Sur l'estimation d'un vecteur moyen sous symétrie sphérique et sous contrainte

par

Othmane KORTBI

Thèse présentée au Département de mathématiques
en vue de l'obtention du grade de Docteur ès science (Ph.D.)

FACULTÉ DES SCIENCES
UNIVERSITÉ DE SHERBROOKE

Sherbrooke (Québec), Canada, Novembre-2011

Your file Votre référence
ISBN: 978-0-494-83295-0

Our file Notre référence
ISBN: 978-0-494-83295-0

Le 10 novembre 2011

le jury a accepté la thèse de Monsieur Othmane Kortbi
dans sa version finale.

Membres du jury

Professeur Éric Marchand
Directeur de recherche
Département de mathématiques

Monsieur Dominique Fourdrinier
Codirecteur de recherche
Université de Rouen

Professeur Bernard Colin
Membre
Département de mathématiques

Professeur Laurent Cavalier
Membre externe
Université Aix-Marseille I

Professeur Taoufik Bouezmarni
Président rapporteur
Département de mathématiques

Résumé

Ce travail est essentiellement centré sur l'estimation, du point de vue de la théorie de la décision, de la moyenne d'une distribution multidimensionnelle à symétrie sphérique. Sous coût quadratique, nous nous sommes concentrés à développer des classes d'estimateurs au moins aussi bons que les estimateurs usuels, puisque ces derniers tendent à perdre leur performance en dimension élevée et en présence de contraintes sur les paramètres.

Dans un premier temps, nous avons considéré les distributions de mélange (par rapport à σ^2) de lois normales multidimensionnelles $\mathcal{N}(\theta, \sigma^2 I_p)$, en dimension p supérieure ou égale à 3. Nous avons trouvé une grande classe de lois a priori (généralisées), aussi dans la classe des distributions de mélange de lois normales, qui génèrent des estimateurs de Bayes minimax. Ensuite, pour étendre nos résultats, nous avons considéré les distributions à symétrie sphérique (pas nécessairement mélange de lois normales) avec paramètre d'échelle connu, en dimension supérieure ou égale à 3 et en présence d'un vecteur résiduel. Nous avons obtenu une classe d'estimateurs de Bayes généralisés minimax pour une grande classe de distributions sphériques incluant certaines distributions mélange de lois normales.

Dans l'estimation de la moyenne θ d'une loi $N_p(\theta, I_p)$ sous la contrainte $\|\theta\| \leq m$ avec $m > 0$, une analyse en dimension finie pour comparer les estimateurs linéaires tronqués δ_a $(0 \leq a < 1)$ avec l'estimateur du maximum de vraisemblance δ_{emv} est donnée. Un cadre asymptotique est développé, ceci nous permet de déduire la sous-classe des estimateurs δ_a qui dominent δ_{emv} et de mesurer avec précision le degré d'amélioration relative en risque. Enfin, dans l'estimation de la moyenne θ d'une loi $N_p(\theta, \sigma^2 I_p)$ où σ est inconnu et sous la contrainte $\frac{\|\theta\|}{\sigma} \leq m$ avec $m > 0$, des résultats de dominance de l'estimateur X et de l'estimateur du maximum de vraisemblance δ_{emv} sont développés. En particulier, nous avons montré que le meilleur estimateur équivariant $\delta_m(x,s) = h_m\left(\frac{\|x\|^2}{s^2}\right)x$ pour $\frac{\|\theta\|}{\sigma} = m$ domine δ_{emv} lorsque $m \leq \sqrt{p}$ et que sa troncature $\delta_{h_{emv} \wedge h_m}$ domine δ_{emv} pour tout (m,p).

Mots-clés : Inférence statistique, statistique bayésienne, analyse multivariée statistique, théorie de la décision, espaces paramétriques contraints, coût quadratique.

Remerciements

Les travaux exposés dans ce manuscrit ont été effectués au Département de mathématiques de l'Université de Sherbrooke. Je tiens à adresser mes sincères remerciements à tous ceux qui, de près ou de loin, se sont intéressés à cette thèse et m'ont apporté leur soutien. En particulier, je remercie le Département de mathématiques, l'ISM et le FQRNT pour leur soutien financier.

Je tiens en premier lieu à remercier Éric Marchand, qui m'a encadré pendant la durée de cette thèse. Il a toujours su m'indiquer de bonnes directions de recherche quand il le fallait, me laisser chercher seul quand il le fallait, et a ce talent de savoir expliquer les concepts les plus techniques en des termes très intuitifs. Je pense avoir énormément appris à son contact, j'ai sincèrement apprécié de travailler avec lui et lui suis reconnaissant pour le temps qu'il ma consacré et toutes les opportunités qu'il m'a données au cours de cette thèse.

Mes remerciements vont également à Dominique Fourdrinier, mon co-directeur de thèse, qui a grandement contribué à ma formation. C'est lui qui m'a repéré quand j'étais en visite à l'université de Rouen et qui m'a initié à la recherche. De plus, c'est grâce à lui que j'ai connu Éric Marchand et William E. Strawderman.

Je tiens à remercier aussi William E. Strawderman, qui m'a accueilli lors de mon stage au Department of Statistics and Biostatistics de l'Université

Rutgers et grandement aidé à faire avancer mes travaux.

Je suis également très reconnaissant à Laurent Cavalier d'avoir accepté d'être rapporteur de cette thèse, ainsi qu'aux autres membres du jury, Bernard Colin et Taoufik Bouezmarni pour juger ce travail.

Un immense merci à mes amis, Nacer, Mohamed Tamazight et Sofiane qui m'ont soutenu dès le début. Merci de m'avoir aidé et encouragé, et pour m'avoir supporté durant toute cette période.

Enfin un énorme, ineffable merci à ma femme Fadhila qui a toujours été présente quand j'en avais besoin. Merci pour son soutien moral, pour ses encouragements, son aide. Sa seule présence a été source de sérénité et d'amour, conditions pour moi nécessaires, sans lesquelles je n'aurais jamais terminé cette thèse.

Table des matières

Liste des figures

Introduction

En 1956, Charles Stein [Ste56] découvrit un phénomène statistique, auquel la communauté statisticienne donnera éventuellement son nom en parlant d'effet Stein ou de paradoxe de Stein. Ce phénomène consiste en la non admissibilité, sous coût quadratique, de l'estimateur du maximum de vraisemblance de la moyenne d'une loi gaussienne p-dimensionnelle lorsque la dimension p est supérieure ou égale à 3.

Le caractère "phénoménal" a trait à l'inefficacité de l'estimateur usuel et à la coupure de dimension ainsi mise en évidence. En effet, si $p = 1$ ou si $p = 2$, le maximum de vraisemblance est admissible sous coût quadratique. En revanche, cette admissibilité est perdue pour tout $p > 2$ (cf. [JS61] et [Ste81]). Ce phénomène signifie que, disposant d'un p-uplet de moyennes, estimer celles-ci individuellement (en les considérant l'une après l'autre) ou les estimer globalement (en les envisageant comme formant un vecteur) n'est pas équivalent. Une différence est ainsi mise en évidence selon que $p < 3$ ou que $p \geq 3$. Le "paradoxe" tient à ce que le regroupement des p estimateurs admissibles peut conduire à un estimateur inadmissible alors même que les observations sous-jacentes sont des quantités indépendantes (pouvant, en pratique, n'avoir aucun lien entre elles).

Le résultat de Stein a des conséquences fondamentales. Il met en évidence que le caractère sans biais d'un estimateur n'est pas (ou n'est plus) la propriété ultime qu'il faille rechercher. Le phénomène souligne le fait que la Théorie de la Décision apporte des critères qui, via la fonction de coût, peuvent s'avérer incompatibles avec les critères statistiques classiques. L'objectif de recherche d'un éventuel estimateur meilleur que tous les autres paraît alors sans objet.

Le paradoxe de Stein a ouvert la voie à de nombreuses recherches en estimation. Ainsi la littérature sur les estimateurs à rétrécissement est abondante. La démarche conduisant à ces estimateurs repose tout d'abord sur le choix d'un estimateur standard (pour le problème d'estimation considéré). On entend par là que cet estimateur possède des propriétés telles que son usage s'impose naturellement (soit, par exemple, la minimaxité qui est présente dans le cas gaussien énoncé plus haut). Le but est alors de déterminer des estimateurs qui améliorent l'estimateur standard, le phénomène étant illustré par une coupure de dimension.

D'autre part, dans ce problème d'estimation de la moyenne d'une distribution multidimensionnelle, les deux approches dominantes sont l'approche de minimaxité et une variante du paradigme de Bayes. La première approche a reçu un développement assez important alors que la seconde approche est la plus utilisée en pratique, grâce à sa grande flexibilité. L'étude de l'interface entre ces deux approches, est un thème clé (voir [BR90]). L'inconvénient dans les deux méthodes est qu'aucune des deux n'implique nécessairement des estimateurs admissibles. Même si l'admissibilité peut conduire à des estimateurs non raisonnables (voir Berger [Ber85]), c'est un critère qui peut devenir désirable quand nous combinons la minimaxité avec les propriétés bayésiennes.

En effet, dans le cas gaussien, sous coût quadratique, l'estimateur de Bayes propre et unique est admissible à condition que le risque bayesien soit fini (cf. [LC99]).

Dans le cas général, notons que Stein [Ste56] et Brown [Bro66] ont montré que le phénomène de Stein survient pour plusieurs lois d'échantillonnage, pour plusieurs fonctions de coût et pour la plupart des problèmes d'estimation ; l'estimation de coût est aussi concernée par ce phénomène (cf. [Joh88], [FW95b] et [FW95a]). Ces derniers auteurs ont aussi exhibé des classes d'estimateurs contenant des estimateurs améliorés pour le même problème dans un cadre distributionnel plus large. Car, en dehors du cas gaussien, moins de choses ont été développées pour construire des estimateurs du minimax explicites améliorant le meilleur estimateur équivariant.

Une autre variante de ce problème de même importance; l'estimation d'une moyenne multidimensionnelle θ avec contrainte. Naturellement, le centre de gravité θ de plusieurs valeurs p-variées réelles issues d'un certain phénomène peut être restreint à un sous-ensemble de l'espace euclidien $\mathbb{R}^p$. Cependant, comme l'estimateur usuel n'exploite pas cette information, il tend à perdre l'admissibilité, et même, parfois la minimaxité (cf. [Kat61]). Comme exemple, considérons l'estimation de la moyenne θ d'une distribution normale unidimensionnelle avec variance connue et sous coût quadratique. Pour l'espace paramétrique $\{\theta : -\infty < \theta < \infty)$, l'estimateur du maximum de vraisemblance de θ est sans biais, admissible et du minimax. Par contre, pour l'espace paramétrique restreint $\{\theta : \theta \geq 0)$, ce dernier estimateur est inadmissible (voir Sacks [Sac63]), mais il est toujours du minimax (voir Katz [Kat61]). D'autre part, lorsque $\theta \in [a, b]$ pour certaines constantes connues

a et b, nous voyons que l'estimateur du maximum de vraisemblance de θ est Bayes et admissible sous la perte $L(\theta, d) = |d - \theta|$ (voir Kucerovsky et al. [KMPS09]).

Notre travail consiste en la détermination explicite d'estimateurs du paramètre de position $\theta \in \Theta \subseteq \mathbb{R}^p$ d'une distribution ayant une fonction de densité de la forme $f\left(\frac{\|x-\theta\|}{\sigma}\right)$, où σ est un paramètre d'échelle et $f(\cdot)$ une certaine fonction positive et qui dominent les estimateurs usuels sous coût quadratique. Pour un espace paramétrique Θ non restreint, notre objectif consiste à construire des estimateurs qui soient à la fois des estimateurs de Bayes et du minimax. La minimaxité des estimateurs bayésiens est souvent peu aisée à obtenir et n'est considérée pour de tels estimateurs qu'a posteriori : étant donné un estimateur bayésien, nous cherchons à montrer sa minimaxité éventuelle. En d'autres termes, nous cherchons des conditions suffisantes sur les lois a priori pour que les estimateurs de Bayes associés soient des estimateurs du minimax. Ensuite, pour un espace paramétrique restreint, nous considérons le problème d'estimation dans le cas où $\Theta = \{\theta : \|\theta\| / \sigma \leq m\}$, où m est une certaine constante positive. Cependant, notre objectif consiste à étudier certaines classes d'estimateurs afin de déterminer les conditions de dominance sur l'estimateur du maximum de vraisemblance du problème en question et d'exhiber leur performance en matière de gain en risque.

Le but de ce travail est aussi d'intégrer les plus récentes avancées dans ces domaines. Ainsi, le cadre gaussien est remplacé par des hypothèses distributionnelles plus larges et le cadre unidimensionnel est remplacé par des distributions multidimensionnelles. Souhaitant donner un cadre mathématiquement rigoureux à notre travail, nous nous proposons tout au début de présenter les outils de base et les résultats récents dans la littérature nécessaires pour notre développement. Le contenu de cette thèse est présenté comme suit :

Dans le chapitre 1, nous commençons par le modèle sans contrainte et nous montrons l'inadmissibilité de l'estimateur du minimax standard X dans le cas gaussien pour $p \geq 3$, en présentant des classes d'estimateurs améliorant l'estimateur X. Nous présentons en particulier les classes d'estimateurs du minimax développées par Stein [Ste81] et Efron et Morris [EM76]. Ensuite, nous présentons les classes d'estimateurs de Bayes du minimax développées par Faith [Fai78] et Fourdrinier et al. [FSW98]. Nous donnons aussi un résultat important de Brown [Bro71], utile pour l'analyse de l'admissibilité. Nous avons aussi considéré l'inadmissibilité de X dans le cas non gaussien pour $p \geq 3$, en présentant les classes d'estimateurs du minimax développées par Strawderman [Str71, Str74] et Berger [Ber75]. Dans le modèle avec contraintes, nous présentons le problème d'estimation de la moyenne θ d'une distribution normale $\mathcal{N}_p(\theta, I_p)$ où θ appartient à une boule de centre 0 et de rayon m, $m > 0$ connu. Suite à l'inadmissibilité de l'estimateur du maximum de vraisemblance, nous présentons des classes d'estimateurs améliorant ce dernier. En particulier, nous présentons les classes développées par Marchand et Perron [MP01] et Fourdrinier et Marchand [FM10]. Nous présentons aussi les résultats de minimaxité dans ce contexte développés par Marchand et Perron [MP02].

Dans le chapitre 2, nous présentons le travail que nous avons réalisé en collaboration avec Dominique Fourdrinier (Professeur à l'Université de Rouen) et William E. Strawderman (Professeur à Rutgers University) et qui a été publié dans *Journal of Multivariate Analysis* (Fourdrinier. D., Kortbi. O. and Strawderman, W. E. (2008). Bayes minimax estimators of the mean of a scale mixture of multivariate normal distributions. *Journal of Multivariate Analysis.*, **99**, 74-93.).

Dans ce travail, nous considérons le problème de l'estimation de la moyenne θ d'une distribution d'un mélange (par rapport à la variance) de lois normales multidimensionnelles, sous coût quadratique. Nous présentons de grandes classes de lois a priori (aussi dans la classe des mélanges de lois normales) qui fournissent des estimateurs de Bayes propres et généralisés du minimax. Ce travail étend les résultats de Strawderman [Str74], d'une manière similaire à celle de Maruyama [Mar03a], mais beaucoup plus dans l'esprit des travaux de Fourdrinier et al. [FSW98] pour le cas gaussien, dans le sens où nous construisons des classes de lois a priori induisant automatiquement la minimaxité. Les caractéristiques de ce travail sont que, dans certains cas, nous sommes capables de construire des estimateurs de Bayes propres satisfaisant les propriétés et les bornes de Strawderman [Str74]. Nous apportons également quelques précisions quant à l'application ou non des résultats de Strawderman dans certain cas. Dans le cas où ils sont inapplicables, nous donnons des estimateurs du minimax en nous basant sur les résultats de Berger [Ber75]. La condition de minimaxité la plus importante est que les densités de mélange de la distribution d'échantillonnage et de la distribution a priori possèdent toutes deux la propriété du rapport de vraisemblance monotone relatif au paramètre d'échelle.

Dans le chapitre 3, nous présentons le travail que nous avons réalisé en collaboration avec Dominique Fourdrinier (Professeur à l'université de Rouen) et William E. Strawderman (Professeur à Rutgers University) et qui a été soumis en 2010 à la revue *Journal of Statistical Planning and Inference.*

Dans ce travail, nous cherchons à étendre nos résultats à un cadre distributionnel plus général, en considérant le problème d'estimation de la moyenne θ d'une distribution à symétrie sphérique $p-$dimensionnelle, où $p \geq 3$, en présence d'un vecteur résiduel et sous coût quadratique. L'outil de base est le développement d'une expression du risque en fonction d'expressions différentielles de la marginale, utile pour prouver la minimaxité. Cette technique a été développée et utilisée par plusieurs auteurs (voir Fourdrinier et al. [FS08a, FS08b] et Maruyama [Mar03a]). Notre modèle diffère des précédents par la supposition que la fonctionnelle qui génère la distribution d'échantillonnage est connue. Autrement dit, nous avons supposé que le paramètre d'échelle $\sigma^2 = E\left[\|X-\theta\|^2\right]/p$ est fixé et connu. Ce développement a mis en évidence l'apport du vecteur résiduel et du paramètre d'échelle dans l'estimation du paramètre de position. Nous prouvons la minimaxité d'estimateurs de Bayes généralisés pour une vaste classe de distributions à symétrie sphérique. Cette classe ne contient pas que les distributions de mélange de lois normales et n'exige pas la propriété du rapport de vraisemblance monotone pour ce type de distributions utilisée dans les travaux de Maruyama [Mar03a] et de Fourdrinier *et al.* [FKS08].

Dans le chapitre 4, nous présentons le travail que nous avons réalisé en collaboration avec Éric Marchand (Professeur à l'Université de Sherbrooke) et qui a été soumis en 2011 à la revue *Journal of Statistical Planning and Inference.*

Dans ce travail, nous considérons le problème d'estimation de la moyenne

θ de la distribution $N_p(\theta, I_p)$ sous coût quadratique et sous la contrainte $\|\theta\| \leq m$, pour une certaine constante $m > 0$. L'efficacité de chaque estimateur dépend de la manière d'interpréter l'information issue de la contrainte. Pour cette raison, il est pertinent de comparer les performances de diverses classes d'estimateurs. En considérant la classe des estimateurs linéaires tronqués $\delta_a(x) = (a\,\|x\| \wedge m)\frac{x}{\|x\|}$, pour $0 \leq a \leq 1$, nous donnons une analyse en dimension finie (incluant le cas unidimensionnel) pour comparer δ_a avec l'estimateur du maximum de vraisemblance δ_{emv}. Nous utilisons l'identité de Stein pour obtenir l'estimateur sans biais de la différence de risques entre δ_a et δ_{emv} et les arguments de changement de signe de Karlin pour obtenir les conditions nécessaires et suffisantes pour que δ_a domine δ_{emv}. Ensuite, en utilisant la décomposition en risques conditionnels (en $\|X\|$), nous obtenons des résultats pour p fini. Nous développons un cadre asymptotique qui nous permet de déterminer précisément la sous-classe des estimateurs δ_a dominant δ_{emv}. Ce cadre nous permet aussi d'évaluer exactement les expressions du gain relatif en amélioration (par rapport à δ_{emv}). Nous donnons aussi, des exemples, des évaluations numériques et des graphiques pour illustrer la théorie.

Dans le chapitre 5, nous présentons le travail que nous avons réalisé en collaboration avec Éric Marchand (Professeur à l'Université de Sherbrooke) et qui fera l'objet d'un Rapport de recherche, au Département de mathématiques, Université de Sherbrooke.

Dans ce travail, nous considérons le problème d'estimation de la moyenne θ de la distribution multidimensionnelle $\mathcal{N}_p(\theta, \sigma^2 I_p)$ sous coût quadratique et sous la contrainte $\frac{\|\theta\|}{\sigma} \leq m$, pour une certaine constante $m > 0$. Nous supposons que le paramètre d'échelle est inconnu. Nous proposons des estimateurs qui dominent X, ainsi que l'estimateur du maximum de vraisemblance δ_{emv}, en particulier, des estimateurs bayésiens. Nous montrons que le meilleur estimateur équivariant (qui est un estimateur de Bayes) $\delta_m(x) = h_m(t)\,x$, où $t = \|x\|^2/s^2$, domine les estimateurs sans biais et δ_{emv} lorsque $m \leq \sqrt{p}$ et que sa troncature $\delta_{h_m \wedge h_{emv}}$domine δ_{emv} pour tout (m,p). Ce développement représente une extension des travaux de Kubokawa [Kub05] puisque ce dernier a considéré le cas unidimensionnel de ce problème et a établi des résultats de dominance de l'estimateur X seulement. Nous étendons aussi avec ce développement les résultats de Marchand et Perron [MP01] pour σ connu. Enfin, nous obtenons une forme potentiellement utile pour une certaine classe d'estimateurs de Bayes pour développer d'autres résultats de dominance.

Enfin, dans le dernier chapitre, nous récapitulons les résultats obtenus lors de ces travaux afin de faire le point sur les sujets en question. Nous ajoutons aussi une discussion sur les problèmes à aborder comme une suite naturelle de ce travail.

Chapitre 1

Préliminaires

1.1 Admissibilité et Minimaxité

Dans cette première partie, en se référant à Berger [Ber85] et Lehmann et Casella [LC99], nous revoyons, dans le cadre de la théorie de décision statistique, quelques outils de base dans le cas d'un problème d'estimation. Soient $\mathcal{X}$ et Θ deux espaces euclidiens quelconques et soit X une variable aléatoire définie sur l'espace d'échantillonnage $\mathcal{X}$ et suivant une distribution P_θ appartenant à une famille $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$. Cette loi P_θ de la famille $\mathcal{P}$ est inconnue au sens où le paramètre θ qui indexe P_θ est supposé inconnu ; nous nous référons à X comme étant un échantillon de la loi P_θ. L'objectif du problème d'estimation est la détermination d'une fonction δ définie sur l'espace d'échantillonnage, afin d'estimer une fonction $g(\theta)$ du paramètre θ. Bien entendu, on ne peut prétendre que $\delta(X)$ égale l'inconnu $g(\theta)$, puisqu'il s'agit d'une variable aléatoire. Pour cela, il est nécessaire de spécifier une mesure de l'écart entre $\delta(X)$ et $g(\theta)$. À cette fin, on introduit une fonction de perte (ou de coût) $L(\theta, \delta)$, qui est une fonction réelle positive de θ et de δ et, qu'usuellement, on suppose telle que $L(\theta, g(\theta)) = 0$ pour tout θ. Un exemple de telles fonctions est le coût quadratique $L(\theta, \delta) = (\delta - g(\theta))^2$ lorsque δ et $g(\theta)$ sont réels ou $L(\theta, \delta) = \|\delta - g(\theta)\|^2$ lorsque δ et $g(\theta)$ sont à valeurs dans $\mathbb{R}^p$ ($\|.\|$ étant la norme euclidienne classique).

Comme, pour un estimateur $\delta(X)$ de $g(\theta)$ et pour $X = x$ observé, le coût encouru $L(\theta, \delta(x))$ dépend de l'observation x, la fonction de coût ne suffit pas par elle-même à évaluer les performances globales de $\delta(X)$. Pour y remédier, une approche classique consiste à considérer le coût "moyen" qui est la fonction risque définie par :

$$R(\theta, \delta) = E_\theta\{L(\theta, \delta(X))\} \tag{1.1.1}$$

où E_θ désigne l'espérance par rapport à la loi P_θ. Pour une classe d'estimateurs obtenus, nous pouvons les comparer en utilisant les fonctions de risque définies par (1.1.1).

Définition 1.1.1. *Pour deux estimateurs δ_1 et δ_2, on dit que δ_1 est au moins aussi bon que δ_2 si :*

$$R(\theta, \delta_1) \leq R(\theta, \delta_2)$$

pour tout $\theta \in \Theta$, cette inégalité étant stricte pour au moins une valeur de θ.

Dans le cadre de la définition 1.1.1, un estimateur δ est dit inadmissible s'il est dominé (autrement dit, l'existence d'un estimateur au moins aussi bon que δ) par au moins un estimateur δ'. Inversement il est dit admissible s'il n'existe aucun estimateur qui le domine. Comme l'admissibilité est une propriété importante, il est intéressant de déterminer des estimateurs admissibles.

Définition 1.1.2. *Une classe d'estimateurs C est dite complète si, pour tout δ n'appartenant pas à C, il existe δ' appartenant à C tel que δ' domine δ.*

Cette définition implique que tout estimateur en dehors de la classe complète est inadmissible. Donc il est raisonnable de chercher l'estimateur optimal parmi les estimateurs de la classe complète. Il faut noter qu'il existe plusieurs estimateurs admissibles (tout estimateur dont la fonction de risque se croise avec toutes les autres est admissible) d'une part et d'autre part l'admissibilité est un critère d'optimalité faible. En effet, pour un coût quadratique, l'estimateur constant $\delta(X) = \theta_0$ est typiquement admissible, puisque sa fonction de risque $R(\theta, \delta) = \|\theta_0 - \theta\|^2 = 0$ pour $\theta = \theta_0$. sans avoir utilisé l'information fournie par X. Ce qui signifie que nous ne pouvons pas résoudre le problème du choix de l'estimateur optimal en se basant seulement sur l'admissibilité. Un deuxième critère de comparaison des estimateurs souvent utilisé est le critère de minimaxité, qui consiste à minimiser le risque dans le cas le plus défavorable.

Définition 1.1.3. *Un estimateur δ^M qui minimise le maximum de la fonction risque, c'est-à-dire tel que :*

$$\inf_{\delta} \sup_{\theta} R(\theta, \delta) = \sup_{\theta} R(\theta, \delta^M)$$

est appelé estimateur du minimax.

Dans la suite, nous donnons des relations entre admissibilité et minimaxité.

Théorème 1.1.4. *Si un estimateur du minimax est unique, alors il est admissible.*

Démonstration. Si l'estimateur est inadmissible, alors il est dominé par autre estimateur au sens du risque. D'où l'existence d'un autre estimateur du minimax. □

Pour les estimateurs dont le risque est constant, nous avons le résultat suivant.

Théorème 1.1.5. *Si un estimateur a un risque constant et s'il est admissible, alors il est aussi un estimateur du minimax.*

Démonstration. Si δ' est un estimateur de risque constant égal à k alors :

$$\sup_{\theta \in \Theta} R(\theta, \delta') = k.$$

Si, de plus, δ' est admissible on a, pour tout autre estimateur δ, l'implication suivante :

$$\forall \theta \in \Theta \; R(\theta, \delta) \leq k \Longrightarrow \forall \theta \in \Theta \; R(\theta, \delta) = k.$$

Il en résulte que :

$$\sup_{\theta \in \Theta} R(\theta, \delta) \leq k \Longrightarrow \sup_{\theta \in \Theta} R(\theta, \delta) = k.$$

et ainsi que k est le maximum du risque du modèle et donc que δ' est un estimateur du minimax. □

S'il existe un unique estimateur du minimax δ, selon le théorème 1.1.4, δ est admissible et il sera le meilleur estimateur pour le critère considéré. Maintenant, s'il n'est pas unique, il existera d'autres estimateurs du minimax améliorant δ. En particulier, si l'estimateur du minimax de risque constant est inadmissible, alors tout autre estimateur du minimax a un risque uniformément plus petit. Notons que si δ est un estimateur du minimax de risque constant égal à k, alors tout autre estimateur du minimax δ' est tel que $\sup_\theta R(\theta, \delta') = k$. C'est le contexte du phénomène de Stein. Dans ce cas, et en se basant sur la minimaxité et l'admissibilité, nous ne pouvons pas déterminer l'unique et le meilleur estimateur, puisque il peut exister plusieurs estimateurs du minimax et admissibles. D'où la pertinence du problème de caractérisation des classes d'estimateurs vérifiant la minimaxité et l'admissibilité.

Pour prouver l'admissibilité ou la minimaxité d'un estimateur δ, l'analyse statistique bayésienne fournit un cadre pertinent.

Définition 1.1.6. *Supposons que Θ soit muni d'une loi a priori propre π (c'est-à-dire telle que $\int_\Theta d\pi(\theta) = 1$). Un estimateur δ minimisant le risque de Bayes défini par :*

$$r(\pi, \delta) = \int_\Theta R(\theta, \delta)\, d\pi(\theta)$$

est appelé estimateur de Bayes par rapport à π.

Cette dernière définition induit une méthode de déduction de l'admissibilité qui est donnée dans le résultat suivant.

Théorème 1.1.7. *Si l'estimateur de Bayes relativement à une loi a priori propre π est unique $\left((P_\theta)_{\theta\in\Theta}\right)$-presque sûrement, alors celui-ci est admissible.*

Démonstration. Soit δ un estimateur au moins aussi bon que l'estimateur de Bayes δ_π, c'est-à-dire tel que :

$$\forall \theta \in \Theta \quad R(\theta, \delta) \leq R(\theta, \delta_\pi)\,.$$

Alors son risque de Bayes vérifie :

$$r(\pi, \delta) = \int R(\theta, \delta) \;\; d\pi(\theta) \leq \int R(\theta. \delta_\pi) \;\; d\pi(\theta) = r(\pi, \delta_\pi)\,.$$

Ceci signifie que l'estimateur δ est un estimateur de Bayes. Par hypothèse d'unicité, il en résulte que, $\left((P_\theta)_{\theta\in\Theta}\right)$-presque sûrement, $\delta = \delta_\pi$ et donc que δ_π est un estimateur admissible. $\square$

Théorème 1.1.8. *Lorsque la fonction de risque est continue en θ pour tout estimateur δ, si π est une distribution absolument continue de densité ν positive sur Θ, un estimateur de Bayes δ^ν associé à π avec risque de Bayes fini est admissible.*

Démonstration. Supposons que δ^ν soit inadmissible, alors il existe un estimateur δ tel que $R(\theta, \delta) \leq R(\theta, \delta^\nu)$ pour tout θ avec une inégalité stricte pour au moins une valeur de θ. La continuité de la fonction risque implique qu'il existe un sous-ensemble ouvert Θ_0 de l'espace des paramètres dans lequel $R(\theta, \delta) < R(\theta, \delta^\nu)$ pour tout θ dans Θ_0. Par conséquent, nous obtenons :

$$\int R(\theta, \delta)\nu(\theta)\, d\theta < \int R(\theta, \delta^\nu)\nu(\theta)\, d\theta < \infty$$

puisque ν est une fonction strictement positive, ce qui est en contradiction avec la définition de δ^ν. □

Dans l'analyse bayésienne, la loi π et la densité ν utilisées précédemment sont appelées respectivement, la distribution a priori et la densité a priori. Autrement dit, θ est une réalisation d'un vecteur aléatoire $\boldsymbol{\theta}$ de distribution a priori π. Un échantillon X est alors décrit à partir de sa loi $P_\theta = P_{X|\theta}$, considérée comme une distribution conditionnelle de X sachant $\boldsymbol{\theta} = \theta$. Cependant, l'échantillon observé $X = x$ est utilisé pour obtenir une actualisation de la distribution a priori, appelée distribution a posteriori. Dans ce cas, notons que la distribution conjointe de X et de $\boldsymbol{\theta}$ est la mesure de probabilité définie sur $\mathcal{X} \times \Theta$ et vérifiant :

$$P(A \times B) = \int_B P_{X|\theta}(A)\, d\pi(\theta), \;\; A \in \mathcal{B}_\mathcal{X}, \;\; B \in \mathcal{B}_\Theta$$

où, $\mathcal{B}_\mathcal{X}$ et $\mathcal{B}_\Theta$ sont respectivement, les tribus boréliennes de $\mathcal{X}$ et Θ. Cependant, nous remarquons que la distribution a posteriori de θ, sachant $X = x$, est exprimée par la distribution conditionnelle $P_{\theta|x}$. En plus, si $P_{X|\theta}$ et π ont des densités de probabilité, par rapport à la mesure de Lebesgue, notées respectivement $f(\cdot|\theta)$ et $\nu(\theta)$, alors la loi marginale M admet une densité marginale donnée par :

$$m(x) = \int_{\Theta} f(x|\theta)\nu(\theta)\, d\theta\,,$$

et la loi a posteriori $P_{\theta|x}$ admet une densité a posteriori (si $m(x) \neq 0$) donnée par :

$$\nu(\theta|x) = \frac{f(x|\theta)\nu(\theta)}{m(x)}\,.$$

En utilisant le concept de loi a posteriori, la détermination des estimateurs de Bayes est basée sur le résultat suivant.

Théorème 1.1.9. *En supposant qu'il existe au moins un estimateur à risque fini, et que pour tout x, il existe une valeur $\delta^\pi(x)$ minimisant l'expression*

$$E\{L(\theta, \delta(x))|X = x\}, \tag{1.1.2}$$

où l'espérance est calculée par rapport à la distribution a posteriori $P_{\theta|x}$. Alors δ^π est un estimateur de Bayes.

Démonstration. Soit δ un estimateur à risque fini. Alors l'expression (1.1.2) est finie, puisque L est positive. Cependant, nous avons :

$$E\{L(\theta, \delta(x))|X = x\} \geq E\{L(\theta, \delta^\pi(x))|X = x\},$$

et en calculant l'espérance, par rapport à la loi marginale M, des deux membres de l'inégalité nous obtenons le résultat voulu. $\square$

L'expression (1.1.2) est appelée coût a posteriori. Ainsi l'estimateur de Bayes est interprété comme un estimateur qui minimise le coût a posteriori.

Corollaire 1.1.10. *En supposant que les conditions du théorème 1.1.9 soient valides et en supposant également que $L(\theta, d) = \|d - g(\theta)\|^2$, alors l'estimateur de Bayes de $g(\theta)$ est donné par $\delta^\pi(x) = E[g(\theta)|X = x]$, où l'espérance E est calculée par rapport à la distribution a posteriori de $\theta \mid x$.*

Démonstration. L'estimateur de Bayes, d'après le théorème 1.1.9, est obtenu en minimisant :

$$E\{(g(\theta) - \delta(x))^2|X = x\}. \tag{1.1.3}$$

Le développement de l'expression (1.1.3) en introduisant $E[g(\theta)|X = x]$ donne :

$$(\delta(x) - E[g(\theta)|X = x])^2 - E[g(\theta)|X = x]^2 + E[g(\theta)^2|X = x],$$

qui atteint son minimum pour $\delta(x) = E[g(\theta)|X = x]$. □

Maintenant, nous allons étendre la définition de l'estimateur de Bayes au cas où π est une mesure telle que $\int d\pi(\theta) = \infty$, appelée distribution a priori impropre. Dans le cas où l'expression (1.1.2) est finie pour tout x, l'estimateur de Bayes associé à cette distribution peut être bien défini.

Définition 1.1.11. *Un estimateur δ^π est dit estimateur de Bayes généralisé par rapport à la mesure $\pi(\theta)$ (même si $\pi(\theta)$ n'est pas une distribution de probabilité propre) si l'espérance a posteriori, $E\{L(\theta, \delta(x))|X = x\}$, atteint son minimum en $\delta(x) = \delta^\pi(x)$ pour tout x.*

Une condition suffisante de minimaxité pour quelques estimateurs est donnée par le résultat suivant, en utilisant les propriétés de limite des estimateurs de Bayes.

Lemme 1.1.12. *Soit $\{\pi_n\}$ une suite de distributions a priori propres et soient δ_n les estimateurs de Bayes associés respectivement à π_n. Si $\lim_{n\to\infty} r(\pi_n, \delta_n) < \infty$ et s'il existe un estimateur δ tel que $\sup_\theta R(\theta, \delta) = \lim_{n\to\infty} r(\pi_n, \delta_n)$ alors δ est un estimateur du minimax.*

Démonstration. Soit δ' un estimateur quelconque. Alors nous avons :

$$\sup_\theta R(\theta, \delta') \geq \int R(\theta, \delta')\, d\pi_n(\theta) \geq r(\pi_n, \delta_n),$$

ceci pour tout n. En passant à la limite, nous en déduisons que :

$$\sup_\theta R(\theta, \delta') \geq \sup_\theta R(\theta, \delta),$$

c'est à dire que δ est un estimateur du minimax. $\square$

Nous allons aussi définir l'estimateur sans biais du risque, ce qui nous permettra de montrer la minimaxité de certains estimateurs.

Définition 1.1.13. *Pour un estimateur quelconque $\delta(X)$ dont le risque est $R(\theta, \delta)$, s'il existe une statistique $\widehat{R}(\delta(X))$, indépendante de θ, telle que :*

$$R(\theta, \delta) = E_\theta[\widehat{R}(\delta(X))],$$

alors cette dernière statistique est appelée un estimateur sans biais du risque.

Lemme 1.1.14. *Soit $\delta'(X)$ un estimateur du minimax. Soit $\delta(X)$ un estimateur quelconque. En désignant par $\widehat{R}(\delta'(X))$ et $\widehat{R}(\delta(X))$ des estimateurs sans biais du risque associés respectivement aux estimateurs $\delta'(X)$ et $\delta(X)$, alors on a :*

(a) $\delta(X)$ est un estimateur du minimax si, pour tout $x \in \mathcal{X}$, $\widehat{R}(\delta(x)) \leq \widehat{R}(\delta'(x))$.

(b) $\delta'(X)$ *est inadmissible si, de plus,* $\widehat{R}(\delta(x)) < \widehat{R}(\delta'(x))$ *sur un ensemble de valeurs de* x *de probabilité non nulle par rapport à la famille* $(P_\theta)_{\theta\in\Theta}$.

Démonstration. Pour la première partie, en calculant l'espérance des deux membres de $\widehat{R}(\delta(X)) \leq \widehat{R}(\delta'(X))$, nous obtenons $R(\theta, \delta) \leq R(\theta, \delta')$ et δ est un estimateur du minimax, puisque δ' est supposé du minimax. La deuxième partie est une déduction directe de la définition de l'inadmissibilité. □

1.2 Inadmissibilité de l'estimateur usuel

1.2.1 Cas gaussien

Soit X un vecteur aléatoire de dimension $p \geq 3$, distribué selon une loi normale multidimensionnelle $\mathcal{N}_p(\theta, I_p)$ (on note par p_θ la fonction de densité correspondante). Nous considérons le problème d'estimation de la moyenne θ par un estimateur $\delta(X)$ sous coût quadratique $\|\delta(X) - \theta\|^2$. L'estimateur $\delta(X)$ est alors évalué par sa fonction risque :

$$R(\theta, \delta) = E_\theta[\|\delta(X) - \theta\|^2] = \int_{\mathbb{R}^p} \|\delta(x) - \theta\|^2 p_\theta(x)\, dx\,.$$

L'estimateur du maximum de vraisemblance et sans biais X de θ, qui est aussi à variance minimale est l'estimateur usuel. Rappelons ici que X est aussi le meilleur estimateur équivariant pour le groupe additif G de transformations sur R^p, $g : X \to X + c$ (Voir Lehmann et Casella [LC99]). En outre, il est clair que le risque de X est constant et égal à p. Les résultats suivants, concernant la minimaxité et l'admissibilité de X, sont bien connus et fondamentaux.

Théorème 1.2.1. *(a) L'estimateur* X *est un estimateur du minimax.*

(b) X *est admissible pour* $p \leq 2$.

(c) X est inadmissible pour $p \geq 3$.

Pour $p \geq 3$, les parties (a) et (c) impliquent que, si un estimateur δ domine X, alors δ est aussi un estimateur du minimax.

Démonstration. Pour la partie (a), nous considérons les distributions normales multidimensionnelles $\mathcal{N}_p(0, mI_p)$ comme distributions a priori de θ. Dans ce contexte, la fonction de densité conjointe de θ et de X est donnée par :

$$\begin{aligned} f(x,\theta) &\propto \exp\left(-\frac{\|x-\theta\|^2}{2}\right)\exp\left(-\frac{\|\theta\|^2}{2m}\right) \\ &\propto \exp\left(-\frac{m+1}{m}\frac{\|\theta - mx/(m+1)\|^2}{2}\right)\exp\left(-\frac{\|x\|^2}{2(m+1)}\right). \end{aligned}$$

Cette dernière expression, nous permet de remarquer que la distribution a posteriori de $\theta|X = x$ est la loi $\mathcal{N}_p(mx/(m+1), m(m+1)^{-1}I_p)$, ce qui permet d'assurer, en utilisant le corollaire 1.1.10, que, sous coût quadratique, l'estimateur de Bayes est $\delta_m(X) = mX/(m+1)$. Or le risque de $\delta_m(X)$ est :

$$\begin{aligned} R(\theta,\delta_m) &= E\left[\|\delta_m(X)-\theta\|^2\right] \\ &= E\left[\|m(m+1)^{-1}(X-\theta) - \theta/(m+1)\|^2\right] \\ &= (m/(m+1))^2 p + (m+1)^{-2}\|\theta\|^2. \end{aligned}$$

Par conséquent, en intégrant par rapport à la loi a priori $\mathcal{N}_p(0, mI_p)$, ceci nous permet de déduire le risque de Bayes de δ_m :

$$r(\pi_m,\delta_m) = \frac{m^2}{(m+1)^2}p + (m+1)^{-2}pm = \frac{m}{m+1}\,p.$$

Comme $r(\pi_m, \delta_m)$ tend vers p, la valeur du risque de X, quand m tend vers l'infini, le lemme 1.1.12 assure la minimaxité de X.

Pour la preuve de la partie (b), rappelons que X est un estimateur de Bayes généralisé par rapport à la mesure de Lebesgue dans $\mathbb{R}^p$ et que la densité marginale $m(x)$ est constante. Comme il est clair que l'intégrale $\int_1^\infty t^{-p/2}\,dt$ est divergente pour $p = 1, 2$, le résultat de Brown, rappelé dans le théorème 1.2.2 ci-dessous, garantit l'admissibilité de X. $\square$

Nous donnons, ici, un résultat important pour l'admissibilité et applicable pour les distributions a priori à symétrie sphérique : il est basé sur le comportement de la marginale $m(x)$. Ce résultat est dû à Brown(1971) [Bro71].

Théorème 1.2.2. (Brown, 1971). *Supposons que la distribution a priori généralisée π soit à symétrie sphérique de sorte que la densité marginale m ne soit fonction de x qu'au travers de $\|x\|^2$, soit $m(x) = m^*(\|x\|^2)$. Alors tout estimateur de Bayes généralisé à risque borné δ_π est admissible si et seulement si*

$$\int_1^\infty t^{-p/2} m^*(t)^{-1}\,dt = \infty\,.$$

Pour prouver la partie (c) du théorème 1.2.1, nous présentons une classe d'estimateurs dominants X. L'inadmissibilité de X pour $p \geq 3$, a été mise en évidence par Stein [Ste56]. Elle représentait un phénomène en ce sens que l'estimateur usuel de plusieurs paramètres regroupés est inadmissible, alors que les composantes de l'estimateur sont séparément admissibles dans l'estimation unidimensionnelle correspondante à chaque paramètre. Après la preuve non constructive de l'inadmissibilité de Stein [Ste56], James et Stein [JS61] ont déterminé une forme explicite d'un estimateur améliorant l'estimateur X donnée par :

$$\delta^{JS}(X) = \left(1 - \frac{p-2}{\|X\|^2}\right) X\,,$$

appelé estimateur de James-Stein. Leur méthode de démonstration est basée sur le fait que la distribution du χ^2 non centrée est représentée par un mélange poissonnien de distributions du χ^2 centrées. Dans cette thèse, nous présentons une méthode simple et puissante inspirée de Stein [Ste81]. Le résultat suivant, appelé identité de Stein, est essentiel. Tout d'abord, nous rappelons la différentiabilité faible, qui est une notion pertinente pour faire de l'intégration par parties dans le cas où les fonctions à intégrer ne sont pas différentiables partout. Les estimateurs de James-Stein représentent un bon exemple, puisqu'ils sont déterminés par les fonctions $g(X) = -\dfrac{a}{\|X\|^2}X$ qui n'existent pas en 0.

Définition 1.2.3. *Une fonction $u(\cdot)$ de $\mathbb{R}^p$ dans $\mathbb{R}$ est faiblement différentiable, si elle est localement intégrable et si elle est telle que, pour tout $i = 1, \ldots, p$, il existe une fonction localement intégrable notée $\dfrac{\partial}{\partial x_i}u(\cdot)$ vérifiant :*

$$\int_{\mathbb{R}^p} u(x)\,\frac{\partial}{\partial x_i}\varphi(x) = -\int_{\mathbb{R}^p} \frac{\partial}{\partial x_i}u(x)\,\varphi(x)\,,$$

pour toute fonction $\varphi(\cdot)$ de $\mathbb{R}^p$ dans $\mathbb{R}$ indéfiniment différentiable et à support compact.

Lemme 1.2.4. *Si $Y \sim \mathcal{N}(\mu, 1)$ et si g une fonction faiblement différentiable telle que $E_\mu[|g'(Y)|] < \infty$, alors*

$$E_\mu[g(Y)(Y-\mu)] = E_\mu[g'(Y)]\,. \tag{1.2.1}$$

Démonstration. Notons par $p_0(y)$ la densité de la loi normale centrée réduite, avec sa fonction dérivée $p_0'(y) = -yp_0(y)$. Dans l'expression (1.2.1), le membre de droite existe par hypothèse sur g et s'écrit :

$$E_\mu[g'(Y)] = \int_{-\infty}^{0} g'(y)p_0(y-\mu)\,dy + \int_{0}^{\infty} g'(y)p_0(y-\mu)\,dy\,, \tag{1.2.2}$$

où la fonction g' désigne la dérivée au sens faible de la fonction g. En utilisant la relation $p_0(y) = \int_y^\infty tp_0(t)\,dt$, le second terme du membre de droite de l'équation (1.2.2) devient :

$$\begin{aligned}\int_0^\infty g'(y)p_0(y-\mu)\,dy &= \int_0^\infty g'(y)\left(\int_y^\infty (z-\mu)p_0(z-\mu)\,dz\right)dy \\ &= \int_0^\infty (z-\mu)p_0(z-\mu)\left(\int_0^\infty g'(y)I_{[y\le z]}\,dy\right)dz \\ &= \int_0^\infty (z-\mu)p_0(z-\mu)(g(z)-g(0))\,dz\,, \quad (1.2.3)\end{aligned}$$

où $I_{[y\le z]}$ est la fonction indicatrice, qui vaut 1 pour $y \le z$ et 0 sinon. Le théorème de Fubini assure la deuxième égalité dans (1.2.3). De manière similaire, nous trouvons :

$$\int_{-\infty}^0 g'(y)p_0(y-\mu)\,dy = \int_{-\infty}^0 (z-\mu)p_0(z-\mu)(g(z)-g(0))\,dz\,. \quad (1.2.4)$$

En combinant (1.2.3) et (1.2.4), nous obtenons l'équation (1.2.1). □

Une version multidimensionnelle de l'identité de Stein découle directement du lemme 1.2.4 (voir [Ste81]). Cette nouvelle version est donnée par le résultat suivant.

Lemme 1.2.5. *Si $Z \sim \mathcal{N}_p(\theta, \sigma^2 I_p)$ et si g est une fonction faiblement différentiable de $\mathbb{R}^p$ dans lui-même telle que $E_\theta\left|\frac{\partial}{\partial z_i}g_i(Z)\right| < \infty$ pour tout $i = 1, ..., p$, alors*

$$E_\theta[(Z-\theta)'g(Z)] = \sigma^2 E_\theta\left[div\,(g(Z))\right]\,. \quad (1.2.5)$$

où $div\,(g(Z)) = \sum_{i=1}^p \frac{\partial}{\partial z_i}\,g_i(Z)$ et $(Z-\theta)'$ est le transposé du vecteur $(Z-\theta)$.

En utilisant l'identité (1.2.5), la fonction de risque d'un estimateur quelconque de type $\delta(X) = X + g(X)$ s'écrit :

$$
\begin{aligned}
R(\theta,\delta) &= E_\theta\left[\|\delta(X)-\theta\|^2\right] \\
&= E_\theta\left[\|X-\theta\|^2\right] + E_\theta\left[\|g(X)\|^2\right] + 2E_\theta\left[(X-\theta)'g(X)\right] \\
&= p + E_\theta\left[\|g(X)\|^2\right] + 2E_\theta\left[div\,(g(X))\right] ,
\end{aligned}
$$

où $g(x) = \{g_i(x)\}$ est par hypothèse faiblement différentiable et telle que $E_\theta\ \left[\|g(X)\|^2\right] < \infty$ et $E_\theta|\frac{\partial}{\partial x_i}\ g_i(X)| < \infty$ pour tout $i = 1, ..., p$. Dans ce cas, l'estimateur sans biais du risque est $p + \|g(X)\|^2 + 2div\,(g(X))$, et en appliquant le lemme 1.1.14 nous obtenons une condition suffisante de minimaxité.

Théorème 1.2.6. (Stein, 1981). *Tout estimateur du type $X+g(X)$, où $g(\cdot)$ est une fonction faiblement différentiable et vérifiant les conditions suivantes :*

(a) $E_\theta\ \left[\|g(X)\|^2\right] < \infty$,

(b) $E_\theta\left|\frac{\partial}{\partial x_i} g_i(X)\right| < \infty$ *pour tout* $i = 1, ..., p$ *et,*

(c) pour tout x *:*

$$
\|g(x)\|^2 + 2div\,(g(x)) \leq 0\,, \tag{1.2.6}
$$

est un estimateur du minimax.

Pour appréhender plus facilement la structure des estimateurs du minimax, nous considérons la classe d'estimateurs, dits estimateurs de Baranchik, ayant la forme suivante

$$
\delta_r(X) = (1 - r(\|X\|^2)/\|X\|^2)X\,, \tag{1.2.7}
$$

qui est une classe d'estimateurs équivariants par rapport aux transformations orthogonales ΓX et $\Gamma\theta$ pour une matrice orthogonale Γ. En posant $g(x) = -r(\|x\|^2)x/\|x\|^2$ dans (1.2.6), nous obtenons le résultat suivant.

Corollaire 1.2.7. (Efron et Morris, 1976). *Pour un estimateur du type* (1.2.7) *tel que* $E_\theta\left[\frac{r^2(\|X\|^2)}{\|X\|^2}\right] < \infty$ *et* $E_\theta\left|\frac{\partial}{\partial x_i}\frac{r(\|X\|^2)}{\|X\|^2}X_i\right| < \infty$ *pour tout* $i = 1, ..., p$, *si* $r(w)$ *est solution de l'inégalité :*

$$r(w)(2(p-2) - r(w))/w + 4r'(w) \geq 0\,, \tag{1.2.8}$$

alors δ_r *est un estimateur du minimax.*

La condition (1.2.8) *est satisfaite dès que*

(a) $r(w)$ *est une fonction monotone croissante.*

(b) $0 \leq r(w) \leq 2(p-2)$ *pour tout* $w \geq 0$.

