1\}^n; f_n(\mathbf{x}_n;\theta) = \left(\frac{2-\theta}{2}\right)^{\sum_{i=1}^n x_i} \left(\frac{\theta}{2}\right)^{n-\sum_{i=1}^n x_i}; \Theta = [0,2]\right\}$.

   $\{f_X(\cdot;\theta),\ \theta \in \Theta\}$ appartiene alla famiglia esponenziale, in particolare è un modello Bernoulliano di parametro $\left(\frac{2-\theta}{2}\right)$.

2. La funzione di verosimiglianza è:

$$L(\theta) = \left(\frac{2-\theta}{2}\right)^{\sum_{i=1}^n x_i} \left(\frac{\theta}{2}\right)^{n-\sum_{i=1}^n x_i}$$

   Una statistica sufficiente per θ è $T(\mathbf{X}) = \sum_{i=1}^n X_i$.

   Per determinare la stima di massima verosimiglianza, scriviamo la log-verosimiglianza

$$l(\theta) = \sum_{i=1}^n x_i \log\left(\frac{2-\theta}{2}\right) + (n - \sum_{i=1}^n x_i)\log\left(\frac{\theta}{2}\right),$$

   calcoliamo la derivata prima rispetto a θ e la eguagliamo a 0

$$l'(\theta) = \frac{-\sum_{i=1}^n x_i}{2-\theta} + \frac{(n - \sum_{i=1}^n x_i)}{\theta} = \Longleftrightarrow \theta = 2 - 2\overline{X}_n.$$

   Dopo aver controllato che la derivata seconda

$$l''(\theta) = \frac{-\sum_{i=1}^n x_i}{(2-\theta)^2} - \frac{(n - \sum_{i=1}^n x_i)}{\theta^2}$$

   è negativa per ogni valore di θ, si può concludere che $\widehat{\theta}_{mv} = 2 - 2\overline{X}_n$

   In corrispondenza del campione osservato $\mathbf{x}_3 = (1,0,1)$ la stima di massima verosimiglianza è $\widehat{\theta}_{mv} = 2 - 2\overline{X}_n = 2 - 2 \cdot \frac{2}{3} = \frac{2}{3}$ e l'informazione osservata di Fisher è

$$I_{x_{oss}}((\widehat{\theta})) = -l''(\widehat{\theta}) = \frac{n\overline{X}_n}{(2\overline{X}_n)^2} + \frac{n - n\overline{X}_n}{(2-2\overline{X}_n)^2} = \frac{n}{4\overline{X}_n(1-\overline{X}_n)} = \frac{3}{4 \cdot \frac{2}{3} \cdot \frac{1}{3}} = 3.37$$

3. L'insieme di verosimiglianza approssimato di livello $q = 0.146$, per $n = 100, \widehat{\theta}_{mv} = \frac{1}{2}, I_{x_{oss}} = 25$ è:
$$\widehat{\theta}_{mv} \pm \sqrt{-2 \log q I^{-1}_{x_{oss}}(\widehat{\theta}_{mv})} = \frac{1}{2} \pm 1.96/\sqrt{25} = [0.108; 0.892]$$

4. Lo stimatore di massima verosimiglianza è $\widehat{\theta}_{mv} = 2 - 2\overline{X}_n$. Calcoliamo innanzi tutto il valore atteso, ricordando che in un modello Bernoulliano di parametro p si ha $\mathrm{E}(X) = p$,
$$\mathrm{E}(\widehat{\theta}_{mv}) = 2 - 2\mathrm{E}(\overline{X}_n) = 2 - 2\mathrm{E}(X) = 2 - 2 \cdot \frac{2-\theta}{2} = \theta.$$
Poiché lo stimatore $\widehat{\theta}_{mv}$ è dunque non distorto, il suo errore quadratico medio è uguale alla varianza, ovvero
$$\mathrm{MSE}(\widehat{\theta}_{mv}) = \mathrm{V}(\widehat{\theta}_{mv}) = \mathrm{V}(2 - 2\overline{X}_n) = 4\mathrm{V}(\overline{X}_n) = \frac{4}{n}\mathrm{V}(X) = \frac{4}{n}\left(\frac{2-\theta}{2}\right)\left(\frac{\theta}{2}\right) = \frac{(2-\theta)\theta}{n}.$$
Osservando che $\mathrm{MSE}(\widehat{\theta}_{mv}) = \frac{(2-\theta)\theta}{n} \to 0$ per $n \to \infty$ si può concludere che lo stimatore è consistente.