Corsi di Inferenza Statistica – Primo Esonero. 9 Aprile 2010

Tutti i Canali.

Cognome e Nome: ______________________ **Canale:** __________

Esercizio 1

Si consideri un campione casuale di n osservazioni da una popolazione X con legge di probabilità:

$$f_X(x;\theta) = \mathbb{P}\big(X = x;\theta\big) = (x+1)(1-\theta)^2\,\theta^x \qquad 0 \le \theta \le 1 \qquad x = 0, 1, \ldots$$

1. Per un generico campione osservato $\mathbf{x}_{oss}$ si ottenga la stima di massima verosimiglianza $\widehat{\theta}_{MLE}$ per il parametro θ, verificando che il valore che azzera la derivata prima è effettivamente un massimo.
2. Si determini il valore di $\widehat{\theta}_{MLE}$ quando $\mathbf{x}_{oss} = (2, 5, 8, 10, 0, 20, 2, 1)$.
3. Dato il campione osservato al punto 2. stabilire quale valore è preferibile tra $\theta_1 = \dfrac{7}{10}$ e $\theta_2 = \dfrac{8}{10}$.
4. Assumendo che $\theta = \widehat{\theta}_{MLE}$, si calcoli la probabiltà dei seguenti eventi: $X = 0$, $X = 1$. Si calcoli infine la probabilità di osservare il campione di 3 elementi $\mathbf{X}_3 = (0, 1, 0)$.

Svolgimento

1. La funzione di verosimiglianza è:
   $L_{\mathbf{x}_{oss}}(\theta) = \prod_{i=1}^n \big[\,(x_i+1)\,(1-\theta)^2\,\theta^{x_i}\big] = \big[\,\prod_{i=1}^n\,(x_i+1)\,\big](1-\theta)^{2n}\,\theta^{\sum x_i}$.
   La funzione di log-verosimiglianza è:
   $\ell(\theta) = \sum_{i=1}^n \log(x_i+1) + 2n\log(1-\theta) + \sum_{i=1}^n x_i \log\theta$, con derivata prima
   $\ell'(\theta) = -\dfrac{2n}{1-\theta} + \dfrac{\sum_{i=1}^n x_i}{\theta}$, che si azzera per $-2\,n\,\theta + (1-\theta)\sum_{i=1}^n x_i = 0$ e cioè per $\widehat{\theta} = \dfrac{\sum_{i=1}^n x_i}{2n + \sum_{i=1}^n x_i}$. La derivata seconda della funzione di log-verosimiglianza è data da:
   $$\ell''(\theta) = -\frac{2n}{(1-\theta)^2} - \frac{\sum_{i=1}^n x_i}{\theta^2} < 0$$
   per ogni valore di θ, quindi $\widehat{\theta}$ è stima di massima verosimiglilanza.
2. Dato che $n = 8$, $\sum_{i=1}^n = 48$ si ha $\widehat{\theta}_{MLE} = \dfrac{48}{16+48} = \dfrac{48}{64} = \dfrac{3}{4}$

3.

$$\frac{L_{\mathbf{x}_{oss}}(\theta_1)}{L_{\mathbf{x}_{oss}}(\theta_2)} = \frac{\left(\frac{3}{10}\right)^{16}\left(\frac{7}{10}\right)^{48}}{\left(\frac{2}{10}\right)^{16}\left(\frac{8}{10}\right)^{48}} = 1.08 > 1.$$

Il valore $\theta_1 = \dfrac{7}{10}$ è il più verosimile.

4.

$$\mathbb{P}(X = 0; \theta = \widehat{\theta}) = (1 - \widehat{\theta})^2 = \left(\frac{1}{4}\right)^2 \qquad \mathbb{P}(X = 1; \theta = \widehat{\theta}) = 2(1 - \widehat{\theta})^2\, \widehat{\theta} = 2\left(\frac{1}{4}\right)^2 \frac{3}{4}.$$

e

$$\mathbb{P}\big(\,\mathbf{X}_3 = (0, 1, 0)\,\big) = 2\left(\frac{1}{4}\right)^6 \frac{3}{4}.$$

Esercizio 2. Si consideri un campione casuale proveniente da una popolazione con legge di probabilità:

