Rivista di Statistica Applicata Vol. 19 n. 2 Giugno 1986

ROSTRA

Rubrica a cura di Benito V. Frosini

La stima dello scarto quadratico medio nel caso normale

Benito V. Frosini, Istituto di Statistica, Università Cattolica del Sacro Cuore di Milano.

1. Introduzione

Nel numero precedente di questa Rivista ci siamo occupati della stima della varianza, in particolare nel caso normale. Salvo eccezione, tale stima non è utilizzata direttamente, non essendo la varianza una misura di variabilità (cioè espressa nella stessa unità di misura del fenomeno cui si riferisce); qualora interessi effettivamente la stima puntuale dello scarto quadratico medio σ, usualmente si ricava una stima di σ per mezzo della radice quadrata della stima già ottenuta per la varianza. È curioso che questo procedimento venga seguito anche da chi si sia precedentemente preoccupato di calcolare una stima mediamente non distorta di σ^2: è noto infatti che una trasformazione non lineare non preserva la non-distorsione media (mentre viene preservata la non-distorsione mediana).

È anche assai strano che sui testi di inferenza non venga generalmente trattata in modo esplicito la stima puntuale di σ, dato che essa ha per gli utilizzatori un interesse enormemente maggiore della stima puntuale della varianza, e non presenta particolari difficoltà dal punto di vista matematico. Pertanto, in questa nota saranno considerati per il caso di una popolazione normale $N(\mu,\sigma^2)$ con entrambi i parametri ignoti (a) lo stimatore di σ con distorsione media nulla e quello con minima perdita media quadratica, e (b) lo stimatore di σ con distorsione mediana nulla e quello con minima perdita media assoluta.

2. Distorsione media e perdita quadratica

Come si è visto nell'articolo già richiamato (Frosini, 1986), lo stimatore mediamente non distorto della varianza non coincide con lo stimatore con minima perdita media quadratica (in una data classe di stimatori); analogamente, lo stimatore con distorsione mediana nulla non coincide con lo stimatore con minima perdita media assoluta (idem). Risultati dello stesso tipo saranno ottenuti per quanto riguarda la stima di σ.

Essendo $(X_1, ..., X_n)$ un campione casuale semplice estratto da una *v.c.* $N(\mu,\sigma^2)$ con entrambi i parametri ignoti, e $\bar{X}$ la sua media aritmetica, si definisce con D la devianza campionaria

$$D = \Sigma_{i=1}^{n} (X_i - \bar{X})^2; \tag{1}$$

è noto che D/σ^2 si distribuisce come una v.c. chi-quadrato con parametro $(n — 1)$, ovvero

$$D \sim \sigma^2\chi^2_{n-1} \tag{2}$$

e quindi, prendendo la radice positiva di D:

$$\sqrt{D} \sim \sigma\, \chi_{n-1}. \tag{3}$$

Dalla (3) risulta naturale considerare stimatori di σ del tipo $c\sqrt{D}$; in questa famiglia di stimatori si trova molto semplicemente quello mediamente non distorto, ponendo

$$E(c\sqrt{D}) = c\,\sigma\, E(\chi_{n-1}) = \sigma$$

da cui

$$c = \{E(\chi_{n-1})\}^{-1}. \tag{4}$$

Indicando con ψ_{n-1} la densità di una *v.c.* χ_{n-1},

$$\psi_{n-1}(x) = 2\{2^{(n-1)/2}\Gamma((n-1)/2)\}^{-1} x^{n-2} exp(-x^2/2) \quad 0<x<\infty$$

e ricordando che

$$\int_0^\infty x^n exp(-r^2x^2)\, dx = \Gamma((n + 1)/2)/(2r^{n+1}),$$

si ricava subito

$$E(\chi_{n-1}) = \int_0^\infty x\psi_{n-1}(x)\, dx = \sqrt{2}\, \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \tag{5}$$

il cui reciproco è il valore di c da sostituire in $c\sqrt{D}$ in modo da ottenere lo stimatore cercato (cfr. Kendall & Stuart, 1967, p. 32 Ex. 17.6; Chapman & Robbins, 1951, p. 584, dove però la media della popolazione normale è assunta nota).

