ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 8 — 1987

Estimateurs à rétrécisseur matriciel différentiable, pour un coût quadratique général

Anne-Marie FRAISSE, Christian ROBERT, Madeleine ROY *

RÉSUMÉ. – Pour l'estimation de la moyenne d'une loi normale multidimensionnelle dont la variance n'est connue qu'à un facteur multiplicatif près, on obtient des conditions suffisantes de domination de l'estimateur des moindres carrés par des estimateurs à rétrécisseur d'une classe sensiblement plus large que celles envisagées dans la plupart des articles consacrés à ce sujet.

Shrinkage Estimators with Differentiable Shrinking Functions for Arbitrary Quadratic Loss

ABSTRACT. – The problem of the estimation of a multivariate normal mean when the variance is known up to a multiplicative factor is considered under an arbitrary quadratic loss. We introduce shrinkage estimators with differentiable shrinking functions under weak algebraic assumptions. We deduce sufficient conditions for minimaxity from an unbiased estimator of the risk. We also consider simpler conditions in particular cases like those of shrinking functions with controlled variations.

* A. M. FRAISSE et M. ROY : Université Paris 5-René Descartes. C. ROBERT : Université de Rouen (UA CNRS 759-Calcul des Probabilités et Statistique) et ENSAE, 3, avenue Pierre-Larousse, 92240 Malakoff.

1 Introduction

Les estimateurs à rétrécisseur ont été introduits par JAMES et STEIN [1961] pour démontrer l'inadmissibilité de l'estimateur des moindres carrés de la moyenne d'un vecteur aléatoire normal de dimension supérieure ou égale à 3. Ils ont été généralisés par BARANCHIK [1964] et BOCK [1975] grâce à la notion de fonction de rétrécissement, puis par BERGER et BOCK [1976] et JUDGE et BOCK [1978] qui ont remplacé le coefficient de rétrécissement par un *endomorphisme de rétrécissement*, modulant le rétrécissement suivant la composante (d'où la dénomination de *rétrécisseur matriciel* adoptée dans le titre). Les résultats obtenus s'adaptent de manière naturelle à l'estimation des coefficients d'une régression linéaire (voir remarque 1). Pour une étude plus détaillée des estimateurs à rétrécisseur, on pourra, par exemple, consulter JUDGE et BOCK ou BERGER [1980].

Dans cet article, nous étendons en particulier les résultats de BERGER [1976] à des modèles où la variance n'est connue qu'à un facteur multiplicatif près, grâce à l'hypothèse de différentiabilité qui permet d'utiliser des méthodes d'intégration par parties introduites dans STEIN [1973] et reprises dans STEIN [1981].

Après avoir donné, en section 3, comme dans BERGER [1976] et EFRON et MORRIS [1976], des décompositions du risque des estimateurs considérés qui permetent d'exhiber des estimateurs sans biais de ce risque, nous en déduisons en section 4 des conditions suffisantes de domination du risque de l'estimateur du maximum de vraisemblance qui sont rendues plus explicites si l'on restreint la classe des endomorphismes de rétrécissement. On aboutit ainsi, en section 5, à un résultat proche de celui de CELLIER, FOURDRINIER et ROBERT [1986].

2 Le modèle

2.1. Notations

Soit V un espace vectoriel de dimension n sur $\mathbb{R}$ et a une forme bilinéaire sur V. On note $\bar{a}$ la forme quadratique associée à a.

Si a est une forme bilinéaire symétrique sur V, on peut la considérer comme un homomorphisme de V dans V*, dual de V. On note a^- un inverse généralisé de a, c'est-à-dire un homomorphisme de V* dans V vérifiant $aa^- a=a$. On a alors $\mathrm{Im}(a^-) \oplus \mathrm{Ker}(a)=\mathrm{V}$.

2.2. Soit y une variable aléatoire à valeurs dans un espace vectoriel E de dimension n sur $\mathbb{R}$, suivant une loi normale $\mathrm{N}(\theta, \sigma^2 v)$, où θ appartient à un sous-espace vectoriel connu Θ de E, de dimension $k\,(k < n)$, v est une forme bilinéaire symétrique définie positive connue sur E* et les paramètres θ et $\sigma (\sigma > 0)$ sont inconnus.

Les estimateurs de θ considérés, applications φ de E dans Θ, sont comparés au moyen d'un coût quadratique défini par une forme bilinéaire symétrique positive q sur Θ, soit $\sigma^{-2}\bar{q}(\varphi(y)-\theta)$. Le risque de φ est alors

$$R_\varphi(\theta, \sigma^2)=\sigma^{-2}\,\mathbb{E}_{\theta,\sigma^2}[\bar{q}(\varphi(y)-\theta)]$$

où, pour toute application mesurable f de E dans $\mathbb{R}$, on note $\mathbb{E}_{\theta,\sigma^2}[f(y)]$ (ou $\mathbb{E}[f(y)]$ si cela ne prête pas à ambiguïté) pour $\int_E f(y)\,N(\theta, \sigma^2 v; dy)$.

On note φ^0 l'estimateur du maximum de vraisemblance, projecteur v^{-1}-orthogonal sur Θ. La loi de $\varphi^0(y)$ est $N(\theta, \sigma^2 v_\Theta)$, où v_Θ^{-1} est la restriction de v^{-1} à Θ. Cet estimateur est minimax et admet un risque constant, égal à $\mathrm{tr}(v_\Theta q)$.