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

1. Scrivere il modello statistico per il campione casuale $\mathbf{X} = (X_1, \dots, X_n)$.
2. Verificare se il modello appartiene alla famiglia esponenziale ed identificare, se esiste, una statistica sufficiente.
3. Determinare la funzione di verosimiglianza, il suo nucleo e l'espressione della stima di massima verosimiglianza $\widehat{\theta}_{MLE}$ per un generico campione osservato $\mathbf{x}_{oss}$.
4. Ottenere l'espressione dell'informazione osservata di Fisher $\mathcal{I}_{\mathbf{x}_{oss}}(\widehat{\theta}_{MLE})$.
5. In corrispondenza ad un campione osservato di numerosità $n = 50$ e $\sum \log(1+x_i) = 106.6$ calcolare il valore di $\widehat{\theta}_{MLE}$, di $\mathcal{I}_{\mathbf{x}_{oss}}(\widehat{\theta}_{MLE})$ e scrivere l'approssimazione normale per la funzione di verosimiglianza.

Svolgimento

1. $\left(\mathcal{X}^n = (\mathbb{R}_+)^n, f_n(\mathbf{x}:\theta) = \prod_{i=1}^n \left[\,\theta(1+x_i)^{-(1+\theta)}\,\right] = \theta^n \prod_{i=1}^n (1+x_i)^{-(1+\theta)}, \Theta = \mathbb{R}_+\right)$
2. Dato che
$$f(x;\theta) = \theta(1+x)^{-(1+\theta)} = \frac{1}{(1+x)} \exp\left\{\log\theta - \theta\log(1+x)\right\}$$
Ponendo $h(x) = (1+x)^{-1}$, $\eta(\theta) = -\theta$, $T(x) = \log(x+1)$ e $B(\theta) = \log\theta$, si verifica che la legge di probabilità della variabile di base appartiene alla famiglia esponenziale. Una statistica sufficiente è data da $\sum_{i=1}^n T(x_i) = \sum \log(1+x_i)$
3. La funzione di verosimiglianza è data da:
$$L_{\mathbf{x}_{oss}}(\theta) = \frac{\theta^n}{\prod_{i=1}^n (1+x_i)} \prod_{i=1}^n (1+x_i)^{-\theta}.$$
Il suo nucleo è la funzione $g(T(\mathbf{x}_{oss}),\theta) = \theta^n \prod_{i=1}^n (1+x_i)^{-\theta}$. La funzione di log verosimiglianza e la sua derivata prima sono: $\ell(\theta) = n\log\theta - \theta\sum_{i=1}^n \log(1+x_i)$ e $\ell'(\theta) = \frac{n}{\theta} - \sum_{i=1}^n \log(1+x_i)$. La derivata della funzione di log verosimiglianza si annulla per $\frac{n}{\theta} - \sum_{i=1}^n \log(1+x_i) = 0$ da questa equazione si ottiene $\widehat{\theta} = \frac{n}{\sum_{i=1}^n \log(1+x_i)}$, che è la stima di massima verosimiglianza, dato che la derivata seconda delle funzione di log verosimiglianza $\ell''(\theta) = -n/\theta^2 < 0$ per tutti i valori di θ.
4. L'informazione osservata di Fisher è data da:
$$\mathcal{I}_{\mathbf{x}_{oss}}(\widehat{\theta}) = -\ell''(\theta)\big|_{\theta=\widehat{\theta}} = \frac{\left[\,\sum_{i=1}^n \log(1+x_i)\right]^2}{n}$$

5. In corrispondenza al campione osservato si ottiene $\widehat{\theta}_{MLE} = \dfrac{50}{106.6} = 0.47$ e $\mathcal{I}_{\mathbf{x}_{oss}}(\widehat{\theta}) = \dfrac{(106.6)^2}{50} = 227.27$. Quindi

$$L_{\mathbf{x}_{oss}}(\theta) \approx \frac{\sqrt{227.27}}{\sqrt{2\,\pi}} \exp\left\{-\frac{227.27}{2}(\theta - 0.47)^2\right\}.$$

Corsi di Inferenza Statistica – Primo Esonero. 9 Aprile 2010.

