Stimatori corretti, stimatori efficaci e disuguaglianza di Cramer Rao

Lucio Demeio
Dipartimento di Ingegneria Industriale e Scienze Matematiche
Università Politecnica delle Marche

Definizione. Sia $\{X_1, X_2, ..., X_n\}$ un campione di rango n proveniente da una popolazione la cui densità $f(x, \theta)$ dipende da un parametro θ che vogliamo stimare. Siano inoltre $\mu = E[X]$ e $\sigma^2 = Var(X)$ la media e la varianza della popolazione. Si chiama *statistica* una qualunque funzione delle variabili del campione, $h(X_1, X_2, ..., X_n)$. Una qualunque statistica che venga usata per stimare il valore di θ viene detta *stimatore* del parametro θ ed useremo la notazione

$$\widehat{\theta} = h(X_1, X_2, ..., X_n).$$

Notiamo che uno stimatore non è un numero, ma una variabile aleatoria.

Stimatori corretti. Dato che uno stimatore è una variabile aleatoria, sorge naturale la richiesta che la media di uno stimatore debba essere uguale al valore del parametro che si vuole stimare tramite quello stimatore. Arriviamo così alla definizione:

Sia $\widehat{\theta}$ uno stimatore del parametro incognito θ. Diremo che $\widehat{\theta}$ è uno *stimatore corretto* o *non distorto* se

$$E[\,\widehat{\theta}\,] = \theta$$

dove il valore di aspettazione va calcolato con la distribuzione $f(x, \theta)$. Lo stimatore di massima verosimiglianza per la media,

$$\widehat{\mu} = \overline{X} = \frac{X_1 + X_2 + ... + X_n}{n},$$

è uno stimatore corretto. Infatti, se μ è la media della distribuzione $f(x, \mu)$, abbiamo

$$E[\overline{X}] = E\left[\frac{X_1 + X_2 + ... + X_n}{n}\right] = \mu.$$

Il problema della distorsione per lo stimatore di massima verosimiglianza per la varianza è invece più complesso. Sia

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n},$$

tale stimatore, dove con μ si indica il valore vero della media. Abbiamo allora

$$E[\widehat{\sigma^2}] = E\left[\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}\right] = \frac{1}{n} \sum_{i=1}^{n} E[(X_i - \mu)^2] = \sigma^2,$$

che indica come $\widehat{\sigma^2}$ sia uno stimatore corretto. Sia invece

$$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n},$$

dove abbiamo usato lo stimatore della media al posto del suo valore vero. Notiamo preliminarmente che

$$\sum_{i=1}^{n} E\left[X_i\,\overline{X}\right] = \frac{1}{n}\sum_{i,j=1}^{n} E\left[X_i\,X_j\right] = \frac{1}{n}\left\{\sum_{i=1}^{n} E\left[X_i^2\right] + \sum_{i\neq j=1}^{n} E\left[X_i\,X_j\right]\right\} =$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n} (\sigma^2+\mu^2) + \sum_{i\neq j=1}^{n} \mu^2\right\} = \sigma^2+\mu^2+(n-1)\,\mu^2 = \sigma^2 + n\,\mu^2$$

$$E\left[\overline{X}^2\right] = Var\left(\overline{X}\right) + \mu^2 = \frac{\sigma^2}{n} + \mu^2$$

Abbiamo allora

$$E[\widehat{\sigma^2}] = E\left[\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[(X_i-\overline{X})^2\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{E\left[X_i^2\right] + E\left[\overline{X}^2\right] - 2\,E\left[X_i\,\overline{X}\right]\right\} =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\sigma^2+\mu^2+\frac{\sigma^2}{n}+\mu^2\right\} - \frac{2}{n}\,(\sigma^2+n\,\mu^2) =$$

$$= \sigma^2 + 2\,\mu^2 + \frac{\sigma^2}{n} - \frac{2}{n}\,(\sigma^2+n\,\mu^2) = \frac{n-1}{n}\sigma^2,$$

che indica come $\widehat{\sigma^2}$ non sia in tal caso uno stimatore corretto.

Stimatori efficaci. È naturale pensare che, se la varianza di uno stimatore è grande, le stime che si ottengono usando quello stimatore possono con grande probabilità discostarsi di molto dal valore medio. Se la varianza è piccola, invece, i valori saranno con garnde probabilità vicini alla media. Arriviamo così alla definizione:

Siano $\widehat{\theta}_1$ e $\widehat{\theta}_2$ due stimatori non distorti del parametro θ. Se

$$Var(\widehat{\theta}_1) < Var(\widehat{\theta}_2)$$

diremo che lo stimatore $Var(\widehat{\theta}_1)$ è più *efficace* (o più *efficace*) di $\widehat{\theta}_2$. Inoltre, se

$$Var(\widehat{\theta}_1) < Var(\widehat{\theta})$$

per ogni altro stimatore corretto $\widehat{\theta}$, diremo che $Var(\widehat{\theta}_1)$ è lo stimatore *a varianza minima*. Vale a tal proposito il seguente

