Politecnico di Milano – Facoltà di Ingegneria Informatica

Anno Accademico 2009/2010

Corso di Statistica (2L) per INF e TEL

Docente: Antonio Pievatolo; esercitazioni: Raffaele Argiento

Esercitazione del 26/03/10

Esercizio 1

Tema d'Esame del 30/06/2005

Una grandezza incognita μ è stata misurata più volte usando due strumenti con precisioni diverse e note. Possiamo schematizzare la situazione dicendo che abbiamo due campioni casuali indipendenti $X_1, X_2, \ldots, X_m$ da una popolazione normale $N(\mu, \sigma_X^2)$ ed $Y_1, Y_2, \ldots, Y_n$ da una popolazione normale $N(\mu, \sigma_Y^2)$; le deviazioni standard σ_X e σ_Y sono note e μ è da stimare. Usiamo, come stimatore di μ, la statistica $T := a\bar{X}_m + (1-a)\bar{Y}_n$, dove a è un opportuno numero reale.

1. Si calcolino la media e la varianza di T.
2. Si trovi a in modo da minimizzare l'errore quadratico medio.
3. Fissato per a il valore trovato al punto precedente qual è la distribuzione di T?

Soluzione

1. $\mathbb{E}(T) = \mathbb{E}(a\bar{X}_m + (1-a)\bar{Y}_n) = \mu$, quindi T è uno stimatore non distorto!
   $\mathrm{Var}(T) = \mathrm{Var}(a\bar{X}_m + (1-a)\bar{Y}_n) = a^2 \frac{\sigma_X^2}{m} + (1-a)^2 \frac{\sigma_Y^2}{n}$.
2. T è non distorto quindi $\mathrm{MSE}_a(T) = \mathrm{Var}(T)$, uno studio di MSE_a in funzione di a mostra che questo è minimo per
$$a = \frac{\frac{\sigma_Y^2}{n}}{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}$$
3. $\bar{X}_m \sim N(\mu, \sigma_X^2/m)$; $\bar{Y}_n \sim N(\mu, \sigma_Y^2/n)$; quindi T è combinazione lineare di variabili aleatorie con distribuzione normale, da ciò (fissato a come al punto precedente) T ha distribuzione normale di media μ e varianza
$$\frac{\frac{\sigma_X^2}{m} \cdot \frac{\sigma_Y^2}{n}}{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}$$

∎

Una nota

Sia
$$f_X(x, \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} x^{\alpha-1} \mathrm{e}^{\frac{x}{\beta}} \mathbb{1}_{(0,+\infty)}(x)$$
la densità di una variabile aleatoria con distribuzione gamma(α, β), $\alpha > 0$ e $\beta > 0$.
Si ha chiaramente $\int_{-\infty}^{\infty} f_X(x, \alpha, \beta) dx = 1$, da cui si ricava:
$$\int_0^{+\infty} x^{\alpha-1} \mathrm{e}^{\frac{x}{\beta}} dx = \Gamma(\alpha)\beta^\alpha \qquad (1)$$

Esercizio 2

Il tempo di risposta di un calcolatore all'input di un terminale si modellizza mediante una v.a. esponenziale di parametro θ incognito. In un esperimento vengono misurati n tempi di risposta $T_1, T_2, \ldots, T_n$

1. Mostrare che
$$\bar{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_i$$
è uno stimatore non distorto di θ.

2. Qual è la legge di $\bar{T}_n$?

3. Si può dire che $\tilde{T}_n = \frac{1}{\bar{T}_n}$ sia uno stimatore non distorto della caratteristica $k(\theta) = \frac{1}{\theta}$?

4. Ricavare da $\tilde{T}_n$ uno stimatore non distorto per $k(\theta)$ e calcolarne l'errore quadratico medio.

Soluzione

1. La media campionaria è uno stimatore non distorto della media di popolazione!

2. Se $T \sim$E(θ) allora $T \sim$gamma$(1, \theta)$. Dato il campione $T_1, \ldots, T_n$ allora $\sum_{i=1}^{n} T_i \sim$gamma(n, θ). In conclusione $\bar{T}_n = \frac{1}{n}\sum_{i=1}^{n} T_i \sim$gamma$(n, \frac{\theta}{n})$

