Statistique et analyse des données

Dominique Cellier
Dominique Fourdrinier

Estimateurs à rétrécisseurs de la moyenne d'une loi normale multidimensionnelle, pour un coût quadratique général

Statistique et analyse des données, tome 10, nº 3 (1985), p. 26-41
<http://www.numdam.org/item?id=SAD_1985__10_3_26_0>

Statistique et Analyse des Données
1985 - vol.10 n°3 pp 26-41

ESTIMATEURS A RETRECISSEURS DE LA MOYENNE D'UNE LOI NORMALE MULTIDIMENSIONNELLE, POUR UN COUT QUADRATIQUE GENERAL.

Dominique CELLIER et Dominique FOURDRINIER

Laboratoire de Calcul des Probabilités
et Statistique (UA CNRS 759)
Université de Rouen
(BP 67. 76130 Mont Saint Aignan)

__Résumé__ . *On observe, dans un espace vectoriel E de dimension finie sur $\mathbb{R}$, un vecteur aléatoire, régi par la loi normale de moyenne inconnue appartenant à un sous espace vectoriel de E, connu, de dimension supérieure ou égale à 3 et de variance connue, ou bien connue à un facteur multiplicatif près.*

Nous étudions une classe d'estimateurs à rétrécisseur de la moyenne, qui dominent l'estimateur du maximum de vraisemblance, relativement à un coût quadratique associé à une forme quadratique positive quelconque.

__Abstract__ . *A random normal vector is observed in a finite dimensional real vector space E. Its mean is unknown but belongs to a known subspace of E of dimension ≥ 3. Its variance is known or unknown up to a multiplicative factor.*

We study a class of shrinkage estimators of the mean which dominate the maximum likelihood estimator with respect to a general quadratic loss.

__Mots clés__ . *Modèle statistique linéaire , Coût quadratique, Estimateurs de James-Stein, Fonction de rétrécissement.*

__Indices de classification STMA__ : *04-160, 04-000.*

Manuscrit reçu le 20 décembre 1985
révisé le 10 avril 1986

I - MODELE .

On observe, dans un espace vectoriel E, de dimension finie n sur $\mathbb{R}$, un vecteur aléatoire y, régi par la loi normale de moyenne θ et de variance $\sigma^2 v$, où

(i) θ est inconnu et appartient à un sous-espace vectoriel Θ de E connu de dimension k,

(ii) v est une forme bilinéaire symétrique sur le dual E^* de E, définie positive et connue (il sera pratique de la considérer comme un isomorphisme linéaire de E^* sur E),

(iii) σ^2 (> 0) est connu ou inconnu suivant le problème étudié (et s'il est inconnu, nous supposons $k < n-1$).

Nous nous proposons d'estimer θ; tout estimateur est donc une application φ de E dans Θ; le critère de comparaison des estimateurs adopté est le risque, relativement à un coût quadratique défini par une forme bilinéaire positive (mais non nécessairement définie positive) q, fixée dans toute la suite de l'étude.

De manière générale dans tout ce texte, si w est une forme bilinéaire symétrique, on note $\bar{w}$ la forme quadratique associée; ainsi le coût subi quand on fournit θ' alors que la "vraie valeur" est θ s'écrit

$$\sigma^{-2}\ \bar{q}(\theta'-\theta) .$$

Nous noterons φ^0 l'estimateur du maximum de vraisemblance de θ, c'est à dire le projecteur v^{-1}-orthogonal de E sur Θ (appelé parfois l'estimateur des moindres carrés généralisé).

Sa variance est une forme bilinéaire symétrique sur Θ^* qu'on notera v_Θ (on remarque que v_Θ^{-1} est la restriction de v^{-1} à Θ).

II - ESTIMATEURS DE JAMES-STEIN GENERALISES .

2.1. Les estimateurs φ auxquels nous nous intéressons ici sont construits à l'aide :

- d'une forme bilinéaire symétrique définie positive b sur Θ,
- d'un endomorphisme c de Θ,
- d'une application h de $\mathbb{R}_+$ dans $\mathbb{R}_+$.