Remarque 1 : Ce modèle est, en fait, celui de la régression linéaire. En effet, si $x_1, \ldots, x_k$ sont k vecteurs de E engendrant Θ, pour $\theta=\sum_{i=1}^{k}\beta_i x_i$, on a $y=\sum_{i=1}^{k}\beta_i x_i+u$, où u suit une loi normale de moyenne 0 et de variance $\sigma^2 v$; l'estimation de $(\beta_1, \ldots, \beta_k)$ revient donc à estimer θ dans la base $(x_1, \ldots, x_k)$ de Θ. Cependant ne pas choisir une base de référence pour Θ permet d'aboutir à des expressions concices des résultats.

2.3. Les estimateurs à rétrécisseur.

Soit b une forme bilinéaire symétrique positive sur Θ, c un endomorphisme de Θ et h une application de $(\mathbb{R}_+)^2$ dans $\mathbb{R}_+$; c sera dit endomorphisme de rétrécissement et h fonction de rétrécissement. On pose l'hypothèse H_1 :

HYPOTHÈSE H_1 : $\mathrm{Ker}\, b \subset \mathrm{Ker}\, c$.

Les estimateurs à rétrécisseur seront de la forme

$$\varphi(y)=\left[id_\Theta-h\left(\bar{b}(\varphi^0(y)), \frac{n-k}{\overline{v^{-1}}(y-\varphi^0(y))}\right)c\right]\varphi^0(y). \tag{1}$$

Dans toute la suite du texte, nous noterons

$$t=\bar{b}(\varphi^0(y)) \qquad \text{et} \qquad s^2=\overline{v^{-1}}(y-\varphi^0(y))/(n-k),$$

qui est l'estimateur usuel de la variance σ^2.

Considérons une base de E, $(e_1, \ldots, e_n)$, v^{-1}- orthonormale, telle que $(e_1, \ldots, e_k)$ forme une base b-orthogonale de Θ, telle que, si $m=rg(b)$, les $k-m$ derniers vecteurs de cette base engendrent $\mathrm{Ker}(b)$. On a alors

$$\varphi^0(y)=\sum_{i=1}^{k} y_i e_i \qquad \text{et} \qquad \theta=\sum_{i=1}^{k}\theta_i e_i,$$

$$t=\sum_{i=1}^{m} b_i y_i^2 \qquad \text{et} \qquad s^2=\frac{1}{n-k}\sum_{i=k+1}^{n} y_i^2,$$

d'où

$$\varphi(y)=\sum_{i=1}^{k} y_i e_i - h\left(\sum_{i=1}^{m} b_i y_i^2,\ (n-k)\left(\sum_{i=k+1}^{n} y_i^2\right)^{-1}\right)\sum_{i=1}^{k}\sum_{j=1}^{k} c_{ij} y_j e_i$$

Les estimateurs introduits en (2.1) généralisent BERGER et BOCK [1976], CELLIER et FOURDRINIER [1985], EFRON et MORRIS [1976], JUDGE et BOCK [1978], STRAWDERMAN [1973] qui ne considèrent que des fonctions de rétrécissement d'une seule variable ou qui supposent les opérateurs b, c, q simultanément diagonalisables dans une base v_Θ^{-1}- orthonormale. CELLIER et FOURDRINIER [1985] introduisent des estimateurs à fonctions de rétrécissement non nécessairement différentiables (ni même continues) tandis que nous supposerons par la suite que h admet des dérivées partielles relativement à ses deux variables. Au prix de cette hypothèse analytique, nous supprimons ainsi, par rapport aux articles ci-dessus, des contraintes assez fortes pesant sur les décompositions spectrales des opérateurs utilisés.

Nous dirons que φ est *régulière* si sa fonction de rétrécissement h vérifie les conditions suivantes (où l'on note h_i' la dérivée partielle de h relativement à la i-ième variable) :

$$\mathbb{E}_{\theta,\sigma}[th^2(t,\ 1/s^2)],\qquad \mathbb{E}_{\theta,\sigma}[th_1'(t,\ 1/s^2)],$$

$$\mathbb{E}_{\theta,\sigma}\left[\frac{t}{s^4}h(t,\ 1/s^2)\,h_1'(t,\ 1/s^2)\right]\quad \text{et} \quad \mathbb{E}_{\theta,\sigma}\left[\frac{t}{s^2}h^2(t,\ 1/s^2)\right]$$

sont finis, quels que soient $\theta\in\Theta$ et $\sigma \geqq 0$.

On verra que ces contraintes assurent la finitude du risque de l'estimateur associé à h (voir proposition 1) et la légitimité des intégrations par parties effectuées dans les différentes démonstrations par application du lemme de Stein (voir appendice).

b n'est pas nécessairement définie mais nous supposerons que $rg(b) \geqslant 3$. Cette contrainte est naturelle puisque, lorsque (H_1) est vérifiée, on voit que, sinon, l'estimateur défini par (1) ne diffère de φ^0 qu'au plus suivant deux directions indépendantes. STEIN [1956] a montré que, pour cette situation, φ^0 ne peut être uniformément amélioré par φ.

3 Risque des estimateurs à rétrécisseur

3.1. Sous l'hypothèse H_1, la restriction de l'endomorphisme de Θ^*, ${}^t cqcb^-$, à $\operatorname{Im} b$ est définie sans ambiguïté [lemme 1 (2)]. Soit Σ son spectre, nous noterons $\sigma_M=\sup\Sigma$ et $\sigma_m=\inf\Sigma$, qui est positif ou nul d'après le lemme 1 (4). Alors

PROPOSITION 1 : Sous l'hypothèse H_1, une condition suffisante pour que le risque de φ soit fini en (θ, σ^2) est que $\mathbb{E}_{\theta, \sigma^2}[th^2(t, 1/s^2)]$ soit fini.

Si, de plus, $rg(c) = rg(b)$ et la restriction de q à $\operatorname{Im} c$ est définie positive, cette condition est aussi nécessaire.