Par conséquent, une classe d'estimateurs meilleurs que X est construite. Remarquons que l'estimateur de James-Stein δ^{JS} appartient à cette classe, puisque l'inégalité (1.2.8) est satisfaite par $r(w) = p-2$. Il faut noter que les conditions suffisantes de minimaxité, (a) et (b) dans le corollaire 1.2.7, ont été déterminées en premier lieu par Baranchik [Bar64a] avant Stein [Ste81].

1.2.2 Cas non gaussien

Dans un cadre distributionnel plus large que le cas gaussien, le phénomène de Stein existe toujours. En effet Brown [Bro66] a montré que le meilleur estimateur invariant du paramètre de position est admissible en dimension $p = 1$, pour une famille de distributions plus large qui contient les distributions à symétrie sphérique et pour des fonctions de coût généralisant le coût quadratique. Le cas $p = 2$ a été traité séparément par Brown [Bro65] où il a montré l'admissibilité du meilleur estimateur invariant du paramètre de position dans le même contexte. En dimension $p \geq 3$, Brown [Bro66] a montré que le meilleur estimateur invariant du paramètre de position est inadmissible, pour une famille de distributions plus large qui contient les distributions à symétrie sphérique en considérant des fonctions de coût convexes et d'autres fonctions de coût non nécessairement convexes.

Toutefois, il faut noter que Brown ne donne pas nécessairement d'estimateur dominant explicite et donc, il est pertinent de le faire dans des contextes distributionnels spécifiques.

1.3 Classes d'estimateurs de Bayes du minimax dans le cas gaussien

1.3.1 Estimateurs de Bayes généralisés et leur admissibilité

Dans cette section, nous construisons certaines classes d'estimateurs de Bayes généralisés et nous considérons leur admissibilité. Toute cette construction est inspirée de la littérature concernant ce problème, en particulier de Fourdrinier et al. [FSW98]. Comme lois a priori, nous considérons une sous-classe importante de distributions sphériques à savoir celle des mélanges par rapport à la variance de lois normales multidimensionnelles, dont la fonction de densité est proportionnelle à

$$\int_0^1 \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \exp\left(-\frac{\lambda}{2(1-\lambda)}\|\theta\|^2\right) \lambda^{-a} h(\lambda)\, d\lambda \qquad (1.3.1)$$

pour tout réel a, où $h(\lambda)$ est une fonction quelconque positive mesurable sur $(0,1)$ telle que $\lim_{\lambda\to 0} h(\lambda) = c_1 > 0$, à condition que cette dernière intégrale existe. Notons que ces densités sont propres si $\int_0^1 \lambda^{-a} h(\lambda)\, d\lambda$ est finie. Ces lois a priori sont interprétées comme étant des distributions a priori hiérarchiques à deux niveaux : pour une valeur fixée de λ, le paramètre θ est distribué selon la loi $\mathcal{N}(0, \lambda^{-1}(1-\lambda)I_p)$ et, au second niveau, λ est distribué selon une densité de la forme $\lambda^{-a}h(\lambda)$. En utilisant le théorème de Fubini pour les fonctions positives, la fonction de densité marginale de X est donnée par

$$\begin{aligned} m_h(x) &\propto \int_0^1 \int_{\mathbb{R}^p} \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \lambda^{-a} h(\lambda) \\ &\quad \times \exp\left(-\frac{\|x-\theta\|^2}{2} - \frac{\lambda}{2(1-\lambda)}\|\theta\|^2\right) d\theta\, d\lambda \\ &= \int_0^1 \int_{\mathbb{R}^p} \left(\frac{\lambda}{1-\lambda}\right)^{p/2} \exp\left(-\frac{\lambda}{2}\|x\|^2\right) \\ &\quad \times \exp\left(-\frac{\|\theta-(1-\lambda)x\|^2}{2(1-\lambda)}\right) \lambda^{-a} h(\lambda)\, d\theta\, d\lambda \\ &\propto \int_0^1 \exp\left(-\frac{\lambda}{2}\|x\|^2\right) \lambda^{p/2-a} h(\lambda)\, d\lambda\,. \end{aligned} \tag{1.3.2}$$

Il est clair que $m_h(x)$ existe pour tout x si et seulement si

$$\int_0^1 \lambda^{p/2-a} h(\lambda)\, d\lambda < \infty\,, \tag{1.3.3}$$

ce qui est satisfait pour $a < p/2 + 1$. Dans ce cas, la dérivation sous le signe intégrale est assurée par le théorème de la convergence dominée de Lebesgue, puisqu'il s'agit d'une fonction intégrable ainsi que toutes ses dérivées partielles. Par conséquent, nous obtenons :

$$\nabla m_h(x) = -\left(\int_0^1 \lambda^{p/2+1-a} \exp\left(-\|x\|^2\lambda/2\right) h(\lambda)\, d\lambda\right) x \tag{1.3.4}$$

et

$$\begin{aligned} \Delta m_h(x) = &\left(\int_0^1 \lambda^{p/2+2-a} \exp\left(-\|x\|^2\lambda/2\right) h(\lambda)\, d\lambda\right) \|x\|^2 \\ &- p\left(\int_0^1 \lambda^{p/2+1-a} \exp\left(-\|x\|^2\lambda/2\right) h(\lambda)\, d\lambda\right)\,. \end{aligned} \tag{1.3.5}$$

où, les opérateurs ∇ et Δ désignent respectivement le vecteur gradient et le Laplacien. Notons que l'estimateur de Bayes (propre ou généralisé) associé à ce cas s'écrit $X + \nabla \log m_h(X)$, ce qui nous permet de conclure que cet estimateur de Bayes est bien défini dès que $m_h(X)$ est définie. Cependant, l'estimateur de Bayes s'écrit :

$$\delta_h(X) = \left(1 - \frac{\int_0^1 \lambda^{p/2+1-a} \exp\left(-\|X\|^2\lambda/2\right) h(\lambda)\, d\lambda}{\int_0^1 \lambda^{p/2-a} \exp\left(-\|X\|^2\lambda/2\right) h(\lambda)\, d\lambda}\right) X\,. \qquad (1.3.6)$$

Ensuite, en posant :

$$r_h(w) = w \frac{\int_0^1 \lambda^{p/2+1-a} \exp\left(-w\lambda/2\right) h(\lambda)\, d\lambda}{\int_0^1 \lambda^{p/2-a} \exp\left(-w\lambda/2\right) h(\lambda)\, d\lambda}\,, \qquad (1.3.7)$$

notons que $\delta_h(X)$ est représenté par $(1 - r_h(\|X\|^2)/\|X\|^2)X$. Quant à la finitude du risque de $\delta_h(X)$, nous avons le résultat suivant.

Lemme 1.3.1. (Fourdrinier et al., 1998). *L'estimateur de Bayes $\delta_h(X)$ (propre ou généralisé) a un risque fini.*

Démonstration. Comme le risque de l'estimateur X est fini et égal à p, alors une application directe de l'inégalité de Schwarz nous permet d'évaluer le risque de l'estimateur de Bayes par :

$$\begin{aligned} R(\theta, \delta_h) &= E[\|X + \nabla \log m_h(X) - \theta\|^2] \\ &= E[\|X - \theta\|^2] + E[\|\nabla \log m_h(X)\|^2] \\ &\quad +2E[(X-\theta)'\ \nabla \log m_h(X)] \\ &\leq p + E[\|\nabla \log m_h(X)\|^2] \\ &\quad +2\left(E[\|X-\theta\|^2]\ E[\|\nabla \log m_h(X)\|^2]\right)^{1/2}. \end{aligned}$$

Ainsi le risque de δ_h est fini si $E[\|\nabla \log m_h(X)\|^2] < \infty$. Or, pour tout x, nous avons :

$$
\begin{aligned}
\|\nabla \log m_h(X)\|^2 &= \frac{\|\nabla m_h(X)\|^2}{m_h(X)^2} \\
&= \|X\|^2 \left(\frac{\int_0^1 \lambda^{p/2+1-a} \exp\left(-\|X\|^2\lambda/2\right) h(\lambda)\, d\lambda}{\int_0^1 \lambda^{p/2-a} \exp\left(-\|X\|^2\lambda/2\right) h(\lambda)\, d\lambda} \right)^2 \\
&\leq \|X\|^2 ,
\end{aligned}
$$

ce qui implique que $E[\|\nabla \log m_h(X)\|^2] \leq E[\|X\|^2] \leq p + \|\theta\|^2 < \infty$ où nous avons utilisé le fait que $\|X\|^2$ suit une loi de χ^2 décentrée de paramètre de décentrage $\|\theta\|^2$. Ceci complète la preuve. □

Maintenant, nous considérons l'admissibilité de l'estimateur $\delta_h(X)$. Pour utiliser le théorème 1.2.2 (de Brown), il suffit d'étudier le comportement de $m_h(x)$ quand $\|x\|^2$ tend vers l'infini. Le théorème de Tauber suivant donne une relation entre le comportement de la queue d'une fonction avec sa transformée de Laplace (voir [Wid41]).

Théorème 1.3.2. (Théorème de Tauber). *Soit g une fonction de $\mathbb{R}_+$ dans $\mathbb{R}_+$ admettant comme transformée de Laplace la fonction*

$$
f(s) = \int_0^\infty \exp(-st) g(t)\, dt .
$$

Si $g(t) \sim t^\gamma$ quand $t \to +0$, alors la fonction $f(s) \sim s^{-\gamma-1}\Gamma(\gamma+1)$ quand $s \to \infty$, où le symbole $\sim$ désigne l'équivalence entre deux fonctions au voisinage d'un point.

En utilisant ce dernier théorème, nous pouvons évaluer $m_h(x)$ par :

$$
\begin{aligned}
m_h(x) &= \int_0^1 \exp\left(-\|x\|^2\lambda/2\right) \lambda^{p/2-a} h(\lambda)\, d\lambda \\
&= 2^{p/2-a+1} \int_0^\infty \exp\left(-\|x\|^2 t\right) t^{p/2-a} h(2t) I_{(0,1/2)}(t)\, dt \\
&\sim c_1 2^{p/2-a+1} \Gamma(p/2-a+1) \|x\|^{-2(p/2-a+1)} ,
\end{aligned} \tag{1.3.8}
$$

où $c_1 = \lim_{t\to 0} h(2t)$. En posant $m_h^*(\|x\|^2) = m_h(x)$, nous avons

$$t^{-p/2} m_h^*(t)^{-1} \sim C^{-1} t^{-a+1}$$

quand $t \to \infty$, avec $C = c_1 2^{p/2-a+1}\Gamma(p/2 - a + 1)$. Alors l'intégrale

$$\int_1^\infty t^{-p/2} m_h^*(t)^{-1}\, dt$$

diverge si et seulement si $a \leq 2$. Donc, nous avons le résultat suivant.

Théorème 1.3.3. (Fourdrinier et al., 1998). *L'estimateur $\delta_h(X)$ est admissible si et seulement si $a \leq 2$.*

En outre l'utilisation du théorème 1.3.2 nous permet de calculer la limite de la fonction $r_h(w)$ quand $w \to \infty$. Comme dans (1.3.8), nous avons :

$$\int_0^1 \exp\left(-\|x\|^2\lambda/2\right) \lambda^{p/2-a+1} h(\lambda)\, d\lambda$$
$$\sim c_1 2^{p/2-a+2}\Gamma(p/2 - a + 2)\|x\|^{-2(p/2-a+2)}\,,$$

ce qui implique, quand w tend vers l'infini, que :

$$\begin{aligned} \lim_{w\to\infty} r_h(w) &= \lim_{w\to\infty} \frac{c_1 2^{p/2-a+2}\Gamma(p/2-a+2)}{c_1 2^{p/2-a+1}\Gamma(p/2-a+1)} w^{1-(p/2-a+2)+(p/2-a+1)} \\ &= 2(p/2 - a + 1)\,. \end{aligned}$$

Par conséquent, nous avons le résultat suivant.

Lemme 1.3.4. *Pour toute fonction $h(\lambda)$ vérifiant $\lim_{\lambda\to 0} h(\lambda) = c_1 > 0$, on a :*

$$\lim_{w\to\infty} r_h(w) = p - 2a + 2\,.$$

1.3.2 Construction d'estimateurs de Bayes du minimax

Dans cette section, nous considérons des conditions pour que des estimateurs de Bayes propres ou généralisés correspondant à une densité a priori de la forme (1.3.1) soient des estimateurs du minimax. Si, de plus, nous voulons déterminer des estimateurs du minimax et admissibles, d'après le théorème 1.3.3, il suffit de restreindre les résultats suivants concernant la minimaxité des estimateurs de Bayes au cas où $a \leq 2$. De plus, nous supposons que :

$$\lim_{\lambda \to 1} h(\lambda) = c_2(\geq 0)\,. \tag{1.3.9}$$

ce qui implique que la fonction $h(\lambda)$ est bornée. Avant de considérer la minimaxité de $\delta_h(X)$, nous étudions les propriétés du comportement de $r_h(w)$.

Théorème 1.3.5. *(a) Si la fonction $\lambda h'(\lambda)/h(\lambda)$ avec $h(\lambda)$ bornée est décroissante par rapport à λ, alors $r_h(w)$ est une fonction monotone croissante par rapport à w.*

(b) Si la fonction $\lambda h'(\lambda)/h(\lambda)$ avec $h(\lambda)$ bornée est croissante par rapport à λ, alors $r_h(w)$ est une fonction croissante par rapport à w de l'origine à un certain point, puis décroissante à partir de ce point.

Il faut noter que l'hypothèse de la partie (a) implique que $h'(\lambda) \leq 0$ et l'hypothèse de la partie (b) implique que $h'(\lambda) \geq 0$ puisque la valeur de la fonction $\lambda h'(\lambda)/h(\lambda)$ est égale à zéro en $\lambda = 0$. Nous donnons aussi un résultat utile dans la démonstration de ce théorème. Il s'agit de l'inégalité de covariance bien connue (voir [LC99]).

Lemme 1.3.6. (Inégalité de covariance). *Soit Y une variable aléatoire et $g(y)$ et $h(y)$ deux fonctions réelles telles que les espérances $E[g(Y)]$, $E[h(Y)]$ et $E[g(Y)h(Y)]$ existent.*

(a) Si la fonction $g(.)$ est croissante et si la fonction $h(\ )$ est décroissante, alors

$$E[g(Y)]E[h(Y)] \geq E[g(Y)h(Y)] \, .$$

(b) Si les deux fonctions sont simultanément, soit décroissantes soit croissantes, alors

$$E[g(Y)]E[h(Y)] \leq E[g(Y)h(Y)] \, .$$

Dans les deux cas, l'égalité est atteinte si et seulement si la fonction $g(\cdot)$ ou la fonction $h(\cdot)$ est constante ou bien si Y a une distribution dégénérée.

Démonstration du théorème 1.3.5. En appliquant une intégration par parties, validée par la condition (1.3.9) au numérateur du membre de droite de (1.3.7), nous obtenons :

$$\begin{aligned}\int_0^1 \lambda^{p/2+1-a} \exp(-w\lambda/2)\, h(\lambda)\, d\lambda &= -\frac{2}{w} \left[\lambda^{p/2+1-a} \exp(-w\lambda/2)\, h(\lambda)\right]_0^1 \\ &+ \frac{p+2-2a}{w} \int_0^1 \lambda^{p/2-a} \exp(-w\lambda/2)\, h(\lambda)\, d\lambda \\ &+ \frac{2}{w} \int_0^1 \lambda^{p/2+1-a} \exp(-w\lambda/2)\, h'(\lambda)\, d\lambda \, . \quad (1.3.10)\end{aligned}$$

Cependant, (1.3.10) permet de réécrire $r_h(w)$ sous la forme :

$$\begin{aligned} r_h(w) = 2(p/2-a+1) &+ 2\, \frac{\int_0^1 \lambda^{p/2+1-a} \exp(-w\lambda/2)\, h'(\lambda)\, d\lambda}{\int_0^1 \lambda^{p/2-a} \exp(-w\lambda/2)\, h(\lambda)\, d\lambda} \\ &- \frac{2c_2 \exp(-w/2)}{\int_0^1 \lambda^{p/2-a} \exp(-w\lambda/2)\, h(\lambda)\, d\lambda} \, . \end{aligned}$$

En posant $\varphi(w) = \varphi_1(w)/\varphi_2(w)$ où :

$$\varphi_1(w) = \exp(w/2)\left(\int_0^1 \lambda^{p/2+1-a} \exp(-w\lambda/2)\, h'(\lambda)\, d\lambda - c_2 \exp(-w/2)\right).$$

et où :

$$\varphi_2(w) = \exp(w/2)\left(\int_0^1 \lambda^{p/2-a} \exp(-w\lambda/2)\, h(\lambda)\, d\lambda\right),$$

et en posant :

$$\psi(w) = \frac{\varphi_1'(w)}{\varphi_2'(w)} = \frac{\int_0^1 \lambda^{p/2+1-a}(1-\lambda)\exp(w(1-\lambda)/2)\, h'(\lambda)\, d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda)\exp(w(1-\lambda)/2)\, h(\lambda)\, d\lambda},$$

il vient $r_h(w) = 2(p/2 - a + 1) + 2\varphi(w)$. De plus :

$$\varphi'(w) = \varphi_2'(w)(\psi(w) - \varphi(w)/\varphi_2(w)\,. \qquad (1.3.11)$$

Notons que $\varphi(0) = -p/2 + a - 1$ et

$$\begin{aligned} \psi(0) &= \frac{\int_0^1 \lambda^{p/2+1-a}(1-\lambda)h'(\lambda)\, d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda)h(\lambda)\, d\lambda} \\ &= -p/2 + a - 1 + \frac{\int_0^1 \lambda^{p/2+1-a} h(\lambda)\, d\lambda}{\int_0^1 \lambda^{p/2-a}(1-\lambda)h(\lambda)\, d\lambda}. \end{aligned}$$

ce qui implique que $\varphi(0) < \psi(0)$. De plus, nous avons en vertu du théorème 1.3.2 :

$$\lim_{w\to\infty} \varphi(w) = \lim_{w\to\infty} \psi(w) = 0\,. \qquad (1.3.12)$$

Maintenant, nous commençons par montrer la partie (a). En utilisant le lemme 1.3.6, nous voyons que la fonction $\psi(w)$ est strictement croissante par rapport à w de $\psi(0)(<0)$ à $\psi(\infty) = 0$. Par conséquent, si nous supposons l'existence d'un point w_0 tel que $\varphi(w_0) = \psi(w_0)$, alors nécessairement nous avons $\varphi'(w_0) \geq \psi'(w_0) > 0$. Ce qui est en contradiction avec (1.3.11). Donc, pour tout $w \geq 0$, nous avons $\varphi(w) < \psi(w)$, c'est à dire que $\varphi(w)$ est une fonction croissante.

Pour la partie (b), nous utilisons encore une fois le lemme 1.3.6 et nous voyons que la fonction $\psi(w)$ est strictement décroissante par rapport à w de $\psi(0)(> 0)$ à $\psi(\infty) = 0$. Il faut noter que, pour w assez grand, $\varphi(w)$ est positif. Donc, si nous supposons que $\varphi(w) \leq \psi(w)$ pour tout w, alors (1.3.12) implique que $\varphi(w) \leq 0$ pour tout w. Ce qui est en contradiction avec le fait que $\varphi(w)$ est positive pour w assez grand. D'où l'existence d'au moins un point w_1 pour lequel $\varphi(w_1) > \psi(w_1)$. Ensuite, comme $\psi(w)$ est décroissante, alors il existe un seul point $w_2 < w_1$ pour lequel $\varphi(w_2) = \psi(w_2)$ et par conséquent, pour tout w tel que $0 < w < w_2$ nous avons $\varphi(w) < \psi(w)$. Pour les points w tel que $w > w_2$, de la même manière que dans la partie (a), nous montrons que $\varphi(w) > \psi(w)$. Ce qui complète la preuve. □

Maintenant, en combinant les conditions (a) et (b) du corollaire 1.2.7, le lemme 1.3.4 et la partie (a) du lemme 1.3.6, nous obtenons le résultat suivant qui est dû à Faith [Fai78].

Proposition 1.3.7. (Faith, 1978). *Si $h(\lambda)$ est une fonction positive bornée telle que la fonction $\lambda h'(\lambda)/h(\lambda)$ soit monotone décroissante, alors l'estimateur de Bayes généralisé δ_h avec $3 - p/2 \leq a < 1 + p/2$ est un estimateur du minimax et $r_h(w)$ est une fonction monotone croissante par rapport à w.*

La condition suffisante de minimaxité représentée par l'inégalité (1.2.8) est impliquée par les conditions (a) et (b) du corollaire 1.2.7. Donc en utilisant la condition (1.2.8), qui est équivalente à (1.2.6), nous pouvons construire une large classe d'estimateurs de Bayes (généralisés et propres) du minimax, ce qui est donné par le résultat suivant.

Théorème 1.3.8. (Fourdrinier et al., 1998). *Soit $h(\lambda)$ une fonction positive et bornée telle que la fonction $\lambda h'(\lambda)/h(\lambda)$ se décompose en $h_1(\lambda)+h_2(\lambda)$ où $h_1(\cdot)$ est une fonction croissante par rapport à λ et $h_1(\lambda) \leq H_1$ alors que $0 \leq h_2(\lambda) \leq H_2$ avec $H_1 + 2H_2 \leq p/2 + a - 3$. Alors l'estimateur de Bayes généralisé δ_h est un estimateur du minimax. De plus, si $\lambda^{-a}h(\lambda)$ est intégrable, alors l'estimateur de Bayes propre qui en découle est un estimateur du minimax.*

Il faut noter que la proposition 1.3.7 est un cas particulier de ce théorème. Rappelons aussi que la condition de croissance de Fourdrinier et al.[FSW98] $\lim_{\lambda\to 0} \lambda^{p/2-a+2}h(\lambda) = 0$ est satisfaite, puisque nous avons supposé que $a < p/2 + 1$ pour l'existence de l'estimateur de Bayes et $\lim_{\lambda\to 1} h(\lambda) = c_2 < \infty$. Enfin, notons que, pour la démonstration de ce théorème l'intégration par parties (1.3.10) est essentielle. Pour cela, la condition "$h(\lambda)$ bornée" est indispensable.

Démonstration. En posant $g(x) = \nabla \log m_h(x)$ dans (1.2.6), la condition suffisante de minimaxité est équivalente à :

$$\frac{\Delta m_h(x)}{\|\nabla m_h(x)\|} - \frac{1}{2}\frac{\|\nabla m_h(x)\|}{m_h(x)} \leq 0\,.$$

En utilisant (1.3.4) et (1.3.5), l'inégalité précédente devient :

$$p - \|x\|^2 \frac{\int_0^1 \lambda^{p/2+2-a} \exp\left(-\|x\|^2\lambda/2\right) h(\lambda)\, d\lambda}{\int_0^1 \lambda^{p/2+1-a} \exp\left(-\|x\|^2\lambda/2\right) h(\lambda)\, d\lambda} + \frac{\|x\|^2}{2} \frac{\int_0^1 \lambda^{p/2+1-a} \exp\left(-\|x\|^2\lambda/2\right) h(\lambda)\, d\lambda}{\int_0^1 \lambda^{p/2-a} \exp\left(-\|x\|^2\lambda/2\right) h(\lambda)\, d\lambda} \geq 0\,. \quad (1.3.13)$$

Nous posons $s = \|x\|^2/2$ et, d'une manière similaire à (1.3.10), nous appliquons une intégration par parties au numérateur du deuxième et troisième terme du membre de gauche dans (1.3.13). Alors cette dernière devient :

$$a+p/2-3+\left[\frac{2c_2\exp(-s)}{\int_0^1 \lambda^{p/2+1-a}\exp(-s\lambda)\,h(\lambda)\,d\lambda}+\frac{c_2\exp(-s)}{\int_0^1 \lambda^{p/2-a}\exp(-s\lambda)\,h(\lambda)\,d\lambda}\right]$$
$$-\frac{\int_0^1 \lambda^{p/2+2-a}\exp(-s\lambda)\,h'(\lambda)\,d\lambda}{\int_0^1 \lambda^{p/2+1-a}\exp(-s\lambda)\,h(\lambda)\,d\lambda}+\frac{\int_0^1 \lambda^{p/2+1-a}\exp(-s\lambda)\,h'(\lambda)\,d\lambda}{\int_0^1 \lambda^{p/2-a}\exp(-s\lambda)\,h(\lambda)\,d\lambda}\geq 0\,. \tag{1.3.14}$$

Il est clair que le terme entre crochets dans (1.3.14) est positif, puisque le second dénominateur est plus grand que le premier. En posant, pour s fixé :

$$g_k(\lambda)=\frac{\lambda^k\exp(-s\lambda)\,h(\lambda)}{\int_0^1 \lambda^k\exp(-s\lambda)\,h(\lambda)\,d\lambda},$$

la condition suffisante de minimaxité de l'estimateur de Bayes δ_h est donnée par :

$$a+p/2-3-2E_{p/2+1-a}\left(\lambda h'(\lambda)/h(\lambda)\right)+E_{p/2-a}\left(\lambda h'(\lambda)/h(\lambda)\right)\geq 0\,, \tag{1.3.15}$$

où, E_k représente l'espérance par rapport à la densité $g_k(\lambda)$. Soit $\lambda h'(\lambda)/h(\lambda)$ $= h_1(\lambda)+h_2(\lambda)$, où $h_1(\cdot)$ est une fonction croissante par rapport à λ et $h_1(\lambda)\leq H_1$ alors que $0\leq h_2(\lambda)\leq H_2$. Comme la famille $(g_k(\lambda))$ admet la propriété de rapport de vraisemblance monotone (croissante) par rapport à k et comme h_1 est croissante et bornée par H_1, alors $2E_{p/2+1-a}[h_1(\lambda)]-E_{p/2-a}[h_1(\lambda)]\leq E_{p/2-a}[h_1(\lambda)]\leq H_1$. De plus, comme $0\leq h_2(\lambda)\leq H_2$, alors $2E_{p/2+1-a}[h_2(\lambda)]-E_{p/2-a}[h_2(\lambda)]\leq 2H_2$. Cependant, le membre de gauche dans (1.3.15) est minoré par $a+p/2-3-H_1-2H_2$, et ce dernier est positif d'après l'hypothèse du théorème . Donc l'inégalité (1.3.15) est vraie et l'estimateur de Bayes est un estimateur du minimax. Si $\lambda^{-a}h(\lambda)$ est intégrable, alors l'estimateur de Bayes propre est du minimax. Ce qui complète la preuve. □

Dans le cas de la partie (b) du théorème 1.3.5, où $\lambda h'(\lambda)/h(\lambda)$ est croissante, nous avons le résultat suivant en posant $h_1(\lambda) = 0$ dans le théorème 1.3.8.

Corollaire 1.3.9. *Supposons que la fonction positive bornée $h(\lambda)$ soit telle que la fonction $\lambda h'(\lambda)/h(\lambda)$ soit monotone croissante. Supposons aussi que $3 - p/2 < a < 1 + p/2$ et que $h'(1) \leq c_2(a + p/2 - 3)$. Alors l'estimateur de Bayes δ_h est un estimateur du minimax et la fonction $r_h(w)$ admet un seul extremum.*

Un autre résultat se déduit facilement du théorème 1.3.8, très utile dans la construction d'estimateurs de Bayes du minimax.

Corollaire 1.3.10. *Soit ψ une fonction continue qui peut être décomposée en $\psi_1 + \psi_2$, où ψ_1 est une fonction croissante et majorée par C, et où $0 < \psi_2 \leq D$ avec $C + 2D \leq 0$. Définissons*

$$h(\lambda) = \exp\left[\frac{1}{2}\int_{\lambda_0}^{\lambda} \frac{2\psi(u) + p + 2a - 6}{u}\, du\right] \quad \forall\ 0 < \lambda < 1\,, \quad (1.3.16)$$

où λ_0 est un réel donné de l'intervalle $]0, 1[$. Alors l'estimateur de Bayes δ_h qui en découle est un estimateur du minimax. De plus, si $\lambda^{-a}h(\lambda)$ est intégrable, alors l'estimateur de Bayes propre est un estimateur du minimax.

Démonstration. Notons que $\lambda h'(\lambda)/h(\lambda) = \psi_1(\lambda) + \psi_2(\lambda) + p/2 + a - 3$. En posant $h_1(\lambda) = \psi_1(\lambda) + p/2 + a - 3$ et $h_2(\lambda) = \psi_2(\lambda)$, et avec $H_1 = C + p/2 + a - 3$ et $H_2 = D$, le résultat se déduit du théorème 1.3.8. □

1.4 Estimateurs de Bayes du minimax dans le cas non gaussien

1.4.1 Certaines distributions sphériques

Dans cette section, nous considérons le problème de la construction d'estimateurs de Bayes du minimax, pour une classe particulière de familles possédant un paramètre de position θ et dont la densité est donnée par :

$$f(\|x-\theta\|^2) = \int \frac{1}{(2\pi\sigma^2)^{p/2}} \exp\left(-\frac{1}{2\sigma^2}\|x-\theta\|^2\right) dG(\sigma), \quad (1.4\ 1)$$

où $G(\cdot)$ est une distribution de probabilité quelconque définie sur $(0, \infty)$. Or, cette classe ne représente pas toutes les familles de paramètres de position à symétrie sphérique unimodale, mais elle est vaste dans le sens où un choix approprié de $G(\cdot)$ rend tous les moments d'ordre supérieur négligeables à partir d'un certain rang. Ce qui explique le fait que cette classe contient les distributions à queues légères et les distributions à queues lourdes. Pour cette classe de distributions, nous avons un résultat analogue à celui de Baranchik [Bar64b, Bar64a].

Théorème 1.4.1. (Strawderman, 1974). *Soit X une observation multidimensionnelle provenant d'une distribution de densité du type* (1.4.1) *avec $p \geq 3$ et telle que $E_0(\|X\|^2) < \infty$ et $E_0(1/\|X\|^2) < \infty$, où E_0 est l'espérance calculée par rapport à la densité $f(\|x\|^2)$. Soit δ un estimateur de la forme :*

$$\delta(X) = (1 - ar(\|X\|^2)/\|X\|^2)X\,,$$

où $r(\cdot)$ est une fonction croissante, $0 \leq r(.) \leq 1$ et telle que $r(w)/w$ est décroissante par rapport à w. Alors $\delta(X)$ est un estimateur du minimax sous coût quadratique pour tout $0 \leq a \leq 2/E_0(1/\|X\|^2)$.

En vue de la détermination d'estimateurs de Bayes, considérons une famille de densités a priori généralisées pour θ, par rapport à la mesure de Lebesgue, donnée par $g(\theta) = \|\theta\|^{2-p+\epsilon}$. Cette famille de lois a priori a été utilisé dans le cas gaussien par Baranchik [Bar64a]. L'estimateur de Bayes généralisé de θ associé à la loi a priori définie ci-dessus est donné par $\delta(X) = (\delta_1(X), \delta_2(X), ..., \delta_p(X))$ avec :

$$\begin{aligned}
\delta_i(X) &= \frac{\int \left[\int \theta_i (e^{-(1/2\sigma^2)\|X-\theta\|^2}/\sigma^p)\|\theta\|^{2-p+\epsilon}\, dG(\sigma)\right]\, d\theta}{\int \left[\int (e^{-(1/2\sigma^2)\|X-\theta\|^2}/\sigma^p)\|\theta\|^{2-p+\epsilon}\, dG(\sigma)\right]\, d\theta} \\
&= \frac{\int \left[e^{-\|X\|^2/2\sigma^2}\sigma^2 \frac{\partial}{\partial x_i} e^{\|X\|^2/2\sigma^2} E(\|\theta\|^{2-p+\epsilon})\right]\, dG(\sigma)}{\int E(\|\theta\|^{2-p+\epsilon})\, dG(\sigma)},
\end{aligned}$$

où, $\|\theta\|^2/\sigma^2$, pour σ^2 et X fixés, a une distribution de χ^2 décentrée avec p degré de liberté et $\|X\|^2/2\sigma^2$ comme paramètre de décentrage. Après calcul, nous obtenons :

$$\begin{aligned}
\delta_i(X) &= \frac{\int \left[e^{-\|X\|^2/2\sigma^2}\sigma^2 \frac{\partial}{\partial x_i} \sum_{k=0}^{\infty} \frac{(\|X\|^2/2\sigma^2)^k}{k!} E(\sigma^2\chi^2_{p+2k})^{(2-p+\epsilon)/2}\right]\, dG(\sigma)}{\int \left[e^{-\|X\|^2/2\sigma^2}\sigma^2 \sum_{k=0}^{\infty} \frac{(\|X\|^2/2\sigma^2)^k}{k!} E(\sigma^2\chi^2_{p+2k})^{(2-p+\epsilon)/2}\right]\, dG(\sigma)} \\
&= X_i\Bigg(1 - \frac{p-2-\epsilon}{2\|X\|^2} \\
&\quad \times \frac{\int \left[e^{-\|X\|^2/2\sigma^2}\|X\|^2\sigma^{2-p+\epsilon} \sum_{k=0}^{\infty} \frac{(\|X\|^2/2\sigma^2)^k\Gamma(k+1+\epsilon/2)}{k!\Gamma((p+2k+2)/2)}\right]\, dG(\sigma)}{\int \left[e^{-\|X\|^2/2\sigma^2}\sigma^{2-p+\epsilon} \sum_{k=0}^{\infty} \frac{(\|X\|^2/2\sigma^2)^k\Gamma(k+1+\epsilon/2)}{k!\Gamma((p+2k+2)/2)}\right]\, dG(\sigma)}\Bigg).
\end{aligned}$$

Maintenant, pour donner un résultat de minimaxité de cette classe d'estimateurs de Bayes généralisés, nous supposons de plus que la distribution de la variable $\eta = 1/\sigma^2$ est absolument continue par rapport à la mesure de Lebesgue avec fonction de densité $h(\eta)$. Le résultat est le suivant.

Théorème 1.4.2. (Strawderman, 1974). *Si nous supposons que la fonction de densité f de la variable $\eta = 1/\sigma^2$ soit telle que :*

(i) $\frac{1}{\beta}\, f(\eta/\beta)$ *possède la propriété du rapport de vraisemblance monotone par rapport à* η *quand elle est vue comme une famille de distributions à paramètre d'échelle.*

(ii) $0 < \int \sigma^2\, dG(\sigma) \leq 2/\int \frac{1}{\sigma^2}\, dG(\sigma) < \infty$.

Alors l'estimateur de Bayes généralisé $\delta(X)$ *est un estimateur du minimax*

Par conséquent, pour cette classe de distributions de X, nous avons une classe d'estimateurs de Bayes associés aux lois a priori (généralisées) qui correspondent à une distribution uniforme des points $\|\theta\|^{2+\epsilon}$ sur la droite des réels positifs, et nous concluons que ces estimateurs de Bayes sont du minimax pour $0 \leq \epsilon \leq p-2$ et pour certaines distributions absolument continues $G(\cdot)$ vérifiant les conditions du théorème 1.4.2.

1.4.2 Une large classe de densités

Dans cette section, nous considérons une classe plus large et plus complète de distributions absolument continues. Cette classe représente une extension de la classe de distributions discutée dans la section précédente. Pour ce type de densités de X, nous donnons un résultat de minimaxité concernant des estimateurs de type :

$$\delta(X) = \left(1 - \frac{r(\|X\|^2)}{\|X\|^2}\right) X\,. \tag{1.4.2}$$

Il faut noter aussi que ce résultat représente une autre approche pour déduire la minimaxité.

Théorème 1.4.3. (Berger, 1975). *Soit* $f(\|x-\theta\|^2)$ *une densité de probabilité, par rapport à la mesure de Lebesgue, telle que* *(a)* $E_0(\|X\|^2) < \infty$ *et* $E_0(\|X\|^{-2}) < \infty$*, où* E_0 *est l'espérance calculée par rapport à la densité* $f(\|x\|^2)$.

(b) L'ensemble de points, W*, de* $(0, \infty)$ *pour lesquelles* $f(\cdot)$ *est discontinue, a une mesure de Lebesgue nulle.*

(c) $c = \inf_{s \in U} \dfrac{\int_s^\infty f(v)\,dv}{f(s)} > 0$*, où* $U = \{s \notin W : f(s) > 0\}$.

Soit un estimateur du type (1.4.2)*, où la fonction* $r(\cdot)$ *est croissante et où* $0 \leq r \leq c(p-2)$. *En supposant aussi que* $p \geq 3$*, alors* δ *est un estimateur du minimax sous coût quadratique.*

Une application directe du théorème 1.4.3 dans le cas des distributions de mélanges de lois normales, nous donne un deuxième résultat de minimaxité avec le théorème 1.4.1 de Strawderman. Il faut noter que ce résultat est pratique lorsque la décroissance de la fonction $r(w)/w$ par rapport à w n'est pas assurée. Nous énonçons ce résultat en posant $v = \sigma^2$ dans (1.4.1).

Corollaire 1.4.4. *Soit* $f(\|x-\theta\|^2)$ *une densité de probabilité, par rapport à la mesure de Lebesgue, de type* (1.4.1) *avec* $p \geq 3$. *Soit* δ *un estimateur du type* (1.4.2)*, où la fonction* $r(\cdot)$ *est croissante. Alors* δ *est un estimateur du minimax sous coût quadratique si* $0 \leq r \leq c(p-2)$ *où* $c = 2E[V^{1-p/2}]/E[V^{-p/2}]$.

Enfin il faut noter que, dans le cas gaussien, Strawderman([Str70] a montré que. pour p égal à 3 ou 4, il n'existe aucun estimateur de Bayes propre du minimax du type (1.4.2). De plus, pour $p = 1, 2$, l'estimateur usuel X est l'unique estimateur du minimax. mais il n'est pas de Bayes. Donc l'existence des estimateurs de Bayes propres du minimax n'est envisageable qu'à partir de $p \geq 5$. Pour cela, dans le chapitre 2, la construction d'estimateurs de Bayes propres du minimax ne sera possible que pour les dimensions $p \geq 5$.

1.5 Estimation avec contraintes

Les problèmes d'estimation de paramètres avec restriction sur l'espace paramétrique ont suscité un intérêt croissant. Plusieurs auteurs se sont intéressés à ces problèmes basés sur des notions classiques d'optimalité, dans la théorie de la décision statistique, à savoir la minimaxité et l'admissibilité.

Nous abordons ici le problème d'estimation sous coût quadratique du paramètre de centralité θ d'une distribution multidimensionnelle à symétrie sphérique, basé sur une observation X avec des contraintes sur l'espace paramétrique Θ. Dans plusieurs cas, il existe de l'information a priori sur les valeurs possibles que peut prendre le vecteur de la moyenne. Cependant, les estimateurs usuels du problème sans contrainte, tels que l'estimateur sans biais $\delta_{SB}(x) = x$, les estimateurs du type James-Stein et ses dérivés ne sont ni minimax, ni admissibles pour les problèmes avec contraintes. Ceci prouve l'existence d'un nombre d'alternatives basées sur l'information a priori définie sur le vecteur de la moyenne.

Les premiers développements de ce problème concernaient l'étude de l'estimateur du maximum de vraisemblance, des résultats sont donnés par Brunk [Bru55, Bru58] et van Eeden [vE56, vE57a, vE57b, vE57c, vE58]. Le problème d'inadmissibilité a été largement étudié, voir les travaux de Sacks [Sac63]. Ensuite, dans une série d'articles, Charras et van Eeden [CE91a, CE91b, CE92, CE94] ont établi une variété de résultats concernant l'inadmissibilité sous le coût quadratique pour des estimateurs frontières dans des espaces paramétriques convexes. Comme une suite naturelle à ce problème, les alternatives intéressantes à explorer sont les estimateurs de Bayes et leurs performances fréquentistes, dans ce sens des développements sont donnés par Katz [Kat61], Farrell [Far64], Casella et Strawderman [CS81] et Berry [Ber90]. D'autres développements historiques liés au problème sont donnés dans l'article de revue de Marchand et Strawderman [MS04] et dans la monographie de van Eeden [vE06].

Une problématique consiste à déterminer des estimateurs qui améliorent les estimateurs classiques tels que l'estimateur du maximum de vraisemblance δ_{emv} ou le meilleur estimateur sans biais à variance minimale dans le problème sans contrainte. Dans le but de trouver des estimateurs efficaces et d'évaluer leur performance. Certainement, l'étude de ce problème traite aussi les questions de minimaxité et d'admissibilité. Dans le cas où l'espace contraint est non-compact, généralement ces estimateurs classiques sont du minimax dans les deux problèmes (sans contrainte et avec contrainte). Par contre, si l'espace contraint est convexe, la troncature des procédures classiques donnent des estimateurs du minimax améliorés, mais pas nécessairement admissibles. Dans ce cas, il est intéressant de trouver des estimateurs de Bayes généralisés du minimax. Le premier résultat est dû à Katz [Kat61], qui a montré, dans le cas gaussien avec une moyenne contrainte à être positive et sous le coût quadratique, que l'estimateur de Bayes généralisé associé à la loi a priori uniforme est du minimax et admissible et qu'il domine les estimateurs classiques. Récemment, Maruyama et Iwasaki [MI05] ont montré que ce dernier estimateur de Bayes peut perdre la minimaxité et l'admissibilité si la variance σ^2 n'est pas clairement spécifiée.

Casella et Strawderman [CS81] ont montré que l'unique estimateur du minimax de la moyenne θ d'une distribution normale unidimensionnelle, sous la contrainte $\theta \in [-m, m]$, est l'estimateur de Bayes associé à la loi a priori uniforme sur le bord, pour des valeurs de m suffisamment petites. Dans plusieurs cas, les estimateurs de Bayes (généralisés) forment une classe complète. Il faut aussi rappeler que, sous coût quadratique, l'estimateur de Bayes issu d'une loi a priori non dégénérée à un point, ne prend jamais ses valeurs sur le bord de l'espace paramétrique convexe C car $E(\theta \mid x)$ prend toutes ses valeurs à l'interieur du convexe C. Dans la littérature, il existe plusieurs résultats basés sur ce phénomène pour déterminer l'inadmissibilité de certains estimateurs qui prennent des valeurs au voisinage ou sur le bord (voir [Moo81, Moo85]).

Nous présentons dans cette section, des préalables et des résultats récents correspondant à nos problèmes. Ces résultats vont nous servir comme point de départ dans nos développements.

1.5.1 Résultat de Hartigan

Soit le modèle $X \sim \mathcal{N}_p(\theta, I_p)$, avec $\theta \in C$ où C est un convexe arbitraire avec un intérieur non-vide dans $\mathbb{R}^p$. Pour estimer θ, Hartigan [Har04] a démontré que l'estimateur de Bayes (généralisé) par rapport à la mesure a priori uniforme sur C domine l'estimateur sans biais X, sous la perte quadratique. Ce résultat remarquable, donné ci-dessous, s'appuie sur une décomposition du risque, sur l'identité de Stein pour obtenir un estimateur sans biais du risque ainsi que sur un théorème de divergence (voir aussi [MS04]).

Théorème 1.5.1. (Hartigan, 2004). *Pour $X \sim \mathcal{N}_p(\theta, I_p)$, où θ est contraint à être dans un convexe C de $\mathbb{R}^p$ avec un intérieur non-vide, l'estimateur de Bayes $\delta_U(X)$ associé à la loi a priori uniforme sur C domine l'estimateur sans biais X, pour le coût quadratique $\|d-\theta\|^2$.*

Démonstration. En écrivant :

$$\delta_U(X) = X + \frac{\nabla_X m(X)}{m(X)} \quad \text{avec} \quad m(X) = (2\pi)^{-p/2} \int_C e^{-\frac{1}{2}\|X-\nu\|^2} \, d\nu \, ,$$

et en utilisant l'identité (1.2.5), nous avons :

$$\begin{aligned}
& R(\theta, \delta_U(X)) - R(\theta, X) \\
&= E_\theta \left[\frac{\|\nabla_X m(X)\|^2}{m^2(X)} + div \left(\frac{\nabla_X m(X)}{m(X)} \right) + \frac{(X-\theta)' \nabla_X m(X)}{m(X)} \right] \\
&= E_\theta \left[\frac{\nabla_X^2 m(X)}{m(X)} + \frac{(X-\theta)' \nabla_X m(X)}{m(X)} \right] \\
&= E_\theta \left[\frac{1}{m(X)} H(X, \theta) \right] \, ,
\end{aligned}$$

où $\nabla_x^2 = \Delta_x$ et $H(x, \theta) = \nabla_x^2 m(x) + (x-\theta)' \nabla_x m(x)$. Il suffit de montrer alors que $H(x, \theta) \leq 0$ pour tout $x \in \mathbb{R}^p$ et $\theta \in C$. Ensuite, en remarquant que $\nabla_x \left(e^{-\frac{1}{2}\|x-\nu\|^2} \right) = -\nabla_\nu \left(e^{-\frac{1}{2}\|x-\nu\|^2} \right)$ et que $\nabla_x^2 \left(e^{-\frac{1}{2}\|x-\nu\|^2} \right) = \nabla_\nu^2 \left(e^{-\frac{1}{2}\|x-\nu\|^2} \right)$, on obtient :

$$\begin{aligned}
(2\pi)^{p/2} H(x, \theta) &= \nabla_x^2 \int_C e^{-\frac{1}{2}\|x-\nu\|^2} \, d\nu + (x-\theta)' \nabla_x \int_C e^{-\frac{1}{2}\|x-\nu\|^2} \, d\nu \\
&= \int_C \nabla_\nu^2 \left(e^{-\frac{1}{2}\|x-\nu\|^2} \right) d\nu - (x-\theta)' \int_C \nabla_\nu \left(e^{-\frac{1}{2}\|x-\nu\|^2} \right) d\nu \\
&= \int_C \nabla_\nu \left[e^{-\frac{1}{2}\|x-\nu\|^2} - (x-\theta)' e^{-\frac{1}{2}\|x-\nu\|^2} \right] d\nu \\
&= \int_C div_\nu \left((\theta-\nu)' e^{-\frac{1}{2}\|x-\nu\|^2} \right) d\nu \, . \qquad (1.5.1)
\end{aligned}$$

Maintenant, en utilisant le théorème de divergence, l'expression (1.5.1) donne :

$$(2\pi)^{p/2} H(x, \theta) = \int_{\partial C} \eta(\nu)' (\theta - \nu) e^{-\frac{1}{2}\|x-\nu\|^2} \, d\sigma(\nu) \, .$$

où $\eta(\nu)$ représente le vecteur normal à C au point $\nu \in \partial C$ et $d\sigma(\nu)$ désigne la mesure de l'aire sur ∂C. Finalement, puisque C est convexe, l'angle entre les directions $\eta(\nu)$ et $(\theta - \nu)$ est obtus, ce qui implique que $\eta(\nu)'(\theta - \nu) \leq 0$, pour tout $\theta \in C$, $\nu \in \partial C$. D'où le résultat voulu. $\square$

Les boules et les cônes convexes dans $\mathbb{R}^p$ représentent deux classes particulièrement intéressantes d'ensembles convexes, pour lesquelles le résultat de Hartigan apporte d'importantes nouvelles informations. Dans le cas des boules ayant un rayon suffisamment petit, Marchand et Perron [MP01] ont montré que la loi uniforme génère des procédures dominant l'estimateur du maximum de vraisemblance, alors que le résultat de Hartigan implique que les estimateurs de Bayes associés à l'uniforme dominent tout le temps l'estimateur sans biais X. De même, nous voyons clairement que le résultat de Katz [Kat61] devient un cas particulier du résultat de Hartigan. Par conséquent, ce résultat représente un agrandissement de l'ensemble des situations où l'estimateur de Bayes uniforme domine l'estimateur X pour un coût quadratique.

Enfin, quoique l'approche de Hartigan soit très générale par rapport au choix de l'espace paramétrique convexe C, elle reste spécifique à la loi normale et à la perte quadratique. Bien qu'il existe des résultats connus dans d'autres situations spécifiques telles le cas unidimensionnel avec $C = [0, \infty[$, des généralisations à d'autres lois sphériques (loi de Student, ou plus généralement des mélanges de lois normales, loi de Kotz, etc.) ou la considération d'autres fonctions de coût représentent un intérêt certain.

1.5.2 Méthode de Kubokawa

Kubokawa [Kub94] a introduit une nouvelle technique, basée sur une expression intégrale de la différence de risques (IERD). Ceci a permis de mettre en évidence un traitement unifié de l'estimation ponctuelle et de l'estimation par intervalle d'un paramètre d'échelle, incluant le cas particulier de l'estimation de la variance d'une loi normale. Il a appliqué cette méthode pour déterminer une classe d'estimateurs améliorant l'estimateur de James-Stein de la moyenne multidimensionnelle. Beaucoup d'autres applications ont suivi à travers les travaux de Kubokawa [Kub98, Kub99], entre autres sur l'estimation des paramètres de position et d'échelle avec contrainte. Il faut noter aussi, que cette méthode possède une robustesse particulière qui réside dans la flexibilité de s'adapter à plusieurs fonctions de coût.

Nous illustrons la méthode (Integral Expression of Risk Difference) de Kubokawa à travers l'exemple suivant. À cet effet, nous considérons une observation X générée à partir de la densité d'une famille possédant un paramètre de position, $f_\theta(x) = f_0(x-\theta)$, où f_0 est une fonction connue telle que $E_0[X] = 0$ et $E_0[X^2] < \infty$. Pour estimer θ, sous le coût quadratique $(d-\theta)^2$ et sous la contrainte $\theta \geq a$ (sans perte de généralité, nous considérons $a = 0$), nous montrons que l'estimateur de Bayes généralisé $\delta_U(X)$ associé à la loi a priori uniforme $\pi(\theta) = 1_{(0,\infty)}(\theta)$ domine l'estimateur usuel $\delta_{SB}(X) = X$. Tout d'abord, nous remarquons que l'estimateur δ_U peut s'écrire comme suit :

$$\begin{aligned}
\delta_U(x) &= x + \frac{\int_0^\infty (\theta - x) f_0(x-\theta)\, d\theta}{\int_0^\infty f_0(x-\theta)\, d\theta} \\
&= x - \frac{\int_{-\infty}^x u\, f_0(u)\, du}{\int_{-\infty}^x f_0(u)\, du} \\
&= x - E_0[X\,|X \leq x] \\
&= x + h_U(x)\,,
\end{aligned}$$

où h_U est une fonction continue, décroissante telle que :

$$h_U(x) \geq -\lim_{x\to\infty} E_0[X\,|X \leq x] = -E_0[X] = 0\,.$$

Nous disposons du résultat suivant.

Théorème 1.5.2. *Pour l'espace paramétrique restreint* $\theta \in \Theta = [0, \infty)$ *et sous le coût quadratique*

(a) Tout estimateur, $\delta_h \neq \delta_{SB}$*, de la forme* $\delta_h(x) = x + h(x)$*, où* h *est une fonction continue, positive et décroissante, domine* $\delta_{SB}(x) = x$ *si* $h(x) \leq h_U(x)$.

(b) L'estimateur δ_U *domine l'estimateur usuel* $\delta_{SB}(x) = x$.