Si σ^2 est connu, φ s'exprime sous la forme

$$\varphi(y) = \varphi^0(y) - h\left[\sigma^{-2}\ \bar{b}\left(\varphi^0(y)\right)\right].c\left(\varphi^0(y)\right) \tag{2.1}$$

Si σ^2 est inconnu, nous remplaçons, dans la formule (2.1), σ^2 par son estimation sans biais usuelle, i.e. $\overline{v^{-1}}(y - \varphi^o(y))/(n-k)$, et ainsi φ s'exprime sous la forme

$$\text{(2.2)} \qquad \varphi(y) = \varphi^o(y) - h\left[\frac{n-k}{\overline{v^{-1}}\left(y - \varphi^o(y)\right)} . \bar{b}\left(\varphi^o(y)\right)\right] . c\left(\varphi^o(y)\right)$$

2.2. Risque des estimateurs de James-Stein généralisés .

2.2.1. Dans ce paragraphe, nous posons, pour toute application mesurable f de E dans $\mathbb{R}$ et pour tout θ de Θ,

$$\mathbb{E}_\theta[f(y)] = \int_E f(y)\ \mathcal{N}(\theta, \sigma^2 v; dy) \quad ,$$

et pour tout y de E,

$$\tilde{h}(y) = h\left[\sigma^{-2}\ \bar{b}\left(\varphi^o(y)\right)\right] \qquad \text{si } \sigma^2 \text{ est connu}$$

$$\tilde{h}(y) = h\left[\frac{n-k}{\overline{v^{-1}}\left(y - \varphi^o(y)\right)} . \sigma^{-2}\ \bar{b}\left(\varphi^o(y)\right)\right] \qquad \text{si } \sigma^2 \text{ est inconnu}$$

2.2.2. Calcul du risque .

Notons, pour tout $\theta \in \Theta$, $r(\varphi,\theta)$ la valeur en θ de la fonction de risque de φ. Nous avons

$$r(\varphi,\theta) = \sigma^{-2}\ \mathbb{E}_\theta\left[\bar{q}\left\{\varphi^o(y) - \tilde{h}(y).c\left(\varphi^o(y)\right) - \theta\right\}\right]$$

$$= \sigma^{-2}\ \mathbb{E}_\theta\left[\bar{q}\left(\varphi^o(y) - \theta\right) - 2q\left(\varphi^o(y) - \theta, \tilde{h}(y).c\left(\varphi^o(y)\right)\right) + \bar{q}\left(\tilde{h}(y).c\left(\varphi^o(y)\right)\right)\right]$$

et en particulier

$$r(\varphi^o,\theta) = \sigma^{-2}\ \mathbb{E}_\theta\left[\bar{q}\left\{\varphi^o(y) - \theta\right\}\right] \qquad ;$$

il est classique que ce dernier risque vaut $tr(v_\theta q)$ et est donc fini.

2.2.3. Condition pour que le risque soit fini .

Lemme

Une condition nécessaire et suffisante pour que le risque en θ d'un estimateur de type (2.1) ou de type (2.2) soit fini est que

$$E_\theta\left[\ \bar{q}\left\{\ \tilde{h}(y).c\left(\varphi^o(y)\right)\right\}\ \right] \quad < \quad +\infty .$$

démonstration

Elle résulte immédiatement du calcul du risque *2.2.2.*, de la finitude de $r(\varphi^o,\theta)$ et de l'inégalité de Schwarz.

2.2.4. Dans toute la suite, nous supposerons que b, c et h vérifient les propriétés suivantes :

(P1) il existe une base $(e_1,\dots,e_n)$ de E, orthonormale pour v^{-1} et dont les k premiers éléments constituent une base de Θ orthogonale pour q et pour b, et dans laquelle l'endomorphisme c diagonalise

(P2) les valeurs propres de c sont toutes positives ou nulles

(P3) l'endomorphisme de Θ^* qc^2b^{-1} n'est pas identiquement nul

(P4) pour tout θ de Θ, $E_\theta\left[\ \bar{q}\left\{\ \tilde{h}(y).c\left(\varphi^o(y)\right)\right\}\ \right] \quad < \ +\infty$.

Remarques sur ces propriétés :

a) l'existence d'une base $(e_1,\dots,e_n)$, orthonormale pour v^{-1} et telle que $(e_1,\dots,e_k)$ soit orthogonale pour q, est une propriété élémentaire d'algèbre euclidienne; (P1) est donc une condition sur le choix de b et c

b) il est élémentaire que l'endomorphisme qc^2b^{-1} est diagonalisable et que ses valeurs propres sont positives ou nulles; la propriété (P3) équivaut donc à

$$(P'3) \qquad \text{pgvp}\ (qc^2b^{-1}) \ > \ 0$$

où pgvp désigne la plus grande valeur propre

c) la propriété (P4) assure la finitude du risque (c.f. *2.2.3.*) et la validité des calculs effectués dans la démonstration du théorème.

II - Théorème .