Démonstration : D'après le lemme 1 (4), on a

$$\forall x \in \Theta, \quad \sigma_m \bar{b}(x) \leqslant \bar{q}(c(x)) \leqslant \sigma_M \bar{b}(x) \tag{2}$$

d'où

$$\sigma_m \mathbb{E}[th^2(t, 1/s^2)] \leqslant \mathbb{E}[\bar{q}(c\varphi^0(y))h^2(t, 1/s^2)] \leqslant \sigma_M \mathbb{E}[th^2(t, 1/s^2)]. \tag{3}$$

Or il résulte élémentairement de l'inégalité de Schwarz (voir par exemple CELLIER et FOURDRINIER [1985]) que le risque est fini en (θ, σ^2) si et seulement si $\mathbb{E}_{\theta, \sigma^2}[h^2(t, 1/s^2)\bar{q}(c\varphi^0(y))]$ est fini, d'où le résultat annoncé, en remarquant (pour la condition nécessaire) que σ_m est différent de 0 si et seulement si la restriction à $\operatorname{Im} b$ de ${}^t cqcb^-$ est régulière, ce qui équivaut à $rg(b) = rg(c)$ et la restriction de q à $\operatorname{Im} b$ définie. □

Remarque 2 : Si $\sigma_M = 0$, ${}^t cqc$ est identiquement nulle et, pour tout (θ, y), $\bar{q}(\varphi^0(y) - \theta) = \bar{q}(\varphi(y) - \theta)$. Donc φ et φ^0 ont même risque. Ceci justifie l'hypothèse :

HYPOTHÈSE H_2 : La restriction à $\operatorname{Im} c$ de q n'est pas nulle.
Sous l'hypothèse H_1, elle équivaut à $\sigma_M \neq 0$.

3.2.

PROPOSITION 2 : Si l'hypothèse H_1 est vérifiée, si φ est régulière, $R_\varphi(\theta, \sigma^2)$ s'écrit

$$\begin{aligned} R_\varphi(\theta, \sigma^2) = {} & R_{\varphi 0}(\theta, \sigma^2) \\ & - 2\mathbb{E}_{\theta, \sigma^2}[h(t, 1/s^2)\, tr(v_\Theta qc) + 2h'_1(t, 1/s^2)\, \overline{bv_\Theta qc}(\varphi^0(y))] \\ & + \mathbb{E}_{\theta, \sigma^2}\left[\bar{q}(c\varphi^0(y))\left\{-4h(t, 1/s^2)h'_2(t, 1/s^2)\frac{1}{(n-k)s^4}\right.\right. \\ & \left.\left. + \frac{n-k-2}{(n-k)s^2}h^2(t, 1/s^2)\right\}\right] \end{aligned}$$

Remarque 3 : Ce résultat, généralisation du théorème 1 de EFFRON et MORRIS [1976], peut être considéré comme la donnée d'un estimateur sans biais du risque de φ.

Démonstration : Reprenant la base de E introduite dans la section 2.3, nous pouvons écrire :

$$(v_\Theta qc)(\varphi^0(y)) = \sum_{i=1}^{k} (d_i y_i + \beta_i) e_i,$$

où, pour tout i, β_i est fonction des composantes de $\varphi^0(y)$ autres que y_i, d_i est une constante.

Posons $\Delta_\varphi(\theta, \sigma^2)=R_\varphi(\theta, \sigma^2)-R_{\varphi 0}(\theta, \sigma^2)$. Alors

$$\Delta_\varphi(\theta, \sigma^2)=-\frac{2}{\sigma^2}\mathbb{E}[h(t, 1/s^2)\,q(\varphi^0(y)-\theta, c\,\varphi^0(y))]$$

$$+\frac{1}{\sigma^2}\mathbb{E}[h^2(t, 1/s^2)\,\bar{q}(c\,\varphi^0(y))]$$

$$=-2\,A_\varphi(\theta, \sigma^2)+B_\varphi(\theta, \sigma^2)$$

avec des notations évidentes. On a

$$A_\varphi(\theta, \sigma^2)=\frac{1}{\sigma^2}\mathbb{E}\left[h(t, 1/s^2)\sum_{i=1}^{k}(y_i-\theta_i)(d_i y_i+\beta_i)\right].$$

On peut remarquer que $\mathbb{E}[|h(t, 1/s^2)(y_i-\theta_i)(d_i y_i+\beta_i)|]$ est fini, pour $1\leqslant i\leqslant k$. En effet, par l'inégalité de Schwarz, on obtient

$$\begin{aligned}&\mathbb{E}[|h(t, 1/s^2)(y_i-\theta_i)(d_i y_i+\beta_i)|]\\&\quad\leqslant[\mathbb{E}[h^2(t, 1/s^2)(d_i y_i+\beta_i)^2]\,\mathbb{E}[(y_i-\theta_i)^2]]^{1/2}.\end{aligned}$$

Et

$$\begin{aligned}\mathbb{E}[h^2(t, 1/s^2)(d_i y_i+\beta_i)^2]&\leqslant\mathbb{E}\left[h^2(t, 1/s^2)\sum_{j=1}^{k}(d_j y_j+\beta_j)^2\right]\\&\leqslant\mathbb{E}[h^2(t, 1/s^2)\,\overline{v_\Theta}(qc\,\varphi^0(y))].\quad\square\end{aligned}$$

On voit alors, par une inégalité que l'on peut établir de la même manière que (3), que ce dernier terme est fini car $\mathbb{E}[th^2(t, 1/s^2)]$ est fini, par hypothèse.