Teorema. Sia $\{X_1, X_2, ..., X_n\}$ un campione di rango n proveniente da una popolazione di densità $f(x,\theta)$ e sia

$$\widehat{\theta} = h(X_1, X_2, ..., X_n)$$

uno stimatore corretto per θ. La varianza di $\widehat{\theta}$ soddisfa allora la **disuguaglianza di Cramer-Rao**

$$Var(\widehat{\theta}) \geq \left\{ n\, E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial \theta} \right)^2 \right] \right\}^{-1} \equiv C(n,\theta) \tag{1}$$

Dimostrazione. Ricordiamo innanzitutto che, data una variabile X con densità $f(x,\theta)$, abbiamo che

$$E[g(X)] = \int g(x)\, f(x,\theta) dx$$

e quindi, con $g(x) = \partial \ln f(X,\theta)/\partial\theta$,

$$E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial \theta} \right)^2 \right] = \int \left(\frac{\partial \ln f(x,\theta)}{\partial \theta} \right)^2 f(x,\theta) dx.$$

Inoltre,

$$1 = \int f(x_j,\theta)\, dx_j \tag{2}$$

$$\theta = E[\widehat{\theta}] = E[h(X_1, X_2, ..., X_n)] =$$
$$= \int \int ... \int h(x_1, x_2, ..., x_n)\, f(x_1,\theta)\, f(x_2,\theta) ... f(x_n,\theta)\, dx_1\, dx_2 ..., dx_n \tag{3}$$

Derivando le due equazioni rispetto a θ otteniamo:

$$\int \frac{\partial f(x_j,\theta)}{\partial\theta} dx_j = \int \frac{\partial f(x_j,\theta)/\partial\theta}{f(x_j,\theta)} f(x_j,\theta)\, dx_j = \int \frac{\partial \ln f(x_j,\theta)}{\partial\theta} f(x_j,\theta)\, dx_j = 0 \tag{4}$$

$$\int \int ... \int h(x_1, x_2, ..., x_n) \left[\sum_{j=1}^{n} \frac{\partial f(x_j,\theta)/\partial\theta}{f(x_j,\theta)} \right] f(x_1,\theta)\, f(x_2,\theta) ... f(x_n,\theta)\, dx_1\, dx_2 ..., dx_n =$$
$$= \int \int ... \int h(x_1, x_2, ..., x_n) \left[\sum_{j=1}^{n} \frac{\partial \ln f(x_j,\theta)}{\partial\theta} \right] f(x_1,\theta) ... f(x_n,\theta)\, dx_1 ..., dx_n = 1 \tag{5}$$

Introduciamo ora le variabili aleatorie $(Y_1, Y_2, ..., Y_n)$ definite come

$$Y_j = \frac{\partial \ln f(X_j,\theta)}{\partial\theta}.$$

e sia $Y = \sum_{j=1}^{n} Y_j$. Dalla (4) abbiamo che $E[Y] = 0$ mentre dalla (5) abbiamo che $E[\widehat{\theta}\, Y] = 1$; inoltre

$$Var(Y) = E[Y^2] - E[Y]^2 = E[Y^2] = \sum_{i,j=1}^{n} E[Y_i\, Y_j] = \sum_{i=1}^{n} E[Y_i^2] + \sum_{i\neq j} E[Y_i\, Y_j] = n\, E[Y_i^2] =$$
$$n\, E\left[\left(\frac{\partial \ln f(X_j,\theta)}{\partial\theta} \right)^2 \right] = n\, E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial\theta} \right)^2 \right].$$

Consideriamo ora il coefficiente di correlazione ρ tra le variabili Y e $\widehat{\theta}$. Abbiamo

$$\rho^2 = \frac{Cov(\widehat{\theta}, Y)^2}{Var(\widehat{\theta})\, Var(Y)} = \frac{(E[\widehat{\theta}\, Y] - E[\widehat{\theta}]\, E[Y])^2}{Var(\widehat{\theta})\, Var(Y)} = \frac{1}{Var(\widehat{\theta})\, Var(Y)}$$

Ricordando che il coefficiente di correlazione obbedisce sempre alla relazione $-1 \leq \rho \leq 1$, otteniamo che

$$\frac{1}{Var(\widehat{\theta})\, Var(Y)} \leq 1$$

e quindi

$$\frac{1}{Var(\widehat{\theta})} \geq \frac{1}{Var(Y)}$$

da cui segue la tesi. Dalla disuguaglianza di Cramer-Rao segue che non è possibile, a parità di rango del campione, avere uno stimatore la cui varianza sia più piccola di scendere sotto il limite dato dalla (1), detto anche *limite di Cramer-Rao*.

Limite di Cramer-Rao per alcune distribuzioni notevoli.

