Atti Accademia Nazionale dei Lincei

Classe Scienze Fisiche Matematiche Naturali

Rendiconti

Philippe Caperaa, Robert Cote

Estimation bayesienne avec données manquantes

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Serie 8, Vol. **54** (1973), n.6, p. 887–891.

Accademia Nazionale dei Lincei

<http://www.bdim.eu/item?id=RLINA_1973_8_54_6_887_0>

Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali. Rendiconti, Accademia Nazionale dei Lincei, 1973.

Calcolo delle probabilità. — *Estimation bayesienne avec données manquantes* (*). Nota di PHILIPPE CAPÉRAÀ e ROBERT CÔTÉ, presentata (**) dal Socio B. SEGRE.

RIASSUNTO. — In questa Nota studiamo, partendo da un modello « Bayesiano », le condizioni sotto le quali è possibile, sostituendo delle osservazioni perdute su una parte di un vettore normale con delle osservazioni supplementari relative all'altra parte di questo vettore, di ottenere una stima della media almeno altrettanto buona di quella ottenuta senza perdita di osservazione.

I. INTRODUCTION

Soit $(\mathbf{X})_n$ un échantillon aléatoire de taille n d'un vecteur aléatoire $\mathbf{X}$ de dimension p, $p < n$, dont la loi de probabilité dépend de t paramètres réels $(\theta_1, \theta_2, \cdots, \theta_t) = \boldsymbol{\theta}' \cdot (\boldsymbol{\theta}'$ signifie le vecteur transposé de $\boldsymbol{\theta})$. Supposons que les $p-q$ dernières composantes de $n-r$ vecteurs de l'échantillon soient détruites. Cette note étudie les conditions rendant possible le remplacement d'observations perdues sur $\mathbf{X}_2$, par des observations prises sur $\mathbf{X}_1$ $(\mathbf{X}' = [\mathbf{X}_1' : \mathbf{X}_2'])$ sans perdre de précision sur l'estimation d'une fonction de $\boldsymbol{\theta}$. Nous noterons le nouvel échantillon ainsi obtenu

$$(\mathbf{X})_{a,r} = \begin{bmatrix} (\mathbf{X}_1)_a \\ (\mathbf{X}_2)_r \end{bmatrix} .$$

Soit $\mathrm{F}[(\boldsymbol{x})_{a,r} \mid \boldsymbol{\theta}]$ la vraisemblance de l'échantillon $(\mathbf{X})_{a,r}$. Notons $\pi(\boldsymbol{\theta})$ la loi a priori du vecteur $\boldsymbol{\theta}$. En utilisant une fonction de perte $\mathrm{L}(\cdot\,;\cdot)$ quadratique, l'estimateur solution de Bayes pour $\pi(\boldsymbol{\theta})$ d'une fonction $\gamma(\boldsymbol{\theta})$ des paramètres sera l'espérance a posteriori de $\gamma(\boldsymbol{\theta})$, soit

$$\mathrm{d}_0 [(\boldsymbol{x})_{a,r}] = \mathrm{E}[\gamma(\boldsymbol{\theta}) \mid (\boldsymbol{x})_{a,r}] . \tag{1}$$

En effet le risque a posteriori de la décision $\mathrm{d}(\cdot)$ par rapport à $\pi(\cdot)$ s'écrit

$$\mathrm{R}(\pi, \mathrm{d}) = \int \left\{ \int \mathrm{L}[\mathrm{d}((\boldsymbol{x})_{a,r}) ; \gamma(\boldsymbol{\theta})] \, \mathrm{d}\pi [\boldsymbol{\theta} \mid (\boldsymbol{x})_{a,r}] \right\} \mathrm{dF}[(\boldsymbol{x})_{a,r}] .$$

Ce risque est minimisé pour $d[(\boldsymbol{x})_{a,r}] = \mathrm{E}[\gamma(\boldsymbol{\theta}) \mid (\boldsymbol{x})_{a,r}]$. Si $r = n = a$, ce résultat est connu ([1], p. 46).

(*) Cette recherche a été supportée par une subvention du C.N.R. du Canada.
(**) Nella seduta del 19 giugno 1973.