On peut ainsi écrire, pour $1\leqslant i\leqslant k$,

$$\begin{aligned}&\mathbb{E}[h(t, 1/s^2)(y_i-\theta_i)(d_i y_i+\beta_i)]\\&=\int_{\mathbb{R}^{n-1}}\int_{-\infty}^{+\infty}h(t, 1/s^2)(y_i-\theta_i)(d_i y_i+\beta_i)\\&\quad\times\frac{e^{-(y_i-\theta_i)^2/2\sigma^2}}{\sqrt{2\pi}\,\sigma}\,dy_i\,N_{n-1}(\theta^i, \sigma^2 I_{n-1};\, dy^i)\end{aligned}$$

où $\theta^i=(\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_k, 0, \ldots, 0)$ et $y^i=(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$. Notons f_i la densité de la loi normale unidimensionnelle de moyenne θ_i et de variance σ^2.

$\mathbb{E}_{\theta,\sigma}[|h(t, 1/s^2|]$ est également fini. En effet, en appliquant à nouveau l'inégalité de Schwarz, on obtient

$$\begin{aligned}\mathbb{E}[|h(t, 1/s^2)|]&=\mathbb{E}[t^{1/2}|h(t, 1/s^2)|\,t^{1/2}]\\&\leqslant[\mathbb{E}[th^2(t, 1/s^2)]\,\mathbb{E}[1/t]]^{1/2}.\end{aligned}$$

Or $\mathbb{E}[th^2(t, 1/s^2)]$ est fini par hypothèse et $\mathbb{E}[1/t]$ est majoré, à une constante multiplicative près, par le moment d'ordre (-1) d'une loi du χ^2 décentré à $rg(b)$ degrés de liberté; ce moment est fini pour $rg(b) \geqslant 3$.

Compte tenu de cette finitude et des contraintes imposées par la régularité de φ, nous pouvons appliquer le lemme de Stein (voir en appendice). Il vient, pour $1 \leqslant i \leqslant k$,

$$\int_{-\infty}^{+\infty} h(t, 1/s^2)(d_i y_i + \beta_i)\frac{y_i - \theta_i}{\sigma^2} f_i(y_i)\, dy_i$$

$$= \int_{-\infty}^{+\infty} \frac{\partial}{\partial y_i}(h(t, 1/s^2)(d_i y_i + \beta_i))\, f_i(y_i)\, dy_i$$

$$= \int_{-\infty}^{+\infty} \{h'_1(t, 1/s^2)\, 2\, b_i y_i (d_i y_i + \beta_i) + d_i h(t, 1/s^2)\}\, f_i(y_i)\, dy_i$$

Donc

$$A_\varphi(\theta, \sigma^2) = \mathbb{E}\left[h(t, 1/s^2)\, tr(v_\Theta qc) + 2h'_1(t, 1/s^2) \sum_{i=1}^{m} y_i b_i (d_i y_i + \beta_i)\right]$$

$$= \mathbb{E}[h(t, 1/s^2)\, tr(v_\Theta qc) + 2h'_1(t, 1/s^2)\, \overline{(bv_\Theta qc)}\, (\varphi^0(y))]$$

D'autre part,

$$B_\varphi(\theta, \sigma^2) = \frac{1}{\sigma^2}\mathbb{E}\left[\left(\sum_{i=k+1}^{n} y_i^2\right)\frac{1}{(n-k)s^2} h^2(t, 1/s^2)\, \bar{q}(c\varphi^0(y))\right].$$

Pour $k+1 \leqq i \leqq n$, il vient, d'après le lemme de Stein,

$$\int_{-\infty}^{+\infty} \frac{1}{(n-k)s^2} h^2(t, 1/s^2)\, y_i \frac{y_i}{\sigma^2} f_i(y_i)\, dy_i$$

$$= \int_{-\infty}^{+\infty} \frac{\partial}{\partial y_i}\left(h^2(t, 1/s^2)\frac{y_i}{(n-k)s^2}\right) f_i(y_i)\, dy_i$$

$$= \int_{-\infty}^{+\infty} \left[\frac{-4y_i^2}{n-k} h(t, 1/s^2)\, h'_2(t, 1/s^2)\frac{1}{(n-k)s^6} + h^2(t, 1/s^2)\frac{1}{(n-k)s^2}\left(1 - \frac{2y_i^2}{(n-k)s^2}\right)\right] f_i(y_i)\, dy_i.$$

D'où

$$B_\varphi(\theta, \sigma^2) = \mathbb{E}\left[\bar{q}(c\varphi^0(y))\left(h^2(t, 1/s^2)\frac{n-k-2}{(n-k)s^2} - 4h(t, 1/s^2)\, h'_2(t, 1/s^2)\frac{1}{(n-k)s^4}\right)\right]$$

qui entraîne

$$\Delta_\varphi(\theta, \sigma^2) = -2\,\mathbb{E}[h(t, 1/s^2)\,tr(v_\Theta qc) + 2\,h'_1(t, 1/s^2)\,\overline{bv_\Theta qc}\,(\varphi^0(y))]$$

$$+\mathbb{E}\left[\bar{q}(c\,\varphi^0(y))\left(h^2(t, 1/s^2)\,\frac{n-k-2}{(n-k)\,s^2}\right.\right.$$

$$\left.\left. -4\,h(t, 1/s^2)\,h'_2(t, 1/s^2)\,\frac{1}{(n-k)\,s^4}\right)\right] \quad \square$$

3.3. Si h est la fonction de rétrécissement, on pose $r(x, z) = xzh(x, z)$. Les estimateurs à rétrécisseur s'écrivent alors

$$\varphi(y) = \left(id_\Theta - \frac{s^2\,r(t, 1/s^2)}{t}\,c\right)\varphi^0(y), \tag{4}$$

cette autre écriture étant classique (Baranchik [1964], Bock [1975], Efron et Morris [1976], Strawderman [1973]). On déduit alors de la Proposition 2:

