

**Estimation par densités prédictives**

par

Jean-Philippe TURCOTTE

mémoire présenté au Département de mathématiques  
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES  
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, février 2013



Library and Archives  
Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*

*ISBN: 978-0-494-93282-7*

*Our file Notre référence*

*ISBN: 978-0-494-93282-7*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Le 4 février 2013

*le jury a accepté le mémoire de Monsieur Jean-Philippe Turcotte  
dans sa version finale.*

Membres du jury

Professeur Éric Marchand  
Directeur de recherche  
Département de mathématiques

Professeur William E. Strawderman  
Codirecteur de recherche  
Rutgers University

Professeur Ernest Monga  
Évaluateur interne au programme  
Département de mathématiques

Professeur Hugo Larochelle  
Président rapporteur  
Département de mathématiques

# SOMMAIRE

L'inférence statistique est un domaine complexe et en constante évolution. Ce mémoire traitera de l'inférence sur la fonction de densité d'une variable aléatoire. Nous partirons de plusieurs résultats connus et développerons une analyse de ces résultats dans le cadre paramétrique avec une approche bayésienne. Nous nous aventurerons aussi dans les problèmes avec espace paramétrique restreint. L'objectif du travail est de trouver les meilleurs estimateurs possibles considérant l'information *a priori* et l'observation de variables tirées d'une densité faisant intervenir le paramètre.

Le chapitre 1 traitera de notions d'inférence bayésienne, de choix de perte évaluant la performance d'un estimateur et possédant des propriétés recherchées. Le chapitre 2 concernera l'estimation ponctuelle du paramètre. En particulier, nous aborderons l'estimateur de James-Stein et trouverons des conditions suffisantes pour la minimaxité et la dominance d'estimateurs en remarquant la forme particulière de ceux-ci. Une condition remontera même à la loi *a priori* utilisée. Le chapitre 3 établira des liens entre l'estimation ponctuelle et l'estimation par densité prédictive pour le cas multinormal. Des conditions seront aussi établies pour la minimaxité et la dominance. Nous comparerons nos estimateurs à l'estimateur de Bayes découlant d'une loi *a priori* non informative et démontrerons les résultats par des exemples. Le chapitre 4 considérera le problème dans un cadre plus général où le paramètre d'intérêt pourra être un paramètre de position ou

d'échelle. Des liens entre ces deux problèmes seront énoncés et nous trouverons des conditions sur la famille de densités étudiée pour trouver des estimateurs minimax. Quelques exemples concluront cette section. Finalement, le chapitre 5 est l'intégrale de l'article déposé en collaboration avec Tatsuya Kubokawa, Éric Marchand et William E. Strawderman, concernant l'ensemble du problème étudié dans ce mémoire, à savoir l'estimation par densité prédictive dans un espace paramétrique restreint.

# REMERCIEMENTS

Je tiens d'abord à remercier mes directeurs de maîtrise. En premier lieu, M. Éric Marchand pour sa disponibilité, sa compréhension, son soutien et pour tout ce que j'ai pu apprendre durant les deux dernières années, mais aussi M. William E. Strawderman pour son aide et le temps qu'il m'a consacré. Je voudrais aussi remercier tous mes confrères et consœurs de travail pour les discussions mathématiques et le soutien. Merci au département de mathématiques de l'Université de Sherbrooke et au Fonds de recherche du Québec Nature et technologies (FRQNT) pour leur aide financière. Finalement, un merci tout spécial à ma famille pour m'avoir épaulé et encouragé, en particulier mon copain Michaël.

Jean-Philippe Turcotte  
Sherbrooke, décembre 2012

# TABLE DES MATIÈRES

<b>SOMMAIRE</b>	<b>iii</b>
<b>REMERCIEMENTS</b>	<b>v</b>
<b>TABLE DES MATIÈRES</b>	<b>vi</b>
<b>LISTE DES FIGURES</b>	<b>viii</b>
<b>NOTATION</b>	<b>xi</b>
<b>INTRODUCTION</b>	<b>1</b>
<b>CHAPITRE 1 — Préliminaires</b>	<b>3</b>
1.1 Définitions . . . . .	3
1.2 Choix de perte . . . . .	5
<b>CHAPITRE 2 — Estimation de Stein</b>	<b>12</b>
2.1 Approche géométrique . . . . .	13

2.2	Estimateur de Bayes empirique . . . . .	14
2.3	Résultats de dominance . . . . .	15
2.4	Estimateurs de Bayes : Résultats de minimaxité . . . . .	18
<b>CHAPITRE 3 — Densités prédictives pour la loi Multinormale</b>		<b>23</b>
3.1	Résultats de minimaxité et de dominance . . . . .	24
3.2	Lien entre l'estimation par densité prédictive et l'estimation ponctuelle .	29
3.3	Estimation avec contrainte . . . . .	33
<b>CHAPITRE 4 — Estimation pour des familles à paramètre de position et d'échelle sous contrainte</b>		<b>44</b>
4.1	Cas d'un paramètre de position sous contrainte . . . . .	44
4.2	Cas d'un paramètre d'échelle sous contrainte . . . . .	52
4.3	Cas particulier : Loi Gamma . . . . .	54
<b>CHAPITRE 5 — Minimavity in Predictive Density Estimation with Parametric Constraints</b>		<b>61</b>
5.1	Introduction . . . . .	62
5.1.1	Preamble . . . . .	62
5.1.2	Outline of Paper . . . . .	63
5.1.3	Brief review of previous findings for normal models . . . . .	65
5.2	General conditions for minimavity . . . . .	67



5.3	Location and scale families : minimaxity and improvements on $\hat{q}^{BI}$ . . . . .	74
5.3.1	Minimaxity for location families . . . . .	74
5.3.2	Improvements on the best equivariant estimator $\hat{q}^{BI}$ . . . . .	75
5.3.3	Case of a scale family . . . . .	82
5.4	Estimation in location-scale families . . . . .	85
5.4.1	Non-bounded case . . . . .	85
5.4.2	Bounded case . . . . .	88
5.4.3	Multidimensional case . . . . .	89
5.5	Concluding remarks . . . . .	91
	<b>CONCLUSION</b>	<b>96</b>
	<b>BIBLIOGRAPHIE</b>	<b>97</b>

# LISTE DES FIGURES

2.1	Illustration géométrique de l'estimation du paramètre $\theta$ à partir de l'observation $X$ . . . . .	13
3.1	Estimateur de Bayes pour la loi <i>a priori</i> $\pi(\theta) = 0,4\mathbf{1}_{\{-5\}\cup\{5\}}(\theta)+0,2\mathbf{1}_{\{0\}}(\theta)$ et pour la loi normale ( $v_x = v_y = 1$ ) avec observation $x = 2, 5$ . . . . .	34
3.2	Estimateur de Bayes pour la loi <i>a priori</i> uniforme sur $[-5, 5]$ pour la loi normale ( $v_x = v_y = 1$ ) avec observation $x = 2$ . . . . .	35
3.3	Estimateur de Bayes pour la loi <i>a priori</i> uniforme sur $[-5, 5]$ pour la loi normale ( $v_x = v_y = 1$ ) avec observation $x = 100$ . . . . .	36
3.4	Différence des risques (lois <i>a priori</i> uniforme sur $[-5, 5]$ et sur les réels avec $v_x = v_y = 1$ ) . . . . .	37
3.5	Estimateur de Bayes pour la loi <i>a priori</i> uniforme sur les réels positifs pour la loi normale ( $v_x = v_y = 1$ ) avec observation $x = -2$ . . . . .	39
3.6	Différence relative des risques (loi <i>a priori</i> uniforme sur les réels positifs et sur les réels avec $v_x = v_y = 1$ ) . . . . .	40
4.1	Densités prédictives pour la loi exponentielle avec $\sigma = 1$ et $x = 2$ . . . . .	50

4.2	Différence relative des risques pour la loi exponentielle avec $\sigma = 1$ et $x = 2$ .	51
4.3	Estimateurs de Bayes pour loi Gamma ( $\alpha_1 = \alpha_2 = 1$ avec observation $x = 1$ ).	57
4.4	Différence relative des risques pour estimateurs de Bayes sous loi <i>a priori</i> $\frac{1}{\beta} \mathbf{1}_{\mathbb{R}^+}(\beta)$ et sous loi <i>a priori</i> $\frac{1}{\beta} \mathbf{1}_{]0,1]}(\beta)$ pour loi Gamma ( $\alpha_1 = \alpha_2 = 1$ ).	59
5.1	Relative difference in risks between $\hat{q}^{BI}$ and $\hat{q}^U$ (normal model with $\mu \geq 0, \sigma_X^2 = \sigma_Y^2 = 1$ ).	80

# NOTATION

$\theta$	: Paramètre d'intérêt
$\delta$	: Estimateur
$p(\cdot), q(\cdot)$	: Fonctions de densités
$\hat{p}$	: Densité prédictive
$\pi(\cdot)$	: Densité de la loi <i>a priori</i> du paramètre $\theta$
$\delta_\pi, \hat{\theta}_\pi$	: Estimateurs de Bayes du paramètre $\theta$ associés à la loi <i>a priori</i> $\pi$
$\hat{p}_\pi$	: Estimateur de Bayes de la densité $p(\cdot)$ associé à la loi <i>a priori</i> $\pi$
$m_\pi(\cdot)$	: Densité marginale sous la loi <i>a priori</i> $\pi$
$L_{KL}$	: Fonction de perte Kullback-Leibler
$L_Q$	: Fonction de perte quadratique
$R_{KL}$	: Fonction de risque Kullback-Leibler
$R_Q$	: Fonction de risque quadratique
$\Delta(\cdot)$	: Différence entre les risques de deux estimateurs
$r_\pi$	: Risque de Bayes associé à la loi <i>a priori</i> $\pi$
$p$	: Dimension
$\nabla$	: Gradient
$\nabla \cdot$	: Divergence
$\nabla^2$	: Laplacien
$\phi(\cdot)$	: Densité d'une loi normale
$\Phi(\cdot)$	: Répartition d'une loi normale
$\mathbf{1}_D(\cdot)$	: Fonction constante égale à 1 sur l'espace $D$

# INTRODUCTION

Dans le domaine de la statistique, l'estimation est un domaine très utile pour comprendre le monde qui nous entoure et prédire de futurs événements. Plusieurs statisticiens se sont penchés sur l'estimation de paramètres de position ou d'échelle tels que la moyenne ou la variance d'une variable  $X$  de loi normale.

Or, si l'objectif est de prédire de futurs comportements de la variable, ne serait-il pas mieux d'estimer la fonction de densité au complet, plutôt que de se restreindre à estimer un paramètre? Dans ce cas, comment peut-on définir la notion de distance entre l'estimation et la densité réelle?

Ajoutons à ce problème l'information *a priori* du paramètre. Par exemple, si on sait que la moyenne d'une variable aléatoire est strictement positive, comment gère-t-on cette information à travers notre réflexion? La statistique bayésienne fournira des pistes de réponses quant aux meilleurs estimateurs, tout en permettant le plus d'objectivité possible par le choix de lois *a priori* non informatives.

Ainsi, nous consacrerons nos efforts sur l'estimation de la fonction de densité d'une variable  $Y$  dans le cadre paramétrique à partir d'une observation  $X$ . Nous débuterons par se choisir une fonction de perte mesurant l'erreur d'une estimation. La perte Kullback-Leibler jouera un rôle important dans ce mémoire. Elle possède des propriétés intéressantes que nous démontrerons.

Par la suite, pour mieux comprendre l'estimation par densité prédictive et établir des liens, nous réviserons plusieurs résultats concernant l'estimation ponctuelle du paramètre d'intérêt sous la perte quadratique. En particulier, nous rappellerons les résultats concernant l'efficacité de l'estimateur de James-Stein qui domine l'estimateur usuel  $X$ . De plus, nous établirons des conditions sur la loi *a priori* pour qu'ils conduisent à un estimateur minimax.

La troisième partie de ce mémoire concernera le problème d'estimation par densité prédictive dans le cas particulier de la loi multinormale. Plusieurs estimateurs seront comparés et des résultats similaires à ceux du problème d'estimation ponctuelle pourront être trouvés. Nous regarderons ensuite ce qui se produit lorsque l'espace paramétrique est restreint à partir d'exemples particuliers. Des lois *a priori* impropres seront à l'étude dans ce contexte.

Finalement, la restriction de l'espace paramétrique peut s'appliquer dans un cadre plus général que la loi multinormale. Nous étudierons donc l'efficacité des estimateurs pour des paramètres de position et d'échelle. Nous nous demanderons quelles sont les conditions nécessaires à imposer sur la famille de densités de  $Y$  pour mener à des résultats de dominance et de minimaxité. Des exemples cloront cette partie.

# CHAPITRE 1

## Préliminaires

Ce chapitre fait un rappel des définitions concernant l'inférence bayésienne et des propriétés de la perte Kullback-Leibler qui sera utilisée dans ce mémoire.

### 1.1 Définitions

Tout au long de ce mémoire, nous faisons référence à certaines notions d'inférence statistique. Pour mieux comprendre ce dont on parle, voici quelques définitions auxquelles on fait référence. Nous travaillons avec un modèle statistique  $X \sim p_\theta$ , où  $\theta \in \Theta$  avec  $\Theta$  un espace paramétrique. Nous dénotons par  $\delta$  des estimateurs, par  $\delta(x)$  des estimés fonctions de l'observation  $X = x$  et par  $L(\theta, \delta(X))$  la perte mesurant l'erreur de notre estimation.

**Définition 1.1.1.** *Pour un problème d'estimation d'un paramètre  $\theta$  ou d'une fonction paramétrique  $g(\theta)$  par un estimateur  $\delta$  avec fonction de perte  $L(\theta, \delta)$ , la fonction de risque est définie par  $R(\theta, \delta) = E^{X|\theta}[L(\theta, \delta(X))]$ .*

Une propriété utile pour choisir un estimateur est la minimaxité.

**Définition 1.1.2.** Pour un problème d'estimation avec risque  $R(\theta, \delta)$ , un estimateur  $\delta_0$  est minimax s'il minimise le risque maximal en  $\theta$ , c'est-à-dire si  $\sup_{\theta} R(\theta, \delta_0) = \inf_{\delta} \sup_{\theta} R(\theta, \delta)$ .

Nous souhaiterons à plusieurs reprises comparer deux estimateurs dans un même contexte. Qu'est-ce qui ferait qu'un estimateur serait plus performant qu'un autre relativement au risque défini ?

**Définition 1.1.3.** Un estimateur  $\delta_1$  domine un autre estimateur  $\delta_2$  pour une perte  $L(\theta, \delta)$  si  $R(\theta, \delta_1) \leq R(\theta, \delta_2)$  pour tout  $\theta$  avec inégalité stricte pour au moins un  $\theta$ .

Dans le cadre d'inférence bayésienne, on utilise l'information détenue sur le paramètre inconnu  $\theta$  pour lui associer une fonction de densité  $\pi(\theta)$ . Le paramètre est alors considéré comme une variable aléatoire.

**Définition 1.1.4.** Pour une loi a priori  $\pi(\theta)$  sur le paramètre  $\theta$  et pour la perte  $L(\theta, \delta)$ , l'estimateur  $\delta_{\pi}$  de  $\theta$  est un estimateur de Bayes s'il minimise la perte espérée a posteriori  $E^{\theta|X}[L(\theta, \delta)]$ .

La classe des estimateurs de Bayes peut être agrandie en généralisant le concept d'estimateur de Bayes.

**Définition 1.1.5.** Un estimateur  $\delta_0$  est dit Bayes par extension lorsqu'il existe une suite de lois a priori propres  $\{\pi_m\}_{m \geq 1}$  telle que  $\lim_{m \rightarrow \infty} r_{\pi_m}(\delta_{\pi_m}) = \lim_{m \rightarrow \infty} r_{\pi_m}(\delta_0) < \infty$  où  $r_{\pi_m}(\delta) = E^{\theta}[R(\theta, \delta)]$  est le risque intégré de Bayes pour la loi a priori  $\pi_m$  et où  $\delta_{\pi_m}$  est l'estimateur de Bayes pour la loi a priori  $\pi_m$ .

Il a été démontré qu'il existe un lien entre la minimaxité d'un estimateur et le fait qu'il soit un estimateur de Bayes, comme le souligne le théorème suivant.



**Théoreme 1.1.6.** *Un estimateur de Bayes par extension  $\delta_0$  dont le risque est  $R(\theta, \delta_0) = c$  (constant) pour tout  $\theta$  est minimax.*

*Démonstration.* Supposons que  $\delta_0$  ne soit pas minimax. Alors,

$$\Rightarrow \exists \delta_1 : \sup_{\theta} R(\theta, \delta_1) < c$$

$$\Rightarrow \exists \epsilon > 0 : \sup_{\theta} R(\theta, \delta_1) + \epsilon < c.$$

Or, pour ce même  $\epsilon$ , il existe une valeur  $m_0 \geq 1$  tel que  $c \leq r_{\pi_{m_0}}(\delta_{\pi_{m_0}}) + \epsilon$ ,

$$\Rightarrow \sup_{\theta} R(\theta, \delta_1) < r_{\pi_{m_0}}(\delta_{\pi_{m_0}}),$$

ce qui n'est pas possible. □

Ainsi, on peut considérer ces définitions et ces résultats au cours de ce travail lorsqu'ils sont utiles.

## 1.2 Choix de perte

Dans le cadre de ce mémoire, nous faisons face à un problème d'estimation de la fonction de densité  $p(\cdot)$  d'une variable ou d'un vecteur aléatoire  $Y$  par une densité prédictive  $\hat{p}(\cdot)$ . Pour mesurer l'efficacité d'une telle estimation, nous devons tout d'abord définir une fonction de perte.

Quels sont les critères d'une bonne fonction de perte ? Quelles caractéristiques peuvent être avantageuses ? Pour commencer, nous voulons qu'elle soit positive et qu'elle soit

nulle pour une estimation qui s'avère exacte. Aussi, il est préférable qu'elle soit définie de façon intrinsèque, c'est-à-dire que si on procède à un changement de variables par une fonction strictement monotone, il n'y ait pas d'impact sur la perte ou sur le choix de l'estimateur.

Voilà pourquoi la fonction de perte de Kullback-Leibler, introduite dans ([15]), semble un choix judicieux pour la suite de ce mémoire. Elle se définit comme suit :

$$L_{KL}(p, \hat{p}) = \int p(y) \log\left(\frac{p(y)}{\hat{p}(y)}\right) d\mu(y) = E^Y \left[ \log\left(\frac{p(Y)}{\hat{p}(Y)}\right) \right],$$

où  $\mu$  une mesure  $\sigma$ -finie sur l'espace échantillonnal de  $Y$  sous-ensemble de  $\mathbb{R}^p$ ,  $p(\cdot)$  est la densité à estimer,  $\hat{p}(\cdot)$  est un estimateur de  $p(\cdot)$  qui sont toutes deux absolument continues par rapport à la mesure  $\mu$ , et où l'espérance est prise par rapport à  $p$ .

Cette perte considère la distance entre les deux densités en chaque point  $y$  à travers leur vraisemblance. Notons aussi que cette fonction de perte peut être décrite par l'expression :

$$L_{KL}(p, \hat{p}) = \int p(y) \left( \frac{\hat{p}(y)}{p(y)} - \log\left(\frac{\hat{p}(y)}{p(y)}\right) - 1 \right) dy, \quad (1)$$

ce qui laisse présager qu'il y a un lien direct entre cette fonction de perte et la perte entropie  $\rho(t) = t - \log(t) - 1$  où  $t = \frac{\hat{p}(y)}{p(y)}$ . On peut donc interpréter cette perte comme la perte entropie moyenne pondérée par la fonction de densité de  $Y$ .

**Proposition 1.2.1.** *La fonction de perte de Kullback-Leibler (i) est positive ou nulle et (ii) est nulle lorsque  $\hat{p} \equiv p$  presque partout.*

*Démonstration.* (i) Comme la fonction  $-\log$  est strictement convexe sur  $]0, \infty[$ , on peut déduire le résultat suivant par l'inégalité de Jensen :

$$\begin{aligned}
L(p, \hat{p}) &= E^Y[\log(\frac{p(y)}{\hat{p}(y)})] \\
&\geq -\log(E^Y[\frac{\hat{p}(y)}{p(y)}]) \\
&= -\log(\int p(y) \frac{\hat{p}(y)}{p(y)} dy) \\
&= -\log(1) \\
&= 0,
\end{aligned}$$

avec égalité si et seulement si  $\frac{\hat{p}(y)}{p(y)}$  est constant avec probabilité un. Donc,  $L(p, \hat{p}) = 0 \Leftrightarrow \hat{p}(y) \equiv p(y)$  presque partout sur le support de  $p(y)$ .  $\square$

**Remarque 1.2.2.** *On peut aussi déduire le résultat du fait que la perte entropie est toujours positive ou nulle. Ainsi, la perte Kullback-Leibler qu'on peut exprimer sous la forme  $L(p, \hat{p}) = \int p(y)(\frac{\hat{p}(y)}{p(y)} - \log(\frac{\hat{p}(y)}{p(y)}) - 1)dy$  (voir notes non publiées de Marchand [17]) a la même propriété.*

La perte utilisée dans ce mémoire possède la caractéristique d'être intrinsèque, c'est-à-dire qu'elle est définie à partir de la distribution d'échantillonnage. Cette propriété la rend invariante aux changements de variables strictement monotones, ce qui n'est pas le cas, entre autres, de la perte quadratique. Un changement de variables ne conduit pas à une estimation différente, comme énoncé dans la prochaine proposition. Une étude des pertes intrinsèques est discutée dans Robert ([19]).

**Proposition 1.2.3.** *Dans le cadre de l'estimation multidimensionnelle, soient  $Y \sim p_Y(\cdot)$  et  $Y_1 = h(Y) \sim p_{Y_1}(\cdot)$  où  $h(\cdot)$  est strictement monotone. Soient  $\hat{p}_Y$  et  $\hat{p}_{Y_1}$  des estimateurs de  $p_Y$  et  $p_{Y_1}$  respectivement. Alors,  $L(p_Y, \hat{p}_Y) = L(p_{Y_1}, \hat{p}_{Y_1})$  où  $\hat{p}_{Y_1}$  est la densité de  $Y'_1 = h(Y')$  avec  $Y' \sim \hat{p}(\cdot)$ .*

*Démonstration.* On a :

$$\begin{aligned}
L(p_{Y_1}(\cdot), \widehat{p}_{Y_1}(\cdot)) &= E^{Y_1}[\log(\frac{p_{Y_1}(Y_1)}{\widehat{p}_{Y_1}(Y_1)})] \\
&= E^{Y_1}[\log(\frac{|(h^{-1}(Y_1))'| p_Y(h^{-1}(Y_1))|}{|(h^{-1}(Y_1))'| \widehat{p}_Y(h^{-1}(Y_1))|})] \\
&= E^Y[\log(\frac{p_Y(Y)}{\widehat{p}_Y(Y)})] \\
&= L(p_Y(\cdot), \widehat{p}_Y(\cdot)).
\end{aligned}$$

□

À partir de maintenant, nous considérons le problème paramétrique suivant :

### Problème (\*)

**Soient  $X|\theta \sim p_X(\cdot|\theta)$  et  $Y|\theta \sim p_Y(\cdot|\theta)$  indépendantes. Nous tentons d'estimer sous la perte Kullback-Leibler  $p_Y(\cdot|\theta)$  à partir d'une ou plusieurs observations de  $X$  considérant qu'on dispose d'une loi *a priori*  $\pi$  pour le paramètre  $\theta \in \mathbb{R}^p$ . Il arrive qu'on utilise une loi *a priori* non informative lorsqu'aucune information sur  $\theta$  n'est disponible.**

Pour le problème d'estimation ponctuelle de  $\theta$  sous la perte quadratique  $L(\theta, \delta(x)) = \|\theta - \delta(x)\|^2$ , on remarque que l'estimateur de Bayes associé à une loi *a priori*  $\pi$  est  $E[\theta|X]$  si  $E[\|\theta\|^2 | X] < \infty$ , car cet estimateur minimise, pour tout  $x$ , l'espérance *a posteriori* de la perte. Ce résultat naturel se rencontre aussi pour le problème (\*), comme démontré dans Aitchison ([1]).

**Proposition 1.2.4.** *Pour le problème (\*), l'estimateur de Bayes  $\widehat{p}_\pi(\cdot; X)$  est la densité conditionnelle de  $Y|X = x$  (dite prédictive) donnée par  $p_{Y|X}(y) = \int p(y|\theta)\pi(\theta|x)d\nu(\theta)$ , où  $\pi(\theta|x)$  est la densité *a posteriori* de  $\theta$  étant donné  $x$  et par rapport à une mesure  $\sigma$ -finie  $\nu$ . Le résultat est le même si on considère plutôt la perte  $L_2(\theta, \widehat{p}) = \int (p(y|\theta) - \widehat{p}(y|x))^2 d\mu(y)$ .*

*Démonstration.* L'estimateur de Bayes associé à une loi *a priori*  $\pi$  sur  $\theta$  est celui qui minimise, pour tout  $x$ , l'espérance *a posteriori* de la perte. On a :

$$\begin{aligned}
E^{\theta|X=x}[L_{KL}(\theta, \hat{p})] &= \int_Y \int_{\Theta} \pi(\theta|x) p(y|\theta) \log\left(\frac{p(y|\theta)}{\hat{p}(y|x)}\right) d\nu(\theta) d\mu(y) \\
&= C - \int_Y p(y|x) \log\left(\frac{\hat{p}(y|x)}{p(y|x)}\right) d\mu(y) + \int_Y p(y|x) \log(p(y|x)) d\mu(y) \\
&\geq C + \int_Y p(y|x) \log(p(y|x)) d\mu(y)
\end{aligned}$$

par l'inégalité de Jensen appliquée à la fonction  $-\log(\cdot)$  avec égalité si et seulement si  $\hat{p}(\cdot|x) \equiv p(\cdot|x)$  presque partout et où  $p(y|x) = \int_{\Theta} p(y|\theta)\pi(\theta|x)d\nu(\theta)$ .

Pour la perte  $L_2$ , on a :

$$\begin{aligned}
E^{\theta|X=x}[L_2(\theta, \hat{p})] &= \int_Y \int_{\Theta} \pi(\theta|x) (p(y|\theta) - \hat{p}(y|x))^2 d\nu(\theta) d\mu(y) \\
&= \int_Y \int_{\Theta} \pi(\theta|x) p(y|\theta)^2 d\nu(\theta) d\mu(y) - 2 \int_Y \int_{\Theta} \pi(\theta|x) p(y|\theta) \hat{p}(y|x) d\nu(\theta) d\mu(y) \\
&\quad + \int_Y \int_{\Theta} \pi(\theta|x) \hat{p}(y|x)^2 d\nu(\theta) d\mu(y) \\
&= C - 2 \int_Y p(y|x) \hat{p}(y|x) d\mu(y) + \int_Y \hat{p}(y|x)^2 d\mu(y)
\end{aligned}$$

On cherche donc à minimiser, pour tout  $y$  :  $-2p(y|x)\hat{p}(y|x) + \hat{p}(y|x)^2$ . En dérivant, on obtient :  $-2p(y|x) + 2\hat{p}(y|x)$ , ce qui est nul ssi  $\hat{p}(y|x) = p(y|x)$ . Puisque la dérivée seconde vaut toujours 2,  $\hat{p}(y|x) = p(y|x) = \int \pi(\theta|x)p(y|\theta)d\theta$  est donc la quantité qui minimise la perte *a posteriori*.  $\square$

Voici deux premiers exemples d'estimateurs de Bayes sous la perte Kullback-Leibler :

**Exemple 1.2.5.** *Considérons le problème (\*) où  $X|\theta \sim N_p(\theta, v_x I_p)$ ,  $Y|\theta \sim N_p(\theta, v_y I_p)$  et  $\theta \sim N(\mu, v_\theta I_p)$ .*

*On trouve à partir de ces informations que  $X \sim N_p(\mu, (v_\theta + v_x)I_p)$ . Ainsi, la loi de  $\theta|X$  est définie par la densité suivante :*

$$\begin{aligned}
\pi(\theta|x) &= \frac{\frac{1}{(4\pi^2 v_\theta v_x)^{\frac{p}{2}}} e^{-\frac{1}{2} \frac{v_\theta + v_x}{v_\theta v_x} \left\| \theta - \frac{v_x \mu + v_\theta x}{v_\theta + v_x} \right\|^2} e^{-\frac{1}{2(v_\theta + v_x)} \|x - \mu\|^2}}{\frac{1}{(2\pi(v_\theta + v_x))^{\frac{p}{2}}} e^{-\frac{1}{2(v_\theta + v_x)} \|x - \mu\|^2}} \\
&= \left( \frac{v_\theta + v_x}{2\pi v_\theta v_x} \right)^{\frac{p}{2}} e^{-\frac{1}{2} \frac{v_\theta + v_x}{v_\theta v_x} \left\| \theta - \frac{v_x \mu + v_\theta x}{v_\theta + v_x} \right\|^2}.
\end{aligned}$$

Donc,  $\theta|X = x \sim N_p(\mu(x), v(x)I_p)$  où  $\mu(x) = \frac{v_x \mu + v_\theta x}{v_\theta + v_x}$  et  $v(x) = \frac{v_\theta v_x}{v_\theta + v_x}$ .

Finalement, par un développement similaire à celui utilisé pour trouver la densité marginale de  $X$  et à celui détaillé dans l'Exemple 3.1.2, on a que  $\hat{p}(\cdot|x) \sim N_p(\mu(x), (v(x) + v_y)I_p)$ .

Ce résultat est intéressant, car on peut établir des parallèles avec l'estimation ponctuelle de  $\theta$  sous la perte quadratique  $\|\delta - \theta\|^2$ . D'abord, il est à remarquer que  $\mu(x)$  est l'estimateur de Bayes de  $\theta$  sous la perte quadratique  $\|d - \theta\|^2$  avec la même loi *a priori*. Ainsi, l'estimateur de Bayes de  $p(\cdot|\theta)$  est la densité d'une loi normale de moyenne  $\mu(x)$  et de variance supérieure à la variance  $v_y$ . Aussi, en choisissant  $\mu = 0$  et  $v_\theta \rightarrow \infty$ , on en déduit l'estimateur de Bayes généralisé  $\hat{p}_U(\cdot|x) \sim N_p(x, (v_x + v_y)I_p)$ . Finalement, on sait que ces estimateurs  $\hat{p}(\cdot|x) \sim N_p(\mu(x), (v(x) + v_y)I_p)$  sont admissibles, car ils sont des estimateurs de Bayes uniques avec un risque intégré de Bayes  $r_\pi$  fini.

**Exemple 1.2.6.** Dans le cadre du problème (\*), si  $X|\theta \sim \text{Poisson}(\theta)$ ,  $Y|\theta \sim \text{Poisson}(\theta)$  et que la loi *a priori* de  $\theta$  est la loi impropre  $\pi(\theta) = \mathbb{1}_{]0, \infty[}(\theta)$ , on obtient pour densité prédictive :

$$\begin{aligned}
\hat{p}_\pi(y|x) &= \frac{\int_0^\infty \frac{\theta^y e^{-\theta}}{y!} \frac{\theta^x e^{-\theta}}{x!} d\theta}{\int_0^\infty \frac{\theta^x e^{-\theta}}{x!} d\theta} \\
&= \frac{1}{y!} \frac{\int_0^\infty \theta^{x+y} e^{-2\theta} d\theta}{\int_0^\infty \theta^x e^{-\theta} d\theta} \\
&= \frac{1}{y!} \frac{\left[ -\sum_{i=0}^{x+y} \frac{(x+y)!}{(x+y-i)!} \frac{\theta^{x+y-i}}{2^{i+1}} e^{-2\theta} \right]_{\theta=0}^\infty}{\left[ -\sum_{i=0}^x \frac{x!}{(x-i)!} \theta^{x-i} e^{-\theta} \right]_{\theta=0}^\infty} \\
&= \binom{x+y}{y} \left(\frac{1}{2}\right)^{x+y+1} \mathbb{1}_{\{0,1,2,\dots\}}(y).
\end{aligned}$$

*On reconnaît ici la fonction de masse d'une loi binomiale négative où  $y$  représente le nombre d'échecs obtenus avant qu'on obtienne le  $(x + 1)^e$  succès et où la probabilité de succès est de  $\frac{1}{2}$ .*

Ainsi, les notions décrites dans ce chapitre seront utilisées pour présenter les résultats principaux de ce mémoire. On étudiera la minimaxité et la dominance de certains estimateurs. La perte Kullback-Leibler sera utilisée dans le cadre du problème d'estimation par densité prédictive.

# CHAPITRE 2

## Estimation de Stein

Dans cette section, nous nous attarderons au problème d'estimation ponctuelle d'un paramètre de position  $\theta$  sous la perte quadratique  $L_Q(\theta, \delta) = \|\theta - \delta\|^2$  à partir d'un ensemble d'observations. Bien que la moyenne échantillonnale tirée d'une loi normale avec moyenne  $\theta$  et matrice de variance-covariance  $I_p$  soit sans biais, minimax et l'estimateur du maximum de vraisemblance, il a été démontré par Stein ([21]) et par James et Stein ([10]), que cet estimateur est inadmissible pour des dimensions  $p \geq 3$ , c'est-à-dire qu'il existe un estimateur dont le risque est inférieur ou égal sur l'espace paramétrique avec une inégalité stricte en au moins un point.

Nous aborderons plus précisément ce problème à partir d'une population de loi multivariée normale  $N_p(\theta, \sigma^2 I_p)$  où  $\sigma^2$  est connue. Nous disposons d'un échantillon  $X_1, X_2, \dots, X_n$ . Puisque  $\bar{X}$  est une statistique exhaustive de loi  $N_p(\theta, \frac{\sigma^2}{n} I_p)$ , on peut considérer une seule observation sans perte de généralité.