Impiegando le note eguaglianze $\Gamma(n) = (n-1)\cdot\Gamma(n-1)$, $\Gamma(1/2) = \sqrt{\pi}$ e $\Gamma(1) = \Gamma(2) = 1$, la media di χ_{n-1} si calcola semplicemente con le seguenti formule:
per n pari

$$\text{se } n = 2 \qquad E(\chi_{n-1}) = \sqrt{2/\pi}$$

$$\text{se } n \geqslant 4 \qquad E(\chi_{n-1}) = \sqrt{\frac{2}{\pi}}\,\frac{2\cdot 4\cdot \ldots \cdot(n-4)\,(n-2)}{1\cdot 3\cdot \ldots \cdot(n-5)\,(n-3)}\;;$$

per n dispari

$$\text{se } n = 3 \qquad E(\chi_{n-1}) = \sqrt{\pi/2}$$

$$\text{se } n \geqslant 5 \qquad E(\chi_{n-1}) = \sqrt{\frac{\pi}{2}}\,\frac{3\cdot 5\cdot \ldots \cdot(n-4)\,(n-2)}{2\cdot 4\cdot \ldots \cdot(n-5)\,(n-3)}\,.$$

Appare subito evidente la seguente relazione:

$$E(\chi_{n-1}) \cdot \mathrm{E}(\chi_n) = n - 1, \tag{6}$$

che offre la formula ricorsiva

$$E(\chi_n) = (n-1)/E(\chi_{n-1}). \tag{7}$$

Già per valori piccoli di n si può tuttavia ottenere un'ottima approssimazione di $E(\chi_n)$ in base alle seguenti considerazioni. Poiché la media del quadrato di una *v.c.* non degenere è maggiore del quadrato della media (o anche applicando la diseguaglianza di Jensen), da

$$E(\chi_n^2) > \{E(\chi_n)\}^2$$

si ottiene la diseguaglianza

$$E(\chi_n) < \sqrt{n}. \tag{8}$$

La v.c. χ_n^2, per la sua definizione come somma dei quadrati di n *v.c.* normali standard indipendenti, è stocasticamente maggiore di χ_{n-1}^2; la stessa relazione vale tra le rispettive funzioni crescenti χ_n e χ_{n-1}, per cui $E(\chi_n)$ è funzione crescente di n (essendo la media aritmetica una media monotona):

$$E(\chi_{n-1}) < E(\chi_n)$$

da cui

$$E(\chi_{n-1}) \cdot E(\chi_n) < \{E(\chi_n)\}^2$$

e infine, per la (6):

$$E(\chi_n) > \sqrt{n-1}. \tag{9}$$

Per la (8) e la (9) vale quindi la doppia limitazione

$$\sqrt{n-1} < E(\chi_n) < \sqrt{n} \tag{10}$$

e si può controllare numericamente l'ottima approssimazione offerta da

$$E(\chi_n) \simeq \sqrt{n-0{,}5}. \tag{11}$$

Passiamo ora all'applicazione del criterio ottimale della minima perdita media quadratica, nella classe degli stimatori di σ del tipo $c\sqrt{D}$. L'errore quadratico medio (*MSE* = *mean square error*) risulta, per la (3):

$$\begin{aligned} MSE(c\sqrt{D}) &= E\{(c\sqrt{D}-\sigma)^2\} = Var(c\sqrt{D}) + \{E(c\sqrt{D}) - \sigma\}^2 \\ &= c^2\sigma^2\, Var(\chi_{n-1}) + \sigma^2\, \{c\cdot E(\chi_{n-1}) - 1\}^2. \end{aligned}$$

Per la varianza di χ_{n-1} si ha:

$$\begin{aligned} Var(\chi_{n-1}) &= E(\chi^2_{n-1}) - \{E(\chi_{n-1})\}^2 \\ &= n - 1 - \{E(\chi_{n-1})\}^2. \end{aligned}$$

L'errore quadratico medio si scriverà quindi, dopo semplificazione:

$$MSE(c\sqrt{D}) = \sigma^2\, \{(n-1)c^2 + 1 - 2c\cdot E(\chi_{n-1})\};$$

differenziando, si ottiene che questa espressione in c è minimizzata per

$$c = E(\chi_{n-1})/(n-1) \tag{12}$$

$$= \{E(\chi_n)\}^{-1} \tag{13}$$

ottenendosi quest'ultima eguaglianza per la (7). Si noterà come, dalla (4), lo stimatore non distorto in questa classe sia invece determinato da c pari al reciproco di $E(\chi_{n-1})$.

