

# Sur une approche décisionnelle pour l'analyse bayésienne et l'estimation de densités prédictives

par

Aziz L'Moudden

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

FACULTÉ DES SCIENCES  
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, octobre 2018

Le 23 octobre 2018

le jury a accepté la thèse de Monsieur Aziz L'Moudden dans sa version finale.

Membres du jury

Professeur Éric Marchand  
Directeur de recherche  
Département de mathématiques

Professeur Jean Vaillancourt  
Membre externe  
Département de sciences de la décision, HEC Montréal

Professeure Patrick Richard  
Membre interne  
Département d'économie, Université de Sherbrooke

Professeur Taoufik Bouezmarni  
Président-rapporteur  
Département de mathématiques

# SOMMAIRE

Cette thèse porte principalement sur l'estimation de densités prédictives pour une inconnue  $Y$  à partir d'un observé  $X$  dont les lois de probabilité dépendent d'un même paramètre  $\theta$ . Elle est le fruit de trois projets sur lesquels j'ai travaillé durant les trois dernières années. Ces projets ont abouti à un article publié, un autre soumis et un Chapitre prometteur.

Le premier article, présenté au Chapitre II, a été publié en 2017 dans la revue *Journal of Statistical Planning and Inference*. Il traite l'étude de densités prédictives pour des modèles Gamma avec restrictions sur le paramètre d'échelle sous la perte Kullback-Leibler. La performance fréquentiste de plusieurs estimateurs est analysée, dont des densités prédictives bayésiennes et des densités prédictives de type plug-in. Des résultats de dominance sont obtenus, ainsi que des densités prédictives minimax. Notamment, on a obtenu une méthode universelle pour améliorer une densité prédictive de type plug-in.

Le deuxième article, soumis pour publication récemment, est présenté au Chapitre III. Nous considérons le problème d'estimation d'une densité prédictive pour des modèles de loi normale multivariée, sous une classe de perte de type  $\alpha$ -divergence incluant les pertes Kullback-Leibler et Hellinger. Ce travail contient un résultat important sur une stratégie qui donne une classe de densités prédictives dominant une densité prédictive de type plug-in (universellement par rapport à la dimension, l'espace paramétrique, la

densité prédictive de type plug-in et la perte choisie). De plus, nos résultats étendent une partie des travaux précédents sur la perte de Kullback-Leibler (Fourdrinier et coll., 2011, Electron. J. Stat), ainsi que ceux portant sur les pertes  $L^1$  et  $L^2$  intégrées, et sont appliqués aux estimateurs de rétrécissement ou de type Stein de  $\theta$ .

Le troisième projet de cette thèse est le Chapitre 4. Ce dernier traite l'analyse bayésienne pour deux cadres généraux de mélanges avec densités :

$$i) \quad f_{\theta}(x) = \sum_k p_{\theta}(k) g_k(x),$$

$$ii) \quad f_{\theta}(x) = \int g_t(x) dH_{\theta}(t).$$

On donne les structures générales pour les densités *a posteriori* et les densités prédictives et quelques résultats de dominances. Plusieurs représentations novatrices sont obtenues, notamment pour des mélanges de lois normales, les lois décentrées du chi-deux, Bêta, Fisher, la loi du coefficient de détermination  $R^2$  dans un contexte de régression multiple, la loi de Kibble, et les problèmes avec contraintes sur  $\theta$ , parmi d'autres.

Enfin, certaines généralisations et développements sont présentés au chapitre 1.

# ABSTRACT

This thesis deals mainly with the estimation of predictive densities for an unknown  $Y$  from an observed  $X$  whose probability laws depend on the same parameter  $\theta$ . It is the result of three projects that I have worked on for the last three years. These projects resulted in a published article, another submitted and a promising Chapter.

The first article, presented in Chapter II, was published in 2017 in the "Journal of Statistical Planning and Inference". It deals with the study of predictive densities for Gamma models with restrictions on the scale parameter under Kullback-Leibler loss. The frequentist performance of several estimators is analyzed, including Bayesian predictive densities and plug-in predictive densities. Dominance results as well as minimax predictive densities are obtained. In particular, we obtain a universal method to improve upon predictive density of plug-in type.