3. Valutiamo la media
$$\begin{aligned}
\mathbb{E}(\tilde{T}_n) &= \mathbb{E}(\frac{1}{\bar{T}_n}) = \int_{-\infty}^{\infty} \frac{1}{t} f_{\bar{T}_n}(t)dt \\
&= \int_0^{\infty} \frac{1}{t}\frac{1}{\Gamma(n)}\frac{1}{(\theta/n)^n} t^{n-1} e^{-\frac{t}{\theta/n}} dt \\
&= \frac{n^n}{\Gamma(n)\theta^n}\int_0^{\infty} t^{n-2} e^{-\frac{nt}{\theta}} dt \\
& \text{applicando la formula (1)} \\
&= \frac{n^n}{\Gamma(n)\theta^n} \times \frac{\Gamma(n-1)\theta^{n-1}}{n^{n-1}} \\
&= \frac{n}{n-1}\frac{1}{\theta}.
\end{aligned}$$
Si ricordi che $\Gamma(\alpha) = (1-\alpha)\Gamma(1-\alpha)$. Si conclude che $\tilde{T}_n$ è uno stimatore distorto, mentre non lo è $H_n = \frac{n-1}{n}\tilde{T}_n$.

4. H_n è uno stimatore non distorto. Si lascia il calcolo del relativo errore quadratico medio come esercizio. Si otterrà:
$$\text{MSE}(H_n) = \text{Var}(H_n) = \frac{1}{n-2}\frac{1}{\theta^2}.$$

∎

Esercizio 1 - Metodo dei momenti - Appello del 10/07/2007

Il reddito mensile di una certa popolazione è una variabile aleatoria continua X con densità di Pareto data da
$$f(x, a, b) = \frac{ab^a}{x^{a+1}}\mathbf{1}_{(b,\infty)}(x), \quad a > 2,\ b > 0$$
I parametri a, b sono entrambi incogniti e per stimarli si analizzano i redditi $X_1, ..., X_n$ di un campione casuale di $n = 100$ individui di questa popolazione, ottenendo un reddito medio campionario pari a $\bar{x} = 1500$€ con varianza campionaria $s^2 = 750000$€^{2}.

1. Calcolate i primi due momenti $\mu_1(a, b) = \mathbb{E}(X)$ e $\mu_2(a, b) = \mathbb{E}(X_2)$ della densità $f(x, a, b)$.

2. Determinate uno stimatore di a e uno di b usando il metodo dei momenti e calcolatene il loro valore sulla base delle realizzazioni campionarie fornite.

Soluzione

1.

$$\begin{aligned}
\mu_1(a,b) &= \mathbb{E}_{a,b}(X) = \int_b^\infty x\frac{ab^a}{x^{a+1}}dx = ab^a\int_b^\infty \frac{1}{x^a}dx = ab^a\left[\frac{x^{-a+1}}{-a+1}\right]_b^\infty = \frac{a}{a-1}b \\
\mu_2(a,b) &= \mathbb{E}_{a,b}(X^2) = \int_b^\infty x^2\frac{ab^a}{x^{a+1}}dx = ab^a\int_b^\infty \frac{1}{x^{a-1}}dx = ab^a\left[\frac{x^{-a+2}}{-a+2}\right]_b^\infty = \frac{a}{a-2}b^2
\end{aligned}$$

Si osservi come i due momenti esistono dato che $a > 2$.

2. Sia $M_1 = \bar{X}$ il momento primo campionario, e $M_2 = \frac{1}{n}\sum_{i=1}^n X_i^2$ il momento secondo campionario. Allora lo stimatore per a e b col metodo dei momenti si ottiene risolvendo il sistema:

$$\begin{cases} \mathbb{E}_{a,b}(X) = M_1 \\ \mathbb{E}_{a,b}(X^2) = M_2 \end{cases} \Rightarrow \begin{cases} \frac{ab}{a-1} = M_1 \\ \frac{ab^2}{a-2} = M_2 \end{cases} \Rightarrow \begin{cases} b = \frac{a-1}{a}M_1 \\ \frac{M_1^2(a-1)^2}{a^2}\frac{a}{a-2} = M_2 \end{cases}.$$

La seconda delle due equazioni del sistema si può semplificare in

$$(M_1^2 - M_2)a^2 - 2a(M_1^2 - M^2) + M_1^2 = 0$$

Tale equazione ammette due soluzioni

$$a = 1 \pm \sqrt{\frac{M_2}{M_2 - M_1^2}}$$

(Si osservi che la quantità sotto il segno di radice è positiva!). Dato il vincolo $a > 2$ si ha una sola soluzione ammissibile e dunque gli stimatori per a e b ottenuti col metodo de momenti sono:

$$\hat{a} = 1 + \sqrt{\frac{M_2}{M_2 - M1^2}}, \qquad \hat{b} = \frac{(\hat{a}-1)}{\hat{a}}M_1.$$

Sostituendo i valori campionari, ricordando che $(M_2 - M_1^2) = \frac{n}{n-1}S^2$, si ottengono le stime

$$\hat{a} = 3.0008, \qquad \hat{b} = 1001.33$$

S^2 varianza media campionaria

∎