3.1. Nous introduisons la constante réelle α suivante

$$\alpha = \frac{2[\text{tr}(v_\Theta qc) - 2\ \text{pgvp}(v_\Theta qc)]}{\text{pgvp}\left(qc^2b^{-1}\right)}$$

<u>Théorème</u> :

Sous les conditions (P1), (P2), (P3) et (P4), une condition suffisante pour que, σ^2 étant connu (respectivement inconnu), l'estimateur de James-Stein généralisé φ défini par (2.1) (respectivement (2.2)) ci-dessus soit uniformément de risque inférieur ou égal à celui de l'estimateur du maximum de vraisemblance φ^0 est que l'application $t \to th(t)$ soit croissante et majorée par α (respectivement par $\alpha.\frac{n-k}{n-k+2}$).

Pour que, de plus, φ défini par (2.1) (respectivement (2.2)) soit uniformément de risque strictement inférieur à celui de φ^0 , il suffit qu'il existe deux nombres réels t_1 et t_2 tels que

$$0 < t_1 < t_2 \quad \text{et} \quad 0 < t_1 h(t_1) \leqslant t_2 h(t_2) < \alpha$$

$$\text{(resp.} \quad 0 < t_1 < t_2 \quad \text{et} \quad 0 < t_1 h(t_1) \leqslant t_2 h(t_2) < \alpha.\frac{n-k}{n-k+2} \text{).}$$

3.2. <u>Comparaison de ce théorème avec certains résultats connus</u> .

Des généralisations des résultats initiaux de James et Stein ([14] et [10]) ont été effectuées dans de nombreuses directions, incluant en particulier des recherches explicites d'estimateurs de Bayes relativement à des lois a priori données (voir par exemple [5]) et aussi, depuis une dizaine d'années, des études portant sur des lois non normales, mais dotées de certaines propriétés de symétrie (voir [2],[7],[8],[13]). Nous détaillons ici uniquement la comparaison de notre résultat avec ceux qui portent sur des lois normales et pour lesquels le coût est défini à l'aide d'une forme quadratique arbitraire; une synthèse d'études effectuées en prenant pour coût l'application $(\theta_1,\theta_2) \to \sigma^{-2}\ \overline{v_\Theta^{-1}}(\theta_1-\theta_2)$ peut être trouvée dans [1].

De manière générale, moins les hypothèses sur les transformations algébriques (b et c) que l'on fait subir aux données sont contraignantes, plus les hypothèses analytiques portant sur la fonction de rétrécissement (h) ont à être restrictives.

Un cadre particulièrement restrictif pour le choix de b est celui où on dispose de peu d'information sur la variance des observations. C'est le cas dans les travaux (voir en particulier [6] et [9]) où on considère n observations indépendantes dans $\mathbb{R}^p$, p>1, la variance commune de ces n observations étant totalement inconnue. A notre connaissance, dans ce cas, la fonction de rétrécissement dépend toujours de $\varphi^0(y)$ par l'intermédiaire de la valeur en $\varphi^0(y)$ de la forme quadratique inverse de la valeur estimée au sens de Wishart.

De même, parmi les travaux situés au même niveau d'information que nous sur la variance, mais pour lesquels le choix des outils algébriques (b et c) est plus restrictif que le nôtre, on peut citer, outre de nombreux résultats pour lesquels $b=v_\Theta^{-1}$ et c est l'identité, [3] où $b=v_\Theta^{-1}q^{-1}v_\Theta^{-1}$ et $c=q^{-1}v_\Theta^{-1}$.

Le choix des conditions (P1),(P2) et (P3) que nous avons adoptées pour b et c est essentiellement celui des travaux de Berger [4], repris dans [11]. Nous nous distinguons de ces auteurs en ceci que nous ne supposons pas la fonction de rétrécissement différentiable ni même continue. Cette généralisation a l'avantage de permettre d'inclure dans le champ de cette étude des estimateurs du type pré-test (voir [12]); par contre elle introduit sur l'application $t \to th(t)$ une condition de majoration plus contraignante que celle proposée dans [11].

Enfin, parmi les travaux pour lesquels le choix des outils algébriques est moins restrictif que le nôtre, il y a lieu de citer [4], pour lequel l'hypothèse (P1) est évitée. Il ne nous a pas été possible, pour le moment, de généraliser, au cas où h n'est pas différentiable, les techniques de calcul figurant dans [4]. Plus généralement encore, en 1981, Stein [15] étudie des fonctions de rétrécissement f définies "directement" sur Θ, c'est à dire ne factorisant pas à travers une forme quadratique définie sur cet espace; il introduit alors des hypothèses portant sur le gradient de Log f .