The second article, submitted for publication recently, is presented in Chapter III. We consider the predictive density estimation problem for multivariate normal law models, under a loss class of  $\alpha$ -divergence including Kullback-Leibler and Hellinger losses. This work contains a significant result on a strategy that gives a class of predictive densities dominating a predictive density of plug-in type (universally compared to the dimension, the parametric space, the predictive density of plug-in type and the chosen loss). In addition, our results extend previous work on the Kullback-Leibler loss (Fourdrinier et

al., 2011, J. Electr., J.), as well as those on  $L^1$  and  $L^2$  losses built-in, and are applied to shrink or Stein estimators of  $\theta$ .

The third project of the thesis is Chapter 4 which deals with Bayesian analysis for two general frameworks of mixtures with densities :

$$i) \quad f_{\theta}(x) = \sum_k p_{\theta}(k) g_k(x),$$

$$ii) \quad f_{\theta}(x) = \int g_t(x) h_{\theta}(t) dt.$$

The general structures for the *a posteriori* densities, the predictive densities and some dominance results are given. Several novel representations are obtained, notably for mixtures of normal laws, the decentered Chi-square laws, Beta, Fisher, the coefficient of determination  $R^2$  law in a multiple regression context, the Kibble law, and the problems with constraints on  $\theta$ , among others.

Finally, some generalizations and developments appear in Chapter 1.

# DÉDICACES

À ma mère Aicha et à mon père Lkhalil.

À ma femme Jalila et à nos enfants Arij et Yahya...

# REMERCIEMENTS

La présente thèse a été réalisée grâce à la Miséricorde d'**Allah**, et à la collaboration de plusieurs personnes que je tiens à remercier.

En premier lieu, j'exprime mes profonds remerciements à mon superviseur de thèse, le professeur Éric Marchand, qui m'a appris beaucoup sur la recherche, la rédaction, et pour ses discussions dans une ambiance agréable. Je le remercie aussi pour son support surtout aux moments difficiles.

Je tiens à remercier aussi le professeur William E. Strawderman, qui m'a grandement aidé à faire avancer mes travaux de thèse, ainsi que le professeur - confrère et ami - Othmane Kortbi.

J'aimerais aussi remercier les professeurs du département de mathématiques de l'Université de Sherbrooke, et particulièrement le professeur Taoufik Bouezmarni pour avoir accepté de lire et d'évaluer ma thèse et d'agir à titre de président-rapporteur.

Je tiens à exprimer ma reconnaissance aux membres du jury, les professeurs Jean Vaillancourt et Patrick Richard qui ont également accepté de lire mon travail.

Je remercie sincèrement mes parents de leur encouragement pour aller jusqu'au bout. je suis reconnaissant à ma femme Jalila, pour son soutien et ses encouragements.

Mon doctorat a été rendu possible grâce au soutien financier du département de



mathématique, le Fond de recherche de mon directeur de recherche et l'ISM.

Je voudrais aussi remercier tous mes confrères de travail, je pense notamment à tous les thésards avec qui j'ai passé de très bons moments. Enfin, que toute personne qui a contribué d'une manière où d'une autre pour l'accomplissement de ce travail trouve ici ma gratitude.

# TABLE DES MATIÈRES

SOMMAIRE	iv
SUMMARY	vi
DÉDICACE	viii
REMERCIEMENTS	ix
TABLE DES MATIÈRES	xi
LISTE DES FIGURES	xv
NOTATIONS	xvii
INTRODUCTION	1
CHAPITRE 1 — Notions préliminaires	7
1.1 Notions préliminaires . . . . .	7
1.1.1 Fonction de perte et de risque . . . . .	8

1.1.2	Lois de probabilité hypergéométrique généralisée . . . . .	10
1.1.3	Lois de Bêta de type <i>II</i> et Gamma de type <i>II</i> . . . . .	11
1.2	Estimation de densités prédictives . . . . .	13
1.2.1	Fonction de perte intrinsèque : Pertes $L^\alpha$ et $\alpha$ -divergence . . . . .	13
1.2.2	Approche fréquentiste : Estimateur par substitution ( <i>plug-in</i> ) . . . . .	18
1.2.3	Approche bayésienne et meilleure densité prédictive équivariante . . . . .	20
1.2.4	Quelques résultats pour la loi Gamma . . . . .	26
1.2.5	Lien entre l'estimation d'une densité prédictive et l'estimation ponctuelle . . . . .	31
1.3	Une stratégie d'évaluation des risques . . . . .	35