Démonstration. Premièrement, vu les propriétés de h_U mentionnées plus haut, nous pouvons déduire la partie (b) de la partie (a). Il suffit donc de montrer la partie (a). Selon les hypothèses du théorème nous déduisons que $\lim_{x\to\infty} h(x) = 0$ et en procédant de la même manière que Kubokawa [Kub94], nous avons :

$$\begin{aligned}(x-\theta)^2 - (x+h(x)-\theta)^2 &= (x+h(y)-\theta)^2 \big|_{y=x}^{\infty} \\ &= \int_x^{\infty} \frac{\partial}{\partial y}(x+h(y)-\theta)^2 \, dy \\ &= 2\int_x^{\infty} h'(y)(x+h(y)-\theta)\, dy\,.\end{aligned}$$

Ceci implique que :

$$\begin{aligned}\Delta_h(\theta) &= E_\theta\left[(X-\theta)^2 - (X+h(X)-\theta)^2\right] \\ &= 2\int_{-\infty}^{\infty}\int_x^{\infty} h'(y)(x+h(y)-\theta) f_0(x-\theta)\, dy\, dx \\ &= 2\int_{-\infty}^{\infty} h'(y)\int_{-\infty}^{y}(x+h(y)-\theta) f_0(x-\theta)\, dx\, dy\,.\end{aligned}$$

Maintenant, pour prouver que $\Delta_h(\theta) \geq 0$ pour tout $\theta \geq 0$, il suffit de montrer que :

$$G_h(y,\theta) = \int_{-\infty}^{y}(x+h(y)-\theta) f_0(x-\theta)\, dx \leq 0$$

pour tout y et $\theta \geq 0$, puisque $h'(y) \leq 0$. Or, cette dernière inégalité est équivalente à

$$\begin{aligned}
\frac{\int_{-\infty}^{y}(x+h(y)-\theta) f_0(x-\theta)\, dx}{\int_{-\infty}^{y} f_0(x-\theta)\, dx} \leq 0 \quad &\Leftrightarrow \quad h(y) \leq \frac{\int_{-\infty}^{y-\theta} u f_0(u)\, du}{\int_{-\infty}^{y-\theta} f_0(u)\, du}, \forall \theta \geq 0 \\
&\Leftrightarrow \quad h(y) \leq \inf_{\theta \geq 0}\{-E_0[X \mid X \leq y-\theta]\} \\
&\Leftrightarrow \quad h(y) \leq -E_0[X \mid X \leq y] = h_U(y),
\end{aligned}$$

pour tout y, puisque la fonction $E_0[X \mid X \leq z]$ est croissante en z. Ce qui complète la preuve le théorème . □

Dans le cas des densités de familles possédant un paramètre de position quelconques $f_0(x-\theta)$, Farrell [Far64] a montré que l'estimateur $\delta_U(X)$ domine l'estimateur à risque minimum $\delta_{SB}(X)$ et qu'il est du minimax, sous une fonction de perte invariante $L(\theta, d) = \rho(d-\theta)$, où ρ est une fonction strictement convexe. De même, il a établi l'admissibilité de δ_U sous coût quadratique.

En utilisant la méthode de Kubokawa, Marchand et Strawderman [MS05a] ont établi une extension du théorème 1.5.2 dans le cas des fonctions de perte plus générale, de la forme $L(\theta, d) = \rho(d-\theta)$. Ils ont aussi montré, pour (f_0, ρ) quelconque, que l'estimateur à risque minimum $\delta_{SB}(X)$ est un estimateur du minimax. Pour des résultats unifiés sur la minimaxité dans les espaces paramétriques restreints voir [MS11]. Par conséquent tous les estimateurs qui dominent δ_{SB} sont du minimax, en particulier l'estimateur de Bayes $\delta_U(X)$ est un estimateur du minimax pour l'espace paramétrique restreint $\Theta = [0, \infty)$. Aussi, Marchand et Strawderman [MS05a, MS05b] ont donné un développement similaire pour les familles à paramètre d'échelle, et ils ont étendu les résultats de dominance pour les espaces paramétriques restreints de la forme $[a, b]$.

1.6 Dominance de l'EMV sur une boule

Dans cette partie, nous considérons sous la perte quadratique le problème de l'estimation du paramètre de position θ d'une distribution normale multidimensionnelle $\mathcal{N}(\theta, I_p)$, basé sur l'observation X et avec un espace paramétrique restreint $\Theta(m) = \{\theta \in \mathbb{R}^p : \|\theta\| \leq m\}$, pour une certaine constante $m > 0$.

Pour ce genre de problème, une alternative immédiate à l'estimateur sans biais $\delta_{SB}(x) = x$ est l'estimateur du maximum de vraisemblance δ_{emv}, qui est la troncature de δ_{SB} sur la boule $\Theta(m)$, donné par $\delta_{emv}(x) = \left(\frac{m}{\|x\|} \wedge 1\right) x$. Bien que cet estimateur domine δ_{SB}, nous savons déjà que les estimateurs du maximum de vraisemblance sont toujours inadmissibles sous le coût quadratique pour les espaces paramétriques contraints (voir Sacks [Sac63]). Dans notre contexte, et tel qu'il est décrit à la page 40, l'inadmissibilité de δ_{emv} se déduit aussi du travail de Charras et Van Eeden [CE91a], quoique leur preuve ne mette pas en évidence d'estimateurs dominants. Pour $p = 1$, Casella et Strawderman [CS81] ont montré, pour $m \leq 1$, que la loi a priori uniforme sur le bord est la moins favorable et que l'estimateur de Bayes correspondant est, non seulement minimax, mais domine aussi δ_{emv}.

D'autres alternatives à explorer pour ce problème, sont les estimateurs de Bayes et leur performance fréquentiste (toujours dans le but de déterminer des estimateurs améliorant δ_{emv}). Kempthorne [Kem88] a présenté des exemples numériques d'estimateurs de Bayes dominant δ_{emv} pour $p = 1$, par le moyen d'une procédure induisant des estimateurs optimaux selon le critère maximin. Deux estimateurs de Bayes sont en particulier intéressants à étudier, à savoir, l'estimateur δ_U associé à la loi uniforme sur la boule $\Theta(m)$

et l'estimateur δ_{BU} associé à la loi uniforme sur le bord $\partial\Theta(m)$, qui est un estimateur du minimax pour les petites valeurs de m.

1.6.1 Les estimateurs de Bayes uniformes

Marchand et Perron [MP01, MP05] et Fourdrinier et Marchand [FM10] se sont intéressés particulièrement à ce type d'estimateurs dans leurs travaux. Parmi leurs résultats, nous présentons ceux qui nous seront utiles pour nos développements. Notons $\|X\|$ et $\|x\|$ respectivement par R et r. Notons aussi $\|\theta\|$ par λ lorsque θ est considéré comme un paramètre et par T lorsque θ est considéré comme une variable aléatoire.

Nous considérons les estimateurs équivariants par rapport aux transformations orthogonales dont la forme est donnée par :

$$\delta_g(x) = \frac{g(r)}{r}x\,.$$

Les résultats de dominance de Marchand et Perron [MP01] sont basés sur la décomposition de la fonction risque en risques conditionnels. Ceci signifie que, si nous pouvons montrer que $E_\theta\left[\|\delta_1(X)-\theta\|^2 - \|\delta_2(X)-\theta\|^2 \,|R\right] \leq 0$, pour tout $\theta \in \Theta(m)$ et pour toutes les valeurs possibles de R, et si nous trouvons $\theta_0 \in \Theta(m)$ tel que $P_{\theta_0}\left[E_{\theta_0}\left[\|\delta_1(X)-\theta_0\|^2 - \|\delta_2(X)-\theta_0\|^2 \,|R\right] < 0\right] > 0$, alors nous pouvons en déduire que δ_1 domine δ_2.

L'expression de l'estimateur de Bayes δ_{BU} s'écrit sous la forme :

$$\delta_{BU}(x) = \frac{g_m(r)}{r}x\,,$$

où, selon Spruill [Spr86], pour tout $r > 0$ et tout $\lambda \geq 0$,

$$g_\lambda(r) = E_\theta\left[\frac{\theta' X}{\|X\|}\,\middle|\, \|X\| = r\right]\,.$$

Dans le cas gaussien où $X \sim \mathcal{N}_p(\theta, I_p)$, l'expression de g_λ est donnée par le résultat suivant.

Lemme 1.6.1. ([Berry, 1990 et Robert, 1990]) *Une expression explicite de g_λ est donnée par :*

$$g_\lambda(r) = \lambda \rho_{(p/2)-1}(\lambda r) \ ,$$

où, pour $\nu > -1$ I_ν désigne la fonction de Bessel modifiée de degré ν et $\rho_\nu(t) = I_{\nu+1}(t)/I_\nu(t)$.

Ainsi, la fonction ρ_ν joue un rôle important dans notre analyse. Des propriétés importantes de cette fonction sont données par Watson [Wat83] pour $\nu \geq 0$, et vérifiables pour $\nu = -1/2$ en utilisant la représentation $\rho_{-1/2} \equiv tanh$. Nous donnons aussi des bornes importantes de la fonction ρ_ν^2 établies par Amos [Amo74].

Lemme 1.6.2. *(a) (Watson, 1983) Pour tout $\nu \geq -1/2$, $\rho_\nu(\cdot)$ est une fonction croissante et concave sur $(0, \infty)$, avec $\lim_{t\to 0^+} \rho_\nu(t) = 0$ et $\lim_{t\to\infty} \rho_\nu(t) = 1$; $\frac{\rho_\nu(t)}{t}$ est une fonction décroissante en t, avec $\lim_{t\to 0^+} \frac{\rho_\nu(t)}{t} = \frac{1}{2(\nu+1)}$. Nous disposons aussi de l'identité $\frac{d}{dt}\rho_\nu(t) = 1 - (1 + 2\nu)\frac{\rho_\nu(t)}{t} - \rho_\nu^2(t)$ et de l'inégalité $\frac{d}{dt}\rho_\nu(t) \leq \frac{\rho_\nu(t)}{t}$.*

(b) (Amos, 1974) Pour tout $\nu \geq 0$ et $t > 0$, on a :

$$L\left(\frac{2(\nu+1)}{t}, \frac{2(\nu+1)}{t}\right) \leq \rho_\nu^2(t) \leq L\left(\frac{2\nu}{t}, \frac{2(\nu+2)}{t}\right), \tag{1.6.1}$$

où $L(a,b) = \{a/2 + \sqrt{1 + (b/2)^2}\}^{-2}$

Maintenant, comme les résultats présentés sont basés sur les risques conditionnels sachant R, il est judicieux, au regard de ce qui précède, de considérer la densité $f_p(\cdot, \lambda)$ de R, qui est la densité de la racine carrée du khi-deux décentrée de paramètre de décentrage λ^2 au travers d'une fonction de Bessel modifiée. La densité $f_p(\cdot, \lambda)$ est donnée par (voir Robert [Rob90]).

Lemme 1.6.3. (Robert, 1990). *Soit $r > 0$ et $\lambda > 0$. Alors*

$$f_p(r, \lambda) = r\left(\frac{r}{\lambda}\right)^{(p/2)-1} I_{(p/2)-1}(\lambda r)\exp\left(-\left(\lambda^2 + r^2\right)/2\right) .$$

De plus, certains des risques conditionnels utilisés dans les résultats de dominance mettent en jeu les espérances suivantes :

$$\alpha(m, \lambda) = E_\theta\left[\rho_{p/2-1}(\lambda r) \mid R > m\right] \text{ et } \beta(m, \lambda) = E_\theta\left[\frac{\lambda}{R}\rho_{p/2-1}(\lambda r) \mid R \leq m\right] ;$$

aussi bien que les fonctions suivantes :

$$\bar{\alpha}(m) = \sup_{0<\lambda\leq m} \alpha(m, \lambda) \quad \text{et} \quad \bar{\beta}(m) = \sup_{0<\lambda\leq m} \beta(m, \lambda) .$$

Des propriétés de $\bar{\alpha}$ et $\bar{\beta}$ sont données dans le prochain résultat.

Lemme 1.6.4. (Marchand et Perron, 2001).

(a) La fonction $\bar{\alpha}(\cdot, \cdot)$ est croissante par rapport à ses deux arguments et, par conséquent, $\bar{\alpha}(m) = \alpha(m, m)$. De plus, $\lim_{m\to 0} \bar{\alpha}(m) = 0$ et $\lim_{m\to\infty}$ $\bar{\alpha}(m) = 1$.

(b) La fonction $\bar{\beta}(m, \cdot)$ est croissante et, par conséquent, $\bar{\beta}(m) = \beta(m, m)$. De plus, $0 \leq \bar{\beta}(m) < m^2/p$ et $\lim_{m\to\infty} \bar{\beta}(m) \geq 1$.

Les résultats de dominance, obtenus en utilisant les décompositions en risques conditionnels, sont donnés dans les théorèmes suivants.

Théorème 1.6.5. (Marchand et Perron, 2001). *Si $g(\cdot)$ est une fonction croissante sur $(m, \infty]$ telle que $2(\bar{\alpha}(m) - 1)m < g(r) < m$, pour tout $r \in (m, \infty]$, alors*

$$E_\theta \left[\|\delta_{emv}(X) - \theta\|^2 - \|\delta_g(X) - \theta\|^2 \mid R > m \right] > 0\,,$$

pour tout $\theta \in \Theta(m)$.

Charras et van Eeden [CE91a] présentent, pour $p = 1$, une classe d'estimateurs de la forme $\delta_\epsilon(x) = g_\epsilon(r)\frac{x}{r}$, avec $g_\epsilon(r) = (m - \epsilon) I_{(m,\infty)}(r) + r I_{[0,m]}(r)$. Ces estimateurs sont équivariants par rapport aux transformations orthogonales, mais non monotones en $\|x\|$. D'autre part, pour ϵ suffisamment petit, nous avons $2(\bar{\alpha}(m) - 1)m < g_\epsilon(r) < m$, pour tout $r > m$. D'où, selon le théorème 1.6.5, pour tout ϵ suffisamment petit, nous avons $E_\theta \left[\|\delta_{emv}(X) - \theta\|^2 - \|\delta_{g_\epsilon}(X) - \theta\|^2 \mid R > m \right] > 0$, pour tout $\theta \in \Theta(m)$.

Théorème 1.6.6. (Marchand et Perron, 2001). *Supposons que $g(r)$ et $r(r - g(r))$ soient des fonctions croissantes pour $r \in (0, m]$. Si $\bar{\beta}(m) < 1$ et $2\left(\bar{\beta}(m) - 1\right)r < g(r) < r$, pour tout $r \in (0, m]$, alors*

$$E_\theta \left[\|\delta_{emv}(X) - \theta\|^2 - \|\delta_g(X) - \theta\|^2 \mid R \leq m \right] > 0\,,$$

pour tout $\theta \in \Theta(m)$.

Le résultat suivant, qui est une combinaison des deux théorèmes précédents, donne la première série de conditions sur la fonction g pour lesquelles l'estimateur associé δ_g domine δ_{emv}.

Corollaire 1.6.7. *Supposons que $g(\cdot)$ soit une fonction croissante telle que $g(r)/r$ soit décroissante sur $(0, \infty]$. Si $0 \leq g(r) \leq r \wedge m$ pour tout $r > 0$ et, si $g(m)/m \geq 2\left(\bar{\alpha}(m) \vee \bar{\beta}(m)\right) - 1$, alors δ_g domine δ_{emv}.*

Théorème 1.6.8. (Marchand et Perron, 2001). *Si, pour tout $r > r_1$, avec $r_1 = \inf\{r : g_m(r)/r < 1, r > 0\}$, on a $2g_m(r) - r \wedge m < g(r) < r$, alors*

$$E_\theta\left[\|\delta_{emv}(X) - \theta\|^2 - \|\delta_g(X) - \theta\|^2 \mid R = r\right] > 0,$$

pour tout $\theta \in \Theta(m)$.

Selon le lemme 1.6.2, nous avons $g_m(r)/r \leq m^2/p$. Cependant, la condition $m \leq \sqrt{p}$, implique que $g_m(r)/r \leq 1$ et, par conséquent, $r_1 = 0$. Autrement dit la condition du théorème 1.6.8 est satisfaite pour la fonction $g \equiv g_m$ et ainsi l'estimateur de Bayes δ_{BU} domine δ_{emv} si $m \leq \sqrt{p}$. Or ce dernier théorème ne s'applique pas à δ_{BU} pour $m > \sqrt{p}$. En revanche, il s'applique sur la version tronquée δ_h avec $h(r) = g_m(r) \wedge r$. Ceci implique qu'un tel δ_h domine δ_{emv}.

En combinant le théorème 1.6.5 et le théorème 1.6.8, nous obtenons le résultat suivant qui fournit une seconde série de conditions menant à la dominance.

Corollaire 1.6.9. *Si les conditions suivantes sont satisfaites :*

(a) g est une fonction croissante,

(b) $0 \leq g(r) \leq r \wedge m$, pour tout $r > 0$,

(c) $m \leq m_1 \wedge \sqrt{p/2}$, où m_1 vérifie l'équation $\bar{\alpha}(m_1) = 1/2$,

alors δ_g domine δ_{emv}.

Comme conséquences immédiates de ce corollaire, nous concluons que tous les estimateurs linéaires $\delta_a(X) = aX$ avec $0 \leq a < 1$ et tronqués en m dominent δ_{emv} pour m petit. En particulier, l'estimateur trivial δ_0 domine δ_{emv} pour m petit.

Afin de déterminer d'autres résultats de dominance, nous considérons les estimateurs de Bayes associés à des lois a priori π sur $\Theta(m)$ qui sont orthogonalement invariantes, c'est-à-dire telles que :

$$\pi(\eta) = \pi\left(g^{-1}(\eta)\right) ,$$

pour toute transformation orthogonale $\eta = g(\theta)$. À travers le théorème suivant, nous caractérisons l'estimateur de Bayes pour une telle loi a priori π.

Théorème 1.6.10. (Marchand et Perron, 2001). *Soit $T = \|\theta\|$. Pour une loi a priori π orthogonalement invariante avec $\pi(\{T=0\}) < 1$, l'estimateur de Bayes δ_π est donné par $\delta_\pi(x) = (g_\pi(r)/r)\, x$ où*

$$g_\pi(r) = \frac{\int_0^m t^{-p/2-2} I_{p/2}(rt) \exp\left(-t^2/2\right) \sigma(dt)}{\int_0^m t^{-p/2-1} I_{p/2-1}(rt) \exp\left(-t^2/2\right) \sigma(dt)} ,$$

et où σ est la loi de T sur $[0,m]$. De plus, la fonction g_π est croissante et admet une limite nulle en 0 ; de plus, $0 \leq g_\pi \leq g_m$ et $g_\pi(r)/r \leq 1$ dès que $m \leq \sqrt{p}$

Ce théorème est applicable pour caractériser l'estimateur de Bayes δ_U associé à la loi a priori uniforme sur $\Theta(m)$. Dans ce cas, la loi de T est donnée par $\sigma(dt) = (p/m^p)\, t^{p-1}\, dt$ et, par conséquent,

$$g_U(r) = rP\left[\chi^2_{p+2}\left(r^2\right) \leq m^2\right] / P\left[\chi^2_p\left(r^2\right) \leq m^2\right] .$$

D'autre part, nous pouvons déduire du théorème 1.6.10 que g_π satisfait aux conditions du corollaire 1.6.9 lorsque $m \leq m_1 \wedge \sqrt{p/2}$. Nous avons ainsi le corollaire suivant.

Corollaire 1.6.11. *Si δ_π est un estimateur de Bayes par rapport à une loi a priori orthogonalement invariante, alors δ_π domine δ_{emv} lorsque $m \leq m_1 \wedge \sqrt{p/2}$.*

Il faut noter, qu'avec ce dernier résultat, nous obtenons la dominance pour toute loi a priori orthogonalement invariante π, dès que la condition $m \leq m_1 \wedge \sqrt{p/2}$ est satisfaite et que cette condition ne nécessite que la connaissance de (m, p).

En considérant la classe des lois a priori possédant des densités proportionnelles à $\exp\left(-h\left(\|\theta\|^2\right)\right)$ et en appliquant le corollaire 1.6.7, nous décrivons les conditions à considérer sur cette classe pour induire la dominance. D'où le résultat suivant.

Théorème 1.6.12. (Marchand et Perron, 2001). *Supposons que la loi a priori π ait une densité de la forme $K \exp\left(-h\left(\|\theta\|^2\right)\right)$ où K est la constante de normalisation. Nous disposons des deux résultats suivants*

(a) Si h est une fonction croissante, alors la fonction $g_\pi(r)/r$ est majorée par 1.

(b) Si h est une fonction convexe, alors la fonction $g_\pi(r)/r$ est décroissante en r. De plus, la fonction $g_\pi(r)/r$ est majorée par 1 si et seulement si

$$\int_0^m t^{p+1} \exp\left(-\left\{h\left(t^2\right) + t^2/2\right\}\right) dt \leq p \int_0^m t^{p-1} \exp\left(-\left\{h\left(t^2\right) + t^2/2\right\}\right) dt$$

En combinant ce dernier théorème et le corollaire 1.6.7, nous obtenons le résultat suivant.

Corollaire 1.6.13. *Supposons que la loi a priori π ait une densité de la forme $K \exp\left(-h\left(\|\theta\|^2\right)\right)$, où K est la constante de normalisation, telle que :*

(i) h soit une fonction convexe,

(ii) la fonction $g_\pi(r)/r$ soit majorée par 1 et

(iii) $g_\pi(m)/m \geq 2\left(\bar{\alpha}(m) \vee \bar{\beta}(m)\right) - 1$

Alors l'estimateur de Bayes δ_π domine δ_{emv}.

Ce résultat est applicable en particulier à l'estimateur δ_U, associé à la loi a priori uniforme sur $\Theta(m)$ et qui correspond à $h = 0$. Le théorème 1.6.12 implique que la fonction $g_U(r)/r$ est majorée par 1. Par conséquent, les conditions (i) et (ii) du corollaire 1.6.13 sont satisfaites dans ce cas. De plus, l'estimateur δ_U domine δ_{emv} pour tout m tel que $g_U(m)/m \geq 2\left(\bar{\alpha}(m) \vee \bar{\beta}(m)\right) - 1$.

Idéalement, notre problème consiste à trouver les lois a priori π telles que δ_π domine δ_{emv}, pour chaque couple (m, p). Pour $m \leq \sqrt{p}$, δ_{BU} domine δ_{emv}. Par contre, δ_{BU} ne domine pas δ_{emv} pour m grand, puisque $R(0, \delta_{BU}) = E_0\left[g_m^2(R)\right]$ est croissante en m avec $R(0, \delta_{BU})/m^2 \to 1$ quand $m \to \infty$ alors que $R(0, \delta_{emv}) \leq p$. D'une manière similaire, des résultats numériques suggèrent que $R(\theta, \delta_U) > R(\theta, \delta_{emv})$ à $\|\theta\| = m$ pour m grand. Comme il a été montré dans les travaux de Gatsonis et al. [GMS87] pour $p = 1$, il semble très plausible que la comparaison des risques à $\|\theta\| = m$, implique une condition suffisante pour décider de la dominance entre δ_U et δ_{emv}. Cependant, le lemme suivant est important pour l'analyse en se basant sur les risques à l'origine et au bord de l'espace paramétrique.

Lemme 1.6.14. *Supposons que π et π' soient deux lois a priori orthogonalement invariantes sur $\Theta(m)$ et supposons que σ et σ' soient leurs lois induites sur la sphère $\|\theta\| = m$. Alors nous disposons des deux résultats suivants.*

(a) Si la densité $\frac{d\sigma}{d\sigma'}$ est croissante alors $g_\pi \geq g_{\pi'}$.

(b) Si $g_\pi \geq g_{\pi'}$ alors $R(\theta, \delta_\pi) \leq R(\theta, \delta_{\pi'})$ lorsque $\|\theta\| = m$ et $R(0, \delta_\pi) \geq R(0, \delta_{\pi'})$.

Maintenant, nous supposons que $m > \sqrt{p}$ et nous considérons la loi a priori π_a caractérisée par $h(t) = -at^2/2$, où la constante positive a est l'unique solution de l'équation

$$\int_0^m t^{p+1} \exp\left\{(a-1)\, t^2/2\right\}\, dt = p \int_0^m t^{p-1} \exp\left\{(a-1)\, t^2/2\right\}\, dt\,.$$

Ce choix permet à π_a de satisfaire les conditions (i) et (ii) du corollaire 1.6.13, mais ceci ne garanti en rien que l'estimateur associé δ_{π_a} domine δ_{emv}. D'autre part, le lemme 1.6.14 implique que $g_{\pi_a}(m) > g_{\pi_0}(m)$, où π_0 est la loi a priori uniforme sur $\Theta(m)$, puisque $\frac{d\pi_a}{d\pi_0}(\theta) = \exp\left\{a\,\|\theta\|^2/2\right\}$. Autrement dit $g_{\pi_a}(m)$ peut prendre des grandes valeurs et ainsi la condition (iii) du corollaire 1.6.13 peut se réaliser. Cependant, nous obtenons une classe d'estimateurs qui dominent δ_{emv} pour $m > \sqrt{p}$.

1.6.2 Les estimateurs de Bayes uniformes sur les sphères

D'autres estimateurs de Bayes intéressants sont à considérer, à savoir les estimateurs de Bayes δ_α associés aux lois a priori π_α uniformes sur les sphères S_α, pour $0 \leq \alpha \leq m$. Le choix est particulièrement intéressant, sachant que la valeur de rétrécissement est calibrée par le choix de α avec comme extrêmes $\delta_m \equiv \delta_{BU}$, et $\delta_0 \equiv 0$ (associé à la loi a priori dégénérée en 0). De plus, la détermination des conditions de dominance des estimateurs δ_α, peut entraîner des conséquences sur d'autres estimateurs de Bayes tels que l'estimateur de Bayes uniforme sur $\Theta(m)$, à travers une analyse plus fine de risque et l'estimation sans biais de risque.

La forme explicite de ces estimateurs est déterminée par le lemme 1.6.1, autrement dit :

$$\delta_\alpha(x) = \frac{1}{r} E_\theta \left[\frac{\theta' X}{\|X\|} \;\middle|\; \|X\| = r \right] x = \frac{\alpha}{r} \rho_{p/2-1}(\alpha r)\, x\,, \qquad (1.6.2)$$

pour tout θ élément de S_α. Tout d'abord, il faut rappeler que les estimateurs δ_α sont admissibles pour θ, sous coût quadratique. Ensuite, il faut mentionner que les propriétés analytiques de la fonction $\rho_{p/2-1}$ citées dans le lemme 1.6.2 jouent un rôle important dans les résultats de dominance. Nous donnons ici d'autres propriétés de la fonction $\rho_{p/2-1}$ développées par [FM10].

Lemme 1.6.15. *(a) Pour tout $p \in \{3, 4, \ldots\}$ et $\alpha \geq 0$, la fonction donnée par $r\,\{1 - \rho_{p/2-1}(\alpha r)\}$ est croissante en r ; $r \geq 0$;*

(b) Pour tout $p \in \{3, 4, \ldots\}$, l'inégalité, $\rho_{p/2-1}(t) + t\,\rho'_{p/2-1}(t) \leq 1$, est satisfaite pour tout $t > 0$;

(c) Pour tout $p \in \{3, 4, \ldots\}$, on a $\lim_{t\to\infty} t\,\{1 - \rho_{p/2-1}(t)\} = \frac{p-1}{2}$.

Or, pour déterminer les conditions de dominance de δ_{emv} par les estimateurs δ_α, il suffit d'étudier le comportement de la différence des risques comme fonction de $\|\theta\| = \lambda$. Cependant, en utilisant l'identité de Stein et l'analyse des estimateurs sans biais de la différence des risques, nous avons le résultat suivant. Nous donnons le résultat avec sa démonstration qui nous servira comme technique pour prouver un résultat analogue au chapitre 4.

Lemme 1.6.16. (Fourdrinier et Marchand, 2010).

(a) Dans le cadre de la perte quadratique, un estimateur sans biais de la différence des risques $R(\theta, \delta_\alpha) - R(\theta, \delta_{emv})$ est donné par :

$$D_\alpha(\|X\|) = D_{\alpha,1}(\|X\|)\,[0 \leq \|X\| \leq m] + D_{\alpha,2}(\|X\|)\,[\|X\| > m]\,,$$

avec :

$$D_{\alpha,1}(r) = 2\alpha^2 + r^2 - 2p - 2\alpha r \rho_{p/2-1}(\alpha r) - \alpha^2 \rho^2_{p/2-1}(\alpha r)\,,$$

et

$$D_{\alpha,2}(r) = 2\alpha^2 - m^2 - \alpha^2 \rho^2_{p/2-1}(\alpha r) + 2mr\{1 - \frac{\alpha}{m}\rho_{p/2-1}(\alpha r)\} - 2(p-1)\frac{m}{r}\,, \tag{1.6.3}$$

où on utilise la notation $[r \in A] = 1_A(r)$, *pour tout ensemble* A.

(b) Pour $p \geq 3$, *et* $0 \leq \alpha \leq m$, $D_\alpha(r)$ *change de signes en fonction de* $r > 0$ *selon l'ordre suivant : (i)* $(-,+)$ *lorsque* $\alpha \leq \sqrt{p}$, *et (ii)* $(+,-,+)$ *lorsque* $\alpha > \sqrt{p}$.

Démonstration. (a) En posant $\delta_{emv}(x) = x + h_{emv}(x)$ avec $h_{emv}(x) = (\frac{m}{r} - 1)x\,[r > m]$, qui est une fonction faiblement différentiable, nous obtenons :

$$2div\,[h_{emv}(x)] + \|h_{emv}(x)\|^2 = \{2(p-1)\frac{m}{r} - 2p + (m-r)^2\}\, 1_{(m\,,\infty)}(r)\,;$$

et, en vertu de l'identité (1.2.5), nous avons $R(\theta, \delta_{emv}) = E_\theta[\eta_{emv}(X)]$ avec :

$$\eta_{emv}(x) = p\,1_{[0,\,m]}(r) \;+\; \{2(p-1)\frac{m}{r} - p + (m-r)^2\}\, 1_{(m,\,\infty)}(r). \tag{1.6.4}$$

D'une manière analogue au développement de [Ber90], les représentations de δ_α et $\frac{d}{dt}\rho_\nu(t)$ données par (1.6.2) et le lemme 1.6.2, de même que l'identité (1.2.5), nous permettent d'écrire que $R(\theta, \delta_\alpha) = E_\theta[\eta_\alpha(X)]$ avec :

$$\eta_\alpha(x) = 2\alpha^2 + r^2 - p - 2\alpha\, r\, \rho_{p/2-1}(\alpha\, r) - \alpha^2 \rho^2_{p/2-1}(\alpha\, r)\,. \tag{1.6.5}$$

Finalement, l'expression de l'estimateur sans biais $D_\alpha(\|X\|)$ se déduit directement à partir de (1.6.4) et (1.6.5).

(b) Nous commençons par énoncer trois propriétés intermédiaires, dont les preuves sont données plus bas.

(I) Les changements de signe de $D_{\alpha,1}(r)$; $r \in [0,m]$; sont ordonnés selon une possibilité parmi les cinq combinaisons suivantes : $(+), (-), (-,+)$. $(+,-), (+,-,+)$;

(II) $\lim_{r\to m^+}\{D_\alpha(r)\} = \lim_{r\to m^-}\{D_\alpha(r)\} + 2$;

(III) quant à la fonction $D_{\alpha,2}(r)$; $r \in (m,\infty)$, soit elle est positive, soit elle change de signe du $-$ au $+$.

En combinant les propriétés (I), (II), and (III), nous en déduisons que les changements de signe de $D_\alpha(r)$, pour $r \in (0,\infty)$, (qui est une fonction continue partout avec un saut au point m) sont ordonnés selon une possibilité parmi les trois combinaisons suivantes : $(+), (-,+), (+,-,+)$. Maintenant, rappelons que δ_α est un estimateur de Bayes et admissible de θ sous coût quadratique. Cependant, parmi les combinaisons citées ci-dessus, $(+)$ est impossible puisque δ_α ne peut être dominé par δ_{emv}. Finalement, nous distinguons les deux autres cas en observant que $D_\alpha(0) = 2\alpha^2 - 2p \leq 0$ si et seulement si $\alpha \leq \sqrt{p}$.

Preuve de (I). Nous calculons la dérivée de $D_{\alpha,1}$, ensuite en se servant de la partie (a) du lemme 1.6.2 nous obtenons :

$$r^{-1}\, D'_{\alpha,1}(r) = 2 - 2\frac{\alpha}{r}\rho_{p/2-1}(\alpha r) - 2\alpha^2\rho'_{p/2-1}(\alpha r) - 2\alpha^3\rho'_{p/2-1}(\alpha r)\frac{\rho_{p/2-1}(\alpha r)}{r}\,.$$

Or, les quantités $r^{-1}\rho_{p/2-1}(\alpha r)$ et $\rho'_{p/2-1}(\alpha r)$ sont positives et décroissantes par rapport à r selon le lemme 1.6.2, ce qui implique que la fonction $r^{-1}\, D'_{\alpha,1}(r)$ est nécessairement croissante par rapport à r, $r \in [0,m]$. Cependant $D'_{\alpha,1}(\cdot)$ possède, sur $[0,m]$, des changements de signe ordonnés selon les combinaisons : $(+), (-)$, ou $(-,+)$. Finalement, nous constatons que $D_{\alpha,1}(\cdot)$ possède au plus deux changements de signe sur $[0,m]$ et, de plus, la combinaison $(-,+,-)$ n'est pas cohérente avec les changements de signe de $D'_{\alpha,1}$.

Preuve de (II). Une évaluation directe de $D_{\alpha,1}(m)$ et $D_{\alpha,2}(m)$, dont les expressions sont données par la partie (a) de ce lemme, donne le résultat.

Preuve de (III) Premièrement, nous vérifions à partir de (1.6.3), de la partie (a) du lemme 1.6.2, et de la partie (c) du lemme 1.6.15 que $\lim_{r\to\infty} D_{\alpha,2}(r)$ est $+\infty$, pour $\alpha < m$; et égale à $p-1$ si $\alpha = m$. De plus, la partie (a) nous permet d'écrire $D_{\alpha,2}(r)$; $r > m$; comme $(1 - \frac{\alpha}{m}\ \rho_{p/2-1}(\alpha r))\ \sum_{i=1}^{3} H_i(\alpha, r)$ avec :

$$
\begin{aligned}
H_1(\alpha, r) &= 2rm\{1 - \frac{(1-\frac{\alpha^2}{m^2})m}{r\{\ 1 - \frac{\alpha}{m}\rho_{p/2-1}(\alpha r)\ \}}\}, \\
H_2(\alpha, r) &= \frac{-2(p-1)\,m}{r\ \{\ 1 - \frac{\alpha}{m}\rho_{p/2-1}(\alpha r)\ \}}
\end{aligned}
$$

et

$$
H_3(\alpha, r) = m^2 + \alpha\, m\, \rho_{p/2-1}(\alpha r).
$$

Cependant, pour établir la propriété (III), il suffit de montrer que chacune des fonctions $H_i(\alpha, \cdot)$, $i = 1, 2, 3$, est croissante sur $(m\,, \infty)$ sous les conditions imposées sur (p, α, m). La fonction $H_3(\alpha, \cdot)$ est clairement croissante en vertu du lemme 1.6.2, et de même pour $H_2(\alpha, \cdot)$ en vertu des lemmes 1.6.2 et 1.6.15 puisque :

$$
r(1 - \frac{\alpha}{m}\rho_{p/2-1}(\alpha r)) = r(1 - \rho_{p/2-1}(\alpha r)) + r\ (1 - \frac{\alpha}{m})\ \rho_{p/2-1}(\alpha r).
$$

Finalement, pour l'analyse de $H_1(\alpha, r), r > m$, nous procédons par dérivation et réarrangement des termes pour obtenir :

$$
\frac{\partial}{\partial r} H_1(\alpha, r) \geq 0 \Leftrightarrow T(m) \geq 0
$$

où, pour $r > m \geq \alpha$,

$$
T(m) = (m - \alpha\, \rho_{p/2-1}(\alpha\, r))^2 - \alpha^2 (m^2 - \alpha^2)\, \rho'_{p/2-1}(\alpha\, r)\,.
$$

Or notons que :

$$T(\alpha) = \alpha^2 \, (1 - \rho_{p/2-1}(\alpha r))^2 \geq 0$$

et

$$\begin{aligned}
\frac{1}{2}\,\frac{\partial T(m)}{\partial m} &= (m - \alpha\, \rho_{p/2-1}(\alpha\, r)) - m\alpha^2\, \rho'_{p/2-1}(\alpha\, r) \\
&\geq (\alpha - \alpha\rho_{p/2-1}(\alpha r)) - m\alpha^2\, \frac{(1 - \rho_{p/2-1}(\alpha r))}{\alpha r} \\
&= \alpha\,(1 - \rho_{p/2-1}(\alpha r))\,(1 - \frac{m}{r}) \\
&\geq 0\,,
\end{aligned}$$

selon la partie (b) du lemme 1.6.15, puisque $r \geq m \geq \alpha$. Ceci prouve que $T(m) \geq T(\alpha) \geq 0$ pour tout $m \geq \alpha$, et que $H_1(\alpha, r)$ est croissante en r. $\square$

Maintenant, sachant que les estimateurs δ_{emv} et δ_α sont équivariants par rapport aux transformations orthogonales et que leurs risques ne dépendent de θ qu'à travers $\|\theta\|$, nous notons la différence de risques comme suit :

$$\Delta_\alpha(\lambda) \;=\; R(\theta, \delta_\alpha) - R(\theta, \delta_{emv})\,;\; \lambda = \|\theta\|\,.$$

D'autre part, puisque la distribution de probabilité de la variable $\|X\|^2$ est $\chi^2_p(\lambda^2)$, les changements de signe potentiels de $\Delta_\alpha(\lambda) \;=\; E_\lambda[D_\alpha(\|X\|)]$ sont contrôlés par les propriétés de variation de $D_\alpha(\cdot)$ en terme de changement de signe (les arguments du changement de signe de Karlin, voir [BJM81]). Cependant les changements de signe possibles de $\Delta_\alpha(\lambda)$, en fonction de $\lambda \in [0, m]$, nous permettent de déduire des conditions nécessaires et suffisantes pour que δ_α domine δ_{emv} sur $\Theta(m)$. Ces conditions sont données par le corollaire suivant.

Corollaire 1.6.17. (Fourdrinier et Marchand, 2010). *Pour tout* $p \geq 3$ *et* $0 \leq \alpha \leq m$, *l'estimateur* δ_α *domine* δ_{emv} *si et seulement si :*

(i) $\Delta_\alpha(m) \leq 0$ *lorsque* $\alpha \leq \sqrt{p}$ *; ou*

(ii) $\Delta_\alpha(0) \leq 0$ *et* $\Delta_\alpha(m) \leq 0$, *lorsque* $\alpha > \sqrt{p}$.

Nous observons à travers ce corollaire que les conditions nécessaires et suffisantes pour que δ_α domine δ_{emv} s'expriment seulement en fonction des risques en $\theta = 0$ (c'est-à-dire au centre de l'espace paramétrique $\Theta(m)$) et en $\theta \in S_m$, la frontière de $\Theta(m)$. L'étape suivante consiste à traduire ces conditions en fonction de (α, m, p), ce qui nous permettra d'avoir des résultats pour tout p, soit, en conséquence, pour de grandes valeurs de p. Pour cela, nous avons besoin du lemme suivant qui donne les représentations de la fonction risque pour $\theta = 0$ et $\theta \in S_m$.

Lemme 1.6.18. (Marchand et Perron, 2005). *Pour tout estimateur équivariant* δ *de la forme* $\delta(X) = g(\|X\|)X$ *pour une certaine fonction positive* g, *on a :*

$$R(0, \delta) = E_0[\,\|\delta(X)\|^2\,]\,, \; \textit{pour } \theta = 0$$

et, pour $\theta \in S_m$,

$$R(\theta, \delta) = R(\theta, \delta_m) + E_\theta[\{\|\,\delta(X)\| \; - \; \|\,\delta_m(X)\|)\}^2]\,.$$

Les représentations données par le lemme 1.6.18 démontrent que plus $\delta(X)$ est rétréci vers 0, plus son risque est petit en 0 et, par conséquent, le risque minimal est atteint par $\delta \equiv 0$. D'autre part, sachant que tout estimateur de Bayes propre δ_π associé à une loi a priori définie sur $\Theta(m)$

retrécit par rapport à δ_m (voir lemme 1.6.14), c'est-à-dire que $\|\delta_\pi\| \leq \|\delta_m\|$, le risque minimal au bord S_m est atteint par $\delta_\pi \equiv \delta_m$.

Dans le cas des estimateurs δ_α, les cas extrêmes correspondent à $\alpha = 0$ et $\alpha = m$. De plus, puisque $\|\delta_\alpha(x)\| = \alpha\rho_{p/2-1}(\alpha\|x\|)$, nous pouvons conclure que l'amplitude du rétrécissement par rapport à δ_m est contrôlée directement par α, et que cette amplitude est strictement décroissante lorsque α croît (voir lemme 1.6.2). Ainsi, nous avons le résultat suivant obtenu à partir de cette dernière analyse et du corollaire 1.6.17.

Corollaire 1.6.19. (Fourdrinier et Marchand, 2010). *Pour $p \geq 3$ et $0 \leq \alpha \leq m$, l'estimateur δ_α domine δ_{emv} si et seulement si :*

(i) $\alpha \geq k_1(p,m)$ lorsque $\alpha \leq \sqrt{p}$; ou

(ii) $\alpha \geq k_1(p,m)$ et $\alpha \leq k_2(p,m)$ lorsque $\alpha > \sqrt{p}$;

où $k_1(p,m)$ et $k_2(p,m)$ sont respectivement les solutions uniques en α des équations $\Delta_\alpha(m) = 0$ et $\Delta_\alpha(0) = 0$.

Nous voyons clairement que le corollaire 1.6.17 et le corollaire 1.6.19 s'appliquent au cas particulier de l'estimateur de Bayes uniforme sur le bord ; ils impliquent que δ_m domine δ_{emv} lorsque $m \leq \sqrt{p}$ (et $p \geq 3$), comme il a été établi par Marchand et Perron [MP01] (mais pour tout $p \geq 1$).

Ces résultats de dominance permettent de pousser davantage les bornes des domaines de dominance de δ_m et de δ_α pour $\alpha < m$. Ceci sera réalisé par le biais d'une analyse asymptotique du comportement des quantités $k_i(p,m); i = 1, 2$. Pour cela, nous avons besoin du lemme suivant donné par Marchand et Perron [MP02].

Lemme 1.6.20. *Soit* $R_p^2 \sim \chi_p^2(d^2p)$ *pour* $d \geq 0$ *et* $p = 1, 2, \ldots$ *Alors, pour tout* $c > 0$*, la suite de variables aléatoires*

$$\rho_{p/2-1}^2(c\sqrt{p}\,R_p) \xrightarrow{P} L\left(\frac{1}{c\sqrt{1+d^2}}, \frac{1}{c\sqrt{1+d^2}}\right), \text{ lorsque } p \to \infty\,,$$

où $L(a,b) = \{a/2 + \sqrt{1+(b/2)^2}\}^{-2}$*. Dans les cas particuliers* $d = 0$ *et* $d = c$*, la suite converge respectivement vers* $\left(\frac{2c}{1+\sqrt{4c^2+1}}\right)^2$ *et* $\frac{d^2}{1+d^2}$*.*

Démonstration. Nous commençons par montrer que R_p^2/p converge en probabilité, lorsque $p \to \infty$, vers $1 + d^2$. Pour cela, nous écrivons $R_p^2 \stackrel{\mathcal{L}}{=} \sum_{i=1}^p R_i$, où les R_i sont des variables indépendantes et identiquement distribuées $\chi_1^2(d^2)$. Enfin, en utilisant la loi faible des grands nombres nous avons le résultat de convergence.

Maintenant, selon (1.6.1), nous avons pour tout $p \geq 3$:

$$L\left(\frac{\sqrt{p}}{cR_p}, \frac{\sqrt{p}}{cR_p}\right) \leq \rho_{p/2-1}^2(c\sqrt{p}\,R_p) \leq L\left(\frac{p-2}{p}\frac{\sqrt{p}}{cR_p}, \frac{p+2}{p}\frac{\sqrt{p}}{cR_p}\right)$$

avec probabilité 1. Le résultat est obtenu en utilisant la convergence de $\frac{R_p^2}{p}$ vers $1+d^2$, ce qui implique la convergence en probabilité des bornes supérieure et inférieure de l'inégalité précédente vers $L\left(\frac{1}{c\sqrt{1+d^2}}, \frac{1}{c\sqrt{1+d^2}}\right)$. Finalement, les résultats donnés dans les cas $d = 0$ et $d = c$ sont obtenus en évaluant $L\left(\frac{1}{c}, \frac{1}{c}\right)$ et $L\left(\frac{1}{d\sqrt{1+d^2}}, \frac{1}{d\sqrt{1+d^2}}\right)$. □

Une illustration de ce résultat pour différentes valeurs de p (voir Figure 1.1), montre la rapidité de la convergence en probabilité de la suite $\rho_{p/2-1}^2(c\sqrt{p}\,R_p)$. Ainsi, nous obtiendrons de bonnes approximations (même avec p modéré) en utilisant ce résultat, puisque, en réalité, c'est la moyenne de la distribution limite que nous utiliserons dans nos calculs. Maintenant, en combinant le lemme 1.6.18 et le lemme 1.6.20, nous avons de nouveaux résultats présentés dans le théorème suivant.

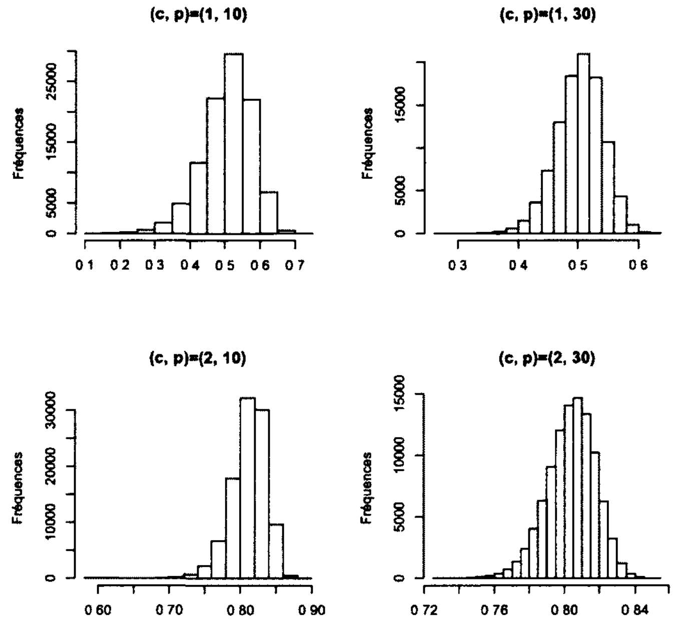

FIGURE 1.1 – Distribution limite pour différentes valeurs de c et de p de la fonction $\rho^2_{p/2-1}(c\sqrt{p}\,R_p)$.

Théorème 1.6.21. (Fourdrinier et Marchand, 2010). *Soient* $k_1(p,m)$ *et* $k_2(p,m)$ *les points donnés dans le corollaire 1.6.19 et soit* $\gamma(d) = 3d^2 + \frac{2d^4}{1+d^2} - \frac{d(1+4d^2)}{\sqrt{1+d^2}}$, *pour* $d > 0$, *on a alors*

(a) *Pour tout* $d \geq 1$,

$$\lim_{p\to\infty} \frac{k_2(p, d\sqrt{p})}{\sqrt{p}} = \sqrt{2}.$$

(b) *Pour tout* $d > 0$,

$$\lim_{p\to\infty} \frac{k_1(p, d\sqrt{p})}{\sqrt{p}} = \sqrt{0 \vee \gamma(d)}.$$

(c) *En particulier,* $\lim_{p\to\infty} \frac{k_1(p,d\sqrt{p})}{\sqrt{p}} = 0$, *pour tout* $d \leq \sqrt{\frac{1}{3}}$.

Pour $m \geq \sqrt{p}$, les résultats précédents montrent que l'estimateur de Bayes δ_m domine δ_{emv} si et seulement si $m \leq k_2(p,m)$, avec $k_2(p,m) \approx \sqrt{2p}$ pour des grandes valeurs de p, indépendamment de m. En comparaison avec la condition suffisante de dominance, $m \leq \sqrt{p}$, donné par Marchand et Perron [MP01], la nouvelle condition est nécessaire et suffisante, et elle est asymptotiquement précise.

D'autre part, il faut noter que l'estimateur de maximum de vraisemblance est manifestement inefficace dans le cas d'espaces paramétriques petits, puisque cet estimateur est dominé par tout estimateur de Bayes associé à une loi a priori orthogonalement invariante. Pour la sous-classe des estimateurs de Bayes δ_α, $\alpha \in [0,m]$, la partie (c) du théorème 1.6.21, nous permet de conclure que tous ces estimateurs δ_α doivent nécessairement dominer δ_{emv} pour les grandes valeurs de p lorsque $\frac{m}{\sqrt{p}} \leq \frac{1}{3}$. Ce résultat s'applique particulièrement sur l'estimateur $\delta_0 \equiv 0$, en utilisant la condition suffisante précédente, et qui devient aussi nécessaire en vertu de la partie (b) du théorème 1.6.21, comme $\gamma(d) > 0$ pour $d > \sqrt{\frac{1}{3}}$.

Pour illustrer davantage ces résultats théoriques, nous considérons le cas $p = 9, m = 4.5$, ce qui donne $d = \frac{m}{\sqrt{p}} = \frac{4.5}{\sqrt{9}} = 1.5$, et $\gamma(1.5) = 3(3/2)^2 + \frac{2(3/2)^4}{1+(3/2)^2} - \frac{(3/2)(1+4(3/2)^2)}{\sqrt{1+(3/2)^2}} \approx 1.5449$. Ce qui implique que

$$k_1(9,4.5) \approx \sqrt{9\,\gamma(1.5)} = 3.7288...,$$

d'après le théorème 1.6.21. De plus, le même théorème implique que

$$k_2(9,4.5) \approx 3\sqrt{2} = 4.2426...$$

Ainsi, le corollaire 1.6.19 nous permet de déduire, pour cet exemple, que l'estimateur δ_α domine δ_{emv} si et seulement si $\alpha \in [k_1(9,4.5), k_2(9,4.5)] \approx$

$[3.7288, 4.2426]$. Cependant, il faut remarquer que les petites et les grandes valeurs de α ne sont pas dans le domaine de dominance, parce qu'ils n'améliorent pas le risque de δ_{emv} sur la frontière S_m et en $\theta = 0$, respectivement. Notons aussi que l'évaluation numérique des valeurs de $k_1(p,m)$ et de $k_2(p,m)$ est possible pour tout couple (p,m). Donc, pour pouvoir comparer, nous avons obtenu $k_1(9, 4.5) \approx 3.835$ et $k_2(9, 4.5) \approx 4.223$, c'est-à-dire que les approximations données pour la taille de l'échantillon $p = 9$ sont proches des valeurs réelles. En rappelant que ces approximations découlent du lemme 1.6.20, nous pouvons remarquer la concordance avec notre illustration (Figure 1.1). Ceci prouve aussi que les approximations utilisés sont aussi bonnes, même pour les petites valeurs de p.