*

Tutti i Canali

Cognome e Nome: ______________________________ **Canale:** __________

Esercizio 1 Si consideri un campione casuale di n osservazioni da una popolazione X con legge di probabilità:

$$f_X(x;\theta) = \mathbb{P}(X = x;\theta) = \frac{1}{2}(x+1)(x+2)\,\theta^3\,(1-\theta)^{\ x} \quad 0 \le \theta \le 1 \qquad x = 0, 1, \ldots$$

1. Per un generico campione osservato $\mathbf{x}_{oss}$ si ottenga la stima di massima verosimiglianza $\widehat{\theta}_{MLE}$ per il parametro θ.
2. Si determini il valore di $\widehat{\theta}_{MLE}$ quando $\mathbf{x}_{oss} = (3, 1, 0, 1, 2, 2)$.
3. Dato il campione osservato al punto 2. stabilire quale valore è preferibile tra $\theta_1 = \dfrac{23}{30}$ e $\theta_2 = \dfrac{17}{30}$.
4. Assumendo che $\theta = \widehat{\theta}_{MLE}$, si calcoli la probabiltà dei seguenti eventi: $X = 0$, $X = 1$. Si calcoli infine la probabilità di osservare il campione di 3 elementi $\mathbf{x} = (0, 1, 0)$.

Svolgimento

1. La funzione di verosimiglianza è:

$$L_{\mathbf{x}_{oss}}(\theta) = \prod_{i=1}^{n}\left[\frac{1}{2}\,(x_i+1)\,(x_i+2)\,\theta^3\,(1-\theta)^{x_i}\right] = \left[\prod_{i=1}^{n}(x_i+1)\,(x_i+2)\right]2^{-n}\,\theta^{3n}\,(1-\theta)^{\sum x_i}.$$

La funzione di log-verosimiglianza è:

$$\ell(\theta) = \sum_{i=1}^{n}\log(x_i+1) + \sum_{i=1}^{n}\log(x_i+2) - n\log 2 + 3n\log\theta + \sum_{i=1}^{n}x_i\log(1-\theta),$$

con derivata prima $\ell'(\theta) = \dfrac{3n}{\theta} - \dfrac{\sum_{i=1}^n x_i}{(1-\theta)}$, che si azzera per

$3\,n\,(1-\theta) - \theta\sum_{i=1}^n x_i = 0$ e cioè per $\widehat{\theta} = \dfrac{3n}{3n + \sum_{i=1}^n x_i}$. La derivata seconda della funzione di log-verosimiglianza è data da:

$$\ell''(\theta) = -\frac{3n}{\theta^2} - \frac{\sum_{i=1}^n x_i}{(1-\theta)^2} < 0$$

per ogni valore di θ, quindi $\widehat{\theta}$ è stima di massima verosimiglilanza.

2. Dato che $n = 6$, $\sum_{i=1}^{n} = 9$ si ha $\widehat{\theta}_{MLE} = \dfrac{18}{18+9} = \dfrac{18}{27} = \dfrac{2}{3}$

3.
$$\frac{L_{\mathbf{x}_{oss}}(\theta_1)}{L_{\mathbf{x}_{oss}}(\theta_2)} = \frac{\left(\frac{23}{30}\right)^{18}\left(\frac{7}{30}\right)^{9}}{\left(\frac{17}{30}\right)^{18}\left(\frac{13}{30}\right)^{9}} = 0.877 < 1.$$

Il valore $\theta_2 = \dfrac{17}{30}$ è il più verosimile.

4.
$$\mathbb{P}(X = 0; \theta = \widehat{\theta}) = \widehat{\theta}^{\,3} = \left(\frac{1}{3}\right)^{3} \qquad \mathbb{P}(X = 1; \theta = \widehat{\theta}) = 3\,\widehat{\theta}^{\,3}\,(1 - \widehat{\theta}) = \left(\frac{1}{3}\right)^{2}\frac{2}{3}.$$

e
$$\mathbb{P}\big(\,\mathbf{X}_3 = (0,1,0)\,\big) = \left(\frac{1}{3}\right)^{8}\frac{2}{3}.$$

Esercizio 2. Si consideri un campione casuale proveniente da una popolazione con legge di probabilità:

$$f_X(x;\theta) = \frac{2\,x}{\theta^2}\;e^{-\dfrac{x^2}{\theta^2}} \qquad\qquad x > 0, \quad \theta > 0$$

1. Scrivere il modello statistico per il campione casuale $\mathbf{X} = (X_1, \ldots, X_n)$.
2. Verificare se il modello appartiene alla famiglia esponenziale ed identificare, se esiste, una statistica sufficiente.
3. Determinare la funzione di verosimiglianza, il suo nucleo e l'espressione della stima di massima verosimiglianza $\widehat{\theta}_{MLE}$ per un generico campione osservato $\mathbf{x}_{oss}$.
4. Ottenere l'espressione dell'informazione osservata di Fisher $\mathcal{I}_{\mathbf{x}_{oss}}(\widehat{\theta}_{MLE})$.
5. In corrispondenza ad un campione osservato di numerosità $n = 100$ e $\sqrt{\sum_{i=1}^n x_i^2} = 29.01$ calcolare il valore di $\widehat{\theta}_{MLE}$, di $\mathcal{I}_{\mathbf{x}_{oss}}(\widehat{\theta}_{MLE})$ e scrivere l'approssimazione normale per la funzione di verosimiglianza.