Dans ce contexte, l'estimateur de James-Stein est donné par  $\delta_{JS}(X) = (1 - \frac{p-2}{\|X\|^2})X$ . Pour  $p \geq 3$ , cet estimateur domine  $X$  uniformément en  $\theta \in \mathbb{R}^p$ . Puisque  $X$  est minimax (voir Casella et Lehmann ([5])), tout estimateur dominant  $X$  est aussi minimax.



Nous élaborerons donc sur ce résultat célèbre et décrirons certains aspects décrits dans la littérature, notamment en ce qui concerne des conditions pour qu'un estimateur  $\delta(X) = X + g(X)$  domine  $X$ . Nous identifierons aussi des conditions de dominance applicables aux estimateurs bayésiens. Cette présentation est adaptée de Strawderman ([23]).

## 2.1 Approche géométrique

Considérons la figure suivante.

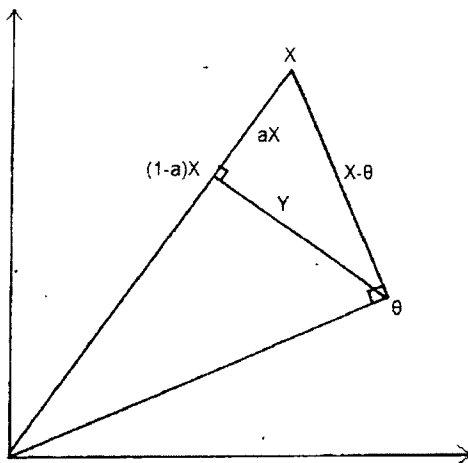


Figure 2.1 – Illustration géométrique de l'estimation du paramètre  $\theta$  à partir de l'observation  $X$ .

Nous savons que  $E[(X - \theta)'\theta] = 0$ , ce qui nous suggère que ces deux vecteurs sont en moyenne orthogonaux. Nous savons aussi que  $E[\|X\|^2] = p\sigma^2 + \|\theta\|^2$ , ce qui pourrait nous faire croire que  $X$  est un estimateur trop long de  $\theta$  dans le sens où  $E[\|X\|^2] > \|\theta\|^2$ . La projection  $(1 - a)X$  de  $\theta$  sur  $X$  serait donc une meilleure approximation. Considérons  $Y$  comme dans la figure 2.1. En supposant que  $X - \theta$  et  $\theta$  soient vraiment orthogonaux et que  $\|X\|^2$  mesure vraiment  $p\sigma^2 + \|\theta\|^2$ , nous avons d'un côté :

$$\begin{aligned}\|Y\|^2 &= \|X - \theta\|^2 - a^2 \|X\|^2 \\ &= p\sigma^2 - a^2 \|X\|^2\end{aligned}$$

et de l'autre

$$\begin{aligned}\|Y\|^2 &= \|\theta\|^2 - (1 - a)^2 \|X\|^2 \\ &= \|X\|^2 - p\sigma^2 - (1 - a)^2 \|X\|^2\end{aligned}$$

En égalant ces deux expressions, on obtient :  $a = \frac{p\sigma^2}{\|X\|^2}$  et  $\hat{\theta} = (1 - a)X = (1 - \frac{p\sigma^2}{\|X\|^2})X$ .

Cet estimateur serait donc plus approprié intuitivement.

## 2.2 Estimateur de Bayes empirique

Partons toujours du problème où on cherche à estimer un vecteur paramétrique  $\theta$  à partir d'une variable aléatoire  $X \sim N_p(\theta, \sigma^2 I_p)$ . Supposons de plus que nous disposons de la loi *a priori* telle que  $\theta \sim N_p(0, bI_p)$  où la constante  $\sigma^2$  est connue alors que  $b$  ne l'est pas.

Dans ce cas, la loi *a posteriori* de  $\theta|X = x$  est une loi normale de moyenne  $\frac{b}{\sigma^2 + b}X = (1 - \frac{\sigma^2}{\sigma^2 + b})X$ . Il s'agit de l'estimateur de Bayes pour la perte quadratique  $L_Q$ . Il reste donc à estimer la valeur de  $b$ . Comme  $(X - \theta)|\theta \sim N_p(0, \sigma^2 I_p)$ ,  $X - \theta$  est indépendant de  $\theta$ . On peut en conclure que  $X = X - \theta + \theta$  a pour distribution marginale une loi normale de moyenne 0 et de matrice de variance-covariance  $(\sigma^2 + b)I_p$ . On peut ainsi souligner que  $\frac{\|X\|^2}{b + \sigma^2} \sim \chi^2(p)$  (Chi-deux avec  $p$  degrés de liberté). De là, on peut utiliser le résultat suivant :  $E[\frac{b + \sigma^2}{\|X\|^2}] = \frac{1}{p - 2}$ . Donc, comme  $b$  est inconnu, on peut approximer l'estimateur  $(1 - \frac{\sigma^2}{\sigma^2 + b})X$  par  $(1 - \frac{(p-2)\sigma^2}{\|X\|^2})X$  qui est un estimateur de Bayes empirique correspondant à  $\delta_{JS}(X)$ .

## 2.3 Résultats de dominance

Les deux sous-sections précédentes nous ont proposé des estimateurs de la forme  $(1 - \frac{a\sigma^2}{\|X\|^2})X$  pour un paramètre de position  $\theta$  qui semblent plus appropriés que  $X$ . Nous retraçons ici le résultat à l'effet que l'estimateur  $X$  est dominé pour la perte quadratique  $L_Q$  par ces estimateurs, soient celui obtenu par l'approche géométrique et l'autre obtenu par l'approche Bayes empirique.

Nous avons besoin du résultat suivant de Stein ([22]). Nous dirons qu'une fonction  $f(x)$  est faiblement différentiable s'il existe une fonction  $\psi(x) \in L^1(\mathbb{R})$  telle que  $f(x) = \int_a^x \psi(x)dx$ . De plus, la divergence d'une fonction  $f(x)$  est  $\nabla \cdot f(x) = \sum_{i=1}^p \frac{\partial}{\partial x_i} f(x)$ .

**Lemme 2.3.1. ([22],[23]) (Lemme de Stein)** *Soit  $X \sim N_p(\theta, \sigma^2 I_p)$  et  $g(x) : \mathbb{R}^p \mapsto \mathbb{R}^p$  une fonction faiblement différentiable telle que  $E[\|g(X)\|^2] < \infty$ . Alors,  $E[(X-\theta)g(X)] = \sigma^2 E[\nabla \cdot g(X)]$ .*

*Démonstration.* Considérons d'abord le cas unidimensionnel où  $X \sim N(0, 1)$  et  $g(x) : \mathbb{R} \mapsto \mathbb{R}$ . L'intégration par parties nous donne :

$$\begin{aligned} E[Xg(X)] &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} xg(x)e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} ([-g(x)e^{-\frac{x^2}{2}}]_{x \rightarrow -\infty}^{\infty} + \int_{-\infty}^{\infty} g'(x)e^{-\frac{x^2}{2}} dx) \\ &= E[g'(X)], \end{aligned}$$

car  $E[\|g(X)\|] < \infty$ . Le cas où  $X \sim N_p(0, I_p)$  et où  $g(X) : \mathbb{R}^p \mapsto \mathbb{R}^p$  se prouve en utilisant ce dernier résultat. On remarque que

$$\begin{aligned}
E[X'g(X)] &= E[\sum X_i g_i(X)] \\
&= \sum E[X_i g_i(X)] \\
&= \sum E[\frac{\partial}{\partial x_i} g_i(X)] \\
&= E[\sum \frac{\partial}{\partial x_i} g_i(X)] \\
&= E[\nabla \cdot g(X)].
\end{aligned}$$

Finalemment, le résultat pour  $X \sim N_p(\theta, \sigma^2 I_p)$  s'obtient en effectuant le changement de variable  $Y = \frac{X-\theta}{\sigma}$ .  $\square$

Le théorème qui suit a bouleversé les croyances sur l'estimateur  $X$  intuitif et considéré efficace pour une moyenne, car il est minimax, sans biais et estimateur du maximum de vraisemblance.

**Théorème 2.3.2. [8]** *Soit  $X \sim N_p(\theta, \sigma^2 I_p)$ . Pour  $p \geq 3$ , l'estimateur  $\delta_a(X) = (1 - \frac{a\sigma^2}{\|X\|^2})X$  de  $\theta$  domine l'estimateur  $X$  pour  $0 < a < 2(p-2)$  et pour la perte quadratique  $\|\delta - \theta\|^2$ . De plus,  $\delta_{p-2}(X)$  est celui qui minimise le risque de cette classe. Ce minimum est uniquement atteint en  $\theta = 0$  et vaut  $2\sigma^2$ , car le risque est croissant en  $\|\theta\|$ .*

*Démonstration.* À l'aide du Lemme 2.3.1, on a :

$$\begin{aligned}
R(\theta, \delta_a) &= E_\theta[\|(1 - \frac{a\sigma^2}{\|X\|^2})X - \theta\|^2] \\
&= E_\theta[\|X - \theta\|^2] + a^2\sigma^4 E_\theta[\frac{1}{\|X\|^2}] - 2a\sigma^2 E_\theta[\frac{X'(X-\theta)}{\|X\|^2}] \\
&= p\sigma^2 + a^2\sigma^4 E_\theta[\frac{1}{\|X\|^2}] - 2a\sigma^2 \sum_{i=1}^p E_\theta[\frac{X_i(X_i-\theta)}{\sum_{j=1}^p X_j^2}] \\
&= p\sigma^2 + a^2\sigma^4 E_\theta[\frac{1}{\|X\|^2}] - 2a\sigma^4 \sum_{i=1}^p E_\theta[\frac{d}{dx_i} \frac{X_i}{\sum_{j=1}^p X_j^2}] \\
&= p\sigma^2 + a^2\sigma^4 E_\theta[\frac{1}{\|X\|^2}] - 2a\sigma^4 \sum_{i=1}^p E_\theta[\frac{\sum_{j=1}^p X_j^2 - 2X_i^2}{(\sum_{j=1}^p X_j^2)^2}] \\
&= p\sigma^2 + a^2\sigma^4 E_\theta[\frac{1}{\|X\|^2}] - 2a\sigma^4 E_\theta[\frac{p\|X\|^2 - 2\|X\|^2}{(\|X\|^2)^2}] \\
&= p\sigma^2 + \sigma^4(a^2 - 2a(p-2)) E_\theta[\frac{1}{\|X\|^2}].
\end{aligned}$$

Comme  $a^2 - 2a(p - 2) = a(a - 2(p - 2))$  est négatif pour  $0 < a < 2(p - 2)$ , on peut conclure que  $\delta_a(X)$  domine  $X$  pour  $a$  dans cet intervalle. De plus, ce polynôme atteint son minimum en  $a = p - 2$ .

Finalement, lorsque  $\theta = 0$ ,  $\frac{\|X\|^2}{\sigma^2}$  a pour distribution une loi du khi-deux avec  $p$  degrés de liberté et le risque est donc de :

$$R(0, \delta_{p-2}) = p\sigma^2 - \sigma^2(p - 2)^2 E\left[\frac{\sigma^2}{\|X\|^2}\right] = \sigma^2\left(p - \frac{(p - 2)^2}{p - 2}\right) = 2\sigma^2$$

Finalement, le risque minimal de l'estimateur  $\delta_{p-2}(X)$  est uniquement atteint pour  $\theta = 0$ , car  $R(\theta, \delta_{p-2}(X)) = p\sigma^2 - \sigma^4(p - 2)^2 E_{\theta}\left[\frac{1}{\|X\|^2}\right]$  est monotone croissante en  $\|\theta\|$ .  $\square$

Tel que décrit par Brandwein et Strawderman ([3]), les deux généralisations qui suivent sont à remarquer.

(i) Le résultat tient pour un changement d'origine. Si on considère la variable  $Y = X - \theta_0$ , l'estimateur  $\delta_{p-2}(Y) = \left(1 - \frac{(p-2)\sigma^2}{\|Y\|^2}\right)Y$  de  $\theta - \theta_0$  domine  $Y$  avec un risque minimal de  $2\sigma^2$  en  $\theta - \theta_0 = 0$ . On peut donc conclure que l'estimateur  $\delta(X) = \theta_0 + \left(1 - \frac{(p-2)\sigma^2}{\|X - \theta_0\|^2}\right)(X - \theta_0)$  de  $\theta$  domine  $X$  avec un risque minimal de  $2\sigma^2$  en  $\theta = \theta_0$ , car  $R(\theta - \theta_0, \delta_{p-2}) = R(\theta, \delta_{p-2} + \theta_0)$  pour tout  $\theta \in \mathbb{R}^p$ . Cette généralisation est utile lorsqu'on a une idée de la valeur de  $\theta$ .

(ii) Sachant que  $\theta$  appartient à un sous-espace linéaire  $\mathbb{V}$  de  $\mathbb{R}^p$  de dimension  $s$ , on peut considérer la projection de  $X$  sur ce sous-espace additionné à l'estimateur de Stein appliquée à la projection de  $X$  sur le sous-espace orthogonal au premier. Le risque minimal est de  $\sigma^2(s + 2)$ . Il faut par contre que  $p - s \geq 3$ . Lindley ([16]) propose un estimateur de ce type où  $\mathbb{V} = \{\theta : \theta_1 = \theta_2 = \dots = \theta_p\}$ .

## 2.4 Estimateurs de Bayes : Résultats de minimaxité

Considérons toujours le problème où  $X \sim N_p(\theta, \sigma^2 I_p)$  et où on cherche à estimer ponctuellement le paramètre  $\theta$  sous la perte quadratique  $L(\theta, \delta) = \|\delta - \theta\|^2$ . Nous nous concentrons ici sur les estimateurs de Bayes. Leur forme particulière nous permettra de trouver des conditions de minimaxité.

**Lemme 2.4.1. ([22],[23])** *Pour un estimateur de la forme  $\delta(X) = X + \sigma^2 g(X)$ , le risque est égal à  $R(\theta, \delta) = p\sigma^2 + E[\sigma^4(\|g(X)\|^2 + 2\nabla \cdot g(X))]$ . Ainsi, si  $\|g(x)\|^2 + 2\nabla \cdot g(x) \leq 0$  pour tout  $x \in \mathbb{R}^p$ , alors  $\delta(X)$  est minimax, car le risque est uniformément inférieur au risque de  $X$  qui est de  $p\sigma^2$ .*

*Démonstration.* On a :

$$\begin{aligned} R(\theta, \delta) &= E[\|X + \sigma^2 g(X) - \theta\|^2] \\ &= E[\|X - \theta\|^2] + \sigma^4 E[\|g(X)\|^2] + 2\sigma^2 E[g(X)'(X - \theta)] \\ &= p\sigma^2 + \sigma^4 E[\|g(X)\|^2 + 2\nabla \cdot g(X)], \end{aligned}$$

par le Lemme 2.3.1. □

**Exemple 2.4.2.** *L'estimateur de James-Stein  $\delta_{p-2}(X) = (1 - \frac{(p-2)\sigma^2}{\|X\|^2})X$  est un exemple du type d'estimateur décrit plus haut. Dans ce cas, on a  $g(X) = -\frac{p-2}{\|X\|^2}X$ .*

Le prochain lemme illustre la forme particulière d'un estimateur de Bayes.

**Lemme 2.4.3. ([22],[23])** *Si  $\theta$  suit une loi a priori  $\pi(\theta)$  telle que la marginale  $m(\cdot)$  existe, alors l'estimateur de Bayes est de la forme  $\delta_\pi(X) = X + \sigma^2 \frac{\nabla m(X)}{m(X)}$ .*

*Démonstration.* On a :

$$\begin{aligned}
\nabla m(X) &= [\frac{\delta}{\delta x_i} \int f(x|\theta)\pi(\theta)d\theta]_{i=1\dots p} \\
&= [\int -\frac{(x_i-\theta_i)}{\sigma^2} f(x|\theta)\pi(\theta)d\theta]_{i=1\dots p} \\
&= [-\frac{x_i}{\sigma^2}m(x) + \frac{m(x)}{\sigma^2}E^{\theta|X}[\theta_i]]_{i=1\dots p} \\
&= \frac{m(X)}{\sigma^2}(-X + \delta_\pi(X)) \\
\Rightarrow \delta_\pi(X) &= X + \sigma^2 \frac{\nabla m(X)}{m(X)}.
\end{aligned}$$

□

Le définition suivante est utile pour comprendre le Théorème 2.4.6.

**Définition 2.4.4.** Une fonction continue  $\phi : \mathbb{R}^p \rightarrow \mathbb{R}$  est dite superharmonique si pour tout  $x_0 \in \mathbb{R}^p$  et pour tout  $r > 0$ , on a  $\phi(x_0) \geq \frac{\Gamma(\frac{p}{2})}{(2\pi)^{\frac{p}{2}} r^{\frac{p}{2}-1}} \int_{S_{r,x_0}} \phi(x) dU_{S_{r,x_0}}(x)$ , où  $S_{r,x_0} = \{x \in \mathbb{R}^p : \|x - x_0\| = r\}$  est la sphère de rayon  $r$  centrée en  $x_0$  et  $U_{S_{r,x_0}}$  est la mesure uniforme sur cette sphère.

**Remarque 2.4.5.** – La quantité  $\frac{(2\pi)^{\frac{p}{2}} r^{\frac{p}{2}-1}}{\Gamma(\frac{p}{2})}$  correspond à l'aire de la sphère  $S_{r,x_0}$ .

- Pour une fonction superharmonique, on peut affirmer qu'en tout point, la valeur de la fonction est supérieure ou égale à la moyenne des valeurs de la fonction sur une sphère autour de ce point de rayon arbitraire.
- Une définition équivalente lorsque les dérivées partielles d'ordre 2 existent serait qu'une fonction  $\phi$  est superharmonique si  $\nabla^2 \phi = \sum_{i=1}^p \frac{\partial^2}{\partial x_i^2} \phi(x) \leq 0$ , où  $\nabla^2 = \nabla \cdot \nabla$  correspond au Laplacien.

**Théorème 2.4.6. ([22],[23])** Si  $\theta$  suit une loi a priori  $\pi(\theta)$  telle que la marginale  $m(\cdot)$  existe, alors  $R(\theta, \delta_\pi) = p\sigma^2 + 4\sigma^4 E[\frac{\nabla^2 \sqrt{m(X)}}{\sqrt{m(X)}}]$  et  $\delta_\pi$  est donc minimax si  $\sqrt{m(X)}$  est une fonction superharmonique.

*Démonstration.* En regard des deux lemmes précédents, on sait que  $R(\theta, \delta_\pi) = p\sigma^2 + \sigma^4 E[\|g(X)\|^2 + 2\nabla \cdot g(X)]$  où  $g(X) = \frac{\nabla m(X)}{m(X)}$ . Remarquons que l'intérieur de l'espérance est alors égale à :

$$\begin{aligned} \left\| \frac{\nabla m(X)}{m(X)} \right\|^2 + 2\nabla \cdot \frac{\nabla m(X)}{m(X)} &= \frac{\|\nabla m(X)\|^2}{m(X)^2} + 2 \frac{m(X)\nabla^2 m(X) - \|\nabla m(X)\|^2}{m(X)^2} \\ &= \frac{2m(X)\nabla^2 m(X) - \|\nabla m(X)\|^2}{m(X)^2}. \end{aligned}$$

Mais, d'un autre côté, nous avons :

$$\begin{aligned} \frac{4\nabla^2 \sqrt{m(X)}}{\sqrt{m(X)}} &= \frac{4}{\sqrt{m(X)}} \nabla \cdot \nabla \sqrt{m(X)} \\ &= \frac{4}{\sqrt{m(X)}} \nabla \cdot \left( \frac{\nabla m(X)}{2\sqrt{m(X)}} \right) \\ &= \frac{2}{\sqrt{m(X)}} \frac{\sqrt{m(X)\nabla^2 m(X)} - \nabla m(X) \cdot \frac{\nabla m(X)}{2\sqrt{m(X)}}}{m(X)} \\ &= \frac{2m(X)\nabla^2 m(X) - \|\nabla m(X)\|^2}{m(X)^2}. \end{aligned}$$

□

Le résultat suivant nous assure de la minimaxité d'un estimateur de Bayes dès que la loi *a priori*  $\pi(\theta)$  est superharmonique ou dès que la loi marginale de  $X$  est superharmonique.

**Proposition 2.4.7.** *Soit une loi a priori  $\pi(\cdot)$  superharmonique telle que la marginale  $m_X(\cdot)$  existe. Alors  $m_X(\cdot)$  est superharmonique et  $\sqrt{m_X(\cdot)}$  est superharmonique.*

*Démonstration.* Supposons  $\pi(\cdot)$  superharmonique. Alors, on a :

$$\begin{aligned} \nabla_x^2 m_X(x) &= \nabla_x^2 \int f_{X|\theta}(x) \pi(\theta) d\theta \\ &= \int \nabla_x \cdot (\nabla_x f_{X|\theta}(x)) \pi(\theta) d\theta \\ &= \int \nabla_x \cdot \left[ -\frac{1}{2\sigma^2} (2(x_i - \theta_i)) f_{X|\theta}(x) \right]_{i=1}^p \pi(\theta) d\theta \\ &= \int \sum_{i=1}^p \left( -\frac{1}{\sigma^2} + \frac{(x_i - \theta_i)^2}{\sigma^4} \right) f_{X|\theta}(x) \pi(\theta) d\theta \\ &= -\frac{p}{\sigma^2} m_X(x) + \frac{1}{\sigma^4} E^\theta [(\theta - X)'(\theta - X) \pi(\theta)] \quad (\text{où } \theta \sim N(x, \sigma^2)) \\ &= -\frac{p}{\sigma^2} m_X(x) + \frac{1}{\sigma^2} E^\theta \left[ \sum_{i=1}^p (\pi(\theta) + (\theta_i - X_i) \frac{\delta}{\delta \theta_i} \pi(\theta)) \right] \quad (\text{par le Lemme de Stein}) \\ &= \frac{1}{\sigma^2} E^\theta [(\theta - X)' \nabla_\theta \pi(\theta)] \\ &= \frac{1}{\sigma^2} E^\theta [\nabla_\theta^2 \pi(\theta)] \quad (\text{par le Lemme de Stein}) \\ &\leq 0. \end{aligned}$$



Ainsi,  $\pi(\theta)$  superharmonique  $\Rightarrow m_X(x)$  superharmonique.

Or, on a aussi que :

$$\begin{aligned}
\nabla^2 \sqrt{m_X(x)} &= \nabla \cdot \left[ \frac{\delta}{\delta x_i} \sqrt{m_X(x)} \right]_{i=1}^p \\
&= \nabla \cdot \left[ \frac{1}{2\sqrt{m_X(x)}} \frac{\delta}{\delta x_i} m_X(x) \right]_{i=1}^p \\
&= \sum_{i=1}^p \left( -\frac{1}{4\sqrt{(m_X(x))^3}} \left( \frac{\delta}{\delta x_i} m_X(x) \right)^2 + \frac{1}{2\sqrt{m_X(x)}} \frac{\delta^2}{\delta x_i^2} m_X(x) \right) \\
&= \sum_{i=1}^p -\frac{1}{4\sqrt{(m_X(x))^3}} \left( \frac{\delta}{\delta x_i} m_X(x) \right)^2 + \frac{1}{2\sqrt{m_X(x)}} \nabla^2 m_X(x) \\
&= \frac{1}{4\sqrt{(m_X(x))^3}} (2m_X(x) \nabla^2 m_X(x) - \|\nabla m_X(x)\|^2) \\
&\leq 0.
\end{aligned}$$

Ainsi,  $m_X(x)$  superharmonique  $\Rightarrow \sqrt{m_X(x)}$  superharmonique.  $\square$

Voici donc une condition suffisante pour obtenir un estimateur minimax.

**Corollaire 2.4.8.** *Dans le cadre du problème d'estimation ponctuelle d'un paramètre  $\theta$  à partir d'une observation  $X|\theta \sim N(\theta, \sigma^2 I_p)$  sous la perte quadratique  $L_Q$ , une condition suffisante pour obtenir un estimateur minimax est que la loi a priori  $\pi(\theta)$  soit superharmonique.*

*Démonstration.* En combinant les résultats du Théorème 2.4.6 et du Corollaire 2.4.8, on obtient le résultat.  $\square$

**Exemple 2.4.9.** *Considérons la famille de lois de la forme  $\pi(\theta) = \frac{1}{\|\theta\|^{2k}}$  où  $k \geq 0$ . On a alors :*

$$\frac{\partial}{\partial x_i} \pi(\theta) = -2k\theta_i \left( \sum_{i=1}^p \theta_i^2 \right)^{-k-1}.$$

Ainsi, le Laplacien équivaut à :

$$\begin{aligned}
\nabla^2 \pi(\theta) &= \sum_{i=1}^p (-2k(\sum_{i=1}^p \theta_i^2)^{-k-1} + 4k(k+1)\theta_i^2(\sum_{i=1}^p \theta_i^2)^{-k-2}) \\
&= \sum_{i=1}^p \frac{-2k}{(\sum_{i=1}^p \theta_i^2)^{k+2}} (\sum_{i=1}^p \theta_i^2 - 2(k+1)\theta_i^2) \\
&= \frac{-2k}{(\sum_{i=1}^p \theta_i^2)^{k+2}} (p \sum_{i=1}^p \theta_i^2 - 2(k+1) \sum_{i=1}^p \theta_i^2) \\
&= \frac{-2k}{(\sum_{i=1}^p \theta_i^2)^{k+1}} (p - 2(k+1)).
\end{aligned}$$

Ainsi,  $\pi(\theta)$  est superharmonique si  $k \leq \frac{p}{2} - 1$ . On peut donc conclure que, pour  $p \geq 3$  et  $k \leq \frac{p}{2} - 1$ , l'estimateur de Bayes associé à cette loi a priori est minimax et par conséquent admissible pour la perte Kullback-Leibler.

En conclusion, il a été démontré, à partir de résultats tirés de la littérature et dans le cadre du problème d'estimation ponctuelle d'un paramètre  $\theta$  à partir d'une observation  $X|\theta \sim N(\theta, \sigma^2 I_p)$  sous la perte quadratique, que l'estimateur  $X$  n'est pas le meilleur choix lorsque  $p \geq 3$ . Il est dominé par l'estimateur de James-Stein et une grande classe d'estimateurs de la forme  $X + g(X)$  et ces estimateurs dominant  $X$  sont minimax. Une condition suffisante a de plus été trouvée pour obtenir un estimateur minimax, à savoir que la loi a priori  $\pi(\theta)$  soit superharmonique.

# CHAPITRE 3

## Densités prédictives pour la loi Multinormale

Dans ce chapitre, nous nous attarderons principalement à l'estimation d'une densité de loi normale, aux conditions de minimaxité, de dominance ainsi qu'aux liens reliant ce problème au problème d'estimation ponctuelle abordé au chapitre 2. Le développement de la section 3.2 est tiré des travaux de George, Liang et Xu ([8]) et de Brown, George et Xu ([4]). Le développement de la section 3.3, quant à lui, présente des résultats de Fourdrinier, Marchand, Righi et Strawderman ([7]) ainsi que de George, Liang et Xu ([8]). Les contributions de ce mémoire se situent au niveau de l'estimation pour un espace paramétrique restreint, des exemples illustrés et de l'assemblage de plusieurs résultats connus.

### Problème (\*\*)

**Considérons le problème où on cherche à estimer la fonction de densité d'une variable aléatoire  $Y \sim N_p(\theta, v_y I_p)$  à partir de l'observation de  $X \sim N_p(\theta, v_x I_p)$  pour prédire un futur comportement de  $Y$ . Les variances  $v_x$  et  $v_y$  sont connues.**

**La perte utilisée est la perte de Kullback-Leibler.**

Nous comparerons l'efficacité d'un estimateur par substitution avec l'estimateur de Bayes découlant de lois *a priori* non informatives. Nous établirons aussi des liens avec des problèmes où il y a une contrainte sur le paramètre  $\theta$ .

### 3.1 Résultats de minimaxité et de dominance

Tout d'abord, si on dispose d'une loi *a priori* de  $\theta$  (paramètre inconnu) absolument continue par rapport à la mesure de Lebesgue, on a vu dans la Proposition 1.2.4 que l'estimateur de Bayes pour la perte Kullback-Leibler est  $\hat{p}(y|x) = \int p(y|\theta)\pi(\theta|x)d\theta$ .

Or, si on n'a aucune information sur la distribution de  $\theta$ , quelle loi *a priori* peut-on choisir ? Voilà la question qui rend certains statisticiens sceptiques. Nous considérerons des lois *a priori* non-informatives pour éviter toute subjectivité du statisticien. De telles lois sont présentées au chapitre 3 du livre de Robert ([20]).

**Définition 3.1.1.** *Une loi a priori dont la mesure de l'ensemble total est infinie mais conduisant à une loi a posteriori de mesure 1 est dite impropre.*

Les lois impropres sont souvent utilisées dans le domaine de la statistique bayésienne, car elles possèdent, mais pas toujours, de bonnes propriétés fréquentistes. Elles représentent souvent des choix par défaut ou appropriés lorsqu'il n'y a pas d'information *a priori* explicite.

**Exemple 3.1.2.** *Dans le cadre du problème (\*\*), en imposant la loi a priori uniforme sur les réels et en considérant que l'estimateur de Bayes se calcule comme énoncé dans le chapitre 1, on obtient :*

$$\begin{aligned}
\widehat{p}_U(y|x) &= \int \frac{1}{(2\pi v_x)^{\frac{p}{2}}} e^{-\frac{1}{2v_x} \|\theta-x\|^2} \frac{1}{(2\pi v_y)^{\frac{p}{2}}} e^{-\frac{1}{2v_y} \|y-\theta\|^2} d\theta \\
&= \frac{1}{(2\pi)^{\frac{p}{2}}} \int \frac{1}{(2\pi v_x v_y)^{\frac{p}{2}}} e^{-\frac{1}{2v_x v_y} (v_y \|x\|^2 + v_x \|y\|^2 + (v_x + v_y) \|\theta\|^2 - 2(v_y x + v_x y)' \theta)} d\theta \\
&= \frac{1}{(2\pi(v_x + v_y))^{\frac{p}{2}}} \int \left( \frac{v_x + v_y}{2\pi v_x v_y} \right)^{\frac{p}{2}} e^{-\frac{(v_x + v_y)}{2v_x v_y} (\|\theta\|^2 - 2\left(\frac{v_y x + v_x y}{v_x + v_y}\right)' \theta + \frac{v_y \|x\|^2 + v_x \|y\|^2}{v_x + v_y})} d\theta \\
&= \frac{1}{(2\pi(v_x + v_y))^{\frac{p}{2}}} e^{-\frac{(v_x + v_y)}{2v_x v_y} \left( \frac{v_y \|x\|^2 + v_x \|y\|^2}{v_x + v_y} - \frac{v_y^2 \|x\|^2 + v_x^2 \|y\|^2 + 2v_x v_y x' y}{(v_x + v_y)^2} \right)} \\
&\quad \cdot \int \left( \frac{v_x + v_y}{2\pi v_x v_y} \right)^{\frac{p}{2}} e^{-\frac{(v_x + v_y)}{2v_x v_y} \left\| \theta - \frac{v_y x + v_x y}{v_x + v_y} \right\|^2} d\theta \\
&= \frac{1}{(2\pi(v_x + v_y))^{\frac{p}{2}}} e^{-\frac{(v_x + v_y)}{2v_x v_y} \frac{(v_x + v_y) v_y \|x\|^2 + (v_x + v_y) v_x \|y\|^2 - v_y^2 \|x\|^2 - v_x^2 \|y\|^2 + 2v_x v_y x' y}{(v_x + v_y)^2}} \\
&= \frac{1}{(2\pi(v_x + v_y))^{\frac{p}{2}}} e^{-\frac{1}{2(v_x + v_y)} \|y-x\|^2},
\end{aligned}$$

qui est la densité d'une loi normale de moyenne  $x$  et de matrice de variances-covariances  $(v_x + v_y)I_p$ . On aurait pu retrouver cette densité en considérant le problème de l'Exemple 1.2.5. Il suffit de prendre  $\mu = 0$  et de faire tendre  $v_\theta$  vers l'infini pour retrouver la loi a priori uniforme sur les réels. Dans ce cas particulier, on a que  $\mu(X) = X$  et que  $v(X) = v_x$ , ce qui nous conduit à la fonction de densité ci-haut. Nous ferons dorénavant référence à cette densité par la notation  $\widehat{p}_U(y|x)$ .

Le résultat suivant appuie l'idée qu'il s'agit d'un bon estimateur.

**Proposition 3.1.3.** *Pour le problème (\*\*), parmi la classe d'estimateurs  $\{\widehat{p}(y|x) \sim N(x, cv_y I_p)\}$ ,  $\widehat{p}_U$  est celui qui minimise le risque pour tout  $\theta \in \mathbb{R}^p$ .*

*Démonstration.* Comme  $Y - X|\theta \sim N_p(0, (v_x + v_y)I_p)$ , on a :

$$\begin{aligned}
R(\theta, \widehat{p}_c) &= E^{X,Y} [\log(\frac{p(Y|\theta)}{\widehat{p}_c(Y|X)})] \\
&= E^{X,Y} [\log(c^{\frac{p}{2}}) - \frac{\|Y-\theta\|^2}{2v_y} + \frac{\|Y-X\|^2}{2cv_y}] \\
&= \frac{p}{2}(\log(c) - 1 + \frac{v_x + v_y}{cv_y})
\end{aligned}$$

minimisé en  $c = \frac{v_x + v_y}{v_y}$ . □

Cet estimateur, minimisant le risque sous la perte Kullback-Leibler parmi les estimateurs de la classe définie plus tôt, a pour effet d'augmenter la variance de la fonction de densité. Au lieu de conserver la variance  $v_y$  de la variable  $Y$ , elle additionne la variance  $v_x$  de la variable  $X$ . Cette inflation est due à l'incertitude sur le paramètre  $\theta$ , interprétable comme une correction par la substitution du paramètre  $\theta$  par  $X$ .

Pour le problème d'estimation par densité prédictive sous la perte Kullback-Leibler, quelle est la performance des estimateurs par substitution où on ne fait que remplacer le paramètre inconnu dans la fonction de densité par un estimateur? Les résultats suivants apportent une réponse.