II. APPLICATION À LA LOI MULTINORMALE

Considérons l'échantillon $(\mathbf{X})_{a,r}$ tiré d'une loi normale $N_p(\mu, \Sigma)$ où Σ est une matrice régulière et connue. Soit la partition naturelle accompagnant celle de $(\mathbf{X})_{a,r}$:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad , \quad \Sigma^{-1} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} .$$

Alors

$$dF\,[(\boldsymbol{x})_{a,r} \mid \mu] \propto \exp. \left\{ -\frac{1}{2} \left[\operatorname{trace} (\Sigma^{-1} U_r) + \operatorname{trace} (\Sigma_{11}^{-1} U_{r,a}) \right] \right\}$$

où $U_r = \sum_{j=1}^{r} (\boldsymbol{x}_j - \mu)(\boldsymbol{x}_j - \mu)'$ et $U_{r,a} = \sum_{j=r+1}^{a} (\boldsymbol{x}_{1j} - \mu_1)(\boldsymbol{x}_{1j} - \mu_1)'$.

Nous envisagerons deux types de connaissance a priori sur μ. D'une part nous prendrons pour $\pi(\mu)$ une loi normale, conjuguée naturelle de la loi de l'échantillon, et l'autre part, nous prendrons $d\pi(\mu) = d\mu$ ce qui traduit une connaissance diffuse de μ. Dans ces deux cas nous chercherons les estimateurs Bayesiens pour $\gamma(\mu) = \mu_2$.

A. *Cas où $\pi(\cdot)$ est une loi normale.*

Soit μ obéissant à une loi normale $N_p(\boldsymbol{m}, V)$ où V est une matrice régulière et connue. Notons

$$\boldsymbol{m} = \begin{bmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \end{bmatrix} \quad , \quad V^{-1} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} .$$

Alors

$$d\pi(\mu) \propto \exp. \left\{ -\frac{1}{2} \operatorname{trace} (V^{-1} W) \right\} \quad \text{où} \quad W = (\boldsymbol{m} - \mu)(\boldsymbol{m} - \mu)' .$$

La solution de Bayes pour estimer μ_2 est donnée par (1), soit

$$d\,[(\boldsymbol{x})_{a,r}] = E\,[\mu_2 \mid (\boldsymbol{x})_{a,r}] .$$

Après calcul, nous obtenons ${}^*d\,[(\boldsymbol{x})_{a,r}] = {}^*l_2' \, {}^*M_2^{-1}$ où

$${}^*l_2' = \left\{ \sum_{j=1}^{r} \boldsymbol{x}_{1j}' \sigma_{12} + \sum_{j=1}^{r} \boldsymbol{x}_{2j}' \sigma_{22} + \boldsymbol{m}_2' v_{12} + \boldsymbol{m}_2' v_{22} - {}^*\mathbf{L}_2 W \right\}$$

$${}^*\mathbf{L}_2 = \left[\sum_{j=1}^{r} \boldsymbol{x}_{ij}' \sigma_{11} + \sum_{j=1}^{r} \boldsymbol{x}_{2j}' \sigma_{21} + \sum_{j=r+1}^{a} \boldsymbol{x}_{1j}' \Sigma_{11}^{-1} + \boldsymbol{m}_1' v_{11} + \boldsymbol{m}_2' v_{21} \right]$$

$$W = M_1^{-1} [r\sigma_{21} + v_{21}]$$

$$M_1 = r\sigma_{22} + v_{11} + (a - r)\,\Sigma_{11}^{-1}$$

$${}^*M_2 = r\sigma_{22} + v_{22} - (r\sigma_{21} + v_{21})\, M_1^{-1} (r\sigma_{12} + v_{12}) .$$

La précision de cette estimation, c'est-à-dire, l'inverse de la matrice des covariances de la loi a posteriori est

$$^*\mathrm{V}^{-1}(a\,,r) = r\sigma_{22} + v_{22} - (r\sigma_{21} + v_{21})\,[r\sigma_{11} + v_{11} + (a - r)\,\Sigma_{11}^{-1}]^{-1}\,(r\sigma_{12} + v_{12})\,.$$

