Politecnico di Milano – Facoltà di Ingegneria Informatica

Anno Accademico 2008/2009

Corso di Statistica (2L) per INF e TEL

Docente: Antonio Pievatolo; esercitazioni: Raffaele Argiento

Esercitazione del 20/03/09

Esercizio 1

Sia $(X_1, \ldots, X_n)$ un campione da una $\mathbf{U}(0, \theta)$, con $\theta > 0$. Per stimare θ vengono proposti due stimatori:

$$A = a(n)X_{(n)}, \qquad B = 2\bar{X}_n;$$

dove $a(n)$ è tale che A sia non distorto, $X_{(n)} = \max\{X_1, \ldots, X_n\}$ e $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ è la media campionaria.

1. Si determini $a(n)$;
2. Quale statistica fra A e B è preferibile come stimatore di θ? Perché?

Soluzione

Si ricordi che se $X \sim \mathbf{U}(0, \theta)$ allora la sua funzione di ripartizione è:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{\theta} & 0 \le x < \theta \\ 1 & x \ge \theta \end{cases}$$

1. Sia $X_{(n)} = \max\{X_1, \ldots, X_n\}$. Se $x \in \mathbb{R}$,

$$\begin{aligned} F_{X_{(n)}}(x) &= \mathbb{P}(X_{(n)} \le x) = \mathbb{P}(\{X_1, \ldots, X_n\} \le x) = \mathbb{P}(\{X_1 \le x\} \cap \cdots \cap \{X_n \le x\}) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \le x) \quad \text{(per l'indipendenza del campione)} \\ &= \prod_{i=1}^n F_{X_i}(x) = F_X(x)^n. \end{aligned}$$

In conclusione

$$F_{X_{(n)}} = \begin{cases} 0 & x < 0 \\ \frac{x^n}{\theta^n} & 0 \le x < \theta \\ 1 & x \ge \theta. \end{cases}$$

Di conseguenza la densità di $X_{(n)}$ è:

$$f_{X_{(n)}} = \frac{d}{dx}F_{X_{(n)}} = \begin{cases} \frac{nx^{n-1}}{\theta^n}, & 0 < x < \theta \\ 0 & \text{altrimenti} \end{cases}$$

Possiamo dunque calcolare la media:

$$\begin{aligned} \mathbb{E}(X_{(n)}) &= \int_{-\infty}^{+\infty} x \cdot f_{X_{(n)}}(x)dx = \int_0^\theta x \cdot \frac{nx^{n-1}}{\theta^n} \\ &= \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n}{n+1}\theta. \end{aligned}$$

In conclusione, affinché A sia non distorto bisogna che

$$a(n) = \frac{n+1}{n}$$

2. Ricordiamo che se $X \sim U(0,\theta)$ allora $\mathbb{E}(X) = \frac{\theta}{2}$, dato che $\bar{X}_n$ è uno stimatore non distorto di $\mathbb{E}(X)$, $2\bar{X}_n$ è uno stimatore non distorto di θ.

   Per decidere quale fra A e B è "preferibile" bisogna valutare l'errore quadratico medio per ciascuno dei due stimatori. Dato che entrambi sono non distorti basta valutarne la varianza! Per quanto riguarda B si ottiene facilmente:

$$\mathrm{Var}(2\cdot\bar{X}_n) = \frac{4}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\theta^2}{3n}.$$

viene fuori dalla definizione della media campionaria!! X_n ha dentro 1/n che tiro fuori dalla varianza e quindi diventa al quadrato!!

   Per valutare la varianza di A ricordiamo che

$$\mathrm{Var}(A) = \frac{(n+1)^2}{n^2}\mathrm{Var}(X_{(n)}) = \frac{(n+1)^2}{n^2}\left[\mathbb{E}(X_{(n)}^2) + \mathbb{E}^2(X_{(n)})\right].$$

   In particolare:

$$\mathbb{E}(X_{(n)}^2) = \int_{-\infty}^{\infty} x^2 f_{X_{(n)}}(x)dx = \frac{n}{n+2}\theta^2,$$

   ottenendo

$$\mathrm{Var}(A) = \frac{1}{n(n+2)}\theta^2.$$

   Dato che $\mathrm{Var}(A) < \mathrm{Var}(B)$ per ogni $\theta > 0$, possiamo concludere che lo stimatore A è preferibile a B!

∎

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$(n,\frac{\theta}{n})$ il fratto n dipende dal fatto che inizialmente ho 1/n

3. Valutiamo la media

$$\begin{aligned}
\mathbb{E}(\tilde{T}_n) &= \mathbb{E}(\frac{1}{\bar{T}_n}) = \int_{-\infty}^{\infty}\frac{1}{t}f_{\bar{T}_n}(t)dt \\
&= \int_0^{\infty}\frac{1}{t}\frac{1}{\Gamma(n)}\frac{1}{(\theta/n)^n}t^{n-1}e^{-\frac{t}{\theta/n}}dt \\
&= \frac{n^n}{\Gamma(n)\theta^n}\int_0^{\infty}t^{n-2}e^{-\frac{nt}{\theta}}dt \\
&\quad \text{con il cambio di variabile } z = \frac{nt}{\theta} \\
&= \frac{n}{\Gamma(n)\theta}\int_0^{\infty}z^{(n-1)-1}e^{-z}dz \\
&= \frac{n}{n-1}\frac{1}{\theta}.
\end{aligned}$$

doveva essere solo 1/theta per non essere distorto!!

Si conclude che $\tilde{T}_n$ è uno stimatore distorto, mentre non lo è $H_n = \frac{n-1}{n}\tilde{T}_n$. quindi per farlo diventare solo theta modifico in modo intelligento...

4. H_n è uno stimatore non distorto. Si lascia il calcolo del relativo errore quadratico medio come esercizio.

se lo stimatore è distorto è la VARIANZA + BIAS

Se bias = a zero non è distorto