**Corollaire 3.1.4.** *Dans le cadre du problème (\*\*), l'estimateur  $\hat{p}_U(y|x)$  domine l'estimateur par substitution  $\hat{p}_0(y|x) \sim N_p(x, v_y I_p)$ , où  $x$  est l'estimateur du maximum de vraisemblance de  $\theta$  et où  $\hat{p}_0(\cdot|x)$  est l'estimateur du maximum de vraisemblance de  $p(\cdot|y)$ .*

*Démonstration.* Le résultat découle de la Proposition 3.1.3.

□

On remarque en effet que la différence des risques de ces deux estimateurs peut s'exprimer comme suit :

$$\begin{aligned} \Delta(\theta) &= R(\theta, \hat{p}_U) - R(\theta, \hat{p}_0) \\ &= \frac{p}{2}(\log(\frac{v_x+v_y}{v_y}) - 1 + 1) - \frac{p}{2}(\log(1) - 1 + \frac{v_x+v_y}{v_y}) \\ &= \frac{p}{2}(\log(\frac{v_x}{v_y} + 1) - \frac{v_x}{v_y}) \\ &< 0, \end{aligned}$$

où la dernière inégalité est aisément justifiable par l'analyse de la fonction  $f(x) = \log(x + 1) - x$  qui est strictement décroissante sur  $]0, \infty[$  avec  $f(0) = 0$ .

Ainsi, pour  $X$  et  $Y$  suivant des lois normales de même moyenne  $\theta$  inconnue et de matrices de variances-covariances connues  $v_x I_p$  et  $v_y I_p$ , l'estimateur de Bayes pour la densité de  $Y| \theta$

découlant de la loi *a priori* uniforme sur les réels domine l'estimateur par substitution pour la perte Kullback-Leibler. Ce résultat se généralise lorsqu'on observe  $n$  variables provenant de la même loi que  $X$  en remarquant que  $\bar{X}$ , statistique exhaustive de cet ensemble de vecteurs aléatoires observés, suit une loi normale de moyenne  $\theta$  et de variance  $\frac{v_x}{n} I_p$  et peut être considérée comme l'unique observation dans la démonstration.

Le prochain résultat nous donne une raison de plus de prendre en considération l'estimateur  $\hat{p}_U$ .

**Théoreme 3.1.5. ([2],[8])** *Dans le cadre du problème (\*\*), l'estimateur  $\hat{p}_U$  est minimax pour la perte Kullback-Leibler.*

Si nous réussissons à démontrer que  $\hat{p}_U$  est un estimateur de Bayes par extension à risque constant, nous pourrions conclure qu'il est minimax. Pour ce faire, nous avons besoin du lemme suivant.

**Lemme 3.1.6. ([8])** *Dans le cadre du problème (\*\*), on a pour tout  $\pi(\theta)$ ,  $\hat{p}_\pi(y|x) = \frac{m_{\pi,W}(w)}{m_{\pi,X}(x)} \hat{p}_U(y|x)$ , où  $W = \frac{v_y X + v_x Y}{v_x + v_y}$ ,  $m_{\pi,W}(\cdot)$  est la densité marginale de  $W$  et  $m_{\pi,X}(\cdot)$  est la densité marginale de  $X$ .*

*Démonstration.* On a :

$$\begin{aligned}
 p_\pi(x, y) &= \int p(x|\theta)p(y|\theta)\pi(\theta)d\theta \\
 &= \int \frac{1}{(2\pi v_x)^{\frac{p}{2}}} e^{-\frac{\|x-\theta\|^2}{2v_x}} \frac{1}{(2\pi v_y)^{\frac{p}{2}}} e^{-\frac{\|y-\theta\|^2}{2v_y}} \pi(\theta)d\theta \\
 &= \int \frac{1}{(2\pi(v_x+v_y))^{\frac{p}{2}}} e^{-\frac{\|y-x\|^2}{2(v_x+v_y)}} \frac{1}{(2\pi v_w)^{\frac{p}{2}}} e^{-\frac{\|w-\theta\|^2}{2v_w}} \pi(\theta)d\theta \\
 &= \hat{p}_U(y|x)m_\pi(w).
 \end{aligned}$$

□

Par ce lemme, on aurait pu retrouver l'estimateur de Bayes présenté dans l'Exemple 1.2.5 en remarquant que  $m_{\pi,X}(\cdot) \sim N_p(\mu, (v_x + v_\theta)I_p)$  et  $m_{\pi,W}(\cdot) \sim N_p(\mu, (v_w + v_\theta))$ .

Revenons au Théorème 3.1.5 à démontrer :

*Démonstration.* Considérons la suite de lois *a priori*  $\pi_n(\theta) \sim N_p(0, nI_p)$ . Ainsi,

$$\begin{aligned} r_{\pi_n}(\widehat{p}_U) - r_{\pi_n}(\widehat{p}_{\pi_n}) &= \int \pi_n(\theta)(R_{KL}(\theta, \widehat{p}_U) - R_{KL}(\theta, \widehat{p}_{\pi_n}))d\theta \\ &= \int \pi_n(\theta)(E[\log(m_{\pi_n}(W))] - E[\log(m_{\pi_n}(X))])d\theta \\ &= \int \int \pi_n(\theta)p(w|\theta)\log(m_{\pi_n}(w))dw d\theta - \int \int \pi_n(\theta)p(x|\theta)\log(m_{\pi_n}(x))dx d\theta, \end{aligned}$$

$$\text{où } m_{\pi_n}(z) = \frac{1}{(2\pi(v+n))^{\frac{p}{2}}} e^{-\frac{1}{2(v+n)}\|z\|^2}.$$

Il est aisé de vérifier que l'intérieur de chaque intégrale est  $O(\frac{1}{n^p})$ , car on a :

$$\begin{aligned} \pi_n(\theta)p(z|\theta)\log(m_{\pi_n}(z)) &= \frac{1}{(2\pi n)^{\frac{p}{2}}} e^{-\frac{\|\theta\|^2}{2n}} \frac{1}{(2\pi v_z)^{\frac{p}{2}}} e^{-\frac{\|z-\theta\|^2}{2v_z}} \log\left(\frac{1}{(2\pi(v_z+n))^{\frac{p}{2}}} e^{-\frac{\|z\|^2}{2(v_z+n)}}\right) \\ &= C \frac{e^{-\frac{\|\theta\|^2}{2n}}}{n^{\frac{p}{2}}} \left(-\frac{\|z\|^2}{2(v_z+n)} + \log((2\pi(v_z+n))^{-\frac{p}{2}})\right) \\ &\leq \frac{C}{n^{\frac{p}{2}}} \left(-\frac{\|z\|^2}{2(v_z+n)} + \frac{1}{(2\pi(v_z+n))^{\frac{p}{2}}}\right) \\ &\leq \frac{C}{n^{\frac{p}{2}}} \frac{1}{(2\pi n)^{\frac{p}{2}}} \\ &= \frac{C}{n^p}. \end{aligned}$$

Ainsi, chaque intégrale tend vers 0 lorsque  $n$  tend vers l'infini. Donc,  $\widehat{p}_U$  est Bayes par extension. De plus, par la Proposition 3.1.3, on sait que  $\widehat{p}_U$  a un risque constant pour la perte Kullback-Leibler. Par le Théorème 1.1.6, il est minimax.  $\square$

En résumé, l'estimateur  $\widehat{p}_U$  est minimax et domine l'estimateur par substitution  $\widehat{p} \sim N_p(x, v_y I_p)$ . Par contre, cet estimateur n'est pas toujours admissible, comme le démontre les prochains résultats.

**Lemme 3.1.7.** *Dans le cadre du problème (\*\*), l'estimateur  $\widehat{p}_1(\cdot|x) \sim N_p(\delta(x), cI_p)$  domine l'estimateur  $\widehat{p}_2(\cdot|x) \sim N_p(x, cI_p)$  si et seulement si  $\delta(X)$  domine  $X$  comme estimateur de  $\theta \in C$  sous la perte  $\|\theta - \delta\|^2$ .*



*Démonstration.* On a :

$$\begin{aligned} R(\theta, \hat{p}_1) - R(\theta, \hat{p}_2) &= E^{X,Y} \left[ \log \left( \frac{(2\pi c)^{-\frac{p}{2}} e^{-\frac{1}{2c} \|Y-X\|^2}}{(2\pi c)^{-\frac{p}{2}} e^{-\frac{1}{2c} \|Y-\delta(X)\|^2}} \right) \right] \\ &= \frac{1}{2c} E^{X,Y} [\|Y - \delta(X)\|^2 - \|Y - X\|^2]. \end{aligned}$$

□

Le prochain corollaire a été démontré par Komaki ([11]). En voici une autre démonstration.

**Corollaire 3.1.8.** *Dans le cadre du problème (\*\*),  $\hat{p}_U$  est inadmissible pour  $p \geq 3$ .*

*Démonstration.* Comme démontré dans le Théorème 2.3.2, on peut trouver un estimateur de Bayes  $\delta_\pi$  qui domine l'estimateur  $X$  pour  $p \geq 3$ . Ainsi, par le Lemme 3.1.7,  $\hat{p}_U$  est inadmissible pour  $p \geq 3$ . □

## 3.2 Lien entre l'estimation par densité prédictive et l'estimation ponctuelle

Le chapitre 2 concernant l'estimation ponctuelle ne fut pas élaboré sans objectif. Dans cette section, nous tenterons d'établir le lien qui unit les deux problèmes d'estimation et ainsi retrouver des résultats similaires dans les deux cas. Le premier théorème met en relation le risque quadratique et le risque Kullback-Leibler.

**Théorème 3.2.1.** ([4],[7]) *Dans le cadre du problème (\*\*) et considérant  $v_w = \frac{v_x v_y}{v_x + v_y}$ ,*

(a) *Si  $\theta \sim \pi(\theta)$ , alors  $R_{KL}(\theta, \hat{p}_U) - R_{KL}(\theta, \hat{p}_\pi) = \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (R_Q^v(\theta, X) - R_Q^v(\theta, \hat{\theta}_{\pi,v})) dv$  où  $\hat{\theta}_{\pi,v}$  est l'estimateur de Bayes de  $\theta$  sous la perte quadratique et sous la loi a priori  $\pi(\cdot)$ .*

(b) *Pour un estimateur par substitution  $\hat{p}_1 \sim N_p(\delta_1(X), v_y I_p)$ , on a  $R_{KL}(\theta, \hat{p}_1) - R_{KL}(\theta, \hat{p}_U) = \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (R_Q^{v_x}(\theta, \delta_1) - R_Q^v(\theta, X)) dv$ .*

Démonstration. (a) Soit  $Z \sim N_p(\theta, vI_p)$  et posons  $Z^* = \frac{Z-\theta}{\sqrt{v}}$ . On a :

$$\begin{aligned} \frac{\partial}{\partial v} E^Z [\log(m_\pi(Z))] &= \frac{\partial}{\partial v} E^{Z^*} [\log(m_\pi(\sqrt{v}Z^* + \theta))] \\ &= E^{Z^*} \left[ \frac{\frac{\partial}{\partial v} m_\pi(\sqrt{v}Z^* + \theta)}{m_\pi(\sqrt{v}Z^* + \theta)} \right] \end{aligned}$$

où

$$\begin{aligned} \frac{\partial}{\partial v} (\sqrt{v}z^* + \theta) &= \frac{\partial}{\partial v} \int \frac{1}{(2\pi v)^{\frac{p}{2}}} e^{-\frac{1}{2v} \|\sqrt{v}z^* + \theta - t\|^2} \pi(t) dt \\ &= \int \left( -\frac{p}{2v} - \frac{1}{2} \left( -\frac{1}{v^2} \|\theta - t\|^2 - \frac{\sqrt{v}}{v^2} z^{*\prime} (\theta - t) \right) \right) f(z|t) \pi(t) dt \\ &= \int \left( -\frac{p}{2v} + \frac{1}{2v^2} (\|z - \sqrt{v}z^* - t\|^2 + \sqrt{v}z^{*\prime} (\theta - t)) \right) f(z|t) \pi(t) dt \\ &= \int \left( -\frac{p}{2v} + \frac{1}{2v^2} (\|z - t\|^2 + v \|z^*\|^2 - 2\sqrt{v}z^{*\prime} (z - t) + \sqrt{v}z^{*\prime} (\theta - t)) \right) \\ &\quad f(z|t) \pi(t) dt \\ &\quad + \sqrt{v}z^{*\prime} (\theta - t)) f(z|t) \pi(t) dt \\ &= \int \left( -\frac{p}{2v} + \frac{1}{2v^2} (\|z - t\|^2 - v \|z^*\|^2 - \sqrt{v}z^{*\prime} (\theta - t)) \right) f(z|t) \pi(t) dt \\ &= \int \left( -\frac{p}{2v} + \frac{1}{2v^2} (\|z - t\|^2 - \|z - \theta\|^2 - (z - \theta)' (\theta - t)) \right) f(z|t) \pi(t) dt \\ &= \int \left( -\frac{p}{2v} + \frac{1}{2v^2} (\|z - t\|^2 - (z - \theta)' (z - t)) \right) f(z|t) \pi(t) dt \\ &= \frac{\partial}{\partial v} m_\pi(z) - \int \frac{1}{2v^2} (z - \theta)' (z - t) f(z|t) \pi(t) dt. \end{aligned}$$

Ainsi,

$$\begin{aligned} E^{Z^*} \left[ \frac{\frac{\partial}{\partial v} m_\pi(\sqrt{v}Z^* + \theta)}{m_\pi(\sqrt{v}Z^* + \theta)} \right] &= E^Z \left[ \frac{\frac{\partial}{\partial v} m_\pi(Z)}{m_\pi(Z)} - \frac{(Z-\theta)' (Z - ET|Z[T])}{2v^2} \right] \\ &= E^Z \left[ \frac{\frac{\partial}{\partial v} m_\pi(Z)}{m_\pi(Z)} - \frac{(Z-\theta)' (Z - (Z + v \frac{\nabla m_\pi(Z)}{m_\pi(Z)}))}{2v^2} \right] \\ &= E^Z \left[ \frac{\frac{\partial}{\partial v} m_\pi(Z)}{m_\pi(Z)} + \frac{1}{2v} v \nabla \cdot \frac{\nabla m_\pi(Z)}{m_\pi(Z)} \right] \\ &= E^Z \left[ \frac{\frac{\partial}{\partial v} m_\pi(Z)}{m_\pi(Z)} + \frac{1}{2} \frac{m_\pi(Z) \nabla^2 m_\pi(Z) - \|\nabla m_\pi(Z)\|^2}{m_\pi(Z)^2} \right] \\ &= E^Z \left[ \frac{\nabla^2 m_\pi(Z)}{2m_\pi(Z)} + \frac{1}{2} \frac{m_\pi(Z) \nabla^2 m_\pi(Z) - \|\nabla m_\pi(Z)\|^2}{m_\pi(Z)^2} \right] \\ &= E^Z \left[ \frac{m_\pi(Z) \nabla^2 m_\pi(Z) - \frac{1}{2} \|\nabla m_\pi(Z)\|^2}{m_\pi(Z)^2} \right] \\ &= -\frac{1}{2v^2} (R_Q^v(\theta, Z) - R_Q^v(\theta, \hat{\theta}_{\pi, v})), \end{aligned}$$

où  $T$  est de même loi que  $\theta$ .

Puisque cette dernière expression correspond à  $\frac{\partial}{\partial v} E^Z [\log(m_\pi(Z))]$ , on peut déduire que

$$\frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (R_Q^v(\theta, X) - R_Q^v(\theta, \hat{\theta}_{\pi, v})) dv = E^W [\log(m_\pi(W))] - E^X [\log(m_\pi(X))] = R_{KL}(\theta, \hat{p}_U) - R_{KL}(\theta, \hat{p}_\pi).$$

(b) D'un côté, nous avons :

$$\frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} R_Q^{v_x}(\theta, \delta_1) dv = \frac{R_Q^{v_x}(\theta, \delta_1)}{2} \left( \frac{1}{v_w} - \frac{1}{v_x} \right) = \frac{R_Q^{v_x}(\theta, \delta_1)}{2v_y}$$

et

$$\frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} R_Q^v(\theta, X) dv = \frac{1}{2} \int_{v_w}^{v_x} \frac{p}{v} dv = \frac{p}{2} \log\left(\frac{v_x + v_y}{v_y}\right).$$

D'un autre côté, la différence des risques de ces deux estimateurs nous donne :

$$\begin{aligned} E^{X,Y}[\log(\frac{\hat{p}_U}{\hat{p}_1})] &= E^{X,Y} \left[ -\frac{p}{2} \log\left(\frac{v_x + v_y}{v_y}\right) + \frac{\|Y - \delta_1(X)\|^2}{2v_y} - \frac{\|Y - X\|^2}{2(v_x + v_y)} \right] \\ &= -\frac{p}{2} \log\left(\frac{v_x + v_y}{v_y}\right) + \frac{1}{2v_y} (E^{X,Y}[\|Y - \theta\|^2] + E^{X,Y}[\|\delta_1(X) - \theta\|^2]) - \frac{p}{2} \\ &= -\frac{p}{2} \log\left(\frac{v_x + v_y}{v_y}\right) + \frac{R_Q^{v_x}(\theta, \delta_1)}{2v_y}. \end{aligned}$$

□

Ce résultat est très intéressant, puisqu'il nous permet de faire des parallèles entre l'estimation par densité prédictive et l'estimation ponctuelle. De (a), nous pouvons conclure qu'un estimateur de Bayes découlant d'une loi *a priori*  $\pi(\theta)$  domine l'estimateur de Bayes découlant de la loi *a priori* uniforme sur les réels  $\hat{p}_\pi \sim N_p(x, (v_x + v_y)I_p)$  si l'estimateur de Bayes sous cette loi  $\pi$  pour le problème d'estimation ponctuelle de  $\theta$  avec perte quadratique et variance connue  $v$  domine l'estimateur  $\delta(X) = X$  pour tout  $v_w \leq v \leq v_x$ .

De plus, on peut conclure à nouveau de (a) que l'estimateur  $\hat{p}_U$  est inadmissible pour  $p \geq 3$ , car l'estimateur  $X$  de  $\theta$  est inadmissible sous la perte quadratique pour  $p \geq 3$ , comme discuté dans l'article de Komaki ([11]).

De (b), on peut considérer le cas particulier où on compare l'estimateur par substitution  $\hat{p}_1 \sim N_p(x, v_y I_p)$  avec l'estimateur  $\hat{p}_U$ . On a :

$$R_{KL}(\theta, \hat{p}_1) - R_{KL}(\theta, \hat{p}_U) = \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (pv_x - pv) dv = \frac{p}{2} \left( \frac{v_x}{v_w} - \log\left(\frac{v_x}{v_w}\right) - 1 \right),$$

comme démontré dans le Corollaire 3.1.4.

Finalement, on peut combiner les deux résultats pour comparer un estimateur de Bayes avec un estimateur par substitution, deux estimateurs de Bayes ou deux estimateurs par substitution. En particulier, on peut affirmer que, parmi les estimateurs par substitution, celui qui substitue le paramètre  $\theta$  par l'estimateur de James-Stein ou tout autre estimateur minimax de la forme  $X + g(X)$  du chapitre 2 domine celui qui substitue  $\theta$  par l'estimateur  $X$  pour  $p \geq 3$ .

**Corollaire 3.2.2.** *Dans le cadre du problème (\*\*), un estimateur de Bayes est minimax si la loi a priori  $\pi(\theta)$  est superharmonique.*

*Démonstration.* En combinant les résultats du Théorème 2.4.6, de la Proposition 2.4.7, du Corollaire 2.4.8 et du Théorème 3.2.1, on obtient le résultat.  $\square$

**Exemple 3.2.3.** *Dans le cadre du problème (\*\*) avec  $p \geq 3$  et pour la loi a priori  $\pi(\theta) = \frac{1}{\|\theta\|^{2k}}$  où  $0 \leq k \leq \frac{p}{2} - 1$ , l'estimateur de Bayes associé  $\hat{p}_\pi$  est minimax. On a démontré la superharmonicité de cette loi a priori dans l'Exemple 2.4.9.*

D'autres applications de ces résultats sont traitées dans la prochaine section.

### 3.3 Estimation avec contrainte

Supposons que la seule information que nous possédions à propos du paramètre de position  $\theta$  soit qu'il appartient à un sous-ensemble non vide  $A$  de  $\mathbb{R}^p$ . Existe-t-il un estimateur dominant  $\hat{p}_U$  pour  $\theta \in A$ ? Regardons quel type d'estimateurs de Bayes nous obtenons dans certains cas particuliers.

**Exemple 3.3.1.** *Pour le problème (\*\*), considérons le cas unidimensionnel et la loi a priori  $\pi(\theta) = p\mathbb{1}_{\{-m\} \cup \{m\}}(\theta) + (1 - 2p)\mathbb{1}_{\{0\}}(\theta)$  pour  $m > 0$  et  $p \leq \frac{1}{2}$ . L'estimateur de Bayes pour ce problème est alors un mélange de lois normales prenant les densités  $\phi_{-m, v_y}$ ,*

*$\phi_{m, v_y}$  et  $\phi_{0, v_y}$  avec probabilités respectives  $\frac{1}{1 + e^{\frac{mx}{v_x}} (e^{\frac{mx}{v_x}} + \frac{1-2p}{p} e^{\frac{m^2}{2v_x}})}$ ,  $\frac{1}{1 + e^{-\frac{mx}{v_x}} (e^{-\frac{mx}{v_x}} + \frac{1-2p}{p} e^{\frac{m^2}{2v_x}})}$  et  $\frac{1}{1 + \frac{2p}{1-2p} e^{-\frac{m^2}{2v_x}} \cosh(\frac{mx}{v_x})}$ , c'est-à-dire*

$$\hat{p}_\pi(y|x) = \frac{p\phi_{-m, v_x}(x)}{p\phi_{-m, v_x}(x) + p\phi_{m, v_x}(x) + (1-2p)\phi_{0, v_x}(x)} \phi_{-m, v_y}(y) + \frac{p\phi_{m, v_x}(x)}{p\phi_{-m, v_x}(x) + p\phi_{m, v_x}(x) + (1-2p)\phi_{0, v_x}(x)} \phi_{m, v_y}(y) + \frac{(1-2p)\phi_{0, v_x}(x)}{p\phi_{-m, v_x}(x) + p\phi_{m, v_x}(x) + (1-2p)\phi_{0, v_x}(x)} \phi_{0, v_y}(y).$$

Pour  $p = 0,4$ ,  $m = 5$  et  $x = 2,5$ , on obtient l'estimateur représenté à la figure 3.1.

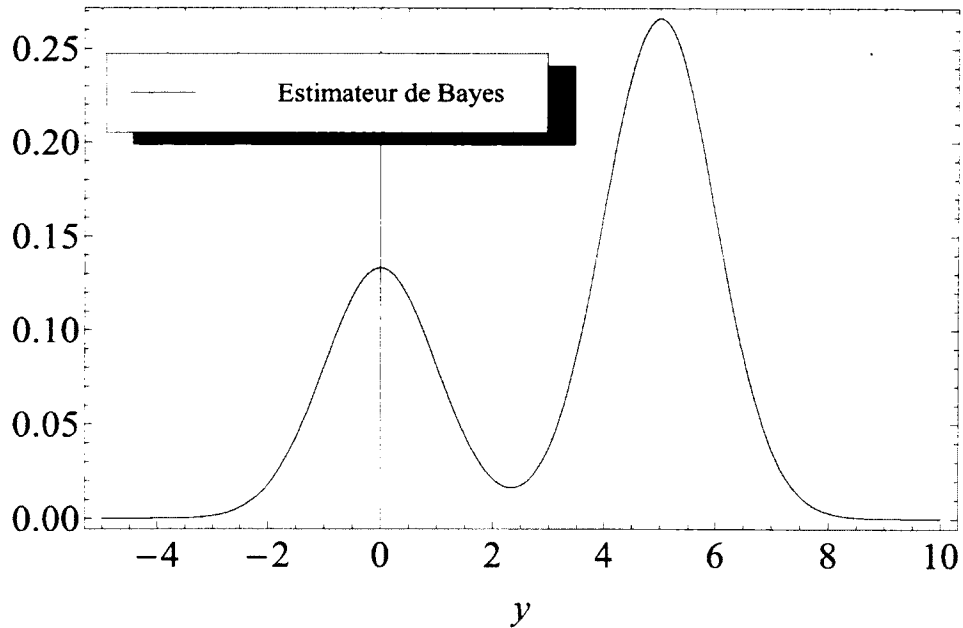


Figure 3.1 – Estimateur de Bayes pour la loi *a priori*  $\pi(\theta) = 0,4\mathbb{1}_{\{-5\} \cup \{5\}}(\theta) + 0,2\mathbb{1}_{\{0\}}(\theta)$  et pour la loi normale ( $v_x = v_y = 1$ ) avec observation  $x = 2, 5$ .

La densité prédictive est influencée par l’incertitude sur la valeur de  $\theta$ , mais ajuste les poids en considérant la probabilité  $p$ .

Casella et Strawderman ([6]) ont démontré, pour le problème d’estimation ponctuelle d’une moyenne  $\theta$  d’une loi normale avec variance connue sous la perte quadratique que la densité *a priori* concentrée aux bornes  $m$  et  $-m$  se trouve à être la loi la moins favorable (au niveau du risque minimal) pour un  $m$  assez petit ( $m \leq c_0\sigma$  avec  $c_0 \approx 1,059$ ) si on sait que  $-m \leq \theta \leq m$  et que l’estimateur de Bayes découlant de cette loi est donc minimax. Voir Marchand et Strawderman [18] pour une revue historique de ce type de problème. Ainsi, par le Théorème 3.2.1, on peut remarquer une similitude entre le résultat

de l'estimation ponctuelle sous la perte quadratique et l'estimation par densité prédictive sous la perte Kullback-Leibler représentée dans cet exemple.

**Exemple 3.3.2.** *Toujours dans le cas unidimensionnel, considérons la loi a priori uniforme sur  $[-m, m]$  de densité  $\pi(\theta) = \frac{1}{2m} \mathbb{1}_{[-m, m]}(\theta)$  pour  $m > 0$ . Considérant le résultat du Lemme 3.1.6 et le fait que  $X|\theta \sim N_p(\theta, v_x I_p)$  et  $W|\theta = \frac{v_y X + v_x Y}{v_x + v_y} \sim N_p(\theta, v_w)$  où  $v_w = \frac{v_x v_y}{v_x + v_y}$ , on a alors pour estimateur de Bayes :*

$$\hat{p}(y|x) = \frac{m_{\pi, W}(w)}{m_{\pi, X}(x)} \hat{p}_U(y|x) = \frac{\Phi(\frac{m-w}{\sqrt{v_w}}) - \Phi(\frac{-m-w}{\sqrt{v_w}})}{\Phi(\frac{m-x}{\sqrt{v_x}}) - \Phi(\frac{-m-x}{\sqrt{v_x}})} \varphi\left(\frac{y-x}{\sqrt{v_x + v_y}}\right)$$

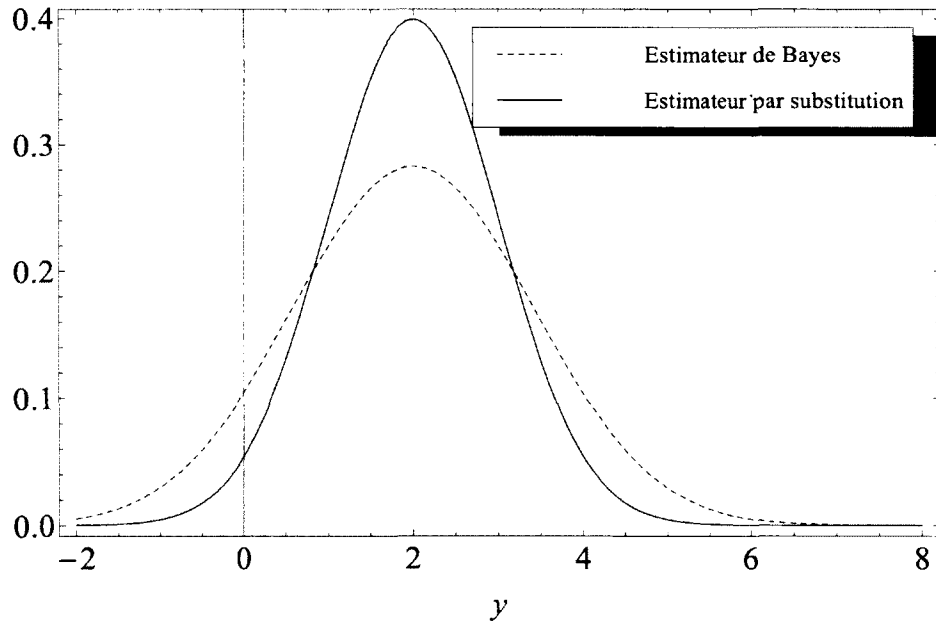


Figure 3.2 – Estimateur de Bayes pour la loi *a priori* uniforme sur  $[-5, 5]$  pour la loi normale ( $v_x = v_y = 1$ ) avec observation  $x = 2$

En comparant l'estimateur par substitution et l'estimateur de Bayes, on voit clairement que ce dernier est influencé par l'incertitude sur la valeur de  $\theta$ . Malgré la contrainte, la variance présente dans la densité de Bayes est supérieure à celle de l'estimateur par substitution.

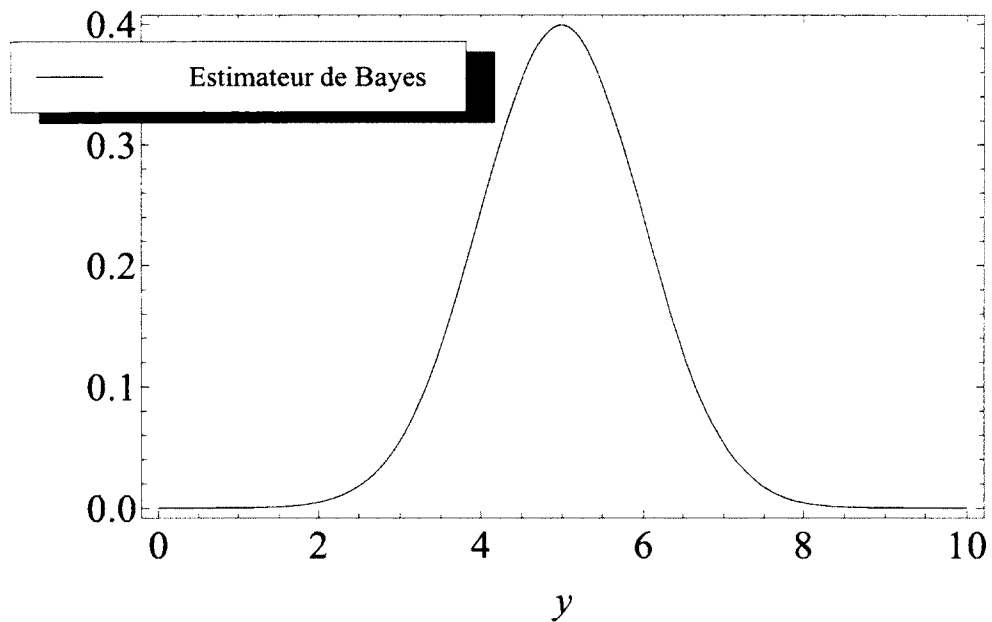


Figure 3.3 – Estimateur de Bayes pour la loi *a priori* uniforme sur  $[-5, 5]$  pour la loi normale ( $v_x = v_y = 1$ ) avec observation  $x = 100$

Dans ce graphe, malgré qu'on ait observé  $x = 100$ , la densité prédictive n'est pas centrée en cette valeur, considérant que  $-5 \leq \theta \leq 5$ . Ainsi, l'estimateur de Bayes pour la loi *a priori* uniforme sur l'intervalle  $[-5, 5]$  a une moyenne bornée par 5, comme on peut le voir dans la figure ci-haut.



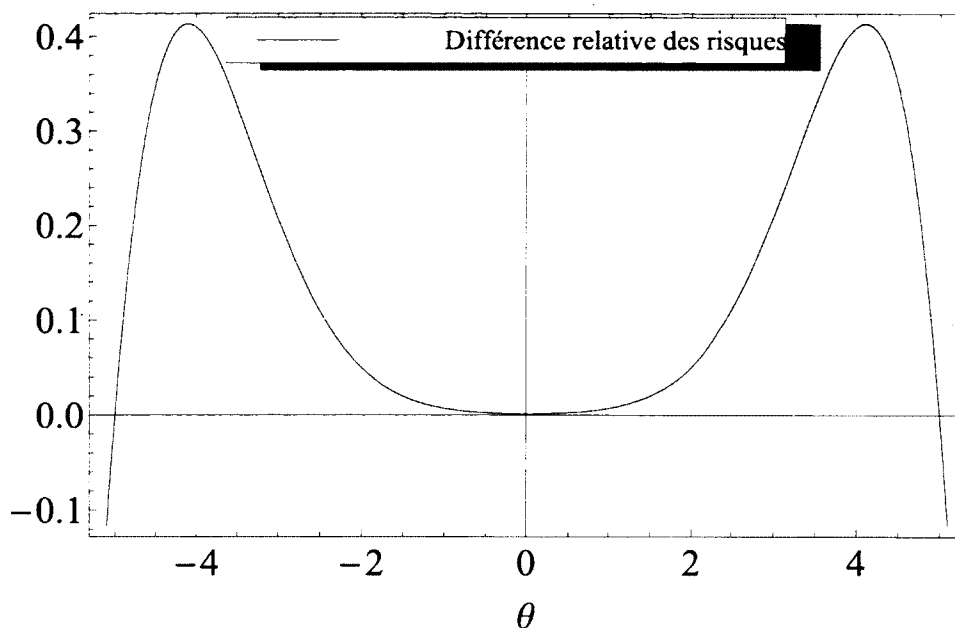


Figure 3.4 – Différence des risques (lois *a priori* uniforme sur  $[-5, 5]$  et sur les réels avec  $v_x = v_y = 1$ )

Le graphe de la figure 3.4, comme le démontre le prochain théorème, illustre le fait que l'estimateur de Bayes associé à la loi *a priori* uniforme sur l'intervalle  $[-5, 5]$  domine l'estimateur  $\hat{p}_U$  pour  $-5 \leq \theta \leq 5$ . En effet, la différence des risques de ces deux estimateurs est toujours positive dans cet intervalle.