**CHAPITRE 2 — On predictive density estimation for Gamma models with parametric constraints. . . . . 40**

2.1	Introduction . . . . .	42
2.2	Bayes estimators . . . . .	44
2.2.1	A general scale family representation of the MRE predictive density estimator . . . . .	47
2.3	Dominance results . . . . .	48
2.4	Minimax estimators and minimax risk . . . . .	51
2.5	Plug-in predictive density estimators . . . . .	53
2.6	Variance expansion improvements on plug-in predictive density estimators . . . . .	58

2.7	Illustrations and final comments . . . . .	61
 <b>CHAPITRE 3 — On Predictive Density Estimation under <math>\alpha</math>-divergence</b>		
	<b>Loss.</b>	<b>67</b>
3.1	Introduction . . . . .	68
3.2	Main results . . . . .	71
3.2.1	Case $\hat{\theta}_a(X) = aX$ . . . . .	72
3.2.2	Case of a non-negative mean with $\hat{\theta}_+(X) = \max\{X, 0\}$ . . . . .	75
3.2.3	General $\hat{\theta}(X)$ . . . . .	79
3.2.4	Behaviour of the cut-off point $k(d, \alpha, \sigma_X^2, \sigma_Y^2)$ and simultaneous dominance . . . . .	84
3.2.5	Example . . . . .	86
3.3	Concluding remarks . . . . .	86
3.4	Appendix . . . . .	88
 <b>CHAPITRE 4 — Sur les densités prédictives bayésiennes de mélanges</b>		
4.1	Introduction . . . . .	92
4.2	Mélanges continus . . . . .	96
4.3	Mélanges discrets . . . . .	101
4.3.1	$X \theta$ et $Y \theta$ de type I . . . . .	102
4.3.2	$X \theta$ et $Y \theta$ de type II . . . . .	110
4.3.3	Inférence bayésienne pour le loi Kibble . . . . .	112

4.4 Discussion . . . . .	116
<b>CONCLUSION</b>	<b>118</b>
4.5 Annexe A . . . . .	119
4.6 Annexe B . . . . .	121

# LISTE DES FIGURES

1.1	Les fonctions $h_\alpha$ pour différentes valeurs de $\alpha$ . . . . .	15
1.2	Diagramme représentatif de la propriété d'invariance bayésienne liant les densités prédictives. . . . .	16
1.3	Risques des densité prédictives $\hat{q}_{\pi_0}$ en rouge et $\hat{q}_{\pi_C}$ ( $C = [1, 4]$ ) en bleu pour $\delta = 0$ , $\alpha_1 = 1$ et $\alpha_2 = 2$ . . . . .	29
2.1	Ratios $\frac{\inf_c \bar{R}(\hat{q}_c)}{\inf_{\hat{q}} \bar{R}(\hat{q})}$ of minimax risks for $\beta \geq a$ . . . . .	58
2.2	Risks of predictive density estimators : $\alpha_1 = 3$ , $\alpha_2 = 5$ , $\beta \geq 1$ . . . . .	62
2.3	Risks of predictive density estimators : $\alpha_1 = 2$ , $\alpha_2 = 1$ , $\beta \leq 1.0$ . . . . .	63
2.4	Risks of predictive density estimators : $\alpha_1 = 3$ , $\alpha_2 = 3$ , $\beta \in [1, 8]$ . . . . .	63
3.1	Risk ratios (3.8) as functions of $\alpha$ for various pairs $(d, r)$ . . . . .	72
3.2	Hellinger risk ratios $\frac{R_0(\theta, q_{\hat{\theta}_a, c})}{R_0(\theta, q_{\hat{\theta}_a, 1})}$ for various $c$ and $r$ . . . . .	76
3.3	Risk ratios $\frac{R_\alpha(\theta, q_{\hat{\theta}_+, c})}{R_\alpha(\theta, q_{\hat{\theta}_+, 1})}$ for $c = \kappa(\alpha, r)$ , and various $c$ and $r$ . . . . .	79
3.4	Cut-off points $k(3, \alpha, 1, \sigma_Y^2 = 1/r)$ as functions of $\alpha$ , for the James-Stein estimator . . . . .	85