Un autre point important à exposer, comme il a été indiqué par les résultats analytiques et numériques de Marchand et Perron (2001), c'est qu'il devient difficile de déterminer des compétiteurs de δ_{emv} lorsque l'espace paramétrique est assez grand. Pour cela, nous considérons encore une fois le cas $p = 9$ dans le but d'étudier l'impact des grandes valeurs de m sur la condition nécessaire et suffisante pour que δ_α domine δ_{emv}. D'un côté, les conditions de dominance exigent que $\alpha \leq k_2(9,m)$, avec $k_2(9,m) \approx 3\sqrt{2}$ selon le théorème 1.6.21. Et de l'autre, il faut que $\alpha \geq k_1(9,m)$, avec $k_1(p,m) \approx 3\sqrt{\gamma(d)}$, et $d = m/3$. Maintenant, comme il est clair que $\gamma(d)$ est une fonction croissante en d, alors l'ensemble des valeurs de α menant à la dominance est non vide si nous avons approximativement $\gamma(d) \leq 2 \Leftrightarrow d \leq d_0 = \gamma^{-1}(2) \approx 1.656$. Cependant, pour des espaces paramétriques assez grand avec $m > d_0\sqrt{p}$, où la contrainte devient moins informative, δ_{emv} est plutôt efficace dans le sens où des compétiteurs dans la classe des estimateurs δ_α sont introuvables.

1.6.3 Minimaxité

Il est question, dans cette section, de la recherche d'estimateurs du minimax du paramètre θ dans notre contexte. Les résultats de DasGupta [Das85] impliquent que l'estimateur de Bayes δ_{BU} est du minimax pour des valeurs assez petites de m, c'est-à-dire pour $m \leq m_0(p)$. Cependant le problème consiste à vérifier si la condition $m \leq \sqrt{p}$ est suffisante pour la minimaxité de δ_{BU} et, par conséquent, à voir si $m_0(p) \geq \sqrt{p}$.

Pour $p = 1$, Casella et Strawderman [CS81] ont rapporté que $m_0(1) \approx 1.05674$. De même, Berry [Ber90] a évalué $m_0(2) \approx 1.09318\sqrt{2}$ et $m_0(3) \approx 1.1015783\sqrt{3}$, ce qui implique que la condition $m \leq \sqrt{p}$ est suffisante pour la minimaxité de δ_{BU} dans les cas $p = 1, 2$ et 3. Ainsi, le développement suivant consiste à montrer que l'estimateur δ_{BU} est du minimax si $m \leq \sqrt{p}$, pour $p \geq 4$.

Pour cela, nous commençons par rappeler que Berry [Ber90] a montré que $\sup_{\theta \in \Theta(m)} \{R(\theta, \delta_{BU})\}$ est atteint, soit à l'origine ($\theta = 0$), soit sur la frontière ($\|\theta\| = m$). D'où, le résultat suivant.

Lemme 1.6.22. (Berry, 1990). *Une condition nécessaire et suffisante pour que δ_{BU} soit du minimax est*

$$R(0, \delta_{BU}) \leq R(\theta, \delta_{BU}) \ \ \textit{avec} \ \ \|\theta\| = m\,. \tag{1.6.6}$$

Maintenant, sachant que $\delta_{BU}(x) = \frac{g_m(r)}{r}x$. où $g_m(r)$ est donnée par le lemme 1.6.1, nous pouvons écrire le risque comme suit :

$$\begin{aligned}
R(\theta, \delta_{BU}) &= E_\theta\left[\|\delta_{BU}(X) - \theta\|^2\right] \\
&= \theta'\theta + E_\theta\left[g_m^2(\|X\|)\right] - 2E_\theta\left[\theta' X \frac{g_m(\|X\|)}{\|X\|}\right] \\
&= \theta'\theta + E_\lambda\left[g_m^2(R)\right] - 2E_\lambda\left[g_m(R)\, g_\lambda(R)\right]\,,
\end{aligned}$$

où $\lambda = \|\theta\|$. Ensuite, vu les propriétés de la fonction $\rho_{(p/2)-1}$ données par le lemme 1.6.2 et vu la propriété de rapport de vraisemblance monotone de la densité de R^2 par rapport au paramètre m, nous déduisons que l'inégalité (1.6.6) est satisfaite si et seulement si $m \leq m_0(p)$ pour une certaine valeur de $m_0(p)$. Finalement, en utilisant les bornes de ρ_ν^2, pour $\nu > 0$, données par (1.6.1), Marchand et Perron [MP02] ont démontré le théorème suivant.

Théorème 1.6.23. (Marchand et Perron, 2002). *$m_0(p) \geq \sqrt{p}$, pour tout $p \geq 1$.*

Néanmoins il est important d'étudier le comportement asymptotique de $m_0(p)$, afin de déterminer une forme explicite de $m_0(p)$ en fonction de $\sqrt{p}$, au moins pour les grandes valeurs de p. Pour cela, en utilisant la paramétrisation $m = d\sqrt{p}$ et $\lambda = c\sqrt{p}$ avec $0 \leq c \leq d$ et en utilisant le lemme 1.6.20, nous avons le résultat suivant.

Théorème 1.6.24. (Marchand et Perron, 2002). *On a,*

$$\lim_{p\to\infty} m_0(p)/\sqrt{p} = \kappa \approx 1.150963925\,,$$

où κ^2 est la racine réelle du polynôme $s^3 - s - 1$.

Des évaluations numériques de $m_0(p)$ sont données pour $4 \leq p \leq 50$ par Marchand et Perron [MP02]. De plus, le risque minimax, dans le cas où δ_{BU} est du minimax, est donné par

$$R(m,p) = m^2 - E_{m^2}\left[m^2\rho^2_{(p/2)-1}(mR)\right]\,.$$

Ensuite, en utilisant les bornes de la fonction $\rho^2_{(p/2)-1}$ données par (1.6.1) et en posant $m = d\sqrt{p}$ et $Y = \sqrt{\frac{R^2}{p}}$, nous obtenons l'inégalité suivante :

$$1 - L\left(\frac{1-p/2}{d\sqrt{1+d^2}}, \frac{1+p/2}{d\sqrt{1+d^2}}\right) \leq \frac{R(m,p)}{m^2} \leq 1 - E_{m^2}\left[L\left(\frac{1}{dY}, \frac{1}{dY}\right)\right]\,,$$

(1.6.7)

où $L(a,b) = \{a/2 + \sqrt{1+(b/2)^2}\}^{-2}$. Cette inégalité nous permet d'avoir des bornes pour le risque minimax dès que $m \leq m_0\,(p)$. D'autre part, en utilisant encore une fois le lemme 1.6.20, nous obtenons la valeur asymptotique du risque minimax comme suit :

$$\lim_{p\to\infty} \frac{R\,(m,p)}{d^2 p} = \frac{1}{d^2+1}\,. \tag{1.6.8}$$

Enfin, rappelons que Marchand [Mar93] a établi une borne inférieure pour le risque minimax donnée par :

$$R\,(m,p) \geq (p-1)\, m^2 / \left(p + m^2\right)\ ,$$

ce qui représente une alternative à celle donnée dans (1.6.7). Pour $m = d\sqrt{p}$, cette deuxième borne inférieure devient $\frac{p-1}{p(1+d^2)}$ et donne asymptotiquement la même limite que (1.6.8).

1.7 La classe des estimateurs linéaires tronqués

Nous considérons ici le même problème d'estimation du paramètre θ, sous coût quadratique, à partir d'une observation de $X \sim \mathcal{N}_p(\theta, I_p)$ $(p \geq 1)$, avec la contrainte $||\theta|| \leq m$, pour une certaine constante $m > 0$. Ce problème d'inférence avec informations additionnelles données par la contrainte est souvent rencontré dans les cas pratiques. Il est alors pertinent d'analyser le gain en performance de diverses classes d'estimateurs, puisque l'efficacité dépend de l'interprétation des informations paramétriques par chaque estimateur.

Une suite naturelle aux travaux de Marchand et Perron [MP01] et Fourdrinier et Marchand [FM10], consiste à étudier d'autres classes d'estimateurs dominant δ_{emv} et à évaluer leur performance. D'autre part, plusieurs résultats de dominance, entre autres les résultats bayésiens, donnés par Marchand et Perron [MP01] et Fourdrinier et Marchand [FM10], nécessitent une condition de la forme $m \leq c(p)$, alors que nous n'avons pas l'affirmation de l'existence de tels estimateurs (bayésiens) pour (m,p) arbitraire.

Un autre problème non résolu, fondamental et difficile consiste en l'évaluation du degré d'amélioration obtenu par rapport à l'estimateur δ_{emv}. Des résultats numériques donnés par [MP01] pour différentes valeurs de (m,p) exposent des gains substantiels réalisés par une grande classe d'estimateurs incluant les estimateurs de Bayes δ_U et δ_{BU}, sur la majeure partie de l'espace paramétrique associé aux valeurs de m petites par rapport à p. D'autre part, ces gains sont beaucoup moins impressionnants ou disparaissent pour les valeurs moyennes ou grandes de m par rapport à p. De même, le résultat de dominance universel donné par [FM10] suggère qu'il y a des gains considérables pour les valeurs de m telles que $m \leq \sqrt{p/3}$. Maintenant, malgré les preuves indirectes données ci-dessus, il y a un manque de résultats analytiques produisant des bornes ou de bonnes approximations de pourcentage du degré d'amélioration.

Pour ces raisons, nous proposons l'étude de la classe des estimateurs linéaires tronqués $\delta_a(x) = (a\|x\| \wedge m)\frac{x}{\|x\|}$, avec $0 \leq a \leq 1$. Cette classe d'estimateurs à rétrécissement (par rapport à δ_{emv}) représente un certain intérêt, puisqu'elle contient l'estimateur du maximum de vraisemblance δ_1, l'estimateur trivial (mais de Bayes) δ_0, et l'estimateur linéaire minimax tronqué δ_{tlx} avec $a = \frac{m^2}{m^2+p}$. Notons aussi que, pour $a \in [0,1)$, δ_a est la troncature de l'estimateur admissible aX sur l'espace paramétrique $\Theta(m) = \{\theta : \|\theta\| \leq m\}$, alors que δ_{emv} est la troncature de l'estimateur inadmissible, pour $p \geq 3$, X sur $\Theta(m)$. [Mar93] a aussi donné des preuves analytiques limitant la perte d'efficacité encourue lorsque nous utilisons l'estimateur linéaire minimax, et donc δ_{tlx}, à la frontière de $\Theta(m)$ au lieu d'utiliser le meilleur estimateur équivariant (pour $\|\theta\| \leq m$) δ_{BU}. Pour les valeurs assez petites de m, tous les estimateurs δ_a ($a \in [0,1)$) dominent δ_{emv}, d'après le résultat universel de dominance. D'autre part, une analyse fine pour $p = 1$ a été établie par [MOPP08] pour δ_a et δ_{tlx}, et par [DvE09] pour δ_0.

Notre étude consiste à établir une analyse plus fine pour toute dimension p finie et à décrire avec précision, dans un contexte asymptotique, la perte ou le gain d'efficacité en ce qui concerne les risques de δ_a et de δ_{emv}. Ainsi, nous considérons une suite de points $\left\{(m,p) = \left(\frac{d}{\sqrt{p}}, p\right), p \geq 1\right\}$, pour d fixe, et nous étudions les conditions de dominance de δ_{emv} par les estimateurs δ_a et en particulier par δ_{tlx} pour p assez grand.

Dans le cas unidimensionnel, nous donnons un résultat (lemme 1.7.2), que nous utiliserons dans nos preuves pour le calcul de risque de l'estimateur linéaire tronqué δ_a lorsque $\|\theta\| = m$. Les fonctions de répartition et de densité da la loi normale centrée réduite sont désignées respectivement par $\Phi(\cdot)$ et $\phi(\cdot)$.

Lemme 1.7.1. *La fonction génératrice des moments de la loi $\mathcal{N}(m,1)$ tronquée sur $]-c,c[$ est donnée par :*

$$\varphi(t) = e^{t^2/2+mt}\frac{\Phi(t+m+c) - \Phi(t+m-c)}{\Phi(m+c) - \Phi(m-c)}, t \in \mathbb{R}.$$

Démonstration. Soit X de loi $\mathcal{N}(m,1)$ tronquée sur $]-c,c[$. Nous avons :

$$\begin{aligned}
\varphi(t) = E\left(e^{tX}\right) &= \frac{\int_{-c}^{c} e^{tx} e^{-\frac{1}{2}(x-m)^2}\,dx}{\int_{-c}^{c} e^{-\frac{1}{2}(x-m)^2}\,dx} \\
&= \frac{e^{t^2/2+mt}\int_{-c}^{c}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}(x-(m+t))^2}\,dx}{\Phi(c-m)-\Phi(-c-m)} \\
&= e^{t^2/2+mt}\frac{\Phi(c-m-t)-\Phi(-c-m-t)}{\Phi(c-m)-\Phi(-c-m)} \\
&= e^{t^2/2+mt}\frac{\Phi(t+m+c)-\Phi(t+m-c)}{\Phi(m+c)-\Phi(m-c)}\,.
\end{aligned}$$

□

Lemme 1.7.2. *Pour la loi $\mathcal{N}(m,1)$ tronquée sur $]-c,c[$, on a :*

$$\int_{-c}^{c} x\;\phi(x-m)\;dx = m\left(\Phi(m+c)-\Phi(m-c)\right) + \phi(m+c)-\phi(m-c)\ , \quad (1.7.1)$$

et

$$\int_{-c}^{c} x^2\;\phi(x-m)\;dx = \left(1+m^2\right)\left(\Phi(m+c)-\Phi(m-c)\right) + (m-c)\,\phi(m+c)-(m+c)\,\phi(m-c)\ . \quad (1.7.2)$$

Démonstration. Du lemme 1.7.1, nous déduisons que :

$$\varphi'(t) = (t+m)\,\varphi(t) + e^{t^2/2+mt}\,\frac{\phi(t+m+c)-\phi(t+m-c)}{\Phi(m+c)-\Phi(m-c)}\,,$$

pour tout $t\in\mathbb{R}$. Ainsi, pour X de loi $\mathcal{N}(m,1)$ tronquée sur $]-c.c[$, nous obtenons le premier moment :

$$E[X] = \varphi'(0) = m + \frac{\phi(m+c)-\phi(m-c)}{\Phi(m+c)-\Phi(m-c)}\,,$$

ce qui induit l'intégrale (1.7.1).

De la même manière, nous trouvons :

$$\varphi''(t) = \left(1 + (t+m)^2\right)\varphi(t) + e^{t^2/2+mt} \times \frac{(t+m-c)\,\phi(t+m+c) - (t+m+c)\,\phi(t+m-c)}{\Phi(m+c) - \Phi(m-c)},$$

pour tout $t \in \mathbb{R}$, puisque $\phi'(z) = -z\phi(z)$. Par conséquent, nous obtenons le second moment :

$$E\left[X^2\right] = 1 + m^2 + \frac{(m-c)\,\phi(m+c) - (m+c)\,\phi(m-c)}{\Phi(m+c) - \Phi(m-c)},$$

ce qui induit l'intégrale (1.7.2). □

1.8 Estimation avec coefficient de variation borné

Ce problème se veut une généralisation dans le cas des distributions multidimensionnelles de dimension $p \geq 2$ de l'estimation de la moyenne μ d'une $\mathcal{N}(\mu, \sigma^2)$ quand le coefficient de variation σ/μ est connu. Le problème en dimension 1 a été abordé par Fisher. Il a été repris par la suite par plusieurs statisticiens dont Efron [EM76], Hinkley [Hin77], Amari [Ama82] et Kariya [Kar89]. L'hypothèse de base dans ce modèle repose sur le fait qu'en pratique, dans certaines situations, l'écart type a tendance à prendre de grandes valeurs de façon directement proportionnelle avec la valeur prise par la valeur de la moyenne.

Dans le cas de l'estimation de la moyenne où le paramètre d'échelle est connu, beaucoup de résultats ont été établis pour $p = 1$ et pour $p > 1$. Nous avons cité dans les sections précédentes un bon nombre de ces résultats. Par contre, si le paramètre d'échelle est inconnu, des travaux basés sur une approche d'invariance ont été développés par Kariya et al. [KGP88], Perron [Per87] et Marchand [Mar94] lorsque le coefficient $\|\mu\| / \sigma$ est connu.

Le modèle de [KGP88] est défini de la manière suivante. Soient $Y_1, Y_2, ..., Y_n$ des vecteurs aléatoires indépendants de même loi de probabilité $\mathcal{N}_p(\mu, \sigma^2 I_p)$, où les paramètres $\mu \in \mathbb{R}^p$ et $\sigma > 0$ sont supposés inconnus et $n > p$. Considérons

$$X = \sqrt{n}\,\overline{Y} \quad \text{et} \quad S^2 = tr\left(\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)\left(Y_i - \overline{Y}\right)'\right).$$

La statistique (X, S^2) forme une statistique exhaustive pour ce problème. De plus, le vecteur aléatoire X est indépendant de la variable aléatoire S^2, X est $\mathcal{N}(\sqrt{n}\mu, \sigma^2 I_p)$ et S^2 est $\sigma^2 \chi^2_{(n-1)p}(0)$.

Le problème consiste à estimer le vecteur $\theta = \sqrt{n}\mu$ dans $\mathbb{R}^p$ sous la fonction de perte $L(\theta, d) = \|\theta - d\|^2 / \sigma^2$. Ce problème reste invariant par rapport au groupe $G = \mathbb{R}_+ \times O_p(\mathbb{R})$, où $O_p(\mathbb{R})$ est le sous-groupe multiplicatif des matrices orthogonales d'ordre p. La transformation induite par G sur la statistique exhaustive (X, S^2), pour un élément $g = (b, A) \in G$, est donnée par

$$(X, S^2) \to (bAX, b^2 S^2).$$

La forme générale des estimateurs équivariants a été donnée par les auteurs cités ci-dessus. Le résultat est donné par la proposition suivante.

Proposition 1.8.1. ((Kariya, Giri et Perron, 1988). *Un estimateur δ, basé sur (X, S^2), est équivariant si et seulement s'il existe une fonction mesurable $h(\cdot)$ telle que $\delta(x, s^2) = h\left(\frac{\|x\|^2}{s^2}\right) x$ pour tout $(x, s^2) \in \mathbb{R}^p \times \mathbb{R}_+$.*

Dans notre cas, nous considérons le problème d'estimation de la moyenne d'une distribution p−variée $\mathcal{N}_p(\theta, \sigma^2 I_p)$, avec $p \geq 1$, où les deux paramètres

θ et σ sont inconnus et assujettis à la condition $\frac{\|\theta\|}{\sigma} \leq m$ pour une certaine constante $m > 0$. Nous considérons la construction d'estimateurs du paramètre θ, sous la fonction de perte $\|\theta - d\|^2 / \sigma^2$, en se basant sur les réalisations de la statistique exhaustive (X, S^2).

Kubokawa [Kub05] a abordé ce problème pour $p = 1$, comme étant une extension des travaux de Marchand et Strawderman [MS05a] avec un paramètre d'échelle inconnu. Il a développé une classe d'estimateurs qui dominent l'estimateur sans biais X. Cette classe d'estimateurs contient, entre autres, les estimateurs de Bayes δ_{BU} et δ_U, par rapport aux lois a priori uniformes respectivement sur $\{-m, m\}$ et sur $[-m, m]$, et aussi des estimateurs tronqués.

Par conséquent, notre objectif sera de construire des estimateurs qui dominent X, ainsi que l'estimateur du maximum de vraisemblance δ_{emv}. Ce qui représentera une extension des travaux de Kubokawa[Kub05] et des travaux de Marchand et Perron [MP01] pour σ connu.

Chapitre 2

Estimation dans le cas non gaussien

Bayes minimax estimators of the mean of a scale mixture of multivariate normal distributions

Dominique Fourdrinier [1], Othmane Kortbi [2] and William E. Strawderman [3]

Abstract

Bayes estimation of the mean of a variance mixture of multivariate normal distributions is considered under sum of squared errors loss. We find broad class of priors (also in the variance mixture of normal class) which result in proper and generalized Bayes minimax estimators. This paper extends the results of Strawderman [Minimax estimation of location parameters for certain spherically symmetric distribution, J. Multivariate Anal. 4 (1974) 255-264] in a manner similar to that of Maruyama [Admissible minimax estimators of a mean vector of scale mixtures of multivariate normal distribution, J. Multivariate Anal. 21 (2003) 69-78] but somewhat more in the spirit of Fourdrinier et al. [On the construction of bayes minimax estimators, Ann. Statist. 26 (1998) 660-671] for the normal case, in the sense that we construct classes of priors giving rise to minimaxity. A feature of this paper is that in certain cases we are able to construct proper Bayes minimax estimators satisfying the properties and bounds in Strawderman [Minimax estimation of location parameters for certain spherically symmetric distribution, J. Multivariate Anal. 4 (1974) 255-264]. We also give some insight into why Strawderman's results

1. Université de Rouen, UMR CNRS 6085, BP 12 76801 Saint Étienne du Rouvray, France.

2. Université de Sherbrooke, Département de mathématiques, Canada.

3. Rutgers University, Department of Statistics, New Brunswick, NJ 08903, USA. The support of NSA Grant 03G-112 is gratefully acknowledged.

do or do not seem to apply in certain cases. In cases where it does not apply, we give minimax estimators based on Berger's [Minimax estimation of location vectors for a wide class of densities, Ann. Statist. 3 (1975) 1318-1328] results. A main condition for minimaxity is that the mixing distributions of the sampling distribution and the prior distribution satisfy a monotone likelihood ratio property with respect to a scale parameter.

AMS 1991 subject classifications. Primary 62C20, 62C15, 62C10.

Keywords and phrases : Minimax estimators, Variance mixture of multivariate normal distributions, Quadratic loss, Bayes estimators.

2.1 Introduction

In this paper we study Bayes minimax estimation of the mean vector of a variance mixture of a multivariate normal distributions under sum of squared errors loss in dimension three and greater. It has been known since Stein [Ste56], in the normal case, and Brown [Bro66], generally, that the best equivariant and minimax estimator is inadmissible for $p \geq 3$. Explicit improvements in the normal case were given by James and Stein [JS61] and by many other authors thereafter. Explicit improvements for other subclasses of distributions were given by several authors starting with Strawderman [Str74] for the case of variance mixtures of normals and Berger [Ber75] for certain general classes of spherically symmetric distributions. See also, for example [Boc85, BS78]. All of these early papers (except [Str74]) in the non-normal setting did not consider (generalized) Bayes minimax estimators.

Strawderman [Str71] gave proper Bayes minimax estimators for the normal case for $p \geq 5$, while Fourdrinier et al. [FSW98] constructed broad classes of proper and generalized Bayes estimators in this case as well.

Strawderman [Str74] gave generalized Bayes minimax estimators for certain variance mixtures of normals. Recently, Maruyama [Mar03a] extended these results and gave classes of proper and generalized Bayes estimators for the same subclass as Strawderman [Str74]. In particular these results are for the case where the mixing distribution has monotone likelihood ratio when considered as a scale family.

In this paper we restrict ourselves to the class of mixing distributions with monotone likelihood ratio. We give broad classes of priors (including Maruyama's), somewhat in the spirit of Fourdrinier et al. [FSW98], that lead to proper and generalized Bayes minimax estimators.

An interesting property of the class of priors, also shared by Maruyama's, is that they too are a variance mixture of normals with monotone likelihood ratio.

The methods of proof in this paper are also generally similar to those of Maruyama with one exception. Both papers consider estimators of the form $\delta(X) = (1 - r(\|X\|^2)/\|X\|^2)X$ where $r(w)$ is non-decreasing, non-negative, and bounded. The proofs of monotonicity and of determination of the bound are quite similar. To determinate minimaxity, one of two results is used in each paper ; (a) the result of Strawderman [Str74] which also requires $r(w)/w$ to be monotone non-increasing but gives a larger upper bound for $r(w)$ or (b) a result of Berger [Ber75] which requires the existence of more moments but does not require that $r(w)/w$ be non-increasing.

Our methods and results differ from those of Maruyama in that we are able to give conditions that guarantee monotonicity of $r(w)/w$ in certain general cases. In these cases it is possible to use the larger bound on $r(w)$ in [Str74] and to find proper Bayes minimax estimators.

Another main difference in the two papers is that the class of priors considered by Maruyama is specific. In our setting, this class is characterized by a mixing density on the variance of the form $h(t) = t^b(1+t)^{a-2-b}$ with $b \geq 0$. Our class of mixing distributions, which contains this class, is given by $h(t)$ such that $h(t) \leq Kt^{-\alpha}$ for $0 < t < t_0$ and $\alpha < 1$, $h(0) < \infty$ and $\lim_{t\to\infty} h(t)/t^{\beta} = c > 0$.

Section 2.2 states the problem and develops the form of the Bayes estimators. Section 2.3 contains the main results. Section 2.4 gives examples illustrating the theory. Two basic examples, a gamma mixture and an inverse gamma mixture (the Student t case) indicate when our method does or does not lead to a proper Bayes minimax estimator with $r(w)/w$ non-increasing. In Section 2.5, we give some concluding remarks. Finally, an Appendix gives certain of the proofs.

2.2 General expression of Bayes estimators

Let X be a random vector in $\mathbb{R}^p$ ($p \geq 3$) distributed as a variance mixture of multivariate normal distributions with mean vector θ. Thus we assume that the density function of X is of the form :

$$f(x) = \int_0^{\infty} \frac{1}{(2\pi v)^{p/2}} \exp\left(-\frac{1}{2}\frac{||x-\theta||^2}{v}\right) dG(v) \tag{2.2.1}$$

where G is the distribution of a known non-negative random variable V. More precisely, most of this paper is devoted to the case where G has a density g

with respect to the Lebesgue measure on $\mathbb{R}_+$ (g is said the mixing density). The goal of this paper is to give sufficient conditions for Bayes estimators δ of θ to be minimax, under the usual quadratic loss function $L(\theta, \delta) = ||\theta - \delta||^2$.

For a prior probability measure π the marginal distribution of X has a density m with respect to the Lebesgue measure in $\mathbb{R}^p$ given by :

$$m(x) = \int_{\mathbb{R}^p} \int_0^\infty \frac{1}{(2\pi v)^{p/2}} \exp\left(-\frac{1}{2}\frac{||x-\theta||^2}{v}\right) dG(v)\, d\pi(\theta). \tag{2.2.2}$$

Upon an application of Fubini's theorem for positive functions, for any $x \in \mathbb{R}^p$, we have :

$$m(x) = \int_0^\infty K(x,v)\, dG(v) \tag{2.2.3}$$

where

$$K(x,v) = \int_{\mathbb{R}^p} \frac{1}{(2\pi v)^{p/2}} \exp\left(-\frac{1}{2}\frac{||x-\theta||^2}{v}\right) d\pi(\theta)\,. \tag{2.2.4}$$

The Bayes estimator $\delta_\pi = \delta_\pi(X)$, which is defined as the minimizer of the Bayes risk, is given for any $x \in \mathbb{R}^p$ in the case of quadratic loss by $\delta_\pi(x) = E[\theta|x]$, where this last expectation is considered with respect of the posterior distribution given x. After classical calculations, we obtain :

$$\delta_\pi(x) = x + \gamma(x) \tag{2.2.5}$$

with

$$\gamma(x) = \frac{\nabla}{m(x)} \int_0^\infty v K(x,v)\, dG(v) \tag{2.2.6}$$

where the symbol $\nabla = (\partial/\partial x_1, ..., \partial/\partial x_p)$ denotes the gradient. Note that, in the special case where G is Dirac measure, we find again the normal case where $\gamma(x) = \nabla \log(m(x))$ considered in Fourdrinier et al. [FSW98].

Recall that the quadratic risk of any estimator $\delta(X) = X + \gamma(X)$, that is :

$$R(\theta, \delta) = E_\theta \left[\|\theta - \delta\|^2 \right]$$

(where E_θ denotes the expectation with respect to (2.2.1)), is finite as soon as the risk of X is finite (that is, if $E[V] < \infty$) if and only if $E\left[\|\gamma(X)\|^2\right] < \infty$ (this can be verified by an application of Schwarz's inequality).

As the sampling density is a variance mixture of normal distributions of the type (2.2.1), a natural choice for a prior distribution π is to assume that it has a density with respect to the Lebesgue measure of the form :

$$\theta \longmapsto \int_0^\infty \frac{1}{(2\pi t)^{p/2}} \exp\left(-\frac{1}{2}\frac{\|\theta\|^2}{t}\right) h(t)\, dt \tag{2.2.7}$$

where h is a function from $\mathbb{R}_+$ into $\mathbb{R}_+$ such that this integral exists (h is the mixing function). Note that this density is proper provided that $\int_0^\infty h(t)\, dt < \infty$.

In this context, the expression of $K(x, v)$ in (2.2.4) can be calculated through an application of Fubini's theorem for positive functions. Indeed, for any $x \in \mathbb{R}^p$ and any $v \geq 0$, we have :

$$\begin{aligned} K(x,v) &= \int_{\mathbb{R}^p} \frac{1}{(2\pi v)^{p/2}} \exp\left(-\frac{1}{2}\frac{\|x-\theta\|^2}{v}\right) \\ &\times \int_0^\infty \frac{1}{(2\pi t)^{p/2}} \exp\left(-\frac{1}{2}\frac{\|\theta\|^2}{t}\right) h(t)\, dt\, d\theta \\ &= \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \exp\left(-\frac{1}{2}\frac{\|x\|^2}{v+t}\right) h(t)\, dt\,. \end{aligned} \tag{2.2.8}$$

Then the marginal density $m(x)$ in (2.2.3) becomes :

$$m(x) = \int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \exp\left(-\frac{1}{2}\frac{\|x\|^2}{v+t}\right) h(t)\, dt\, dG(v)\,. \tag{2.2.9}$$

Through the Lebesgue dominated convergence theorem we obtain, according to (2.2.6),

$$\gamma(x) = \frac{1}{m(x)} \int_0^\infty v \, \nabla K(x,v) \, dG(v) \,. \tag{2.2.10}$$

Substituing $K(x,v)$ by its value given in (2.2.8), it follows that

$$\gamma(x) = -\, \frac{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \frac{v}{v+t} \exp\left(-\frac{1}{2}\frac{\|x\|^2}{v+t}\right) h(t)\, dt\, dG(v)}{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \exp\left(-\frac{1}{2}\frac{\|x\|^2}{v+t}\right) h(t)\, dt\, dG(v)}\, x \,. \tag{2.2.11}$$

Finally, the Bayes estimator resulting from a variance mixture of normal distributions prior is of the form :

$$\delta_h(x) = x - \frac{r(\|x\|^2)}{\|x\|^2} x \tag{2.2.12}$$

where

$$r(w) = w \, \frac{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \frac{v}{v+t} \exp\left(-\frac{1}{2}\frac{w}{v+t}\right) h(t)\, dt\, dG(v)}{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \exp\left(-\frac{1}{2}\frac{w}{v+t}\right) h(t)\, dt\, dG(v)} \,. \tag{2.2.13}$$

It is worth noting that the Bayes estimator has always finite risk. Indeed, for any $x \in \mathbb{R}^p$, we have :

$$\begin{aligned} \|\gamma(x)\|^2 &= \|x\|^2 \left[\frac{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \frac{v}{v+t} \exp\left(-\frac{1}{2}\frac{\|x\|^2}{v+t}\right) h(t)\, dt\, dG(v)}{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \exp\left(-\frac{1}{2}\frac{\|x\|^2}{v+t}\right) h(t)\, dt\, dG(v)} \right]^2 \\ &\le \|x\|^2 \end{aligned} \tag{2.2.14}$$

using $0 < \dfrac{v}{v+t} \le 1$. Thus, it follows that :

$$E_\theta \left[\|\gamma(X)\|^2\right] \le E_\theta \left[\|X\|^2\right] \le pE\left[V\right] + \|\theta\|^2 < \infty \tag{2.2.15}$$

since we have supposed that $E\left[V\right] < \infty$.

To prove the minimaxity of the Bayes estimator δ_h in (2.2.12), we cannot rely on an unbiased estimator of its risk, in contrast to Stein [Ste81] in the normal case. Strawderman [Str74] and Berger [Ber75], to prove minimaxity of an estimator of the type (2.2.12) for variance mixture of normal distributions, use the fact the function r is bounded from above and non-decreasing. They also need the property that the function $w \longmapsto r(w)/w$ is non-increasing. However it is worth noting that Berger [Ber75], proposing the class of densities $x \longmapsto f\left(\|x-\theta\|^2\right)$ such that $c = \inf_{t\geq 0} \int_t^\infty f(u)\,du/f(t) > 0$, does not use this last monotonicity condition. Furthermore it is interesting to notice that this class of densities contains some variance mixtures of normal distributions. Indeed, it is easy to show that a density of the form (2.2.1) belongs to this class with $c = 2\,E\left[V^{1-p/2}\right]/E\left[V^{-p/2}\right]$. This fact was used by Maruyama [Mar03a]. We recall below an adaptation of the results of Strawderman and Berger.

Theorem 2.2.1. *Let f be a density with respect to Lebesgue measure of the type* (2.2.1) *with $p \geq 3$. Let δ_h be an estimator of the form* (2.2.12) *where r is non-decreasing. Then δ_h is a minimax estimator of θ under quadratic loss if,*

either (a) $0 \leq r \leq c(p-2)$ with $c = 2\,E\left[V^{1-p/2}\right]/E\left[V^{-p/2}\right]$,

or (b) $0 \leq r \leq c^*(p-2)$ with $c^* = 2/E\left[V^{-1}\right]$ and $r(w)/w$ is non-increasing.

It is easy to see that $c^* > c$.

2.3 Minimax Bayes estimators

Throughout this section we assumed that the mixing distribution G has a density g. We will see that, for a sampling distribution of the form (2.2.1) and for a prior of the form (2.2.7), the main condition for obtaining minimaxity of the corresponding Bayes estimator is that both the mixing distribution g and the mixing (possibly improper) density h have monotone non-decreasing likelihood ratio when considered as a scale parameter family. It was noticed by Maruyama [Mar03a] that (expressed in terms of the function h) this property is respectively equivalent to

$$h(s_1t_1)h(s_2t_2) \leq h(s_1t_2)h(s_2t_1) \tag{2.3.1}$$

for any $s_1 \leq s_2$ and $t_1 \leq t_2$ and to the fact that the function

$$t \longmapsto t\,h'(t)/h(t) \tag{2.3.2}$$

is non-increasing (if h is absolutely continuous). Actually, in the following lemma, this monotone likelihood property implies the monotonicity of the function r in (2.2.13) under a mild growth condition on h.

Lemma 2.3.1. *If both h and g have monotone increasing likelihood ratio when considered as a scale parameter family and if $h(0) < \infty$, $h(t)$ is $o(t^{p/2-1})$ for t in a neighborhood of infinity and is absolutely continuous, then the function :*

$$r(w) = w\,\frac{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \frac{v}{v+t} \exp\left(-\frac{1}{2}\frac{w}{v+t}\right) h(t)\,dt\,g(v)\,dv}{\int_0^\infty \int_0^\infty \frac{1}{(2\pi(v+t))^{p/2}} \exp\left(-\frac{1}{2}\frac{w}{v+t}\right) h(t)\,dt\,g(v)\,dv} \tag{2.3.3}$$

is non-decreasing.

The proof of Lemma 2.3.1 is postponed to the appendix.

The following lemma (whose proof is also given in the appendix) gives conditions on h and g such that $\lim_{w\to\infty} r(w)$ can be determined. Under the additional conditions of Lemma 2.3.1, this limit is the upper bound of r.

Lemma 2.3.2. *Assume that the mixing density g of the sampling distribution in (2.2.1) is such that :*

$$E[V] = \int_0^\infty v g(v)\, dv < \infty \quad \text{and} \quad E[V^{-1}] = \int_0^\infty v^{-1} g(v)\, dv < \infty . \tag{2.3 4}$$

Assume also that the mixing density h of the (possibly improper) prior distribution in (2.2.7) satisfies

$$\lim_{t\longrightarrow\infty} \frac{h(t)}{t^\beta} = c \tag{2.3.5}$$

for some $\beta < p/2 - 1$ and some $c > 0$, and assume that :

$$E[V^{1-\beta}] = \int_0^\infty v^{1-\beta} g(v)\, dv < \infty . \tag{2.3.6}$$

Then

$$\lim_{w\longrightarrow\infty} r(w) = (p - 2 - 2\beta) E[V]$$

provided there exist $K > 0$, $t_0 > 0$ and $\alpha < 1$ such that :

$$h(t) \leq K t^{-\alpha} \text{ for } 0 < t < t_0 . \tag{2.3.7}$$

Note that Condition (2.3.5) implies that $h(t) \leq K t^\beta$ for $t_0 \leq t < \infty$ (it is clear that we can use the same K as in (2.3.7)).

Combining Theorem 2.2.1.a, Lemma 2.3.1 and Lemma 2.3.2 gives immediately our first domination result.

Theorem 2.3.1. *Assume that the mixing density g of the sampling distribution in* (2.2.1) *is such that :*

$$E[V] = \int_0^\infty v\,g(v)\,dv < \infty \text{ and } E[V^{-p/2}] = \int_0^\infty v^{-p/2}\,g(v)\,dv < \infty. \tag{2.3.8}$$

Assume also that the mixing density h of the (possibly improper) prior distribution in (2.2.7) *is absolutely continuous and satisfies Condition* (2.3.5), *for some $\beta < p/2 - 1$ and for $0 < c < \infty$. Assume finally that h and g have monotone increasing likelihood ratio when considered as a scale parameter family.*

Then, if Condition (2.3.7) *is satisfied for some $K > 0$, $t_0 > 0$ and $\alpha < 1$, the (generalized or proper) Bayes estimator δ_h with respect to the prior distribution corresponding to the mixing distribution h is minimax provided that β satisfies*

$$-(p-2)\left[\frac{E[V^{-\frac{p}{2}+1}]}{E[V]E[V^{-\frac{p}{2}}]} - \frac{1}{2}\right] \le \beta\,. \tag{2.3.9}$$

It is easy to see that Condition (2.3.8) implies Condition (2.3.6) since $\beta < p/2 - 1$. Note that $\beta < -1$ corresponds to the proper priors. Note also, since $E[V^{-\frac{p}{2}+1}]/(E[V]\ E[V^{-\frac{p}{2}}]\ ) \le 1$, we have :

$$-(p-2)\left[\frac{E[V^{-\frac{p}{2}+1}]}{E[V]E[V^{-\frac{p}{2}}]} - \frac{1}{2}\right] \ge -\frac{p-2}{2}\,.$$

Hence, in the case of a proper prior, it is necessary that $-(p-2)/2 < -1$ which is equivalent to $p > 4$. This is related to the known fact in the normal case that $p > 4$ is required for the existence of a proper Bayes minimax estimator (see [Str70]).

Comment (Admissibility) : For priors with mixing distribution h satisfying (2.3.5) and (2.3.7) an argument as in Maruyama [Mar03a] using Brown

[Bro79] and a Tauberian theorem suggests that the resulting generalized Bayes estimator is admissible if $\beta \leq 0$. A referee called to our attention that, recently, Maruyama and Takemura [MT06] have verified this under additional conditions which imply, in our setting, that $E_\theta[\|X\|^3] < \infty$.

As mentioned in Section 2.2, the bound $c^*(p-2)$ in part (b) of Theorem 2.2.1 for minimaxity is greater than or equal to the bound $c(p-2)$ in part (a). However this increase in the bound is obtained at the cost of the assumption of monotonicity of $r(w)/w$. In general, it does not appear that the assumptions of Lemma 2.3.1 are sufficient to guarantee this monotonicity. The reason relies in the fact that both densities g and h have a monotone likelihood ratio property in the same direction. This will become clearer in the discussion below and in the examples in Section 2.4.

Here is one way to deal with the monotonicity of $r(w)/w$. Note first that, according to (2.3.3), we have :

$$\frac{r(w)}{w} = E_w\left[\frac{V}{V+T}\right]$$

where E_w is the expectation with respect to the density

$$f(t, v|w) \propto \left(\frac{1}{2\pi(v+t)}\right)^{p/2} \exp\left(-\frac{1}{2}\frac{w}{v+t}\right) h(t)g(v)\,.$$

Through the change of variable $\lambda = v/(v+t)$ and $z = v+t$, we have :

$$f(\lambda, z|w) \propto z^{-p/2+1} \exp\left(-\frac{1}{2}\frac{w}{z}\right) h(z(1-\lambda))g(\lambda z)\,,$$

so that the density of λ given z and w does not depend on w since

$$f(\lambda|z, w) = \frac{f(\lambda, z|w)}{\int_0^1 f(\lambda, z|w)\, d\lambda} = c(z)h(z(1-\lambda))g(\lambda z) = f(\lambda|z)\,. \tag{2.3.10}$$

Further the family of densities given by :

$$f(z|w) = \int_0^1 f(\lambda, z|w)\, d\lambda = K(z) z^{-p/2+1} \exp\left(-\frac{1}{2}\frac{w}{z}\right),$$

for some constant $K(z)$, has monotone increasing likelihood ratio property with respect to the parameter w since, for $w_1 < w_2$, we have :

$$\frac{f(z\,|w_2)}{f(z\,|w_1)} = \exp\left(-\frac{1}{2}\frac{1}{z}(w_2 - w_1)\right).$$

Hence the monotonicity of $r(w)/w$ follows provided the function $E[\lambda|z, w]$ $= E[\lambda|z]$ is non-increasing in z. The monotone decreasing likelihood ratio property for the density $f(\lambda|z)$ in (2.3.10) is a natural sufficient condition to insure this monotonicity. Note however that $f(\lambda|z) \propto h(z(1-\lambda))g(z\lambda)$ where $h(z(1-\lambda))$ and $g(z\lambda)$ have monotone likelihood ratio in opposite direction according to Lemma 2.3.1. As we will see in Section 2.4, there are examples where $f(\lambda|z)$ has decreasing monotone likelihood ratio in z (Example 3.4.2) and examples where $f(\lambda|z)$ does not (Example 3.4.1).

The following theorem is useful in situations where decreasing monotone likelihood ratio property holds for $f(\lambda|z)$.

Theorem 2.3.2. *Under the conditions of Theorem 2.3.1, assume that $f(\lambda|z)$ in (2.3.10) has decreasing monotone likelihood ratio in z. Then minimaxity of δ_h is satisfied as soon as*

$$-(p-2)\left[\frac{1}{E[V^{-1}]E[V]} - \frac{1}{2}\right] \leq \beta < \frac{p}{2} - 1\,. \tag{2.3.11}$$

Proof. The proof follows from Theorem 2.2.1.b, Lemmas 2.3.1 and 2.3.2 and the above discussion. □

The minimaxity results of Theorem 2.3.1 and 2.3.2 rely on the monotone likelihood ratio property of the densities g and h. This class of densities is essentially described in the following lemma.

Lemma 2.3.3. *Let φ be a non-increasing function on $\mathbb{R}_+$ such that the indefinite integral of $\varphi(u)/u$ exists. Then the function h defined by :*

$$h(t) = C\exp\left\{\int_\alpha^t \frac{\varphi(u)}{u}\,du\right\} \tag{2.3.12}$$

with $C > 0$ and $\alpha > 0$ has monotone non-decreasing likelihood ratio when considered as a scale parameter family.

Proof. Let $0 < \sigma_1 < \sigma_2$. For any $t \in \mathbb{R}_+$, we have :

$$\begin{aligned}\frac{1/\sigma_2 h(t/\sigma_2)}{1/\sigma_1 h(t/\sigma_1)} &= \frac{\sigma_1}{\sigma_2}\exp\left\{\int_{t/\sigma_2}^{t/\sigma_1} \frac{-\varphi(u)}{u}\,du\right\}\\ &= \frac{\sigma_1}{\sigma_2}\exp\left\{\int_{1/\sigma_2}^{1/\sigma_1} \frac{-\varphi(tv)}{v}\,dv\right\}\end{aligned}$$

which is non-decreasing in t by monotonicity of φ. □

Note that the fact that Lemma 2.3.3 gives virtually all smooth functions with monotone likelihood ratio follows since $\varphi(t) = th'(t)/h(t)$ and thus h satisfies (2.3.2).

The monotone likelihood ratio property of $f(\lambda\,|z)$ needed in Theorem 2.3.2 involves the two densities g and h. For densities of the form (2.3.12), this link is made explicit in the following lemma.

Lemma 2.3.4. *Let φ (respectively φ_0) a derivable non-increasing function on $\mathbb{R}_+$ such that the indefinite integral of $\varphi(u)/u$ (respectively of $\varphi_0(u)/u$) exists. Then, for*

$$h(t) = C\exp\left\{\int_\alpha^t \frac{\varphi(u)}{u}\,du\right\} \text{ with } C > 0 \text{ and } \alpha > 0$$

and

$$g(v) = C_0\exp\left\{\int_{\alpha_0}^v \frac{\varphi_0(u)}{u}\,du\right\} \text{ with } C_0 > 0 \text{ and } \alpha_0 > 0\,,$$

the density $f(\lambda\,|z)$ given in (2.3.10) *has decreasing monotone likelihood ratio in z if and only if the derivatives of φ and φ_0 satisfy*

$$\varphi_0'(z\lambda) \le \varphi'(z(1-\lambda)) \tag{2.3.13}$$

for any $z \ge 0$ and any $\lambda \in]0,1[$.

Proof. Let $0 \le z_1 < z_2$. According to (2.3.10), we have to show that the function

$$\frac{h(z_2(1-\lambda))\; g(z_2\lambda)}{h(z_1(1-\lambda))\; g(z_1\lambda)}$$

is non-increasing in λ which reduces to

$$\frac{d}{d\lambda}\left\{\int_\alpha^{z_2(1-\lambda)} \frac{\varphi(u)}{u}\,du - \int_\alpha^{z_1(1-\lambda)} \frac{\varphi(u)}{u}\,du + \int_{\alpha_0}^{z_2\lambda} \frac{\varphi_0(u)}{u}\,du - \int_{\alpha_0}^{z_1\lambda} \frac{\varphi_0(u)}{u}\,du\right\} \le 0$$

that is to

$$\frac{\varphi(z_1(1-\lambda))}{1-\lambda} - \frac{\varphi_0(z_1\lambda)}{\lambda} \le \frac{\varphi(z_2(1-\lambda))}{1-\lambda} - \frac{\varphi_0(z_2\lambda)}{\lambda}.$$

The latter inequality expresses that the function

$$z \longmapsto \frac{\varphi(z(1-\lambda))}{1-\lambda} - \frac{\varphi_0(z\lambda)}{\lambda}$$

is non-decreasing in z, for any $\lambda \in]0,1[$, and hence that

$$\varphi'(z(1-\lambda)) - \varphi_0'(z\lambda) \ge 0$$

for any $z \ge 0$ and any $\lambda \in]0,1[$. $\square$

2.4 Examples

In this section, we consider sampling distributions where the mixing density g is a gamma or an inverse gamma density. We start with the latter which corresponds to a generalized Student t.

Example 3.4.1. Inverse gamma : Let $g(v) \propto 1/v^{a_0+1} \exp(-b_0/v)$ with $a_0 > 1$ and $b_0 > 0$. Let also the mixing prior h be a (possibly) generalized inverse gamma, that is, $h(t) \propto 1/t^{a+1} \exp(-b/t)$ with $a > -p/2$ and $b > 0$. Note that h and g have monotone increasing likelihood ratio when considered as a scale parameter family.

(1) Minimaxity using Theorem 2.3.1 : It is clear that Conditions (2.3.8) and (2.3.5) are satisfied with $a_0 > 1$ and $\beta = -(a+1)$. It is also clear that Condition (2.3.7) holds for any $\alpha < 1$. Finally a simple calculation shows that

$$\frac{E[V^{-\frac{p}{2}+1}]}{E[V]E[V^{-\frac{p}{2}}]} = \frac{a_0 - 1}{p/2 + (a_0 - 1)}$$

so that Condition (2.3.9) reduces to

$$a \leq (p-2)\left[\frac{a_0 - 1}{p/2 + a_0 - 1} - \frac{1}{2}\right] - 1 \tag{2.4.1}$$

which guarantees the minimaxity of the (generalized or proper) Bayes estimator δ_h.

Note that the condition $E[V^{1-\beta}] = E[V^{a+2}] < \infty$ is equivalent to $a_0 - 2 > a$. Thus $a_0 \neq a$ and the mixing densities g and h cannot be equal. This is consistent with the fact that, when the sampling and the prior distributions are identical, necessarily, the Bayes estimator is $X/2$.

Note also that properness of h requires $a > 0$ so that

$$0 < a \leq (p-2)\left[\frac{a_0 - 1}{p/2 + (a_0 - 1)} - \frac{1}{2}\right] - 1$$

and this double inequality can hold if and only if $p \geq 5$ and

$$a_0 > 1 + \frac{p}{2} \frac{p}{p-4} .$$

In particular, if the sampling distribution is a $p-$variate Student t with n_0 degrees of freedom, the mixing density g is the inverse gamma $(n_0/2, n_0/2)$, that is $g(v) \propto v^{-\frac{n_0+2}{2}} \exp(-n_0/2v)$. Similarly, if the prior is Student t distribution with n degrees of freedom, we have $h(t) \propto t^{-\frac{n+2}{2}} \exp(-n/2t)$. This corresponds to $a_0 = n_0/2$, $b_0 = n_0/2$, $a = n/2$ and $b = n/2$.

The condition $a_0 > 1$ corresponds to $n_0 > 2$ and the condition $a > -p/2$ is of course satisfied since it is a "true" Student t distribution. The condition for minimaxity (2.4.1) becomes :

$$n \leq \frac{2(n_0 - 2)}{p + n_0 - 2}(p - 2) - p . \tag{2.4.2}$$

Note also that the above remark $a_0 - 2 > a$ corresponds to $n_0 > n + 4$.

The properness condition imposes that the left-hand side of (2.4.2) is positive which implies

$$n_0 > 2 + \frac{p^2}{p-4} .$$

It is worth noting that, for large n_0, Condition (2.4.2) becomes approximatively $n \leq p - 4$ which corresponds to the condition for proper Bayes minimaxity under a normal sampling distribution given in [FSW98].

(2) Minimaxity using Theorem 2.3.2 : Coming back to the general case of inverse gamma densities, it is clear that Theorem 2.3.2 does not apply with the choice of a prior mixing density in the same class as the sampling mixing density. Indeed it is easy to show that, for $z_1 < z_2$,

$$\frac{f(\lambda|\, z_2)}{f(\lambda|\, z_1)} = \psi(z_1, z_2) \exp\left(\left(\frac{1}{z_1} - \frac{1}{z_2}\right)\left(\frac{b}{1-\lambda} + \frac{b_0}{\lambda}\right)\right)$$

(for some function ψ) which is non-monotone in λ.