Corollaire 4 : Si l'hypothèse H_1 est vérifiée et si φ est réguliére, $\mathrm{R}_\varphi(\theta, \sigma^2)$ est égal à

$$\mathrm{R}_{\varphi 0}(\theta, \sigma^2) - 2\,\mathbb{E}\left[\frac{s^2}{t}\left\{r(t, 1/s^2)\,tr(v_\Theta qc)\right.\right.$$

$$\left.\left. +2\,\overline{bv_\Theta qc}\,(\varphi^0(y))\left(r'_1(t, 1/s^2) - \frac{r(t, 1/s^2)}{t}\right)\right\}\right]$$

$$+\mathbb{E}\left[\frac{\bar{q}(c\,\varphi^0(y))}{(n-k)\,t^2}\{s^2\,r^2(t, 1/s^2)\,(n-k+2)\right.$$

$$\left. -4\,r(t, 1/s^2)\,r'_2(t, 1/s^2)\}\right].$$

4. Condition suffisante de domination

4.1. Soit $a = bv_\Theta qc$. Comme $\mathrm{Ker}\,b \subset \mathrm{Ker}\,a$, d'après le lemme 1 (remarque 10), le spectre de ab^- coincide avec celui de l'endomorphisme $v_\Theta qc$. Soit Λ le spectre de la restriction de ab^- à $\mathrm{Im}\,b$ [indépendant du choix de b^- d'après le lemme 1 (2)]. C'est aussi l'ensemble obtenu en ôtant $(k - rg(b))$ valeurs propres nulles au spectre de $v_\Theta qc$. Soit

$$\lambda_m = \inf \Lambda \qquad \text{et} \qquad \lambda_M = \sup \Lambda.$$

On a, d'après le lemme 1 (4), pour tout y de E tel que $t=\bar{b}(\varphi^0(y)) \neq 0$,

$$\lambda_m \leqq \frac{\overline{bv_\Theta qc}(\varphi^0(y))}{t} \leqq \lambda_M. \tag{5}$$

Sous les hypothèses H_1 et H_2, définissons

$$R_0 = 2\frac{tr\,(v_\Theta qc)-2\lambda_M}{\sigma_M}\,\frac{n-k}{n-k+2}$$

Si $\lambda_M \geqslant 0$, pour que R_0 soit positif, il est bien nécessaire que $\mathrm{rg}(b) \geqslant 3$, puisque $tr\,(v_\Theta qc)-2\lambda_M \leqslant (\mathrm{rg}(b)-2)\lambda_M$.

Il vient, en notant $r_i'^+ = \sup(0, r_i')$ et $r_i'^- = \sup(0, -r_i')$ $(i=1, 2)$,

PROPOSITION 5 : Si les hypothèses H_1 et H_2 sont vérifiées, si φ est régulière et si, pour presque tout (t, u),

$$\begin{aligned} 0 \leqslant\ & \sigma_M \frac{n-k+2}{n-k} r(t, u)[R_0 - r(t, u)] \\ & +4[t(\lambda_m r_1'^+(t, u) - \lambda_M r_1'^-(t, u)) \\ & + \frac{r(t, u)u}{n-k}(\sigma_m r_2'^+(t, u) - \sigma_M r_2'^-(t, u))], \end{aligned} \tag{6}$$

alors φ domine φ^0.

Démonstration : D'après (5),

$$\frac{r_1'(t, 1/s^2)\,\overline{bv_\Theta qc}(\varphi^0(y))}{t} \geqslant \lambda_m r_1'^+(t, 1/s^2) - \lambda_M r_1'^-(t, 1/s^2),$$

et, d'après (2),

$$\frac{r_2'(t, 1/s^2)\,\bar{q}(c\,\varphi^0(y))}{t} \geqslant \sigma_m r_2'^+(t, 1/s^2) - \sigma_M r_2'^-(t, 1/s^2).$$

Le résultat découle alors du Corollaire 4. □

Remarque 4 : Si $r(t, u)$ est, pour tout t, une fonction croissante de u et, pour tout u, une fonction croissante de t si $\lambda_m \geqslant 0$, décroissante de t si $\lambda_M \leqslant 0$, et si $0 \leqq r(t, u) \leqslant R_0$, alors, d'après la proposition ci-dessus, φ domine φ^0 (cette remarque n'apporte pas de solution au cas où $\lambda_m \lambda_M \leqslant 0$).

4.2. Étude de fonctions de rétrécissement particulières.

4.2.1. Considérons la condition

HYPOTHÈSE H_3 : Il existe une fonction g de $\mathbb{R}_+$ dans $\mathbb{R}_+$, telle que, pour tout (t, u) de $(\mathbb{R}_+)^2$, $r(t, u) = g(tu)$ et $\Delta = \{x;\, g(x) = R_0\}$ est de mesure nulle. De plus, $R_0 \lambda_m \lambda_M \neq 0$.

Ce cas, où la fonction de rétrécissement ne dépend que d'une seule variable, est celui le plus fréquemment envisagé dans la littérature

(BARANCHIK [1964], CELLIER et FOURDRINIER [1985], EFRON et MORRIS [1976], JAMES et STEIN [1961]).