1. Distribuzione esponenziale. Abbiamo:

$$\begin{aligned}
&f(x,\lambda) = \lambda\, e^{-\lambda x},\ x>0 \qquad \ln f(x,\lambda) = \ln\lambda - \lambda x\\
&\left(\frac{\partial \ln f(x,\lambda)}{\partial\lambda}\right)^2 = \left(\frac{1}{\lambda} - x\right)^2\\
&E\left[\left(\frac{\partial \ln f(X,\lambda)}{\partial\lambda}\right)^2\right] = \int_0^\infty \left(\frac{1}{\lambda} - x\right)^2 \lambda\, e^{-\lambda x}\, dx = \frac{1}{\lambda^2}
\end{aligned}$$

e quindi

$$C(n,\lambda) = \frac{\lambda^2}{n}$$

2. Distribuzione normale. Per la stima di μ abbiamo:

$$\begin{aligned}
&f(x,\mu,\sigma) = \frac{1}{\sqrt{2\,\pi}\,\sigma}\, e^{(x-\mu)^2/(2\,\sigma^2)} \qquad \ln f(x,\mu,\sigma) = -\ln(\sqrt{2\,\pi}\,\sigma) + \frac{(x-\mu)^2}{2\,\sigma^2}\\
&\left(\frac{\partial \ln f(x,\mu,\sigma)}{\partial\mu}\right)^2 = \frac{(x-\mu)^2}{\sigma^4}\\
&E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial\mu}\right)^2\right] = \int_{-\infty}^\infty \frac{(x-\mu)^2}{\sigma^4}\,\frac{1}{\sqrt{2\,\pi}\,\sigma}\, e^{(x-\mu)^2/(2\,\sigma^2)} dx = \frac{1}{\sigma^2}
\end{aligned}$$

da cui, svolgendo l'integrale, si ricava che

$$C(n,\mu,\sigma) = \frac{\sigma^2}{n}$$

Per la stima di σ abbiamo:

$$\begin{aligned}
&f(x,\mu,\sigma) = \frac{1}{\sqrt{2\,\pi}\,\sigma}\, e^{(x-\mu)^2/(2\,\sigma^2)} \qquad \ln f(x,\mu,\sigma) = -\ln(\sqrt{2\,\pi}\,\sigma) + \frac{(x-\mu)^2}{2\,\sigma^2}\\
&\left(\frac{\partial \ln f(x,\mu,\sigma)}{\partial\sigma}\right)^2 = \frac{(x^2 - 2\,x\,\mu + \mu^2 - \sigma^2)^2}{\sigma^6} = \frac{[(x-\mu)^2 - \sigma^2]^2}{\sigma^6}\\
&E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial\sigma}\right)^2\right] = \int_{-\infty}^\infty \frac{[(x-\mu)^2 - \sigma^2]^2}{\sigma^6}\,\frac{1}{\sqrt{2\,\pi}\,\sigma}\, e^{(x-\mu)^2/(2\,\sigma^2)}\, dx = \frac{2}{\sigma^2}
\end{aligned}$$

da cui, svolgendo l'integrale, si ricava che

$$C(n,\sigma) = \frac{\sigma^2}{2\,n}$$

3. Distribuzione di Bernoulli. Abbiamo:

$$\begin{aligned} f(k,p) &= p, \qquad \text{per} \quad k=1 \\ f(k,p) &= 1-p, \qquad \text{per} \quad k=0. \end{aligned}$$

$$\begin{aligned} \ln f(k,p) &= \ln p, \qquad \text{per} \quad k=1 \\ \ln f(k,p) &= \ln(1-p), \qquad \text{per} \quad k=0. \end{aligned}$$

$$\begin{aligned} \left(\frac{\partial \ln f(x,p)}{\partial p}\right)^2 &= \frac{1}{p^2}, \qquad \text{per} \quad k=1 \\ \left(\frac{\partial \ln f(x,p)}{\partial p}\right)^2 &= \frac{1}{(1-p)^2}, \qquad \text{per} \quad k=0 \\ E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial p}\right)^2\right] &= \frac{1}{p^2}\,p + \frac{1}{(1-p)^2}\,(1-p) = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)} \end{aligned}$$

da cui

$$C(n,p) = \frac{p(1-p)}{n}$$

4. Distribuzione di Poisson. Abbiamo:

$$\begin{aligned} f(k,\lambda) &= e^{-\lambda}\,\frac{\lambda^k}{k!},\ k \geq 0 \qquad \ln f(k,\lambda) = -\lambda + k\,\ln\lambda \\ \left(\frac{\partial \ln f(k,\lambda)}{\partial \lambda}\right)^2 &= \left(-1 + \frac{k}{\lambda}\right)^2 \\ E\left[\left(\frac{\partial \ln f(X,\theta)}{\partial \lambda}\right)^2\right] &= \sum_{k=0}^{\infty} \left(-1 + \frac{k}{\lambda}\right)^2 e^{-\lambda}\,\frac{\lambda^k}{k!} = \\ &= \sum_{k=0}^{\infty} \frac{(k-\lambda)^2}{\lambda^2} e^{-\lambda}\,\frac{\lambda^k}{k!} = \frac{1}{\lambda} \end{aligned}$$

e quindi

$$C(n,\lambda) = \frac{\lambda}{n}.$$

Sfida: si invitano gli studenti a calcolare l'ultima sommatoria.