**Théorème 3.3.3.** *Dans le cadre du problème (\*\*), l'estimateur de Bayes  $\hat{p}_{\pi_U}$  découlant de la loi a priori uniforme sur un intervalle  $[-m, m]$  domine l'estimateur  $\hat{p}_U$  pour  $\theta \in [-m, m]$ .*

*Démonstration.* Par le Lemme 3.1.6, on a :

$$\begin{aligned}
\Delta(\theta) &= R(\theta, \hat{p}_U) - R(\theta, \hat{p}_{\pi_U}) \\
&= E^W[\log(\Phi(\frac{m-W}{\sqrt{v_w}}) - \Phi(\frac{-m-W}{\sqrt{v_w}}))] - E^X[\log(\Phi(\frac{m-X}{\sqrt{v_x}}) - \Phi(\frac{-m-X}{\sqrt{v_x}}))] \\
&= E^Z[\log(\Phi(\frac{m-\theta}{\sqrt{v_w}} - Z) - \Phi(\frac{-m-\theta}{\sqrt{v_w}} - Z)) - \log(\Phi(\frac{m-\theta}{\sqrt{v_x}} - Z) - \Phi(\frac{-m-\theta}{\sqrt{v_x}} - Z))],
\end{aligned}$$

où on a effectué les changements de variables  $w = \sqrt{v_w}z + \theta$  et  $x = \sqrt{v_x}z + \theta$  pour prendre l'espérance sur une normale centrée et réduite. Or, comme  $v_x > v_w$ , alors on a  $\frac{m-\theta}{\sqrt{v_x}} - Z < \frac{m-\theta}{\sqrt{v_w}} - Z$  et  $\frac{-m-\theta}{\sqrt{v_x}} - Z > \frac{-m-\theta}{\sqrt{v_w}} - Z$  pour tout  $\theta \in [-m, m]$ . La différence des risques est alors positive, car les fonctions  $\log$  et  $\Phi$  sont strictement croissantes.  $\square$

**Exemple 3.3.4.** *Toujours dans le cas unidimensionnel, considérons la loi a priori impropre  $\pi(\theta) = \mathbb{1}_{\mathbb{R}^+}(\theta)$ . Par le Lemme 3.1.6, on a alors pour estimateur de Bayes :*

$$\hat{p}(y|x) = \frac{\Phi(\frac{w}{\sqrt{v_w}})}{\Phi(\frac{x}{\sqrt{v_x}})} \phi\left(\frac{y-x}{\sqrt{v_x+v_y}}\right)$$

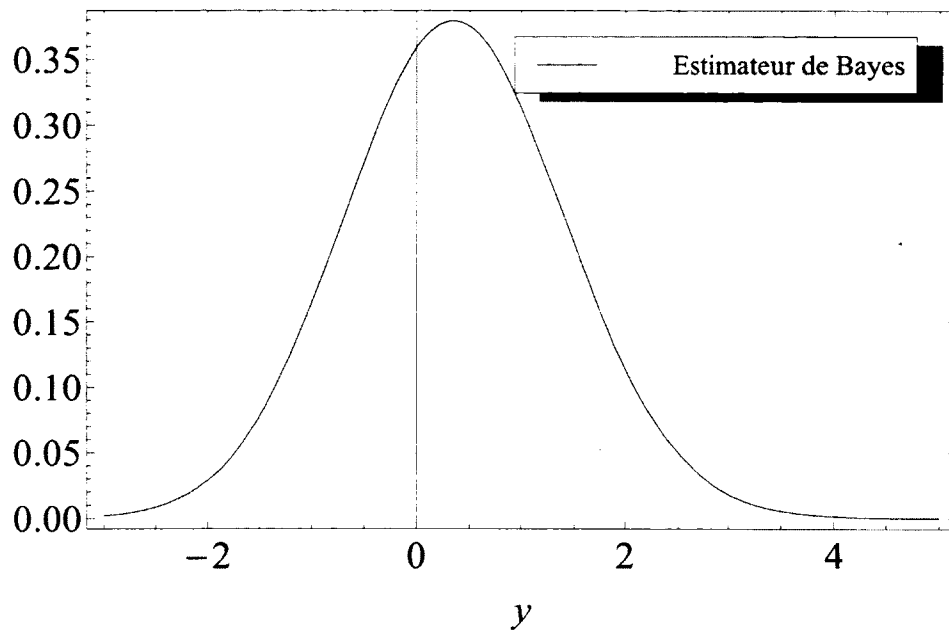


Figure 3.5 – Estimateur de Bayes pour la loi *a priori* uniforme sur les réels positifs pour la loi normale ( $v_x = v_y = 1$ ) avec observation  $x = -2$

Dans la figure 3.5, on remarque que malgré qu'on eut observé une valeur  $x$  négative, la densité prédictive s'en tient aux densités à moyenne positive considérant le domaine d'appartenance du paramètre  $\theta$ . La moyenne de cette densité prédictive est alors bornée par 0, peu importe l'observation.

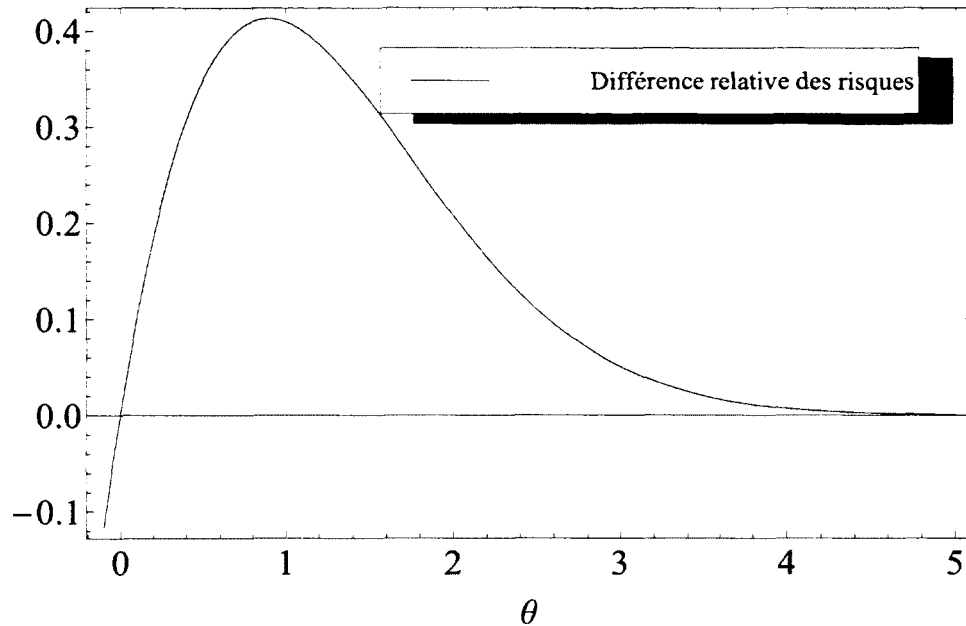


Figure 3.6 – Différence relative des risques (loi *a priori* uniforme sur les réels positifs et sur les réels avec  $v_x = v_y = 1$ )

Une fois de plus, on remarque dans la figure 3.6 la dominance de l'estimateur de Bayes pour cette loi *a priori* sur l'estimateur  $\hat{p}_U$  pour  $\theta \geq 0$ . La figure illustre que la différence des risques entre ces deux estimateurs est positive sur les réels positifs. Démontrons ce résultat.

**Théoreme 3.3.5.** *Dans le cadre du problème (\*\*), l'estimateur de Bayes  $\hat{p}_{\pi_U}$  découlant de la loi a priori uniforme sur les réels positifs domine l'estimateur  $\hat{p}_U$  pour  $\theta \in \mathbb{R}^+$ .*

*Démonstration.* On a :

$$\begin{aligned}\Delta(\theta) &= E^W[\log(\Phi(\frac{W}{\sqrt{v_w}}))] - E^X[\log(\Phi(\frac{X}{\sqrt{v_x}}))] \\ &= E^Z[\log(\Phi(Z + \frac{\theta}{\sqrt{v_w}})) - \log(\Phi(Z + \frac{\theta}{\sqrt{v_x}}))],\end{aligned}$$

où le même changement de variable que dans le Théorème 3.1.6 a été effectué. Comme  $v_x > v_w$ , alors  $Z + \frac{\theta}{\sqrt{v_x}} \leq Z + \frac{\theta}{\sqrt{v_w}}$  pour tout  $\theta \geq 0$ . La différence des risques est alors positive, car les fonctions  $\log$  et  $\Phi$  sont croissantes.  $\square$

On peut généraliser ce résultat de dominance aux ensembles convexes avec intérieur non-vide et au cas multidimensionnel.

**Théorème 3.3.6. [7]** *Soit  $C \subset \mathbb{R}^p$  un convexe avec intérieur non vide et  $\pi_C(\theta) = \mathbf{1}_C(\theta)$ . Alors  $\hat{p}_{\pi_C}$  domine  $\hat{p}_U$  pour  $\theta \in C$  et pour le risque Kulback-Leibler.*

*Démonstration.* La différence des risques correspond à :

$$\begin{aligned}\Delta(\theta) &= E^W[\log(m_{\pi_C(W)})] - E^X[\log(m_{\pi_C(X)})] \\ &= \int_{\mathbb{R}^p} p(w|\theta) \log(\int_C p(w|\theta_1) d\theta_1) dw - \int_{\mathbb{R}^p} p(x|\theta) \log(\int_C p(x|\theta_1) d\theta_1) dx \\ &= \int_{\mathbb{R}^p} \phi(z) \log(\frac{\int_{C_1} \phi(s) ds}{\int_{C_2} \phi(s) ds}) dz,\end{aligned}$$

où la dernière égalité est due aux changements de variables  $z = \frac{w-\theta}{\sqrt{v_w}}$ ,  $s = \frac{\theta_1-w}{\sqrt{v_w}}$ ,  $z = \frac{x-\theta}{\sqrt{v_x}}$  et  $s = \frac{\theta_1-x}{\sqrt{v_x}}$  et où  $C_1 = \{\frac{1}{\sqrt{v_w}}(s - \theta) - z : s \in C\}$  et  $C_2 = \{\frac{1}{\sqrt{v_x}}(s - \theta) - z : s \in C\}$ . Ainsi, si  $C_2 \subset C_1$ , l'intégrande est positive pour tout  $z \in \mathbb{R}^p$ . Il suffit donc de montrer que  $C_2 \subset C_1$ .

Soit  $c_1 \in \frac{1}{\sqrt{v_x}}(C - \theta)$ . Alors, il existe un  $c \in C$  tel que  $c_1 = \frac{1}{\sqrt{v_x}}(c - \theta)$ . On peut en déduire que  $c_1 = \frac{1}{\sqrt{v_w}}(c' - \theta)$  où  $c' = \frac{\sqrt{v_w}}{\sqrt{v_x}}c + (1 - \frac{\sqrt{v_w}}{\sqrt{v_x}})\theta$ . Comme  $v_w < v_x$ ,  $c'$  est une combinaison convexe de  $c$  et  $\theta$  qui sont tous deux éléments de  $C$ . Ainsi,  $c_1 \in \frac{1}{\sqrt{v_w}}(C - \theta)$ . Donc,  $C_2 \subset C_1$  et  $\Delta(\theta) > 0$ .  $\square$

Voici une autre démonstration du Théorème 3.3.6 utilisant le Théorème 3.2.1 et un résultat de Hartigan ([9]).

*Démonstration.* Nous savons, par le Théorème 3.2.1, que pour le problème d'estimation par densité prédictive et pour la perte Kullback-Leibler, la différence des risques entre l'estimateur  $\hat{p}_U$  et  $\hat{p}_{\pi_C}$  peut s'exprimer comme suit :

$$\frac{1}{2} \int_{v_u}^{v_x} \frac{1}{v^2} (R_Q^v(\theta, X) - R_Q^v(\theta, \hat{\theta}_{\pi_C, v})) dv.$$

Pour le problème d'estimation ponctuelle de  $\theta$  où  $X \sim N_p(\theta, vI_p)$  sous la perte quadratique, Hartigan ([9]) a montré que pour  $\theta \in C$  convexe avec intérieur non vide, l'estimateur de Bayes  $\hat{\theta}_{\pi_C, v}$  découlant de la loi *a priori* uniforme sur  $C$  domine  $X$  pour tout  $v > 0$ .

Or, pour tout  $v$  et pour tout  $\theta \in C$ ,  $R_Q^v(\theta, X) - R_Q^v(\theta, \hat{\theta}_{\pi_C, v}) \geq 0$  avec inégalité stricte pour au moins un  $\theta$ . Ainsi, en combinant ces résultats, on peut en déduire que  $\hat{p}_{\pi_C}$  domine  $\hat{p}_U$  pour  $\theta \in C$ .  $\square$

Le résultat étant démontré pour le problème d'estimation par densité prédictive sous la perte Kullback-Leibler, on peut déduire un résultat similaire pour l'estimation ponctuelle sous la perte quadratique donnant une autre démonstration du résultat de Hartigan.

**Corollaire 3.3.7. ([7])** *Pour le problème d'estimation ponctuelle de  $\theta$  où  $X \sim N_p(\theta, v_x I_p)$  sous la perte quadratique, l'estimateur de Bayes découlant d'une loi *a priori* uniforme sur un ensemble convexe  $C$  avec intérieur non vide domine  $X$ .*

*Démonstration.* On a montré que la différence des risques dans le contexte du Théorème 3.3.6 était positif. Or, à partir du fait que

$$\Delta(\theta) = \frac{1}{2} \int_{v_u}^{v_x} \frac{1}{v^2} (R_Q^v(\theta, X) - R_Q^v(\theta, \hat{\theta}_{\pi_C, v})) dv \geq 0$$

pour  $\theta \in C$  et pour tout  $v_x, v_w$  positifs, on peut donc en déduire que  $R_Q^v(\theta, X) - R_Q^v(\theta, \hat{\theta}_{\pi_C, v})$  est positif presque partout.  $\square$

En résumé, nous avons trouvé, dans le cadre de la loi normale, un estimateur  $\hat{p}_U$  possédant des propriétés de minimaxité et de dominance. Un lien entre l'estimation ponctuelle et l'estimation par densité prédictive a été établi, ce qui nous a conduit à de meilleurs estimateurs, notamment lorsque le paramètre  $\theta$  subit une contrainte. Ce mémoire ajoute alors une meilleure compréhension du problème lorsque l'espace paramétrique est restreint.

# CHAPITRE 4

## Estimation pour des familles à paramètre de position et d'échelle sous contrainte

Une grande partie de ce chapitre est adaptée de l'article de Kubokawa, Marchand, Strawderman et Turcotte [14]. Cette contribution de recherche, à paraître dans le *Journal of Multivariate Analysis* est reproduite au chapitre 5. Nous élaborerons sur le problème d'estimation par densité prédictives pour des familles de lois à paramètres de position et d'échelle où le paramètre doit respecter une contrainte. Quelques illustrations seront présentées.

### 4.1 Cas d'un paramètre de position sous contrainte

Dans ce chapitre, nous tenterons de trouver des estimateurs performants pour des familles à paramètre de position  $\theta$ . En particulier, nous aborderons le problème où  $\theta$  est restreint



à un sous-ensemble des réels. Nous nous concentrerons sur les conditions nécessaires à la famille de densités de la variable aléatoire, conditions qui étaient respectées au chapitre 3 dans le cas de la loi normale.

Soient deux variables aléatoires indépendantes  $X \sim p_\theta \in P = \{p_\theta(x - \theta) : \theta \in \mathbb{R}\}$  et  $Y \sim q_\theta \in Q = \{q_\theta(y - \theta) : \theta \in \mathbb{R}\}$  où  $p_0(\cdot)$  et  $q_0(\cdot)$  sont des fonctions de densités. Si nous sommes en présence d'un vecteur aléatoire  $X = (X_1, X_2, \dots, X_n)$ , un développement analogue est possible en conditionnant sur le maximal invariant  $(X_1 - X - 2, \dots, X_1 - X_n)$  (voir chapitre 5).

Supposons de plus que  $\theta$  est restreint à l'intervalle  $[b, \infty[$  et que la condition suivante est satisfaite :

(\*)  $q_0(y - \theta)$  est continuellement différentiable et  $\frac{q_0(y-\theta)}{q_0(y)}$  est non décroissante en  $y$  pour  $\theta \geq 0$ .

Cette condition est équivalente à dire que la famille de densités  $q_\theta$  a un rapport de vraisemblance monotone croissant en  $y$ .

Nous voulons démontrer la dominance (et par conséquent la minimaxité) de l'estimateur de Bayes  $\hat{p}_U$  découlant de la loi *a priori* uniforme sur les réels positifs sur l'estimateur de Bayes découlant de la loi *a priori* uniforme sur les réels  $\hat{p}_0$ . Ce dernier est le meilleur estimateur invariant par rapport à des changements de position et aussi minimax (Barron et Liang [2]). Pour ce faire, nous avons besoin du lemme suivant.

**Lemme 4.1.1. [14]** *Sous la condition (\*), on a*

(i)  $q_0(y)$  est logconcave en  $y$ .

(ii) Pour  $\theta \geq 0$ , la fonction  $A_\theta(y|x) = \frac{\int_{-\infty}^0 p_0(x+w-\theta)q_0(y+w-\theta)dw}{\int_{-\infty}^0 p_0(x+w)q_0(y+w)dw}$  est non décroissante en  $y$ .

*Démonstration.* (i) La condition (\*) nous assure que

$$\begin{aligned}
& \frac{\partial}{\partial y} \left( \frac{q_0(y-\theta)}{q_0(y)} \right) \geq 0 \\
\Rightarrow q'_0(y-\theta)q_0(y) - q_0(y-\theta)q'_0(y) & \geq 0 \\
\Rightarrow \frac{q'_0(y-\theta)}{q_0(y-\theta)} & \geq \frac{q'_0(y)}{q_0(y)}.
\end{aligned}$$

(ii) Nous devons montrer que  $\frac{\partial}{\partial y} A_\theta(y|x) \geq 0$ . Or, cette inégalité est vraie si et seulement si

$$\begin{aligned}
& \int_{-\infty}^0 p_0(x+w-\theta)q'_0(y+w-\theta)dw \int_{-\infty}^0 p_0(x+w)q_0(y+w)dw \\
& \geq \int_{-\infty}^0 p_0(x+w-\theta)q_0(y+w-\theta)dw \int_{-\infty}^0 p_0(x+w)q'_0(y+w)dw.
\end{aligned}$$

En effectuant le changement de variable  $w \rightarrow w - \theta$ , on obtient :

$$\frac{\int_{-\infty}^{-\theta} p_0(x+w)q'_0(y+w)dw}{\int_{-\infty}^{-\theta} p_0(x+w)q_0(y+w)dw} \geq \frac{\int_{-\infty}^0 p_0(x+w)q'_0(y+w)dw}{\int_{-\infty}^0 p_0(x+w)q_0(y+w)dw}.$$

En posant la partie de gauche comme la fonction  $f(\theta)$ , il suffit de montrer que  $\frac{d}{d\theta} f(\theta) \geq 0$ . Il faut donc que l'expression suivante soit positive :

$$-p_0(x-\theta)q'_0(y-\theta) \int_{-\infty}^{-\theta} p_0(x+w)q_0(y+w)dw + p_0(x-\theta)q_0(y-\theta) \int_{-\infty}^{-\theta} p_0(x+w)q'_0(y+w)dw.$$

Cette expression se ramène sous la forme :

$$p_0(x-\theta)q_0(y-\theta) \int_{-\infty}^{-\theta} p_0(x+w)q_0(y+w) \left( \frac{q'_0(y+w)}{q_0(y+w)} - \frac{q'_0(y-\theta)}{q_0(y-\theta)} \right) dw.$$

Or, (i) nous assure que la quantité intégrée est positive ou nulle pour tout  $w \in [-\infty, -\theta]$ , ce qui conclut la démonstration.  $\square$

Passons maintenant au résultat principal de cette section concernant les estimateurs de Bayes lorsque l'espace paramétrique est restreint. Nous adaptons la méthode IERD (Integral expression for risk difference) de Kubokawa ([12],[13]) pour en faire une démonstration élégante.

**Théorème 4.1.2. [14]** *Supposons la condition (\*) respectée et  $\pi(\theta) = \mathbf{1}_{[b,\infty[}(\theta)$ . Alors, l'estimateur de Bayes  $\hat{p}_\pi(\cdot|x)$  est minimax pour  $\theta \geq b$ . De plus,  $\hat{p}_\pi(\cdot|x)$  et  $\hat{p}_U(\cdot|x)$  ont des risques égaux si et seulement si  $\theta = b$ .*

Pour faire la démonstration de ce théorème, nous avons besoin d'un résultat connu sur la covariance.

**Lemme 4.1.3.** *Soient  $X$  une variable aléatoire,  $g(\cdot)$  et  $h(\cdot)$  deux fonctions croissantes telles que  $E[g(X)h(X)]$ ,  $E[g(X)]$  et  $E[h(X)]$  existent. Alors,  $Cov(g(X), h(X)) \geq 0$  avec égalité stricte ssi  $g(\cdot)$  où  $h(\cdot)$  est constante avec probabilité un.*

*Démonstration.* Soient  $X_1$  et  $X_2$  deux copies indépendantes de  $X$ . Alors,

$$\begin{aligned} & E[(h(X_1) - h(X_2))(g(X_1) - g(X_2))] \\ &= E[h(X_1)g(X_1)] + E[h(X_2)g(X_2)] - E[h(X_1)g(X_2)] - E[h(X_2)g(X_1)] \\ &= 2E[h(X)g(X)] - 2E[h(X)]E[g(X)] \\ &= 2Cov(h(X), g(X)). \end{aligned}$$

Comme la variable aléatoire  $Y = (h(X_1) - h(X_2))(g(X_1) - g(X_2))$  est toujours positive ou nulle, alors  $E[Y] = Cov(h(X), g(X)) \geq 0$ . En particulier, on a que  $E[Y] = 0$  ssi  $g(\cdot)$  ou  $h(\cdot)$  est constante avec probabilité 1.  $\square$

Revenons au Théorème 4.1.2 à démontrer :

*Démonstration.* Sans perte de généralité, nous posons  $b = 0$ , car pour  $b \neq 0$ , il suffit de poser  $\theta' = \theta - b$ . Puisque  $\hat{p}_U(\cdot|x)$  est minimax, il suffit de montrer que  $\hat{p}_\pi(\cdot|x)$  domine  $\hat{p}_U(\cdot|x)$  pour  $\theta \geq 0$ . On pose

$$\begin{aligned}\Delta(\theta) &= R_{KL}(\theta, \widehat{p}_U(\cdot|x)) - R_{KL}(\theta, \widehat{p}_\pi(\cdot|x)) \\ &= \int \int p_0(x - \theta) q_0(y - \theta) (\log(\widehat{p}_\pi(y|x)) - \log(\widehat{p}_U(y|x))) dx dy.\end{aligned}$$

Or, on a :

$$\begin{aligned}\log(\widehat{p}_\pi(y|x)) - \log(\widehat{p}_U(y|x)) &= \log\left(\frac{\int_0^\infty q_0(y-a)p_0(x-a)da}{\int_0^\infty p_0(x-a)da}\right) - \log\left(\frac{\int_{-\infty}^\infty q_0(y-a)p_0(x-a)da}{\int_{-\infty}^\infty p_0(x-a)da}\right) \\ &= \int_{-\infty}^0 \frac{d}{dt} \left( \log\left(\frac{\int_t^\infty q_0(y-a)p_0(x-a)da}{\int_t^\infty p_0(x-a)da}\right) \right) dt \\ &= \int_{-\infty}^0 \left( \frac{p_0(x-t)}{\int_t^\infty p_0(x-a)da} - \frac{p_0(x-t)q_0(y-t)}{\int_t^\infty p_0(x-a)q_0(y-a)da} \right) dt.\end{aligned}$$

En posant  $x = x - t$  et  $y = y - t$ , on obtient :

$$\begin{aligned}\Delta(\theta) &= \int \int \int_{-\infty}^0 p_0(x+t-\theta) q_0(y+t-\theta) \left( \frac{p_0(x)}{\int_{-\infty}^0 p_0(x+w)dw} - \frac{p_0(x)q_0(y)}{\int_{-\infty}^0 p_0(x+w)q_0(y+w)dw} \right) dt dx dy \\ &= \int \int p_0(x) q_0(y) \frac{\int_{-\infty}^0 p_0(x+w-\theta) q_0(y+w-\theta) dw}{\int_{-\infty}^0 p_0(x+w) q_0(y+w) dw} \left( \frac{\int_{-\infty}^0 p_0(x+w) q_0(y+w) dw}{\int_{-\infty}^0 p_0(x+w) dw q_0(y)} - 1 \right) dy dx \\ &= E^X [E^Y [A_\theta(Y|X) \left( \frac{\int_{-\infty}^0 p_0(X+w) q_0(Y+w) dw}{\int_{-\infty}^0 p_0(X+w) dw q_0(Y)} - 1 \right) ]].\end{aligned}$$

où  $A_\theta(y|x)$  est définie au Lemme 4.1.1. Or, comme  $A_\theta(y|x)$  et  $\frac{q_0(y+w)}{q_0(y)}$  sont non décroissantes en  $y$ , l'inégalité de covariance stipule que :

$$\begin{aligned}\text{Cov}(A_\theta(Y|x), \frac{\int_{-\infty}^0 p_0(x+w) q_0(Y+w) dw}{\int_{-\infty}^0 p_0(x+w) dw q_0(Y)} - 1) &\geq 0 \\ \Rightarrow E^Y [A_\theta(Y|x) \left( \frac{\int_{-\infty}^0 p_0(x+w) q_0(Y+w) dw}{\int_{-\infty}^0 p_0(x+w) dw q_0(Y)} - 1 \right)] &\geq E^Y [A_\theta(Y|x)] E^Y \left[ \frac{\int_{-\infty}^0 p_0(x+w) q_0(Y+w) dw}{\int_{-\infty}^0 p_0(x+w) dw q_0(Y)} - 1 \right].\end{aligned}$$

Ainsi, la différence des risques est bornée comme suit :

$$\begin{aligned}\Delta(\theta) &\geq E^X [E^Y [A_\theta(y|x)] E^Y \left[ \frac{\int_{-\infty}^0 p_0(X+w) q_0(Y+w) dw}{\int_{-\infty}^0 p_0(X+w) dw q_0(Y)} - 1 \right]] \\ &= 0,\end{aligned}$$

où la dernière égalité est due à une inversion des intégrales et à un changement de variable dans la deuxième espérance.

De plus, on peut remarquer que nous avons égalité ssi  $A_\theta(Y|X)$  ou  $\frac{\int_{-\infty}^0 p_0(X+w)q_0(Y+w)dw}{\int_{-\infty}^0 p_0(X+w)dwq_0(Y)} - 1$  est constante avec probabilité 1 (voir Lemme 4.1.3), ce qui se produit ssi  $\theta = 0$ , car, dans ce cas,  $A_0(y|x) = 1$ .  $\square$

**Remarque 4.1.4.** *On peut étendre ce résultat pour  $\theta \in ]-\infty, c]$ . En effet, les changements de variables  $Y = -X$  et  $\theta' = -\theta$  mènent, pour des estimateurs de Bayes équivalents, à des risques égaux. De plus, on a que  $\theta'$  est un paramètre de position pour la variable  $Y$ .*

Pour illustrer les résultats de cette section, voici un exemple respectant les conditions du contexte décrites au début de ce chapitre.

**Exemple 4.1.5.** *Considérons le cas où  $X, Y \sim \text{Exp}(\theta, \sigma)$  dont la fonction de densité s'exprime comme suit :*

$$f_{T|\theta}(t) = \frac{1}{\sigma} e^{-\frac{(t-\theta)}{\sigma}} \mathbf{1}_{(\theta, \infty)}(t).$$

L'estimateur de Bayes découlant de la loi a priori non informative  $\pi(\theta) = \mathbf{1}_{\mathbb{R}}(\theta)$  est :

$$\begin{aligned} \hat{p}_U(y|x) &= \frac{\int_{-\infty}^{\infty} \frac{1}{\sigma^2} e^{-\frac{1}{\sigma}(x-\theta+y-\theta)} \mathbf{1}_{(\theta, \infty)}(x) \mathbf{1}_{(\theta, \infty)}(y) d\theta}{\int_{-\infty}^{\infty} \frac{1}{\sigma} e^{-\frac{1}{\sigma}(x-\theta)} \mathbf{1}_{(\theta, \infty)}(x) d\theta} \\ &= \frac{\int_{-\infty}^{\min(x,y)} \frac{1}{\sigma} e^{-\frac{1}{\sigma}(x+y-2\theta)} d\theta}{\int_{-\infty}^x e^{-\frac{1}{\sigma}(x-\theta)} d\theta} \\ &= \frac{[\frac{1}{2} e^{-\frac{1}{\sigma}(x+y-2\theta)}]_{\theta \rightarrow -\infty}^{\min(x,y)}}{[\sigma e^{-\frac{1}{\sigma}(x-\theta)}]_{\theta \rightarrow -\infty}^x} \\ &= \frac{1}{2\sigma} e^{-\frac{1}{\sigma}(x+y-2\min(x,y))} \\ &= \frac{1}{2\sigma} e^{-\frac{1}{\sigma}|x-y|}, \end{aligned}$$

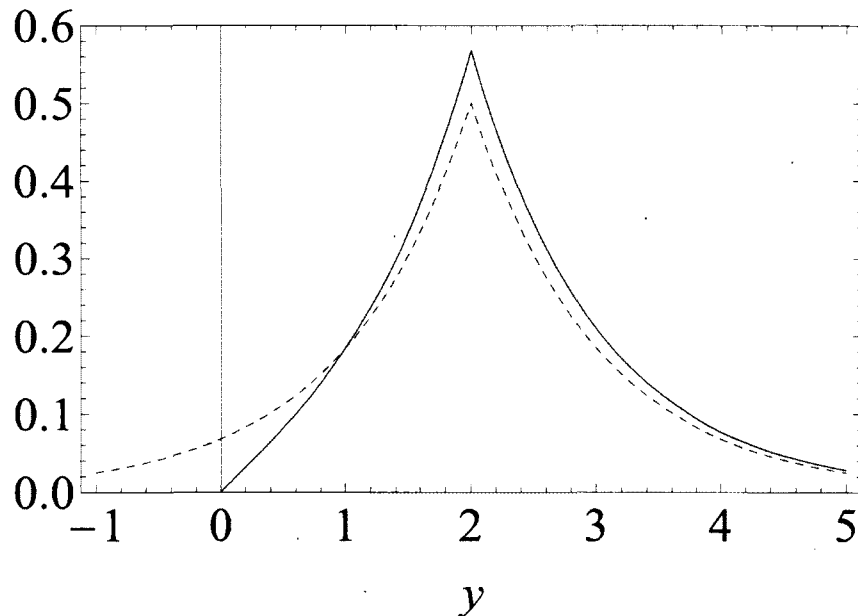
qui se trouve à être la densité d'une variable aléatoire de loi de Laplace avec paramètre d'échelle  $\sigma$ .

D'un autre côté, si on restreint le paramètre  $\theta$  aux réels positifs, l'estimateur de Bayes découlant de la loi a priori  $\pi(\theta) = \mathbf{1}_{[0, \infty[}(\theta)$  est :

$$\begin{aligned}
\widehat{p}_\pi(y|x) &= \frac{\int_0^\infty \frac{1}{2\sigma} e^{-\frac{1}{\sigma}(x-\theta+y-\theta)} \mathbb{1}_{(\theta, \infty)}(x) \mathbb{1}_{(\theta, \infty)}(y) d\theta}{\int_0^\infty \frac{1}{\sigma} e^{-\frac{1}{\sigma}(x-\theta)} \mathbb{1}_{(\theta, \infty)}(x) d\theta} \\
&= \frac{\left[ \frac{1}{2} e^{-\frac{1}{\sigma}(x+y-2\theta)} \right]_{\theta=0}^{\min(x,y)}}{\left[ \sigma e^{-\frac{1}{\sigma}(x-\theta)} \right]_{\theta=0}^x} \\
&= \frac{\frac{1}{2\sigma} e^{-\frac{1}{\sigma}(x+y-2\min(x,y))} - e^{-\frac{1}{\sigma}(x+y)}}{1 - e^{-\frac{x}{\sigma}}} \mathbb{1}_{(0, \infty)}(y) \\
&= \frac{1}{2\sigma} \frac{e^{-\frac{1}{\sigma}|x-y|} - e^{-\frac{1}{\sigma}(x+y)}}{1 - e^{-\frac{x}{\sigma}}} \mathbb{1}_{(0, \infty)}(y).
\end{aligned}$$

On peut aussi écrire cette densité sous la forme suivante.

$$\widehat{p}_\pi(y|x) = \begin{cases} \frac{1}{2\sigma} e^{-\frac{y}{\sigma}} (1 + e^{\frac{x}{\sigma}}) & \text{pour } y \geq x \\ \frac{1}{2\sigma} \frac{e^{\frac{y}{\sigma}} - e^{-\frac{y}{\sigma}}}{e^{\frac{x}{\sigma}} - 1} & \text{pour } 0 < y < x \end{cases}$$



-----	Bayes sous loi a priori uniforme sur réels
—————	Bayes sous loi a priori uniforme sur réels positifs

Figure 4.1 – Densités prédictives pour la loi exponentielle avec  $\sigma = 1$  et  $x = 2$ .

