Rostra
a cura di
By
Benito V. Frosini

ALCUNE CONSIDERAZIONI ELEMENTARI SUL CONCETTO DI EFFICIENZA NEL CASO MULTIPARAMETRICO

Renato Leoni

Dipartimento Statistico dell'Università di Firenze.

1.

In una esposizione dei principali metodi di stima parametrica puntuale classica, è abbastanza usuale considerare dapprima il caso in cui la funzione che descrive in termini analitici la distribuzione del carattere X nella popolazione dipenda da un unico parametro incognito *(caso uniparametrico),* riservando ad un secondo momento l'esame del caso in cui i parametri incogniti siano due o più *(caso multiparametrico).*

Coerentemente a questa impostazione – che, com'è ovvio, presenta sul piano didattico numerosi vantaggi – i concetti di correttezza, efficienza, sufficienza, ecc. vengono definiti anzitutto per il caso uniparametrico ed estesi successivamente a quello multiparametrico. Tuttavia, mentre per la maggior parte di questi concetti l'estensione al caso multiparametrico è del tutto banale, per altri la questione presenta qualche maggiore difficoltà.

Supponiamo, ad esempio, che la funzione che descrive in termini analitici la distribuzione del carattere X nella popolazione dipenda dalla m–upla di parametri incogniti $(\theta_1,....,\theta_m)$, appartenente allo spazio parametrico Ω.

È immediato estendere direttamente il concetto di correttezza dal caso uniparametrico a quello multiparametrico e dire che la m–upla $(T_1,....,T_m)$ di stimatori di $(\theta_1,....,\theta_m)$ è *corretta* se e soltanto se (sse), qualunque sia $(\theta_1,....,\theta_m) \in \Omega$, risulta $E(T_1) = \theta_1,....,E(T_m) = \theta_m$.

Siano adesso $(T_1^{(1)},....,T_m^{(1)})$ e $(T_1^{(2)},....,T_m^{(2)})$ due m–uple di stimatori corretti di $(\theta_1,....,\theta_m)$. Anche questa volta si potrebbe pensare di estendere direttamente il concetto di efficienza dal caso uniparametrico a quello multiparametrico e dire che $(T_1^{(1)},....,T_m^{(1)})$ è *non meno efficiente* di $(T_1^{(2)},....,T_m^{(2)})$ sse, qualunque sia $(\theta_1,....,\theta_m) \in \Omega$, risulta

$$Var\left(T_1^{(1)}\right) \le Var\left(T_1^{(2)}\right),....,Var\left(T_m^{(1)}\right) \le Var\left(T_m^{(2)}\right). \tag{1}$$

In questo modo, tuttavia si verrebbero a trascurare le covarianze, generalmente non nulle, esistenti tra le componenti di ciascuna m–upla di stimatori. Se vogliamo, come sembra ragionevole, tener conto di questo ulteriore elemento nell'esprimere un giudizio di efficienza comparativa tra $(T_1^{(1)},....,T_m^{(1)})$ e $(T_1^{(2)},....,T_m^{(2)})$ diviene naturale considerare le loro *matrici delle covarianze* (o *matrici delle varianze–covarianze)*

$$\boldsymbol{V}_1 = \begin{bmatrix} Var\left(T_1^{(1)}\right) & \dots & Cov\left(T_1^{(1)},T_m^{(1)}\right) \\ \dots & \dots & \dots \\ Cov\left(T_1^{(1)},T_m^{(1)}\right) & \cdots & Var\left(T_m^{(2)}\right) \end{bmatrix},$$

$$\boldsymbol{V}_2 = \begin{bmatrix} Var\left(T_1^{(2)}\right) & \dots & Cov\left(T_1^{(2)},T_m^{(2)}\right) \\ \dots & \dots & \dots \\ Cov\left(T_1^{(2)},T_m^{(2)}\right) & \cdots & Var\left(T_m^{(2)}\right) \end{bmatrix},$$

e dire che la prima m–upla di stimatori è non meno efficiente della seconda sse – sulla base di una prefissata relazione di ordine nell'insieme $C(\boldsymbol{V})$ delle matrici delle covarianze, da denotarsi con il simbolo $\le$ – risulta $\boldsymbol{V}_1 \le \boldsymbol{V}_2$.

La questione che in via preliminare occorre risolvere riguarda quindi il significato da attribuire alla scrittura $\boldsymbol{V}_1 \le \boldsymbol{V}_2$, ovverosia il criterio con cui stabilire un ordinamento nell'insieme $C(\boldsymbol{V})$.