3.3. Remarques .

a) Il sera pratique, dans la démonstration de ce théorème, de recourir à l'expression matricielle, à l'aide de la base $(e_1,\dots,e_n)$ dont la propriété (P1) assure l'existence, des différentes formes bilinéaires symétriques et endomorphismes utilisés.

Bien sûr les matrices de v et v^{-1} sont égales à I_n , matrice unité d'ordre n; les matrices de q, b et c, diagonales, sont notées respectivement

$$Q = \begin{pmatrix} q_1 & & & \\ & . & 0 & \\ & & . & \\ & 0 & . & \\ & & & q_k \end{pmatrix}, \quad B = \begin{pmatrix} b_1 & & & \\ & . & 0 & \\ & & . & \\ & 0 & . & \\ & & & b_k \end{pmatrix} \quad \text{et } C = \begin{pmatrix} c_1 & & & \\ & . & 0 & \\ & & . & \\ & 0 & . & \\ & & & c_k \end{pmatrix}$$

où, pour tout i $(1 \leqslant i \leqslant k)$, $q_i \geqslant 0$, $b_i > 0$ et $c_i \geqslant 0$ (d'après (P2)).
Alors

$$\mathrm{pgvp}(v_\Theta qc) = \max_{1\leqslant i\leqslant k} q_i\, c_i$$

$$\mathrm{tr}(v_\Theta qc) = \sum_{i=1}^{k} q_i\ c_i$$

$$\mathrm{pgvp}(qc^2b^{-1}) = \max_{1\leqslant i\leqslant k} q_i\ c_i^2\ b_i^{-1} \quad .$$

(P3) exprime donc qu'il existe i ($1 \leqslant i \leqslant k$) tel que $q_i > 0$ et $c_i > 0$.
Nous noterons d'autre part $y_1,\ldots,y_n$ les composantes de y dans la base $(e_1,\ldots,e_n)$. Alors

$$\varphi^0(y) = (y_1,\ldots,y_k)$$

$$\overline{v^{-1}}\left(y-\varphi^0(y)\right) = \sum_{j=k+1}^{n} y_j^2$$

$$\varphi(y) = (\varphi_1(y),\ldots,\varphi_k(y))$$

où, pour tout i ($i \leqslant i \leqslant k$), φ_i est donné par la formule

$$\varphi_i(y) = (1 - h(t).c_i)y_i$$

où, dans le cas où σ^2 est connu,

$$t = \sigma^{-2} \sum_{j=1}^{k} b_j\ y_j^2 \tag{3.1}$$

et, dans le cas où σ^2 est inconnu,

$$t = \frac{n-k}{\displaystyle\sum_{j=k+1}^{n} y_j^2} \cdot \sum_{j=1}^{k} b_j\ y_j^2 \tag{3.2}$$

b) Pour qu'il existe au moins un estimateur de James-Stein généralisé, satisfaisant aux conditions de ce théorème, et non identique à φ^0, il faut que l'application h puisse être prise non identiquement nulle, et donc que α soit strictement positif, autrement dit que

$$\mathrm{tr}(v_\Theta qc) > 2\ \mathrm{pgvp}(v_\Theta qc)$$

ou encore que

$$\sum_{i=1}^{k} q_i\, c_i \;>\; 2 \max_{1\leqslant i\leqslant k} q_i\, c_i \qquad ;$$

cette condition n'est bien sûr réalisable que si $k > 2$.

3.4. <u>Démonstration du théorème</u> .

Fixons θ dans Θ. Nous noterons, pour toute application mesurable f de $\mathbb{R}^n$ dans $\mathbb{R}$,

$$\mathbb{E}_\theta[f(y_1,\dots,y_n)] = \frac{1}{\left(\sqrt{2\pi}\,\sigma\right)^n}\cdot\int_{\mathbb{R}^n} f(y_1,\dots,y_n)\exp\left(-\frac{\sum_{i=1}^{n}(y_i-\theta_i)^2}{2\sigma^2}\right)dy_1\dots dy_n$$

où les θ_i sont les composantes de θ dans la base $(e_1,\dots,e_n)$.

D'autre part, nous posons

$$\Delta_\varphi(\theta) = r(\varphi^0,\theta) - r(\varphi,\theta)\ .$$

L'expression du risque de φ en θ, vue en *2.2.2.*, et le recours à l'expression matricielle (c.f. remarque a) ci-dessus) nous permettent d'écrire

$$\Delta_\varphi(\theta) = A_\varphi(\theta) - B_\varphi(\theta)$$

où

$$A_\varphi(\theta) = \frac{2}{\sigma^2}\,\mathbb{E}_\theta\left[\sum_{i=1}^{k} q_i\, c_i\, y_i\,(y_i-\theta_i)\, h(t)\right]$$

$$B_\varphi(\theta) = \frac{1}{\sigma^2}\,\mathbb{E}_\theta\left[\sum_{i=1}^{k} q_i\, c_i^2\, y_i^2\, h^2(t)\right]$$

Notre démonstration s'appuiera alors sur une minoration de $A_\varphi(\theta)$ et une majoration de $B_\varphi(\theta)$.