Svolgimento

1.

$$\left(\mathcal{X}^n = (\mathbb{R}_+)^n, f_n(\mathbf{x}:\theta) = \prod_{i=1}^n \left[\frac{2x_i}{\theta^2}\exp\left\{-\frac{x_i^2}{\theta^2}\right\}\right] = \frac{2^n}{\theta^{2n}}\left[\prod_{i=1}^n x_i\right]\,e^{-\dfrac{\sum_{i=1}^n x_i^2}{\theta^2}}, \Theta = \mathbb{R}_+\right)$$

2. Dato che

$$f(x;\theta) = \frac{2\,x}{\theta^2}\;e^{-\dfrac{x^2}{\theta^2}} = 2\,x\exp\left\{-2\log\theta - \frac{x^2}{\theta^2}\right\}$$

Ponendo $h(x) = 2x$, $\eta(\theta) = -\theta^{-2}$, $T(x) = x^2$ e $B(\theta) = -2\log\theta$, si verifica che la legge di probabilità della variabile di base appartiene alla famiglia esponenziale. Una statistica sufficiente è data da $\sum_{i=1}^n T(x_i) = \sum x_i^2$

3. La funzione di verosimiglianza è data da:

$$L_{\mathbf{x}_{oss}}(\theta) = \left[2^n\,\prod_{i=1}^n x_i\right]\theta^{-2n}\,e^{-\dfrac{\sum_{i=1}^n x_i^2}{\theta^2}}.$$

Il suo nucleo è la funzione

$$g(T(\mathbf{x}_{oss}),\theta) = \theta^{-2n}\,e^{-\dfrac{\sum_{i=1}^n x_i^2}{\theta^2}}.$$

La funzione di log verosimiglianza e la sua derivata prima sono:

$$\ell(\theta) = -2n \log \theta - \frac{\sum_{i=1}^n x_i^2}{\theta^2} \qquad \text{e} \qquad \ell'(\theta) = -\frac{2n}{\theta} + \frac{2\sum_{i=1}^n x_i^2}{\theta^3}.$$

La derivata della funzione di log verosimiglianza si annulla per $-n + \frac{\sum_{i=1}^n x_i^2}{\theta^2} = 0$. Da questa equazione si ottiene $\widehat{\theta} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}$, che è la stima di massima verosimiglianza, dato che la derivata seconda delle funzione di log verosimiglianza calcolata in $\theta = \widehat{\theta}$ è $\ell''(\theta)\big|_{\theta=\widehat{\theta}} = \frac{2n}{\theta^2} - \frac{6}{\theta^4}\sum_{i=1}^n x_i^2\big|_{\theta=\widehat{\theta}} = \frac{2n\frac{\sum x_i^2}{n} - 6\sum x_i^2}{(\sum x_i^2/n)^4} = \frac{-4\sum x_i^2}{(\sum x_i^2/n)^4} < 0$.

4. L'informazione osservata di Fisher è data da:

$$\mathcal{I}_{\mathbf{x}_{oss}}(\widehat{\theta}) = -\ell''(\theta)\big|_{\theta=\widehat{\theta}} = \frac{4n^4}{(\sum x_i^2)^3}$$

5. In corrispondenza al campione osservato si ottiene $\widehat{\theta}_{MLE} = \frac{29.01}{10} = 2.9$ e $\mathcal{I}_{\mathbf{x}_{oss}}(\widehat{\theta}) = \frac{4 \cdot 10^4}{(29.01^2)^3} = 0.67$. Quindi

$$L_{\mathbf{x}_{oss}}(\theta) \approx \frac{\sqrt{0.67}}{\sqrt{2\,\pi}} \exp\left\{-\frac{0.67}{2}(\theta - 2.9)^2\right\}.$$