Lo scopo di questa nota è quello di svolgere alcune elementari considerazioni sul modo con cui tale questione viene correntemente risolta e di porre in evidenza le conseguenze più immediate che la soluzione prescelta comporta nei riguardi del concetto di efficienza.

2.

Consideriamo anzitutto due vettori $\boldsymbol{x}$ e $\boldsymbol{y}$ appartenenti ad R^m. Com'è noto, si dice che $\boldsymbol{x}$ è *minore od eguale* a $\boldsymbol{y}$ e si scrive $\boldsymbol{x} \leq \boldsymbol{y}$, *sse* $y_j - x_j \geq 0$ per ogni $j=1,....,m$, ovvero *sse* $\boldsymbol{y}-\boldsymbol{x}$ appartiene all'insieme C dei vettori di R^m a componenti non negative[(1)].

Ciò premesso, vediamo in che modo è possibile utilizzare la definizione ora richiamata per stabilire un ordinamento nell'insieme delle matrici delle covarianze.

Supponiamo, per semplicità, che $\boldsymbol{V}_1$ e $\boldsymbol{V}_2$ siano definite positive [(2)].

Esiste allora una matrice non singolare $\boldsymbol{P}$ tale che

$$\boldsymbol{P'V_1P} = \boldsymbol{D} = diag(\mu_1,....,\mu_m), \boldsymbol{P'V_2P} = \boldsymbol{I} = diag(1,....,1) \tag{2}$$

con $\mu_j > 0$ per ogni $j=1,....,m$. Indicando con μ e $\boldsymbol{u}$ i vettori le cui componenti sono date ordinatamente dagli elementi che entrano nella diagonale principale rispettivamente di $\boldsymbol{D}$ ed $\boldsymbol{I}$, la riduzione a forma canonica operata in (2) associa dunque la matrice $\boldsymbol{V}_1$ al vettore μ ed, analogamente, la matrice $\boldsymbol{V}_2$ al vettore $\boldsymbol{u}$. Diviene pertanto naturale servirsi dell'ordinamento stabilito per i vettori di R^m per definire un ordinamento nell'insieme delle matrici delle covarianze e dire che $\boldsymbol{V}_1 \leq \boldsymbol{V}_2$ sse $\mu \leq \boldsymbol{u}$, ovvero sse $\boldsymbol{u}-\mu \in C$. D'altra parte, dire che $\boldsymbol{u}-\mu \in C$ equivale ad affermare che la matrice $\boldsymbol{V}_2-\boldsymbol{V}_1$ è definita non negativa, cosicché possiamo anche esprimere la condizione ora indicata dicendo che $\boldsymbol{V}_1 \leq \boldsymbol{V}_2$ sse $\boldsymbol{V}_2-\boldsymbol{V}_1$ è definita non negativa [(3)].

3.

L'accoglimento del suddetto criterio di ordinamento nell'insieme delle matrici delle covarianze comporta alcune conseguenze nei riguardi del concetto di efficienza che vogliamo adesso porre in evidenza.

La prima è del tutto ovvia e consiste nel definire $(T_1^{(1)},....,T_m^{(1)})$ *non meno efficiente* di $(T_1^{(2)},....,T_m^{(2)})$ sse, qualunque sia $(\theta_1,....,\theta_m) \in \Omega$, $\boldsymbol{V}_2-\boldsymbol{V}_1$ è definita non

---

(1) *Si noti che C è un cono convesso – cioè se $\boldsymbol{z}_1, \boldsymbol{z}_2 \in C$, anche $c_1 \boldsymbol{z}_1 + c_2 \boldsymbol{z}_2 \in C$ per tutti i $c_1, c_2 \geq 0$. Pertanto, la condizione di cui sopra può essere espressa dicendo che $\boldsymbol{x} \leq \boldsymbol{y}$ sse $\boldsymbol{y}-\boldsymbol{x}$ appartiene al cono convesso C. È ovvio che si tratta di un ordinamento parziale.*

(2) *Si ricorda che, in generale, una matrice delle covarianze $\boldsymbol{V}$ è definita non negativa, cioè a dire la forma quadratica $\boldsymbol{a'Va}$ ad essa associata o è definita positiva ($\boldsymbol{a'Va}>0$ per ogni $\boldsymbol{a} \neq \boldsymbol{0}$) oppure è semidefinita positiva ($\boldsymbol{a'Va} \geq 0$ per ogni $\boldsymbol{a} \in R^m$ e $\boldsymbol{a'Va}=0$ per almeno un $\boldsymbol{a} \neq \boldsymbol{0}$).*