Minoration de $A_\varphi(\theta)$.

Pour tout i ($1 \leqslant i \leqslant k$), posons

$$\alpha_i = \mathbb{E}_\theta[h(t)y_i(y_i-\theta_i)] \quad .$$

Nous intégrons par parties relativement à la variable y_i , en utilisant le lemme 1 (c.f. appendice) pour

$$g(y_i) = y_i\ h(t) \quad \text{et} \quad f(y_i) = \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{(y_i-\theta_i)^2}{2\sigma^2}\right) \quad ;$$

Comme $k \geqslant 3$, on a p.s. pour tout i ($1 \leqslant i \leqslant k$), $\sum_{j=1}^{k} b_j y_j^2 - b_i y_i^2 > 0$. Il résulte alors du lemme 2 que

$$\alpha_i \geqslant \sigma^2\ \mathbb{E}_\theta\left[h(t)\left(1 - \frac{2b_i y_i^2}{\sum_{j=1}^{k} b_j\ y_j^2}\right)\right]$$

et ainsi

$$A_\varphi(\theta) \geqslant \mathbb{E}_\theta\left[h(t)\ 2\left(\sum_{i=1}^{k} q_i\ c_i - 2\ \frac{\sum_{i=1}^{k} q_i\ c_i\ b_i\ y_i^2}{\sum_{j=1}^{k} b_j\ y_j^2}\right)\right]$$

$$\geqslant 2\left\{\sum_{i=1}^{k} q_i\ c_i - 2 \max_{1\leqslant i\leqslant k} q_i\ c_i\right\} \mathbb{E}_\theta[h(t)] \quad .$$

Majoration de $B_\varphi(\theta)$.

$$B_\varphi(\theta) \quad = \quad \sigma^{-2} \; \mathbb{E}_\theta\left[th^2(t) . \frac{\sum_{i=1}^{k} b_i \; y_i^2}{t} . \frac{\sum_{i=1}^{k} q_i \; c_i^2 \; y_i^2}{\sum_{i=1}^{k} b_i \; y_i^2} \right]$$

$$\leqslant \quad \max_{1 \leqslant i \leqslant k} \; q_i c_i^2 b_i^{-1} \; . \sigma^{-2} \; \mathbb{E}_\theta \left[th^2(t) . \frac{\sum_{i=1}^{k} b_i \; y_i^2}{t} \right] .$$

Dans le cas où σ^2 est connu, nous avons $\dfrac{\sum_{i=1}^{k} . b_i \; y_i^2}{t} = \sigma^2$ et ainsi

$$B_\varphi(\theta) \quad \leqslant \quad \max_{1 \leqslant i \leqslant k} \; q_i c_i^2 b_i^{-1} \quad \mathbb{E}_\theta \left[t \; h^2(t) \right] \quad .$$

Dans le cas où σ^2 est inconnu, nous avons $\dfrac{\sum_{i=1}^{k} b_i \; y_i^2}{t} = \dfrac{\sum_{j=k+1}^{n} y_j^2}{n-k}$ donc

$$B_\varphi(\theta) \quad \leqslant \quad \max_{1 \leqslant i \leqslant k} \; q_i c_i^2 b_i^{-1} \; . \frac{1}{n-k} \; . \sigma^{-2} \; \mathbb{E}_\theta \left[t \; h^2(t) . \sum_{j=k+1}^{n} y_j^2 \right] .$$

Or

$$\sigma^{-2}\ \mathbb{E}_\theta\left[t\ h^2(t)\ \sum_{j=k+1}^{n} y_j^2\right]$$

$$= \sigma^{-2}\int_{\mathbb{R}^k}\ \sum_{j=k+1}^{n}\left\{\int_{\mathbb{R}^{n-k}} y_j^2\ t\ h^2(t)\ \mathcal{N}_{n-k}(0,\sigma^2 I_{n-k}, dy_{k+1},\ldots,dy_n)\right\}$$

$$\mathcal{N}_k(\theta,\sigma^2 I_k, dy_1,\ldots,dy_k)\ .$$

Comme n-k ≥ 2, on a p.s. pour tout j ($k+1 \leq j \leq n$), $\sum_{i=k+1}^{n} y_i^2 - y_j^2 > 0$. Il résulte alors du lemme 3 que