La figure 4.1 illustre les densités prédictives associées aux lois *a priori* uniforme sur les réels et uniforme sur les réels positifs. On remarque que lorsque le paramètre  $\theta$  est restreint aux réels positifs, la densité l'est aussi. De plus, cette densité présente des similitudes avec la densité d'une variable aléatoire de Laplace. La densité prédictive associée à la loi *a priori* uniforme sur les réels en est un exemple.

Pour  $\sigma = 1$ , la différence des risques de ces deux estimateurs est représentée par le graphique suivant :

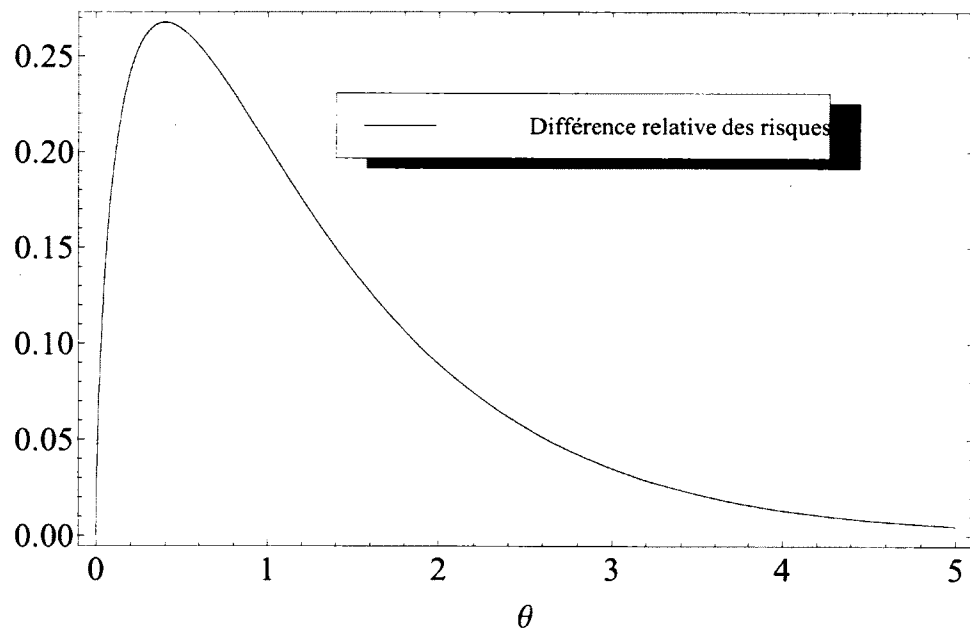


Figure 4.2 – Différence relative des risques pour la loi exponentielle avec  $\sigma = 1$  et  $x = 2$ .

On voit ici une application directe du Théorème 4.1.2. Pour  $\theta \in [0, \infty[$ , l'estimateur de Bayes découlant de la loi *a priori* uniforme sur les réels positifs domine l'estimateur de Bayes découlant de la loi *a priori* uniforme sur les réels.

Le Théorème 4.1.2 peut aussi s'appliquer au problème du Théorème 3.3.6 avec le convexe  $C = ]0, \infty[$ . Il s'agit là d'une troisième démonstration de ce théorème dans le cadre univarié, puisque la fonction de densité d'une loi normale respecte la condition (\*).

## 4.2 Cas d'un paramètre d'échelle sous contrainte

Le résultat de dominance pour un paramètre de position étant démontré, peut-on conclure de même pour des familles à paramètre d'échelle? Nous pouvons répondre positivement en constatant que les transformations  $X' = \log(X)$ ,  $Y' = \log(Y)$  et  $\theta' = \log(\theta)$  conduisent aux densités suivantes :  $p'_{\theta'}(x') = e^{x'-\theta'} p_0(e^{x'-\theta'})$  et  $q'_{\theta'}(y') = e^{y'-\theta'} q_0(e^{y'-\theta'})$ , où  $\theta'$  est un paramètre de position. Comme la perte Kullback-Leibler est définie de façon intrinsèque (voir Proposition 1.2.3), estimer la densité de  $Y|\theta$  est équivalent à estimer la densité de  $Y'|\theta'$ .

**Proposition 4.2.1.** *Considérons le problème d'estimation de la densité de  $Y|\theta$  à partir d'une observation de même loi que  $X|\theta$  sous la perte Kullback-Leibler avec  $\theta$  paramètre d'échelle et le problème similaire déduit des transformations énoncées plus haut. Pour une loi a priori  $\pi(\theta)$  sur  $\theta$ , l'estimateur de Bayes  $\hat{p}_\pi(\cdot|x')$  (où  $\pi'(\theta')$  représente la loi a priori de  $\theta'$ ) pour le deuxième problème est la densité de la variable aléatoire  $T' = \log(T)$ , où  $T$  a pour densité l'estimateur de Bayes  $\hat{p}_\pi(\cdot|x)$  pour le premier problème. De plus, les pertes respectives de ces deux estimateurs sont égales.*

*Démonstration.* Dans le chapitre 1, nous avons montré que les deux problèmes sont équivalents pour une transformation  $h(y)$  strictement monotone sur  $Y$ . Montrons le résultat pour ce cas particulier où  $h(y) = \log(y)$  en appliquant les mêmes transformations sur  $X$  et sur  $\theta$ .

Soient  $T \sim \hat{p}_\pi(\cdot|X)$  et  $T' = \log(T)$ . On a alors :



$$\begin{aligned}
p_{T'|X}(t') &= e^{t'} \widehat{p}_\pi(e^{t'}|X) \\
&= e^{t'} \frac{\int p_\theta(x) q_\theta(e^{t'}) \pi(\theta) d\theta}{\int p_\theta(x) \pi(\theta) d\theta} \\
&= \frac{\int e^{x'} p_\theta(e^{x'}) e^{t'} q_\theta(e^{t'}) \pi(e^{\theta'}) e^{\theta'} d\theta'}{\int e^{x'} p_\theta(e^{x'}) \pi(e^{\theta'}) e^{\theta'} d\theta'} \\
&= \frac{\int p_{\theta'}(x') q_{\theta'}(t') \pi'(\theta') d\theta'}{\int p_{\theta'}(x') \pi'(\theta') d\theta'};
\end{aligned}$$

qui est l'estimateur de Bayes du deuxième problème. La perte respective de ces deux estimateurs est la même, comme démontré à la Proposition 1.2.3.  $\square$

Ainsi, les résultats de la première section de ce chapitre s'appliquent aux familles à paramètre d'échelle avec des espaces paramétriques de la forme  $]0, c]$  ou  $[b, \infty[$ .

**Théorème 4.2.2.** *Soient deux variables aléatoires indépendantes  $X \sim p_\theta \in P_\theta = \{\frac{1}{\theta} p_0(\frac{x}{\theta}) : \theta \in \mathbb{R}^+\}$  et  $Y \sim q_\theta \in Q_\theta = \{\frac{1}{\theta} q_0(\frac{y}{\theta}) : \theta \in \mathbb{R}^+\}$ , où  $p_0(\cdot)$  et  $q_0(\cdot)$  sont des fonctions de densités. Supposons de plus que  $\pi(\theta) = \frac{1}{\theta} \mathbf{1}_{[b, \infty[}(\theta)$ . Si  $\frac{1}{\theta} q_0(\frac{y}{\theta})$  est continuellement différentiable et  $\frac{\frac{1}{\theta} q_0(\frac{y}{\theta})}{q_0(y)}$  est non décroissant en  $y$  pour  $\theta \geq 1$ , alors l'estimateur de Bayes  $\widehat{p}_\pi$  est minimax pour  $\theta \geq b$ .*

*Démonstration.* Effectuons les changements de variables  $Y' = \log(Y)$  et  $\theta' = \log(\theta)$ . La Proposition 4.2.1 nous assure que le problème d'estimation de la densité de  $Y'|\theta'$  est équivalent au problème d'estimation de la densité de  $Y|\theta$ . On se retrouve donc avec un problème d'estimation où  $\theta'$  est un paramètre de position. Nous noterons la densité de  $Y'|\theta'$  par  $q'_0(y' - \theta')$ .

Comme  $\frac{1}{\theta} q_0(\frac{y}{\theta})$  est continuellement différentiable et  $\frac{\frac{1}{\theta} q_0(\frac{y}{\theta})}{q_0(y)}$  est non décroissant en  $y$  pour  $\theta \geq 1$ , alors  $q'_0(y' - \theta')$  est continuellement différentiable et  $\frac{q'_0(y' - \theta')}{q'_0(y')}$  est non décroissant en  $y'$  pour  $\theta' \geq 0$ , car nous avons proposé un changement de variables monotone croissant. Ainsi, par le Théorème 4.1.2, l'estimateur de Bayes  $\widehat{p}_{\pi'}$  est minimax pour  $\theta' \geq \log(b)$ . Par l'équivalence assurée par la Proposition 4.2.1, l'estimateur de Bayes  $\widehat{p}_\pi$  est minimax pour  $\theta \geq b$ .  $\square$

### 4.3 Cas particulier : Loi Gamma

Dans cette section, nous analyserons plus particulièrement le cas de la loi Gamma en illustrant la performance de certains estimateurs.

Considérons le problème d'estimation de la densité de  $Y \sim \text{Gamma}(\alpha_2, \beta)$  sous la perte Kullback-Leibler à partir d'une observation  $X \sim \text{Gamma}(\alpha_1, \beta)$  avec  $\alpha_1, \alpha_2 > 0$  connus et  $\beta$  inconnu, où la fonction de densité d'une variable aléatoire  $T$  de loi Gamma( $\alpha, \beta$ ) est  $f(t) = \frac{t^{\alpha-1} e^{-\frac{t}{\beta}}}{\beta^\alpha \Gamma(\alpha)} \mathbf{1}_{(0, \infty)}(t)$ .

**Proposition 4.3.1.** *La performance d'un estimateur par substitution  $\hat{p}_1(\cdot|x) \sim \text{Gamma}(\alpha_2, \delta_1(x))$  est équivalente à la performance de l'estimateur ponctuel  $\delta_1(X)$  de  $\beta$  sous la perte entropie  $L_E(\beta, \delta(x)) = \frac{\beta}{\delta(x)} - \log(\frac{\beta}{\delta(x)}) - 1$ .*

*Démonstration.* On a :

$$\begin{aligned} R_{KL}(\theta, \hat{p}_1) &= E^{X,Y}[\log(\frac{p(Y|\beta)}{\hat{p}_1(Y|X)})] \\ &= E^{X,Y}[\log(\frac{Y^{\alpha_2-1} e^{-\frac{Y}{\beta}}}{\beta^{\alpha_2} \Gamma(\alpha_2)} \frac{\delta_1(X)^{\alpha_2} \Gamma(\alpha_2)}{Y^{\alpha_2-1} e^{-\frac{Y}{\delta_1(X)}}}] \\ &= E^{X,Y}[Y(\frac{1}{\delta_1(X)} - \frac{1}{\beta}) + \alpha_2 \log(\frac{\delta_1(X)}{\beta})] \\ &= \alpha_2 E^X[\frac{\beta}{\delta_1(X)} - \log(\frac{\beta}{\delta_1(X)}) - 1], \end{aligned}$$

car  $E[Y] = \alpha_2 \beta$ . □

**Corollaire 4.3.2.** *Dans le problème défini plus tôt avec  $\alpha_1 > 1$ , l'estimateur par substitution  $\hat{p}_1(\cdot|x) \sim \text{Gamma}(\alpha_2, \frac{x}{\alpha_1-1})$  est le meilleur parmi les estimateurs par substitution de la forme  $\hat{p}(\cdot|x) \sim \text{Gamma}(\alpha_2, ax)$  sous la perte Kullback-Leibler.*

*Démonstration.* Par la Proposition 4.3.1, il suffit de montrer que l'estimateur ponctuel  $\delta_1(X) = \frac{X}{\alpha_1-1}$  est le meilleur parmi les estimateurs du type  $\delta(X) = aX$  sous la perte entropie pour  $\alpha_1 > 1$ .

$$\begin{aligned} E_\beta[L_E(\beta, aX)] &= E_\beta\left[\frac{\beta}{aX} - \log\left(\frac{\beta}{aX}\right) - 1\right] \\ &= \frac{1}{a(\alpha_1-1)} + \log(a) + E_\beta\left[\frac{X}{\beta}\right] - 1, \end{aligned}$$

car  $E_\beta\left[\frac{1}{X}\right] = \frac{1}{(\alpha_1-1)\beta}$ . Aussi, puisque  $\frac{X}{\beta} \sim \text{Gamma}(\alpha_1, 1)$  (indépendante de  $\beta$ ) et que  $E[|\log(X)|] < \infty$ , on obtient un minimum en  $a = \frac{1}{\alpha_1-1}$ .  $\square$

La proposition suivante nous donne des formes closes pour, respectivement, l'estimateur de Bayes sous une loi *a priori* impropre non-informative et pour une loi *a priori* quelconque.

**Proposition 4.3.3.** a) Pour la loi *a priori* non informative  $\pi_0(\beta) = \frac{1}{\beta} \mathbf{1}_{\mathbb{R}^+}(\beta)$ , l'estimateur de Bayes est  $\hat{p}_0(y|x) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{1}{x} \left(\frac{y}{x}\right)^{\alpha_2-1} \left(1 + \frac{y}{x}\right)^{-(\alpha_1+\alpha_2)} \mathbf{1}_{\mathbb{R}^+}(y)$ .

b) Pour une loi *a priori*  $\pi(\beta)$  par rapport à une mesure  $\nu$ , l'estimateur de Bayes est  $\hat{p}_\pi(y|x) = \hat{p}_0(y|x) \frac{z}{x} \frac{m_{\pi,Z}(z)}{m_{\pi,X}(x)}$ , où  $m_{\pi,X}(\cdot)$  et  $m_{\pi,Z}(\cdot)$  sont les densités marginales de  $X$  et  $Z = X + Y$  respectivement.

*Démonstration.* (a) L'estimateur de Bayes dans ce contexte est :

$$\begin{aligned} \hat{p}_0(y|x) &= \frac{\int_0^\infty \frac{y^{\alpha_2-1} e^{-\frac{y}{\beta}}}{\beta^{\alpha_2} \Gamma(\alpha_2)} \frac{x^{\alpha_1-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha_1} \Gamma(\alpha_1)} \frac{1}{\beta} d\beta}{\int_0^\infty \frac{x^{\alpha_1-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha_1} \Gamma(\alpha_1)} \frac{1}{\beta} d\beta} \\ &= \frac{\left(\frac{1}{x+y}\right)^{\alpha_1+\alpha_2} \frac{y^{\alpha_2-1} x^{\alpha_1-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty t^{\alpha_1+\alpha_2-1} e^{-t} dt}{\frac{1}{x\Gamma(\alpha_1)} \int_0^\infty t^{\alpha_1-1} e^{-t} dt} \\ &= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{1}{x} \left(\frac{y}{x}\right)^{\alpha_2-1} \left(1 + \frac{y}{x}\right)^{-(\alpha_1+\alpha_2)}, \end{aligned}$$

où la deuxième égalité se trouve en effectuant les changements de variables  $t = \frac{x+y}{\beta}$  et  $t = \frac{x}{\beta}$ .

b) Dans ce contexte, l'estimateur de Bayes est :

$$\begin{aligned}
\widehat{p}_\pi(y|x) &= \frac{\int_0^\infty \frac{y^{\alpha_2-1} e^{-\frac{y}{\beta}}}{\beta^{\alpha_2} \Gamma(\alpha_2)} \frac{x^{\alpha_1-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha_1} \Gamma(\alpha_1)} \pi(\beta) d\nu(\beta)}{\frac{m_X(x)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}} \\
&= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{y^{\alpha_2-1} x^{\alpha_1-1}}{(x+y)^{\alpha_1+\alpha_2-1}} \frac{m_Z(z)}{m_X(x)} \\
&= \widehat{p}_0(y|x) \frac{z}{x} \frac{m_Z(z)}{m_X(x)}.
\end{aligned}$$

□

Il est à remarquer que la fonction de densité en (a) est une fonction de densité de Fisher avec paramètre d'échelle  $\frac{\alpha_1 x}{\alpha_2}$  et degrés de liberté  $2\alpha_1$  et  $2\alpha_2$ , car la fonction de densité d'une variable aléatoire  $Z$  suivant une loi de Fisher de degrés de liberté  $d_1$  et  $d_2$  est :

$$f_Z(z) = \frac{\Gamma(\frac{d_1}{2} + \frac{d_2}{2})}{\Gamma(\frac{d_1}{2})\Gamma(\frac{d_2}{2})} \frac{d_1^{\frac{d_1}{2}} d_2^{\frac{d_2}{2}} y^{\frac{d_1}{2}-1}}{(d_1 y + d_2)^{\frac{d_1}{2} + \frac{d_2}{2}}}.$$

Pour la partie (b), dans le cas particulier où  $\beta \in ]0, 1]$  et où on utilise la troncature de la loi *a priori* non informative  $\pi(\beta) = \frac{1}{\beta} \mathbf{1}_{]0,1]}(\beta)$ , on a :  $\widehat{p}_\pi(y|x) = \widehat{p}_0(y|x) \frac{\overline{F}_{\alpha_1+\alpha_2,1}(z)}{\overline{F}_{\alpha_1,1}(x)}$  où la fonction  $\overline{F}_{\alpha,1}(\cdot)$  est la fonction de survie d'une loi gamma de paramètres  $\alpha$  et 1. Ce résultat se prouve en remarquant que, pour  $T|\beta$  de loi Gamma( $\alpha, \beta$ ) :

$$t \cdot m_T(t) = \int_0^1 \frac{t^\alpha e^{-\frac{t}{\beta}}}{\beta^{\alpha+1} \Gamma(\alpha)} \mathbf{1}_{]0,\infty[}(t) d\beta = \overline{F}_{\alpha,1}(t)$$

par le changement de variable  $u = \frac{t}{\beta}$ . Si, de plus,  $\alpha_1 = \alpha_2 = 1$ , on a une forme close pour l'estimateur. Il s'agit de :

$$\widehat{p}_\pi(y|x) = \widehat{p}_0(y|x) \frac{(z+1)e^{-z}}{e^{-x}} \mathbf{1}_{]0,\infty[}(x) \mathbf{1}_{]0,\infty[}(z) = \frac{x}{(x+y)^2} (x+y+1) e^{-y} \mathbf{1}_{(]0,\infty[)^2}(x,y).$$

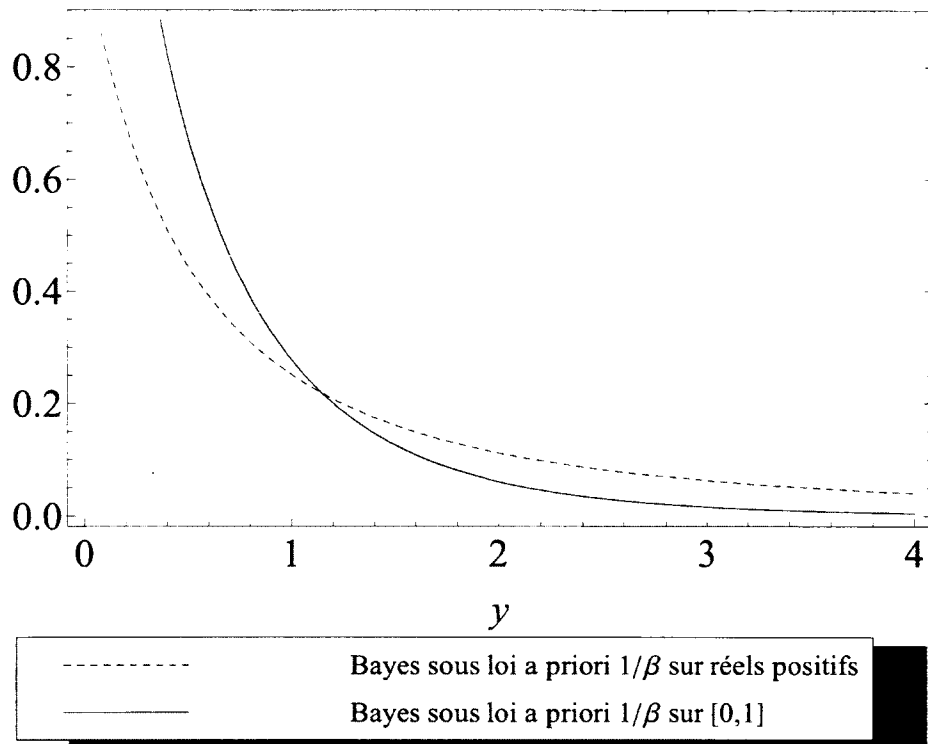


Figure 4.3 – Estimateurs de Bayes pour loi Gamma ( $\alpha_1 = \alpha_2 = 1$  avec observation  $x = 1$ ).

La figure 4.3 illustre les estimateurs de Bayes associés aux lois *a priori* non informatives  $\pi(\beta) = \frac{1}{\beta} \mathbf{1}_{\mathbb{R}^+}(\beta)$  et  $\pi(\beta) = \frac{1}{\beta} \mathbf{1}_{]0,1]}(\beta)$ . Lorsque  $\beta$  est restreint à  $]0,1]$ , cette restriction est prise en considération dans l'estimateur. Il est construit avec un paramètre d'échelle plus petit.

La question revient. Est-ce que cet estimateur domine l'estimateur de Bayes découlant de la loi *a priori* non informative  $\pi_0(\beta) = \frac{1}{\beta} \mathbf{1}_{\mathbb{R}^+}(\beta)$ ? C'est ce que le prochain théorème tente de démontrer.

**Théorème 4.3.4.** *Dans le cadre du problème énoncé au début de cette section, l'estimateur de Bayes découlant de la loi a priori non informative  $\pi(\beta) = \frac{1}{\beta} \mathbf{1}_{]0,c]}(\beta)$  avec*

$c < \infty$  domine l'estimateur de Bayes découlant de la loi a priori non informative  $\pi_0(\beta) = \frac{1}{\beta} \mathbf{1}_{\mathbf{R}^+}(\beta)$  pour  $\beta \in ]0, c]$ .

*Démonstration.* On sait, à partir des résultats démontrés plus tôt (Théorème 4.1.2 et Proposition 4.2.1), que cette dominance est assurée pour tout  $\alpha_1, \alpha_2$ .  $\square$

Mais, peut-on le démontrer directement? Le théorème nous dit que  $\Delta(\beta) \geq 0$  pour tout  $\beta \in ]0, c]$ , mais nous n'avons pas, pour l'instant, de démonstration générale directe. Cependant, nous en avons une pour  $\alpha_1 = \alpha_2 = 1$ .

*Démonstration.* Sans perte de généralité, supposons  $c = 1$ . Il faut alors montrer que la différence des risques  $\Delta(\beta)$  est positive. Or, comme  $m_{\pi, X}(x) = \frac{1}{x} \bar{F}_{\alpha_1, 1}(X)$  et  $m_{\pi, Z}(z) = \frac{1}{z} \bar{F}_{\alpha_1 + \alpha_2, 1}(z)$ , on peut déduire à partir de la partie (b) de la Proposition 4.3.3 que :

$$\begin{aligned} \Delta(\beta) &= E^{X, Y} \left[ \log \left( \frac{\bar{F}_{\alpha_1 + \alpha_2, 1}(Z)}{\bar{F}_{\alpha_1, 1}(X)} \right) \right] \\ &= E^T [\log(\bar{F}_{\alpha_1 + \alpha_2, 1}(\beta T))] - E^{T'} [\log(\bar{F}_{\alpha_1, 1}(\beta T'))], \end{aligned}$$

où  $T \sim \Gamma(\alpha_1 + \alpha_2, 1)$  et  $T' \sim \Gamma(\alpha_1, 1)$ .

Une condition suffisante serait de montrer que la fonction  $H(\alpha) = E_{\alpha}^T [\log(\bar{F}_{\alpha, 1}(\beta T))]$  est croissante en  $\alpha$  pour tout  $\beta \in ]0, 1]$ . Or, il est aisé de remarquer que  $\lim_{\beta \rightarrow 0^+} \Delta(\beta) = 0$ . Notons de plus que  $\Delta(1) = 0$ , car dans ce cas,  $\bar{F}_{\alpha_1 + \alpha_2, 1}(T) \sim U(0, 1)$  et  $\bar{F}_{\alpha_1, 1}(T') \sim U(0, 1)$ .

Donc, une condition suffisante est la concavité de  $\Delta(\beta)$  en  $\beta$  pour tout  $\alpha_1, \alpha_2$ . Pour  $\alpha_1 = \alpha_2 = 1$ , on a que

$$\Delta(\beta) = H(2) - H(1),$$

$$H(1) = E_{1, 1}^T [\log(e^{-\beta T})] = E_{1, 1}^T [-\beta T] = -\beta,$$

$$H(n) = E_{n,1}^T[\log(\sum_{k=0}^{n-1} \frac{(\beta T)^k}{k!}) - \beta T] = E_{n,1}^T[\log(\sum_{k=0}^{n-1} \frac{(\beta T)^k}{k!})] - n\beta.$$

$$H(2) = E_{2,1}^T[\log(1 + \beta T)] - 2\beta.$$

Ainsi,  $\frac{\delta}{\delta\beta}(H(2) - H(1)) = E_{2,1}^T[\frac{T}{1+\beta T}] - 1$  et  $\frac{\delta^2}{\delta\beta^2}(H(2) - H(1)) = E_{2,1}^T[\frac{-T^2}{(1+\beta T)^2}] < 0$ .

Donc, dans ce cas particulier,  $\Delta(\beta)$  est concave en  $\beta$ . Si on ajoute cette information au fait que  $\lim_{\beta \rightarrow 0^+} \Delta(\beta) = 0$  et que  $\Delta(1) = 0$ , alors on a dominance de l'estimateur  $\hat{p}_\pi$  sur  $\hat{p}_0$  pour la perte Kullback-Leibler, car  $\Delta(\beta)$  doit être positif ou nul sur  $]0, 1]$ .  $\square$

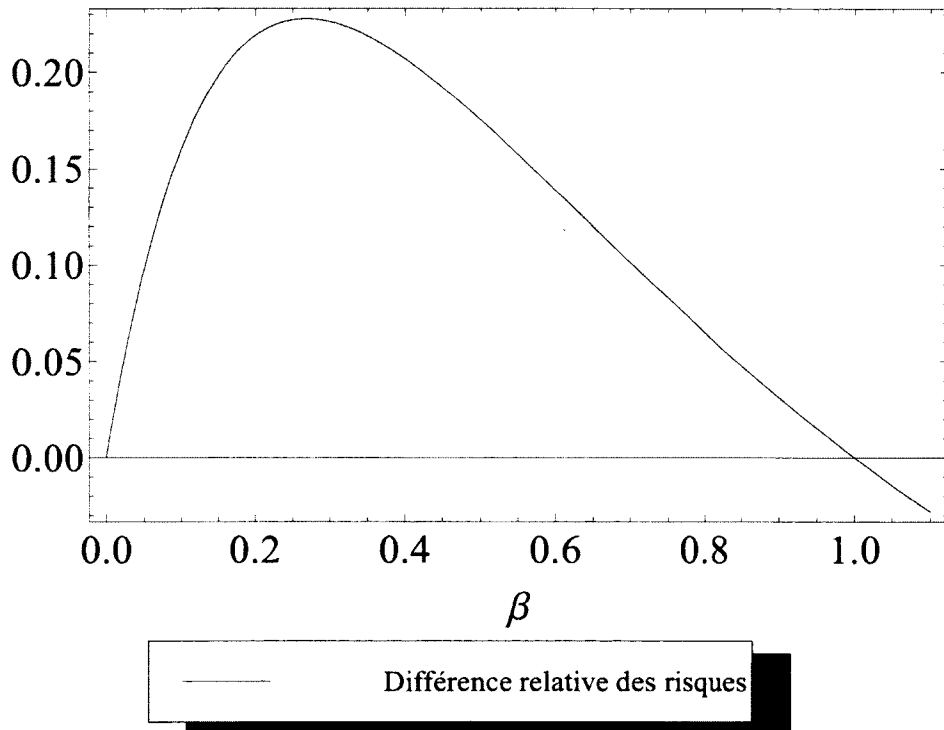


Figure 4.4 – Différence relative des risques pour estimateurs de Bayes sous loi *a priori*  $\frac{1}{\beta} \mathbb{1}_{\mathbb{R}^+}(\beta)$  et sous loi *a priori*  $\frac{1}{\beta} \mathbb{1}_{]0,1]}(\beta)$  pour loi Gamma ( $\alpha_1 = \alpha_2 = 1$ ).

On constate à l'observation de cette figure que la dominance est assurée pour  $\beta$  dans l'intervalle d'appartenance. Le gain relatif se situe aux alentours de 20 % à son maximum.

Il est à remarquer, comme le mentionne la Remarque 4.1.4, que le résultat s'étend au cas où l'espace paramétrique est de la forme  $[b, \infty[$  avec  $b > 0$  par un changement de variable  $X \rightarrow \frac{1}{X}$  en se rappelant que la perte Kullback-Leibler est intrinsèque tel que vu à la Proposition 1.2.3.

Pour conclure, ce chapitre a résumé les contributions principales du mémoire concernant l'estimation par densités prédictives lorsque l'espace paramétrique est restreint dans le cadre de familles de lois particulières à paramètre de position ou d'échelle.



## CHAPITRE 5

# Minimaxity in Predictive Density Estimation with Parametric Constraints

Ce chapitre contient l'article soumis en collaboration avec Kubokawa, Marchand et Strawderman. Il reprend plusieurs résultats traités dans ce mémoire, à savoir les conditions générales de minimaxité et de dominance d'estimateurs bayésiens pour des familles de densités à paramètres de position et d'échelle à l'intérieur d'un espace restreint. Plusieurs exemples sont illustrés pour mieux comprendre les résultats.

### **Abstract**

This paper is concerned with estimation of a predictive density with parametric constraints under Kullback-Leibler loss. When an invariance structure is embedded in the problem, general and unified conditions for the minimaxity of the best equivariant predictive density estimator are derived. These conditions are applied to check minimaxity in various

restricted parameter spaces in location and/or scale families. Further, it is shown that the generalized Bayes estimator against the uniform prior over the restricted space is minimax and dominates the best equivariant estimator in a location family when the parameter is restricted to an interval of the form  $[a_0, \infty)$ . Similar findings are obtained for scale parameter families. Finally, the presentation is accompanied by various observations and illustrations, such as normal, exponential location, and gamma model examples.

*AMS 2000 subject classifications* : 62C20, 62C86, 62F10, 62F15, 62F30.

*Key words and phrases* : Bayes estimators, decision theory, dominance, Kullback-Leibler loss, invariance, location family, location-scale family, minimaxity, order restriction, predictive density, restricted parameter space, scale family.

## 5.1 Introduction

### 5.1.1 Preamble

We consider here predictive density estimation for continuous models with

$$X \sim p_\theta(\cdot), Y \sim q_\theta(\cdot), \quad (5.1.1)$$

where the parameter  $\theta$  is restricted. We seek efficient estimators  $\hat{q}(\cdot|X)$  of  $q_\theta$  based on  $X$  under Kullback-Leibler loss

$$L_{KL}(q_\theta, \hat{q}) = \int q_\theta(y) \log \frac{q_\theta(y)}{\hat{q}(y)} dy, \quad (5.1.2)$$

and as measured by the Kullback-Leibler risk

$$R_{KL}(\theta, \hat{q}) = E_\theta^{X,Y} L_{KL}(q_\theta, \hat{q}(Y|X)). \quad (5.1.3)$$

Such a framework includes normal models with, for instance,  $X \sim N_p(\mu, \sigma_X^2 I_p)$ ,  $Y \sim N_p(\mu, \sigma_Y^2 I_p)$  with  $\mu$  restricted to a convex subset of  $\mathbb{R}^p$  as studied recently by Fourdrinier et al. (2011). Our findings will focus on two fundamental questions :

- (A) whether the best equivariant procedure  $\hat{q}^{BI}$  is minimax for both the unrestricted version and the restricted version of the problem ;

(B) whether the Bayes estimator  $\hat{q}^U$  with respect to the truncated (onto the restricted parameter space) right Haar invariant measure improves upon uniformly on  $\hat{q}^{BI}$ .

Part (A) requires an invariance structure which we will expand on in Section 2. Point estimation *unrestricted* parameter space versions of (A), with affirmative answers in many situations, date back to Girshick and Savage (1951), Kiefer (1957), Hora and Buehler (1966, 1967), among others. Point estimation *restricted parameter* versions of (A) and (B), with affirmative answers, date back to Katz (1961) who showed under squared error loss that the Bayes estimator with respect to the flat prior on  $[0, \infty)$ , for normal models with mean  $\mu$  and known variance, dominates the best equivariant estimator and is minimax for the restricted parameter space  $\mu \in [0, \infty)$ . There are several related results in the literature (e.g., Farrell, 1964; Kubokawa, 2004; Marchand and Strawderman, 2005A,B; Tsukuma and Kubokawa, 2008) for restricted (unbounded) parameter spaces, with a quite general minimax result given recently by Marchand and Strawderman (2012). As further illustrated by the work of Casella and Strawderman (1981), Marchand and Perron (2001), Hartigan (2004), Marchand and Strawderman (2004), Kubokawa (2005A,B), and van Eeden (2006) among others, frequentist properties like minimaxity of best equivariant estimators, restricted maximum likelihood estimators or Bayesian estimators depend on the model, the loss, but also intimately on the nature of the parametric restriction.

Predictive density estimation addresses the challenging and ambitious problem of estimating the whole distribution of a future observation  $Y$ . This has become a field of active study with early findings due to Aitchison (1975). In particular, for Gaussian models under Kullback-Leibler loss, fascinating connections with *Stein estimation* have been developed, as recently reviewed by George, Liang, and Xu (2012), and as expanded upon below in subsection 1.3.

### 5.1.2 Outline of Paper

In this paper, we investigate minimaxity of the best equivariant predictive density estimator in location and/or scale families with parametric constraints under Kullback-

Leibler loss. In Section 5.2, we treat a setup with a general invariance structure given by Hora and Buehler (1966, 67), where the parameter space is restricted to a subset of multi-dimensional Euclidean space. Using similar arguments as in Girshick and Savage (1951), we derive unified conditions under which the best equivariant estimator is minimax. These conditions are available for both restricted and non-restricted cases, and in a sense, the minimaxity result is an extension of findings by Liang and Barron (2004), who showed minimaxity when the parameter space is unrestricted. Minimaxity under parametric constraints for a given type of problem can thus be tested by checking those unified conditions.