(3) *Poiché l'insieme $C(\boldsymbol{V})$ delle matrici delle covarianze, come facilmente si verifica, è un cono convesso – cioè se $\boldsymbol{Z}_1, \boldsymbol{Z}_2 \in C(\boldsymbol{V})$ anche $c_1 \boldsymbol{Z}_1 + c_2 \boldsymbol{Z}_2 \in C(\boldsymbol{V})$ per tutti i $c_1, c_2 \geq 0$ – la condizione di cui sopra può essere espressa dicendo che $\boldsymbol{V}_1 \leq \boldsymbol{V}_2$ sse $\boldsymbol{V}_2-\boldsymbol{V}_1$ appartiene al cono convesso $C(\boldsymbol{V})$. Ovviamente anche in questo caso si tratta di un ordinamento parziale.*

negativa. Ma dire che $\boldsymbol{V}_2-\boldsymbol{V}_1$ è definita non negativa equivale ad affermare che, per ogni $\boldsymbol{a}\in R^m$, $\boldsymbol{a}'\boldsymbol{V}_1\boldsymbol{a}\leq \boldsymbol{a}'\boldsymbol{V}_2\boldsymbol{a}$, ovvero che la varianza di una qualsiasi combinazione lineare della prima m–upla di stimatori è minore od eguale alla varianza della stessa combinazione lineare della seconda m–upla di stimatori.

Il fatto, inoltre, che $\boldsymbol{V}_2-\boldsymbol{V}_1$ sia definita non negativa implica che gli elementi sulla diagonale principale di tale matrice – della forma $Var(T_j^{(2)})-Var(T_j^{(1)})$ – siano non negativi. Ne consegue che si avrà anche

$$\sum_j Var\left(T_j^{(1)}\right)=tr(\boldsymbol{V}_1)\leq tr(\boldsymbol{V}_2)=\sum_j Var\left(T_j^{(2)}\right), \tag{3}$$

ovverosia che la *varianza globale* di $(T_1^{(1)},....,T_m^{(1)})$ è minore o eguale alla *varianza globale* di $(T_1^{(2)},....,T_m^{(2)})$.

Risultando inoltre dalla (2)

$$\boldsymbol{V}_1=\left(\boldsymbol{P}^{-1}\right)'\boldsymbol{D}\boldsymbol{P}^{-1},\boldsymbol{V}_2=\left(\boldsymbol{P}^{-1}\right)'\boldsymbol{P}^{-1}$$

e quindi

$$det(\boldsymbol{V}_1)=\left[det\left(\boldsymbol{P}^{-1}\right)\right]^2det(\boldsymbol{D}),det(\boldsymbol{V}_2)=\left[det\left(\boldsymbol{P}^{-1}\right)\right]^2,$$

si ha che

$$\frac{det(\boldsymbol{V}_1)}{det(\boldsymbol{V}_2)}=det(\boldsymbol{D})=\Pi_j\mu_j.$$

D'altro canto, poiché è $0<\mu_j\leq 1$ per ogni $j=1,2,....,m$ [4], possiamo concludere che

$$det(\boldsymbol{V}_1)\leq det(\boldsymbol{V}_2), \tag{4}$$

ovverosia che la *varianza generalizzata* di $(T_1^{(1)},....,T_m^{(1)})$ è minore o eguale alla *varianza generalizzata* di $(T_1^{(2)},....,T_m^{(2)})$.

4.

La definizione di efficienza qui sopra riportata è suscettibile di ricevere un'interessante interpretazione geometrica di cui, a conclusione di questa nota, ci proponiamo di fornire gli elementi essenziali.

A questo fine, supponiamo R^m dotato di prodotto scalare standard e consideriamo lo spazio euclideo di punti E^m ad esso associato. Scegliamo poi in E^m un

---

[4] *Infatti, come si ricorderà, dire che $\boldsymbol{V}_2-\boldsymbol{V}_1$ è definita non negativa equivale ad affermare che $\boldsymbol{u}-\mu\in C$, ovvero che $1-\mu_j\geq 0$ per ogni j = 1,2,....,m, da cui l'asserto.*

riferimento cartesiano (ortonormale) costituito dall'origine $O \in E^m$ e dalla base di R^m formata dagli m vettori canonici $\boldsymbol{u}_1,....,\boldsymbol{u}_m$ [5].