$$\sigma^{-2}\ \mathbb{E}_\theta\left[t\ h^2(t)\ \sum_{j=k+1}^{n} y_j^2\right]$$

$$\leq \int_{\mathbb{R}^k}\ \sum_{j=k+1}^{n}\left\{\int_{\mathbb{R}^{n-k}}\left(1+\frac{2y_j^2}{\sum_{i=k+1}^{n} y_i^2}\right) t\ h^2(t)\ \mathcal{N}_{n-k}(0,\sigma^2 I_{n-k}, dy_{n-k},\ldots,dy_n)\right\}$$

$$\mathcal{N}_k(\theta,\sigma^2 I_k, dy_1,\ldots,dy_k)$$

$$= (n-k+2)\ \mathbb{E}_\theta\left[t\ h^2(t)\right]$$

et donc que

$$B_\varphi(\theta) \leq \max_{1\leq i\leq k} q_i c_i^2 b_i^{-1}\ .\ \frac{n-k+2}{n-k}\ .\ \mathbb{E}_\theta\left[t\ h^2(t)\right]\ .$$

Minoration de $\Delta_\varphi(\theta)$.

Elle résulte immédiatement de ce qui précède car nous pouvons écrire

- dans le cas où σ^2 est connu

$$\Delta_\varphi(\theta) \geqslant \mathbb{E}_\theta\left[h(t)\left\{2\left(\sum_{i=1}^{k} q_i\ c_i - 2 \max_{1\leqslant i\leqslant k} q_i\ c_i\right) - \max_{1\leqslant i\leqslant k} q_i\ c_i^2\ b_i^{-1}.th(t)\right\}\right]$$

$$(3.3) \qquad \geqslant \max_{1\leqslant i\leqslant k} q_i\ c_i^2\ b_i^{-1}\ .\mathbb{E}_\theta[h(t)\{\alpha - th(t)\}]$$

- dans le cas où σ^2 est inconnu

$$\Delta_\varphi(\theta) \geqslant \mathbb{E}_\theta\left[h(t)\left\{2\left(\sum_{i=1}^{k} q_i\ c_i - 2 \max_{1\leqslant i\leqslant k} q_i\ c_i\right) - \max_{1\leqslant i\leqslant k} q_i\ c_i^2\ b_i^{-1}.\ \frac{n-k+2}{n-k}\ .th(t)\right\}\right]$$

$$(3.4) \qquad \geqslant \max_{1\leqslant i\leqslant k} q_i\ c_i^2\ b_i^{-1}.\ \frac{n-k+2}{n-k}.\mathbb{E}_\theta\left[h(t)\left\{\frac{n-k}{n-k+2}\ \alpha - th(t)\right\}\right]$$

La majoration de l'application $t \to th(t)$ permet alors de conclure que $\Delta_\varphi(\theta) \geqslant 0$ c'est à dire que le risque de φ est inférieur ou égal à celui de φ^0.

Les deux conditions suffisantes (relatives respectivement au cas σ^2 connu et au cas σ^2 inconnu) qui permettent d'assurer que le risque de φ est strictement inférieur à celui de φ^0 découlent directement des inégalités (3.3) et (3.4).

IV - APPENDICE .

Lemme 1

Soit f une fonction numérique continue sur $\mathbb{R}$ et de limite nulle en $-\infty$ et $+\infty$. Soit g une fonction numérique à variation bornée sur $\mathbb{R}$.

Alors les intégrales de Stieltjes

$$\int_{-\infty}^{+\infty} f(x)\ dg(x) \quad \text{et} \quad \int_{-\infty}^{+\infty} g(x)\ df(x)$$

existent et vérifient

$$\int_{-\infty}^{+\infty} f(x)\ dg(x) \quad = \quad - \int_{-\infty}^{+\infty} g(x)\ df(x)$$

Il s'agit d'un résultat classique d'intégration par parties de l'intégrale de Stieltjes d'une fonction continue et bornée sur $\mathbb{R}$ par rapport à une fonction à variation bornée sur $\mathbb{R}$.

<u>Lemme 2</u>

Soit h une application de $\mathbb{R}_+$ dans $\mathbb{R}_+$ telle que l'application $t \to th(t)$ soit croissante et majorée sur $\mathbb{R}_+$. Soit f une application positive continue sur $\mathbb{R}$ et tendant vers 0 au voisinage de $+\infty$ et de $-\infty$.

