

ESAME DI STATISTICA MATEMATICA A.A. 2017-18 (18.09.2018)

ESERCIZIO 1. Consideriamo un campione casuale associato alla densità di probabilità

$$f(x; \theta) = c(\theta)x^{-3}\mathbb{1}(x \geq \theta),$$

dove il parametro θ è una costante positiva, e $c(\theta)$ è la costante di normalizzazione (da determinare).

- (a) Determinare lo stimatore di θ con il metodo dei momenti.
- (b) Determinare lo stimatore di massima verosimiglianza di θ e lo stimatore di massima verosimiglianza di $\log(\theta)$.
- (c) Supponendo ora che θ sia uniformemente distribuito nell'intervallo $[1, 2]$, determinare lo stimatore bayesiano di θ .

ESERCIZIO 2. Si modella approssimativamente il tempo necessario per percorrere l'intero grande raccordo anulare come $60+X$ minuti, dove X è una variabile aleatoria esponenziale di parametro 20. Immaginiamo che una persona faccia 3 volte il giro del grande raccordo anulare. Assumendo che i tempi di percorrenza dei vari giri siano variabili aleatorie indipendenti, determinare:

- (a) la funzione generatrice dei momenti del tempo totale impiegato per fare 3 volte il giro del grande raccordo anulare,
- (b) la varianza del tempo totale impiegato per fare 3 volte il giro del grande raccordo anulare.

ESERCIZIO 3. Consideriamo un campione casuale $X_1, \dots, X_n$ di variabili uniformemente distribuite in $[0, \theta]$. Dire se la variabile aleatoria $e^{\bar{X}_n}$ presenta delle fluttuazioni gaussiane se opportunamente riscalata e centrata, giustificando la risposta, ed in caso affermativo calcolarle.

ESERCIZIO 4. I valori di un campione gaussiano di media e varianza non note e di ampiezza 5 sono

10, 12, 15, 9, 10

Determinare un intervallo di confidenza al 95% per la varianza.

ESERCIZIO 5. Una ditta di hamburger afferma che i propri hamburger hanno una percentuale di grasso minore del 20%. La percentuale di grasso in 9 hamburger campionati è la seguente:

18, 23, 19, 24, 18.5, 17, 26, 25, 23.

Prendendo l'affermazione della ditta come ipotesi nulla e approssimando la percentuale di grasso negli hamburger con delle variabili gaussiane di deviazione standard 2, dire se l'ipotesi va accettata o rifiutata con un test avente livello di significatività del 5%.

SOLUZIONI

ESERCIZIO 1. Siccome $\int_{\theta}^{\infty} x^{-3} dx = \frac{1}{2\theta^2}$, deve valere $c(\theta) = 2\theta^2$.

(a) Il valore medio di f è dato da

$$c(\theta) \int_{\theta}^{\infty} x^{-2} dx = \frac{c(\theta)}{\theta} = 2\theta.$$

Richiedo quindi che $\bar{X}_n = 2\hat{\theta}$. Ottengo che lo stimatore col metodo dei momenti è dato da $\hat{\theta} = \frac{\bar{X}_n}{2}$.

(b) La funzione di likelihood è data da

$$L_n(x_1, \dots, x_n; \theta) = c(\theta)^n (x_1 \cdot x_2 \cdots x_n)^{-3} \mathbb{1}(x_1, \dots, x_n \geq \theta).$$

Quindi, considerati $x_1, \dots, x_n$ come numeri positive fissati, la suddetta funzione si massimizza per $\theta = \min\{x_1, \dots, x_n\}$. Lo stimatore di massima verosimiglianza è quindi $\hat{\theta} = \min\{X_1, X_2, \dots, X_n\}$. Siccome $\log$ è monotona crescente su $(0, \infty)$ abbiamo

$$\widehat{\log \theta} = \log(\hat{\theta}) = \log(\min\{X_1, X_2, \dots, X_n\}).$$

(c) La distribuzione a priori di θ è data da

$$f(\theta | x_1, \dots, x_n) = \frac{c(\theta)^n (x_1 \cdot x_2 \cdots x_n)^{-3} \mathbb{1}(x_1, \dots, x_n \geq \theta) \mathbb{1}(\theta \in [1, 2])}{\int_1^2 c(y)^n (x_1 \cdot x_2 \cdots x_n)^{-3} \mathbb{1}(x_1, \dots, x_n \geq y) dy}$$