Occupiamoci anzitutto della rappresentazione in E^m della matrice delle covarianze $\boldsymbol{V}$ di una generica m–upla $(T_1,....,T_m)$ di stimatori corretti di $(\theta_1,....,\theta_m)$. Il modo più immediato è ovviamente quello di far corrispondere a $\boldsymbol{V}$ l'ellissoide $\boldsymbol{a}'\boldsymbol{Va}=1$. Tuttavia, nell'ipotesi che $\boldsymbol{V}$ sia definita positiva, potremmo anche considerare la sua inversa $\boldsymbol{V}^{-1}$ ed associare a questa l'ellissoide $\boldsymbol{a}'\boldsymbol{V}^{-1}\boldsymbol{a}=1$. Il primo di tali ellissoidi è chiamato *ellissoide di inerzia;* il secondo, *ellissoide di concentrazione* [6].

Benché sia formalmente equivalente riferirsi all'uno o all'altro ellissoide [7], conviene però, in generale, fare ricorso all'ellissoide di concentrazione in quanto è con riferimento a questo che sono più direttamente identificabili in E^m scarti quadratici medi e covarianze [8].

Allo scopo di chiarire quanto ora affermato, consideriamo il punto $\boldsymbol{a}_j=\boldsymbol{Vu}_j\,(1/\boldsymbol{u}'_j\boldsymbol{Vu}_j)^{1/2}$ $(j=1,2,....,m)$, appartenente all'ellissoide $\boldsymbol{a}'\boldsymbol{V}^{-1}\boldsymbol{a}=1$. Il gradiente di $\boldsymbol{a}'\boldsymbol{V}^{-1}\boldsymbol{a}$, calcolato in $\boldsymbol{a}_j$, è $2\boldsymbol{V}^{-1}\boldsymbol{a}_j=2\boldsymbol{u}_j\,(1/Var(T_j))^{1/2}$ ed ha quindi la stessa direzione di $\boldsymbol{u}_j$. Ciò significa che l'iperpiano tangente all'ellissoide $\boldsymbol{a}'\boldsymbol{V}^{-1}\boldsymbol{a}=1$ in $\boldsymbol{a}_j$ è ortogonale all'asse di direzione $\boldsymbol{u}_j$. Ne consegue che la proiezione ortogonale di $\boldsymbol{a}_j$ su tale asse è $\boldsymbol{u}_j(Var(T_j))^{1/2}$ e si trova quindi ad una distanza dall'origine pari allo scarto quadratico medio di T_j. A sua volta, la proiezione ortogonale di $\boldsymbol{a}_j$ sull'asse individuato da $\boldsymbol{u}_{j^*}(j \neq j^*\,;\,j^*=1,2,....,m)$ è

$$\boldsymbol{u}_{j*}\frac{Cov\left(T_{j*},T_j\right)}{\left(Var\left(T_j\right)\right)^{1/2}}=\boldsymbol{u}_{j*}\frac{Cov\left(T_{j*},T_j\right)\left(Var\left(T_{j*}\right)\right)^{1/2}}{\left(Var\left(T_{j*}\right)\right)^{1/2}\left(Var\left(T_j\right)\right)^{1/2}}=\boldsymbol{u}_{j*}r_{j*j}\left(Var\left(T_{j*}\right)\right)^{1/2}$$

e si trova pertanto ad una distanza dall'origine pari alla covarianza tra T_{j^*} e T_j divisa per lo scarto quadratico medio di T_j ovvero al coefficiente di correlazione lineare r_{j^*j} tra T_{j^*} e T_j moltiplicato per lo scarto quadratico medio di T_{j^*} [9].

---

(5) *$\boldsymbol{u}_j$ indica il vettore le cui componenti sono tutte eguali a 0 tranne la j–esima posta eguale ad 1.*

(6) *Più specificatamente, ellissoide di inerzia o di concentrazione riferito all'origine.*

(7) *Tra i due ellissoidi sussiste una stretta relazione: ad esempio, essi presentano gli stessi assi principali e semiassi la cui lunghezza è l'una il reciproco dell'altra.*

(8) *Qualora la forma della distribuzione di $(T_1,....,T_m)$ sia multinormale, è ovvio che sussiste un'ulteriore motivazione per usare l'ellissoide di concentrazione.*