Sous l'hypothèse H_3, on définit les fonctions F_1 et F_2 par

$$\forall x \notin \Delta, \quad F_i(x) = \frac{\lambda_i x^{\alpha_i} g(x) \operatorname{sgn}(R_0 - g(x))}{R_0 \, | R_0 - g(x) |^{a_i}} \qquad (i=1,\ 2)$$

avec

$$\lambda_1 = \lambda_m,\ \lambda_2 = \lambda_M,$$

$$a_1 = 1 + \frac{\sigma_m R_0}{(n-k)\lambda_m},\ a_2 = 1 + \frac{\sigma_M R_0}{(n-k)\lambda_M}$$

et $\alpha_i = \dfrac{(n-k+2) R_0 \sigma_M}{4(n-k)\lambda_i} \qquad (i=1,\ 2).$

PROPOSITION 6 : Si les hypothèses H_1, H_2, H_3 sont vérifiées, si φ est régulière et si, pour tout $x \notin \Delta$, on a

$$g'(x) \geqslant 0 \quad \text{et} \quad F_1'(x) \geqq 0, \qquad ou \qquad g'(x) \leqslant 0 \quad \text{et} \quad F_2'(x) \leqslant 0,$$

alors φ domine φ^0.

Remarque 5 : On déduit ainsi des résultats précédents une généralisation du théorème 2 de EFRON et MORRIS [1976] et de 10.2.1 (*b*) de JUDGE et BOCK [1978].

Démonstration : On a, pour $i=1,\ 2$ et pour tout $x \notin \Delta$,

$$F_i'(x) = \frac{\lambda_i x^{\alpha_i - 1} [\alpha_i g(x)(R_0 - g(x)) + R_0 x g'(x) + (a_i - 1) x g(x) g'(x)]}{R_0 \, | R_0 - g(x) |^{a_i + 1}}$$

Or $r_1'(t,\ u) = u g'(tu)$ et $r_2'(t,\ u) = t g'(tu)$. L'inégalité (6) s'écrit alors

$$\begin{cases} 0 \leqslant \sigma_M \dfrac{n-k+2}{n-k} g(tu)[R_0 - g(tu)] \\ \qquad + 4\left(t \lambda_m u g'(tu) + \dfrac{g(tu) u}{n-k} \sigma_m t g'(tu) \right) \qquad \text{si} \quad g'(tu) \geqq 0, \\ 0 \leqslant \sigma_M \dfrac{n-k+2}{n-k} g(tu)[R_0 - g(tu)] \\ \qquad + 4\left(t \lambda_M u g'(tu) + \dfrac{g(tu) u}{n-k} \sigma_M t g'(tu) \right) \quad \text{sinon.} \end{cases}$$

La proposition 5 permet alors de conclure. □

4.2.2. **Cas particulier** A la suite de JAMES et STEIN [1961], un cas fréquemment envisagé est celui où la fonction r, définie en 3.3, est constante (voir VINOD et ULLAH [1981] et STEIN [1981]). On obtient ainsi la sous-classe de (4)

$$(7) \qquad \varphi_\rho(y) = \left(\mathrm{id}_\Theta - \rho \frac{s^2}{t} c \right) \varphi^0(y), \qquad \rho \in \mathbb{R}_+.$$

Lorsque $b = v_\Theta^{-1}$, ces estimateurs constituent une généralisation directe de l'estimateur « historique » de James et Stein [1961] (pour lequel $c = \mathrm{id}_\Theta$). Les variables t et s^2 étant indépendantes, les conditions de régularité sont satisfaites si $\mathbb{E}(1/t)$, $\mathbb{E}(s^2)$ et $\mathbb{E}(s^4)$ sont finis. Nous avons vu dans la démonstration de la proposition 2 que, pour $rg(b) \geqq 3$, la première de ces conditions est remplie; les deux autres découlent du fait que $(n-k)s^2/\sigma^2$ suit une loi du χ^2 à $(n-k)$ degrés de liberté. Il vient, d'après le corollaire 4,

$$R_\varphi(\theta, \sigma^2) = R_{\varphi^0}(\theta, \sigma^2)$$

$$-2\rho\,\mathbb{E}\left[\frac{tr(v_\Theta qc)\,s^2}{t} - 2\frac{b\overline{v_\Theta}\,qc(\varphi^0(y))\,s^2}{t^2}\right]$$

$$+\rho^2\,\mathbb{E}\left[\frac{\bar{q}(c\,\varphi^0(y))\,s^2}{t^2}\right]\frac{n-k+2}{n-k}.$$

A (θ, σ^2) fixé, $R_\varphi(\theta, \sigma^2)$, fonction de ρ, décroît avant

$$\rho_m(\theta, \sigma^2) = \frac{n-k}{n-k+2}$$

$$\times\frac{tr(v_\Theta qc)\,\mathbb{E}[s^2/t] - 2\,\mathbb{E}[\overline{bv_\Theta\,qc}(\varphi^0(y))\,s^2/t^2]}{\mathbb{E}[\bar{q}(c\,\varphi^0(y))\,s^2/t^2]}$$

et croît après cette valeur; on vérifie immédiatement que

$$\forall(\theta, \sigma^2), \quad \frac{R_0}{2} \leqq \rho_m(\theta, \sigma^2) \leqq \frac{R_1}{2},$$

où $R_1 = 2\dfrac{tr(v_\Theta qc) - 2\lambda_m}{\sigma_m}\,\dfrac{n-k}{n-k+2}$, si $\sigma_m \neq 0$ (dans le cas contraire, on remplace σ_m par la plus petite valeur propre non nulle de ${}^t cqcb^-$). Remarquons que, si R_1 est négatif, les coefficients ρ optimaux sont négatifs, ce qui est exclu. Nous supposerons donc ci-dessous que R_1 est positif [on peut s'y ramener en changeant éventuellement c en $(-c)$]. Alors

Proposition 7 : Parmi les estimateurs φ_ρ, ceux tels que ρ soit compris entre $\sup\left(0, \dfrac{R_0}{2}\right)$ et $\dfrac{R_1}{2}$ forment une classe complète.