3.5	Risk ratios $\frac{R_\alpha(\theta, q_{\hat{\theta}_{JS^+}, c})}{R_\alpha(\theta, q_{\hat{\theta}_{JS^+}, 1})}$ for $c^2 = k(d, \alpha = 0, \sigma_X^2 = 1, \sigma_Y^2 = 1)$ as functions of $\ \theta\ $ . . . . .	87
4.1	La loi a posteriori $\pi(\theta   x)$ . . . . .	106
4.2	La densité prédictive $\hat{q}_\pi(y; x)$ . . . . .	107
4.3	La loi a posteriori $\pi(\theta   r_1^2)$ pour plusieurs valeurs de $r_1^2$ . . . . .	109
4.4	La densité prédictive $\hat{q}_\pi(r_2^2; r_1^2)$ pour plusieurs valeurs de $r_1^2$ . . . . .	109

# NOTATIONS

$\theta$  : Paramètre d'intérêt

$\hat{\theta}$  : Estimateur de  $\theta$

$p_\theta(\cdot)$ ,  $q_\theta(\cdot)$  : Fonctions de densités

$\hat{q}$  : Densité prédictive

$\hat{\theta}_\pi$  : Estimateur de Bayes du paramètre  $\theta$  associés à la loi a priori  $\pi$

$\hat{q}_\pi$  : Densité prédictive bayésienne de  $q(\cdot)$  associé à la loi a priori  $\pi$

$q_{\hat{\theta}}$  : Estimateur de plug-in de la densité  $q(\cdot)$  associé à l'estimateur  $\hat{\theta}$

$m_\pi(\cdot)$  : Densité marginale sous la loi a priori  $\pi$

$R_{KL}$ ,  $L_{KL}$  : Fonctions de risque et de perte Kullback-Leibler

$R_\alpha$ ,  $L_\alpha$  : Fonctions de risque et de perte  $\alpha$ -divergence

$R^\alpha$ ,  $L^\alpha$  : Fonctions de risque et de perte  $\mathbb{L}^\alpha$

$R_Q$ ,  $L_Q$  : Fonctions de risque et de perte quadratique

$L_0$  : Fonction de perte entropie

$\Delta(\cdot)$  : Différence entre les risques de deux estimateurs

$\phi(\cdot)$ ,  $\Phi(\cdot)$  : Densité et fonction de répartition d'une loi normale

$Ga(\cdot)$  : Loi Gamma

$\mathbb{I}_D(\cdot)$  : Fonction indicatrice d'un sous-ensemble  $D$

${}_pF_q(\cdot)$  : Fonction hypergéométrique généralisée  ${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; \cdot)$ .



# Introduction

L'analyse prédictive est un domaine de la statistique qui sert à prévoir des événements futurs à partir des données historiques et actuelles. Elle occupe un rôle majeur dans le développement des méthodes statistiques telles que les séries chronologiques, les processus stochastiques et la résolution des problèmes pratiques actuels comme les problèmes de régression (Lawless et Fredette, 2005), de compression de données et de sélection de modèles (Liang et Barron, 2004).

Cette thèse a pour objectif d'élaborer une densité qui sera utile pour prédire le comportement de futurs échantillons  $Y$  à partir des données antérieures observées  $X$  (Aitchison et Dunsmore, 1975). La forme la plus élaborée, qu'on appellera densité prédictive, est d'estimer la densité conditionnelle  $f(y|x)$  (Geisser, 1971) de manière optimale en fonction de la qualité d'ajustement appropriée. On considère les vecteurs aléatoires conditionnellement indépendantes suivantes

$$X|\theta \sim p_\theta(x) \text{ et } Y|\theta \sim q_\theta(y), \quad x, y \in \mathbb{R}^d$$

où la dépendance repose sur le paramètre inconnu  $\theta \in \Theta \subseteq \mathbb{R}^p$ . Le défi statistique est d'obtenir un substitut pour la densité  $q_\theta$  inconnue dans le cadre paramétrique. Il existe à la fois des approches fréquentiste et bayésienne, bien que le terme lié à la densité prédictive semble plus associé à la perspective bayésienne. Nous abordons trois thèmes de l'estimation de densité prédictive et de leurs calculs de risque correspondants.