B. *Cas où* $\pi(\cdot)$ *traduit une connaissance diffuse de* μ.

Le résultat (1) peut être obtenu pour une mesure $\pi(\cdot)$ non finie mais telle que tout ensemble borné de $\Theta = \{\theta : \theta_i \in \mathrm{R}\ \ i = 1\,, 2\,, \cdots, t\}$ ait une mesure finie [4]. Alors nous obtenons $^{**}d\,[(\mathbf{X})_{a,r}] = {}^{**}l_2'\,{}^{**}\mathrm{M}_2^{-1}$ où

$$^{**}\mathrm{M}_2 = r\sigma_{22} - r^2\,\sigma_{21}\,[r\sigma_{11} + (a - r)\,\Sigma_{11}^{-1}]^{-1}\,\sigma_{12}$$

$$^{**}l_2' = \sum_{j=1}^{r} \boldsymbol{x}_{1j}'\,\sigma_{12} + \sum_{j=1}^{r} \boldsymbol{x}_{2j}'\,\sigma_{22} - {}^{**}\mathrm{L}_2$$

$$^{**}\mathbf{L}_2 = \sum_{j=1}^{r} \boldsymbol{x}_{1j}'\,\sigma_{11} + \sum_{j=1} \boldsymbol{x}_{2j}'\,\sigma_{21} + \sum_{j=r+1} \boldsymbol{x}_{1j}'\,\Sigma_{11}^{-1}\,{}^{**}\mathrm{M}_2$$

et $\quad ^{**}\mathrm{V}^{-1}(a\,,r) = {}^{**}\mathrm{M}_2\,.$

III. ETUDE DE $^*\mathrm{V}^{-1}(a\,,r)$

Les $a - n$ observations additionnelles sont prises sur $\mathbf{X}_1$ pour récupérer, en tout ou en partie, la précision perdue par les $n - r$ observations détruites sur $\mathbf{X}_2$. Nous allons chercher les conditions pour qu'il existe un nombre entier a tel que $^*\mathrm{V}^{-1}(a\,,r) \geq {}^*\mathrm{V}^{-1}(n\,,n)$. Les résultats suivants nous seront alors utiles:

$\forall r \in \mathbf{R}$ (fixé), nous obtenons

a) $\quad \boldsymbol{z}'\,{}^*\mathrm{V}^{-1}(k\,,r)\,\boldsymbol{z} \geq \boldsymbol{z}'\,{}^*\mathrm{V}^{-1}(o\,,r)\,\boldsymbol{z} \qquad \forall k \geq 1\,,\ \forall \boldsymbol{z} \in \mathbf{R}^{p-q}$

b) $\quad \lim\limits_{k\to\infty} {}^*\mathrm{V}^{-1}(k\,,r) = [r\sigma_{22} + v_{22}]$

c) $\quad \forall \boldsymbol{z} \in \mathbf{R}^{p-q}$ et $r < a$, $\mathrm{Q}(a) = \boldsymbol{z}'\,{}^*\mathrm{V}(a\,,r)\,\boldsymbol{z}$ est une fonction continue strictement décroissante de a.

Le résultat *a)* est essentiellement basé sur le lemme suivant: ([2], p. 329, Théorème 12.2.14).

LEMME 1. *Si* A *et* B *sont deux matrices symmétriques et régulières telles que* $\forall \boldsymbol{z} \neq \mathbf{o}\,,\ \boldsymbol{z}'\,\mathrm{A}\boldsymbol{z} > \boldsymbol{z}'\,\mathrm{B}\boldsymbol{z}$ alors $\boldsymbol{z}'\,\mathrm{A}^{-1}\,\boldsymbol{z} < \boldsymbol{z}'\mathrm{B}^{-1}\,\boldsymbol{z}$.

Le résultat *c)* peut être prouvé par des arguments semblables à ceux de *a)* et le résultat *b)* est basé essentiellement sur le lemme suivant:

LEMME 2. *Soient* A, B et $\mathrm{A} + c\mathrm{B}$ *des matrices régulières avec* $c \geq 1$. *Alors si* A *et* B *ne dépendent pas de* c *nous avons* $\lim\limits_{c\to+\infty} [\mathrm{A} + c\mathrm{B}]^{-1} = 0$.

Le résultat *a)* souligne clairement qu'on peut, par des observations additionnelles sur $\mathbf{X}_1$ accroître la précision de l'estimation de μ_2. De plus, si $[r\sigma_{22} + v_{22}] - {}^*\mathrm{V}^{-1}(n\,,n)$ est une matrice définie positive i.e. si

(2) $\qquad (r - n)\,\sigma_{22} + (n\sigma_{21} + v_{21})\,[n\sigma_{11} + v_{11}]^{-1}\,(n\sigma_{12} + v_{12})$

est une matrice définie positive, alors en utilisant le résultat c) on peut conclure qu'il existe $a_0 > n$ tel que

$$^*V^{-1}(a_0, r) = {}^*V^{-1}(n, n).$$

Pour $a_0^* = [a_0] + 1$, où $[a_0]$ est la partie entière de a_0, nous avons

$$\forall\, \boldsymbol{z},\ \boldsymbol{z}'\, {}^*V^{-1'}(a_0^*, r)\, \boldsymbol{z} \geq \boldsymbol{z}'\, {}^*V^{-1}(n, n)\, \boldsymbol{z}$$