Pour $a > 0$ et $b > 0$ fixés, notons g l'application définie sur $\mathbb{R}$ par

$$(\forall x \in \mathbb{R}) \qquad g(x) \quad = \quad x\ h\left(ax^2+b\right) .$$

Alors

$$\int_{-\infty}^{+\infty} f(x)\ dg(x) \quad \geqslant \quad \int_{-\infty}^{+\infty} f(x)\quad h\left(ax^2+b\right) \left(1 - \frac{2ax^2}{ax^2+b}\right)\ dx .$$

<u>démonstration</u>

Il est clair que g est à variation bornée sur $\mathbb{R}$. Par conséquent, en vertu du lemme 1, $\int_{-\infty}^{+\infty} f(x)\ dg(x)$ existe. Alors

$$\int_{-\infty}^{+\infty} f(x)\ dg(x) \quad = \quad \int_{-\infty}^{+\infty} f(x)\ d\left(\left(ax^2+b\right)\ h\left(ax^2+b\right)\ \ \frac{x}{ax^2+b}\right)$$

$$= \quad I_1 + I_2 + J$$

où

$$I_1 \quad = \quad \int_{-\infty}^{0} f(x)\ \ \frac{x}{ax^2+b}\ d\left(\left(ax^2+b\right)\ h\left(ax^2+b\right)\right)$$

$$I_2 \quad = \quad \int_{0}^{+\infty} f(x)\ \ \frac{x}{ax^2+b}\ d\left(\left(ax^2+b\right)\ h\left(ax^2+b\right)\right)$$

$$J \quad = \quad \int_{-\infty}^{+\infty} f(x)\ \left(ax^2+b\right)\ h\left(ax^2+b\right)\ d\left(\frac{x}{ax^2+b}\right)$$

I_1 est l'intégrale de Stieltjes d'une fonction négative par rapport à une fonction décroissante sur $\mathbb{R}_+$ ($t \to th(t)$ est croissante sur $\mathbb{R}_+$); donc $I_1 \geqslant 0$. I_2 est l'intégrale de Stieltjes d'une fonction positive par rapport à une fonction croissante sur $\mathbb{R}_+$; donc $I_2 \geq 0$. Finalement $I_1+I_2 \geq 0$.

Ainsi $\int_{-\infty}^{+\infty} f(x)\, dg(x) \geq J$ et un simple calcul de J donne le deuxième membre de l'inégalité cherchée.

Lemme 3

Soit h une application de $\mathbb{R}_+$ dans $\mathbb{R}_+$ telle que l'application $t \to th(t)$ soit croissante et majorée sur $\mathbb{R}_+$. Soient a et b deux constantes réelles strictement positives. Alors

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x^2 \frac{a}{x^2+b} h^2\left(\frac{a}{x^2+b}\right) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

$$\leq \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{a}{x^2+b} h^2\left(\frac{a}{x^2+b}\right) \left(1+\frac{2x^2}{x^2+b}\right) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \,.$$

démonstration

Posons

$$I = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} g(x) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

où, pour tout x de $\mathbb{R}$,

$$g(x) = x^2 \frac{a}{x^2+b} h^2\left(\frac{a}{x^2+b}\right) .$$

Remarquons que, g étant une application positive, l'intégrale I existe toujours (valant éventuellement $+\infty$) et est égale à

$$\lim_{\substack{\gamma \to -\infty \\ \delta \to +\infty}} \int_{\gamma}^{\delta} g(x) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \,.$$

L'application $x \to g(x) \exp-\frac{x^2}{2\sigma^2}$ est le produit de l'application continue $x \to x \exp-\frac{x^2}{2\sigma^2}$ et de l'application k: $x \to \frac{ax}{x^2+b} h^2\left(\frac{a}{x^2+b}\right)$.

Or, sur tout intervalle $[\gamma,\delta]$ de $\mathbb{R}$, k est à variation bornée puisqu'elle peut s'écrire

$$k(x) = \left[\frac{a}{x^2+b} h\left(\frac{a}{x^2+b}\right)\right]^2 . \frac{x^3+bx}{a}$$

et que, par hypothèse sur h, elle est le produit de deux applications monotones par morceaux. Nous pouvons donc intégrer par parties sur $[\gamma,\delta]$ puis, grâce à la remarque précédente, passer à la limite en $-\infty$ et $+\infty$. Ceci nous donne

$$I = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dk(x)$$

l'annulation du terme entre crochets $\left[k(x) \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)\right]_{-\infty}^{+\infty}$ provenant de la majoration de l'application $t \to th(t)$ par une constante β:

$$\begin{aligned} k(x) \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) &= \frac{\sigma}{\sqrt{2\pi}} \left[\frac{a}{x^2+b} h\left(\frac{a}{x^2+b}\right)\right]^2 \frac{x^3+bx}{a} \exp\left(-\frac{x^2}{2\sigma^2}\right) \\ &\leqslant \frac{\sigma}{\sqrt{2\pi}} \frac{\beta^2}{a} (x^3+bx) \exp\left(-\frac{x^2}{2\sigma^2}\right) . \end{aligned}$$