## I- Les densités prédictives par substitution (plug-in)

Les densités de type plug-in sont des estimations naturelles de densité prédictive. Elles se situent dans la même classe paramétrique que la densité  $q_\theta$  et sont construites en remplaçant le paramètre inconnu  $\theta$  par un estimé  $\hat{\theta}(X)$  basé sur  $X \sim p_\theta(\cdot)$ , dans la densité  $q_\theta(\cdot)$ , ce qui donne  $q_{\hat{\theta}}$ . Cette approche simple couramment utilisée peut conduire souvent à des estimateurs de densité prédictive de mauvaise qualité (Aitchison, 1975 ; Geisser, 1993 ; Komaki, 1996 ; Barberis, 2000 ; Tanaka et Komaki, 2005 ; Tanaka, 2006). Par exemple, Aitchison (1975) a montré que les estimateurs de densité de plug-in de maximum de vraisemblance pour les modèles Gamma et pour les modèles normales sont uniformément dominés sous le perte Kullback-Leibler par des estimateurs prédictifs bayésiens basés sur des lois *a priori* non-informatives usuelles. Fourdrinier et coll. (2011) ont proposé une méthode pour améliorer cet estimateur de densité prédictive par l'expansion de la variance. De plus, une des problématiques reliées aux densités de plug-in est l'ignorance de l'incertitude sur  $\theta$  en le traitant comme si  $\theta$  était égal à  $\hat{\theta}$ . En revanche, l'approche bayésienne résout directement l'incertitude de ce paramètre, en donnant à  $\theta$  une distribution préalable, ce qui devient intégré dans l'estimateur de densité.

## II- Les densités prédictives de Bayes

L'approche bayésienne introduite sur les inconnus révèle une réponse naturelle et immédiate. Par exemple, supposons que nous observons les données  $X \sim p(x|\theta)$  et nous souhaitons prédire  $Y \sim q(y|\theta)$  où  $\theta$  est un paramètre commun inconnu. Étant donné une loi *a priori*  $\pi$  pour  $\theta$ , il résulte une estimation naturelle de  $q(y|\theta)$  est la densité prédictive qui minimise la perte espérée *a posteriori* associée à une fonction de perte  $L$  (Berger, 1985). Les principaux défis pour la mise en œuvre de cette approche est le choix d'une loi *a priori* (Liang et coll. (2008)) ainsi que la fonction de perte appropriée. Cette approche a été fructueuse pour un certain nombre de problèmes. En particulier, les travaux de

Komaki (2001), Liang et Barron (2004), George, Liang et Xu (2006) et Brown, George et Xu (2008) ont établi des conditions de minimaxité ainsi que des résultats probants pour la densité prédictive bayésienne multivariée. Pour les distributions au-delà de la normale, des nouveaux résultats de risque KL pour les densités prédictives bayésiennes ont été développés par Aslan (2006), Hartigan (1998), Komaki (1996, 2001, 2004) et Sweeting, Datta, et Ghosh (2006) parmi bien d'autres.

### **III- Les densités prédictives gaussiennes et relations avec l'estimation ponctuelle**

Considérons le cas où  $X \sim N_d(\theta, \sigma_X^2 I_d)$  indépendant de  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$  avec les variances  $\sigma_X^2$  et  $\sigma_Y^2$  connues. L'objectif est d'obtenir une densité prédictive pour  $Y$  à partir de  $X$ . L'article de revue de George, Liang et Xu (2012, Stat. Science) décrit plusieurs contributions récentes, dont celles de Komaki (2001), Liang et Barron (2004), George, Liang et Xu, Brown (2006), George et Xu (2008), parmi d'autres. Récemment, des liens fondamentaux entre les problèmes d'estimation de densité prédictive et d'estimation ponctuelle sont mis de l'avant dans la littérature vis-à-vis les différentes analyses statistiques suivantes : analyse bayésienne et la minimaxité. Par exemple, il y a une connexion entre les travaux de Strawderman (1971) et George Liang et Xu (2006), Brown et Hwang (1982) et Brown George et Xu (2008) à propos de l'admissibilité des estimateurs sans contrainte. Également, dans le cas des paramètres sous contraintes, on trouve Pinsker (1980), Xu et Liang (2010) et des travaux récents de Fourdrinier et coll. (2011), de Kubokawa et coll (2013) et de Marchand et Sadeghkhani (2017) qui s'intéressent aux thématiques telles que des modèles différents et la présence d'information additionnelle.

## IV- Notre contributions

Cette thèse est divisée en quatre chapitres. Au premier chapitre, nous présentons dans un cadre paramétrique quelques notions de base pour obtenir des densités prédictives efficaces, comme la notion d’admissibilité, les fonctions de pertes, les méthodes de plug-in et bayésiennes. Nous donnons quelques exemples faisant intervenir plusieurs techniques, dont l’expansion de la variance, les fonctions de coût duaux en estimation ponctuelle, l’estimation sous contraintes, et une stratégie d’évaluation des risques. Ce Chapitre sera utile pour les Chapitres suivants.