Section 5.3 deals with a location or scale family. In Section 5.3.1, minimaxity of the best location equivariant estimator is verified under a one-sided restriction of the location parameter in a location family. In section 3.2, we make use of a novel variation of the IERD method introduced by Kubokawa (1994A,B) and Kubokawa and Saleh (1998) to prove that the generalized Bayes estimator against the uniform prior over the restricted space dominates the best location equivariant estimator if the target density to be predicted has a monotone likelihood ratio property. It is interesting to note that the density of the observation does not have to have a monotone likelihood ratio and need not be of the same family as the target density. Analogous findings for scale parameter families are obtained in Section 5.3.3. Various other observations, detailed examples for normal, exponential and gamma models, and a non-minimaxity result for a compact interval restriction, complement the presentation.

In Section 5.4, we treat various restrictions in location-scale families and investigate minimaxity of the best location-scale equivariant estimator. Section 5.4.1 considers the cases that the location and scale parameters are in one-sided open spaces, and Section 5.4.2 investigates cases with a compact interval restriction for the location parameter and an unknown scale. Through several examples of parametric restrictions given in Sections 5.4.1 and 5.4.2, we demonstrate how to use the conditions given in Section 5.2. Minimaxity in the cases of ordered location or scale parameters in multidimensional distributions is shown in Section 5.4.3.

### 5.1.3 Brief review of previous findings for normal models

We conclude this introduction with a brief review on developments under a multivariate normal distribution with unknown mean vector and known covariance matrices which are multiples of identity, since most decision-theoretic results have been studied in this model and since such a review is helpful for the overall presentation of our findings. Let  $\mathbf{X}$  and  $\mathbf{Y}$  be mutually independent random vectors such that  $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, v_x \mathbf{I})$  and  $\mathbf{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, v_y \mathbf{I})$  for known constants  $v_x$  and  $v_y$ . The density functions of  $\mathbf{X}$  and  $\mathbf{Y}$  are denoted by  $f(\mathbf{x} - \boldsymbol{\mu}|v_x)$  and  $f(\mathbf{y} - \boldsymbol{\mu}|v_y)$ . The problem is to predict the density  $f(\mathbf{y} - \boldsymbol{\mu}|v_y)$  based on  $\mathbf{X}$  in terms of the following risk relative of the Kullback-Leibler (KL) divergence

$$R_{KL}(\boldsymbol{\mu}, \hat{f}) = E_{\mathbf{X}} \left[ \int f(\mathbf{y} - \boldsymbol{\mu}|v_y) \log \left( \frac{f(\mathbf{y} - \boldsymbol{\mu}|v_y)}{\hat{f}(\mathbf{y}|\mathbf{X}, v_x, v_y)} \right) d\mathbf{y} \right],$$

where  $\hat{f}(\mathbf{y}|\mathbf{X}, v_x, v_y)$  is a predictive density estimator of  $f(\mathbf{y} - \boldsymbol{\mu}|v_y)$ . Since this model is invariant under location transformations, the best equivariant estimator of  $f(\mathbf{y} - \boldsymbol{\mu}|v_y)$  is the generalized Bayes estimator against the uniform prior with respect to Lebesgue measure. As expressed in (5.2.3), the best equivariant estimator is given by

$$\hat{f}^{BI}(\mathbf{y} - \mathbf{x}|v_x, v_y) = \frac{\int f(\mathbf{s}|v_x) f(\mathbf{y} - \mathbf{x} + \mathbf{s}|v_y) d\mathbf{s}}{\int f(\mathbf{s}|v_x) d\mathbf{s}} = f(\mathbf{y} - \mathbf{x}|v_x + v_y).$$

Liang and Barron (2004) showed that  $\hat{f}^{BI}(\mathbf{y} - \mathbf{x}|v_x, v_y)$  is minimax. Concerning the admissibility of  $\hat{f}^{BI}(\mathbf{y} - \mathbf{x}|v_x, v_y)$  in the case of normal distributions, Komaki (2001) showed that it is inadmissible when  $p \geq 3$ , namely, it is improved on by a generalized Bayes estimator against a shrinkage prior. Brown, George and Xu (2008) showed that it is admissible when  $p = 1, 2$ . These are noteworthy results in the sense that the so called Stein inadmissibility result in point estimation is inherited by the problem of estimation of a predictive normal density function. George, Liang and Xu (2006) extended Komaki's result, and along with Brown, George and Xu (2008), showed that several decision-theoretic results for point estimation of a multivariate normal mean with a known variance still hold for the predictive density estimation problem. Kato (2009) succeeded in deriving a minimax and improved generalized Bayes predictive density estimator in the case of unknown variance.

Brown, George and Xu (2008) derived an interesting identity which expresses the relationship between point estimation and predictive density estimation. Let  $R_Q^v(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}})$  be the risk function of a point estimator  $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}(\mathbf{z})$  under a normal distribution  $\mathcal{N}_p(\boldsymbol{\mu}, v\mathbf{I})$ , namely,

$$R_Q^v(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}) = \int \|\hat{\boldsymbol{\mu}}(\mathbf{z}) - \boldsymbol{\mu}\|^2 f(\mathbf{z} - \boldsymbol{\mu}|v) d\mathbf{z},$$

for the Euclidean norm  $\|\cdot\|$ . Let  $\hat{\boldsymbol{\mu}}_v^\pi$  be the Bayes point estimator of  $\boldsymbol{\mu}$  for a prior distribution  $\pi(\boldsymbol{\mu})$  in terms of the loss  $\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2$ . Also, let  $\hat{f}^\pi(\mathbf{y}|\mathbf{x}, v_x, v_y)$  be the Bayes estimator of the predictive density. Then, Brown, Brown, George and Xu (2008) showed that

$$R_{KL}(\boldsymbol{\mu}, \hat{f}^{BI}) - R_{KL}(\boldsymbol{\mu}, \hat{f}^\pi) = \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} [R_Q^v(\boldsymbol{\mu}, \mathbf{X}) - R_Q^v(\boldsymbol{\mu}, \hat{\boldsymbol{\mu}}_v^\pi)] dv, \quad (5.1.4)$$

for  $v_w = v_x v_y / (v_x + v_y)$ . This implies that dominance properties in point estimation can be automatically inherited by predictive density estimation. An essential point in the above identity is that in the normal distribution, the following representation due to George, Liang and Xu (2006) holds :

$$\hat{f}^\pi(\mathbf{y}|\mathbf{x}, v_x, v_y) = \frac{m_\pi(\mathbf{W}; v_w)}{m_\pi(\mathbf{X}; v_x)} \hat{f}^{BI}(\mathbf{y}|\mathbf{x}, v_x, v_y), \quad (5.1.5)$$

where  $m_\pi(\mathbf{W}; v_w)$  and  $m_\pi(\mathbf{X}; v_x)$  are marginal densities of  $\mathbf{W}$  and  $\mathbf{X}$  for  $\mathbf{W} = (v_y \mathbf{X} + v_x \mathbf{Y}) / (v_x + v_y)$ . Using this equality, Fourdrinier, *et al.* (2011) extended identity (5.1.4) to plug-in estimators of the predictive density.

Identity (5.1.4) can be applied when the parameter space  $\boldsymbol{\theta}$  is restricted to a convex cone  $C$ , or more generally to a convex set. In the framework of point estimation under a constraint and squared error loss, Hartigan (2004) proved that  $\mathbf{X}$  is improved on by the generalized Bayes estimator against the uniform prior over  $C$ , and Tsukuma and Kubokawa (2008) showed that  $\mathbf{X}$  is minimax under the constraint. As developed in Fourdrinier *et al.* (2011), combining these results and the identity (5.1.4) implies that these properties hold for the estimation of the predictive density.

The inferences are valid for normal distributions where key property (5.1.4) can be derived from the equality (5.1.5). The equality (5.1.5) holds under normality with known

variances, but it does not hold in the case of unknown variances. Thus, it is not clear whether a decision-theoretic property in point estimation is inherited by estimation of the predictive density under normality with unknown variances or for another distribution.

## 5.2 General conditions for minimaxity

In this section, we treat general parametric distributions in which an invariance structure is embedded, and derive general conditions for minimaxity of the best equivariant estimator. The conditions will be used for checking minimaxity in location and/or scale families.

Let  $X$  be an observable random variable and  $Y$  be a future random variable. Let  $(\mathcal{X} \times \mathcal{Y}, \mathcal{B}_X \times \mathcal{B}_Y)$  be a measurable space of  $(X, Y)$  and  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$  be a family of identifiable probability measures with parameter space  $\Theta$ . We assume the following conditions.

**(A1)** There exist a group  $\mathcal{G}$  and a measurable space  $(\mathcal{G}, \mathcal{B}_G)$  on which there exists a left invariant Haar measure  $\gamma$  satisfying

$$\gamma(gG) = \gamma(G) \quad \text{for all } g \in \mathcal{G} \text{ and all } G \in \mathcal{B}_G.$$

Each  $g \in \mathcal{G}$  induces a one-to-one transformation  $\bar{g}$  from  $\Theta$  onto itself defined by  $P_{\bar{g}\theta}(gA) = P_\theta(A)$  for any  $A \in \mathcal{B}_X \times \mathcal{B}_Y$  and any  $\theta \in \Theta$ . The induced space  $\bar{\mathcal{G}} = \{\bar{g} : \bar{g} \in \mathcal{G}\}$  is measurable.

**(A2)** There exists a one-to-one correspondence  $X \leftrightarrow (t_x, u_x)$  between  $\mathcal{X}$  and  $\mathcal{G} \times \mathcal{U}_X$  such that  $gX$  corresponds to  $(gt_x, u_x)$  and  $\mathcal{U}_X$  is a measurable space. Also, there exists a one-to-one correspondence  $Y \leftrightarrow (t_y, u_y)$  between  $\mathcal{Y}$  and  $\mathcal{G} \times \mathcal{U}_Y$  such that  $gY$  corresponds to  $(gt_y, u_y)$  and  $\mathcal{U}_Y$  is a measurable space. The statistics  $u_x$  and  $u_y$  are maximal invariant under the transformation  $\mathcal{G}$ .

**(A3)** There exists a one-to-one correspondence  $\theta \leftrightarrow \bar{g}_\theta$  between  $\Theta$  and  $\bar{\mathcal{G}}$  such that  $\bar{g}\theta$  corresponds to  $\bar{g}\bar{g}_\theta$  for all  $\bar{g} \in \bar{\mathcal{G}}$ . The correspondence of  $\bar{g}_\theta$  in  $\mathcal{G}$  is denoted by  $g_\theta$ .

(A4) There exist conditional probability density functions  $p(g_\theta^{-1}t_x|u_x)$  and  $q(g_\theta^{-1}t_y|u_y)$  given  $u_x$  and  $u_y$  such that for all  $A \in \mathcal{B}_X$ ,  $B \in \mathcal{B}_Y$ ,

$$\begin{aligned} P_\theta[A] &= \int_A p(g_\theta^{-1}t_x|u_x)p_x(u_x)\gamma(dt_x)\gamma_x(du_x), \\ P_\theta[B] &= \int_B q(g_\theta^{-1}t_y|u_y)q_y(u_y)\gamma(dt_y)\gamma_y(du_y), \end{aligned}$$

where  $p_x(\cdot)$  is a marginal density function of  $u_x$  with respect to a measure  $\gamma_x(\cdot)$  on  $\mathcal{U}_X$ , and  $p_y(\cdot)$  and  $\gamma_y(\cdot)$  are defined similarly.

We define a measure  $\nu(\cdot)$  by

$$\nu(dg) = \gamma(dg^{-1}).$$

This is a right invariant Haar measure. Since  $\gamma(\cdot)$  is left invariant, it is noted that  $\gamma(hdg) = \gamma(dg)$  and  $\gamma((dg)h) = \Delta(h)\gamma(dg)$  for  $h, g \in \mathcal{G}$ , where  $\Delta(\cdot)$  is a modular function.

Now we can set up the problem of estimating the joint predictive density  $q(g_\theta^{-1}t_y|u_y)q_y(u_y)$  based on  $(t_x, u_x)$ . When we estimate  $q_\theta$  by a density  $\hat{q}(t_y|u_y, t_x, u_x)q_y(u_y)$ , we evaluate the performance using the Kullback-Leibler (KL) divergence in (5.1.2) and we may write

$$\begin{aligned} L_{KL}(\theta, \hat{q}(\cdot|\cdot, t_x, u_x)) &= \int q(g_\theta^{-1}t_y|u_y)q_y(u_y) \log \left( \frac{q(g_\theta^{-1}t_y|u_y)q_y(u_y)}{\hat{q}(t_y|u_y, t_x, u_x)q_y(u_y)} \right) \gamma(dt_y)\gamma_y(du_y) \\ &= E^{u_y} \left[ \int q(g_\theta^{-1}t_y|u_y) \log \left( \frac{q(g_\theta^{-1}t_y|u_y)}{\hat{q}(t_y|u_y, t_x, u_x)} \right) \gamma(dt_y) \right]. \end{aligned}$$

Then, the risk function is

$$R_{KL}(\theta, \hat{q}) = E[L_{KL}(\theta, \hat{q}(\cdot|\cdot, t_x, u_x))] = E^{u_x, u_y}[R_{KL}(\theta, \hat{q}|u_x, u_y)], \quad (5.2.1)$$

where  $E^{u_x, u_y}[\cdot]$  is the expectation with respect to the marginal distribution of  $(u_x, u_y)$ , and  $R_{KL}(\theta, \hat{q}|u_x, u_y)$  is the conditional risk function given  $(u_x, u_y)$  equal to

$$\begin{aligned} R_{KL}(\theta, \hat{q}|u_x, u_y) &= \int p(g_\theta^{-1}t_x|u_x)\gamma(dt_x) \\ &\quad \times \left\{ \int q(g_\theta^{-1}t_y|u_y) \log \left( \frac{q(g_\theta^{-1}t_y|u_y)}{\hat{q}(t_y|u_y, t_x, u_x)} \right) \gamma(dt_y) \right\}. \end{aligned} \quad (5.2.2)$$



This demonstrates that estimation of the joint density function  $q(g_\theta^{-1}t_y|u_y)q_y(u_y)$  can be reduced to that of estimating the conditional density function  $q(g_\theta^{-1}t_y|u_y)$  as long as estimators of the form  $\hat{q}(t_y|u_y, t_x, u_x)q_y(u_y)$  are considered.

Since the problem has an invariance structure, we can derive the best equivariant estimator. Conditional predictive density equivariant estimators under the transformation  $\mathcal{G}$  satisfy

$$\hat{q}(gt_y|u_y, gt_x, u_x) = \hat{q}(t_y|u_y, t_x, u_x) \quad \text{for all } g \in \mathcal{G},$$

which implies that a class of (nonrandomized) equivariant estimators is given by

$$\mathcal{Q}_I = \left\{ \hat{q}_I(t_x^{-1}t_y|u_y, u_x) \mid \int \hat{q}_I(s|u_y, u_x)\gamma(ds) = 1 \right\}.$$

The best equivariant estimator is given in the following proposition.

**Proposition 5.2.1.** *Assume conditions (A1) to (A4). Then, the best equivariant estimator of  $q(g_\theta^{-1}t_y|u_y)$  is given by*

$$\begin{aligned} \hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x) &= \int p(t|u_x) q(tt_x^{-1}t_y|u_y) \gamma(dt) \\ &= \frac{\int p(g^{-1}t_x|u_x) q(g^{-1}t_y|u_y) \nu(dg)}{\int p(g^{-1}t_x|u_x) \nu(dg)}. \end{aligned} \quad (5.2.3)$$

**Proof.** Note that the conditional risk function of  $\hat{q}_I(t_x^{-1}t_y|u_y, u_x)$  is free from  $\theta$ , and from (5.2.2), it is expressed as

$$\begin{aligned} R_{KL}(\hat{q}_I|u_x, u_y) &= \int \int p(t_x|u_x) q(t_y|u_y) \log \left( \frac{q(t_y|u_y)}{\hat{q}_I(t_x^{-1}t_y|u_y, u_x)} \right) \gamma(dt_x) \gamma(dt_y) \\ &= \int \int p(t_x|u_x) q(t_x s|u_y) \log \left( \frac{q(t_x s|u_y)}{\hat{q}_I(s|u_y, u_x)} \right) \gamma(dt_x) \gamma(ds), \end{aligned} \quad (5.2.4)$$

where  $s = t_x^{-1}t_y$  and  $\gamma(t_x ds) = \gamma(ds)$  for the left invariant measure  $\gamma(\cdot)$ . With the alternative rewriting

$$\begin{aligned} R_{KL}(\hat{q}_I|u_x, u_y) &= \int \int p(t_x|u_x) q(t_x s|u_y) \gamma(dt_x) \log \left( \frac{\int p(t_x|u_x) q(t_x s|u_y) \gamma(dt_x)}{\hat{q}_I(s|u_y, u_x)} \right) \gamma(ds) \\ &\quad + \int \int p(t_x|u_x) q(t_x s|u_y) \log \left( \frac{q(t_x s|u_y)}{\int p(t_x|u_x) q(t_x s|u_y) \gamma(dt_x)} \right) \gamma(dt_x) \gamma(ds), \end{aligned}$$

it is seen that the best equivariant predictive density estimator is

$$\hat{q}^{BI}(s|u_y, u_x) = \int p(t|u_x) q(ts|u_y) \gamma(dt). \quad (5.2.5)$$

Making the transformation  $t = g^{-1}t_x$ , we see that

$$\gamma(d(g^{-1}t_x)) = \gamma((dg^{-1})t_x) = \Delta(t_x)\gamma(g^{-1}) = \Delta(t_x)\nu(dg).$$

Since  $1 = \int p(t|u_x)\gamma(dt) = \int p(g^{-1}t_x|u_x)\Delta(t_x)\nu(dg)$ , it is seen that

$$\Delta(t_x) = 1 / \int p(g^{-1}t_x|u_x)\nu(dg).$$

Substituting  $s = t_x^{-1}t_y$  into (5.2.5) and using the above arguments shows that  $\hat{q}^{BI}(s|u_y, u_x)$  is expressed as (5.2.3).  $\square$

As seen from the form in (5.2.3), the best equivariant estimator is the generalized Bayes predictive density estimator against the right invariant measure  $\nu(dg)$ . Liang and Barron (2004) showed that the best equivariant estimator  $\hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)$  is *minimax* if the group  $\mathcal{G}$  is *amenable*, namely, if there is a sequence of probability measures  $\gamma_j(\cdot)$  on  $\mathcal{G}$  that is asymptotically invariant in the sense that  $\lim_{j \rightarrow \infty} \int \{\psi(ag) - \psi(a)\}\gamma_j(da) = 0$  for every  $g \in \mathcal{G}$  and every bounded measurable function  $\psi$  on  $\mathcal{G}$ . However, the best equivariant estimator is not necessarily minimax when the parameter space is restricted.

We now provide unified conditions for the minimaxity of the best equivariant predictive density estimator. Although the conditions can be applied to both cases that parameters are restricted and non-restricted, they lead to new findings in restricted cases only, since minimaxity in non-restricted cases follows from the result of Liang and Barron (2004).

**(A5)**  $\Theta$  is restricted, and this restriction is equivalently expressed as  $g_\theta \in P$ . Also, it is assumed that  $P \subset \mathcal{G} \subset \mathbb{R}^r$ ; namely,  $\mathcal{G}$  is a subset of  $r$  dimensional Euclidean space and  $P$  is a restricted space of  $\mathcal{G}$ .

**(A6)** There exist sequences of subsets  $P_k(\subset P)$  and one-to-one functions  $h_k(\cdot)$  between  $P_k \leftrightarrow \Xi \subset \mathbb{R}^r$  with  $\xi = h_k(g_\theta)$  for  $g_\theta \in P_k$ , where  $P_k$ ,  $h_k(\cdot)$  and  $\Xi$  satisfy the following conditions :

**(A6-1)**  $\cup_{k=k_0}^{\infty} P_k = P$  for some  $k_0 \geq 1$ .

**(A6-2)** Let  $V(P_k) = \int_{P_k} \nu(dg_\theta)$ . Let  $\gamma_k(\cdot)$  be an induced measure defined by  $\gamma_k(A) = \nu(h_k^{-1}(A))$  for  $A \in \Xi$ . Then,  $h_k(P_k) = \Xi = \prod_{i=1}^r [-1 + a_{i,k}, 1 + b_{i,k}]$  and

$$\int_{h_k(P_k)} f(\xi_k) \gamma_k(d\xi_k) / V(P_k) \geq \frac{1}{2^r + c_k} \int I\left(\xi \in \prod_{i=1}^r [-1 + a_{i,k}, 1 + b_{i,k}]\right) f(\xi) d\xi, \quad (5.2.6)$$

where  $f(\cdot) > 0$ ,  $\xi_k = h_k(g_\theta)$ ,  $I(\cdot)$  is the indicator function, and  $\lim_{k \rightarrow \infty} a_{i,k} = \lim_{k \rightarrow \infty} b_{i,k} = \lim_{k \rightarrow \infty} c_k = 0$  for  $i = 1, \dots, r$ .

**(A6-3)** For any small enough  $\varepsilon > 0$  and any  $\xi \in \prod_{i=1}^r [-1 + a_{i,k} + \varepsilon, 1 + b_{i,k} - \varepsilon]$ , there exists a sequence of subsets  $P_k^*$  such that  $P_k^*$  does not depend on  $\xi$ ,  $\cup_{k=k_1}^\infty P_k^* = \mathcal{G}$  for some  $k_1 \geq 1$  and

$$P_k^* \subset \{[h_k^{-1}(\xi)]^{-1}g; g \in P_k\}.$$

**Théoreme 5.2.2.** *Assume conditions (A1) to (A6-3). Then, the best equivariant estimator  $\hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)$  is minimax in estimation of the conditional density  $q(g_\theta^{-1}t_y|u_y)$  in terms of the conditional risk (5.2.2).*

**Proof.** We can show this theorem along the same lines as in Kubokawa (2004) who modified the method of Girshick and Savage (1951). Consider the sequence of prior distributions given by

$$\pi_k(g_\theta)\nu(dg_\theta) = \begin{cases} \{V(P_k)\}^{-1}\nu(dg_\theta) & \text{if } g_\theta \in P_k \\ 0 & \text{otherwise.} \end{cases}$$

This yields the Bayesian predictive densities

$$\hat{q}_k^\pi(t_y|u_y, t_x, u_x) = \int_{P_k} p(g^{-1}t_x|u_x)q(g^{-1}t_y|u_y)\nu(dg) \int_{P_k} p(g^{-1}t_x|u_x)\nu(dg)$$

with conditional Bayes risks

$$\begin{aligned} r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) &= \frac{1}{V(P_k)} \int_{P_k} \int \int p(g_\theta^{-1}t_x|u_x)q(g_\theta^{-1}t_y|u_y) \\ &\quad \times \log\left(\frac{q(g_\theta^{-1}t_y|u_y)}{\hat{q}_k^\pi(t_y|u_y, t_x, u_x)}\right) \gamma(dt_x)\gamma(dt_y)\nu(dg_\theta). \end{aligned}$$

Since  $r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) \leq r_k(\pi_k, \hat{q}^{BI}|u_x, u_y) = R_0(u_x, u_y)$ , it is sufficient to show that  $\liminf_{k \rightarrow \infty} r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) \geq R_0(u_x, u_y)$ . Making the transformations  $s_x = g_\theta^{-1}t_x$  and

$s_y = g_\theta^{-1} t_y$  yields

$$r_k(\pi_k, \hat{q}_k^\pi | u_x, u_y) = \frac{1}{V(P_k)} \int_{P_k} \int \int p(s_x | u_x) q(s_y | u_y) \log \left( \frac{q(s_y | u_y)}{\hat{q}_k^\pi(g_\theta s_y | u_y, g_\theta s_x, u_x)} \right) \nu(dg_\theta) \\ \times \gamma(ds_x) \gamma(ds_y) \nu(dg_\theta), \quad (5.2.7)$$

where  $\hat{q}_k^\pi(g_\theta s_y | u_y, g_\theta s_x, u_x)$  is expressed as

$$\hat{q}_k^\pi(g_\theta s_y | u_y, g_\theta s_x, u_x) = \frac{\int_{P_k} p(g^{-1} g_\theta s_x | u_x) q(g^{-1} g_\theta s_y | u_y) \nu(dg)}{\int_{P_k} p(g^{-1} g_\theta s_x | u_x) \nu(dg)}.$$

Now, make the transformation  $g_1 = g_\theta^{-1} g$  with  $\nu(dg) = \Delta(g_\theta) \nu(dg_1)$  in order to rewrite  $\hat{q}_k^\pi(g_\theta s_y | u_y, g_\theta s_x, u_x)$  as

$$\hat{q}_k^\pi(g_\theta s_y | u_y, g_\theta s_x, u_x) = \frac{\int_{g_\theta g_1 \in P_k} p(g_1^{-1} s_x | u_x) q(g_1^{-1} s_y | u_y) \nu(dg_1)}{\int_{g_\theta g_1 \in P_k} p(g_1^{-1} s_x | u_x) \nu(dg_1)}.$$

In view of the assumptions, there exists a transformation  $\xi_k = h_k(g_\theta)$  satisfying the condition **(A6)**. Note that  $g_\theta g_1 \in P_k$  is equivalent to  $h_k^{-1}(\xi_k) g_1 \in P_k$ , or

$$g_1 \in \{[h_k^{-1}(\xi_k)]^{-1} g; g \in P_k\} \equiv \tilde{P}_k(\xi_k).$$

Then, the Bayes estimator  $\hat{g}_k^\pi(g_\theta s_y | u_y, g_\theta s_x, u_x)$  is rewritten as

$$\hat{g}_k^\pi(h_k^{-1}(\xi_k) s_y | u_y, h_k^{-1}(\xi_k) s_x, u_x) = \frac{\int_{g_1 \in \tilde{P}_k(\xi_k)} p(g_1^{-1} s_x | u_x) q(g_1^{-1} s_y | u_y) \nu(dg_1)}{\int_{g_1 \in \tilde{P}_k(\xi_k)} p(g_1^{-1} s_x | u_x) \nu(dg_1)}, \quad (5.2.8)$$

and the conditional Bayes risk (5.2.7) is rewritten as

$$r_k(\pi_k, \hat{q}_k^\pi | u_x, u_y) = \frac{1}{V(P_k)} \int_{h_k(P_k)} \int \int p(s_x | u_x) q(s_y | u_y) \\ \times \log \left( \frac{q(s_y | u_y)}{\hat{q}_k^\pi(h_k^{-1}(\xi_k) s_y | u_y, h_k^{-1}(\xi_k) s_x, u_x)} \right) \gamma_k(d\xi_k) \gamma(ds_x) \gamma(ds_y).$$

It is noted that from **(A6-2)**, for any small  $\varepsilon > 0$ ,

$$h_k(P_k) = \prod_{i=1}^r [-1 + a_{i,k}, 1 + b_{i,k}] \supset \prod_{i=1}^r [-1 + a_{i,k} + \varepsilon, 1 + b_{i,k} - \varepsilon] \equiv I_{k,\varepsilon}.$$

Then from (5.2.6), the conditional Bayes risk is evaluated as

$$r_k(\pi_k, \hat{q}_k^\pi | u_x, u_y) \geq \frac{1}{2^r} \int I(\xi \in I_{k,\varepsilon}) \int \int p(s_x | u_x) q(s_y | u_y) \\ \times \log \left( \frac{q(s_y | u_y)}{\hat{q}_k^\pi(h_k^{-1}(\xi) s_y | u_y, h_k^{-1}(\xi) s_x, u_x)} \right) d\xi \gamma(ds_x) \gamma(ds_y).$$

For  $\xi \in I_{k,\varepsilon}$ , from **(A6-3)**, it can be seen that  $\hat{q}_k^\pi(h_k^{-1}(\xi)s_y|u_y, h_k^{-1}(\xi)s_x, u_x) \rightarrow \hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)$  as  $k \rightarrow \infty$ . Hence, Fatou's lemma is used to bound the Bayes risks as

$$\begin{aligned}
\liminf_{k \rightarrow \infty} r_k(\pi_k, \hat{q}_k^\pi | u_x, u_y) &\geq \frac{1}{2^r} \int \liminf_{k \rightarrow \infty} I(\xi \in I_{k,\varepsilon}) \int \int p(s_x|u_x)q(s_y|u_y) \\
&\quad \times \liminf_{k \rightarrow \infty} \log \left( \frac{q(s_y|u_y)}{\hat{q}_k^\pi(h_k^{-1}(\xi)s_y|u_y, h_k^{-1}(\xi)s_x, u_x)} \right) \gamma(ds_x)\gamma(ds_y) d\xi \\
&= \frac{1}{2^r} \int_{|-1+\varepsilon, 1-\varepsilon|^r} d\xi \\
&\quad \times \int \int p(s_x|u_x)q(s_y|u_y) \log \left( \frac{q(s_y|u_y)}{\hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)} \right) \gamma(ds_x)\gamma(ds_y) \\
&= (1-\varepsilon)^r R(\theta, \hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)) = (1-\varepsilon)^r R_0(u_x, u_y)
\end{aligned}$$

From the arbitrariness of  $\varepsilon > 0$ , it follows that  $\liminf_{k \rightarrow \infty} r_k(\pi_k, \hat{q}_k^\pi | u_x, u_y) \geq R_0(u_x, u_y)$ , completing the proof of Theorem 5.2.2.  $\square$

In the above proof, the Bayes risk is given by  $r_k(\pi_k, \hat{q}_k^\pi) = E^{u_x, u_y}[r_k(\pi_k, \hat{q}_k^\pi | u_x, u_y)]$ . It is easy to see that  $r_k(\pi_k, \hat{q}_k^\pi) \leq E^{u_x, u_y}[r_k(\pi_k, \hat{q}^{BI} | u_x, u_y)] = E^{u_x, u_y}[R_0(u_x, u_y)]$ . On the other hand, Fatou's lemma is used to evaluate the Bayes risk as  $\liminf_{k \rightarrow \infty} r_k(\pi_k, \hat{q}_k^\pi) \geq E^{u_x, u_y}[\liminf_{k \rightarrow \infty} r_k(\pi_k, \hat{q}_k^\pi | u_x, u_y)] \geq E^{u_x, u_y}[R_0(u_x, u_y)]$ . Thus, we get the following corollary.

**Corollaire 5.2.3.** *Assume conditions **(A1)** to **(A6-3)**. Then, the best equivariant estimator  $\hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)q_y(u_y)$  is minimax for the estimation of the joint density  $q(g_\theta^{-1}t_y|u_y)q_y(u_y)$  in terms of the Kullback-Leibler risk (5.2.1).*

As we will show in various situations, Theorem 5.2.2 includes both non-restricted and restricted cases and thus provides a unified result for the minimaxity of the best equivariant estimator.

## 5.3 Location and scale families : minimaxity and improvements on $\hat{q}^{BI}$

### 5.3.1 Minimaxity for location families

We first deal with the estimation of a density with a restricted location parameter. Let  $\mathbf{X} = (X_1, \dots, X_{n_1})$  be a random variable having a density  $f(\mathbf{x} - \mu)$  for  $\mathbf{x} - \mu = (x_1 - \mu, \dots, x_{n_1} - \mu)$ , and let  $\mathbf{Y} = (Y_1, \dots, Y_{n_2})$  be a random variable having a density  $g(\mathbf{y} - \mu)$  for  $\mathbf{y} - \mu = (y_1 - \mu, \dots, y_{n_2} - \mu)$ , where the location parameter is restricted to the one-sided parameter space

$$A = \{\mu \mid \mu \geq a_0\} \quad \text{for known } a_0.$$

Let  $u_x = (x_2 - x_1, \dots, x_{n_1} - x_1)$  and  $u_y = (y_2 - y_1, \dots, y_{n_2} - y_1)$  be the maximal invariants. The location models are expressed as  $p(x_1 - \mu | u_x) = f(x_1 - \mu, u_x + x_1 - \mu) / p_x(u_x)$  and  $q(y_1 - \mu | u_y) = g(y_1 - \mu, u_y + y_1 - \mu) / q_y(u_y)$  for  $p_x(u_x) = \int f(t, u_x + t) dt$  and  $q_y(u_y) = \int g(t, u_y + t) dt$ , where  $u_x + a$  means  $u_x + a = (x_2 - x_1 + a, \dots, x_{n_1} - x_1 + a)$  for a scalar  $a$ .

When the parameter  $\mu$  is not restricted, it follows from (5.2.3) that the best equivariant estimator for predicting the density  $q(y_1 - \mu | u_y) q_y(u_y)$  is  $\hat{q}^{BI}(\mathbf{y} | \mathbf{x}) = \hat{q}^{BI}(y_1 - x_1 | u_y, u_x) q_y(u_y)$ , where

$$\hat{q}^{BI}(y_1 - x_1 | u_y, u_x) = \frac{\int_{-\infty}^{\infty} p(x_1 - a | u_x) q(y_1 - a | u_y) da}{\int_{-\infty}^{\infty} p(x_1 - a | u_x) da}, \quad (5.3.1)$$

which is minimax without the restriction  $A$ . When  $\mu$  is restricted to  $A$ , we can show the minimaxity of  $\hat{q}^{BI}(\mathbf{y} | \mathbf{x})$ .

**Théoreme 5.3.1.** *The best equivariant estimator  $\hat{q}^{BI}(\mathbf{y} | \mathbf{x})$  in the location problem is minimax for estimation of the predictive density under the restricted parameter space  $A$  relative to  $L_{KL}$ -loss; and the minimax risk is given by  $R_0 = R(\mu, \hat{q}^{BI})$ .*

**Proof.** It is sufficient to check conditions (A6)-(A6-3) in Theorem 5.2.2. In this case,  $P = \{\mu \geq a_0\}$ ,  $\mathcal{G} = \mathbb{R}$ ,  $\gamma(d\mu) = \nu(d\mu) = d\mu$ ,  $P_k = \{\mu | a_0 < \mu < a_0 + k\}$  and  $V(P_k) = k$ .