(9) *Si osservi che anche la varianza generalizzata di $(T_1,....,T_m)$ è suscettibile di ricevere un'espressiva interpretazione geometrica. Si può infatti dimostrare che il quadrato del volume del parallelotopo circoscritto all'ellissoide di concentrazione $\boldsymbol{a}'\boldsymbol{V}^{-1}\boldsymbol{a}=1$ è proporzionale al det $(\boldsymbol{V})$.*

Dopo questa premessa, vediamo quale è l'interpretazione, in termini di ellissoidi di concentrazione, della definizione di efficienza che abbiamo dato in precedenza. Al riguardo, occorre anzitutto osservare che, se $\boldsymbol{V}_2-\boldsymbol{V}_1$ è definita non negativa, anche $\boldsymbol{V}_1^{-1}-\boldsymbol{V}_2^{-1}$ è definita non negativa [(10)]. Ne consegue che $\boldsymbol{a}'\boldsymbol{V}_1^{-1}\boldsymbol{a} \geq \boldsymbol{a}'\boldsymbol{V}_2^{-1}\boldsymbol{a}$, ovvero che i punti appartenenti all'ellissoide $\boldsymbol{a}'\boldsymbol{V}_1^{-1}\boldsymbol{a}=1$ sono *non esterni* ai punti dell'ellissoide $\boldsymbol{a}'\boldsymbol{V}_2^{-1}\boldsymbol{a}=1$. Possiamo allora dire che $(T_1^{(1)},....,T_m^{(1)})$ *è non meno efficiente* di $(T_1^{(2)},....,T_m^{(2)})$, sse qualunque sia $(\theta_1,....,\theta_m)\, E\Omega$, l'ellissoide $\boldsymbol{a}'\boldsymbol{V}_1^{-1}\boldsymbol{a}=1$ è *non esterno* all'ellissoide $\boldsymbol{a}'\boldsymbol{V}_2^{-1}\boldsymbol{a}=1$.

NOTA BIBLIOGRAFICA

Per una esposizione del procedimento di riduzione a forma canonica di due forme quadratiche, a cui si fa riferimento al punto 2, il lettore può consultare

Leoni R., Lauro N., Introduzione all'analisi statistica multidimensionale, Napoli 1984, pagg. 274–282.

Il concetto di ellissoide di concentrazione è stato introdotto, sotto la denominazione di "ellissoide indicatore", da

Darmois G., Sur les limites de la dispersion de certaines estimations, Revue de l'Institut International de Statistique, 1945.

Il medesimo concetto – sia pure sotto denominazioni a volte diverse e con varianti inessenziali – si trova espresso in

Cramér H., Mathematical Methods of Statistics, Princeton 1954, pagg. 283–285 e pagg. 300–301.

De Finetti B., Teoria delle probabilità, Torino 1970, pagg. 201–203 e pagg. 541–543.

Dempster A.P., Continuos Multivariate Analysis, London 1969, pagg. 123–127.

Malinvaud E., Metodi statistici dell'econometria, Torino 1971, pagg. 182–186.

SUMMARY

In this paper, some remarks on the covariance matrices ordering are developed and applied to the concept of efficiency in the multiparametric case.

[(10)] *Dalla (2) risulta*

$$\boldsymbol{V}_1^{-1}-\boldsymbol{V}_2^{-1}=\boldsymbol{P}\boldsymbol{D}^{-1}\boldsymbol{P}'-\boldsymbol{P}\boldsymbol{P}'=\boldsymbol{P}\left(\boldsymbol{D}^{-1}-\boldsymbol{I}\right)\boldsymbol{P}'$$

e quindi

$$\boldsymbol{a}'\left(\boldsymbol{V}_1^{-1}-\boldsymbol{V}_2^{-1}\right)\boldsymbol{a}=\boldsymbol{a}'\boldsymbol{P}\left(\boldsymbol{D}^{-1}-\boldsymbol{I}\right)\boldsymbol{P}'\boldsymbol{a}=b'\left(\boldsymbol{D}^{-1}-\boldsymbol{I}\right)\boldsymbol{b}=\sum_j\left[\left(1/\mu_j\right)-1\right]b_j^2\boldsymbol{I}.$$

Ma, nell'ipotesi che $\boldsymbol{V}_2-\boldsymbol{V}_1$ *sia definita non negativa è* $\mu_j\leq 1$ *(Cfr. nota (4)), ossia* $1\leq (1/\mu_j)$*; pertanto* $\boldsymbol{V}_1^{-1}-\boldsymbol{V}_2^{-1}$ *è definita non negativa.*