Chiamo $m_n = \min(x_1, \dots, x_n)$. Allora

$$f(\theta | x_1, \dots, x_n) = \frac{\theta^{2n} \mathbb{1}(1 \leq \theta \leq m_n \wedge 2)}{\int_1^2 y^{2n} \mathbb{1}(y \leq m_n) dy} = \theta^{2n} \mathbb{1}(\theta \leq m_n \wedge 2) \frac{2n+1}{(2 \wedge m_n)^{2n+1} - 1}.$$

Ne deriva che

$$\int \theta f(\theta | x_1, \dots, x_n) d\theta = \frac{2n+1}{(2 \wedge m_n)^{2n+1} - 1} \int_1^{2 \wedge m_n} \theta^{2n+1} d\theta = \frac{2n+1}{2n+2} \frac{(2 \wedge m_n)^{2n+2} - 1}{(2 \wedge m_n)^{2n+1} - 1}$$

Lo stimatore bayesiano è quindi dato da

$$\hat{\theta} = \frac{2n+1}{2n+2} \frac{(2 \wedge m_n)^{2n+2} - 1}{(2 \wedge m_n)^{2n+1} - 1} \quad m_n := \min(X_1, \dots, X_n). \quad (1)$$

ESERCIZIO 2.

Il tempo impiegato è dato da $Z = 180 + X_1 + X_2 + X_3$, dove X_1, X_2, X_3 sono i.i.d. esponenziali di parametro 20. Allora $U := X_1 + X_2 + X_3$ è v.a di tipo Gamma($\alpha = 3, \lambda = 20$).

(a)

$$M_Z(t) = E[e^{Zt}] = e^{180t} E[e^{Ut}] = \begin{cases} e^{180t} \left(\frac{20}{20-t}\right)^3 & t < 20, \\ +\infty & \text{altrimenti} \end{cases}$$

Nota: se uno non ricorda la funzione generatrice dei momenti della Gamma($\alpha = 3, \lambda = 20$) poteva semplicemente fare il cubo (per indipendenza) della funzione generatrice dei momenti della v.a. esponenziale.

(b) $Var(Z) = Var(X_1 + X_2 + X_3) = Var(X_1) + Var(X_2) + Var(X_3) = 3/400$ (la seconda identità segue dall'indipendenza).

ESERCIZIO 3.

Vi sono fluttuazioni gaussiane. Per il CLT $\sqrt{n}(\bar{X}_n - \mu) \rightarrow \mathcal{N}(0, \sigma^2)$. Ho $\mu = \frac{\theta}{2}$. $\mathbb{E}(X_1^2) = \frac{1}{\theta} \int_0^\theta x^2 dx = \theta^2/3$. Quindi

$$\sigma^2 = \frac{1}{3}\theta^2 - \frac{\theta^2}{4} = \frac{\theta^2}{12}.$$

Quindi

$$\sqrt{n}(\bar{X}_n - \frac{\theta}{2}) \rightarrow \mathcal{N}(0, \frac{\theta^2}{12}).$$

Posto $f(x) = e^x$, per il metodo delta

$$\sqrt{n}(f(\bar{X}_n) - f(\theta/2)) \rightarrow \mathcal{N}\left(0, \frac{\theta^2}{12} f'(\theta/2)^2\right)$$

cioè

$$\sqrt{n}(e^{\bar{X}_n} - e^{\theta/2}) \rightarrow \mathcal{N}\left(0, \frac{\theta^2}{12} e^\theta\right)$$

ESERCIZIO 4.

Chiamata s la deviazione campionaria, $n = 5$ e $\alpha = 0.05$, l'intervallo cercato è dato da

$$I = \left(\frac{(n-1)s^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2, n-1}^2} \right).$$

Abbiamo $s^2 = 5.7$, $\chi_{\alpha/2, n-1}^2 = \chi_{0.025, 4}^2 = 11.143$, $\chi_{1-\alpha/2, n-1}^2 = \chi_{0.975, 4}^2 = 0.484$. Quindi

$$I = \left(\frac{4 * 5.7}{11.143}, \frac{4 * 5.7}{0.484} \right) = (2.046.., 47, 107..).$$

ESERCIZIO 5. Posto $n = 9$, $\alpha = 0.05$, $\sigma = 2$, $\mu_0 = 20$ il test è il seguente: rifiuto l'ipotesi nulla se e solo se $\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > z_\alpha$.

Siccome $z_\alpha = z_{0.05} \approx 1.645$ rifiuto H_0 se e solo se $\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} > 1.645$. Ho $\bar{X}_n = 21.5$, quindi

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{21.5 - 20}{2/\sqrt{9}} = 2.25$$

Ne deriva che rifiuto l'ipotesi nulla.