In fact note that g (respectively h) corresponds, in Lemma 2.3.3, to the choice of $\varphi_0(v) = -(a_0+1) + b_0/v$ (respectively $\varphi(t) = -(a+1) + b/t$) and it is easy to check that Condition (2.3.13) in Lemma 2.3.4 is not satisfied. To find a density h corresponding to the choice of the density g in order that Lemme 2.3.4 applies reduces to exhibit a function φ such that :

$$0 \leq -\ \varphi'(z(1-\lambda)) \leq \frac{b_0}{z^2\lambda^2} \tag{2.4.3}$$

for any $z \geq 0$ and any $\lambda \in [0,1]$. For fixed $u \geq 0$, Condition (2.4.3) implies that

$$0 \leq -\ \varphi'(u) \leq \frac{b_0}{(z-u)^2}$$

for any $z \geq 0$ and hence, when z goes to infinity, it follows that $-\varphi'(u) = 0$. Thus the function φ is constant, that is, $\varphi(u) = \gamma$ and necessarily, according to (2.3.12), we have that $h(t) = C\ t^{\gamma}$ which is improper for any γ. Now Conditions (2.3.5) and (2.3.7) impose that $\gamma = \beta \geq -\ \alpha > -1$.

Minimaxity of δ_h will follow from (2.3.11), that is, from the existence of a non-empty interval of the form :

$$\left[-(p-2)\left(\frac{a_0-1}{a_0} - \frac{1}{2}\right),\ \frac{p}{2} - 1\right) \tag{2.4.4}$$

for the value of β. Using the fact that $\beta > -1$ the following cases arise.

When $p = 3$ or 4, it is easy to check that $-\ (p-2)((a_0-1)/a_0 - 1/2) > -1$ since $1/a_0 > (p-4)/2(p-2)$. Thus the range of values of β is exactly the interval in (2.4.4). This is still the case, when $p \geq 5$, as soon as $a_0 < 2(p-2)/(p-4)$. However, when $p \geq 5$ and $a_0 \geq 2(p-2)/(p-4)$, the range of value of β reduces to $(-1,\ p/2-1)$.

Hence, for the case of an inverse gamma mixing distribution, g, there exists a mixing distribution, h, of the form $h(t) \propto t^{\beta}$ which results in a minimax (improper) Bayes estimator. Note that this result implies that, in the Student case mentioned above, minimaxity of δ_h is guaranteed for any $n_0 \geq 3$.

Example 3.4.2. Gamma : Let $g(v) \propto v^{a_0-1} \exp(-b_0 v)$ with $a_0 > 0$ and $b_0 > 0$. Clearly g has increasing monotone likelihood ratio property as a scale parameter family.

(1) Minimaxity using Theorem 2.3.1 : The choice of a function h also proportional to a gamma density, that is, $h(t) \propto t^{a-1} \exp(-bt)$ with $b > 0$ seems natural but, for any $\beta \in \mathbb{R}$, $c = \lim_{t \to \infty} h(t)/t^{\beta} = 0$ and Theorem 2.3.1 does not apply since Condition (2.3.5) needs $c > 0$.

Now it is worth noting that the choice of an inverse gamma type density for h allows to use Theorem 2.3.1. Indeed, for $h(t) \propto 1/t^{a+1} \exp(-b/t)$ with $b > 0$ and $a > -p/2$, satisfies Condition (2.3.5) with $\beta = -(a+1)$. It is also clear that, for $a_0 > p/2$ Condition (2.3.8) is satisfied. Finally a simple calculation shows that

$$\frac{E[V^{-\frac{p}{2}+1}]}{E[V]E[V^{-\frac{p}{2}}]} = \frac{a_0 - p/2}{a_0}$$

so that Condition (2.3.9) reduces to

$$\frac{p-2}{2}\left[1 - \frac{p}{a_0}\right] \geq a + 1 . \tag{2.4.5}$$

Then it follows from Theorem 2.3.1 that the Bayes estimator δ_h in (2.2.12) is minimax provided that (2.4.5) is satisfied. Note that, properness of the estimator δ_h needs that the right-hand side of (2.4.5) be greater than 1, which requires that $a_0 > p\,(p-2)/(p-4)$, provided $p \geq 5$.

(2) Minimaxity using Theorem 2.3.2 : Note that the densities h and g correspond respectively to the functions $\varphi(t) = -(a+1) + b/t$ and $\varphi_0(v) = a_0 - 1 - b_0 v$ in Lemma 2.3.4. Obviously Condition (2.3.13) is not satisfied since it reduces to

$$- b_0 \leq \frac{- b}{(z(1-\lambda))^2}$$

for any z and any λ. Hence, we cannot apply Theorem 2.3.2 for this choice of h.

The choice of $\varphi(t) = \nu - \mu \frac{t/\sigma}{1 + t/\sigma}$, with $\nu \in \mathbb{R}$, $\mu \geq 0$ and $\sigma > 0$, leads to the density $h(t) \propto (t/\sigma)^{\nu}(1 + t/\sigma)^{-\mu}$ according to (2.3.12). Thus Condition (2.3.13) expresses that, for any $z \geq 0$ and any $\lambda \in]0,1[$,

$$- b_0 \leq \frac{- \mu/\sigma}{(1 + z(1-\lambda)/\sigma)^2}$$

and is satisfied as soon as $b_0 \geq \mu/\sigma$.

In this case Theorem 2.3.2 does apply with $\beta = \nu - \mu$ and $\nu > -1$. Condition (2.3.11) becomes :

$$- (p-2)\left[\frac{1}{2} - \frac{1}{a_0}\right] \leq \nu - \mu < \frac{p}{2} - 1 \tag{2.4.6}$$

This interval of values of $\beta = \nu - \mu$ is non-void for all $a_0 > 1$. Hence it is always possible to find pairs (μ, ν) with $\mu > 0$ and $\nu > -1$ which satisfy (2.4.6). It remains for any such choice to choose a scale parameter, σ, for the mixing density h such that $b_0 \geq \mu/\sigma$.

Propriety of h requires in addition that $\beta = \nu - \mu < -1$. This condition in turn implies that $-\ (p-2)[1/2 - 1/a_0] < -1$, since $a_0 > 1$, p must be at least 5, and $a_0 > 2(p-2)/(p-4)$. Under this condition it is always possible to apply Theorem 2.3.2 to obtain (many) proper Bayes minimax estimators corresponding to priors of the form $h(t) = t^{\nu}(1 + t/\sigma)^{-\mu}$.

It should be noted that when $\sigma = 1$, this class of priors corresponds to the class of Maruyama. It is also interesting to note (and this is how we initially found the class) that this class arises as an inverse gamma mixture of gamma densities.

The class of priors h, to which Theorem 2.3.2 applies, is however much broader, as is the class of mixing densities g. This is the subject of our final example.

Example 3.4.3. Mixtures with bounded ψ : Suppose g is any mixing density such that :

$$\psi_0(v) = \frac{d}{dv}[\ v\ \frac{d}{dv}g(v)/g(v)\ ] = \frac{d}{dv}[\ \varphi_0(v)\ ]$$

satisfies $-\infty < \psi_0(v) \leq -\ b_0 < 0$. The gamma class of Example 3.4.2 is such a density. The inverse gamma class of Example 3.4.1 is not.

Let $h_1(t)$ be any mixing density satisfying $\lim_{t\to\infty} h_1(t)/t^\beta = c$ and such that, $0 \geq \psi_1(t) = \frac{d}{dt}[\ \varphi_1(t)\ ] > -b_1$. Then, if $h_\sigma(t) = \frac{1}{\sigma}h_1(t/\sigma)$, an easy calculation gives $\psi_\sigma(t) = \frac{d}{dt}\varphi_\sigma(t) = \frac{1}{\sigma}\psi_1(t/\sigma)$. Hence $\psi_\sigma(t) = \frac{1}{\sigma}\psi_1(t/\sigma) > -\ b_1/\sigma \geq -\ b_0$ as soon as $\sigma \geq b_1/b_0$.

The mixing density $h_1(t) = t^\nu(1+t)^{-\mu}$ of Example 3.4.2 is such a density. It is interesting that scaling the prior mixing density does not affect the value β, but it does affect the lower bound of the function $\psi_\sigma(t)$.

It is easy to construct other such examples of $h_1(t)$. One such example is :

$$h_1(t) = t^\nu(1+t)^{-\mu_1}(2+t)^{-\mu_2}\ with\ \nu > -1, \mu_1 + \mu_2 > 0$$

for which

$$0 > \psi_1(t) = -\ \mu_1\left(\frac{1}{1+t}\right)^{-2} - \mu_2\left(\frac{1}{2+t}\right)^{-2} \geq -\mu_1 - \frac{\mu_2}{4}$$

and for which $\beta = \nu - \mu_1 - \mu_2$.

In such cases Theorem 2.3.2 applies. It results in proper Bayes estimators whenever $p \geq 5$, $E[V]\ E[V^{-1}] > 2(p-2)/(p-4)$ so that the left-hand side of (2.3.11) is less than -1. In this case, choosing

$$-(p-2)\left[\frac{1}{E[V^{-1}]E[V]} - \frac{1}{2}\right] \leq \beta < -1\,,$$

and a scale parameter σ such that $-\ b_1/\sigma > -\ b_0$, the resulting procedure is proper Bayes and minimax. Choosing the same scale parameter and $-1 \leq \beta < p/2 - 1$ gives a generalized (non-proper) Bayes minimax estimator. As noted above $\beta \leq 0$ corresponds to admissible estimator.

2.5 Conclusions

We have studied Bayes minimax estimation for the case of variance mixture of normal distributions in dimension 3 and higher. We have assumed throughout that the prior distribution is also a variance mixture of normals and that both sampling and prior mixing distributions have monotone non-decreasing likelihood ratio when considered as a scale parameter family. We, and Maruyama [Mar03a], use the monotone likelihood ratio property to establish monotonicity of the function $r(w)$.

Our minimaxity results (and Maruyama's) rely on the results of Strawderman [Str74] when $r(w)/w$ is non-increasing, and on a result of Berger [Ber75] when $r(w)/w$ cannot be shown to be non-increasing. In either case $r(w)$ is required to be non-decreasing, non-negative and bounded by a constant. This constant is always larger for the Strawderman case than for the Berger case.

The class of mixing priors in Maruyama, in our setting (his parameterization is slightly different than ours), is given by $h(t) \propto t^b(1+t)^{a-2-b}$ with $b > -1$. These priors are proper for $a < 1$ and improper for $a \geq 1$.

Our class is a generalization of Maruyama's in that we assume $h(t) \leq Kt^{-\alpha}$ for $0 < t < t_0$ and $\alpha < 1$, $h(0) < \infty$ and $\lim_{t\to\infty} h(t)/t^\beta = c > 0$. These priors are proper for $\beta < -1$. Each paper establishes that $\lim_{w\to\infty} r(w) = (p-2-2\beta)E[V]$ (however, Maruyama's V is our V^{-1}).

Both papers use this bound, the monotonicity of r and the Strawderman or Berger result to obtain minimaxity of the resulting proper or generalized Bayes estimator. With (our β) equal to (Maruyama's $a-2$), the conditions for minimaxity in the two papers agree.

A point of departure in the present paper is that we study in some detail, conditions under which $r(w)/w$ is non-increasing. In fact we give broad classes of examples where $r(w)/w$ is non-increasing and where the resulting proper Bayes estimator is minimax. This development depends on the fact that $h(t) = C\exp\left(\int_\alpha^t \frac{\varphi(u)}{u}\,du\right)$ (respectively $g(v) = C_0\exp\left(\int_{\alpha_0}^v \frac{\varphi_0(u)}{u}\,du\right)$) with $\varphi(u)$ (respectively $\varphi_0(u)$) non-increasing, then h (respectively g) has non-increasing monotone likelihood ratio as a scale family. Using these expressions we show that $r(w)/w$ is non-increasing whenever

$$\sup \varphi_0'(u) \leq \inf \varphi'(u)\,. \tag{2.5.1}$$

In particular, if the function g is such that $\varphi_0'(v) \leq -\, b_0 < 0$, we construct functions h such that (2.5.1) holds. One way to do this is to consider a scaled version of Maruyama's class of priors and to choose the scale parameter sufficiently large.

This method leads to proper Bayes minimax estimators (with $r(w)/w$ decreasing) in the case where $g(v)$ is a gamma distribution (with exponential tails) and it fails to do so in the case where $g(v)$ is an inverse gamma distribution (with polynomial tails).

As in Maruyama, in the inverse gamma case, we give generalized Bayes minimax estimators with decreasing $r(w)/w$. Both Maruyama and we give proper Bayes minimax estimators based on the Berger result.

Acknowledgments

The autors are grateful to two referees for their careful reading and for their comments which improved the manuscript.

2.6 Appendix

We first recall here the FKG inequality established by Fortuin et al. [FKG71] which will be used in the proof of Lemma 2.3.1.

Lemma 2.6.1. *Let ξ denote a probability density function with respect to a σ-finite measure ν on $\mathbb{R}^n$. For any two points $y = (y_1, ..., y_n)$ and $z = (z_1, ..., z_n)$, define*

$$y \wedge z = (y_1 \wedge z_1, ..., y_n \wedge z_n)$$

$$y \vee z = (y_1 \vee z_1, ..., y_n \vee z_n)$$

(where $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$) and suppose that ξ satisfies

$$\xi(y)\xi(z) \leq \xi(y \wedge z)\xi(y \vee z).$$

If the functions $f(y)$ and $g(y)$ are non-decreasing in each arguement and if f, g and fg are integrable with respect to ξ, then

$$\int_{\mathbf{R}^n} f(y)g(y)\xi(y)\,d\nu(y) \geq \int_{\mathbf{R}^n} f(y)\xi(y)\,d\nu(y) \int_{\mathbf{R}^n} g(y)\xi(y)\,d\nu(y).$$

Proof of Lemma 2.3.1. Through the change of the variables $u = \dfrac{1}{v}$ and $t = \dfrac{1}{v}\dfrac{1-\lambda}{\lambda}$ in the integral (2.2.13), the function r characterising the Bayes estimator can be expressed, for any $w \geq 0$, as

$$r(w) = w\frac{\phi_1(w)}{\phi_0(w)} \tag{2.6.1}$$

where, for $i = 0, 1$,

$$\phi_i(w) = \int_0^1 \int_0^\infty \lambda^{\frac{p}{2}-2+i} u^{\frac{p}{2}-3} \exp\left(-\frac{1}{2}\lambda u w\right) h\left(\frac{1}{u}\frac{1-\lambda}{\lambda}\right) g\left(\frac{1}{u}\right) du\, d\lambda. \tag{2.6.2}$$

Then, to study the monotonicity of r, differentiating (2.6.1) with respect to w, we have :

$$r'(w) = \frac{\phi_1(w)}{\phi_0(w)} - \frac{1}{2}w\frac{A_1(w)\phi_0(w) - A_0(w)\phi_1(w)}{\phi_0^2(w)} \tag{2.6.3}$$

where, for $i = 0, 1$,

$$A_i(w) = \int_0^\infty \eta_i(u) u^{\frac{p}{2}-2} g\left(\frac{1}{u}\right) du \tag{2.6.4}$$

with

$$\eta_i(u) = \int_0^1 \lambda^{\frac{p}{2}-1+i} \exp\left(-\frac{1}{2}\lambda u w\right) h\left(\frac{1}{u}\frac{1-\lambda}{\lambda}\right) d\lambda. \tag{2.6.5}$$

Now integrating by parts in (2.6.5) and applying $h(0) < \infty$ and the growth condition $h(t) = o(t^{p/2-1})$ for t in a neighborhood of infinity to the bracketed term yield

$$\begin{aligned}\eta_i(u) = \frac{2}{uw}\Bigg\{&-h(0)\exp\left(-\frac{1}{2}uw\right)\\ &+ (\frac{p}{2} - 1 + i)\int_0^1 \lambda^{\frac{p}{2}-2+i} \exp\left(-\frac{1}{2}\lambda u w\right) h\left(\frac{1}{u}\frac{1-\lambda}{\lambda}\right) d\lambda\\ &\qquad - \frac{1}{u}\int_0^1 \lambda^{\frac{p}{2}-3+i} \exp\left(-\frac{1}{2}\lambda u w\right) h'\left(\frac{1}{u}\frac{1-\lambda}{\lambda}\right) d\lambda\Bigg\}.\end{aligned} \tag{2.6.6}$$

Hence (2.6.4) and (2.6.6) give, for $i = 0, 1$,

$$A_i(w) = \frac{2}{w}\left[\left(\frac{p}{2} - 1 + i\right)\phi_i(w) - B_i(w) - C(w)\right] \tag{2.6.7}$$

where

$$B_i(w) = \int_0^1 \int_0^\infty \lambda^{\frac{p}{2}-3+i} u^{\frac{p}{2}-4} \exp\left(-\frac{1}{2}\lambda u w\right) h'\left(\frac{1}{u}\frac{1-\lambda}{\lambda}\right) g\left(\frac{1}{u}\right) du\, d\lambda \tag{2.6.8}$$

and

$$C(w) = h(0)\int_0^\infty u^{\frac{p}{2}-3} \exp\left(-\frac{1}{2}uw\right) g\left(\frac{1}{u}\right) du\,. \tag{2.6.9}$$

Finally, combining (2.6.3) and (2.6.7) gives, after some algebra, the following expression :

$$r'(w) = \frac{1}{\phi_0^2(w)}\Big[\phi_0(w)B_1(w) - \phi_1(w)B_0(w) + C(w)(\phi_0(w) - \phi_1(w))\Big]. \tag{2.6.10}$$

As we want to prove that $r'(w) \geq 0$, it follows from (2.6.10) that it suffices that

$$\frac{B_1(w)}{\phi_0(w)} \geq \frac{\phi_1(w)}{\phi_0(w)}\frac{B_0(w)}{\phi_0(w)} \tag{2.6.11}$$

since, according to (2.6.2) and to (2.6.9), $\phi_0(w) \geq \phi_1(w)$ and $C(w) \geq 0$ respectively. This can be proved through a new expression of $\phi_i(w)$ and $B_i(w)$. Using the change of variable $u = \frac{1-\lambda}{\lambda}t$ in (2.6.2) and (2.6.8) and setting

$$G_w(t, \lambda) = t^{\frac{p}{2}-3}(1-\lambda)^{\frac{p}{2}-2} \exp\left(-\frac{1}{2}(1-\lambda)tw\right) h\left(\frac{1}{t}\right) g\left(\frac{1}{t}\frac{\lambda}{1-\lambda}\right) \tag{2.6.12}$$

it can be shown that, for $i = 0, 1$,

$$\phi_i(w) = \int_0^\infty \int_0^1 \lambda^i G_w(t, \lambda)\, d\lambda\, dt$$

and

$$B_i(w) = \int_0^\infty \int_0^1 \frac{\lambda^i}{1-\lambda} \frac{1}{t} \frac{h'(\frac{1}{t})}{h(\frac{1}{t})} G_w(t, \lambda)\, d\lambda\, dt.$$

Thus inequality (2.6.11) can be interpreted as follows : with respect to the density $\xi_w : (\lambda, t) \longmapsto G_w(t, \lambda)/\phi_0(w)$, its left-hand side appears as the expectation of the product of the functions $\varphi : (\lambda, t) \longmapsto \dfrac{1}{1-\lambda} \dfrac{1}{t} \dfrac{h'(\frac{1}{t})}{h(\frac{1}{t})}$ and $\psi : (\lambda, t) \longmapsto \lambda$ while its right-hand side is the product of the respective expectations of φ and ψ. Note that both functions $\varphi(\lambda, t)$ and $\psi(\lambda, t)$ are non-decreasing in both arguments λ and t (the monotonicity in t of $\varphi(\lambda, t)$ coming from the monotone likelihood ratio assumption of h as a scale parameter family mentioned above).

Then inequality (2.6.11) will follow from the FKG inequality recalled in Lemma 2.6.1. Indeed, fix $w \geq 0$, $\lambda_1 \leq \lambda_2$ and $t_1 \leq t_2$. First it is clear that

$$\exp\left(-\frac{1}{2}(1-\lambda_1)t_2 w\right) \exp\left(-\frac{1}{2}(1-\lambda_2)t_1 w\right) \leq \exp\left(-\frac{1}{2}(1-\lambda_1)t_1 w\right) \exp\left(-\frac{1}{2}(1-\lambda_2)t_2 w\right). \quad (2.6.13)$$

Now, by monotone likelihood ratio property of g considered as a scale parameter family and since the functions $\lambda \longmapsto \lambda/(1-\lambda)$ and $t \longmapsto 1/t$ are respectively increasing in λ and decreasing in t, we have according to (2.3.1)

$$g\left(\frac{1}{t_2}\frac{\lambda_1}{1-\lambda_1}\right) g\left(\frac{1}{t_1}\frac{\lambda_2}{1-\lambda_2}\right) \leq g\left(\frac{1}{t_1}\frac{\lambda_1}{1-\lambda_1}\right) g\left(\frac{1}{t_2}\frac{\lambda_2}{1-\lambda_2}\right). \quad (2.6.14)$$

Therefore, according to (2.6.12), it follows from (2.6.13) and (2.6.14) that

$$\xi_w(t_1, \lambda_2)\ \xi_w(t_2, \lambda_1) \leq \xi_w(t_1, \lambda_1)\ \xi_w(t_2, \lambda_2).$$

Finally Lemma 2.6.1 gives Inequality (2.6.11) which proves that the function r is non-decreasing. $\square$

Proof of Lemma 2.3.2. We give the proof of the case $\beta < 0$, the case $\beta \geq 0$ being similar and somewhat simpler. Dividing numerator and denominator of the right-hand side of (2.3.3) by w^β and using the change of variable $t = v(w-s)/s$, we obtain :

$$r(w) = \frac{\int_0^\infty \int_0^\infty M_1(w,s,v)\,ds\,dv}{\int_0^\infty \int_0^\infty M_0(w,s,v)\,ds\,dv} \tag{2.6.15}$$

where, for $i = 0, 1$,

$$M_i(w,s,v) = w^{-\beta} 1_{[0,w]}(s) s^{\frac{p}{2}-2+i} \exp\left(-\frac{s}{2v}\right) h\left(v\frac{w-s}{s}\right) v^{-p/2+1} g(v). \tag{2.6.16}$$

We now bound the integral of M_i to apply the Lebesgue dominated convergence theorem. Before, for simplicity, let :

$$\lambda_0(v) = \lambda_0 = \frac{v/t_0}{1+v/t_0},$$

$$K_0(w,v) = \int_0^{\lambda_0 w} w^{-\beta} s^{\frac{p}{2}-2+i} \exp\left(-\frac{s}{2v}\right) h\left(v\frac{w-s}{s}\right) ds$$

and

$$K_1(w,v) = \int_{\lambda_0 w}^{w} w^{-\beta} s^{\frac{p}{2}-2+i} \exp\left(-\frac{s}{2v}\right) h\left(v\frac{w-s}{s}\right) ds$$

so that

$$\int_0^\infty \int_0^\infty M_i(w,s,v)\,ds\,dv = \int_0^\infty \Big[K_0(w,v) + K_1(w,v)\Big] v^{-p/2+1} g(v)\,dv\,.$$

Note now that, for $0 < s < w$, under Condition (2.3.7), we have :

$$h\left(v\frac{w-s}{s}\right) \leq K \left(v\frac{w-s}{s}\right)^{-\alpha}$$

if $0 < \lambda_0 < s/w < 1$. Therefore

$$\begin{aligned} K_1(w,v) &\leq K \int_{\lambda_0 w}^{w} w^{-\beta} s^{p/2-2+\imath} \exp\left(-\frac{s}{2v}\right) \left(v\frac{w-s}{s}\right)^{-\alpha} ds \\ &= K w^{-\beta+p/2-1+\imath}\, v^{-\alpha} \int_{\lambda_0}^{1} \lambda^{p/2-2+\alpha+\imath}(1-\lambda)^{-\alpha} \exp\left(-\frac{\lambda w}{2v}\right) d\lambda \end{aligned} \tag{2.6.17}$$

through the change of variable $s = \lambda w$. Using $w^B e^{-Aw} \leq B^B e^B A^{-B}$, it follows that, for some constant C,

$$\begin{aligned} K_1(w,v) &\leq KCv^{-\beta-\alpha+p/2+\imath-1} \int_{\lambda_0}^{1} \lambda^{\beta+\alpha-1}(1-\lambda)^{-\alpha}\, d\lambda \\ &\leq KCv^{-\beta-\alpha+p/2+\imath-1} \left(\frac{v/t_0}{1+v/t_0}\right)^{\beta+\alpha-1} \int_0^1 (1-\lambda)^{-\alpha}\, d\lambda\,, \end{aligned}$$

since $\beta < 0$ and $\alpha - 1 < 0$. Finally, after some algebra, we have, for some constant μ,

$$\begin{aligned} \int_0^\infty K_1(w,v) v^{-p/2+1} g(v)\, dv &\leq \mu \int_0^\infty \left(1+\frac{v}{t_0}\right)^{1-\beta-\alpha} v^{\imath-1} g(v)\, dv \\ &\leq \mu\, 2^{1-\beta-\alpha} \int_0^\infty \left[1+\left(\frac{v}{t_0}\right)^{1-\beta-\alpha}\right] v^{\imath-1} g(v)\, dv \end{aligned} \tag{2.6.18}$$

where the right-hand side of (2.6.18) is finite since $\int_0^\infty v^{\imath-\beta-\alpha} g(v)\, dv < \infty$ and $\int_0^\infty v^{\imath-1} g(v)\, dv < \infty$ by assumption.

We now consider the case $0 < s/w < \lambda_0 < 1$. We have (see comment after Lemma 2.3.2) :

$$h\left(v\frac{w-s}{s}\right) \leq K\left(v\frac{w-s}{s}\right)^{\beta}$$

which imply that

$$\begin{aligned} K_0(w,v) &\leq K \int_0^{\lambda_0 w} w^{-\beta} s^{p/2-2+\imath} \exp\left(-\frac{s}{2v}\right) \left(v\frac{w-s}{s}\right)^{\beta} ds \\ &\leq Kv^{\beta} \int_0^{\lambda_0 w} s^{p/2-2-\beta+\imath} \exp\left(-\frac{s}{2v}\right) \left(1-\frac{s}{w}\right)^{\beta} ds \\ &\leq Kv^{\beta} \left(\frac{1}{1+v/t_0}\right)^{\beta} \int_0^{\lambda_0 w} s^{p/2-2-\beta+\imath} \exp\left(-\frac{s}{2v}\right) ds \end{aligned}$$

since $\beta < 0$. Hence

$$K_0(w.v) \leq K v^\beta \left(\frac{1}{1+v/t_0}\right)^\beta \Gamma(p/2-1-\beta+i)(2v)^{p/2-1-\beta+i}$$

and then, for some constant K', we have :

$$\begin{aligned}\int_0^\infty K_0(w,v)\, v^{-p/2+1} g(v)\, dv &\leq K' \int_0^\infty (1+v/t_0)^{-\beta}\, v^i g(v)\, dv \\ &\leq K'\, 2^{-\beta} \int_0^\infty \left[1+(v/t_0)^{-\beta}\right] v^i g(v)\, dv \quad (2.6.19)\end{aligned}$$

where the right-hand side of (2.6.19) is finite since $\int_0^\infty v^{i-\beta} g(v)\, dv < \infty$.

The finitness of the integrals in (2.6.18) and (2.6.19) allows to apply the Lebesgue dominated convergence theorem. Then, according to (2.6.16) and (2.3.5)

$$\begin{aligned}\lim_{w\to\infty} \int_0^\infty \int_0^\infty M_i(w,s,v)\, ds\, dv &= \int_0^\infty \int_0^\infty \lim_{w\to\infty} M_i(w,s,v)\, ds\, dv \\ &= c \int_0^\infty \int_0^\infty s^{\frac{p}{2}-\beta-2+i} e^{-\frac{s}{2v}} v^{\beta-\frac{p}{2}+1} g(v)\, ds\, dv \\ &= c\, 2^{p/2-\beta-1+i} \Gamma(p/2-\beta-1+i) E[V^i]\,.\end{aligned}$$

Finally (2.6.15) gives :

$$\lim_{w\to\infty} r(w) = 2\left(\frac{p}{2}-\beta-1\right) E[V]$$

which is the desired result. □

Chapitre 3

Estimation avec vecteur résiduel sous symétrie sphérique

Generalized Bayes minimax estimators of location vectors for spherically symmetric distributions with residual vector

Dominique Fourdrinier [1], **Othmane Kortbi** [2]
and William E. Strawderman [3]

Abstract

We consider estimation of the mean vector, θ, of a spherically symmetric distribution with known scale parameter under quadratic loss and when a residual vector is available. We show minimaxity of generalized Bayes estimators corresponding to superharmonic priors with a non decreasing Laplacian of the form $\pi(\|\theta\|^2)$, under certain conditions on the generating function $f(\cdot)$ of the sampling distributions. The class of sampling distributions includes certain variance mixtures of normals.

AMS 1991 subject classifications. Primary 62C20, 62C15, 62C10.

Keywords and phrases : Spherical symmetry, Bayes estimators, Minimax estimators, Quadratic loss.

3.1 Introduction

This paper is devoted to the study of minimaxity of generalized Bayes estimators, under quadratic loss, of the location parameter of a spherically symmetric distribution. More specifically, we consider the model where (X, U) is

1. Université de Rouen, LITIS, EA 4108, France. *The support of the ANR grant 08-EMER-002 is gratefully acknowledged.*

2. Université de Sherbrooke, Département de mathématiques, Canada.

3. Rutgers University, Department of Statistics, USA.

a $p+k$ random vector having a spherically symmetric distribution around the $p+k$ vector $(\theta, 0)$ with θ being the unknown parameter of interest. Here dim X= dim $\theta = p$ and dim U = dim $0 = k$. Suppose the density function of (X, U) is of the form :

$$f\left(\|x-\theta\|^2+\|u\|^2\right) , \tag{3.1.1}$$

where $f(\cdot)$ is known, and consider generalized Bayes estimators of θ for priors of the form :

$$\pi\left(\|\theta\|^2\right) , \tag{3.1.2}$$

under the quadratic loss function :

$$L(\theta, \delta) = \|\delta - \theta\|^2 . \tag{3.1.3}$$

The fundamental tool used in this work is the development of an expression for the risk based on a differential expression involving the marginal which has become a basic tool to prove minimaxity. This differential expression has been extended to non normal models such as (3.1.1) by several authors (see Fourdrinier *et al.* [FSW03, FS08b, FS08a, FSW06], Maruyama [Mar03a]). The main difference between the setup of this paper and the others is that here $f(\cdot)$ is known and so the scale parameter $\sigma^2 = E\|X-\theta\|^2/p$ is fixed and known. However the residual vector U is also available. Because of this, results herein are intermediate between the previous results ; however distributional robustness results for generalized Bayes minimax estimators falls somewhat closer to those for a known scale without a residual vector.

The general line of research pertinent to this work is the development of (generalized) Bayes minimax estimators. In the case of normal distribution with known scale, see for example Fourdrinier et al. [FSW98]. When the scale is unknown, see Strawderman [Str73]. For variance mixture of normals with known scale and no residual vector, see Strawderman [Str74], Fourdrinier *et al.* [FKS08], and Maruyama [Mar03a]. For general spherically symmetric distributions, with no residual vector see Fourdrinier *et al.* [FS08a] and for spherically symmetric distributions with residual vector and unknown scale, see Fourdrinier *et al.* [FS10], and Maruyama [Mar03b].

In this paper, we establish minimaxity of generalized Bayes estimators for broad classes of spherically symmetric distributions which are not restricted to variance mixtures of normals. Minimaxity results are obtained for unimodal spherically symmetric superharmonic priors of the form (3.1.2) under the additional condition of monotonicity of the Laplacian of the prior. This class contains the fundamental harmonic prior $||\theta||^{2-p}$ and, more generally, priors of the form $(a + ||\theta||^2)^{-b}$ for $0 < b < (p-2)/2$ and $a \geq 0$.

Conditions on the sampling density are that $\int_s^\infty f(t)\, dt / 2\, f(s)$ is bounded below by a certain positive function $c(s)$, that $f'(t)/f(t)$ is non decreasing, and that

$$\int_0^\infty f(v+s)\, v^{p/2}\, dv \leq 4\, c(s) \int_0^\infty -f'(v+s)\, v^{p/2}\, dv < \infty$$

for all $s > 0$. This class includes variance mixtures of normal distributions satisfying mild conditions on the mixing density.

In Section 3.2, we develop the model, give preliminary risk calculations and we establish the main result. Section 3.3 gives examples to illustrate the theory. In Section 3.4, we give some concluding remarks.

3.2 Main result

3.2.1 Risk considerations

Any estimator $\delta(X, U)$ of θ is evaluated by its risk function associated to the loss (3.1.3), that is by :

$$R(\theta, \delta) = E_\theta\left[\|\delta(X, U) - \theta\|^2\right] \tag{3.2.1}$$

where E_θ denotes the expectation with respect to the density (3.1.1). For the rest of this paper, we assume that :

$$E_\theta\left[\|X - \theta\|^2\right] < \infty \tag{3.2.2}$$

which guarantees that the standard estimator X has finite risk and is minimax. As any estimator δ can be written as $\delta(X, U) = X + g(X, U)$, the finiteness of its risk is guaranteed by :

$$E_\theta\left[\|g(X, U)\|^2\right] < \infty\,, \tag{3.2.3}$$

and δ will be itself minimax provided the difference in risk

$$\begin{aligned}\Delta_\theta(\delta) &= E_\theta\left[\|\delta(X, U) - \theta\|^2 - \|X - \theta\|^2\right] \\ &= E_\theta\left[2(X - \theta)' g(X, U) + \|g(X, U)\|^2\right] \end{aligned} \tag{3.2.4}$$

is non positive.

To express the risk difference between δ and X, we introduce first the function F defined, for any $t > 0$. by :

$$F(t) = \frac{1}{2}\int_t^\infty f(l)\, dl\,. \tag{3.2.5}$$

We have the following lemma whose proof is similar to the version found in Fourdrinier *et al.* [FSW03].

Lemma 3.2.1. *Let g be a weakly differentiable function from $\mathbb{R}^{p+k}$ into $\mathbb{R}^p$. Then*

$$E_\theta\left[(X-\theta)'\,g(X,U)\right] = E_\theta\left[\frac{F\left(\|X-\theta\|^2+\|U\|^2\right)}{f\left(\|X-\theta\|^2+\|U\|^2\right)}\ div_X\left(g(X,U)\right)\right] \tag{3.2.6}$$

provided these expectations exist.

The next result follows immediately from the Lemma 3.2.1.

Lemma 3.2.2. *Assume that $E_\theta\left[\|g(X,U)\|^2\right] < \infty$. The risk difference in (3.2.4) equals*

$$\Delta_\theta(\delta) = E_\theta\left[2\frac{F\left(\|X-\theta\|^2+\|U\|^2\right)}{f\left(\|X-\theta\|^2+\|U\|^2\right)}\ div_X\left(g(X,U)\right) + \|g(X,U)\|^2\right]. \tag{3.2.7}$$

A sufficient condition for minimaxity of estimators δ with finite risk is given by the following lemma when $p \geq 3$. For that, let use the notation $s = \|u\|^2$.

Lemma 3.2.3. *Assume that F/f is bounded below by a positive function $c(s)$ and that $E_\theta\left[\|g(X,U)\|^2\right] < \infty$ for all $\theta \in \mathbb{R}^p$. Then the estimator $\delta(X,U)$ is minimax provided that :*

$$2c(s)\ div_x\left(g(x,u)\right) + \|g(x,u)\|^2 \leq 0 \tag{3.2.8}$$

for all $x \in \mathbb{R}^p$ and all $u \in \mathbb{R}^k$.

The proof follows immediately from Lemma 3.2.2.

3.2.2 Form of the Bayes Estimators

The Bayes estimator under loss (3.1.3), for the priors given by (3.1.2) is of the form :

$$\delta_\pi (X,U) = X + \frac{\int_{\mathbb{R}^p} (\theta - X) f \left(\|X-\theta\|^2 + \|U\|^2\right) \pi \left(\|\theta\|^2\right) \, d\theta}{\int_{\mathbb{R}^p} f \left(\|X-\theta\|^2 + \|U\|^2\right) \pi \left(\|\theta\|^2\right) \, d\theta} . \tag{3.2.9}$$

Now, define for any $(x,u) \in \mathbb{R}^p \times \mathbb{R}^k$,

$$M(x,u) = \int_{\mathbb{R}^p} F\left(\|x-\theta\|^2 + \|u\|^2\right) \pi \left(\|\theta\|^2\right) \, d\theta \tag{3.2.10}$$

and

$$m(x,u) = \int_{\mathbb{R}^p} f\left(\|x-\theta\|^2 + \|u\|^2\right) \pi \left(\|\theta\|^2\right) \, d\theta \tag{3.2.11}$$

Then the generalized Bayes estimator δ_π defined above can be written as :

$$\delta_\pi (X,U) = X + \frac{\nabla_X M(X,U)}{m(X,U)} \tag{3.2.12}$$

where ∇_X denote the gradient with respect of the variable X.

Applying Lemma 3.2.3 to the Bayes estimator δ_π gives the following result.

Lemma 3.2.4. *Assume that F/f is bounded below by a positive function $c(s)$ and that $E_\theta \left[\| \nabla_X M(X,U)/m(X,U) \|^2 \right] < \infty$ for all $\theta \in \mathbb{R}^p$. Then the estimator $\delta_\pi (X,U)$ in (3.2.12) is minimax provided that :*

$$2c(s) \, \frac{\Delta_x M(x,u)}{m(x,u)} - 2c(s) \, \frac{\nabla_x M(x,u)' \, \nabla_x m(x,u)}{m^2(x,u)} + \frac{\|\nabla_x M(x,u)\|^2}{m^2(x,u)} \leq 0 \tag{3.2.13}$$

for all $x \in \mathbb{R}^p$ and all $u \in \mathbb{R}^k$.

Proof. The lemma follows from the expression of $div_x(\nabla_x M(x,u)/m(x,u))$, where we denote by Δ_x the Laplacian with respect to the variable x. $\square$

Note that it follows from the spherical symmetry of f and π that, the functions $m\,(x,u)$ and $M\,(x,u)$ depend on (x,u) through $t = \|x\|^2$ and s. Then, we write (with a slight abuse of notation) $m\,(x,u) = m\,(t,s)$ and $M\,(x,u) = M\,(t,s)$ so that $\nabla_x m(x,u) = 2\,m'(t,s)\,x$ and $\nabla_x M(x,u) = 2\,M'(t,s)\,x$.

3.2.3 Minimaxity of generalized Bayes Estimators

First, we will need the following result which essentially gathers results in Lemma 3.1 - 3.3 of Fourdrinier and Strawderman [FS08a].

Lemma 3.2.5. *Let $m\,(x,u)$ and $M\,(x,u)$ are as defined in (3.2.11) and (3.2.10) and let $\cdot$ denote the inner product in $\mathbb{R}^p$. Then we have :*

1.
$$x \cdot \nabla_x m\,(x,u) = -2\int_0^\infty H\left(v, \|x\|^2\right) v^{p/2}\, f'(v+s)\, dv$$

and

$$x \cdot \nabla_x M\,(x,u) = \int_0^\infty H\left(v, \|x\|^2\right) v^{p/2} f\,(v+s)\; dv$$

where, for $v > 0$,

$$H\left(v, \|x\|^2\right) = \lambda\,(B) \int_{B_{\sqrt{v},x}} x \cdot \theta\, \pi'(||\theta||^2)\, dV_{\sqrt{v},x}\,(\theta) \tag{3.2.14}$$

and $V_{\sqrt{v},x}$ is the uniform distribution on the ball $B_{\sqrt{v},x}$ of radius $\sqrt{v}$ centered at x and $\lambda\,(B)$ is the volume of the unit ball.

2. *For any $x \in \mathbb{R}^p$, the function $H\left(v, \|x\|^2\right)$ in (3.2.14) is non decreasing in v provided that $\Delta\pi(||\theta||^2)$ is non decreasing in $\|\theta\|^2$.*

3. *For any $v > 0$ and any $x \in \mathbb{R}^p$, the function $H\left(v, \|x\|^2\right)$ in (3.2.14) is non positive provided that $\pi'\left(\|\theta\|^2\right) \leq 0$.*

Now, our main result is given in the following theorem.

Theorem 3.2.1. *Suppose $\theta \in \mathbb{R}^p$ has a superharmonic prior $\pi(\|\theta\|^2)$, where $\pi(\|\theta\|^2)$ is non increasing and $\Delta\pi(\|\theta\|^2)$ is non decreasing in $\|\theta\|^2$. Then the Bayes estimator δ_π is minimax under quadratic loss (3.1.3) provided that $\frac{f'}{f}$ is non decreasing, $\frac{F(v+s)}{f(v+s)}$ bounded below by a positive function $c(s)$ and for all $s > 0$*

$$\int_0^\infty f\,(v+s)\, v^{p/2}\, dv \leq 4c\,(s) \int_0^\infty -f'\,(v+s)\, v^{p/2}\, dv < \infty\,. \qquad (3.2.15)$$

Proof. Due to the superharmonicity of $\pi(\|\theta\|^2)$, for any $(x, u) \in \mathbb{R}^p \times \mathbb{R}^k$, we have $\Delta_x M(x, u) \leq 0$ so that, by Lemma 3.2.4, it suffices to prove that

$$-2c(s)\ \nabla_x M(x, u)'\nabla_x m(x, u) + \|\nabla_x M(x, u)\|^2 \leq 0 \qquad (3.2.16)$$

for all x and all u. Using notation for $m\,(x, u)$ and $M\,(x, u)$ defined above, (3.2.16) reduces to

$$-2c(s)\ M'(t, s)\, m'(t, s) + (M'(t, s))^2 \leq 0\,. \qquad (3.2.17)$$

Now, since we have supposed that $\pi(\|\theta\|^2)$ is non increasing in $\|\theta\|^2$, Lemma 3.2.5 implies that $M'(t, s) \leq 0$. So, Inequality (3.2.17) reduces to

$$-2\,c(s)\, m'(t, s) + M'(t, s) \geq 0\,, \qquad (3.2.18)$$

which can be written as

$$-2\,c(s)\, x \cdot \nabla_x m(x, u) + x \cdot \nabla_x M(x, u) \geq 0\,. \qquad (3.2.19)$$

Using again Lemma 3.2.5, the left hand side of the last inequality may be expressed as

$$\begin{aligned} 4\,c(s) \int_0^\infty H(v, \|x\|^2)\, v^{p/2} f'(v+s)\, dv + \int_0^\infty H(v, \|x\|^2)\, v^{p/2} f(v+s)\, dv \\ = 4\,c(s)\, E\Big[H(v, \|x\|^2)\, \frac{f'(v+s)}{f(v+s)}\Big] + E\big[H(v, \|x\|^2)\big] \quad (3.2.20) \end{aligned}$$

where E denotes the expectation with respect to the density proportional to $v^{p/2} f\,(v+s)$. As the function $\frac{f'(v+s)}{f(v+s)}$ is assumed to be non decreasing in v and as $H(v, \|x\|^2)$ is also non decreasing in v according to Lemma 3.2.5, since $\Delta\pi(\|\theta\|^2)$ is assumed to be non decreasing in $\|\theta\|^2$, the first expectation in (3.2.20) satisfies

$$E\left[H(v, \|x\|^2) \frac{f'(v+s)}{f(v+s)}\right] \geq E\left[H(v, \|x\|^2)\right] E\left[\frac{f'(v+s)}{f(v+s)}\right]$$

by the covariance inequality. Therefore Inequality (3.2.19) will be satisfied as soon as

$$4c(s) E\left[\frac{f'(v+s)}{f(v+s)}\right] + 1 \leq 0 \qquad (3.2.21)$$

since $H\left(v, \|x\|^2\right) \leq 0$ by Lemma 3.2.5. Finally, Inequality (3.2.21) is equivalent to the condition given in the theorem. □

It is worth noting that, since $f'(v+s)/f(v+s)$ is non decreasing, it follows that $F(v+s)/f(v+s)$ is non decreasing as well and hence the lower bound function $c(s)$ can be replaced by $F(s)/f(s)$.

3.3 Examples

3.3.1 Examples of sampling distributions

Example (S1).

Assume that

$$f(v+s) = K\left(\frac{1}{v+s+a}\right)^b, \tag{3.3.1}$$

where $b > 1$, $a > 0$ and K is the normalizing constant. Note that :

$$\frac{f'(v+s)}{f(v+s)} = \left(\frac{-b}{v+s+a}\right) \tag{3.3.2}$$

is non decreasing in v for all value of $b > 0$. For the lower bound function, as noted above, we can take :

$$c(s) = \frac{F(s)}{f(s)} = \frac{s+a}{2(b-1)}. \tag{3.3.3}$$

Also, we have :

$$\begin{aligned} E\left[\frac{f'(v+s)}{f(v+s)}\right] &= -b\,\frac{\int_0^\infty v^{p/2}\left(\frac{1}{v+s+a}\right)^{b+1} dv}{\int_0^\infty v^{p/2}\left(\frac{1}{v+s+a}\right)^{b} dv} \\ &= \frac{-b}{s+a}\,\frac{\int_0^\infty z^{p/2}\left(\frac{1}{z+1}\right)^{b+1} dz}{\int_0^\infty z^{p/2}\left(\frac{1}{z+1}\right)^{b} dz} \\ &= \frac{-b}{s+a}\,\frac{B\left(p/2+1, b-p/2\right)}{B\left(p/2+1, b-p/2-1\right)} \end{aligned} \tag{3.3.4}$$

where $B(\alpha, \beta)$ is the beta function with parameters $\alpha > 0$ and $\beta > 0$. Then (3.3.4) becomes :

$$\begin{aligned} E\left[\frac{f'(v+s)}{f(v+s)}\right] &= \frac{-b}{s+a}\,\frac{\Gamma(b-p/2)}{\Gamma(b+1)}\,\frac{\Gamma(b)}{\Gamma(b-p/2-1)} \\ &= -\frac{b-p/2-1}{s+a}. \end{aligned} \tag{3.3.5}$$

So, the last condition in the theorem becomes :

$$2\frac{b-p/2-1}{b-1} \geq 1 \tag{3.3.6}$$

which is satisfied for $p \leq b-1$.

Example (S2). Variance mixtures of normals.

Assume that

$$f(v+s) = K\int_0^\infty w^{-(p+k)/2}\exp\left(-\frac{v+s}{2w}\right)h(w)\,dw\,, \tag{3.3.7}$$

where h is a mixing density and K is the normalizing constant. Then, we have :

$$\begin{aligned} \frac{f'(v+s)}{f(v+s)} &= \frac{-\frac{1}{2}\int_0^\infty w^{-(p+k)/2-1}\exp\left(-\frac{v+s}{2w}\right)h(w)\,dw}{\int_0^\infty w^{-(p+k)/2}\exp\left(-\frac{v+s}{2w}\right)h(w)\,dw} \\ &= -\frac{1}{2}E_v\left[W^{-1}\right], \end{aligned} \tag{3.3.8}$$

where E_v denotes the expectation with respect to the density $f_v(w) \propto w^{-(p+k)/2}$ $\exp\left(-\frac{v+s}{2w}\right)h(w)$ which has increasing monotone likelihood ratio in v. Since $-w^{-1}$ is increasing with respect to w, the monotonicity of $\frac{f'(v+s)}{f(v+s)}$ follows. Again, for the lower bound function $c(s)$ we take :

$$\begin{aligned} \cdot\, c(s) = \frac{F(s)}{f(s)} &= \frac{\int_0^\infty w^{-(p+k)/2+1}\exp\left(-\frac{s}{2w}\right)h(w)\,dw}{\int_0^\infty w^{-(p+k)/2}\exp\left(-\frac{s}{2w}\right)h(w)\,dw} \\ &= \frac{E_s\left[W^{-p/2+1}\right]}{E_s\left[W^{-p/2}\right]}, \end{aligned} \tag{3.3.9}$$

where E_s denotes the expectation with respect to the density $f_s(w) \propto w^{-k/2}$ $\exp\left(-\frac{s}{2w}\right)h(w)$. Now, to verify the last condition of the theorem, we use Fubini's theorem to obtain :

$$\begin{aligned} \int_0^\infty v^{p/2}f(v+s)\,dv = K\int_0^\infty\int_0^\infty v^{p/2}\exp\left(-\frac{v}{2w}\right)\,dv \\ \times\, w^{-(p+k)/2}\exp\left(-\frac{s}{2w}\right)h(w)\,dw \\ = K2^{p/2+1}\Gamma\left(p/2+1\right)\int_0^\infty w^{-k/2+1}\exp\left(-\frac{s}{2w}\right)h(w)\,dw\,. \end{aligned}$$

Similarly,

$$\int_0^\infty v^{p/2} f'(v+s)\, dv = -\frac{K}{2} \int_0^\infty \int_0^\infty v^{p/2} \exp\left(-\frac{v}{2w}\right) dv$$
$$\times\, w^{-(p+k)-1/2} \exp\left(-\frac{s}{2w}\right) h(w)\, dw$$
$$= -K 2^{p/2} \Gamma\left(p/2+1\right) \int_0^\infty w^{-k/2} \exp\left(-\frac{s}{2w}\right) h(w)\, dw\,.$$

Then, we have :

$$E\left[\frac{f'(v+s)}{f(v+s)}\right] = -\frac{1}{2} \frac{\int_0^\infty w^{-k/2} \exp\left(-\frac{s}{2w}\right) h(w)\, dw}{\int_0^\infty w^{-k/2+1} \exp\left(-\frac{s}{2w}\right) h(w)\, dw} = -\frac{1}{2E_s\left[W\right]}\,, \tag{3.3.10}$$

and so the last condition in the theorem reduces to

$$g(s) = \frac{E_s\left[W\right] E_s\left[W^{-p/2}\right]}{E_s\left[W^{-p/2+1}\right]} \leq 2\,, \tag{3.3.11}$$

for all $s > 0$. Also, we have the next result.

Lemma 3.3.1. *Assume that the mixing density h of the sampling distribution in (3.3.7) is such that :*

$$-t\frac{h'(t)}{h(t)} \leq \alpha \tag{3.3.12}$$

for some $\alpha \geq 2 - (p+k)/2$. Assume also that the mixing density h satisfies

$$\lim_{t\to\infty} t^\beta h(t) = c \tag{3.3.13}$$

for some $\beta > 2 - k/2$ and some constant $c \neq 0$. Then

$$g(s) \leq \frac{p+k+2\alpha-4}{k+2\beta-4}\,, \tag{3.3.14}$$

provided there exist $K > 0$ such that $t^\beta h(t) \leq K$ for all $t > 0$.