3. Distorsione mediana e perdita assoluta

Nell'ambito della classe di stimatori del tipo $c\sqrt{D}$, si ricava quello con distorsione mediana nulla ponendo

$$Me(c\sqrt{D}) = c\sigma \cdot Me(\chi_{n-1}) = \sigma$$

da cui

$$c = \{Me(\chi_{n-1})\}^{-1} = \{Me(\chi^2_{n-1})\}^{-1/2} \tag{14}$$

valendo l'approssimazione (cfr. Frosini, 1986):

$$Me(\chi^2_{n-1}) \simeq n - 1 - 2/3.$$

L'applicazione del criterio ottimale della minima perdita media assoluta, nella stessa classe di stimatori, richiede la minimizzazione della seguente quantità (in cui MAE = *mean absolute error*):

$$MAE(c\sqrt{D}) = E\{|c\sqrt{D} - \sigma|\} = E\{|c\sigma \cdot \chi_{n-1} - \sigma|\} = \sigma E\{|c\chi_{n-1} - 1|\}.$$

Quest'ultima media si può scrivere

$$\begin{aligned} E\{|c\chi_{n-1} - 1|\} &= \int_0^\infty |cx - 1| \; \psi_{n-1}(x)\, dx \\ &= \int_0^{1/c} (1 - cx)\psi_{n-1}(x)\, dx + \int_{1/c}^\infty (cx - 1)\psi_{n-1}(x)\, dx \\ &= c \cdot E(\chi_{n-1}) - 1 - 2\int_0^{1/c} (cx - 1)\psi_{n-1}(x)\, dx; \end{aligned}$$

derivando rispetto a c ed eguagliando a zero si ottiene l'equazione

$$\frac{1}{E(\chi_{n-1})} \int_0^{1/c} x\psi_{n-1}(x)\, dx = \frac{1}{2}, \tag{15}$$

dove si legge che $1/c$ è la mediana della *v.c.* "primo momento incompleto della *v.c.* χ_{n-1}" (cfr. la formula analoga (10) in Frosini (1986), concernente la stima della varianza). Poiché si può scrivere

$$\begin{aligned} x\psi_{n-1}(x) &= 2\{2^{(n-1)/2}\Gamma((n - 1)/2)\}^{-1}x^{n-1}exp(-x^2/2) \\ &= E(\chi_{n-1}) \cdot \psi_n(x), \end{aligned} \tag{16}$$

sostituendo nella (15) si ha

$$\int_0^{1/c} \psi_n(x)\, dx = \frac{1}{2} \tag{17}$$

da cui $1/c = Me(\chi_n)$, e quindi

$$c = \{Me(\chi_n)\}^{-1} = \{Me(\chi_n^2)\}^{-1/2} \tag{18}$$
$$c \simeq (n - 1 - 2/3)^{-1/2}$$

Si osservi come il risultato conseguito differisca da quello in formula (14) concernente lo stimatore con distorsione mediana nulla.

Riferimenti bibliografici

Chapman D.G. e Robbins H., Minimum variance estimation without regularity assumptions, *Annals of Mathematical Statistics*, 22, 1951, 581-586.
Frosini B.V., Distorsione e criteri ottimali nella stima della varianza nel caso normale, *Rivista di Statistica Applicata*, 19, 1986.
Kendall M.G. e Stuart A., *The advanced theory of statistics,* vol. 2, Griffin, London, 1967.

Summary

On the estimation of the standard deviation in the normal case

Starting from the observation that estimating the standard deviation is more common and useful than estimating the variance, it is suggested that also this type of estimation problem should be included — as a rule — among the major examples of point estimation in textbooks. In this paper point estimators of the standard deviation in the normal case are obtained by imposing (in a given class of estimators) (a) mean unbiasedness, (b) median unbiasedness, (c) minimum mean square error, (d) minimum mean absolute error.