Take  $\xi_k = h_k(\mu) = (2/k)(\mu - a_0) - 1$ . Then,  $h_k(P_k) = [-1, 1]$ ,  $\gamma_k(d\xi_k) = (k/2)d\xi_k$  and  $\int_{h_k(P_k)} f(\xi_k)\gamma_k(d\xi_k)/V(P_k) = (1/2) \int_{[-1,1]} f(\xi)d\xi$ , which satisfies condition **(A6-2)**. For any  $\xi \in [-1 + \varepsilon, 1 - \varepsilon]$ , it is noted that  $\mu = h_k^{-1}(\xi) = a_0 + (k/2)(\xi + 1)$ , so that  $\{[h_k^{-1}(\xi)]^{-1}g; g \in P_k\} = \{\mu - a_0 - (k/2)(\xi + 1); a_0 < \mu < a_0 + k\} = (-(k/2)(\xi + 1), (k/2)(1 - \xi)) \supset (-k\varepsilon, k\varepsilon) \equiv P_k^*$ . Since  $\lim_{k \rightarrow \infty} P_k^* = \mathbb{R}$ , condition **(A6-3)** is satisfied, and the minimaxity of  $\hat{q}^{BI}$  is established.  $\square$

### 5.3.2 Improvements on the best equivariant estimator $\hat{q}^{BI}$

Although the best equivariant predictive density is minimax, it is not reasonable from a Bayesian or optimization perspective because the prior distribution is taken over whole the space of  $\mu$ . This suggests that the unrestricted uniform prior Bayes predictive density is likely to be inadmissible and may be improved upon by other (necessarily minimax) predictive densities. A reasonable alternative is the generalized Bayes predictive density against the uniform prior over the restricted space  $A$ , given by  $\hat{q}^U(\mathbf{y}|\mathbf{x}) = \hat{q}^U(y_1, |x_1, u_y, u_x)q_y(u_y)$ , where

$$\hat{q}^U(y_1|x_1, u_y, u_x) = \frac{\int_{a_0}^{\infty} p(x_1 - a|u_x)q(y_1 - a|u_y)da}{\int_{a_0}^{\infty} p(x_1 - a|u_x)da}. \quad (5.3.2)$$

We will indeed establish the minimaxity of the uniform prior Bayes predictive density  $\hat{q}^U(\mathbf{y}|\mathbf{x})$  under the following logconcavity or increasing monotone likelihood ratio property :

**(C1)** The density  $q(y_1 - \mu|u_y)$  is a continuously differentiable function such that  $q(y_1 - \mu|u_y)/q(y_1 - a_0|u_y)$  is nondecreasing in  $y_1$  for  $\mu > a_0$ .

**Lemme 5.3.2.** *Assume that  $q(y_1 - \mu|u_y)$  satisfies condition **(C1)**. Define  $A(y_1|x_1, u_x, u_y, \mu)$  by*

$$A(y_1|x_1, u_x, u_y, \mu) = \frac{\int_{-\infty}^0 p(x_1 + w - \mu|u_x)q(y_1 + w - \mu|u_y)dw}{\int_{-\infty}^0 p(x_1 + w|u_x)q(y_1 + w|u_y)dw}. \quad (5.3.3)$$

*Then, the following properties hold :*

(i)  $q'(y_1|u_y)/q(y_1|u_y)$  is nonincreasing in  $y_1$ , where  $q'(y_1|u_y) = \nabla_{y_1} q(y_1|u_y)$  for  $\nabla_{y_1} = \partial/\partial y_1$ ;

(ii) For  $\mu > 0$ ,  $A(y_1|x_1, u_x, u_y, \mu)$  is nondecreasing in  $y_1$ .

**Proof.** Property (i) follows from the fact that  $\nabla_{y_1} \{q(y_1 - \mu|u_y)/q(y_1|u_y)\} \geq 0$ . For establishing (ii), we shall show that  $\nabla_{y_1} A(y_1|x_1, u_x, u_y, \mu) \geq 0$  under assumption (C1).

Carrying out the differentiation, we see that this inequality is equivalent to

$$\begin{aligned} & \frac{\int_{-\infty}^0 p(x_1 + w - \mu|u_x)q'(y_1 + w - \mu|u_y)dw}{\int_{-\infty}^0 p(x_1 + w - \mu|u_x)q(y_1 + w - \mu|u_y)dw} \\ & \geq \frac{\int_{-\infty}^0 p(x_1 + w|u_x)q'(y_1 + w|u_y)dw}{\int_{-\infty}^0 p(x_1 + w|u_x)q(y_1 + w|u_y)dw}, \end{aligned}$$

or

$$\begin{aligned} & \frac{\int_{-\infty}^{-\mu} p(x_1 + w|u_x)q'(y_1 + w|u_y)dw}{\int_{-\infty}^{-\mu} p(x_1 + w|u_x)q(y_1 + w|u_y)dw} \\ & \geq \frac{\int_{-\infty}^0 p(x_1 + w|u_x)q'(y_1 + w|u_y)dw}{\int_{-\infty}^0 p(x_1 + w|u_x)q(y_1 + w|u_y)dw}. \end{aligned} \quad (5.3.4)$$

Hence from (5.3.4), it is sufficient to show that

$$\frac{\partial \int_{-\infty}^{-\mu} p(x_1 + w|u_x)q'(y_1 + w|u_y)dw}{\partial \mu \int_{-\infty}^{-\mu} p(x_1 + w|u_x)q(y_1 + w|u_y)dw} \geq 0. \quad (5.3.5)$$

In fact, this derivative is proportional to

$$\begin{aligned} & -p(x_1 - \mu|u_x)q'(y_1 - \mu|u_y) \int_{-\infty}^{-\mu} p(x_1 + w|u_x)q(y_1 + w|u_y)dw \\ & + p(x_1 - \mu|u_x)q(y_1 - \mu|u_y) \int_{-\infty}^{-\mu} p(x_1 + w|u_x)q'(y_1 + w|u_y)dw, \end{aligned}$$

which is rewritten as

$$\begin{aligned} & p(x_1 - \mu|u_x)q(y_1 - \mu|u_y) \int_{-\infty}^{-\mu} p(x_1 + w|u_x)q(y_1 + w|u_y) \\ & \times \left\{ \frac{q'(y_1 + w|u_y)}{q(y_1 + w|u_y)} - \frac{q'(y_1 - \mu|u_y)}{q(y_1 - \mu|u_y)} \right\} dw. \end{aligned} \quad (5.3.6)$$

From property (i), note that  $\nabla_{y_1} q(y_1|u_y)/q(y_1|u_y)$  is nonincreasing in  $y_1$ . Hence, the integrand in (5.3.6) is not negative, and the inequality (5.3.5) holds. This proves Lemma 5.3.2.  $\square$



Using this lemma, we prove the following theorem.

**Théoreme 5.3.3.** *Assume condition (C1). Then, the uniform prior Bayes predictive density  $\hat{q}^U(\mathbf{y}|\mathbf{x})$  is minimax under the restriction  $\mu \geq a_0$ . The risks of  $\hat{q}^U(\cdot)$  and  $\hat{q}^{BI}(\cdot)$  coincide if and only if  $\mu = a_0$ .*

**Proof.** Let  $a_0 = 0$  without any loss generality. Since  $\hat{q}^{BI}(\mathbf{y}|\mathbf{x})$  is a minimax estimator with a constant risk, we shall show that  $\hat{q}^U(\mathbf{y}|\mathbf{x})$  improves on  $\hat{q}^{BI}(\mathbf{y}|\mathbf{x})$ . From (5.2.1), it is sufficient to show the improvement in terms of the conditional risk (5.2.2). The IERD method developed by Kubokawa (1994A,B) is useful for the purpose. The conditional risk difference of the two predictive densities  $\hat{q}^{BI}(\mathbf{y}|\mathbf{x})$  and  $\hat{q}^U(\mathbf{y}|\mathbf{x})$  is written as

$$\begin{aligned}\Delta(\mu) &= R_{KL}(\mu, \hat{q}^{BI}|u_x, u_y) - R_{KL}(\mu, \hat{q}^U|u_x, u_y) \\ &= \int \int p(x_1 - \mu|u_x)q(y_1 - \mu|u_y) \left\{ \log \hat{q}^U(\mathbf{y}|\mathbf{x}) - \log \hat{q}^{BI}(\mathbf{y}|\mathbf{x}) \right\} dx_1 dy_1.\end{aligned}$$

Observe that

$$\begin{aligned}\log \hat{q}^U(\mathbf{y}|\mathbf{x}) - \log \hat{q}^{BI}(\mathbf{y}|\mathbf{x}) &= \log \frac{\int_0^\infty q(y_1 - a|u_y)p(x_1 - a|u_x)da}{\int_0^\infty p(x_1 - a|u_x)da} - \log \frac{\int_{-\infty}^\infty q(y_1 - a|u_y)p(x_1 - a|u_x)da}{\int_{-\infty}^\infty p(x_1 - a|u_x)da} \\ &= \int_{-\infty}^0 \frac{d}{dt} \left( \log \frac{\int_t^\infty q(y_1 - a|u_y)p(x_1 - a|u_x)da}{\int_t^\infty p(x_1 - a|u_x)da} \right) dt \\ &= \int_{-\infty}^0 \left\{ \frac{p(x_1 - t|u_x)}{\int_t^\infty p(x_1 - a|u_x)da} - \frac{q(y_1 - t|u_y)p(x_1 - t|u_x)}{\int_t^\infty q(y_1 - a|u_y)p(x_1 - a|u_x)da} \right\} dt.\end{aligned}$$

which permits us to write

$$\begin{aligned}\Delta(\mu) &= \int \int p(x_1 - \mu|u_x)q(y_1 - \mu|u_y) dx_1 dy_1 \\ &\quad \times \int_{-\infty}^0 \left\{ \frac{p(x_1 - t|u_x)}{\int_t^\infty p(x_1 - a|u_x)da} - \frac{q(y_1 - t|u_y)p(x_1 - t|u_x)}{\int_t^\infty q(y_1 - a|u_y)p(x_1 - a|u_x)da} \right\} dt.\end{aligned}$$

Making the transformation  $w = -a + t$  with  $dw = -da$  gives that  $\int_t^\infty p(x_1 - a|u_x)da = \int_{-\infty}^0 p(x_1 - t + w|u_x)dw$  and  $\int_t^\infty q(y_1 - a|u_y)p(x_1 - a|u_x)da = \int_{-\infty}^0 q(y_1 - t + w|u_y)p(x_1 - t + w|u_x)dw$ . Then, making the transformations  $x = x_1 - t$  and  $y = y_1 - t$  yields,

$$\begin{aligned}\Delta(\mu) &= \int \int \int_{-\infty}^0 p(x + t - \mu|u_x)q(y + t - \mu|u_y) dt \\ &\quad \times \left\{ \frac{p(x|u_x)}{\int_{-\infty}^0 p(x + w|u_x)dw} - \frac{q(y|u_y)p(x|u_x)}{\int_{-\infty}^0 q(y + w|u_y)p(x + w|u_x)dw} \right\} dx dy.\end{aligned}$$

Replacing  $t$  with  $w$ , we can get the expression

$$\begin{aligned}
\Delta(\mu) &= \iint p(x|u_x) \frac{\int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)dw}{\int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)dw} \\
&\quad \times \left\{ \frac{\int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)dw}{\int_{-\infty}^0 p(x+w|u_x)dw} - q(y|u_y) \right\} dx dy \\
&= \iint p(x|u_x) \int A(y|x, u_x, u_y, \mu) \\
&\quad \times \left\{ \frac{\int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)dw}{\int_{-\infty}^0 p(x+w|u_x)dwq(y|u_y)} - 1 \right\} q(y|u_y) dy dx.
\end{aligned}$$

Let  $B(y|x, u_x, u_y) = \int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)dw / \{ \int_{-\infty}^0 p(x+w|u_x)dwq(y|u_y) \} - 1$ . Denote an expectation with the density  $q(y|u_y)$  by  $E_q[\cdot]$ . From Lemma 5.3.2, it follows that  $A(y|x, u_x, u_y, \mu)$  is nondecreasing in  $y$  for  $\mu > 0$ . Since  $q(y+w|u_y)/q(y|u_y)$  is nondecreasing in  $y$ , it is seen that  $B(y|x, u_x, u_y)$  is nondecreasing in  $y$ . Thus, for  $\mu > 0$

$$\begin{aligned}
&\int A(y|x, u_x, u_y, \mu) \left\{ \frac{\int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)dw}{\int_{-\infty}^0 p(x+w|u_x)dwq(y|u_y)} - 1 \right\} q(y|u_y) dy \\
&= E_q[A(Y|x, u_x, u_y, \mu)B(Y|x, u_x, u_y)] \\
&\geq E_q[A(Y|x, u_x, u_y, \mu)]E_q[B(Y|x, u_x, u_y)] \tag{5.3.7} \\
&= E_q[A(y|x, u_x, u_y, \mu)] \left\{ \frac{\int_{-\infty}^0 p(x+w|u_x) \int q(y+w|u_y)dydw}{\int_{-\infty}^0 p(x+w|u_x)dw} - \int q(y|u_y)dy \right\},
\end{aligned}$$

where the inequality in (5.3.7) follows from the well known covariance inequality since both functions  $A(y|x, u_x, u_y, \mu)$  and  $B(y|x, u_x, u_y)$  are nondecreasing in  $y$  (see Wijsman (1984) for example). Since  $\int q(y+w|u_y)dy = \int q(y|u_y)dy = 1$ , it follows that

$$\frac{\int_{-\infty}^0 p(x+w|u_x) \int q(y+w|u_y)dydw}{\int_{-\infty}^0 p(x+w|u_x)dw} - 1 = 0,$$

showing that  $\Delta(\mu) \geq 0$  for all  $\mu \geq 0$ . Observe that  $A(y|x, u_x, u_y, 0)$  is constant(= 1) in  $y$ , so that  $\Delta(0) = 0$  as seen with the above expansion with an equality replacing the inequality in (5.3.7). Finally, the covariance inequality in (5.3.7) is strict when  $\mu > 0$  and the proof of Theorem 5.3.3 is therefore complete.  $\square$

## Other improvements on $\hat{q}^{BI}$

Theorem 5.3.3 establishes a general comparison between the generalized Bayes estimator  $\hat{q}^U$  and the best equivariant estimator  $\hat{q}^{BI}$ , with the former dominating the latter under the simple condition that  $q$  be logconcave. It is of interest to seek classes of other dominating procedures. Although we will not explore this issue in depth here, it is nevertheless pertinent to make the following observation which generates many other dominating procedures. The next result follows from the strict concavity of the log function on  $(0, \infty)$ , or alternatively from the strict convexity with respect to  $\hat{q}$  of the loss  $L_{KL}(q_\theta, \hat{q})$ .

**Lemme 5.3.4.** *Let  $\alpha \in (0, 1)$ . Let  $\hat{q}_i, i = 0, 1, 2$  be estimators such that  $\hat{q}_1 \neq \hat{q}_2$ . If  $R_{KL}(\theta, \hat{q}_i) \leq R_{KL}(\theta, \hat{q}_0)$  for  $i = 1, 2$  and for all  $\theta \in \Theta$ , then  $R_{KL}(\theta, \alpha\hat{q}_1 + (1 - \alpha)\hat{q}_2) \leq R_{KL}(\theta, \hat{q}_0)$ , with equality at a given  $\theta_0$  if and only if  $R_{KL}(\theta_0, \hat{q}_i) = R_{KL}(\theta_0, \hat{q}_0)$  for  $i = 1, 2$ .*

The above result implies directly that convex linear combinations of  $\hat{q}_{BI}$  and  $\hat{q}_U$  dominate  $\hat{q}_{BI}$  in the context of Theorem 5.3.3 by taking  $\hat{q}_0 = \hat{q}_1 = \hat{q}^{BI}$  and  $\hat{q}_2 = \hat{q}^U$ . Finally, since Theorem 5.3.3 applies for the conditional risks, the weights can be made to depend on the maximal invariants  $u_x$  and  $u_y$  and it thus follows that estimators  $\alpha(u_x, u_y)\hat{q}^U(y|x_1, u_y, u_x)q_y(u_y) + (1 - \alpha(u_x, u_y))\hat{q}^{BI}(y|x_1, u_y, u_x)q_y(u_y)$  with  $\alpha(\cdot, \cdot) \in (0, 1)$  are also minimax.

## Examples

We proceed with instructive examples and illustrations.

**Exemple 5.3.5.** *(normal models) The results above apply to the particular setup :*

$$X|\mu \sim \mathcal{N}(\mu, \sigma_X^2), Y|\mu \sim \mathcal{N}(\mu, \sigma_Y^2). \quad (5.3.8)$$

*with the restriction  $\mu \geq a_0$ . Namely, Theorem 5.3.1 tells us that  $\hat{q}^{BI}(\cdot|X) \sim \mathcal{N}(X, \sigma_X^2 + \sigma_Y^2)$  remains minimax under the restriction  $\mu \geq a_0$ , while Theorem 5.3.3 implies that the generalized Bayes estimator  $\hat{q}^U$  is also minimax, and dominates  $\hat{q}^{BI}$  under the restriction  $\mu \geq a_0$ . Figure 1 compares the risks of these two estimators for  $a_0 = 0, \sigma_X^2 = 1, \sigma_Y^2 = 1$ .*

The curve measures the relative difference in risks (i.e.,  $\frac{R_{\kappa I}(\mu, \hat{q}^{BI}) - R_{\kappa I}(\mu, \hat{q}^U)}{R_{\kappa I}(\mu, \hat{q}^{BI})}$ ). Observe that the risks coincide indeed at the lower boundary of the parameter space and at  $\mu = \infty$  and that the gains are appreciable, particularly around one standard deviation from the boundary where they fluctuate around 40%.

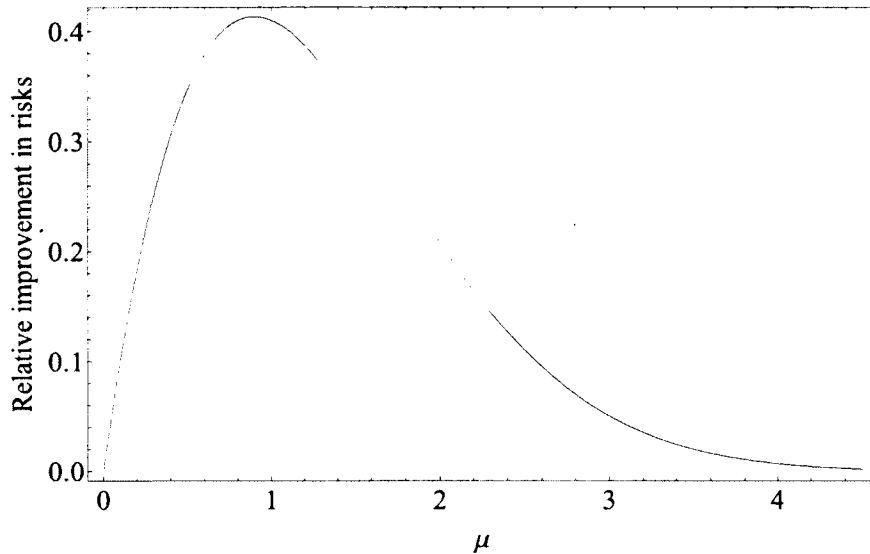


Figure 5.1 – Relative difference in risks between  $\hat{q}^{BI}$  and  $\hat{q}^U$  (normal model with  $\mu \geq 0$ ,  $\sigma_X^2 = \sigma_Y^2 = 1$ )

For the specific normal case illustrated here, the above dominance and minimax results are not new and were previously obtained through a different route by Fourdrinier et al. (2011) by methods which are also applicable for the multivariate case. Interestingly, yet another proof of the dominance result can be derived by a more direct and instructive approach. We now expand on this, considering the more general problem  $\mu \in [a_0, a_0 + m)$ , with  $m = \infty$  corresponding to the lower bounded case and setting hereafter  $a_0 = 0$  without loss of generality. Making use of (5.1.5), the uniform Bayes estimator  $\hat{q}^U$  with respect to the flat prior on  $[0, m)$  is given by

$$\begin{aligned} \hat{q}^U(Y|X) &= \frac{m_U(W; \sigma_W^2)}{m_U(X; \sigma_X^2)} \hat{q}^{BI}(Y|X) \\ &= \left\{ \frac{\Phi\left(\frac{W}{\sigma_W}\right) - \Phi\left(\frac{W-m}{\sigma_W}\right)}{\Phi\left(\frac{X}{\sigma_X}\right) - \Phi\left(\frac{X-m}{\sigma_X}\right)} \right\} \hat{q}^{BI}(Y|X), \end{aligned}$$

with  $W = {}^d (\sigma_Y^2 X + \sigma_X^2 Y) / (\sigma_X^2 + \sigma_Y^2) \sim \mathcal{N}(\mu, \sigma_W^2)$  for  $\sigma_W^2 = (\sigma_X^2 \sigma_Y^2) / (\sigma_X^2 + \sigma_Y^2)$ . Conse-

quently, the difference in risks may be expressed as

$$\begin{aligned}\Delta(\mu) &= R_{KL}(\mu, \hat{q}^{BI}) - R_{KL}(\mu, \hat{q}^U) = E^{X,Y} \log \left( \frac{\hat{q}^U(Y|X)}{\hat{q}^{BI}(Y|X)} \right) \\ &= E^{X,Y} \left\{ \log \left( \Phi\left(\frac{W}{\sigma_W}\right) - \Phi\left(\frac{W-m}{\sigma_W}\right) \right) - \log \left( \Phi\left(\frac{X}{\sigma_X}\right) - \Phi\left(\frac{X-m}{\sigma_X}\right) \right) \right\}.\end{aligned}$$

Here, set  $W' = W/\sigma_W \sim \mathcal{N}(\mu/\sigma_W, 1)$ ,  $X' = X/\sigma_X \sim \mathcal{N}(\mu/\sigma_X, 1)$  and observe that  $W' \stackrel{d}{=} X' + \delta$ , with  $\delta = \mu(1/\sigma_W - 1/\sigma_X) \geq 0$  for  $\mu \geq 0$  with equality iff  $\mu = 0$ , given that  $\sigma_W < \sigma_X$ . Hence,

$$\Delta(\mu) = E^{X'} \left\{ \log \left( \Phi(X' + \delta) - \Phi(X' + \delta - \frac{m}{\sigma_W}) \right) - \log \left( \Phi(X') - \Phi(X' - \frac{m}{\sigma_X}) \right) \right\} \geq 0,$$

for all  $\mu \in [0, m]$ , since  $\Phi(\cdot)$  is strictly increasing on  $\mathbb{R}$  and  $x' + \delta \geq x'$  and  $x' + \delta - m/\sigma_W \leq x' - m/\sigma_X$  for all  $x' \in \mathbb{R}$ , and with equality occurring only if  $\mu = 0$  and  $m = \infty$ . We have thus shown directly that the uniform Bayes procedure  $\hat{q}^U$  dominates  $\hat{q}^{BI}$  for the normal model in (5.3.8) with the restriction  $\mu \in [a_0, a_0 + m)$ . This offers an alternative to Fourdrinier et al.'s proof. Notwithstanding this development (as well as the next Remark), the search for efficient Bayesian procedures under a compact interval constraint which merits further study will not be pursued here. Recent advances for point estimation versions of this problem were obtained by Kubokawa (2005B), as well as Marchand and Payandeh (2011).

**Remarque 5.3.6.** (non-minimality of  $\hat{q}^{BI}$  in the compact interval case)

In the previous example for the compact interval case with  $m < \infty$ , observe that  $\Delta(\mu) > 0$  for all  $\mu \in [a_0, a_0 + m]$ , which implies in turn that  $\inf_{\mu \in [a_0, a_0 + m]} \Delta(\mu) > 0$  and that  $\hat{q}^{BI}$  is not minimax, in contrast to the unbounded lower bounded case. This provides an analog of a familiar point estimation version of this argument (e.g., Lehmann and Casella, 1998, page 327). Moreover, the non-minimality argument is more general under condition **(C1)** in the context of Theorem 5.3.3 as seen by the following elements of proof :

- Theorem 5.3.3 implies that  $\hat{q}^{U1}$  dominates  $\hat{q}^{BI}$  for the restriction  $\mu \in [a_0, a_0 + m]$  where  $\hat{q}^{U1}$  is the generalized Bayes predictive density with respect to the flat prior on  $[a_0, \infty)$  with equality in risks iff  $\mu = a_0$  ;
- Theorem 5.3.3 implies that  $\hat{q}^{U2}$  dominates  $\hat{q}^{BI}$  for the restriction  $\mu \in [a_0, a_0 + m]$  where  $\hat{q}^{U2}$  is the generalized Bayes predictive density with respect to the flat prior on  $[-\infty, a_0 + m]$  with equality in risks iff  $\mu = a_0 + m$  ;

- Paired with the above, Lemma 5.3.4 implies that the predictive density estimator  $\frac{1}{2}\hat{q}^{U1} + \frac{1}{2}\hat{q}^{U2}$  dominates  $\hat{q}^{BI}$  strictly for  $\mu \in [a_0, a_0 + m]$ ;
- Consequently, as in the first paragraph of this Remark,  $\hat{q}^{BI}$  cannot be minimax for  $\mu \in [a_0, a_0 + m]$  when  $q$  satisfies condition **(C1)**.

**Example 5.3.7.** The results of this section also apply to Exponential location models with  $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$  i.i.d.  $Exp(\mu, \sigma)$ ,  $\mu \geq 0$  and known  $\sigma$ , with density  $\sigma^{-1} \exp\{-(t - \mu)/\sigma\} \mathbf{1}_{(\mu, \infty)}(t)$ . Here the order statistics  $X_{(1)}$  and  $Y_{(1)}$  form a sufficient statistic, and we can take  $\sigma = 1$  without loss of generality, so that it suffices to consider the setup

$$X \sim Exp(\mu, n_1^{-1}), Y \sim Exp(\mu, n_2^{-1}). \quad (5.3.9)$$

Evaluating (5.3.1) and (5.3.2), we obtain with a little bit of manipulation

$$\hat{q}^{BI}(y|x) = \frac{n_1 n_2}{n_1 + n_2} \left\{ e^{-n_2|x-y|} \mathbf{1}_{[x, \infty)}(y) + e^{-n_1|x-y|} \mathbf{1}_{(-\infty, x)}(y) \right\},$$

and

$$\hat{q}^U(y|x) = \hat{q}^{BI}(y|x) \left\{ \frac{e^{(n_1+n_2)x} - 1}{e^{(n_1+n_2)x} - e^{n_2x}} \mathbf{1}_{[x, \infty)}(y) + \frac{1 - e^{-(n_1+n_2)y}}{1 - e^{-(n_1)y}} \mathbf{1}_{(0, x)}(y) \right\},$$

Observe that  $\hat{q}^{BI}$  is an asymmetric Laplace distribution (and symmetric Laplace for  $n_1 = n_2$ ), while  $\hat{q}^U$  is a skewed version of  $\hat{q}^{BI}$ . Theorems 5.3.1 and 5.3.3 apply and tell us that both  $\hat{q}^{BI}$  and  $\hat{q}^U$  are minimax under the restriction  $\mu \geq 0$ , with  $\hat{q}^U$  dominating  $\hat{q}^{BI}$ .

### 5.3.3 Case of a scale family

We next consider estimation of the predictive density with a restricted scale parameter. Let  $\mathbf{X} = (X_1, \dots, X_{n_1})$  be a positive random variable having a density  $\sigma^{-n_1} f(\sigma^{-1} \mathbf{x})$  for  $\sigma^{-1} \mathbf{x} = (\sigma^{-1} x_1, \dots, \sigma^{-1} x_{n_1})$ , and let  $\mathbf{Y} = (Y_1, \dots, Y_{n_2})$  be a random variable having a density  $\sigma^{-n_2} g(\sigma^{-1} \mathbf{y})$  for  $\sigma^{-1} \mathbf{y} = (\sigma^{-1} y_1, \dots, \sigma^{-1} y_{n_2})$ , where the scale parameter is lower bounded belonging to the restricted parameter space

$$B = \{\sigma \mid \sigma \geq b_0\}, \quad \text{for known positive } b_0.$$

Let  $t_x = |x_1|$ ,  $u_x = (x_1/|x_1|, x_2/|x_1|, \dots, x_{n_1}/|x_1|)$  and  $t_y$  and  $u_y$  are defined similarly. The joint densities  $\sigma^{-n_1} f(\sigma^{-1} \mathbf{x}) d\mathbf{x}$  and  $\sigma^{-n_2} g(\sigma^{-1} \mathbf{y}) d\mathbf{y}$  are expressed as, respectively,  $p(\sigma^{-1} t_x | u_x) p_x(u_x) \gamma(dt_x) \gamma_x(du_x)$  and  $q(\sigma^{-1} t_y | u_y) q_y(u_y) \gamma(dt_y) \gamma_y(du_y)$ , where  $\gamma(d\sigma) = d\sigma/\sigma$ , and  $p_x(u_x)$  and  $q_y(u_y)$  are marginal densities of  $u_x$  and  $u_y$ .

Note that  $\sigma^{-1} t_x = \exp\{\log t_x - \log \sigma\}$  and  $d \log t_x = dt_x/t_x$ . Since the restriction  $B$  is written as  $\log \sigma > \log b_0$ , all the results given in the previous subsection hold for the restricted scale problem. The results corresponding to Theorems 5.3.1 and 5.3.3 are described below.

When the parameter  $\sigma$  is not restricted, it follows from (5.2.3) that the best equivariant estimator for predicting the density  $q(\sigma^{-1} t_y | u_y) q_y(u_y)$  is  $\hat{q}^{BI}(\mathbf{y} | \mathbf{x}) = \hat{q}^{BI}(t_x^{-1} t_y | u_y, u_x) q_y(u_y)$ , where

$$\hat{q}^{BI}(t_x^{-1} t_y | u_y, u_x) = \frac{\int_0^\infty p(b^{-1} t_x | u_x) q(b^{-1} t_y | u_y) b^{-1} db}{\int_0^\infty p(b^{-1} t_x | u_x) b^{-1} db}, \quad (5.3.10)$$

which is minimax without the restriction  $B$ . Even if  $\sigma$  is restricted on  $B$ , the minimaxity of  $\hat{q}^{BI}(\mathbf{y} | \mathbf{x})$  still holds.

**Théoreme 5.3.8.** *The best equivariant estimator  $\hat{q}^{BI}(\mathbf{y} | \mathbf{x})$  is minimax for estimation of the predictive density under the restricted parameter space  $B$  relative to the  $L_{KL}$ -loss, and the minimax risk is given by  $R_0 = R(\sigma, \hat{q}^{BI})$ .*

Although the best equivariant predictive density is minimax, it is not reasonable because the prior distribution is taken over whole the space of  $\sigma$ . This suggests that  $\hat{q}^{BI}$  is likely to be inadmissible and to be improved upon by other (minimax) predictive densities. A reasonable choice is the generalized Bayes predictive density against the invariant prior over the restricted space  $B$ , given by  $\hat{q}^U(\mathbf{y} | \mathbf{x}) = \hat{q}^U(t_y | t_x, u_y, u_x) q_y(u_y)$ , where

$$\hat{q}^U(t_y | t_x, u_y, u_x) = \frac{\int_{b_0}^\infty p(b^{-1} t_x | u_x) q(b^{-1} t_y | u_y) b^{-1} db}{\int_{b_0}^\infty p(b^{-1} t_x | u_x) b^{-1} db}. \quad (5.3.11)$$

To establish the minimaxity of the invariant prior Bayes predictive density  $\hat{q}^U(\mathbf{y} | \mathbf{x})$ , we assume the following condition analogous to (C1) :

(C2) The density  $q(\sigma^{-1} t_y | u_y)$  is a continuously differentiable function such that the ratio of the densities  $q(\sigma^{-1} t_y | u_y) / q(b_0^{-1} t_y | u_y)$  is nondecreasing in  $t_y$  for  $\sigma > b_0$ .

**Théoreme 5.3.9.** *Assume condition (C2). Then, the Bayes predictive density  $\hat{q}^U(\mathbf{y}|\mathbf{x})$  is minimax under the restriction  $\sigma \geq b_0$ , and the risks of  $\hat{q}^U$  and  $\hat{q}^{BI}$  coincide if and only if  $\sigma = b_0$ .*

Lemma 5.3.2 used for proving Theorem 5.3.3 is expressed in the scale case as follows :

**Lemme 5.3.10.** *Assume that  $q(\sigma^{-1}t_y|u_y)$  satisfies the condition (C2). Then, the following properties hold :*

(i)  $t_y\{\nabla_{t_y}q(t_y|u_y)\}/q(t_y|u_y)$  is nonincreasing in  $t_y$ , where  $\nabla_{t_y} = \partial/\partial t_y$ .

(ii) Define  $B(t_y|t_x, u_x, u_y, \sigma)$  by

$$B(t_y|t_x, u_x, u_y, \sigma) = \frac{\int_0^1 w^{-1}p(\sigma^{-1}wt_x|u_x)q(\sigma^{-1}wt_y|u_y)dw}{\int_0^1 w^{-1}p(wt_x|u_x)q(wt_y|u_y)dw}. \quad (5.3.12)$$

Then for  $\sigma > b_0$ ,  $B(t_y|t_x, u_x, u_y, \mu)$  is nondecreasing in  $t_y$ .

We can show Theorem 5.3.9 directly using Lemma 5.3.10, though we have here applied Theorem 5.3.3 to the scale case. We conclude this section with an application to Gamma models.