Proof. Let, for any β and for any $s > 0$ $I_\beta(s)$ be equal to :

$$I_\beta(s) = \int_0^\infty w^{\beta - k/2} \exp\left(-\frac{s}{2w}\right) h(w)\, dw\,, \tag{3.3.15}$$

then we can see that $g(s) = \frac{I_1(s)}{I_0(s)} \frac{I_{(-p/2)}(s)}{I_{(-p/2+1)}(s)}$. Now, using integration by parts we can see that :

$$\begin{aligned} I_{-p/2}(s) = \frac{2}{s}\left(\frac{p+k}{2} - 2\right) \int_0^\infty t^{(p+k)/2-3} \exp\left(-\frac{st}{2}\right) h\left(\frac{1}{t}\right) dt \\ - \frac{2}{s} \int_0^\infty t^{(p+k)/2-4} \exp\left(-\frac{st}{2}\right) h'\left(\frac{1}{t}\right) dt\,. \end{aligned}$$

So, using condition (3.3.12) we can see that :

$$\frac{I_{-p/2}(s)}{I_{-p/2+1}(s)} \le \frac{1}{s}\left(p + k + 2\alpha - 4\right)\,.$$

Also, we can write :

$$\frac{I_1(s)}{I_0(s)} = s \frac{\int_0^\infty t^{k/2-3} \exp\left(-\frac{t}{2}\right) h\left(\frac{s}{t}\right) dt}{\int_0^\infty t^{k/2-2} \exp\left(-\frac{t}{2}\right) h\left(\frac{s}{t}\right) dt}\,,$$

which imply in turn that

$$\begin{aligned} g(s) &\le \frac{1}{s}\left(p + k + 2\alpha - 4\right) \times s\, \frac{\int_0^\infty t^{k/2-3} \exp\left(-\frac{t}{2}\right) h\left(\frac{s}{t}\right) dt}{\int_0^\infty t^{k/2-2} \exp\left(-\frac{t}{2}\right) h\left(\frac{s}{t}\right) dt} \\ &\le \left(p + k + 2\alpha - 4\right) \lim_{s\to\infty} \frac{\int_0^\infty t^{k/2-3} \exp\left(-\frac{t}{2}\right) h\left(\frac{s}{t}\right) dt}{\int_0^\infty t^{k/2-2} \exp\left(-\frac{t}{2}\right) h\left(\frac{s}{t}\right) dt}\,, \end{aligned} \tag{3.3.16}$$

since $I_1(s)/I_0(s)$ is increasing in s. Now the condition $t^\beta h(t) \le K$ for all $t > 0$, allows us to apply the Lebesgue dominated convergence theorem. Then, according to (3.3.13), we have :

$$\begin{aligned} \lim_{s\to\infty} \frac{\int_0^\infty t^{k/2-3} \exp\left(-\frac{t}{2}\right) h\left(\frac{s}{t}\right) dt}{\int_0^\infty t^{k/2-2} \exp\left(-\frac{t}{2}\right) h\left(\frac{s}{t}\right) dt} &= \frac{\int_0^\infty t^{k/2+\beta-3} \exp\left(-\frac{t}{2}\right) \lim_{s\to\infty} \left(\frac{s}{t}\right)^\beta h\left(\frac{s}{t}\right) dt}{\int_0^\infty t^{k/2+\beta-2} \exp\left(-\frac{t}{2}\right) \lim_{s\to\infty} \left(\frac{s}{t}\right)^\beta h\left(\frac{s}{t}\right) dt} \\ &= \frac{1}{2} \frac{\Gamma\left(k/2 + \beta - 2\right)}{\Gamma\left(k/2 + \beta - 1\right)} \\ &= \frac{1}{k + 2\beta - 4}\,. \end{aligned} \tag{3.3.17}$$

Combining (3.3.16) and (3.3.17), we have (3.3.14). $\square$

Finally, according to Lemma 3.3.1, Inequality 3.3.11 reduces to

$$\frac{p+k+2\alpha-4}{k+2\beta-4} \leq 2\,. \tag{3.3.18}$$

Therefore, we can give general conditions for Bayes estimators to be minimax in this case of sampling distributions.

Corollary 3.3.1. *Assume that a prior π satisfies conditions of the Theorem 3.2.1. Assume also that f is given by (3 3.7) such that $E_0\left[W^{-p/2}\right] < \infty$ and that the mixing density h satisfy (3.3.12) and (3.3.13) Then the corresponding generalized Bayes estimator is minimax provided that (3.3.18) is satisfied.*

Example (S3). Generalized multivariate Student distribution.

We consider the particular case of the inverse gamma mixing, when $1/W \sim \text{Gamma}(\alpha, \beta)$ with $\alpha > 0$ and $\beta > 0$. Thus, the mixing density is given by :

$$h\,(w) \propto w^{-\alpha-1} \exp\left(-\beta/w\right)\,,$$

which in turn implies that

$$f\,(v+s) \propto \left(\frac{1}{v+s+2\beta}\right)^{(p+k)/2+\alpha}\,.$$

Note that, the case of $a = 2\beta$ and $b = \frac{p+k}{2} + \alpha$, corresponds to the Example **(S1)**. Also, note that the case of $\alpha = \beta = n/2$, corresponds to the multivariate t with n degrees of freedom.

Now to prove minimaxity, we have to apply Corollary 3.3.1. For that, we have for $k/2 + m > -\alpha$ and for any $s > 0$

$$E_s\left[W^{-m}\right] = \frac{2^m}{(s+2\beta)^m} \frac{\Gamma(k/2+m+\alpha)}{\Gamma(k/2+\alpha)} < \infty\,.$$

Also, we can see that $-t\frac{h'(t)}{h(t)} \leq \alpha + 1$ and $\lim_{t\to\infty} t^{\alpha+1}h(t) = c$, where c is the normalizing constante of h. Then, the condition (3.3.18) is satisfied if $p \leq k + 2\alpha - 2$. For the multivariate t distribution with n degrees of freedom, the condition (3.3.18) is satisfied in the cases when $n \geq p - k + 2$.

Example (S4). Kotz distribution.

In this example, we consider the class of spherical distributions with generating function f given by :

$$f(v+s) = K(v+s)^m \exp\left(-\frac{v+s}{2}\right), \tag{3.3.19}$$

where $m > -p/2$ and K is the constant of normalisation.

Note that, for the Kotz distribution, when $m > 0$, it is not a variance mixture of normals, while, when $m < 0$, it is (the mixing distribution is the $\beta(p/2 + m, -m)$ distribution). However, in the last case, Condition (3.3.13) cannot be satisfied since the beta distribution has compact support. So, Corollary 3.3.1 cannot be applied.

But, we can see that

$$\frac{f'(v+s)}{f(v+s)} = \frac{m}{v+s} - \frac{1}{2}, \tag{3.3.20}$$

which is non increasing for $m > 0$ and non decreasing for $m < 0$ with respect to v. Thus, we can apply the Theorem 3.2.1 only in the case $m < 0$. In this case, we can see that

$$\frac{F(s)}{f(s)} = \frac{s}{2}\int_0^\infty (1+u)^m \exp(-us/2)\, du. \tag{3.3.21}$$

In order to determine the conditions for which Inequality (3.2.21) is satisfied, we use integration by parts to obtain :

$$\int_0^\infty \left(\frac{m}{v+s} - \frac{1}{2}\right) v^{p/2} (v+s)^m \exp\left(-(v+s)/2\right)\, du$$
$$= -\frac{p}{2} \int_0^\infty v^{p/2-1} (v+s)^m \exp\left(-(v+s)/2\right)\, du$$
$$= -\frac{p}{2} s^{p/2+m} \int_0^\infty u^{p/2-1} (1+u)^m \exp\left(-us/2\right) \exp\left(-s/2\right)\, du\,.$$

Also, we have :

$$\int_0^\infty v^{p/2} (v+s)^m \exp\left(-(v+s)/2\right)\, du$$
$$= s^{p/2+m+1} \int_0^\infty u^{p/2} (1+u)^m \exp\left(-us/2\right) \exp\left(-s/2\right)\, du\,,$$

which in turn implies

$$E\left[\frac{f'(v+s)}{f(v+s)}\right] = -\frac{p}{2s} \frac{\int_0^\infty u^{p/2-1} (1+u)^m \exp\left(-us/2\right)\, du}{\int_0^\infty u^{p/2} (1+u)^m \exp\left(-us/2\right)\, du}\,.$$

Then, Inequality (3.2.21) reduces to

$$\frac{J_0(s) J_{p/2-1}(s)}{J_{p/2}(s)} \geq \frac{1}{p} \tag{3.3.22}$$

for all $s > 0$, where $J_\beta(s) = \int_0^\infty u^\beta (1+u)^m \exp\left(-us/2\right)\, du$. This last integral can be written as :

$$J_\beta(s) = \Gamma(\beta+1) U(\beta+1, m+\beta+2, s/2) \tag{3.3.23}$$

(see [AS64], page 504-515) where U is the Kummer's function of the second kind. Now, we will use the following recurrence relation for the function U,

$$(b-a)U(a,b,z) + U(a-1,b,z) = zU(a,b+1,z)$$

for the case $b - a \leq 0$, which imply that $U(a-1,b,z) \geq zU(a,b+1,z)$. In our setting, $b - a = m + 1$. So if we assume that $m + 1 \leq 0$, then

$$\frac{J_{p/2-1}(s)}{J_{p/2}(s)} = \frac{\Gamma(\frac{p}{2})\; U(\frac{p}{2}, \frac{p}{2}+m+1, \frac{s}{2})}{\Gamma(\frac{p}{2}+1)\; U(\frac{p}{2}+1, \frac{p}{2}+m+2, \frac{s}{2})} \geq \frac{s}{p}\,.$$

On the other hand, $J_0(s)$ is a decreasing function with respect to s and for large s it can be written as $\frac{2}{s}\left(1+O\left(\frac{1}{s}\right)\right)$, so we have for all $s>0$:

$$J_0(s)\frac{J_{p/2-1}(s)}{J_{p/2}(s)} \geq \frac{2}{p}\left(1+O\left(\frac{1}{s}\right)\right) \geq \frac{2}{p}$$

Then, the condition 3.2.21 is satisfied for all $-p/2 < m \leq -1$.

Example (S5).

We consider the class of spherical distributions with generating function f given by :

$$f\left(v+s\right) = K\exp\left(-\alpha\left(v+s\right)+\beta\varphi\left(v+s\right)\right)\,, \qquad (3.3.24)$$

where $\alpha>0$, $\beta>0$, $\varphi(0)<\infty$, $\varphi'(t)\geq -M$, where $M>0$ and K is the constant of normalisation. Also, we assume that φ is a positive, non increasing and convex function.

Clearly $f'(v+s)/f(v+s) = -\alpha+\beta\varphi'(v+s)$ is non decreasing in v. Note that, $\int_0^\infty f\left(v+s\right)v^{p/2}\,dv<\infty$ and also $\int_0^\infty -f'\left(v+s\right)v^{p/2}\,dv<\infty$ since $|\varphi'|$ is bounded. Finally, condition (3.2.21) becomes :

$$\begin{aligned}\int_0^\infty f\left(v+s\right)v^{p/2}\,dv \leq 4\alpha c(s)&\int_0^\infty f\left(v+s\right)v^{p/2}\,dv\\ &+4\beta c(s)\int_0^\infty -\varphi'(v+s)f\left(v+s\right)v^{p/2}\,dv\,,\end{aligned}$$

which is satisfied if we have $4\alpha c(s)\geq 1$ for all $s>0$. Then, as we show below, the generalized Bayes estimator is minimax for any $\alpha>0$ and for any function φ satisfying the above conditions with $\beta\leq\alpha\,\frac{\log(2)}{M}$. Indeed, we have for all $s>0$:

$$c(s) = \frac{F\left(s\right)}{f\left(s\right)} = \frac{1}{2\alpha\exp\left(\beta\varphi(s)\right)}E_\alpha\left[\exp\left(\beta\varphi(u+s)\right)\right]\,,$$

where E_α denotes the expectation with respect to the density $\alpha \exp(-\alpha u)$. As the function $\exp(\beta\varphi(u+s))$ is convex in u, we have by Jensen inequality :

$$c(s) \geq \frac{1}{2\alpha \exp(\beta\varphi(s))} \exp\left(\beta\varphi(\frac{1}{\alpha}+s)\right).$$

Hence $4\alpha c(s) \geq 1$ for all $s > 0$ as soon as

$$2\exp\left(\beta\left(\varphi(\frac{1}{\alpha}+s) - \varphi(s)\right)\right) \geq 1$$

which is equivalent to

$$\beta \leq \frac{\log(2)}{\varphi(s) - \varphi(\frac{1}{\alpha}+s)} \tag{3.3.25}$$

all $s > 0$. Using the mean value theorem, we have $\varphi(s) - \varphi(\frac{1}{\alpha}+s) = -\varphi'(k_s)\frac{1}{\alpha} \leq M\frac{1}{\alpha}$ by assumption on φ. Then, $\beta \leq \alpha\ \frac{\log(2)}{M}$ implies (3.3.25) for all $s > 0$.

3.3.2 Examples of priors

Several examples of priors which satisfy the conditions of Theorem 3.2.1 are given in Fourdrinier and Strawderman [FS08a]. Thus, as a first example, these conditions are satisfied by the following priors $\pi(\|\theta\|^2) = (A + \|\theta\|^2)^{-k}$ when $A \geq 0$ and $0 \leq k \leq p/2 - 1$. Note that the case where $A = 0$ and $k = p/2 - 1$ corresponds to the fundamental harmonic prior. Note also that this class of priors arises as a scale mixture of normals.

More generally, the mixture of normals with respect to the inverse of the variance

$$\pi(\|\theta\|^2) = \int_0^\infty \left(\frac{u}{2\pi}\right)^{p/2} \exp\left(-\frac{u}{2}\|\theta\|^2\right) h(u)\, du$$

satisfy the required conditions as soon as

$$u\frac{h'(u)}{h(u)} \leq -2\,.$$

Other examples can be found in [FS08a] including priors which are non necessarily scale mixture of normals.

3.4 Concluding Remarks

In this work, we have proved minimaxity of generalized Bayes estimators corresponding to superharmonic priors of the form $\pi(\|\theta\|)^2$ with non decreasing Laplacian, for a broad class of spherically symmetric distributions when a residual vector is present. This subclass contains certain variance mixtures of normal distributions.

In this paper and in Fourdrinier and Strawderman [FS08a] the minimax generalized Bayes estimators depend on the form of the underlying density. On the other hand, when the scale is unknown and a residual vector is available, the form of the generalized Bayes estimators is independent of the form of the density provided the prior distribution has a certain form, see Fourdrinier and Strawderman [FS10] and Maruyama [Mar03b]. For the model of the current paper however, the form of the Bayes estimator generally depends on the density $f(\cdot)$. In terms of distributional robustness of the generalized Bayes minimax estimators, the results of this paper fall closer to those in [FS08a] then to those in [FS10] and [Mar03b]. In [FS08a] the scale parameter is known and there is no residual vector, wheras in [FS10] and [Mar03b] the scale parameter is unknown but a residual vector is available. Here, of course, we are treating the intermediate case of a known scale parameter, but with the supplemental sample information of a residual vector.

Maruyama [Mar03a] and Fourdrinier *et al.* [FKS08] gave classes of priors including some proper priors for which the resulting Bayes estimators are minimax. In our setting, due to the superharmonicity supposition, our priors cannot be proper (see [FSW98]). Furthermore, in the case of mixtures of normal sampling distributions, our mixing distributions are not required to have monotone likelihood ratio as in the above papers.

Chapitre 4

Estimateurs linéaires tronqués d'une moyenne multidimensionnelle

Truncated linear estimation of a bounded multivariate normal mean

Othmane KORTBI and Éric MARCHAND [1]

Abstract

We consider the problem of estimating the mean θ of a $N_p(\theta, I_p)$ distribution with squared error loss and under the constraint $\|\theta\| \leq m$, for some constant $m > 0$. Using Stein's identity to obtain unbiased estimates of risk, Karlin's sign change arguments, and conditional risk analysis, we compare the risk performance of truncated linear estimators with that of the maximum likelihood estimator δ_{mle}. We obtain for fixed (m, p) sufficient conditions for dominance. An asymptotic framework is developed, where we demonstrate that the truncated linear minimax estimator dominates δ_{mle}, and where we obtain simple and accurate measures of relative improvement in risk. Numerical evaluations illustrate the effectiveness of the asymptotic framework for approximating the risks for moderate or large values of p.

AMS 2010 subject classifications. 62C20, 62C99, 62F10, 62F30, 62H12.

Keywords and phrases : Restricted parameters, point estimation, squared error loss, dominance, truncated linear estimators, truncated linear minimax, maximum likelihood, multivariate normal, asymptotic analysis.

1. Département de mathématiques, Université de Sherbrooke, Canada.

4.1 Introduction

Consider independently and normally distributed observables : $X_i \sim N(\theta_i, \sigma^2)$, $i = 1, \ldots, p$, with the additional information that $\sum_{i=1}^{p} \frac{(\theta_i - \tau_i)^2}{\sigma^2} \leq m^2$, with known $\tau_1, \ldots, \tau_p, \sigma^2, m$. From a practical point of view, a constraint as the one above signifies that the squared standardized deviations $(|\frac{\theta_i - \tau_i}{\sigma}|)$ are on average bounded by $\frac{m^2}{p}$. Inferential problems with information or constraints of the above type certainly arise in practical settings, and it is thus relevant to analyze the comparative performance of various estimators that may well vary in how efficiently they capitalize on the parametric information. For recent decision-theoretic descriptions of such restricted parameter space problems, we refer to the paper of Marchand and Strawderman [MS04], as well as the monograph of van Eeden [vE06], among others.

We proceed by setting $\sigma^2 = 1$ and $\tau_i = 0, i = 1, \ldots, p$, without loss of generality, and writing our model as $X \sim N_p(\theta, I_p)$ with $X = (X_1, \ldots, X_p)'$ and $\theta = (\theta_1, \ldots, \theta_p)' \in \Theta(m) = \{\theta \in \Re^p : \|\theta\| \leq m\}$. We are concerned here with estimating θ under quadratic loss $\|\delta - \theta\|^2$. Previous work of Marchand and Perron [MP01], as well as Fourdrinier and Marchand [FM10], focused on the determination of estimators that improve upon the benchmark, but inadmissible, maximum likelihood estimator given by $\delta_{\text{mle}}(x) = (m \wedge \|x\|)\frac{x}{\|x\|}$. Here is a key result synthesized from [FM10] and [MP01] which we will refer to.

Theorem 4.1.1. *For sufficiently small m, say $m \leq c_1(p)$, all equivariant estimators $\delta(X)$ with respect to orthogonal transformations taking values in $\Theta(m)$ with $\|\delta(x)\|$ nondecreasing in $\|x\|$, as well as all Bayes estimators δ_π with respect to an orthogonally invariant prior π supported on $\Theta(m)$ dominate*

δ_{mle}. *Furthermore, we have* $\lim_{p\to\infty} \frac{c_1(p)}{\sqrt{p}} = \sqrt{\frac{1}{3}}$.

We refer to [MP01] and [FM10] for various other dominance results, such as those pertaining to Bayes estimators with respect to : the boundary uniform prior (δ_{BU}), uniform priors on spheres centered at the origin, the fully uniform prior on $\Theta(m)$ (δ_U), and other absolutely continuous priors on $\Theta(m)$.

Another unresolved, fundamental, and difficult issue concerns the degree to which δ_{mle} can be improved upon. There is certainly numerical evidence, provided in [MP01] for various pairs (m,p), that suggests substantial gains over a large part of parameter space for small m relative to p achieved by the Bayesian estimators δ_U and δ_{BU}, among others. As well, the universal dominance result in Theorem 4.1.1 strongly hints at large gains for quite small m. On the other hand, these gains are either much less impressive or vanish for moderate or large m relative to p. Notwithstanding the above indirect evidence, there is a lack of crisp analytical results providing bounds or good approximations for the degree or the percentage improvement.

To meet this challenge, we study the class of truncated linear estimators $\delta_a(x) = (a\,\|x\| \wedge m)\frac{x}{\|x\|}$, with $0 \le a \le 1$. The class of shrinking (with respect to δ_{mle}) estimators are of interest : their form is simple and appealing, the class contains the mle δ_1, the trivial (but Bayes) estimator δ_0, and the truncated linear minimax estimator δ_{tlx} with $a = \frac{m^2}{m^2+p}$. Furthermore, such estimators are perhaps just as, or more, plausible than δ_{mle}. For instance, for $a \in [0,1)$, δ_a truncates an admissible Bayes (with respect to the prior $\theta \sim N_p(0, (a/(1-a))I_p)$ estimator aX onto the parameter space $\Theta(m)$, while δ_{mle} truncates an inadmissible for $p \ge 3$ procedure X onto $\Theta(m)$. Also, [Mar93] provides analytical evidence limiting the loss of efficiency incurred with the linear minimax procedure, and hence with δ_{tlx}, at the boundary of $\Theta(m)$ as opposed to the optimal best equivariant estimator (for $\|\theta\| = m$) δ_{BU}. For small enough m, the universal dominance result tells us that all such δ_a's, $a \in [0, 1)$, dominate δ_{mle}, while finer analyses for $p = 1$ were carried in [MOPP08] for δ_a and δ_{tlx}, and in [DvE09] for δ_0.

We provide below further analysis for all (finite) dimensions p and we describe precisely in an asymptotic framework the relative gain or loss in efficiency in the risk comparison of δ_a and δ_{mle}. Moreover, we determine exactly which estimators δ_a improve on δ_{mle} ; and δ_{tlx} turns out to always be such an improvement. More precisely, we consider a sequence $\{(m,p) = (d\sqrt{p}\,, p), p \geq 1\}$ for fixed $d > 0$, and show that there exists, for all $d > 0$, a large enough value $p_0(d)$ and values of a (which depend on d) such that δ_a dominates δ_{mle} for all $p \geq p_0(d)$. We also study sequences $\{(m, \|\theta\|, p) = (d\sqrt{p}\,, c\sqrt{p}\,, p), p \geq 1\}$, for $c \in [0, d]$, $d > 0$, obtaining sharp asymptotic analysis concerning the relative difference in risks, and solid approximations for finite p. Observe that Theorem 4.1.1's cutoff point is proportional to $\sqrt{p}$, which translates to the average squared standardized deviation $(\frac{\theta_i - \tau_i}{\sigma})^2$ being less than or equal to $1/3$ for universal dominance. Analogously, the asymptotic setup incorporates growth conditions with m and $\|\theta\|$ proportional to $\sqrt{p}$, with findings hence being amenable to similar interpretations. We also point out as an additional motivating factor that an increasing number of modern statistical inference problems relate to large number of parameters (see for instance [DL11] and the references therein).

This paper is organized as follows. In Section 4.2, we make use of Stein's ([Ste81]) technique for estimating unbiasedly the difference in risks between

δ_a and δ_{mle} along with sign change arguments, to obtain necessary and sufficient conditions for δ_a to dominate δ_{mle}. Capitalizing on this condition (i.e., Corollary 4.2.1) and making use of a conditional (on $\|X\|$) risk decomposition introduced in [MP01], Subsection 4.3.1 includes findings for finite p. Subsection 4.3.2 exploits the asymptotic framework to produce inferences and these include a precise determination of dominating δ_a's, along exact expressions for the asymptotic relative difference in risks. Throughout, examples, numerical evaluations, and visual representations are provided. Section 4.4 contains concluding remarks.

4.2 An unbiased estimator for the difference of risks

Throughout this manuscript, we shall denote $\|X\|$, $\|x\|$ and $\|\theta\|$ by R, r and λ, respectively. An estimator $\delta(X)$ of θ is evaluated by its risk function

$$R(\theta, \delta) = E_\theta\left[\|\delta(X) - \theta\|^2\right], \theta \in \Theta(m), \tag{4.2.1}$$

where E_θ denotes the expectation with respect to $X \sim \mathcal{N}_p(\theta, I_p)$. We will make use of Stein's identity ([Ste81]) for an unbiased estimator of risk which allows us to write the risk for an estimator $\delta(X) = X + h(X)$, with h any function weakly differentiable and $E_\theta\left[h(\|X\|^2)\right] < \infty$, as

$$R(\theta, \delta) = E_\theta\left[\|\delta(X) - \theta\|^2)\right] = E_\theta\left[p + 2divh(X) + \|h(X)\|^2\right], \tag{4.2.2}$$

with div being the divergence operator. We proceed with the derivation of an unbiased estimator $D_a(\|X\|)$ of the risk difference $R(\theta, \delta_{mle}) - R(\theta, \delta_a)$, as well as a sign change analysis of $D_a(\cdot)$.

Lemma 4.2.1. **(a)** *For $0 \leq a \leq 1$, an unbiased estimator of the difference in risks $R(\theta, \delta_{mle}) - R(\theta, \delta_a)$ is given by :*

$$D_a(\|X\|) = D_{a,1}(\|X\|)\ [0 \leq \|X\| \leq m] + D_{a,2}(\|X\|)\ [m \leq \|X\| \leq m/a]\,,$$

with

$$D_{a,1}(r) = 2p - 2ap - (a-1)^2\, r^2\,,$$

and

$$D_{a,2}(r) = 2(p-1)\frac{m}{r} + (m-r)^2 - 2ap - (a-1)^2\, r^2. \qquad (4.2.3)$$

(b) *For $p \geq 1$, and $0 \leq a < 1$, $D_a(r)$ changes signs as a function of $r \in [0, \infty)$ once from positive to negative.*

Proof. See the Appendix.

With the estimators δ_a and δ_{mle} being equivariant (with respect to orthogonal transformations), and hence with their risks constant on spheres centered at the origin, we denote their difference in risks as

$$\Delta_a^p(\lambda) = R(\theta, \delta_{mle}) - R(\theta, \delta_a)\,, \qquad (4.2.4)$$

for $\theta \in S_\lambda = \{\theta \in \Re^p : \|\theta\| = \lambda\}$, $\lambda \in [0, m]$, emphasizing the role of the dimension p. Recall also that, the probability distribution of $\|X\|^2$ is $\chi_p^2(\lambda^2)$, so that the potential sign changes of $\Delta_a(\lambda) = E_\theta\left[D_a(\|X\|)\right]$ are controlled by the variational properties of $D_a(\cdot)$ in terms of sign changes (e.g., [Kar57]; [BJM81]). Therefore, according to Lemma 4.2.1 and since $\Delta_a^p(0) = E_0[\|\delta_{mle}(X)\|^2 - \|\delta_a(X)\|^2] > 0$, we deduce that $\Delta_a^p(\lambda)$ is, as function of $\lambda \in [0, m]$, either positive, or changes signs once from positive to negative. So, we conclude this section with the following.

Corollary 4.2.1. *For any $0 \le a < 1$, δ_a dominates δ_{mle} if and only if*

$$\Delta_a^p(m) \ge 0 .$$

4.3 Analytical risk comparisons

In this section, we focus on : **(A)** dominance results obtained as an application of Corollary 4.2.1, achieved by further risk analysis at the boundary S_m of the parameter space both for finite p (Subsection 4.3.1) and within an asymptotic framework when $p \to \infty$ (Subsection 4.3.2) ; and **(B)** approximations for the actual difference in risks $\Delta_a^p(\lambda)$; $\lambda \in [0, m]$ (Subsection 4.3.2). In what follows, we write $\|\delta_a(x)\| = g_a(r) = m \wedge ar$, $\|\delta_{\text{mle}}(x)\| = g_1(r) = m \wedge r$ and, similarly, we note and recall (e.g., [MP01]) that $\|\delta_{\pi_\lambda}(x)\| = g_{\pi_\lambda}(r) = \lambda\, \rho_{p/2-1}(\lambda r)$, where $\rho_\nu(t) = I_{\nu+1}(t)/I_\nu(t), t > 0, \nu \ge -1/2$, I_ν representing the modified Bessel function of order ν. Some further properties of ρ_ν, presented in Lemma 4.5.1 of the Appendix, will be used below. We require the following risk and difference in risks representations, achieved by conditioning on $R = \|X\|$, and bringing into play the Bayes estimator δ_{π_λ} associated with a uniform prior on S_λ. These representations were introduced by Marchand and Perron [MP01], and were further exploited in [MP05] and [FM10].

Lemma 4.3.1. **(a)** *For an equivariant estimator $\delta_g(x) = g(\|x\|)\frac{x}{\|x\|}$, we have for $\theta \in S_\lambda$*

$$R(\theta, \delta_g) = R(\theta, \delta_{\pi_\lambda}) + E_\theta\left[\{\|\delta_g(X)\| - \|\delta_{\pi_\lambda}(X)\|\}^2\right], \qquad (4.3.1)$$

where δ_{π_λ} is the Bayes estimator with respect to uniform prior on the sphere S_λ.

(b) *For $\theta \in S_\lambda$, the difference in risks between the estimators δ_{mle} and δ_a is given by*

$$\Delta_a^p(\lambda) = E_\theta\left\{(g_{mle}(R) - g_a(R))\,(g_{mle}(R) + g_a(R) - 2g_{\pi_\lambda}(R))\right\}\,. \tag{4.3.2}$$

Proof. Part (b) follows from part (a). Part (a) is given in [MP01].

4.3.1 Finite p

We seek analytical results describing the subset of parameter space values $\{\theta \in \Theta(m) : \Delta_a^p(\|\theta\|) \geq 0\}$, as well as dominance situations (δ_a dominates δ_{mle}) where $\Delta_a^p(\lambda) \geq 0$ for all $\lambda \in [0, m]$. Beyond the universal cutoff point of Theorem 4.1.1, the following inferences may be derived from (4.3.2) for cases where $\lambda = \|\theta\| < \sqrt{p}$.

Theorem 4.3.1. **(a)** *For $a \in (0,1)$, $\lambda < \sqrt{p}$, we have $\Delta_a^p(\lambda) \geq 0$ whenever $a \geq (\frac{2\lambda^2}{p} - 1) \vee (\frac{\lambda}{m}\rho_{p/2-1}(\lambda m))$;*

(b) *For $a \in (0,1)$, $m < \sqrt{p}$, δ_a dominates δ_{mle} whenever $a \geq (\frac{2m^2}{p} - 1) \vee \rho_{p/2-1}(m^2)$.*

Proof. It suffices to establish **(a)**, since **(b)** follows from **(a)** as $\rho(\cdot)$ is increasing and positive on $(0, \infty)$. Since $g_a(r) \leq g_{mle}(r)$ for all $a \in (0,1), r > 0$, it follows from (4.3.2) that $\Delta_a^p(\lambda) \geq 0$ as soon as

$$g_{mle}(r) + g_a(r) \geq 2g_{\pi_\lambda}(r) = 2\lambda\; \rho_{p/2-1}(\lambda r) \;\text{ for all}\, r > 0. \tag{4.3.3}$$

Now consider cases : **(i)** $r \in (0, m]$, **(ii)** $r \in (m, m/a)$, and **(iii)** $r \geq m/a$. In **(iii)**, (4.3.3) is satisfied since $g_{mle}(r) + g_a(r) = 2m$ for $r \geq m/a$, while

$\rho_{p/2-1}(\cdot)$ is bounded above by 1 (see Lemma 4.5.1 of the Appendix). In **(i)**, condition (4.3.3) is equivalent to :

$$\begin{aligned}
& r(1+a) && \geq 2\lambda\rho_{p/2-1}(\lambda r) \text{ for all } r \in (0,m] \\
\Leftrightarrow \quad & (1+a) && \geq 2\lambda^2 \sup_{r\in(0,m]} \{\frac{\rho_{p/2-1}(\lambda r))}{\lambda r}\} \\
\Leftrightarrow \quad & a && \geq \frac{2\lambda^2}{p} - 1,
\end{aligned}$$

using the properties of Lemma 4.5.1. Finally in **(ii)**, condition (4.3.3) becomes

$$\frac{m}{r} + a \geq 2\lambda^2 \frac{\rho_{p/2-1}(\lambda r))}{\lambda r}, \quad \text{for all } r \in (m, m/a),$$

which is satisfied whenever

$$\begin{aligned}
& \inf_{r\in(m,m/a)}\{\tfrac{m}{r} + a\} \geq 2\lambda^2 \sup_{r\in(m,m/a)}\{\tfrac{\rho_{p/2-1}(\lambda r))}{\lambda r}\} \\
\Leftrightarrow \quad & a \geq \tfrac{\lambda}{m}\rho_{p/2-1}(\lambda m),
\end{aligned}$$

using Lemma 4.5.1 again. This establishes the lemma. □

Before proceeding with an illustration, we point out that the actual comparison of risks is stronger than stated in Theorem 4.3.1 as it applies to all conditional on $\|X\| = r$ risk differences.

Exemple 4.3.1. *The performances of δ_{tlx} and δ_{mle} are compared in Figure 4.1 for two combinations of (m,p) with δ_{tlx} seen to improve on δ_{mle} for a large part of the parameter space $\Theta(m)$. These represent situations where part (b) of Theorem 4.3.1 does not apply as $m \geq \sqrt{p}$. However, part (a) of the Theorem applies for $a = \frac{m^2}{m^2+p}$, and we find by evaluation that δ_{tlx} has a smaller risk than δ_{mle} when $\lambda \leq \sqrt{11/13} \approx 0.9199$ for $(m,p) = (1.5, 1)$, and when $\lambda \leq \sqrt{21/8} \approx 1.6202$ for $(m,p) = (3,3)$. These are sufficient conditions of course and the actual regions where δ_{tlx} improves on δ_{mle} stretches beyond* *these limits, to about $\lambda \le 1.18$ and $\lambda \le 2.72$ respectively. Observe as well that the graphs illustrate Section 4.2's sign change analysis for $\Delta_a^p(\|\theta\|)$, as well as Corollary 4.2.1.*

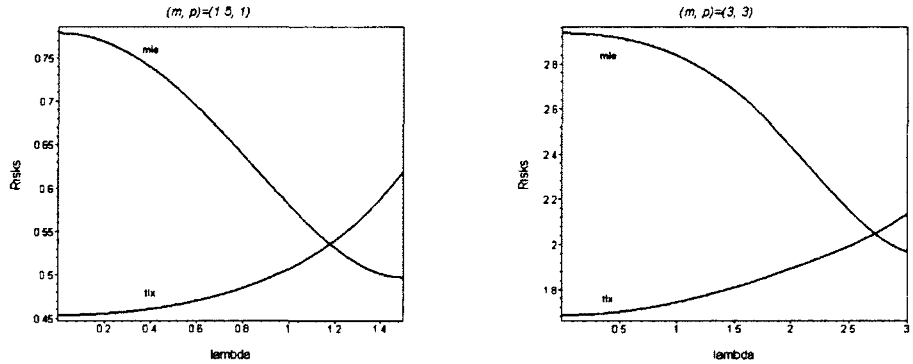

FIGURE 4.1 – Risks of δ_{mle} and δ_{tlx} for $(m,p) = (1.5, 1)$ and $(m,p) = (3,3)$.

Now consider the set $C(m,p) = \{a \in [0,1) : \Delta_a^p(m) \ge 0\}$, representing the subclass of estimators δ_a that dominate δ_{mle} for a given (m,p) in view of Corollary 4.2.1. Part (b) of Theorem 4.3.1 produces a clear result when $m < \sqrt{p}$: $C(m,p)$ includes an interval of the form $[a(m,p),1)$. For instance, we have for $(m,p) = (1,3)$: $a(1,3) = \rho_{1/2}(1) = \coth(1) - 1 = 2/(e^2-1) \approx 0.3130$ (see Lemma 4.5.1). This is illustrated in Figure 4.2 where the actual set $C(1,3)$ is, of course, larger and coincides almost with the full interval $[0,1)$. For $m \ge \sqrt{p}$, as Theorem 4.3.1 suggests, it seems plausible that $C(m,p)$ also includes, and is given by an interval of the form $[a(m,p),1)$ when it is not empty. Such a situation is shown in Figure 4.2 and where we have evaluated $C(6,9) \approx [0.833,1)$. We pursue in this direction by identifying instances where $\Delta_a^p(m)$ decreases in a locally at $a=1$, and thus guarantees that δ_a dominates δ_{mle} for values of a such that $1-a$ is sufficiently small.

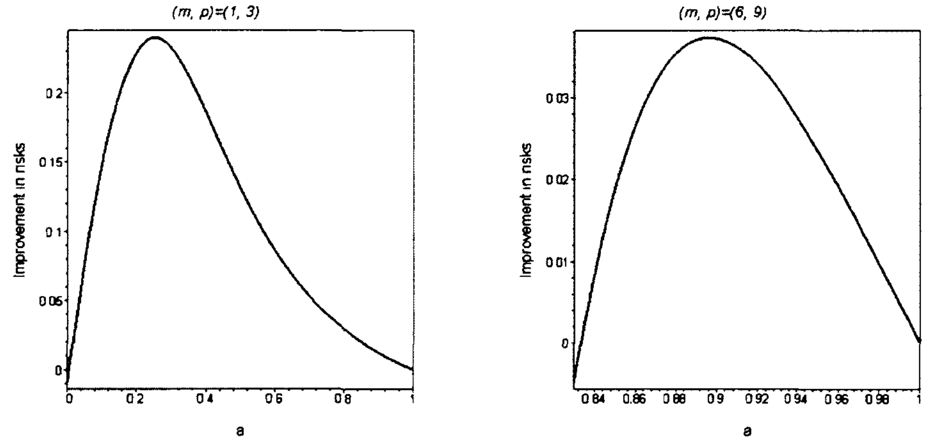

FIGURE 4.2 – Boundary risk difference $\Delta_a^p(m)$ between δ_{mle} and δ_a for $(m,p) = (6,9)$ and $(m,p) = (1,3)$.

Theorem 4.3.2. *For all $p \geq 1$, there exists $m_p > \sqrt{p}$ such that for any $m \leq m_p$, there exists $a(m,p)$ for which δ_a dominates δ_{mle} for all $a \in [a(m,p), 1)$. The cutoff point m_p is given as the smallest positive root of $H_p(\cdot)$, with $H_p(m) = E[(m\,Y\,\rho_{p/2-1}(mY) - Y^2)\,1_{(0,m)}(Y)]$, and $Y \sim \sqrt{\chi_p^2(m^2)}$.*

Proof. By virtue of Corollary 4.2.1, it suffices to focus on the behavior of $\Delta_a^p(m)$ for fixed (m,p) and in a neighbourhood of $a = 1$. From (4.3.2), we obtain :

$$\frac{\partial}{\partial a}\Delta_a^p(m) = 2\int_0^{m/a} r f_{p,m}(r)\left[m\rho_{p/2-1}(mr) - ar\right]\,dr\,,$$

where $f_{p,m}(\cdot)$ represents the density of R (i.e., of Y) when $\|\theta\| = m$. Hence, at $a = 1$, we have :

$$\begin{aligned}\frac{\partial}{\partial a}\Delta_a^p(m)|_{a=1^-} &= 2\int_0^m r^2 f_{p,m}(r)\left[m\frac{\rho_{p/2-1}(mr)}{r} - 1\right]\,dr \\ &= 2H_p(m)\ .\end{aligned}$$

Now, using Lemma 4.5.1 of the Appendix, we have for all $p \geq 1$ and any $m \leq \sqrt{p}$:

$$m\frac{\rho_{p/2-1}(mr)}{r} - 1 < \frac{m^2}{p} - 1 \leq 0,$$

for all $0 < r \leq m$, which in turn implies that $H_p(m) < 0$ for $0 \leq m \leq \sqrt{p}$. The result follows from continuity of $H_p(\cdot)$. $\square$

Exemple 4.3.2. *For $p = 9$, we obtain numerically $m_9 \approx 7.408$ as the smallest positive root of $H_9(\cdot)$. Theorem 4.3.2 tells us that for any $m \leq 7.408$ (approx.), there exists dominating estimators δ_a of δ_{mle} for small enough positive $1 - a$. For instance, if we consider $m = 5$, we obtain numerically $a(5,9) \approx 0.6979$, and hence the estimators δ_a with $a \in [0.6979, 1)$ (approx.) dominate δ_{mle}. These include δ_{tlx} since $a_{tlx} = 25/34 \geq a(5,9)$.*

We conjecture but have been unable to prove that, for all p, $C(m,p)$ is empty for $m > m_p$. For the univariate case $p = 1$, we go further into the analysis with regards to Theorem 4.3.2's cutoff point m_1, as well as properties of $R(m, \delta_a)$ and $C(m,1)$. We will require the following expressions concerning the risk $R(m, \delta_a)$ and its derivative $\frac{d}{da}R(m, \delta_a)$. We denote hereafter ϕ and Φ as the standard normal pdf and cdf respectively.

Lemma 4.3.2. *For $p = 1$ and $a \in (0,1)$, we have :*

$$\begin{aligned} R(m,\delta_a) &= \left(a^2\left(1+m^2\right) - 2am^2\right)(B - A) \\ &\quad + m\left(a^2 - 3a\right)D - m\left(a^2 - a\right)C + m^2(4 - A - 3B), \end{aligned} \tag{4.3.4}$$

and

$$\frac{d}{da}R(m,\delta_a) = \left(2a + 2m^2(a-1)\right)(B - A) + 2m(a-2)D - 2maC,$$

where $A = \Phi(m - m/a)$, $B = \Phi(m + m/a)$, $C = \phi(m - m/a)$, *and* $D = \phi(m + m/a)$.

Proof. See the appendix. $\square$

Here is a follow up to Theorem 4.3.2 for the univariate case.

Theorem 4.3.3. *For* $p = 1$, *Theorem 4.3.2 applies with* $m_1 \approx 1.1404$ *being the unique root in* $(0, \infty)$ *of* $H_1(m) = -m\phi(0) + m\phi(2m) - \frac{1}{2} + \Phi(2m)$. *Furthermore if* $m > m_1$, *then the set* $C(m, 1) = \{a \in [0, 1) : \Delta_a^1(m) \geq 0\}$ *does not contain an interval of the form* $[a(m, 1), 1)$.

Proof. From Lemma 4.3.2, we obtain $\frac{\partial}{\partial a}\Delta_a^1(m)|_{a=1^-} = 2H_1(m)$. Observe now that $H_1(0) = 0$, $H_1'(m) = \phi(2m) - \phi(0) + 4m^2\phi(2m)$, and $H_1''(m) = 4m\,\phi(2m)(1-4m^2)$. Since $H_1'(0) = 0$, and $H_1'(m) \to \phi(0) > 0$ when $m \to \infty$, we infer that $H_1'(m)$ changes signs once from positive to negative on $(0, \infty)$, and we conclude indeed that $H_1(m)$ has a unique positive root m_1, which implies that $\frac{\partial}{\partial a}\Delta_a^1(m)|_{a=1^-} \leq 0$ if and only if $m \leq m_1$, establishing the result. $\square$

Lemma 4.3.3. *The function* $R(m, \delta_a)$ *is decreasing in* a *if* $0 \leq a \leq \frac{m^2}{m^2+1}$ *and increasing when* $m < m_1 \approx 1.1404$ *and* $b(m) \leq a < 1$, *with* $b(m) = \frac{m^2(\Phi(2m)-1/2)+2m\phi(2m)}{(1+m^2)(\Phi(2m)-1/2)+m\phi(2m)-m\phi(0)}$.

Proof. See the Appendix.

As an illustration of the above univariate results, we take $m = 0.95$. Theorems 4.3.2 and 4.3.3 tell us that there are necessarily some δ_a's, with $1 - a$ small enough and positive, that dominate δ_{mle}. An evaluation leads to dominance for $a \in [a(0.95, 1), 1) \approx [0.5180, 1)$, but this does not include $\delta_{tlx} = \delta_{361/761}$.

The value $b(m)$ provides, for $m \leq m_1$, an explicit upper bound for $a(m,1)$. Here, for instance, we obtain $b(0.95) \approx 0.733$.

Previous findings related to the univariate case were given by Dou and van Eeden [DvE09] and Marchand et al. [MOPP08]. Dou and van Eeden have shown that the trivial estimator $\delta_0 = 0$ dominates δ_{mle} if and only if $0 < m \leq c_0 \approx 0.52044$. By virtue of Lemma 4.3.3's decreasing property and Corollary 4.2.1, Dou and van Eeden's result implies automatically that δ_a dominates δ_{mle} for $p = 1$, $m \leq c_0$, $a \in (0, a_0]$ with $a_0 > m^2/(m^2+1)$. This includes the truncated linear minimax δ_{tlx}, but [MOPP08] have shown by a different route the stronger result that δ_{tlx} dominates δ_{mle} whenever $m \leq c_1 \approx 0.60936$.

4.3.2 An asymptotic framework and inferences for large p

In this section, we not only investigate Corollary 4.2.1's dominance condition in an asymptotic framework, but we also provide accurate approximations for the risk differences $\Delta_a^p(\lambda)$ in (4.2.4), as well as for the relative risk differences

$$w_a^p(m,\lambda) = \frac{\Delta_a^p(\lambda)}{R(\theta, \delta_{mle})}\,, \text{ with } \theta \in S_\lambda\,. \tag{4.3.5}$$

To achieve this, we consider a sequence of problems $\{(m,\lambda,p) = (d\sqrt{p}, c\sqrt{p}, p)\,;\ p \geq 1\}$, with $d > 0$ and $0 \leq c \leq d$, and study the corresponding limiting value of $w_a^p(d\sqrt{p}, c\sqrt{p})$ as p grows to $+\infty$. We will denote this value as :

$$h(a,d,c) \;=\; \lim_{p\to\infty} w_a^p(d\sqrt{p}, c\sqrt{p})\,. \tag{4.3.6}$$

For assessing whether or not $\delta_a\,;\, a \in [0,1)\,;$ dominates δ_{mle}, it is true, by virtue of Corollary 4.2.1, that we only need to consider $w_a^p(d\sqrt{p}, c\sqrt{p})$, or $h(a,d,c)$, for $c = d$, but our objective here is also to quantify explicitly the relative gains in risk. In our findings below, δ_a is taken to be independent of p. For instance, observe for the truncated linear estimator, that $\delta_{tlx}(x) = \delta_{\frac{d^2}{d^2+1}}(x)$ with the setting $m = d\sqrt{p}$. We now require the following technical result.

Lemma 4.3.4. *For $c \geq 0$, $p = 1, 2, \ldots$, $R_p^2 \sim \chi_p^2(c^2 p)$, and $m = d\sqrt{p}$ with $d > 0$,*

(a) The sequence of random variables $(m \wedge aR_p)/\sqrt{p}$, $a \geq 0$, converges as $p \to \infty$, in probability to $d \wedge a\sqrt{c^2+1}$;

(b) The sequence of random variables $\rho_{p/2-1}(c\sqrt{p}R_p)$ converges, as $p \to \infty$, in probability to $c/\sqrt{c^2+1}$.

Proof. Part (a) follows since R_p^2/p converges, as $p \to \infty$, in probability to c^2+1, while part (b) was given by [MP02]. □

Applying Lemma 4.3.4 allows us to calculate $lim_{p\to\infty} \frac{\Delta_a^p(c\sqrt{p})}{d^2 p}$ in terms of c, d and a, and to infer dominance results in the corresponding asymptotic setting.

Theorem 4.3.4. *For $d > 0$, $0 \leq c \leq d$, and $a \in [0, 1)$, we have :*

$$\lim_{p\to\infty} \frac{\Delta_a^p(c\sqrt{p})}{d^2 p} = \begin{cases} (1-a)\,\frac{c^2+1}{d^2}\,(a - \frac{c^2-1}{c^2+1}) & \text{if } c < \sqrt{(d^2-1)_+} \\ \left(1 - a\frac{\sqrt{c^2+1}}{d}\right)_+ (1 + a\frac{\sqrt{c^2+1}}{d} - \frac{2c^2}{d\sqrt{c^2+1}}) & \text{if } c \geq \sqrt{(d^2-1)_+}; \end{cases}$$

where $y_+ = y \vee 0$, $y \in \Re$.

Proof. Applying (4.3.2), we have for $\theta \in S_\lambda$:

$$\lim_{p\to\infty} \frac{\Delta_a^p(c\sqrt{p})}{d^2 p} = \lim_{p\to\infty} E_\theta \left[\left(\frac{m \wedge R}{d\sqrt{p}} - \frac{m \wedge aR}{d\sqrt{p}} \right) \times \left(\frac{m \wedge R}{d\sqrt{p}} + \frac{m \wedge aR}{d\sqrt{p}} - 2\frac{c}{d}\rho_{p/2-1}(c\sqrt{p}R) \right) \right].$$

Now, from Lemma 4.3.4, we obtain :

$$\lim_{p\to\infty}\frac{\Delta_a^p(c\sqrt{p})}{d^2p} = \Bigg[\left((1\wedge\frac{\sqrt{c^2+1}}{d}) - (1\wedge\frac{a\sqrt{c^2+1}}{d})\right) \times \left((1\wedge\frac{\sqrt{c^2+1}}{d}) + (1\wedge\frac{a\sqrt{c^2+1}}{d}) - 2\frac{c}{d}\frac{c}{\sqrt{c^2+1}}\right)\Bigg],$$

which leads to the results. □

Corollary 4.3.1. *For $a \in [0,1)$, we have $\lim_{p\to\infty}\frac{\Delta_a^p(c\sqrt{p})}{p} \geq 0$, for all $c \in [0,d]$, if and only if $a \in \left[\frac{2d^2}{d^2+1} - \frac{d}{\sqrt{d^2+1}}, 1\right)$. Furthermore, the inequality is strict for all $c \in [0,d]$ if and only if $a \in \left(\frac{2d^2}{d^2+1} - \frac{d}{\sqrt{d^2+1}}, \frac{d}{\sqrt{d^2+1}}\right)$, and this includes δ_{tlx}. Finally, for $d < 1$ and $a \in [d,1)$, the estimators δ_a and δ_{mle} are asymptotically equivalent with $\lim_{p\to\infty}\frac{\Delta_a^p(c\sqrt{p})}{p} = 0$ for all $c \in [0,d]$.*

Proof. The results follow by working directly with Theorem 4.3.4's expression for $\lim_{p\to\infty}\frac{\Delta_a^p(c\sqrt{p})}{p}$. The inferences specific to δ_{tlx} are a direct consequence of the more general result since $\frac{d^2}{d^2+1} \in \left(\frac{2d^2}{d^2+1} - \frac{d}{\sqrt{d^2+1}}, \frac{d}{\sqrt{d^2+1}}\right)$ for all $d > 0$. □

Corollary 4.3.1 exhibits, through Theorem 4.3.4's expression for the exact asymptotics risk difference, the asymptotic dominance, as well as strict asymptotic dominance, relative to a sequence $(m = d\sqrt{p}, p)$ with $p \to \infty$, of a subclass of truncated linear estimators δ_a over δ_{mle}. This subclass of improvements includes the truncated linear minimax estimator δ_{tlx} for all d. Coupled with the fixed p analytical and numerical results in Section 4.3.1 and below as well, this finding clearly indicates that δ_{tlx} and other asymptotically dominating δ_a's will necessarily outperform δ_{mle} for large enough p. Observe as well that Corollary 4.3.1's lower bound $\frac{2d^2}{d^2+1} - \frac{d}{\sqrt{d^2+1}}$ is less or equal to 0 for $d \leq \sqrt{1/3}$, so that the dominance condition is universal applying to all

δ_a's with $a \in [0, d)$, in accordance with Theorem 4.1.1's universal dominance cutoff point.