L'inégalité cherchée résulte alors de l'expression de dk(x). Nous avons

$$dk(x) = 2x\, h\left(\frac{a}{x^2+b}\right) d\left(\frac{a}{x^2+b} h\left(\frac{a}{x^2+b}\right)\right) + \frac{a}{x^2+b} \left(1 + \frac{2x^2}{x^2+b}\right) h^2\left(\frac{a}{x^2+b}\right) dx$$

Le premier terme de cette somme donne, dans le calcul de I, une quantité négative, soit

$$\frac{2\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x\, h\left(\frac{a}{x^2+b}\right) \exp\left(-\frac{x^2}{2\sigma^2}\right) d\left(\frac{a}{x^2+b} h\left(\frac{a}{x^2+b}\right)\right) \leqslant 0 \; ;$$

en effet l'intégration sur $\mathbb{R}_-$ correspond à l'intégrale de Stieltjes d'une fonction négative par rapport à une fonction croissante et l'intégration sur $\mathbb{R}_+$ correspond à l'intégrale de Stieltjes d'une fonction positive par rapport à une fonction décroissante.

Finalement nous obtenons le résultat annoncé

$$I \leqslant \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{a}{x^2+b} h^2\left(\frac{a}{x^2+b}\right) \left(1 + \frac{2x^2}{x^2+b}\right) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx .$$

REFERENCES BIBLIOGRAPHIQUES

[1] BEN MANSOUR,D., "Présentation, dans les cas classiques et bayésiens, des estimateurs de James-Stein généralisés", Thèse de 3° cycle de l'Université de Rouen, 1983.

[2] BERGER, J.O., "Minimax estimation of location vectors for a wide class of densities", The Annals of Statistics, 1975,Vol. 3, n°6, pp. 1318-1328.

[3] BERGER, J.O., "Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss", The Annals of Statistics, 1976, Vol. 4, n°1, pp. 223-226.

[4] BERGER, J.O., "Minimax estimation of a multivariate normal mean under arbitrary quadratic loss", Journal of Multivariate Analysis, 1976, 6, pp.256-264.

[5] BERGER, J.O., "A robust generalized Bayes estimator and confidence region for a multivariate normal mean", The Annals of Statistics, 1980, Vol. 8, n°4, pp. 716-761.

[6] BERGER,J.O., BOCK, M.E., BROWN, L.D., CASELLA, G. and GLESER, L., "Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix", The Annals of Statistics, 1977, Vol. 5, n°4, pp. 763-771.

[7] BRANDWEIN, A.C., "Minimax estimation of the mean of spherically symmetric distribution under general quadratic loss", Journal of Multivariate Analysis, 1979, 9, pp.579-588.

[8] BRANDWEIN, A.C. and STRAWDERMAN, W.E., "Minimax estimation of location parameters for spherically symmetric distributions with concave loss", The Annals of Statistics, 1980, Vol. 8, n°2, pp. 279-284.

[9] GLESER, L.J., "Minimax estimation of a normal mean vector when the covariance is unknown", The Annals of Statistics, 1979, Vol. 7, n°4, pp. 838-846.

[10] JAMES, W. and STEIN, C., "Estimation with quadratic loss", Proc. Fourth Berkeley Symp. Math. Statist. Prob., University of California Press, 1961, 1, pp. 361-379.

[11] JUDGE, G.G. and BOCK, M.E., The statistical implications of pre-test and Stein rule estimators in econometrics, North-Holland, 1978.

[12] SCLOVE, S.L.,MORRIS, C. and RADHAKRISHMAN, R., "Non-optimality of preliminary test estimators for the multinormal mean", The Annals of Mathematical Statistics, 1980, Vol. 43, n°5, pp. 1481-1490.

[13] SHINOZAKI, N., "Simultaneous estimation of location parameters under quadratic loss", The Annals of Statistics, 1984, Vol. 1, n°4, pp. 322-335.

[14] STEIN, C., "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proc. Third Berkeley Symp. Math. Statist. Prob., University of California Press, 1955, 1, pp. 197-206.

[15] STEIN, C., "Estimation of the mean of a multivariate normal distribution", The Annals of Statistics, 1981, Vol. 9, n°6, pp. 1135-1151.