**Exemple 5.3.11.** *An interesting application consists of Gamma distributions for  $X$  and  $Y$  with*

$$X|\sigma \sim \text{Gamma}(\alpha_1, \sigma), Y|\sigma \sim \text{Gamma}(\alpha_2, \sigma), \quad (5.3.13)$$

with  $\alpha_1, \alpha_2$  known, and the lower bound restriction  $\sigma \geq b_0 (> 0)$ . We have assumed without loss of generality that the samples for  $X$  and  $Y$  are of size one by sufficiency of the sums in such Gamma models. Evaluating (5.3.10) and (5.3.11), we obtain the elegant representations

$$\hat{q}^{BI}(y|x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{1}{x} \left(\frac{y}{x}\right)^{\alpha_2-1} \left(1 + \frac{y}{x}\right)^{-(\alpha_1+\alpha_2)} \mathbf{1}_{(0,\infty)}(y),$$

and

$$\hat{q}^U(y|x) = \hat{q}^{BI}(y|x) \frac{\bar{F}_{\alpha_1+\alpha_2}\left(\frac{x+y}{b_0}\right)}{\bar{F}_{\alpha_1}\left(\frac{x}{b_0}\right)},$$



where  $\bar{F}_\cdot(\cdot)$  is the survival function of a  $\text{Gamma}(\gamma, 1)$  distribution. Observe that  $\hat{q}^{BI}$  is the density of a Fisher distribution with scale parameter  $\frac{\alpha_2}{\alpha_1}x$ , and shape parameters  $2\alpha_2$  (d.f. numerator) and  $2\alpha_1$  (d.f. denominator), while  $\hat{q}^U$  is a skewed version of  $\hat{q}^{BI}$ .

The findings of this section apply. First,  $\hat{q}^{BI}$  is minimax for the unrestricted parameter space and remains minimax in presence of the lower bound  $b_0$  on the scale parameter. Second, since Gamma densities form a family with an increasing monotone likelihood ratio, condition **(C2)** is satisfied and the Bayes procedure  $\hat{q}^U$  dominates  $\hat{q}^{BI}$  by virtue of Theorem 5.3.9. Finally, we point out that analogous results hold here for the case where the scale parameter  $\sigma$  is upper bounded, say  $\sigma \in (0, c_0)$ . In such cases, we consider the transformed problem with  $X' = X$  and  $Y' = \frac{1}{Y}$  and consider the setup of Theorem 5.3.9 with  $b_0 = \frac{1}{c_0}$ ,  $p_\theta$  being the density of  $X'$  and  $q_\theta$  being the density of  $Y'$ . Since inverse Gamma distributions have logconcave densities as well, and the Kullback-Leibler loss is intrinsic, Theorem 5.3.9 indeed applies.

## 5.4 Estimation in location-scale families

In this section, we treat location-scale families with location and/or scale parameters constrained, and investigate minimaxity of the best equivariant estimators using Theorem 5.2.2.

### 5.4.1 Non-bounded case

We begin with the univariate case. Let  $\mathbf{X} = (X_1, \dots, X_{n_1})$  be a random variable having a density  $\sigma^{-n_1} f((\mathbf{x} - \mu)/\sigma)$  for  $(\mathbf{x} - \mu)/\sigma = ((x_1 - \mu)/\sigma, \dots, (x_{n_1} - \mu)/\sigma)$ , and let  $\mathbf{Y} = (Y_1, \dots, Y_{n_2})$  be a random variable having a density  $\sigma^{-n_2} g((\mathbf{y} - \mu)/\sigma)$  for  $(\mathbf{y} - \mu)/\sigma = ((y_1 - \mu)/\sigma, \dots, (y_{n_2} - \mu)/\sigma)$ , where the location and scale parameters are restricted to the space

$$C = \{(\mu, \sigma) | \mu > c_0\sigma + a_0, \sigma > b_0\}, \quad (5.4.1)$$

where  $a_0, b_0$  and  $c_0$  are constants such that  $b_0 \geq 0$  and  $-\infty \leq a_0, c_0 < \infty$ . The unrestricted case is described by  $b_0 = c_0 = 0$  and  $a_0 = -\infty$ . Let  $t_x = (|x_2 - x_1|, x_1)$ ,  $u_x = ((x_2 - x_1)/|x_2 - x_1|, \dots, (x_{n_1} - x_1)/|x_2 - x_1|)$  and let  $t_y$  and  $u_y$  be defined similarly. Let  $\mathcal{G} = \mathbb{R}^+ \times \mathbb{R}$  and define the product by  $(a, b)(\sigma, \mu) = (a\sigma, a\mu + b)$ . This implies that  $(\sigma, \mu)^{-1} = (1/\sigma, -\mu/\sigma)$  and  $(\sigma, \mu)^{-1}(|x_2 - x_1|, x_1) = (|x_2 - x_1|/\sigma, (x_1 - \mu)/\sigma)$ . Then,  $\sigma^{-n_1} f((\mathbf{x} - \mu)/\sigma) d\mathbf{x}$  and  $\sigma^{-n_2} g((\mathbf{y} - \mu)/\sigma) d\mathbf{y}$  are expressed as  $p((\sigma, \mu)^{-1} t_x | u_x) p_x(u_x) \gamma(dt_x) \gamma_x(du_x)$  and  $q((\sigma, \mu)^{-1} t_y | u_y) q_y(u_y) \gamma(dt_y) \gamma_y(du_y)$ , respectively, where  $\gamma(d(\sigma, \mu)) = (d\mu d\sigma)/\sigma^2$ .

When the parameters are not restricted, it follows from (5.2.3) that the best equivariant predictive density estimator of  $q((\sigma, \mu)^{-1} t_x | u_x) q_y(u_y)$  is given by  $\hat{q}^{BI}(t_x^{-1} t_y, u_y | u_x) = \hat{q}^{BI}(t_x^{-1} t_y | u_y, u_x) q_y(u_y)$ , where

$$\hat{q}^{BI}(t_x^{-1} t_y | u_y, u_x) = \frac{\int p((b, a)^{-1} t_x | u_x) q((b, a)^{-1} t_y | u_y) \nu(d(b, a))}{\int p((b, a)^{-1} t_x | u_x) \nu(d(b, a))}, \quad (5.4.2)$$

and where  $\nu(d(b, a)) = (dad b)/b^2$ . Using Theorem 5.2.2, we analyze the question of minimaxity of the best equivariant estimator under the restriction  $C$ .

**[1] Case of  $a_0 > -\infty$  and  $b_0 > 0$ .** This case implies that both  $\mu$  and  $\sigma$  are restricted from one side.

**Théoreme 5.4.1.** *Assume that  $a_0$  and  $b_0$  satisfy that  $a_0 > -\infty$  and  $b_0 > 0$ . Then, the best equivariant estimator  $\hat{q}^{BI}(t_x^{-1} t_y, u_y | u_x)$  is minimax in the estimation of the predictive density under the restricted parameter space  $C$  relative to the  $L_{KL}$ -loss, and the minimax risk is given by  $R_0 = R((\sigma, \mu), \hat{q}^{BI})$ .*

**Proof.** For  $c_0 = 0$ , we define the sequence  $d_k = k$ , while for  $c_0 \neq 0$  we take  $d_k = \log k$ . Such a sequence admits the following behaviour when  $k \rightarrow \infty$ ,

(a)  $(k/d_k) d_k^{\varepsilon/2} \rightarrow \infty$  for any  $\varepsilon > 0$  when  $c_0 = 0$ ,

(b)  $d_k/k \rightarrow 0$  and  $d_k \rightarrow \infty$  when  $c_0 \neq 0$ .

We proceed by verifying conditions **(A6)-(A6-3)** in Theorem 5.2.2. In this case,  $P = \{(\sigma, \mu) | a_0 + c_0\sigma < \mu, b_0 < \sigma\}$ ,  $\mathcal{G} = \mathbb{R}^+ \times \mathbb{R}$ , we set  $P_k = \{(\sigma, \mu) | a_0 + c_0\sigma < \mu < a_0 + c_0\sigma + k, b_0 < \sigma < b_0 d_k\}$  and  $V(P_k) = k \log d_k$  where  $d_k$  is defined above. Take  $\xi_1 =$

$(2/\log d_k) \log(\sigma/b_0) - 1$  and  $\xi_2 = (2/k)(\mu - a_0 - c_0\sigma) - 1$ . Letting  $\boldsymbol{\xi} = (\xi_1, \xi_2) = h_k((\sigma, \mu))$ , we see that  $h_k(P_k) = [-1, 1]^2$ ,  $\gamma_k(d\boldsymbol{\xi}) = \{(k \log d_k)/4\}d\boldsymbol{\xi}$  and  $\int_{h_k(P_k)} f(\xi_k)\gamma_k(d\boldsymbol{\xi})/V(P_k) = (1/4) \int_{[-1, 1]^2} f(\boldsymbol{\xi})d\boldsymbol{\xi}$ , which satisfies condition **(A6-2)**. For any  $\boldsymbol{\xi} \in [-1 + \varepsilon, 1 - \varepsilon]^2$ , let  $(b, a) = h_k^{-1}(\boldsymbol{\xi})$ . Then,  $b = b_0 d_k^{(1+\xi_1)/2}$  and  $a = (k/2)(1 + \xi_2) + a_0 + c_0 b_0 d_k^{(1+\xi_1)/2}$  so that  $\{[h_k^{-1}(\boldsymbol{\xi})]^{-1}(\sigma, \mu); (\sigma, \mu) \in P_k\} = \{(\sigma/b, (\mu - a)/b); (\sigma, \mu) \in P_k\}$  and  $\sigma/b, (\mu - a)/b$  satisfy the inequalities

$$\begin{aligned} d_k^{-(1+\xi_1)/2} &< \frac{\sigma}{b} < d_k^{(1-\xi_1)/2}, \\ c_0 \frac{\sigma}{b} - \frac{d_k^{-(1+\xi_1)/2}}{b_0} \left\{ \frac{k}{2}(1 + \xi_2) + c_0 b_0 d_k^{(1+\xi_1)/2} \right\} \\ &< \frac{\mu - a}{b} < c_0 \frac{\sigma}{b} + \frac{d_k^{-(1+\xi_1)/2}}{b_0} \left\{ \frac{k}{2}(1 - \xi_2) - c_0 b_0 d_k^{(1+\xi_1)/2} \right\}. \end{aligned}$$

Note that  $1 - \xi_i > \varepsilon$  and  $1 + \xi_i > \varepsilon$  for  $i = 1, 2$ . The first inequality is satisfied by  $d_k^{-\varepsilon/2} < \sigma/b < d_k^{\varepsilon/2}$ , which can be expanded to  $(0, \infty)$  as  $k \rightarrow \infty$  if  $d_k \rightarrow \infty$  as  $k \rightarrow \infty$ . Also, the second inequality is satisfied by

$$\frac{k}{d_k^{(1+\xi_1)/2}} \left\{ c_0 \frac{\sigma}{b} \frac{d_k^{(1+\xi_1)/2}}{k} - \frac{\varepsilon}{2b_0} - \frac{c_0}{k} \right\} < \frac{\mu - a}{b} < \frac{k}{d_k^{(1+\xi_1)/2}} \left\{ c_0 \frac{\sigma}{b} \frac{d_k^{(1+\xi_1)/2}}{k} + \frac{\varepsilon}{2b_0} - \frac{c_0}{k} \right\}.$$

Here, it is noted that

$$\begin{aligned} \frac{\sigma}{b} \frac{d_k^{(1+\xi_1)/2}}{k} &< d_k^{(1-\xi_1)/2} \frac{d_k^{(1+\xi_1)/2}}{k} < \frac{d_k}{k}, \\ \text{and } \frac{k}{d_k^{(1+\xi_1)/2}} &= \frac{k}{d_k} d_k^{(1-\xi_1)/2} > \frac{k}{d_k} d_k^{\varepsilon/2}. \end{aligned}$$

Since  $d_k$  satisfies the condition (a) or (b), it can be seen that the lower end point of  $(\mu - a)/b$  goes to  $-\infty$ , and the upper point goes to  $\infty$ . This verifies condition **(A6-3)**, and the minimaxity of  $\hat{q}^{BI}$  is established.  $\square$

**[2] Case of  $a_0 = -\infty$  and  $b_0 > 0$ .** Although we can show the minimaxity directly by the same arguments as in the proof of Theorem 5.4.1, we here give a simple proof based on Theorem 5.3.8. Since  $\mu$  is not restricted and the problem is invariant under a location transformation, we can consider location equivariant estimators, which depend on  $x_1$  and  $y_1$  through  $y_1 - x_1$ . Thus, the risk function of the location equivariant estimator does not depend on  $\mu$ . Then, the problem can be reduced to the estimation in the scale family

with the restriction  $\sigma > b_0$ . Hence from Theorem 5.3.8, it follows that best equivariant estimator is minimax. This is summarized as follows.

**Théoreme 5.4.2.** *Assume that  $\mu$  is not restricted, but  $\sigma$  is restricted to  $\sigma > b_0$ . Then, the best equivariant estimator  $\hat{q}^{BI}(t_x^{-1}t_y, u_y|u_x)$  is minimax in the estimation of the predictive density under the restricted parameter space.*

**[3] Case of  $a_0 > -\infty$  and  $b_0 = 0$ .** This case implies that  $\mu$  is restricted as  $\mu > a_0$  and  $\sigma$  is not restricted. By considering  $x' = x - a_0$ , we can set  $a_0 = 0$  without loss of generality and the problem becomes invariant (as in the previous case) under a scale transformation. We are thus led to the following.

**Théoreme 5.4.3.** *Assume that  $\sigma$  is not restricted, but  $\mu$  is such that  $\mu \geq a_0$ . Then, the best equivariant estimator  $\hat{q}^{BI}(t_x^{-1}t_y, u_y|u_x)$  is minimax in the estimation of the predictive density under the restricted parameter space.*

## 5.4.2 Bounded case

Concerning the estimation of the predictive density, we have already seen that the best location equivariant estimator  $\hat{q}^{BI}$  (Example 5.3.5 and Remark 5.3.6) is generally not minimax for estimating a location parameter bounded to a compact interval. However, the result of Kubokawa (2005) suggests minimaxity in the case of an unknown scale, and the following theorem shows that this suggestion is correct.

Let us consider the following restriction under the same location-scale families as treated in the previous subsection :

$$D = \{(\sigma, \mu) | a_1 < \mu < a_2, 0 < \sigma < b_0\},$$

where  $a_1$  and  $a_2$  are bounded constants and  $b_0$  is a positive constant.

**Théoreme 5.4.4.** *Assume that  $(\mu, \sigma)$  is restricted to  $D$ . Then, the best equivariant estimator  $\hat{q}^{BI}(t_x^{-1}t_y, u_y|u_x)$  is minimax for the estimation of the predictive density under the restricted parameter space.*

**Proof.** We shall check conditions **(A6)**-**(A6-3)** in Theorem 5.2.2. In this case,  $P = \{(\sigma, \mu) | a_1 < \mu < a_2, 0 < \sigma < b_0\}$ ,  $P_k = \{(\sigma, \mu) | a_1 < \mu < a_2, b_0/k < \sigma < b_0\}$  for  $kb_0 > 1$ , and  $V(P_k) = (a_2 - a_1) \log k$ . Take  $\xi_1 = (2/\log k) \log(\sigma/b_0) + 1$  and  $\xi_2 = \{2/(a_2 - a_1)\} \{\mu - (a_1 + a_2)/2\}$ . Letting  $\boldsymbol{\xi} = (\xi_1, \xi_2) = h_k((\sigma, \mu))$ , we see that  $h_k(P_k) = [-1, 1]^2$ ,  $\gamma_k(d\boldsymbol{\xi}) = \{(a_2 - a_1) \log k/4\} d\boldsymbol{\xi}$  and  $\int_{h_k(P_k)} f(\boldsymbol{\xi}_k) \gamma_k(d\boldsymbol{\xi}) / V(P_k) = (1/4) \int_{[-1, 1]^2} f(\boldsymbol{\xi}) d\boldsymbol{\xi}$ , which satisfies condition **(A6-2)**. For any  $\boldsymbol{\xi} \in [-1 + \varepsilon, 1 - \varepsilon]^2$ , let  $(b, a) = h_k^{-1}(\boldsymbol{\xi})$ . Then,  $b = b_0 k^{(\xi_1 - 1)/2}$  and  $a = \{(a_2 - a_1)/2\} \xi_2 + (a_1 + a_2)/2$  so that  $\{[h_k^{-1}(\boldsymbol{\xi})]^{-1}(\sigma, \mu); (\sigma, \mu) \in P_k\} = \{(\sigma/b, (\mu - a)/b); (\sigma, \mu) \in P_k\}$  and  $\sigma/b, (\mu - a)/b$  satisfy the inequalities

$$k^{-(1+\xi_1)/2} < \frac{\sigma}{b} < k^{(1-\xi_1)/2},$$

$$-\frac{a_2 - a_1}{2b_0} (1 + \xi_2) k^{-(\xi_1 - 1)/2} < \frac{\mu - a}{b} < \frac{a_2 - a_1}{2b_0} (1 - \xi_2) k^{-(\xi_1 - 1)/2},$$

both of which are satisfied by  $k^{-\varepsilon/2} < \sigma/b < k^{\varepsilon/2}$  and

$$-\frac{a_2 - a_1}{2b_0} \varepsilon k^{-\varepsilon/2} < \frac{\mu - a}{b} < \frac{a_2 - a_1}{2b_0} \varepsilon k^{-\varepsilon/2}.$$

Hence, condition **(A6-3)** is satisfied and the minimaxity of  $\hat{q}^{BI}$  is established.  $\square$

Note that minimaxity still holds under the restriction  $D_0 = \{(\sigma, \mu) | a_1 < \mu < a_2, 0 < \sigma\}$ . However, we could not show minimaxity for the restriction  $D_1 = \{(\sigma, \mu) | a_1 < \mu < a_2, b_0 < \sigma\}$ , since we cannot take a sequence so that the lower and upper bounds of  $(\mu - a)/b$  can be expanded to the whole real line in the proof of Theorem 5.4.4. We conjecture that the best equivariant estimator is not minimax under the restriction  $D_1$ . From Kubokawa (2005), we also guess that the best equivariant estimator is not minimax for the restriction  $\{(\sigma, \mu) | a_1 < \mu/\sigma < a_2, \sigma > 0\}$ .

### 5.4.3 Multidimensional case

As an extension to a multidimensional model, we consider density functions of the forms  $p(\boldsymbol{\sigma}^{-1}(\mathbf{t}_x - \boldsymbol{\mu}), \boldsymbol{\sigma}^{-1} \mathbf{s}_x | \mathbf{u}_x) p_x(\mathbf{u}_x)$  and  $q(\boldsymbol{\sigma}^{-1}(\mathbf{t}_y - \boldsymbol{\mu}), \boldsymbol{\sigma}^{-1} \mathbf{s}_y | \mathbf{u}_y) q_y(\mathbf{u}_y)$  where  $\mathbf{u}_x$  and  $\mathbf{u}_y$  are location-scale invariant statistics,

$$\boldsymbol{\sigma}^{-1}(\mathbf{t}_x - \boldsymbol{\mu}) = \left( \frac{t_{x,1} - \mu_1}{\sigma_1}, \dots, \frac{t_{x,p} - \mu_p}{\sigma_p} \right) \quad \text{and} \quad \boldsymbol{\sigma}^{-1} \mathbf{s}_x = \left( \frac{s_{x,1}}{\sigma_1}, \dots, \frac{s_{x,p}}{\sigma_p} \right),$$

and  $\sigma^{-1}(\mathbf{t}_y - \boldsymbol{\mu})$  and  $\sigma^{-1}\mathbf{s}_y$  are defined similarly.

**[1] Ordered restriction of locations.** We first treat the constraint given by

$$M_1 = \{(\boldsymbol{\sigma}, \boldsymbol{\mu}) \mid \mathbf{B}\boldsymbol{\mu} \leq \boldsymbol{\alpha}, \sigma_1 = \cdots = \sigma_p = \sigma\},$$

where  $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_q)'$  is a  $q \times p$  known matrix for  $q \leq p$ ,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)'$  is a known vector, and the inequality  $\mathbf{B}\boldsymbol{\mu} \leq \boldsymbol{\alpha}$  means that  $\mathbf{b}'_i \boldsymbol{\mu} \leq \alpha_i$  for  $i = 1, \dots, q$ . This restriction means that the location parameters are restricted to the polyhedral convex cone and includes the positive orthant restriction  $\mu_i \geq 0$ ,  $i = 1, \dots, p$ , the simple order restriction  $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_p$ , and the tree order restriction  $\mu_1 \leq \mu_i$ ,  $i = 2, \dots, k$ .

Combining the arguments as in the proof of theorem 2.1 in Tsukuma and Kubokawa (2008) and the proof of Theorem 5.4.3, we can show the minimaxity of the best equivariant estimator.

**Théoreme 5.4.5.** *Assume that  $(\boldsymbol{\sigma}, \boldsymbol{\mu})$  is restricted to the polyhedral convex cone  $M_1$  with unrestricted unknown scale  $\sigma$ . Then, the best equivariant estimator is minimax in the estimation of the predictive density under the restricted parameter space.*

**[2] Ordered restriction of scales.** We next consider the constraint given by

$$M_2 = \{(\boldsymbol{\sigma}, \boldsymbol{\mu}) \mid \boldsymbol{\mu} \in \mathbb{R}^p, \mathbf{B}\boldsymbol{\eta} \leq \boldsymbol{\alpha}\},$$

where  $\boldsymbol{\eta} = (\eta_1, \dots, \eta_p)'$  for  $\eta_i = \log \sigma_i$ , and  $\mathbf{B}$  and  $\boldsymbol{\alpha}$  are the same as defined in  $M_1$ . This restriction means that  $\boldsymbol{\eta}$  is restricted on the polyhedral convex cone and includes the positive orthant restriction  $\sigma_i \geq 1$ ,  $i = 1, \dots, p$ , the simple order restriction  $\sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_p$  and the tree order restriction  $\sigma_1 \leq \sigma_i$ ,  $i = 2, \dots, k$ .

Since  $\boldsymbol{\mu}$  is not restricted and the problem is invariant under location transformations, we can consider location equivariant estimators, which depend on  $\mathbf{t}_x$  and  $\mathbf{t}_y$  through  $\mathbf{t}_y - \mathbf{t}_x$ . Thus, the risk function of the location equivariant estimator does not depend on  $\boldsymbol{\mu}$ . Then, the problem can be reduced to estimation in the scale family with the restriction  $\mathbf{B}\boldsymbol{\eta} \leq \boldsymbol{\alpha}$ . Hence from the arguments as in the proof of Tsukuma and Kubokawa (2008), it follows that the best equivariant estimator is minimax.

**Théoreme 5.4.6.** *Assume that  $(\sigma, \mu)$  is restricted into the polyhedral convex cone  $M_2$  with unrestricted location parameters  $\mu$ . Then, the best equivariant estimator is minimax in the estimation of the predictive density under the restricted parameter space.*

## 5.5 Concluding remarks

We have demonstrated that, for many restricted parameter space problems, the best equivariant predictive density  $\hat{q}^{BI}$  under Kullback-Leibler loss remains minimax, with constant risk matching the minimax risk. We point out that versions of Theorem 2.1, 3.1, 3.3, and 4.1 also follow from the results of Marchand and Strawderman (2012).

For lower (or upper) bounded location or scale parameter problems, we have introduced a novel adaptation of Kubokawa's IERD technique to show that the generalized Bayes procedure  $\hat{q}^U$  with respect to the truncation of the right Haar invariant measure onto the restricted parameter space dominates  $\hat{q}^{BI}$  and is thus minimax. These findings are analogous to various point estimation results previously established. It seems plausible, but more research is required, that similar minimax results and  $\hat{q}^{BI}$ - $\hat{q}^U$  comparisons hold for other choices of loss, such as for  $\alpha$ -divergence losses (e.g., Csiszár, 1967; Corcuera and Guummole, 1999). Finally, further analysis of the efficiency of Bayes estimators for other restricted parameter spaces, such as for univariate compact interval restrictions (see the end of Example 5.3.5), represent challenging and interesting problems for further research.

## Acknowledgments

The authors are grateful to the Associate Editor and referees for their valuable comments and helpful suggestions. Tatsuya Kubokawa's research is supported in part by Grant-in-Aid for Scientific Research Nos. 19200020 and 21540114 from the Japan Society for the Promotion of Science, Eric Marchand's research is supported in part by the Natural Sciences and Engineering Research Council of Canada, and William Strawderman's research is partially supported by a grant from the Simons Foundation (#209035).



# Bibliographie

- [1] Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, **62**, 547-554.
- [2] Brown, L.D., George, E.I., and Xu, X. (2008). Admissible predictive density estimation. *Annals of Statistics*, **36**, 1156-1170.
- [3] Casella, G. and Strawderman, W.E. (1981). Estimating a bounded mean. *Annals of Statistics*, **9**, 870-879.
- [4] Corcuera, J. M. and Giummole, F. (1999). A generalized Bayes rule for prediction. *Scandinavian Journal of Statistics*, **26**, 265-279.
- [5] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungary*, **2**, 299-315.
- [6] Farrell, R.H. (1964). Estimators of a location parameter in the absolutely continuous case. *Annals of Mathematical Statistics*, **35**, 949-998.
- [7] Fourdrinier, D., Marchand, É., Righi, A. and Strawderman, W.E. (2011). On improved predictive density estimation with parametric constraints. *Electronic Journal of Statistics*, **5**, 172-191.
- [8] George, E. I., Liang, F. and Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Annals of Statistics*, **34**, 78-91.
- [9] George, E. I., Liang, F. and Xu, X. (2012). From minimax shrinkage estimation to minimax shrinkage prediction. *Statistical Science*, **27**, 82-94.
- [10] Girshick, M.A. and Savage, L.J. (1951). Bayes and minimax estimates for quadratic loss functions. In *Proc. Second Berkeley Symp. Math. Statist. Probab.*, **1**, 53-74. University of California Press, Berkeley.

- [11] Hartigan, J. (2004). Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, **67**, 285-288.
- [12] Hora, R.B., and Buehler, R.J. (1966). Fiducial theory and invariant estimation. *Annals of Mathematical Statistics*, **37**, 643-656.
- [13] Hora, R.B., and Buehler, R.J. (1967). Fiducial theory and invariant prediction. *Annals of Mathematical Statistics*, **38**, 795-801.
- [14] Kato, K. (2009). Improved prediction for a multivariate normal distribution with unknown mean and variance. *Annals of the Institute of Statistical Mathematics*, **61**, 531-542.
- [15] Katz, M. (1961). Admissible and minimax estimates of parameters in truncated spaces. *Annals of Mathematical Statistics*, **32**, 136-142.
- [16] Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Annals of Mathematical Statistics*, **28**, 573-601.
- [17] Komaki, F. (2001). A shrinkage predictive distribution for multivariate normal observables. *Biometrika*, **88**, 859-864.
- [18] Kubokawa, T. (1994A). A unified approach to improving equivariant estimators. *Annals of Statistics*, **22**, 290-299.
- [19] Kubokawa, T. (1994B). Double shrinkage estimation of ratio of scale parameters. *Annals of the Institute of Statistical Mathematics*, **46**, 95-119.
- [20] Kubokawa, T. (2004). Minimavity in estimation of restricted parameters. *Journal of the Japanese Statistical Society*, **34**, 229-253.
- [21] Kubokawa, T. (2005A). Estimation of a mean of a normal distribution with a bounded coefficient of variation. *Sankhyā*, **67**, 499-525.
- [22] Kubokawa, T. (2005B). Estimation of bounded location and scale parameters. *Journal of the Japanese Statistical Society*, **35**, 221-249.
- [23] Kubokawa, T. and Saleh, A.K.MD.E. (1998). Estimation of location and scale parameters under order restrictions. *Journal of Statistical Research*, **28**, 41-51.
- [24] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*. Springer-Verlag, 2nd edition.

- [25] Liang, F. and Barron, A. (2004). Exact minimax strategies for predictive density estimation, data compression, and model selection. *IEEE Trans. Inform. Theory*, **50**, 2708-2726.
- [26] Marchand, É., and Payandeh Najafabadi, A.T. (2011). Bayesian improvements of a MRE estimator of a bounded location parameter. *Electronic Journal of Statistics*, **5**, 1495-1502.
- [27] Marchand, É. and Perron, F. (2001). Improving on the MLE of a bounded normal mean. *Annals of Statistics*, **29**, 1078-1093.
- [28] Marchand, É. and Strawderman, W.E. (2004). Estimation in restricted parameter spaces : A review. *Festschrift for Herman Rubin*, IMS Lecture Notes-Monograph Series, **45**, 21-44.
- [29] Marchand, É., and Strawderman, W. E. (2005A). On improving on the minimum risk equivariant estimator of a location parameter which is constrained to an interval or a half-interval. *Annals of the Institute of Statistical Mathematics*, **57**, 129-143.
- [30] Marchand, É., and Strawderman, W. E. (2005B). On improving on the minimum risk equivariant estimator of a scale parameter under a lower bound constraint. *Journal of Statistical Planning and Inference*, **134**, 90-101.
- [31] Marchand, É., and Strawderman, W.E. (2012). A unified minimax result for restricted parameter spaces. *Bernoulli*, **18**, 635-643.
- [32] Tsukuma, H. and Kubokawa, T. (2008). Stein phenomenon in estimation of means restricted to a polyhedral convex cone. *Journal of Multivariate Analysis*, **99**, 141-164.
- [33] van Eeden, C. (2006). *Restricted parameter space problems - Admissibility and minimaxity properties*. Lecture Notes in Statistics, **188**, Springer.
- [34] Wijsman, R.A. (1985). A useful inequality on ratios of integrals, with application to maximum likelihood estimation. *Journal of the American Statistical Association*, **80**, 472-475.

# CONCLUSION

Nous avons trouvé, tout au long de ce mémoire, pour la perte Kullback-Leibler, les lois *a priori* utilisées et les familles de densités étudiées pour trouver de bons estimateurs pour la fonction de densité d'une variable aléatoire, notamment selon le critère du minimax, ou dominant des estimateurs naturels (tels les estimateurs par substitution), qui sont équivariants, etc. Des liens ont été faits entre l'estimation ponctuelle et l'estimation par densité prédictive, notamment pour le cas de la loi multinormale et de la loi gamma. Finalement, nous avons trouvé des conditions sur la famille de lois de la variable aléatoire  $Y|\theta$  pour obtenir un estimateur de Bayes minimax lorsque l'espace paramétrique de  $\theta$  (paramètre de position ou d'échelle) est restreint. Des avenues possibles de recherche pourraient considérer des espaces paramétriques de la forme  $[a, b]$  où  $a > -\infty$  et  $b < \infty$ , des familles de densités qui ne sont pas à rapport de vraisemblance monotone croissantes ou des espaces paramétriques non convexes. Le cas multivarié a été travaillé au chapitre 3 et dans les résultats généraux, mais il serait intéressant de trouver d'autres exemples que ceux illustrés dans ce mémoire pour mieux comprendre l'ampleur des résultats. On pourrait penser aux lois sphériques et elliptiques. D'autres pertes que la perte Kullback-Leibler pourraient aussi être étudiées.

# Bibliographie

- [1] Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, **62**, 547-554.
- [2] Barron, A. et Liang, F. (2004). Exact minimax strategies for predictive density estimation, data compression, and model selection. *IEEE Transaction on Information Theory*, **50**, 2708-2726.
- [3] Brandwein, A.C. et Strawderman, W.E. (1990). Stein estimation : the spherically symmetric case. *Statistical Science*, **5**, 356-369.
- [4] Brown, L., George, E. et Xu, X. (2008). Admissible predictive density estimation. *The Annals of Statistics*, **36**, 1156-1170.
- [5] Casella, G. et Lehmann, E.L. (1998). Theory of point estimation. *Springer Texts in Statistics*
- [6] Casella, G. et Strawderman, W.E. (1981). Estimating a bounded normal mean. *The Annals of Statistics*, **9**, 870-878.
- [7] Fourdrinier, D., Marchand, É., Righi, A. et Strawderman, W.E. (2011). On improved predictive density estimation with parametric constraints. *Electronic Journal of Statistics*, **5**, 172-191.
- [8] George, E., Liang, F. et Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *The Annals of Statistics*, **34**, 78-91.
- [9] Hartigan, J.A. (2004). Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, **67**, 285-288.

- [10] James, W. et Stein, C. (1961). Estimation with quadratic loss. *Proc. Third Berkeley Symp. Math. Statist. Probab.*, **1**, 361-380.
- [11] Komaki, F. (2001). A shrinkage predictive distribution for multivariate normal observables. *Biometrika*, **88**, 859-864.
- [12] Kubokawa, T. (1994A). A unified approach to improving equivariant estimators. *The Annals of Statistics*, **22**, 290-299.
- [13] Kubokawa, T. (1994B). Double shrinkage estimation of ratio of scale parameters. *Annals of the Institute of Statistical Mathematics*, **46**, 95-119.
- [14] Kubokawa, T., Marchand, É., Strawderman, W.E. et Turcotte, J.P. (2012). Minimality in predictive estimation with parametric constraints. Rapport de recherche #110, Département de mathématiques, Université de Sherbrooke.
- [15] Kullback, S. et Leibler, R.A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.
- [16] Lindley, D.V. (1962). Discussion of «Confidence sets for the mean of a multivariate normal distribution,» by C.M. Stein. *J. Roy. Statist. Soc. Ser. B*, **24**, 285-287.
- [17] Marchand, É. (2012). Predictive density estimation : An introduction. *Notes non publiées*.
- [18] Marchand, É. et Strawderman, W.E. (2004). Estimation in a restricted parameter spaces : A review. *Festschrift for Herman Rubin*, IMS Lecture Notes-Monograph Series, **45**, 21-44.
- [19] Robert, C. (1996). Intrinsic losses. *Theory and decision*, Kluwer Academic Publishers, **40**, 191-214.
- [20] Robert, C. (2007). The Bayesian choice. *Springer Texts in Statistics*, 105-163.
- [21] Stein, C. (1956). Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.*, **1**, 197-206.

[22] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, **9**, 1135-1151.

[23] Strawderman, W.E. (2003). On minimax estimation of a normal mean vector for general quadratic loss. *Festschrift for Constance van Eeden*, IMS Lecture Notes-Monograph Series, **42**, 3-14.