We now turn to the relative degree of improvement attained by such δ_a's as measured by the quantity $h(a, d, c)$ given in (4.3.6).

Theorem 4.3.5. *For $d > 0$, $0 \le c \le d$, and $a \in [0, 1)$, we have :*

$$h(a,d,c) = \begin{cases} (1-a)\,(c^2+1)\,(a - \frac{c^2-1}{c^2+1}) & \textit{if } c < \sqrt{(d^2-1)_+}\,; \\ 1 - (\frac{c^2(a-1)^2+a^2}{c^2+d^2-\frac{2c^2d}{\sqrt{c^2+1}}}) & \textit{if } c \ge \sqrt{(d^2-1)_+} \textit{ and } c < \sqrt{(\frac{d^2}{a^2}-1)_+}\,; \\ 0 & \textit{if } c \ge \sqrt{(\frac{d^2}{a^2}-1)_+}\,. \end{cases}$$

Furthermore, $h(a, d, \cdot)$ decreases on $[0, d]$ whenever $d \ge d_0 = (\frac{\sqrt{5}-1}{2})^{1/2} \approx 0.7862$, or $d < d_0$ and $a \le d/(d+1)$.

Proof. The monotonicity property is established in the Appendix. Since $R(\theta, \delta_{mle}) = R(\theta, X) + \{R(\theta, \delta_{mle}) - R(\theta, X)\}$, Lemma 4.3.1 permits us to write :

$$R(\theta, \delta_{mle}) = p + E_\theta\left[(g_{mle}(R) - R)\,(g_{mle}(R) + R - 2g_{\pi_\lambda}(R))\right]\,. \quad (4.3.7)$$

for all $\theta \in S_\lambda$. Making use of Lemma 4.3.4, as in the proof of Theorem 4.3.4, we obtain from (4.3.7) :

$$\lim_{p\to\infty} \frac{R(c\sqrt{p}, \delta_{mle})}{p} = \begin{cases} 1 & \text{if } \sqrt{c^2+1} < d \\ 1 + (d - \sqrt{c^2+1})(d + \sqrt{c^2+1} - \frac{2c^2}{\sqrt{c^2+1}}) & \text{if } \sqrt{c^2+1} \ge d. \end{cases}$$

Finally, the result follows by combining Theorem 4.3.4 with the above, and rearranging terms. □

For the particular case of the truncated linear minimax estimator δ_{tlx}, we obtain the following.

Corollary 4.3.2. *For $d > 0$ and $0 \leq c \leq d$, let $\kappa(d,c) = h(\frac{d^2}{d^2+1}, d.c)$ represent the asymptotic relative risk difference between the estimators δ_{tlx} and δ_{mle}. Then, we have :*

$$\kappa(d,c) = \begin{cases} \frac{2d^2-c^2+1}{(d^2+1)^2} & \text{if } c < \sqrt{(d^2-1)_+}\ ; \\ 1 - \frac{1}{(d^2+1)^2}\frac{(c^2+d^4)\sqrt{c^2+1}}{(c^2+d^2)\sqrt{c^2+1}-2c^2d} & \text{if } c \geq \sqrt{(d^2-1)_+}\,. \end{cases}$$

Furthermore, $\kappa(d,c)$ is, for fixed $d > 0$, decreasing in $c \in [0,d]$ with minimum value $k(d,0) = \frac{d^2+1+(d^2\wedge d^4)}{(d^2+1)^2}$, and maximum value $k(d,d) = \frac{1}{2}(1 - \frac{d}{\sqrt{d^2+1}})$.

Proof. The given expression for $\kappa(d,c)$ and its monotonicity are obtained by substituting $a = d^2/(d^2+1)$ in Theorem 4.3.5. □

Exemple 4.3.3. *The expressions above for the asymptotic relative degree of improvement $h(a,d,c)$ or $\kappa(d,c)$ are remarkably simple. As an illustration, take $d = 1$, so that $m^2 = p$ and $\delta_{tlx}(x) = (\frac{1}{2}\|x\| \wedge \sqrt{p})\frac{x}{\|x\|}$. Corollary 4.3.1 indicates that δ_{tlx} dominates δ_{mle} asymptotically, and Corollary 4.3.2 tells us that the relative degree of improvement is asymptotically equal to $\kappa(1,c) = 1 - \frac{1}{4}\frac{(c^2+1)^{3/2}}{(c^2+1)^{3/2}-2c^2}$, decreases in c on $(0,1)$, with a maximum of $\kappa(1,0) = 3/4$, and a minimum of $\kappa(1,1) = \frac{\frac{3}{4}\sqrt{2}-1}{\sqrt{2}-1} \approx 0.146$. In other words, the truncated linear minimax estimator improves on the maximum likelihood estimator asymptotically, as $m^2 = p$ and $p \to \infty$, by as much as 75%, and by at least 14.6%. Observe that such gains are attenuated for larger parameter spaces, and augmented for smaller parameter spaces, as seen by the decreasing monotonicty properties of $k(d,0)$ and $k(d,d)$ as functions of d. In comparison to the case $d = 1$, we have for instance with $d = 2$: $k(2.0) = 9/25$ and $k(2,2) = \frac{1}{2} - \frac{\sqrt{5}}{5} \approx 0.0528$. Finally, the curves $k(d,c)$, $c \in [0,d]$, are presented in Figures 4.5 and 4.6 for $d = 1, 2$.*

Remarque 4.3.4. *Assessing the comparative performance of the estimators δ_a, it is easy to infer from Theorem 4.3.5 that, for fixed c, d, $h(\cdot, d, c)$ increases on $[0, \frac{c^2}{c^2+1}]$ and decreases on $[\frac{c^2}{c^2+1}, 1]$ with a unique maximum at $a^*(c) = \frac{c^2}{c^2+1}$. From this, we infer that, asymptotically, the subclass $\{\delta_a : 0 \leq a \leq \frac{d^2}{d^2+1}\}$ is minimal complete among the class $\{\delta_a : 0 \leq a \leq 1\}$, and furthermore that δ_{tlx} dominates δ_a (asymptotically) for $a \in (\frac{d^2}{d^2+1}, 1]$. These features are illustrated in both Figure 4.3 and 4.4 where, for $d = 2$ and $d = 0.5$ respectively, the degrees of improvement in risk on δ_{mle} are compared. For instance in Figure 4.4 where $d = 0.5$, we see that $\delta_{tlx} = \delta_{0.2}$ dominates both $\delta_{0.30}$ and $\delta_{0.45}$, but not $\delta_{0.10}$, the latter being a member of the admissible subclass with $a \in [0, 0.2]$. Notice as well, for $d = 0.5$, that the strict asymptotic dominance is achieved by all δ_a with $a < 0.5$ by virtue of Corollary 4.3.1.*

Remarque 4.3.5. *As a follow up to the previous remark, we expand on a further "asymptotic minimax type" property of δ_{tlx} To do so, consider our estimation problem with weighted squared error loss $L^*(\theta, \delta) = \frac{(\delta-\theta)^2}{q(\theta)}$, with $q(\theta) = R(\theta, \delta_{mle})$ and associated risk $R^*(\theta, \delta) = E_\theta(L^*(\theta, \delta(X)))$. By virtue of Theorem 4.3.5, we have for the sequence $\{(m, \lambda, p) = (d\sqrt{p},\ c\sqrt{p},\ p); p \geq 1\}$, $\lim_{p\to\infty} R^*(c\sqrt{p}, \delta_a) = 1 - h(a, d, c)$ and the worst case scenario*

$$\sup_{c\in[0,d]} \{\lim_{p\to\infty} R^*(c\sqrt{p}, \delta_a)\} = 1 - h(a, d, d),$$

for $a \in [0, d^2/(d^2+1)]$, i.e. for the complete class members given in Remark 4.3.4. Finally, as seen in Remark 4.3.4, the choice $a = d^2/(d^2+1)$ minimizes $1 - h(a, d, d)$ in a, so that we obtain the optimality property

$$\inf_{a\in[0,1)} \{\sup_{c\in[0,d]} \{\lim_{p\to\infty} R^*(c\sqrt{p}, \delta_a)\}\} = \sup_{c\in[0,d]} \{\lim_{p\to\infty} R^*(c\sqrt{p}, \delta_{tlx})\}\,.$$

for the truncated linear minimax estimator δ_{tlx}.

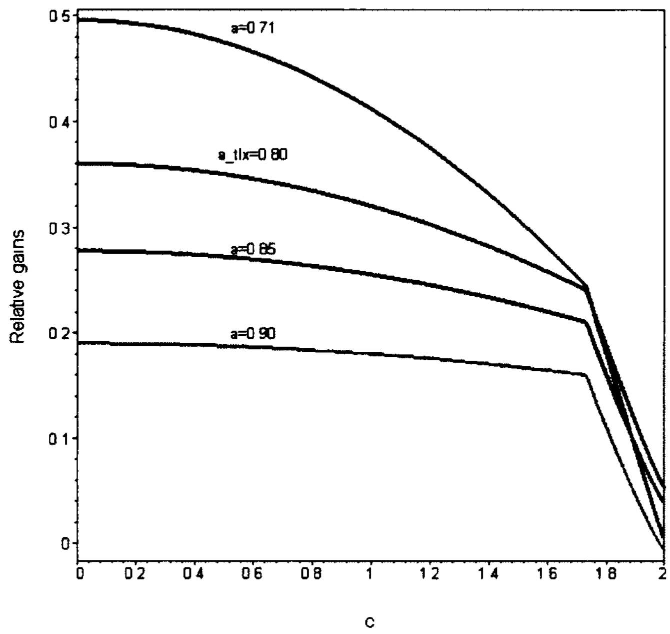

FIGURE 4.3 – Graphs of the ratio of related improvement in risks for $d = 2$ with $a = 0.71, a_{tlx} = 0.8, a = 0.85$ and $a = 0.90$.

Remarque 4.3.6. *In practice for fixed (m,p) and a given estimator δ_a, we are concerned with the behaviour of the relative risk difference $w_a^p(m,\lambda)$, $\lambda \in [0,m]$. To this end, numerical evidence suggests that $w_a^p(d\sqrt{p}, c\sqrt{p})$ is either fairly or very well approximated by $h(a,d,c)$ for large p, and even for moderate p. From a theoretical point of view, we point out that the exhibited quality of the approximation reflects the speed of Lemma 4.3.4's convergence, or more precisely through the expectations arising in the risk functions of δ_{tlx} and δ_{mle}. Illustrations are provided in Figures 4.5 and 4.6, for $\delta_a = \delta_{tlx}$, $p = 9, 16$ and $d = 1, 2$. The approximation $\kappa(d,c)$ produces slight overestimates which are attenuated for larger p. For instance with $d = 1, c = 0.6$,* *we have relative gains of* 0.5028 *for* $(m, p, \lambda) = (3, 9, 1.8)$, *of* 0.5223 *for* $(m, p, \lambda) = (4, 16, 2.4)$, *of* 0.5367 *for* $(m, p, \lambda) = (6, 36, 3.6)$, *in comparison with* $\kappa(1, 0.6) = 0.5422$. *Please note however that* δ_{tlx} *does not always dominate* δ_{mle} *so that the asymptotics, if misinterpreted, will necessarily produce a contrary dominance conclusion for some pairs* (m, p). *Such an occurrence arises for* $(m, p) = (6, 9)$ *as illustrated in Figure 4.6, and as previously seen in Section 4.3.1 where we obtained that* δ_a *dominates* δ_{mle} *iff* $a \in C(6, 9) \approx [0.833, 1)$ *with* $\delta_{tlx} = \delta_{0.8}$.

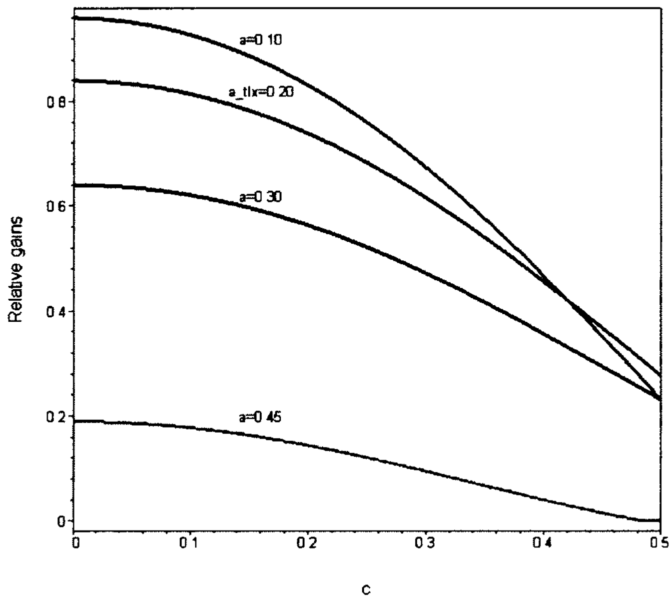

FIGURE 4.4 – Graphs of the ratio of related improvement in risks for $d = 0.5$ with $a = 0.10, a_{tlx} = 0.20. a = 0.30$ and $a = 0.45$.

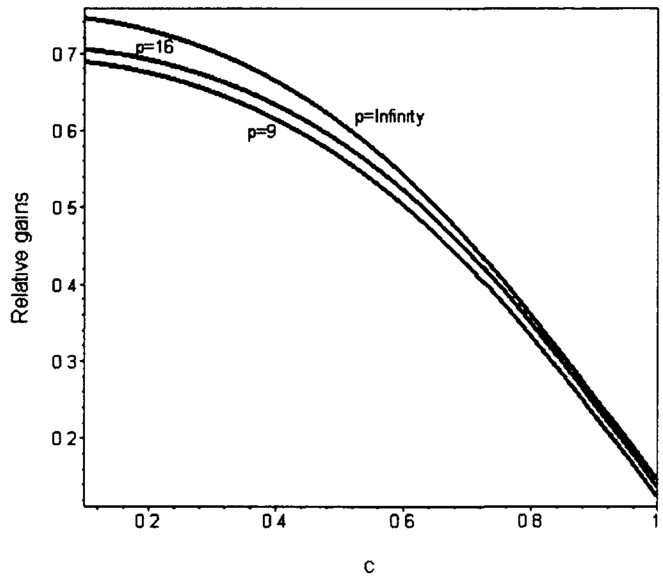

FIGURE 4.5 – The relative gains in risk $w_a^p(d\sqrt{p}, c\sqrt{p})$ for $d = 1$, $\delta_a = \delta_{tlx}$, $p = 9, 16, \infty$.

4.4 Concluding remarks

Section 4.2 and Subsection 4.3.1 provides finite dimension necessary and sufficient analysis (including the univariate case) in comparing truncated linear estimators δ_a with δ_{mle}. We point out that the truncated linear estimators δ_a ; $a > 0$; are themselves inadmissible as they take values on the boundary of the parameter space, and that they can be improved upon (e.g., [Moo85] and [MP01]). However, they remain nevertheless appealing and useful benchmarks in how to shrink efficiently and our findings provide sharp dominance results, namely in the the asymptotic setting, and precise expressions for the relative gains in asymptotic risk, which will be amplified by estimators that improve upon δ_a.

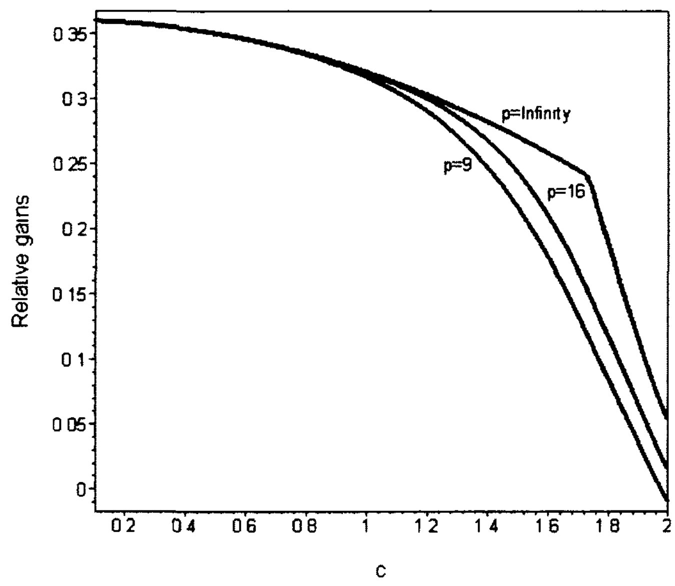

FIGURE 4.6 – The relative gains in risk $w_a^p(d\sqrt{p}, c\sqrt{p})$ for $d = 2$, $\delta_a = \delta_{tlx}$, $p = 9, 16, \infty$.

The asymptotic setting and techniques, which we have made use of and extended to analyze the risks over the entire parameter space, were previously introduced and exploited by Marchand and Perron [MP02], as well as Fourdrinier and Marchand [FM10], for studying the risk on the boundary of the parameter space $\Theta(m)$. The former studied conditions for which the boundary uniform Bayes estimator δ_{BU} is minimax, while the latter give conditions for which δ_{BU}, and other Bayes estimators with respect to uniform priors on spheres, dominate δ_{mle}. It would be most interesting to proceed with a similar asymptotic setting to assess the performance of the fully uniform Bayes estimator δ_U, or other Bayes estimators. Partial results were given in [Gue03], namely the asymptotic dominance in our setting for $d < 1$. We conjecture that further and stronger dominance results hold, and we feel that similar quantification of the relative gains in risk are possible, and that the asymptotic setting holds much promise. As well, in a related result, a particular case of a result of Hartigan ([Har04]) implies that δ_U improves upon the benchmark estimator X, but we conjecture that δ_U does not or does not always improve upon the linear minimax procedure $\frac{m^2}{m^2+p}X$.

Finally, we point out that several challenges remain for the problem of determining improvements on δ_{mle}. Indeed, although dominating estimators can be provided for any pair (m, p), the problem of the specification of priors π that lead to dominating Bayesian estimators δ_π remains unsolved in general with solutions provided for m relatively small enough (see [MP01]). An analogous and more difficult challenge, in view namely of the findings in this paper, resides in findings Bayesian improvements of truncated linear estimators δ_a, and namely of the truncated linear minimax estimator δ_{tlx}. Moreover, to our knowledge, there exists little or no findings with regards to extensions with unknown σ^2.

Acknowledgments

The research work of Éric Marchand is partially supported by NSERC of Canada. Othmane Kortbi is a Ph.D. student at the Université de Sherbrooke. During his studies, he benefited from financial support from several sources but he wishes to thank especially the ISM (Institut de sciences mathématiques) and the CRM (Centre de recherches mathématiques).

4.5 Appendix

The function $\rho_{(p/2)-1}$ plays a key role and the next lemma recalls some useful properties, given in [Amo74], and [Wat83] for p > 1 and readily verified otherwise for $p = 1$ by using the representation $\rho_{-1/2} \equiv tanh$.

Lemma 4.5.1. *For all $p \geq 1$, $\rho_{(p/2)-1}(t)$ is increasing in t, with $\rho_{(p/2)-1}(0) = 0$ and $\rho_{(p/2)-1}(t) \to 1$ as $t \to \infty$. Also, $\rho_{(p/2)-1}(t)/t$ is decreasing in t, with $\rho_{(p/2)-1}(t)/t \to 1/p$ as $t \to 0$. We have as well $\rho_{p/2}(t) = \frac{1}{\rho_{(p/2)-1}(t)} - \frac{p}{t}$, and in particular $\rho_{1/2}(t) = coth(t) - 1/t$.*

Proof of Lemma 4.2.1.

(a) Since $\delta_a(x) = x + h_a(x)$ with $h_a(x) = \{(a-1)\,[0 \leq r \leq m/a] + (\frac{m}{r} - 1)\,[r > m/a]\}\,x$, we obtain :

$$\begin{aligned} 2divh_a(x) + \|h_a(x)\|^2 = &\left\{2ap - 2p + (a-1)^2 r^2\right\}\, 1_{[0,\frac{m}{a}]}(r) \\ &+ \left\{2(p-1)\frac{m}{r} - 2p + (m-r)^2\right\} 1_{(\frac{m}{a},\infty)}(r)\,; \end{aligned}$$

and, by virtue of Stein's identity (4.2.2), $R(\theta, \delta_a) = E_\theta[\eta_a(X)]$ with

$$\begin{aligned} \eta_a(X) = &\left\{2ap - p + (a-1)^2 r^2\right\}\, 1_{[0,\frac{m}{a}]}(r) \\ &+ \left\{2(p-1)\frac{m}{r} - p + (m-r)^2\right\} 1_{(\frac{m}{a},\infty)}(r)\,. \end{aligned} \tag{4.5.1}$$

Finally, the given expression for the unbiased estimator $D_a(\|X\|)$ follows directly from (4.5.1) as $\delta_{mle} \equiv \delta_1$ with our notation.

(b) First, observe that $D_a(\cdot)$ decreases on $[0, m)$ with $D_a(0) = 2p(1-a) > 0$, and is thus either positive, or changes signs once from positive to negative on $[0, m)$. Secondly, $D_a(\cdot)$ is convex on $(m, m/a]$ with $D_a(\frac{m}{a}) = -2a \leq 0$, which implies that $D_a(\cdot)$ is either negative, or changes signs once from positive to negative on $(m, m/a]$. Thirdly, we have $\lim_{r\to m^+} D_a(r) = \lim_{r\to m^-} D_a(r) - 2$, in other words a drop discontinuity at m. The result follows from the above properties and since $D_a(\cdot)$ is everywhere continuous on $[0,\infty)$ except at m and m/a. $\square$

Proof of Lemma 4.3.2.

With $\delta_a(X) = sgn(X)(a\|X\| \wedge m)$ for $p=1$, the risk function can be directly expressed as :

$$\begin{aligned}
R(m,\delta_a) &= \int_{-\frac{m}{a}}^{\frac{m}{a}} |ax-m|^2 \phi(x-m)\,dx + \int_{-\infty}^{-\frac{m}{a}} 4m^2\phi(x-m)\,dx \\
&= a^2\int_{-\frac{m}{a}}^{\frac{m}{a}} x^2\phi(x-m)\,dx + m^2\left(\Phi\left(\frac{m}{a}-m\right) - \Phi\left(-\frac{m}{a}-m\right)\right) \\
&\quad -2am\int_{-\frac{m}{a}}^{\frac{m}{a}} x\phi(x-m)\,dx + 4m^2\Phi\left(-\frac{m}{a}-m\right) \\
&= a^2\int_{-\frac{m}{a}}^{\frac{m}{a}} x^2\phi(x-m)\,dx - 2am\int_{-\frac{m}{a}}^{\frac{m}{a}} x\phi(x-m)\,dx + m^2(4-A-3B).
\end{aligned}$$

Lemma 4.3.2's expression for the risk $R(m,\delta_a)$ follows with the evaluations

$$\int_{-\frac{m}{a}}^{\frac{m}{a}} x^2\phi(x-m)\,dx = \left(1+m^2\right)(B-A) + D\left(m-\frac{m}{a}\right) - C\left(m+\frac{m}{a}\right),$$

$$\int_{-\frac{m}{a}}^{\frac{m}{a}} x\phi(x-m)\,dx = m(B-A) + D - C,$$

and by collecting terms. Finally, the given expression for the derivative arises from straightforward computations and by using the relationships $A' = \frac{m}{a^2}C$, $B' = \frac{-m}{a^2}D$, $C' = \frac{-m^2}{a^2}\left(1-\frac{1}{a}\right)C$ and $D' = \frac{m^2}{a^2}\left(1+\frac{1}{a}\right)D$. $\square$

Proof of Lemma 4.3.3.

At first, we can deduce from Lemma 4.3.2 that

$$\frac{d}{da}R\left(m,\delta_a\right)\le 0 \quad \text{if} \quad \left(2a+2m^2\left(a-1\right)\right)\le 0$$

which is equivalent to $a\le\frac{m^2}{m^2+1}$. Secondly, after assuming that $a\ge\frac{m^2}{m^2+1}$, the same theorem lets us to deduce that

$$\begin{aligned}\frac{d}{da}R\left(m,\delta_a\right)\ge\left(2a+2m^2\left(a-1\right)\right)\left(\Phi\left(2m\right)-1/2\right)&\\ +2m\left(a-2\right)\phi\left(2m\right)-2ma\phi\left(0\right)&\ ,\end{aligned}$$

since $B-A=\Phi\left(m+m/a\right)-\Phi\left(m-m/a\right)$ is a decreasing function in a, and both $D=\phi\left(m+m/a\right)$ and $C=\phi\left(m-m/a\right)$ are increasing in a. So, $\frac{d}{da}R\left(m,\delta_a\right)\ge 0$ if $\left(2a+2m^2\left(a-1\right)\right)\left(\Phi\left(2m\right)-1/2\right)+2m\left(a-2\right)\phi\left(2m\right)-2am\phi\left(0\right)\ge 0$ which is equivalent to $a\ge b(m)$, since $\left(1+m^2\right)\left(\Phi\left(2m\right)-1/2\right)+m\phi\left(2m\right)-m\phi\left(0\right)\ge -H_1\left(m\right)\ge 0$ for all $m\le m_1$. This establishes the desired result. $\square$

Proof of Theorem 4.3.5's monotonicity property.

Note that, when $d>1$ and c lies in $\left(0,\sqrt{d^2-1}\right)$, the decreasing property is obvious. Otherwise, we consider separately the cases : **(i)** $a\le\frac{d}{d+1}$ and **(ii)** $a>\frac{d}{d+1}$.

(i) We may write :

$$h(a,d,c)=1-\left[\frac{\frac{(a-1)^2c^2+a^2}{c^2+d^2}}{1-\frac{2c^2d}{(c^2+d^2)\sqrt{c^2+1}}}\right].$$

and the monotonicity follows since $\frac{(a-1)^2c^2+a^2}{c^2+d^2}$ is increasing for $c>0$, given that $a\le\frac{d}{d+1}$, and $\frac{2c^2d}{(c^2+d^2)\sqrt{c^2+1}}$ is increasing for $c\in\left(0,\sqrt{\frac{d^2}{2}+\frac{d}{2}\sqrt{d^2+8}}\right)$, with $\sqrt{\frac{d^2}{2}+\frac{d}{2}\sqrt{d^2+8}}>d$.

(ii) For $a > \frac{d}{d+1}$, we write :

$$h(a,d,c) = 1 - \frac{(a-1)^2c^2 + a^2}{c^2 + d^2 - \frac{2c^2 d}{\sqrt{c^2+1}}},$$

and observe that the monotone decreasing property follows if $c^2 + d^2 - \frac{2c^2 d}{\sqrt{c^2+1}}$ is decreasing for $c \in (0, d)$. This in turn is equivalent to

$$\begin{aligned}
\left(c^2+1\right)^{3/2} - d\left(c^2+2\right) \leq 0 \quad &\Leftrightarrow \quad d \geq \frac{\left(c^2+1\right)^{3/2}}{(c^2+2)}, \text{ for all } c \in (0,d),\\
&\Leftrightarrow \quad d \geq \sup_{c\in(0,d)} \frac{\left(c^2+1\right)^{3/2}}{(c^2+2)} = \frac{\left(d^2+1\right)^{3/2}}{(d^2+2)}\\
&\Leftrightarrow \quad d^4 + d^2 - 1 \geq 0,
\end{aligned}$$

which yields the result. $\square$

Chapitre 5

Estimation avec coefficient de variation borné

Estimation de la moyenne d'une distribution normale multidimensionnelle avec coefficient de variation borné

Othmane KORTBI et Éric MARCHAND [1]

Résumé

Dans cet article, nous considérons l'estimation de la moyenne θ de la distribution multidimensionnelle $\mathcal{N}_p(\theta, \sigma^2 I_p)$, où le paramètre d'échelle σ est inconnu, avec coût quadratique et sous la contrainte $\frac{\|\theta\|}{\sigma} \leq m$ pour une certaine constante $m > 0$. Nous proposons des estimateurs qui dominent X, ainsi que l'estimateur du maximum de vraisemblance δ_{emv}. Notamment, nous montrons que l'estimateur de Bayes par rapport à la loi a priori $\pi(\theta, \sigma)$ telle que la loi de $\theta\,|\sigma$ est uniforme sur la sphère de rayon $m\sigma$ et σ suit la mesure $\frac{1}{\sigma} 1_{]0,\infty[}(\sigma)$ domine les estimateurs sans biais et δ_{emv} lorsque $m \leq \sqrt{p}$. Ce développement représente à la fois une extension des travaux de Kubokawa [Kub05] qui a considéré la cas unidimensionnel de ce problème et a développé des résultats de dominance de l'estimateur X et ceux de Marchand et Perron [MP01] qui obtiennent la dominance de l'estimateur Bayésien, par rapport à δ_{emv}, associé à la loi a priori uniforme sur la sphère de rayon $m\sigma$ avec σ connu. Nous avons établi des résultats de dominance en nous basant sur les propriétés analytiques des fonctions $h(\cdot)$ génératives des estimateurs construits $\delta_h(x, s) = h(t)x$, où $t = \|x\|^2 / s^2$. Enfin, nous obtenons une forme potentiellement utile pour les estimateurs de Bayes par rapport aux lois a priori $\pi(\theta, \sigma)$ telles que la loi de $\theta\,|\sigma$ est uniforme sur la sphère de rayon $m\sigma$ et σ suit les mesures $\frac{1}{\sigma^{1-l}} 1_{]0,\infty[}(\sigma)$ pour $l < k + p + 2$.

AMS 2010 subject classifications. 62C20, 62C15, 62C10.

1. Département de mathématiques, Université de Sherbrooke, Canada.

Mots-clés : Estimateur équivariant, Estimateur de Bayes, coefficient de variation, fonction hypergéométrique confluente, paramètres restreints, loi normale multidimensionnelle, coût quadratique.

5.1 Introduction.

Dans cet article, nous développons le problème d'estimation sous coût quadratique de la moyenne d'une distribution $p-$variée $\mathcal{N}_p(\theta, \sigma^2 I_p)$ avec $p \geq 2$, où les deux paramètres θ et σ sont inconnus et contraints à la condition $\frac{\|\theta\|}{\sigma} \leq m$, pour une certaine constante $m > 0$. Notre but consiste à estimer paramètre θ, en nous basant sur les réalisations de la statistique (X, S^2) issue du modèle canonique : $X \sim N(\theta, \sigma^2 I_p)$, $S^2 \sim \sigma^2 \chi_k^2(0)$, où $\chi_k^2(0)$ est la loi de khi deux centrée avec k degrés de liberté et X et S^2 indépendantes. Notamment, l'objectif est de proposer des estimateurs qui dominent l'estimateur sans biais X, ainsi que l'estimateur du maximum de vraisemblance δ_{emv}.

Dans le cas où le paramètre d'échelle σ est connu, Marchand et Strawderman [MS05a] ont montré que l'estimateur de Bayes δ_U associé à la loi a priori uniforme sur $[-m\sigma, m\sigma]$ domine l'estimateur X, pour tout $m > 0$. Une extension des travaux de Marchand et Strawderman [MS05a] a été faite par Kubokawa [Kub05] en supposant que la variance est inconnue. Cette extension représente un développement de notre problème pour $p = 1$. Il a développé une classe d'estimateurs qui dominent l'estimateur sans biais X. Cette classe d'estimateurs contient entre autres l'estimateur de Bayes δ_U par rapport à la loi a priori uniforme sur $\{(\theta, \sigma) : |\theta| \leq m\sigma, \sigma > 0\}$, ainsi que d'autres estimateurs tronqués. Par ailleurs, il a montré que l'estimateur de Bayes δ_{BU} par rapport à la loi a priori uniforme sur $\{(\theta, \sigma) : |\theta| = m\sigma, \sigma > 0\}$

domine X seulement lorsque $m \leq 1$.

D'autre part, Marchand et Perron [MP01] ont étudié ce problème dans le cas où le paramètre d'échelle est connu. Ils ont construit de vastes classes d'estimateurs dominant l'estimateur δ_{emv}, en développant des techniques basées sur la décomposition en risques conditionnels. En particulier, ils ont montré que l'estimateur de Bayes δ_{BU} par rapport à la loi uniforme sur la sphère centrée à l'origine et de rayon m domine δ_{emv} si $m/\sigma \leq \sqrt{p}$. D'autres estimateurs de Bayes dominant δ_{emv} pour des valeurs de m relativement petites, incluant l'estimateur de Bayes δ_U par rapport à la loi uniforme sur la boule centrée à l'origine et de rayon m, ont été donnés par Marchand et Perron [MP01].

Ensuite, Kariya, Giri et Perron [KGP88] ont déterminé le meilleur estimateur équivariant (MEE) pour ce problème avec la contrainte $\frac{\|\theta\|}{\sigma} \doteq \lambda$, pour une certaine constante fixée λ. Sous cette contrainte, l'estimateur du maximum de vraisemblance, δ_{emv}, est équivariant et par conséquent le MEE domine δ_{emv} dans ce cas.

Dans la section 5.2, nous abordons le problème d'estimation de θ sous la contrainte $\frac{\|\theta\|}{\sigma} \leq m$, en construisant des estimateurs au moins aussi bons que les estimateurs classiques de θ, à savoir l'estimateur sans biais X et l'estimateur du maximum de vraisemblance δ_{emv}. Des candidats intéressants à considérer, sont les estimateurs équivariants par rapport aux transformations, sur le groupe multiplicatif $G = \mathbb{R}_+ \times O_p(\mathbb{R})$, où $O_p(\mathbb{R})$ est le sous-groupe des matrices orthogonales d'ordre p, $(X, S^2) \to (bAX, b^2S^2)$. En particulier, nous considérons le meilleur estimateur équivariant (MEE), lorsque $\frac{\|\theta\|}{\sigma} = \lambda$, noté par δ_λ pour $0 \leq \lambda \leq m$. Pour cela, nous utilisons les techniques de calcul développées par Kariya, Giri et Perron [KGP88] et Marchand [Mar94] pour déterminer la forme explicite des estimateurs δ_λ. Ainsi, nous avons $\delta_\lambda(x,s) = h(t)\,x$, où $t = \|x\|^2/s^2$ et la fonction générative $h(\cdot)$ s'exprime en termes de fonctions hypergéométriques confluentes. Ensuite, cette écriture va être très utile pour le développement de plusieurs propriétés analytiques liées aux comportement des estimateurs δ_λ, δ_{emv}, en se basant sur des propriétés du rapport de deux fonctions hypergéométrique confluentes ${}_1F_1$.

Dans la section 5.3, nous établissons un résultat de dominance, le théorème 5.3.1, basé sur des conditions suffisantes sur les fonctions $h(\cdot)$ génératives des estimateurs $\delta_h(X,S) = h(t)X$, où $t = \|X\|^2/S^2$. Les propriétés analytiques des fonctions $h(\cdot)$ développées dans la section précédente, nous permettent d'établir des résultats de dominance des estimateurs X et δ_{emv}. En premier lieu, nous montrons que δ_{emv} domine X. Ensuite, nous montrons que le meilleur estimateur équivariant δ_m domine δ_{emv} lorsque $m \leq \sqrt{p}$, et que l'estimateur tronqué $\delta_{h_{emv}\wedge h_m}$ domine δ_{emv} pour tout (m,p). De plus, nous montrons que δ_m domine tous les estimateurs δ_h tels que $h_m \leq h$ pour tout (m,p). Notons que, le résultat de dominance de δ_{emv} par δ_m, qui est aussi un estimateur de Bayes (voir [Per87]), représente une extension du résultat de Kubokawa [Kub05] au cas multidimensionnel et aussi du résultat de Marchand Perron [MP01] au cas où la variance est connue. Dans la section 5.4, nous donnons une conclusion et des perspectives futures. Dans la section annexe, nous donnons la forme explicite et quelques propriétés d'estimateurs de Bayes potentiellement utiles pour déterminer d'autres résultats de dominance.

5.2 Définitions et préliminaires.

Nous considérons le problème d'estimation du paramètre θ sous la contrainte $\frac{\|\theta\|}{\sigma} \leq m$, pour une certaine constante positive m. Pour cela, nous traitons le modèle canonique suivant :

$$X \sim N(\theta, \sigma^2 I_p), \quad S^2 \sim \sigma^2 \chi^2_k(0), \quad X \text{ et } S^2 \text{ indépendantes}, \qquad (5.2.1)$$

où $p \geq 1$ et $k > 0$. Nous admettons aussi que les paramètres inconnus (θ, σ) sont restreints au sous-ensemble de $\mathbb{R} \times \mathbb{R}_+$:

$$\Theta(m) = \left\{ (\theta, \sigma) : \frac{\|\theta\|}{\sigma} \leq m\,, 0 < \sigma < \infty \right\}, \qquad (5.2.2)$$

où m est une certaine constante positive connue.

Ce modèle s'applique pour le modèle linéaire $Y = Z\beta + \epsilon$ avec la contrainte $\|\beta\| \leq m\sigma$, où Y est un vecteur de dimension n, Z une matrice de dimension (n,p) telle que $Z'Z = I_p$ et ϵ un vecteur aléatoire de dimension n distribué selon la loi $\mathcal{N}(0, \sigma^2 I_n)$ avec σ inconnu. La statistique exhaustive du problème $(\hat{\beta}, W)$ est formée par $\hat{\beta} = (Z'Z)^{-1} Z'Y = Z'Y$ et $W = (Y - Z\hat{\beta})'(Y - Z\hat{\beta})$ avec $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2 I_n)$, $W \sim \sigma^2 \chi^2_{n-p}(0)$, $\hat{\beta}$ et W indépendantes et le paramètre à estimer est β.

Sous la fonction de perte $L\left((\theta, \sigma), d\right) = \|d - \theta\|^2 / \sigma^2$, l'objectif est de proposer de bons estimateurs $\delta(X, S)$ du paramètre θ améliorant l'estimateur X et l'estimateur du maximum de vraisemblance δ_{emv}, en utilisant l'information donnée par la contrainte et l'évaluation de tels estimateurs se fait par le biais du risque

$$R((\theta, \sigma), \delta) = \frac{1}{\sigma^2} E_{\theta,\sigma}\left[\|\delta(X, S) - \theta\|^2\right], \text{ pour } (\theta, \sigma) \in \Theta(m)\,. \quad (5.2.3)$$

Dans la prochaine section, nous développons des expressions explicites de quelques estimateurs efficaces candidats pour une meilleure estimation de θ.

Nous développons aussi, pour chaque estimateur, des propriétés analytiques utiles pour comparer ces estimateurs entre eux.

5.2.1 Meilleur estimateur équivariant lorsque $\frac{\|\theta\|}{\sigma} = \lambda$.

Le problème est invariant par rapport au groupe $G = \mathbb{R}_+ \times O_p(\mathbb{R})$, où $O_p(\mathbb{R})$ est le sous-groupe multiplicatif des matrices orthogonales d'ordre p. La transformation induite par G sur la statistique exhaustive (X, S^2), pour un élément $g = (b, A) \in G$, est donnée par :

$$(X, S^2) \to (bAX, b^2S^2).$$

La forme générale des estimateurs équivariants (non aléatoires) a été donnée par Kariya et al. [KGP88]. Nous donnons ce résultat dans le lemme suivant.

Lemme 5.2.1. ***(Kariya, Giri et Perron, 1988).*** *Un estimateur δ, basé sur (X, S^2), est équivariant si et seulement s'il existe une fonction mesurable $h(\cdot)$ telle que $\delta(x, s^2) = h\left(\frac{\|x\|^2}{s^2}\right) x$ pour tout $(x, s^2) \in \mathbb{R}^p \times \mathbb{R}_+$.*

Avec $\frac{\|\theta\|}{\sigma}$ un maximal invariant, le risque d'un tel estimateur équivariant sera constant sur les orbites $\left\{(\theta, \sigma) : \frac{\|\theta\|}{\sigma} = \lambda\right\}$. Par ailleurs, cette écriture nous permet de représenter le risque de tout estimateur équivariant. Notons par δ_h tout estimateur équivariant de la forme $\delta_h(X, S^2) = h(T)X$ avec $T = \|X\|^2 / S^2$.

Lemme 5.2.2. *Le risque de tout estimateur équivariant δ_h, lorsque $\frac{\|\theta\|}{\sigma} = \lambda$, est donné par :*

$$R((\theta, \sigma), \delta_h) = \lambda^2 + E^T[g(T)].$$

où $T = X'X/S^2$ *et*

$$g(T) = h^2(T)\, E_{\theta,\sigma}\left(\frac{X'X}{\sigma^2}\middle| T\right) - 2h(T)\, E_{\theta,\sigma}\left(\frac{\theta' X}{\sigma^2}\middle| T\right). \quad (5.2.4)$$

Démonstration. En utilisant (5.2.3) et en conditionnant par rapport à T, la fonction de risque $R((\theta,\sigma),\delta_h)$ de tout estimateur équivariant δ_h s'écrit :

$$\begin{aligned} R((\theta,\sigma),\delta_h) &= \frac{\theta'\theta}{\sigma^2} + \frac{1}{\sigma^2} E_{\theta,\sigma}\left[h^2\left(\frac{\|X\|^2}{S^2}\right) X'X - 2h\left(\frac{\|X\|^2}{S^2}\right)\theta' X\right] \\ &= \lambda^2 + E^T\left[g(T)\right]. \quad \square \end{aligned}$$

Maintenant, pour minimiser le risque par rapport aux fonctions $h_\lambda(\cdot)$ afin de trouver le meilleur estimateur équivariant pour $\lambda = \frac{\|\theta\|}{\sigma}$ fixé, il suffit de minimiser la fonction donnée par (5.2.4), $g(t)$ pour tout $t > 0$. Le choix optimal de h est noté h_λ et il est donné par le résultat suivant.

Lemme 5.2.3. *Le meilleur estimateur équivariant, du paramètre θ tel que $\frac{\|\theta\|}{\sigma} = \lambda$, est de la forme $\delta_\lambda(X,S) = h_\lambda(t)X$ avec :*

$$h_\lambda(t) = \frac{E_{\theta,\sigma}\left[\theta' X \middle| \frac{\|X\|^2}{S^2} = t\right]}{E_{\theta,\sigma}\left[X'X \middle| \frac{\|X\|^2}{S^2} = t\right]}. \quad (5.2.5)$$

Ce lemme nous permet d'obtenir la forme explicite du meilleur estimateur équivariant, lorsque $\frac{\|\theta\|}{\sigma} = \lambda$, selon le modèle (5.2.1).

Lemme 5.2.4. ***(Kariya, Giri et Perron, 1988)***. *Pour $\frac{\|\theta\|}{\sigma} = \lambda$, le meilleur estimateur équivariant, $\delta_\lambda(x,s) = h_\lambda(t)\,x$, est déterminé par :*

$$h_\lambda(t) = \frac{\lambda^2}{p}\frac{F\left(\frac{k+p}{2}+1, \frac{p}{2}+1, \frac{\lambda^2 t}{2(1+t)}\right)}{F\left(\frac{k+p}{2}+1, \frac{p}{2}, \frac{\lambda^2 t}{2(1+t)}\right)}, \quad (5.2.6)$$

pour tout $0 \le \lambda \le m$, *où* $F(a,b,z) = \sum_{i=0}^{\infty}\frac{(a)_i}{(b)_i}\frac{z^i}{i!}$, $z \in \mathbb{R}$, *avec* $(c)_i = \prod_{j=0}^{i-1}(c+j)$ *pour* $i = 1, 2, \ldots$ *et* $(c)_0 = 1$.

Démonstration. Puisque δ_λ a un risque constant pour $\frac{\|\theta\|}{\sigma} = \lambda$, pour déterminer h_λ il suffit de calculer les espérances dans l'équation (5.2.5) pour $\theta_0 = (\lambda, 0, ..., 0)$ et $\sigma_0 = 1$. Nous avons, en prenant $\theta_0' x = \lambda x_1$ et $\sigma_0 = 1$:

$$h_\lambda(t) = \lambda \frac{\int_{\mathbf{R}^p} x_1 f_{(X,T)}(x,t)\, dx}{\int_{\mathbf{R}^p} \|x\|^2 f_{(X,T)}(x,t)\, dx}, \tag{5.2.7}$$

avec le changement de variable $t = \frac{\|x\|^2}{s^2}$. La fonction de densité du vecteur (X,T) pour tout $t > 0$, est donnée par :

$$\begin{aligned} f_{(X,T)}(x,t) &= k \exp\left[-\frac{1}{2}\left(\frac{1+t}{t}\|x\|^2 + \lambda^2 - 2\lambda x_1\right)\right] \|x\|^k t^{-k/2-1} \\ &= k \exp\left[-\frac{1}{2}\left(\frac{1+t}{t}\|x\|^2 + \lambda^2\right)\right] \|x\|^k t^{-k/2-1} \sum_{i \geq 0} \frac{(\lambda x_1)^i}{i!}, \end{aligned}$$

où k est la constante de normalisation. Maintenant, en utilisant le changement de variables aux coordonnées polaires :

$$x_1 = r^{1/2} sin\,(\varphi_1)$$

$$x_j = r^{1/2} \Pi_{i=1}^{j-1} cos\,(\varphi_i)\, sin\,(\varphi_j) \quad \text{pour} \quad 1 < j < p$$

$$x_p = r^{1/2} \Pi_{i=1}^{p-1} cos\,(\varphi_i)\ ,$$

dont le jacobien est donné par :

$$J = \frac{1}{2}\, r^{p/2-1}\, \Pi_{j=1}^{p-2} cos^{p-1-j}\,(\varphi_j)$$

et le domaine d'intégration de $(r, \varphi_1, \ldots, \varphi_{p-1})$ est $\mathbb{R}_+ \times [-\pi/2, \pi/2]^{p-2} \times$

$[0, 2\pi]$, et en utilisant la relation suivante :

$$\int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} \int_{0}^{2\pi} \frac{(\sin(\theta_1))^{j+1} \prod_{l=1}^{p-2} (\cos(\theta_l))^{p-l-1}}{2\left(\Gamma(\frac{1}{2})\right)^{p-1}} \, d\theta_{p-1} \ldots d\theta_1$$

$$= \begin{cases} \frac{\Gamma(\frac{j}{2}+1)}{\Gamma(\frac{j+p+1}{2})} & \text{si } j \text{ est impair, } j \geq -1 \\ 0 & \text{si } j \text{ est pair, } j \geq 0, \end{cases}$$

et les identités suivantes :

$$(2l+1)!\, \Gamma(1/2) = 2^{2l+1}\, l!\, \Gamma(l+3/2)\,, \ (l \text{ entier positif})$$

et

$$(2l)!\, \Gamma(1/2) = 2^{2l}\, l!\, \Gamma(l+1/2)\,, \ (l \text{ entier positif})$$

nous obtenons :

$$\begin{aligned} h_\lambda(t) &= \frac{\lambda^2}{2} \frac{\sum_{i \geq 0} \frac{\Gamma(\frac{k+p}{2}+i+1)}{\Gamma(\frac{p}{2}+i+1)\, i!} \left[\frac{\lambda^2 t}{2(1+t)}\right]^i}{\sum_{i \geq 0} \frac{\Gamma(\frac{k+p}{2}+i+1)}{\Gamma(\frac{p}{2}+i)\, i!} \left[\frac{\lambda^2 t}{2(1+t)}\right]^i} \\ &= \frac{\lambda^2}{p} \frac{F\left(\frac{k+p}{2}+1, \frac{p}{2}+1, \frac{\lambda^2 t}{2(1+t)}\right)}{F\left(\frac{k+p}{2}+1, \frac{p}{2}, \frac{\lambda^2 t}{2(1+t)}\right)}. \quad \square \end{aligned}$$

Le lemme suivant nous sera utile.

Lemme 5.2.5. *Pour tout $a > 0$, $b > 0$ et $c = 0, 1$, La fonction*

$$K^{a,b,c}(z) = \frac{F(a-c+1, b-c+1, z)}{F(a+1, b, z)}$$

est décroissante en z, $z > 0$.

Démonstration. Notons que, pour tout $z > 0$,

$$K^{a,b,c}(z) = (ca + (1-c)b)\, E_z^{a,b}\left[\frac{1}{(ca+(1-c)b)+L}\right],$$

où l'espérance $E_z^{a,b}$est calculée par rapport à la densité $f_z^{a,b}(l) \propto \frac{(a+1)_l}{(b)_l} \frac{z^l}{l!}$ $1_{\{0,1,\ldots\}}(l)$ qui possède la propriété du rapport de vraisemblance monotone croissant par rapport au paramètre z. Ensuite, comme la fonction $\frac{1}{(ca+(1-c)b)+l}$ est décroissante en l pour $l \in \{0,1,\ldots\}$, alors $K^{a,b,c}(z)$ est décroissante en z. □

En combinant les lemmes 5.2.4 et 5.2.5, nous obtenons les propriétés suivantes de h_λ.

Lemme 5.2.6. *La fonction $h_\lambda(\cdot)$ possède les propriétés suivantes*

(a) $h_\lambda(\cdot)$ *est décroissante sur* $(0, \infty)$, *pour tout* λ ;

(b) $\sup_{t\geq 0} h_\lambda(t) = h_\lambda(0) = \frac{\lambda^2}{p}$;

(c) $\inf_{t\geq 0} h_\lambda(t) = \frac{\lambda^2}{p} \frac{F\left(\frac{k+p}{2}+1, \frac{p}{2}+1, \frac{\lambda^2}{2}\right)}{F\left(\frac{k+p}{2}+1, \frac{p}{2}, \frac{\lambda^2}{2}\right)}$.

Démonstration. Les parties (b) et (c) proviennent d'évaluations directes de (5.2.6) et de la partie (a). La partie (a) découle directement du lemme 5.2.5. □

Ensuite, en utilisant quelques propriétés de la fonction hypergéométrique confluente, nous avons obtenu d'autres propriétés de la fonction $h_\lambda(t)$ par rapport au paramètre λ. D'où le lemme suivant, très utile pour nos résultats de dominance.

Lemme 5.2.7. *La fonction $h_\lambda(t)$ possède les propriétés suivantes*

(a) $h_\lambda(t)$ *est croissante en* λ *pour tout* $t > 0$;

(b) $\lim_{\lambda\to\infty} h_\lambda(t) = \frac{1+t}{t}$, *pour tout* $t > 0$.

Démonstration. Pour la première propriété, nous utilisons la relation de récurrence suivante (voir [AS64]) :

$$wF(a+1, b+1, w) = bF(a+1, b, w) - bF(a, b, w)$$

pour $a, b, w > 0$. Ceci nous permet d'écrire :

$$h_\lambda(t) = \frac{1+t}{t}\left(1 - \frac{F(\frac{p+k}{2}, \frac{p}{2}, w)}{F(\frac{p+k}{2}+1, \frac{p}{2}, w)}\right),$$

et le résultat découle du lemme 5.2.5.