Si $\rho \geqq R_1$, φ^0 domine φ_ρ. Quand $R_0 \geqq 0$, φ_ρ domine φ^0 pour $\rho \in [0, R_0]$.

Remarque 6 : Lorsque $R_0 R_1 \leqq 0$, on ne peut rien dire sur la différence des risques entre φ_ρ et φ^0, pour $\rho \in [0, R_1/2]$. Pour certaines valeurs de (θ, σ^2), φ^0 peut donc avoir un risque inférieur à φ_ρ, quel que soit ρ.

Remarque 7 : Au contraire de James et Stein [1961] et de Stein [1981], l'expression de $\rho_m(\theta, \sigma^2)$ montre qu'il n'existe généralement pas un estimateur φ_ρ qui soit optimal uniformément en (θ, σ^2) (Stein [1981] considère le cas particulier où $b = v_\Theta^{-1} c^2$).

5 Estimateurs contrôlés différentiables

5.1. Les conditions que nous avons pu établir dans le paragraphe précédent apparaissent soit trop générales et donc difficilement interprétables (proposition 5), soit trop particulières (propositions 6 et 7) du fait des hypothèses restrictives sur la fonction de rétrécissement. Les contraintes que nous allons introduire ci-dessous sont plus fortes que celles de la proposition 5 mais elles présentent l'avantage de caractériser l'évolution de la fonction de rétrécissement, en nécessitant une réduction du cadre algébrique du modèle. On impose en effet à l'endomorphisme $v_{\Theta} qc$ d'avoir toutes ses valeurs propres positives ou nulles, autrement dit,

HYPOTHÈSE H_4 : $\lambda_m \geqslant 0$.

Remarquons que cette hypothèse impose à c d'avoir toutes ses valeurs propres positives ou nulles.

La définition ci-dessous est introduite dans CELLIER, FOURDRINIER et ROBERT [1986].

DÉFINITION 1 : Soit $\lambda \in \mathbb{R}_+$. Une application f de $\mathbb{R}_+$ dans $\mathbb{R}_+$ est dite *contrôlée à hauteur* λ si $t^{\lambda} f(t)$ est croissante.

5.2. On obtient le résultat suivant.

PROPOSITION 8 : Si $rg(b) \geqslant 3$ et si les hypothèses H_1, H_2 et H_4 sont vérifiées, une condition suffisante pour que l'estimateur φ domine φ^0 est qu'il existe deux nombres réels, μ_1 et μ_2, tels que, pour tout $u > 0$, $h(., u)$ soit contrôlée à hauteur μ_1, pour tout $t > 0$, $h(t, .)$ soit contrôlée à hauteur μ_2 et que $tu\, h(t, u)$ soit uniformément majorée par

$$R_2 = 2\frac{tr(v_{\Theta} qc) - 2\mu_1 \lambda_M}{\sigma_M} \frac{n-k}{n-k+4\mu_2-2}$$

Remarque 8 : Fondé sur ALAM [1973] qui a introduit le contrôle par des fonctions puissances, le résultat ci-dessus est très proche de CELLIER, FOURDRINIER et ROBERT [1986], où les hypothèses algébriques sur b, c sont plus contraignantes qu'ici, mais où, par contre, le rétrécisseur n'est pas supposé différentiable.

Remarque 9 : On a $R_2 \leqslant R_0$; de plus, $R_2 = R_0$ dans le cas où μ_1 et μ_2 valent 1, valeur minimale pour que les hypothèses de la proposition 8 soient compatibles (voir CELLIER, FOURDRINIER et ROBERT [1986]).

Démonstration : Établissons que les conditions de la proposition 5 sont remplies.

L'hypothèse sur la majoration de $tu\, h(t, u)$ suffit à montrer les conditions de régularité. Pour que $\mathbb{E}[th^2(t, 1/s^2)]$ soit fini, il suffit en effet que $\mathbb{E}[s^4/t]$

soit fini et la finitude de $\mathbb{E}\left[\frac{t}{s^2}h^2(t, 1/s^2)\right]$ découle de celle de $\mathbb{E}[s^2/t]$ (voir section 4.2.2). De plus, cette hypothèse permet d'intégrer par parties les intégrales auxquelles on a appliqué le lemme de Stein dans la démonstration de la proposition 2 et d'en déduire que

$$\mathbb{E}[th'_1(t, 1/s^2)] \qquad \text{et} \qquad \mathbb{E}\left[\frac{t}{s^4}h(t, 1/s^2)\,h'_2(t, 1/s^2)\right]$$

sont finies.

On a

$r(t, u)\,t^{\mu_1-1}u^{\mu_2-1}$ est croissante séparément par rapport à t et u donc

$$r'_1(t, u) \geqslant -(\mu_1-1)\frac{r(t, u)}{t} \qquad \text{et} \qquad r'_2(t, u) \geqslant -(\mu_2-1)\frac{r(t, u)}{u}.$$

D'où

$$\begin{aligned}\lambda_m r_1'^{+}(t, u)-\lambda_M r_1'^{-}(t, u) &= \inf(\lambda_m r'_1(t, u), \lambda_M r'_1(t, u)) \\ &\geqslant -\lambda_M(\mu_1-1)\frac{r(t, u)}{t},\end{aligned}$$

$$\begin{aligned}\sigma_m r_2'^{+}(t, u)-\sigma_M r_2'^{-}(t, u) &= \inf(\sigma_m r'_2(t, u), \sigma_M r'_2(t, u)) \\ &\geqslant -\sigma_M(\mu_2-1)\frac{r(t, u)}{u},\end{aligned}$$

La proposition 5 nous conduit alors au résultat. □

Nous rappelons dans le lemme ci-dessous quelques propriétés algébriques utiles.