Dans le deuxième chapitre, nous présentons le travail que nous avons réalisé en collaboration avec mon directeur de recherche Éric Marchand, Othmane Kortbi (Professeur à Al Ain University) et William E. Strawderman (Professeur à Rutgers University). Ce travail a été publié dans *Journal of Statistical Planning and Inference* en 2017.

Dans ce travail, on s’intéresse à l’estimation d’une densité prédictive pour  $Y \sim \text{Ga}(\alpha_2, \beta)$  sous la perte de Kullback-Leibler basée sur  $X \sim \text{Ga}(\alpha_1, \beta)$ , et pour les situations où il existe une contrainte paramétrique de la forme  $\beta \in C = (a, b)$ . Nous obtenons des représentations pour les densités prédictives de Bayes et la densité prédictive équivalente du risque minimal dans le problème sans contraintes. Il fût déjà démontré que la densité de Bayes généralisé associée à la troncature de la loi a priori non informative sur  $C$  domine la densité prédictive équivariante, et est minimax chaque fois que  $a = 0$  ou  $b = \infty$ . Des comparaisons analytiques des densités prédictives du plug-in,  $\text{Ga}(\alpha_2, \hat{\beta})$ , qui incluent la densité prédictive de mee (meilleur estimateur équivariant), sont obtenues avec des résultats s’appliquant également pour l’estimation ponctuelle sous double perte d’entropie  $\frac{\beta}{\hat{\beta}} - \log\left(\frac{\beta}{\hat{\beta}}\right) - 1$ . Des évaluations numériques confirment qu’une telle densité prédictive est beaucoup moins efficace que certaines alternatives Bayésiennes concernant le cas de la restriction paramétrique. Enfin, il est démontré que les améliorations de l’expansion de variance de densités prédictives plug-in de forme  $\text{Ga}\left(\frac{\alpha_2}{k}, k\hat{\beta}\right)$  peuvent toujours être

trouvées pour un sous-ensemble des valeurs de  $k > 1$  et  $\hat{\beta}$  un estimateur non-dégénéré.

Le troisième chapitre est constitué d'un article soumis pour publication. Nous étudions l'efficacité des densités prédictives sous la perte  $\alpha$ -divergence  $L_\alpha$  pour estimer la densité de  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$  basé sur  $X \sim N_d(\theta, \sigma_X^2 I_d)$ . Nous identifions un grand nombre de cas où l'amélioration d'une densité de plug-in peut être obtenue en augmentant la variance, étendant ainsi les résultats antérieurs applicables à la perte de Kullback-Leibler. Les résultats et les preuves sont unifiés par rapport à la dimension  $d$ , les variances  $\sigma_X^2$  et  $\sigma_Y^2$ , le choix de la perte  $L_\alpha$ ;  $\alpha \in (-1, 1)$ . Les résultats s'appliquent également à un grand nombre de densités de plug-in, ainsi qu'à des espaces de paramètres restreints avec  $\theta \in \Theta \subset \mathbb{R}^d$ . Les résultats théoriques sont accompagnés de diverses observations, illustrations et implications traitant par exemple la robustesse par rapport aux variances du modèle et de la dominance simultanée par rapport à la perte.

Enfin, dans le quatrième et dernière chapitre, nos résultats se rapportent à deux classes de distributions de mélanges qui sont

- (A) Les familles de mélange continues de densité de la forme

$$p_\theta(x) = \int f_t(x) dH_\theta(t).$$

- (B) Les familles de mélange discrets de densité de la forme

$$p_\theta(x) = \sum_k q_\theta(k) f_k(x).$$

Nous décrivons les densités prédictives bayésiennes pour le coût Kullback-Leibler, ainsi que les lois a *posteriori* associées à des lois a priori  $\pi$ . Nous montrons que les distributions a *posteriori* et prédictive sont aussi des modèles de mélanges du même type que  $Y|\theta$ .

Ainsi, pour les familles (A), nous proposons des estimateurs de densité prédictive pour  $Y|\theta$ , avec  $\theta$  est un paramètre de position ou d'échelle avec des exemples de loi multinormale et loi Gamma.