Donc si la condition (2) est satisfaite, il sera possible de remplacer $n - r$ observations sur $\mathbf{X}_2$ par $a - n$ observations sur $\mathbf{X}_1$ de telle façon que l'estimation de μ_2 ait une précision au moins aussi grande que celle qui aurait été obtenue en utilisant l'échantillon $(\mathbf{X})_n$.

Regardons maintenant ce qu'implique la condition (2). Puisque $n > r$, (2) est satisfaite seulement si la matrice $(n\sigma_{21} + v_{21})\,[n\sigma_{11} + v_{11}]^{-1}\,(n\sigma_{12} + v_{12})$ est définie positive donc de rang $(p - q)$. En notant que $(n\sigma_{12} + v_{12})$ est de dimension $(p - q, q)$, alors la condition (2) entraîne que $p - q \leq q$. Ceci signifie que pour obtenir une précision aussi grande pour l'estimateur de μ_2 par l'échantillon $(\mathbf{X})_{a,r}$ que par l'échantillon $(\mathbf{X})_n$ il faut que la dimension de $\mathbf{X}_2$ soit au plus égale à celle de $\mathbf{X}_1$.

Remarque. Des résultats semblables ont été obtenus pour la précision $^{**}V^{-1}(a, r)$, et l'existence de $a_0 \in R$ tel que

$$^{**}V^{-1}(a_0, r) = {}^{**}V^{-1}(n, n)$$

est assurée si la matrice $[(r - n)\sigma_{22} + n\sigma_{21}\,\sigma_{11}^{-1}\,\sigma_{12}]$ est définie positive.

IV. Détermination de a_0

Pour déterminer analytiquement a_0 il faut résoudre l'équation en a suivante:

$$(r\sigma_{21} + v_{21})\,[r\sigma_{11} + v_{11} + (a - r)\,\Sigma_{11}^{-1}]^{-1}\,(r\sigma_{12} + v_{12}) = \tag{3}$$
$$= (n\sigma_{21} + v_{21})\,[n\sigma_{11} + v_{11}]^{-1}\,(n\sigma_{12} + v_{12}) - (n - r)\,\sigma_{22}.$$

Posant $A = r\sigma_{12} + v_{12}$

$$Y = [r\sigma_{11} + v_{11} + (a - r)\,\Sigma_{11}^{-1}]^{-1}$$
$$B = (n\sigma_{21} + v_{21})\,[n\sigma_{11} + v_{11}]^{-1}\,(n\sigma_{12} + v_{12}) - (n - r)\,\sigma_{22}$$

l'équation (3) s'écrit

$$A'\,YA = B. \tag{4}$$

Puisque le rang de A est $(p - q)$ le système (4) est compatible et l'expression générale des solutions est

$$Y = (A')^+\,BA^+ + W - (A')^+\,AWAA^+$$

où A^+ est l'inverse de Moore–Penrose de A ([3], p. 26).

Sachant que a_0 est unique, il existe donc une solution unique Y et par suite une matrice unique W telle que l'équation (3) soit satisfaite pour a_0. Il semble impossible, en général, de déterminer analytiquement la matrice Y et par conséquent le a_0 satisfaisant (3). Cependant on peut aisément déterminer a_0 par approximations successives. Quelques résultats ont été obtenus par cette méthode pour $p \leq 4$.

REFERENCES

[1] T. S. FERGUSON, *Mathematical Statistics: A decision Theoretic Approach*, Academic Press, N. Y. (1967).

[2] F. A. GRAYBILL, *Introduction to Matrices with Applications in Statistics*, Wadsworth Publishing Company, Inc., Belmont, California (1969).

[3] C. R. RAO et S. K. MITRA, *Generalized Inverses of Matrices and its Applications*, John Wiley and Sons, Inc., Toronto (1971).

[4] W. E. STRAWDERMAN et A. COHEN, *Admissibility of estimators of the mean vector of a multivariety normal distribution with quadratic loss*, « The Annals of Mathematical Statistics », *42*, 270–296 (February 1971).