LEMME 1 : Soit a et b deux formes bilinéaires sur un espace vectoriel de dimension n; b est symétrique positive; on suppose que $\operatorname{Ker} b \subset \operatorname{Ker} a$ et $\operatorname{Ker} b \subset \operatorname{Ker} {}^t a$. Alors

(1) $\operatorname{Im} a \subset \operatorname{Im} b$; $\operatorname{Im} {}^t a \subset \operatorname{Im} b$.

(2) Les restrictions à $\operatorname{Im} b$ de ab^- et de ${}^t ab^-$ ne dépendent pas du choix de l'inverse généralisé b^- de b.

(3) Les spectres de ab^- et de ${}^t ab^-$ coincident et ne dépendent pas du choix de b^-; ce spectre commun se partitionne en Λ, spectre de la restriction à $\operatorname{Im} b$, et en $n-rg(b)$ valeurs propres nulles.

(4) $$\operatorname*{Inf}_{x \notin \operatorname{Ker} b} \frac{\bar{a}(x)}{\bar{b}(x)} = \inf \Lambda \quad \text{et} \quad \sup_{x \notin \operatorname{Ker} b} \frac{\bar{a}(x)}{\bar{b}(x)} = \sup \Lambda.$$

La démonstration de ce lemme est élémentaire. On peut, par exemple, considérer un couple de bases duales $(e_1, \ldots, e_n)$ et $(e_1^*, \ldots, e_n^*)$, où $e_1, \ldots, e_m$ sont b-orthonormaux et $e_{m+1}, \ldots, e_n$ engendrent $\operatorname{Ker} b$ [avec $m = rg(b)$]. De plus, pour (4), il est utile de choisir la base $(e_1, \ldots, e_n)$ orthogonale relativement à la forme bilinéaire polaire de $\bar{a}$, ce qui est toujours réalisable.

Remarque 10 : Le lemme ci-dessus s'applique en particulier pour $a = bd$, où d est un endomorphisme de V tel que $\operatorname{Ker} b \subset \operatorname{Ker} d$. On vérifie élémentairement, en utilisant les bases précédemment définies, que le spectre de bdb^- coincide avec le spectre de d.

Le lemme ci-dessous est démontré dans STEIN [1981] (p. 1136) :

LEMME 2 (Stein) : Soit Y une variable aléatoire suivant une loi normale de moyenne θ et de variance σ^2 et g une fonction de $\mathbb{R}$ dans $\mathbb{R}$ dérivable telle que $\mathbb{E}\,|g'(\mathrm{Y})|$ soit fini. Alors

$$\mathbb{E}\left[\frac{\mathrm{Y}-\theta}{\sigma^2} g(\mathrm{Y})\right] = \mathbb{E}\,[g'(\mathrm{Y})].$$

• Références bibliographiques

ALAM, K. (1973). — « A Family of Admissible Minimax Estimators of the Mean of a Multivariate Normal Distribution », *Ann. Stat.*, 1, (3), p. 517-525.

BARANCHIK, A. (1964). — « Multiple Regression and Estimation of the Mean Vector of a Multivariate Normal Distribution », *Stanford Univ.-Tech. Report*, n° 51.

BERGER, J. (1976). — « Minimax Estimation of a Multivariate Normal Mean under Arbitrary Quadratic Loss », *J. Mult. Anal.*, 6, p. 256-264.

BERGER, J. (1980). — « A Robust Generalized Bayes Estimator and Confidence Regions for a Multivariate Normal Mean », *Ann. Stat.*, 8, (4), p. 716-761.

BERGER, J. et BOCK, M. (1976). — « Eliminating Singularities of Stein-Type Estimators of Location Vectors », *J.R.S.S.*, 39, (2), p. 166-170.

BOCK, M. (1975). — « Minimax Estimators of the Mean of a Multivariate Normal Distribution », *Ann. Stat.*, 3, (1), p. 209-218.

CELLIER, D. et FOURDRINIER, D. (1985). — « Estimateurs à rétrécisseurs de la moyenne d'une loi normale multidimensionnelle pour un coût quadratique général », *Statistique et Analyse des Données*, 10, (3), p. 26-41.

CELLIER, D., FOURDRINIER, D. et ROBERT, C. (1986). — « Estimateurs à rétrécisseur contrôlé », *Univ. de Rouen-Doc. Travail*, 1986-01.

EFRON, K. et MORRIS, B. (1976). — « Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution », *Ann. Stat.*, 4, (1), p. 11-21.

JAMES, W. et STEIN, C. (1961). — « Estimation with Quadratic Loss », *Proc. Fourth Berk. Symp. Math. Stat. Prob.*, 1, p. 361-379.

JUDGE, G. et BOCK, M. (1978). — *The Statistical Implications of Pre-test and Stein-rule estimators in Econometrics*, North-Holland.

STEIN, C. (1956). — « Inadmissibility of the Usual Estimator of the Mean of a Multivariate Normal Distribution », *Proc. Third Berk. Symp.*, 1, p. 197-206.

STEIN, C. (1973). — « Estimation of the Mean of a Multivariate Normal Distribution », *Proc. Prague Symp. Asymp. Stat.*, 1, p. 345-381.

STEIN, C. (1981). — « Estimation of the Mean of a Multivariate Normal Distribution », *Ann. Stat.*, 9, (6), p. 1135-1151.

STRAWDERMAN, W. (1973). — « Proper Bayes Minimax Estimators of the Multivariate Normal Mean Vector for the Case of Common Unknown Variances », *Ann. Stat.*, 1, (6), p. 1189-1194.

VINOD, H. et ULLAH, A. (1981). — *Recent Advances in Regression Methods*, M. Decker.