Pour la seconde propriété, il faut noter que pour z grand, la fonction F peut s'écrire selon [AS64] comme suit :

$$F(a,b,z) = \frac{\Gamma(b)}{\Gamma(a)}\exp(z)z^{a-b}\left(1 + O\left(\frac{1}{z}\right)\right), \tag{5.2.8}$$

où $O\left(\frac{1}{z}\right)$ est une fonction bornée par $\frac{1}{z}$ au voisinage de l'infini. D'où la propriété. □

5.2.2 Estimateur du maximum de vraisemblance pour $\Theta(m)$.

Dans le même contexte, nous déterminons l'estimateur du maximum de vraisemblance δ_{emv}. Le résultat est donné dans le lemme suivant.

Lemme 5.2.8. *Pour le modèle* (5.2.1) *avec* $\|\theta\|/\sigma \leq m$, *l'estimateur du maximum de vraisemblance,* δ_{emv}, *est déterminé par :*

$$\delta_{emv}(X,S) = \left(\frac{m^2}{2(p+k)}\left(\sqrt{1+4\frac{(k+p)}{m^2}\frac{1+T}{T}} - 1\right)\wedge 1\right)X, \tag{5.2.9}$$

où $T = \frac{\|X\|^2}{S^2}$.

Démonstration. Selon le modèle (5.2.1), le logarithme de la vraisemblance est donné par :

$$\ln L(\theta,\sigma) \propto -\frac{1}{2\sigma^2}\left(\|X-\theta\|^2 + S^2\right) - \frac{p+k}{2}\ln\sigma^2.$$

Pour une valeur fixe de σ, la vraisemblance par rapport à θ est maximisée pour :

$$\hat{\theta}_\sigma = \begin{cases} X & \text{si} \quad \|X\| \leq m\sigma \, ; \\ m\sigma\frac{X}{\|X\|} & \text{si} \quad \|X\| \geq m\sigma \, . \end{cases}$$

Par conséquent, nous avons bien :

$$\sup_{(\theta,\sigma)\in\Theta(m)} L(\theta,\sigma) = \sup_{\sigma>0} L\left(\hat{\theta}_\sigma . \sigma\right) \, .$$

Ensuite, nous voyons que :

$$\begin{aligned} -\ln L\left(\hat{\theta}_\sigma,\sigma\right) &\propto \frac{1}{2\sigma^2}\left(\left\|X-\hat{\theta}_\sigma\right\|^2 + S^2\right) + \frac{p+k}{2}\ln\sigma^2 \\ &= \begin{cases} \frac{S^2}{2\sigma^2} + \frac{p+k}{2}\ln\sigma^2 & \text{si} \quad \|X\| \leq m\sigma \, ; \\ \frac{S^2}{2\sigma^2} + \frac{p+k}{2}\ln\sigma^2 + \frac{1}{2}\left(m^2 + \frac{\|X\|^2}{\sigma^2} - 2m\frac{\|X\|}{\sigma}\right) & \text{si} \quad \|X\| \geq m\sigma \, . \end{cases} \end{aligned}$$

Maintenant, si le minimum de $-\ln L\left(\hat{\theta}_\sigma,\sigma\right)$ est atteint sur $\left(\frac{\|X\|}{m},\infty\right)$, alors il sera atteint au point $\hat{\sigma} = \sqrt{\frac{S^2}{p+k}}$. Ceci est possible dès que $\frac{S^2}{p+k} \geq \frac{X'X}{m^2}$. Par conséquent, $\hat{\theta}_{emv} = X$ lorsque $\frac{S^2}{p+k} \geq \frac{X'X}{m^2}$.

Par ailleurs, lorsque $\frac{S^2}{p+k} < \frac{X'X}{m^2}$, le minimum de $-\ln L\left(\hat{\theta}_\sigma,\sigma\right)$ est atteint au point $\hat{\sigma_0} = \frac{m\,\|X\|}{2(p+k)}\left(\sqrt{1+4\frac{(k+p)}{m^2}\left(\frac{S^2}{\|X\|^2}+1\right)}-1\right)$. Or, comme $\hat{\theta}_{emv} = m\frac{X}{\|X\|}\hat{\sigma_0}$, alors nous obtenons le résultat cité ci-dessus. □

Il faut noter que $\|\delta_{emv}(X,S)\| \leq \|X\|$ avec probabilité un. D'autre part, nous remarquons que la fonction $h_{emv}(t)$ est décroissante par rapport à t, croissante par rapport à m et que

$$\lim_{t\to\infty} h_{emv}(t) = \frac{\sqrt{1+4\gamma}-1}{2\gamma} \, ,$$

sachant que $\gamma = \frac{p+k}{m^2} > 0$. De plus, nous avons les résultats suivants qui servent à comparer le meilleur estimateur équivariant avec l'estimateur du maximum de vraisemblance.

Lemme 5.2.9. *(a) Pour tout* $m > 0$ *et tout* $t \geq \frac{m^2}{p+k}$*, on a* $h_m(t) \leq h_{emv}(t)$.

(b) Pour tout $m \leq \sqrt{p}$ *et tout* $t \leq \frac{m^2}{p+k}$*, on a* $h_m(t) \leq h_{emv}(t) = h_{SB}(t) = 1$ *avec égalité si et seulement si* $t = 0$ *et* $m = \sqrt{p}$.

Démonstration. En vertu de (5.2.9) et (5.2.6), il suffit de vérifier que

$$\frac{F\left(\frac{k+p}{2}+1, \frac{p}{2}+1, z\right)}{F\left(\frac{k+p}{2}+1, \frac{p}{2}, z\right)} \leq \frac{p}{2(p+k)}\left(\sqrt{1+\frac{2(k+p)}{z}}-1\right) \tag{5.2.10}$$

où $z = \frac{m^2 t}{2(1+t)}$ et donc pour $z \geq \frac{m^4}{2(m^2+p+k)}$. L'inégalité (5.2.10) est équivalente à

$$\frac{E_z\left[\frac{(a)_L}{(b+1)_L}\right]}{E_z\left[\frac{(a)_L}{(b)_L}\right]} \leq \frac{p}{2(p+k)}\left(\sqrt{1+\frac{2(k+p)}{z}}-1\right), \tag{5.2.11}$$

où $a = \frac{k+p}{2}+1$, $b = \frac{p}{2}$, et où l'espérance E_z est calculé par rapport à la loi $L \sim Poisson(z)$. Or, en utilisant l'inégalité de covariance, nous pouvons remarquer que

$$\begin{aligned} E_z\left[\frac{(a)_L}{(b)_L}\right] &= E_z\left[\frac{(a)_L}{(b+1)_L}\frac{b+L}{b}\right] \\ &\geq E_z\left[\frac{(a)_L}{(b+1)_L}\right] E_z\left[\frac{b+L}{b}\right] \\ &= E_z\left[\frac{(a)_L}{(b+1)_L}\right]\frac{b+z}{b}, \end{aligned} \tag{5.2.12}$$

puisque les fonctions $\frac{(a)_l}{(b+1)_l}$ et $\frac{b+l}{b}$ sont croissantes en l ; $l = 0, 1, \ldots$, sachant que $a \geq b+1$. L'inégalité (5.2.11) est alors satisfaite si

$$\frac{1}{\frac{p}{2}+z} \leq \frac{1}{p+k}\left(\sqrt{1+\frac{2(k+p)}{z}}-1\right), \tag{5.2.13}$$

qui est équivalente à

$$\left(\frac{1}{\frac{p}{2}+z}+\frac{1}{p+k}\right)^2 \leq 1+\frac{2(k+p)}{z}. \tag{5.2.14}$$

Il est clair qu'il suffit que l'inégalité (5.2.14) soit satisfaite pour $p = 1$ et $k = 1$. Finalement, nous vérifions aisément que (5.2.14) tient pour $p = 1$, $k = 1$, d'où le résultat. □

5.3 Résultats de dominance.

Dans cette section, nous utiliserons les propriétés analytiques des fonctions $h(\cdot)$ génératives des estimateurs développés dans la section précédente, afin d'établir des résultats de dominance. Ainsi, nous construisons des classes d'estimateurs améliorant les estimateurs X et δ_{emv}.

La croissance de $h_\lambda(t)$ en fonction de λ, pour tout $t > 0$, nous permet d'avoir un premier résultat fondamental pour la suite. Ce résultat est donné par le théorème suivant.

Théorème 5.3.1. *Par rapport au modèle* (5.2.1) *et la contrainte* (5.2.2), *si* $h_1(t) \leq h_2(t)$ *et* $\frac{h_2(t)+h_1(t)}{2} \geq h_m(t)$ *pour tout* $t > 0$ *tel que* $h_2(t) \neq h_1(t)$, *alors l'estimateur* $\delta_{h_1}(x,s) = h_1(t)\,x$ *domine l'estimateur* $\delta_{h_2}(x,s) = h_2(t)\,x$.

Démonstration. En utilisant l'équation (5.2.5) et en posant

$$b(t) = E_{\theta,\sigma}\left[\frac{\theta' X}{\sigma^2}\,\middle|\,\frac{\|X\|^2}{S^2} = t\right] \text{ et } a(t) = E_{\theta,\sigma}\left[\frac{X'X}{\sigma^2}\,\middle|\,\frac{\|X\|^2}{S^2} = t\right],$$

nous pouvons réécrire le risque d'un estimateur équivariant quelconque, $\delta_h(X,S) = h(t)X$ pour $\lambda = \frac{\|\theta\|}{\sigma}$, comme suit :

$$R((\theta,\sigma),\delta_h) = \lambda^2 + E_\lambda\left[a(T)\left(h(T) - h_\lambda(T)\right)^2 - h_\lambda^2(T)\right]. \quad (5.3.1)$$

Ceci nous permet de calculer la différence des risques entre les estimateurs équivariants δ_{h_1} et δ_{h_2}. Ce qui donne :

$$\begin{aligned} R((\theta.\sigma),\delta_{h_1}) - R((\theta,\sigma),\delta_{h_2}) &= E_\lambda\left[a(T)\left(h_1(T) - h_2(T)\right)\right. \\ &\qquad \left. \times \left(h_1(T) + h_2(T) - 2h_\lambda(T)\right)\right] \\ &= E_\lambda\left[\Delta_\lambda(T)\right]. \end{aligned} \tag{5.3.2}$$

En combinant cette égalité avec la partie (a) du lemme 5.2.7, nous voyons sous les conditions du théorème que $\Delta_\lambda(T) \leq 0$ avec probabilité un. □

Le lemme 5.2.9 nous permet, au travers une application directe du théorème 5.3.1 à l'estimateur sans biais $\delta_{SB}(X,S) = X$, à l'estimateur du maximum de vraisemblance $\delta_{emv}(X,S)$ et à l'estimateur de Bayésien $\delta_m(X,S)$ d'avoir le résultat principal suivant.

Corollaire 5.3.2. *(a) Pour tout (m,p), l'estimateur du maximum de vraisemblance, δ_{emv}, domine l'estimateur sans biais X.*

(b) Pour tout (m,p) tels que $m \leq \sqrt{p}$, δ_m domine δ_{emv} et δ_{SB}.

Par ailleurs, nous pouvons construire d'autres estimateurs dominant les estimateurs classiques pour tout (m,p). Le résultat suivant nous permet de construire de tels estimateurs en se basant sur la troncature.

Corollaire 5.3.3. *(a) Sous le modèle (5.2.1), l'estimateur $\delta_{h\wedge h_m}$ domine δ_h tant et aussi longtemps que $\nu\{t > 0 : h_m(t) < h(t)\} > 0$, où ν est la mesure de Lebesgue sur $\mathbb{R}$.*

(b) Si $h_m \leq h$, alors l'estimateur δ_m domine δ_h tant et aussi longtemps que $\nu\{t > 0 : h_m(t) < h(t)\} > 0$, où ν est la mesure de Lebesgue sur $\mathbb{R}$.

Démonstration. La partie (b) découle de la partie (a). Pour la partie (a), l'expression de la différence des risques (5.3.2) entre les estimateurs $\delta_{h\wedge h_m}$ et

δ_h devient

$$R((\theta,\sigma),\delta_{h\wedge h_m}) - R((\theta,\sigma),\delta_h) = E_\lambda \left[a(T)\left(h_m(T) - h(T)\right)\right.$$
$$\left. \times \left(h_m(T) + h(T) - 2h_\lambda(T)\right) 1_{\{t>0:h_m(t)<h(t)\}}\left(T\right)\right] .$$

De plus, $\Delta_\lambda(T) < 0$ avec probabilité un dans ce cas. □

Remarque. Si δ_h est telle que $h > 0$ et $h(0) < m^2/p$, alors $\delta_{h'}$ avec $h' = h\wedge h_m$ n'est pas dominé par δ_m. Pour montrer cette propriété, il suffit de poser $h_1 = h'$, $h_2 = h_m$ et $\lambda = 0$ dans (5.3.2) et d'observer que la différence des risques donnée par :

$$R((0,\sigma),\delta_{h'}) - R((0,\sigma),\delta_{h_m}) = E_0\left[a(T)\left(h'^2(T) - h_m^2(T)\right)\right]$$

est négative pour tout $\sigma > 0$. Par conséquent, tous les estimateurs linéaires tronqués $\delta_{a\wedge h_m}(X) = (a \wedge h_m(t))\, X$ avec $a < m^2/p$ ne sont pas dominés par δ_m. En particulier, l'estimateur sans biais tronqué $= (1 \wedge h_m(t))\, X$ n'est pas dominé par δ_m pour $m^2 > p$. De même, il faut noter que la troncature de δ_m avec $\delta_{lx}(X,S) = \frac{m^2}{m^2+p}X$, le meilleur estimateur au point de vu du minimax de la forme aX dominant X, n'est pas dominé par δ_m pour tout (m,p).

Une autre application du théorème 5.3.3, nous permet d'avoir les résultats suivants. Puisque $h_{SB}(t) \geq h_{emv}(t) \geq h_m(t)$ avec des inégalités strictes pour tout $t > t_0$ (t_0 donné par le lemme 5.2.9).

Corollaire 5.3.4. *Pour tout $m > 0$ et pour tout $p \geq 1$, nous avons*

(a) $\delta_{h_{emv}\wedge h_m}$ domine $\delta_{h_{emv}}$.

(a) $\delta_{h_{SB}\wedge h_m}$ domine X.

5.4 Conclusion.

Dans la section 5.2, nous avons développé les formes explicites de quelques estimateurs du paramètre θ d'une distribution normale multidimensionnelle ($p \geq 2$), sous la contrainte $\frac{\|\theta\|}{\sigma} \leq m$. La comparaison entre les fonctions génératives de ces estimateurs, nous a permet de développé des résultats de dominance. En particulier, nous avons montré que le meilleur estimateur équivariant δ_m domine l'estimateur δ_{emv} lorsque $m \leq \sqrt{p}$. Ce dernier résultat représente une extension, au cas multidimensionnel du résultat de Kubokawa [Kub05] et au cas de la variance connue de Marchand et Perron [MP01]. Nous avons aussi pu montrer que l'estimateur $\delta_m(x, s) = h_m(t)\, x$ domine tous les estimateurs $\delta_h(x, s) = h(t)\, x$ tels que $h_m(t) \leq h(t)$ pour tout $t \geq 0$. De plus, nous avons construit un estimateur qui domine δ_{emv} pour tout (m, p), en considérant la troncature de δ_m avec δ_{emv}.

Autres estimateurs intéressants à considérer sont les estimateurs de Bayes. Pour cela, nous avons considéré deux classes de lois a priori particulières représentées par les distributions dont les densités sont données par :

$$\pi_{BU}^{l}\left(\theta, \sigma^2\right) = 1_{S_{m\sigma}}(\theta)\, \sigma^{l-2}\, d\theta\, d\sigma^2$$

et

$$\pi_{U}^{l}\left(\theta, \sigma^2\right) = 1_{B_{m\sigma}}(\theta)\, \sigma^{l-2}\, d\theta\, d\sigma^2 .$$

Leurs estimateurs de Bayes associés sont notés respectivement par δ_{BU}^{l} et δ_{U}^{l}. Il reste à faire une étude plus fine selon les valeurs de l sur les estimateurs δ_{BU}^{l} et δ_{U}^{l} afin de déterminer une classe d'estimateurs de Bayes dominant les estimateurs classiques en fonction de (m, p). Pour cela, nous avons déterminé la forme générale explicite d'une certaine classe d'estimateurs de Bayes, en particulier celle des estimateurs δ^l_{BU} et δ^l_U, dans la section Annexe. Des travaux en cours visent à développer des techniques de risques conditionnels par rapport à $\frac{\|X\|^2}{S^2} \leq m^2$ et $\frac{\|X\|^2}{S^2} \geq m^2$ pour étendre les résultats de dominance. Autres perspectives futures consistent à reprendre les travaux de Marchand [Mar94] en considérant la contrainte $\theta'\Sigma^{-1}\theta \leq m$, où la matrice de variance covariance n'est pas nécessairement diagonale ou à étendre le modèle gaussien, en considérant des distributions admettant des densités de la forme $f\left(\|X-\theta\| + S^2\right)$ pour une certaine fonction mesurable $f(\cdot)$.

5.5 Annexe.

Voici des précisions et propriétés de différents estimateurs de Bayes. Dans le même contexte, nous déterminons les estimateurs de Bayes, δ_π, associés aux lois a priori définies sur l'espace $A = \{(\theta, \sigma) : \|\theta\| \leq m\sigma, 0 < \sigma < \infty\}$ et ayant des densités de la forme :

$$\pi.\left(\theta, \sigma^2\right) = \pi_\sigma(\theta)\, \sigma^{l-2}\, d\theta\, d\sigma^2\,, \quad l < k+p+2\,. \tag{5.5.1}$$

La forme de tels estimateurs est donnée dans le lemme suivant.

Lemme 5.5.1. *Les estimateurs de Bayes associés aux lois a priori $\pi\left(\theta, \sigma^2\right)$ du type (5.5.1), sous coût quadratique, sont donnés par :*

$$\delta_\pi(X,S) = X + \frac{\nabla_X \int_0^\infty \sigma^{l-k-p-2}\, e^{-\frac{S^2}{2\sigma^2}} M^\pi_m(X)\; d\sigma^2}{\int_0^\infty \sigma^{l-k-p-4}\, e^{-\frac{S^2}{2\sigma^2}} M^\pi_m(X)\; d\sigma^2}\,, \tag{5.5.2}$$

où ∇_X désigne le vecteur gradient par rapport à la variable X et

$$M^\pi_m(X) = \int_{B_{m\sigma}} e^{-\frac{1}{2\sigma^2}\|X-\theta\|^2} \pi_\sigma(\theta)\; d\theta\,. \tag{5.5.3}$$

Démonstration. Sous la fonction de perte $\frac{\|d-\theta\|^2}{\sigma^2}$ et le modèle (5.2.1), les estimateurs de Bayes associés aux lois a priori π du type (5.5.1) sont donnés par :

$$\begin{aligned}
\delta_\pi(X,S) &= \frac{E_\theta\left[\,\theta/\sigma^2\,|X,S\right]}{E_\theta\left[\,1/\sigma^2\,|X,S\right]} \\
&= \frac{\int_0^\infty \int_{B_{m\sigma}} \theta\; e^{-\frac{1}{2\sigma^2}\|X-\theta\|^2} e^{-\frac{S^2}{2\sigma^2}}\;\pi_\sigma(\theta)\,\sigma^{l-k-p-4}\,d\theta\,d\sigma^2}{\int_0^\infty \int_{B_{m\sigma}} e^{-\frac{1}{2\sigma^2}\|X-\theta\|^2} e^{-\frac{S^2}{2\sigma^2}}\;\pi_\sigma(\theta)\,\sigma^{l-k-p-4}\,d\theta\,d\sigma^2} \\
&= X + \frac{\int_0^\infty \sigma^{l-k-p-4}\; e^{-\frac{S^2}{2\sigma^2}} \int_{B_{m\sigma}} (\theta - X)\;\, e^{-\frac{1}{2\sigma^2}\|X-\theta\|^2}\pi_\sigma(\theta)\;d\theta\,d\sigma^2}{\int_0^\infty \sigma^{l-k-p-4}\; e^{-\frac{S^2}{2\sigma^2}} \int_{B_{m\sigma}} e^{-\frac{1}{2\sigma^2}\|X-\theta\|^2}\pi_\sigma(\theta)\;d\theta\,d\sigma^2}\,,
\end{aligned}$$

où $B_{m\sigma}$ est la boule centrée à l'origine et de rayon $m\sigma$. Il faut noter que ces estimateurs de Bayes existent pour tout $l < k+p+2$. Maintenant, en utilisant (5.5.3) nous pouvons voir que

$$\sigma^2\;\nabla_X M_m^\pi(X) = \int_{B_{m\sigma}} (\theta - X)\;\, e^{-\frac{1}{2\sigma^2}\|X-\theta\|^2}\pi_\sigma(\theta)\;d\theta\,. \tag{5.5.4}$$

D'où le résultat. $\square$

Par la suite, nous déterminons la forme explicite des estimateurs de Bayes correspondants à deux types de distributions a priori de la forme (5.5.1). La première classe de distributions est définie par :

$$\pi_{BU}^l\left(\theta,\sigma^2\right) = 1_{S_{m\sigma}}(\theta)\,\sigma^{l-2}\,d\theta\,d\sigma^2$$

et la seconde classe est définie par :

$$\pi_U^l\left(\theta,\sigma^2\right) = 1_{B_{m\sigma}}(\theta)\,\sigma^{l-2}\,d\theta\,d\sigma^2\,,$$

où $B_{m\sigma}$ et $S_{m\sigma}$ sont respectivement la boule et la sphère centrées à l'origine et de rayon $m\sigma$. Alors, nous obtenons les résultats suivants.

Théorème 5.5.2. *L'estimateur de Bayes, $\delta_{BU}^{l}(X,S) = h_m^l(t)X$ avec $t = \frac{\|X\|^2}{S^2}$, associé à la loi a priori π_{BU}^l est déterminé par :*

$$h_m^l(t) = \frac{m^2}{p} \frac{F\left(\frac{k+p-l}{2}+1, \frac{p}{2}+1, \frac{m^2 t}{2(1+t)}\right)}{F\left(\frac{k+p-l}{2}+1, \frac{p}{2}, \frac{m^2 t}{2(1+t)}\right)}, \tag{5.5.5}$$

où F est la fonction hypergéométrique confluente.

Nous donnons ici, un résultat donné par Fourdrinier et Wells [FW96], utile dans la Démonstration du théorème 5.5.2.

Lemme 5.5.3. *Pour $X \sim \mathcal{N}_p(\theta, I_p)$, on a*

$$\int_{S_r} e^{X'\theta} \, d\mathcal{U}_r(\theta) = \Gamma(p/2)\, 2^{(p-2)/2} \left(\|X\| \, r\right)^{(2-p)/2} I_{(p-2)/2}\left(\|X\| \, r\right),$$

où $\mathcal{U}_r$ est la distribution uniforme sur la sphère S_r de rayon r et $I_\nu(.)$ désigne la fonction de Bessel modifiée de degré ν.

Démonstration du théorème 5.5.2. Il suffit de calculer la fonction $M_m^\pi(X)$, ensuite déduire le résultat voulu à partir du lemme 5.5.1. Cependant, pour π_{BU}^l nous obtenons :

$$\begin{aligned} M_m^{BU}(X) &= \int_{S_{m\sigma}} e^{-\frac{1}{2\sigma^2}\|X-\theta\|^2} \, d\theta \qquad (5.5.6)\\ &= e^{-\frac{m^2}{2}} e^{-\frac{\|X\|^2}{2\sigma^2}} \int_{S_{m\sigma}} e^{\frac{X'}{\sigma}\frac{\theta}{\sigma}} \, d\theta \\ &= k\, m^{p-1}\, e^{-\frac{m^2}{2}} e^{-\frac{\|X\|^2}{2\sigma^2}} \int_{S_{m\sigma}} e^{\frac{X'}{\sigma}\frac{\theta}{\sigma}} \, d\mathcal{U}_r(\theta) \\ &= k_1\, m^{p/2}\, e^{-\frac{m^2}{2}}\, \sigma^{(p-2)/2}\, \|X\|^{(2-p)/2} e^{-\frac{\|X\|^2}{2\sigma^2}} I_{(p-2)/2}\left(m\|X\|/\sigma\right), \end{aligned}$$

pour une certaine constante positive k_1. Il faut noter que la dernière égalité est obtenue à partir du lemme 5.5.3. Maintenant, en dérivant par rapport à X et en utilisant l'identité suivante sur les fonctions de Bessel

$$\frac{d}{dy} I_\nu(y) = \frac{\nu}{y} I_\nu(y) + I_{\nu+1}(y).$$

nous obtenons :

$$\nabla_X M_m^{BU}(X) = -\frac{X}{\sigma^2} M_m^{BU}(X) + \\ k_1\ m^{p/2+1}\ e^{-\frac{m^2}{2}}\ \sigma^{p/2-2} \frac{X}{\|X\|^{p/2}}\ e^{-\frac{\|X\|^2}{2\sigma^2}} I_{p/2}(m\|X\|/\sigma)\,. \quad (5.5.7)$$

Cependant, en utilisant le lemme 5.5.1 nous obtenons selon (5.5.2) :

$$\delta_{BU}^l(X,S) = m\frac{X}{\|X\|} \frac{\int_0^\infty \sigma^{l-k-p/2-4}\ e^{-\frac{\|X\|^2+S^2}{2\sigma^2}} I_{p/2}(m\|X\|/\sigma)\, d\sigma^2}{\int_0^\infty \sigma^{l-k-p/2-5}\ e^{-\frac{\|X\|^2+S^2}{2\sigma^2}} I_{p/2-1}(m\|X\|/\sigma)\, d\sigma^2}\,.$$

Finalement, la relation suivante donnée par Abramowitz et Stegun [AS64]

$$F\left(a,c,\frac{u}{z}\right) = \frac{\Gamma(c)}{\Gamma(a)}\, z^a u^{\frac{1-c}{2}} \int_0^\infty y^{a-\frac{1}{2}-\frac{c}{2}} e^{-zy} I_{c-1}(2\sqrt{uy})\, dy\,, \quad (5.5.8)$$

nous permet de déduire le résultat voulu. □

Théorème 5.5.4. *L'estimateur de Bayes, δ_U^l, associé à la loi a priori π_U^l est donné par :*

$$\delta_U^l(X,S) = h_U^l(t)\, X = E\left[h_R^l(t) \,\middle|\, \frac{\|X\|^2}{S^2} = t\right] X\,, \quad (5.5.9)$$

où l'espérance E est calculée par rapport à la densité tronquée sur $(0,m)$, $f_l(r) \propto r^{p-1}e^{-\frac{r^2}{2}} F\left(\frac{k+p-l}{2}+1, \frac{p}{2}, \frac{r^2 t}{2(1+t)}\right) 1_{(0,m)}(r)$.

Démonstration. De la même manière que la Démonstration précédente, nous calculons la fonction $M_m^\pi(X)$ avec la distribution π_U^l cette fois-ci, ce qui nous donne :

$$\begin{aligned} M_m^U(X) &= \int_{B_{m\sigma}} e^{-\frac{1}{2\sigma^2}\|X-\theta\|^2}\, d\theta \\ &= \int_0^m \int_{S_{r\sigma}} e^{-\frac{1}{2\sigma^2}\|X-\theta\|^2}\, d\theta\, dr \\ &= \int_0^m M_r^{BU}(X)\, dr\,, \end{aligned}$$

où $M_r^{BU}(X)$ est donné par (5.5.6). Il faut noter que la seconde égalité est obtenue en conditionnant par rapport au rayon. Ensuite, nous remarquons que

$$\begin{aligned}
\nabla_X M_m^U(X) &= \int_0^m \nabla_X M_r^{BU}(X)\, dr \\
&= \int_0^m -\frac{X}{\sigma^2} M_r^{BU}(X)\, dr + \\
&\quad k_1 \int_0^m r^{p/2+1}\, e^{-\frac{r^2}{2}}\, \sigma^{p/2-2} \frac{X}{\|X\|^{p/2}}\, e^{-\frac{\|X\|^2}{2\sigma^2}} I_{p/2}(r\,\|X\|/\sigma)\, dr\,.
\end{aligned}$$

selon l'équation (5.5.7). Encore une fois, en utilisant le lemme 5.5.1 et en utilisant le théorème de Fubini pour les fonctions positives, nous obtenons :

$$\delta_U^l(X,S) = \frac{X}{\|X\|} \frac{\int_0^m r^{\frac{p}{2}+1} e^{-\frac{r^2}{2}} \int_0^\infty \sigma^{l-k-\frac{p}{2}-4}\, e^{-\frac{\|X\|^2+S^2}{2\sigma^2}} I_{\frac{p}{2}}\left(\frac{r\|X\|}{\sigma}\right) d\sigma^2\, dr}{\int_0^m r^{\frac{p}{2}} e^{-\frac{r^2}{2}} \int_0^\infty \sigma^{l-k-\frac{p}{2}-5}\, e^{-\frac{\|X\|^2+S^2}{2\sigma^2}} I_{\frac{p}{2}-1}\left(\frac{r\|X\|}{\sigma}\right) d\sigma^2\, dr}.$$

Finalement, la formule (5.5.8) nous permet d'écrire :

$$\delta_U^l(X,S) = \frac{1}{p} \frac{\int_0^m r^{p+1} e^{-\frac{r^2}{2}} F\left(\frac{k+p-l}{2}+1, \frac{p}{2}+1, \frac{r^2 t}{2(1+t)}\right) dr}{\int_0^m r^{p-1} e^{-\frac{r^2}{2}} F\left(\frac{k+p-l}{2}+1, \frac{p}{2}, \frac{r^2 t}{2(1+t)}\right) dr}\, X\,,$$

ce qui implique le résultat voulu. □

La premiere remarque est que pour $l = 0$, nous avons $\delta_{BU}^0 = \delta_m$. Autrement dit, le meilleur estimateur équivariant pour $\frac{\|\theta\|}{\sigma} = m$, δ_m, est un estimateur de Bayes associé à la loi a priori π_{BU}^O. Ceci est un résultat qui a été déja mis en évidence par Perron [Per87].

Ensuite, pour les estimateurs de Bayes associés aux lois a priori π_U^l, nous pouvons conclure que $h_U^l(t) \le h_m^l(t)$ et que $h_U^l(0) = m^2/p$, pour tout l.

Conclusion

Dans cette conclusion, nous mettons l'accent sur nos résultats originaux dans cette thèse.

Nous avons essentiellement considéré l'estimation de la moyenne d'une distribution $p-$dimensionnelle à symétrie sphérique, sous coût quadratique, dans le but de construire des classes plus larges d'estimateurs au moins aussi bons que les estimateurs usuels.

Dans le chapitre 2, nous avons considéré l'estimation de Bayes de la moyenne d'une distribution mélange de lois normales avec $p \geq 3$. Nous avons trouvé une classe de lois a priori (propres et généralisées) composée aussi de distributions de mélange de lois normales engendrant des estimateurs du minimax. Globalement, ce travail représente une extension de Strawderman [Str74] et de Maruyama [Mar03b] et les techniques utilisées sont standards. De plus, la condition de régularité clé pour la minimaxité est décrite à travers une propriété du rapport de vraisemblance monotone des distributions de mélange pour les lois d'échantillonnage et les lois a priori.

Dans le chapitre 3, nous avons prouvé la minimaxité des estimateurs de Bayes généralisés correspondants aux lois a priori superharmoniques de la forme $\pi(||\theta||)^2$ dont le Laplacien est croissant, pour une grande classe de lois à symétrie sphérique en présence d'un vecteur résiduel. Cette sous-classe contient certaines distributions de mélange de lois normales.

Dans ce contexte, les estimateurs de Bayes généralisés du minimax obtenus dépendent de la densité sous-jacente. Par contre, lorsque le paramètre d'échelle est inconnu, la forme des estimateurs de Bayes généralisés est indépendante de la forme de la densité sous réserve que les lois a priori aient une certaine forme (voir Fourdrinier et Strawderman [FS10] et Maruyama [Mar03b]). Autrement dit, en terme de robustesse distributionnelle des estimateurs de Bayes généralisés du minimax, nos résultats sont plus proches de ceux présentés dans [FS08a] que de ceux présentés dans [FS10] et [Mar03b]. D'autre part, en raison de la condition de superharmonicité, nos lois a priori ne peuvent être propres (voir [FSW98]). De plus, pour les distributions de mélange de lois normales comme lois d'échantillonnage, les distributions de mélange ne requièrent pas la propriété de rapport de vraisemblance utilisée dans les travaux de Maruyama [Mar03a] et ceux de Fourdrinier *et al.* [FKS08].

Dans le chapitre 4, nous avons développé une condition nécessaire et suffisante en dimension finie pour comparer la performance des estimateurs linéaires tronqués δ_a ($0 \leq a < 1$) avec celle de l'estimateur du maximum de vraisemblance δ_{emv} dans l'estimation de la moyenne de la loi $\mathcal{N}_p(\theta, I_p)$ sous la contrainte $\|\theta\| \leq m$ avec $m > 0$. Nous avons obtenu des résultats de dominance dans le cadre asymptotique et nous avons obtenu également des expressions précises du gain relatif en risque asymptotique.

Une suite naturelle de ce travail sera de procéder, dans un cadre asymptotique similaire, à l'évaluation de la performance de l'estimateur de Bayes δ_U par rapport à la loi a priori uniforme sur la boule $\{\theta : \|\theta\| \leq m\}$ ou autres estimateurs de Bayes. D'autre part, bien qu'il existe des résultats de dominance pour tout (m, p), le problème de trouver des lois a priori π qui génèrent des estimateurs de Bayes dominants reste non résolu. En général, des solutions sont trouvées pour m relativement petit (voir [MP01]).

Dans le chapitre 5, nous avons construit des estimateurs équivariants pour estimer la moyenne θ de la distribution $\mathcal{N}_p(\theta, \sigma^2 I_p)$ avec σ inconnu et sous la contrainte $\frac{\|\theta\|}{\sigma} \leq m$ avec $m > 0$. Nous avons développé des résultats de dominance de l'estimateur X et de l'estimateur du maximum de vraisemblance δ_{emv}. En particulier, nous avons montré que le meilleur estimateur équivariant δ_m, lorsque $\frac{\|\theta\|}{\sigma} = m$, domine δ_{emv} si $m \leq \sqrt{p}$. Ceci nous a permis d'obtenir la dominance de δ_{emv} pour tout (m, p) en considérant la troncature $\delta_{h_m \wedge h_{emv}}$. Une suite de ce travail consistera à déterminer des estimateurs de Bayes dominant X ou δ_{emv}. En vue de cela, nous avons d'ores et déjà mis en évidence la forme explicite des estimateurs de Bayes associés aux lois a priori $\pi(\theta, \sigma)$ telles que la loi de $\theta\,|\sigma$ est uniforme sur la boule $\{\theta : \|\theta\| \leq m\sigma\}$ ou uniforme sur la sphère $\{\theta : \|\theta\| = m\sigma\}$. Autres perspectives consistent à considérer ce problème avec d'autres modèles à symétrie sphérique et avec vecteur résiduel.

Références

[Ama82] S. Amari. Differential geometry of curved exponential families-curvatures and information loss. *Annals of Statistics*, 10 :357–385, 1982.

[Amo74] D. E. Amos. Computation of modified Bessel functions and their ratios. *Mathematics of computation*, 28(125) :239–251, 1974.

[AS64] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, 1964.

[Bar64a] A. J. Baranchik. A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics*, 41 :642–645, 1964.

[Bar64b] A.J. Baranchik. Multiple regression and estimation of the mean of a multivariate normal distribution. *Stanford University*, 51 :Technical Report, 1964.

[Ber75] J. O. Berger. Minimax estimation of location vectors for a wide class of densities. *Annals of Statistics*, 3 :1318–1328, 1975.

[Ber85] J. O. Berger. *Decision Theory and Bayesian Analysis*. Springer, New York, 1985.

[Ber90] C. Berry. Minimax estimation of a bounded normal mean vector. *Journal of Multivariate Analysis*, 35 :130–139, 1990.

[BJM81] L. Brown, I. Johnstone, and B. MacGibbon. Variation diminishing transformations : A direct approach to total positivity and its statistical applications. *Journal of the American Statistical Association*, 376 :824–832, 1981.

[Boc85] M. E. Bock. Minimax estimators that shift towards a hypersphere for location vectors of spherically symmetric distributions. *Journal of Multivariate Analysis*, 2 :127–147, 1985.

[BR90] J. O. Berger and C. Robert. Subjective hierachical Bayes estimation of multivariate normal mean : On the frequentist interface. *Annals of Statistics*, 18 :617–651, 1990.

[Bro65] L. D. Brown. On the admissibility of invariant estimators of two dimensional location parameters. Mimeographed notes. Birkbeck College, London, 1965.

[Bro66] L. D. Brown. On the admissibility of invariant estimators of one or more location parameters. *Annals of Mathematical Statistics*, 37 :1087–1136, 1966.

[Bro71] L. D. Brown. Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Annals of Mathematical Statistics*. 42 :855–903, 1971.

[Bro79] L. D. Brown. A heuristic method for determining admissibility of estimators with applications. *Annals of Statistics*. 7 :960–994, 1979.

[Bru55] H. D. Brunk. Maximum likelihood estimates of monotone parameters. *Annals of Mathematical Statistics*, 26 :607–616. 1955.

[Bru58] H. D. Brunk. On the estimation of parameters restricted by inequalities. *Annals of Mathematical Statistics*, 29 :437–454, 1958.

[BS78] A. Brandwein and W. E. Strawderman. Minimax estimators of location parameters for spherically symmetric unimodal distributions under quadratic loss. *Annals of Statistics*, 6 :377–416, 1978.

[CE91a] A. Charras and C. Van Eeden. Bayes and admissibility properties of estimators in truncated parameter spaces. *La Revue Canadienne de Statistique*, 19 :121–134, 1991.

[CE91b] A. Charras and C. Van Eeden. Limit of Bayes estimators in convex truncated parameter spaces. *Statistics & Probability Letters*, 11 :479–483, 1991.

[CE92] A. Charras and C. Van Eeden. Bayes properties of estimators of location parameters in truncated parameter spaces. *Statistics and Decisions*, 10 :81–86, 1992.

[CE94] A. Charras and C. Van Eeden. Inadmissibility for squared loss when the parameter to be estimated is restricted to the interval $[a, \infty)$. *Statistics and Decisions*, 12 :257–266, 1994.

[CS81] G. Casella and W. E. Strawderman. Estimating a bounded normal mean. *Annals of Statistics*, 9 :870–878. 1981.

[Das85] A. DasGupta. Bayes minimax estimation in multiparameter families when the parameter space is restricted to a bounded convex set. *Sankhya Ser*, A 47 :326–332, 1985.

[DL11] A. DasGupta and S.N. Lahiri. Density in high and ultra high dimensions. regularization, and the L_1 asymptotics (to appear). *A Festschrift for William E. Strawderman. IMS Collections*, 2011.

[DvE09] Y. Dou and C. van Eeden. Comparaisons of the performances of estimators of a bounded normal mean under squared-error loss. *REVSTAT*, 7(3) :203–226, 2009.

[EM76] B. Efron and C. Morris. Families of minimax estimators of the mean of a multivariate normal distribution. *Annals of Statistics*, 4 :11–21, 1976.

[Fai78] R.E. Faith. Minimax Bayes estimators of a multivariate normal mean. *Journal of Multivariate Analysis*, 8 :372–379, 1978.

[Far64] R. H. Farrell. Estimators of a location parameter in the obsolutely continous case. *Annals of Mathematical Statistics*, 35 :949–998, 1964.

[FKG71] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22 :89–103, 1971.

[FKS08] D. Fourdrinier, O. Kortbi, and W. E. Strawderman. Bayes minimax estimators of the mean of a scale mixture of multivariate normal distributions. *Journal of Multivariate Analysis*, 99(1) :74–93, 2008.

[FM10] D. Fourdrinier and É. Marchand. On Bayes estimators with uniform priors on spheres and their comparative performance with maximum likelihood estimators for estimating bounded multivariate normal means. *Journal of Multivariate Analysis*. 101 :1390–1399, 2010.

[FS08a] D. Fourdrinier and W E. Strawderman. Generalized bayes minimax estimators of location vectors for spherically symmetric distributions. *Journal of Multivariate Analysis*, 99 :735–750, 2008.

[FS08b] D. Fourdrinier and W. E. Strawderman. A unified and generalized set of shrinkage bounds on minimax stein estimates. *Journal of Multivariate Analysis*, 99(10) :2221–2233, 2008.

[FS10] D. Fourdrinier and W. E. Strawderman. Robust generalized Bayes estimator of location vector for spherically symmetric distribution with unknown scale. *A Festschrift for Lawrence D. Brown, IMS Lecture Note Series*, 6 :249–262, 2010.

[FSW98] D. Fourdrinier, W. E. Strawderman, and M. T. Wells. On the construction of Bayes minimax estimators. *Annals of Statistics*, 26 :660–671, 1998.

[FSW03] D. Fourdrinier, W. E. Strawderman, and M. T. Wells. Robust shrinkage for elliptically symmetric distributions with unknown covariance matrix. *Journal of Multivariate Analysis*, 85 :24–39, 2003.

[FSW06] D. Fourdrinier, W. E. Strawderman, and M. T. Wells. Estimation of a location parameter with restrictions of "vague information" for spherically symmetric distributions. *Annals of the Institute of Statistical Mathematics*, 58 :73–92, 2006.

[FW95a] D. Fourdrinier and M. T. Wells. Estimation of a loss function for spherically symmetric distributions in the general linear model.

Annals of Statistics, 23(2) :571–592. 1995.

[FW95b] D. Fourdrinier and M. T. Wells. Loss estimation for spherically symmetric distributions. *Journal of Multivariate Analysis*, 53(2) :311–331, 1995.

[FW96] D. Fourdrinier and M. T. Wells. Spherically symmetric Bayes estimators for a linear subspace of a normal law. In Oxford Sci. Publ., editor, *Bayesian Statistics 5*, pages 569–579. Oxford University Press, New York, 1996.

[GMS87] C. Gatsonis, B. MacGibbon, and W. E. Strawderman. On the estimation of a restricted normal mean. *Statistics & Probability Letters*, 6 :21–30, 1987.

[Gue03] N. R. Gueye. *Problèmes d'estimation de paramètres avec restriction sur l'espace des paramètres*. Thèse de doctorat, Université de Montréal, Montréal, Canada, 2003.

[Har04] J. Hartigan. Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, 32 :136–142, 2004.

[Hin77] D. V. Hinkley. Conditional inference about normal mean with known coefficient of variation. *Biometrika*, 64 :105–108, 1977.

[Joh88] I. Johnstone. On admissibility of some unbiased estimates of loss. In *Statistical decision theory and related topics IV (S. S. Gupta, J. O. Berger Ed.)*, volume 1. pages 361–379. Springer-Verlag, New York, 1988.

[JS61] W. James and C. Stein. Estimation of quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Sta-*

tistics and Probability, volume 1, pages 361–379. University of California Press, Berkeley, 1961.

[Kar57] S. Karlin. Pólya type distributions. II. *Annals of Mathematical Statistics*, 28 :281–309, 1957.

[Kar89] T. Kariya. Equivariant estimation in a model with an ancillary statistic. *Annals of Statistics*, 17(2) :920–928, 1989.

[Kat61] M. W. Katz. Admissible and minimax estimates of parameters in truncated spaces. *Annals of Mathematical Statistics*, 32 :136–142, 1961.

[Kem88] P. J. Kempthorne. Dominating inadmissible procedures using compromise decision theory. In *Statistical Decision Theory and Related Topics IV (S. Gupta and J. O. Berger, eds.)*, volume 1, pages 381–396. Springer, New York, 1988.

[KGP88] T. Kariya, N. C. Giri, and F. Perron. Equivariant estimation of a mean vector of $\mathrm{N}(\mu, \sum)$ with $\mu \sum^{-1} \mu = 1$ or $\sum^{-1/2} \mu = c$ or $\sum = \sigma^2 \mu' \mu \mathrm{I}$. *Journal of Multivariate Analysis*, 27 :270–283, 1988.

[KMPS09] D. Kucerovsky, É Marchand, A. T. Payandeh, and W. E. Strawderman. On the Bayesianity of maximum likelihood estimators of restricted location parameters under absolute value error loss. *Statistics and Decisions*, 27 :145–168, 2009.

[Kub94] T. Kubokawa. A unified approach to improving equivariant estimators. *Annals of Statistics*, 22 :290–299, 1994.

[Kub98] T. Kubokawa. The Stein phenomenon in simultaneous estimation : A review. In *Applied Statistical Science III (S. E. Ahmed,* *M. Ahsanullah, and B. K. Sinha)*, pages 143–173. NOVA Science Publishers. New York, 1998.

[Kub99] T. Kubokawa. Shrinkage and modification techniques in estimation of variance and the related problems : A review. *Communications in Statistics : Theory and Methods*, 28 :613–650, 1999.

[Kub05] T. Kubokawa. Estimation of a mean of a normal distribution with a bounded coefficient of variation. *The Indian Journal of Statistics*, 67 :499–525, 2005.

[LC99] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. 2nd Ed., Springer, New York, 1999.

[Mar93] É. Marchand. Estimation of a multivariate mean with constraints on the norm. *La Revue Canadienne de Statistique*, 21 :359–366, 1993.

[Mar94] É. Marchand. On the estimation of the mean of a $\mathrm{N}_p(\mu, \sum)$ population with $\mu' \sum^{-1} \mu$ known. *Statistics & Probability Letters*. 21 :69–75, 1994.

[Mar03a] Y. Maruyama. Admissible minimax estimators of a mean vector of scale mixtures of multivariate normal distributions. *Journal of Multivariate Analysis*, 84 :274–283, 2003.

[Mar03b] Y. Maruyama. A robust generalized Bayes estimator improving on the james-stein estimator for spherically symmetric distributions. *Statistics and Decisions*, 21 :69–78, 2003.

[MI05] Y. Maruyama and K. Iwasaki. Sensitivity of minimaxity and admissibility in the estimation of a positive normal mean. *Annals of the Institute of Statistical Mathematics*, 57 :145–156, 2005.

[Moo81] J. J. A. Moors. Inadmissibility of linearly invariant estimators in truncated parameter spaces. *Journal of the American Statistical Association*, 76 :910–915, 1981.

[Moo85] J. J. A. Moors. *Estimation in truncated parameter spaces*. Ph.D. thesis, Tilburg University, 1985.

[MOPP08] É. Marchand, I. Ouassou, A. T. Payandeh, and F. Perron. On the estimation of a restricted location parameter for symmetric distributions. *Journal of the Japanese Statistical Society*, 38(2) :293–309, 2008.

[MP01] É. Marchand and F. Perron. Improving on the mle of a bounded normal mean. *Annals of Statistics*, 29 :1066–1081, 2001.

[MP02] É. Marchand and F. Perron. On the minimax estimator of a bounded normal mean. *Statistics & Probability Letters*, 58 :327–333, 2002.

[MP05] É. Marchand and F. Perron. Improving on the mle of a bounded location parameter for spherical distributions. *Journal of Multivariate Analysis*, 92 :227–238, 2005.

[MS04] É. Marchand and W. E. Strawderman. Estimation in restricted parameter spaces : A review. In *A Festschrift for Herman Rubin, IMS Lecture Notes-Monograph Series*, volume 45, pages 21–44. 2004.

[MS05a] É. Marchand and W. E. Strawderman. On improving on the minimum risk equivariant estimator of a location parameter which is constrained to an interval or half-interval. *Annals of the Institute of Statistical Mathematics*, 57(1) :129–143, 2005.

[MS05b] É. Marchand and W. E. Strawderman. On improving on the minimum risk equivariant estimator of a scale parameter under lower-bound constraint. *Journal of Statistical Planning and Inference*, 134 :90–101, 2005.

[MS11] É. Marchand and W. E. Strawderman. A unified result for restricted parameter spaces. *To appear in Bernoùlli*, 2011.

[MT06] Y. Maruyama and A. Takemura. Admissibility and minimaxity of generalized Bayes estimators for spherically symmetric family. Technical Report METR 2005-13, Department of Mathematical Informatics, The University of Tokyo, 2006.

[Per87] F. Perron. *Application de l'invariance en théorie de la décision.* Thèse de doctorat, Université de Montréal, Montréal, Canada, 1987.

[Rob90] C. Robert. Modified Bessel functions and their applications in probability and statistics. *Statistics & Probability Letters*, 9 :155–161, 1990.

[Sac63] J. Sacks. Generalized Bayes solutions in estimation problems. *Annals of Mathematical Statistics*, 34 :751–768, 1963.

[Spr86] M. Spruill. Some approximate restricted Bayes estimators of a normal mean. *Statistics and Decisions*, 4 :337–351, 1986.

[Ste56] C. Stein. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 197–206. University of California Press, Berkeley, 1956.

[Ste81] C. Stein. Estimation of the mean of multivariate normal distribution. *Annals of Statistics*, 9 :1135–1151, 1981.

[Str70] W. E. Strawderman. On the existence of proper Bayes minimax estimators of the mean of a multivariate normal distribution. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 51–55. University of California Press, Berkeley, 1970.

[Str71] W. E. Strawderman. Proper Bayes minimax estimators of the multivariate normal mean. *Annals of Mathematical Statistics*, 42 :385–388, 1971.

[Str73] W. E. Strawderman. Proper bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances. *Annals of Statistics*, 1 :1189–1194, 1973.

[Str74] W. E. Strawderman. Minimax estimation of location parameters for certain spherically symmetric distribution. *Journal of Multivariate Analysis*, 4 :255–264, 1974.

[vE56] C. van Eeden. Maximum likelihood estimation of ordered probabilities. *Proc. Kon. Nederl. Akad. Wetensch. Ser. A*, 59 :444–455, 1956.

[vE57a] C. van Eeden. A least-squares inequality for maximum likelihood estimates of ordered parameters. *Proc. Kon. Nederl. Akad. Wetensch. Ser. A*, 60 :513–521, 1957.

[vE57b] C. van Eeden. Maximum likelihood estimation of partially or completely ordered parameters. *Proc. Kon. Nederl. Akad. Wetensch. Ser. A*, 60 :128–136 and 201–211, 1957.

[vE57c] C. van Eeden. Note on two methods for estimating ordered parameters of probability distributions. *Proc. Kon. Nederl. Akad. Wetensch. Ser. A*, 60 :506–512, 1957.

[vE58] C. van Eeden. *Testing and Estimating Ordered Parameters of Probability distributions.* PhD. thesis, University of Amsterdam, Amsterdam, The Netherlands, 1958.

[vE06] C. van Eeden. *Restricted Parameter Space Estimation Problems : Admissibility and Minimaxity Properties*, volume 188. Lecture Notes in Statistics, Springer, New York, 2006.

[Wat83] G. S. Watson. *Statistics on Spheres.* Wiley, New York, 1983.

[Wid41] D. V. Widder. *The Laplace Transform.* Princeton Mathematical Series, v. 6. Princeton University Press, Princeton, New York, 1941.