Pour la famille (B), plusieurs cas de mélanges intéressants sont étudiés, tels la distribution bivariée Gamma de Kibble, la distribution  $\chi^2$  décentrée, la distribution Fisher décentrée et la distribution du coefficient de détermination  $R^2$ . Des nouvelles représentations des lois à posteriori et densités prédictives sont obtenues.

Enfin, l'éventail de lois *a priori* étudié est vaste et inclut, par exemple, des lois *a priori* adaptées à des contraintes d'ordre dont la loi de Mackay.

# CHAPITRE 1

## Notions préliminaires

Ce chapitre est divisé en trois sections. Dans la première section, nous donnerons des définitions, propriétés, théorèmes et lemmes permettant de démontrer nos résultats dans les Chapitres 2, 3 et 4. Puis, nous présentons le problème de l'estimation d'une densité prédictive d'une variable aléatoire dans la deuxième section. Enfin, dans la troisième section, nous développons quelques techniques mathématiques de calcul de risques pour la loi normale multivariée. Ces résultats préliminaires servent dans les trois chapitres suivants.

### 1.1 Notions préliminaires

Nous faisons référence à certaines notions de base d'inférence statistique dans le cas d'un problème d'estimation. De bonnes références sont Berger(1985) et Lehmann et Casella (1999). Nous travaillons avec un modèle statistique  $X \sim p_\theta \in \{\mathbb{P}_\theta, \theta \in \Theta \subseteq \mathbb{R}^d\}$  une famille de densités avec  $\theta \in \Theta$  supposé inconnu. Nous dénotons par  $\delta(X)$  un estimateur d'une fonction paramétrique  $h(\theta)$ .

### 1.1.1 Fonction de perte et de risque

Puisque  $\delta(X)$  est une variable aléatoire, il est nécessaire de spécifier une mesure de l'écart entre  $\delta(X)$  et  $h(\theta)$ . À cette fin, on introduit une fonction de perte (ou de coût) mesurable à valeurs réelles  $L(\theta, \delta)$ . Elle est définie selon le problème étudié et constitue l'armature du problème statistique. Or, la perte  $L(\theta, \delta(x))$ , pour  $X = x$  observée, dépend de l'observation  $x$ , la fonction de perte ne suffit pas par elle-même à évaluer les performances globales de  $\delta(X)$ . Pour y remédier, on considère le coût "moyen" qui est la fonction risque définie comme suit.

**Définition 1.1** (*Risque fréquentiste*)

*Le risque fréquentiste est défini par :*

$$R(\theta, \delta) = \mathbb{E}^{X|\theta} \left[ L(\theta, \delta(X)) \right], \theta \in \Theta.$$

Nous souhaiterons à plusieurs reprises comparer des estimateurs d'une classe d'estimateurs, et nous pouvons le faire en utilisant la fonction de risque.

Dans un cadre bayésien où l'on dispose d'une densité a priori  $\pi(\theta)$  pour  $\theta$  par rapport à une mesure  $\sigma$ -finie  $\nu$  et d'une observation de  $X$ , on a les étends de définition suivants.

**Définition 1.2** (*Perte espérée et estimateur de Bayes*)

*À partir d'une loi a priori  $\pi$  sur  $\Theta$  d'une fonction de perte  $L(\theta, \delta)$ , la perte espérée a posteriori associée à  $\delta(x)$  est définie par*

$$\mathbb{E}^{\theta|X=x} \left[ L(\theta, \delta(x)|x) \right] = \int_{\Theta} L(\theta, \delta(x)) \pi(\theta|x) d\nu(\theta).$$

*Un estimateur de Bayes  $\delta_{\pi}(X)$  est donné par*

$$\delta_{\pi}(x) = \arg \min_{\delta} \mathbb{E}^{\theta|x} \left[ L(\theta, \delta(x)) \right].$$



On dit que  $\delta_\pi(X)$  est un estimateur de Bayes généralisé si  $\pi$  est une loi *a priori* impropre, mais que la loi *a posteriori*  $\pi(\cdot|x)$  existe.

Nous examinons ici un exemple de détermination d'estimateurs bayésiens en utilisant la fonctions de perte  $L(\theta, \delta) = \|\delta - h(\theta)\|^2$ .

**Lemme 1.3** *Soit  $Z$  une variable aléatoire telle que :  $E[Z^2] < \infty$ . Alors la fonction  $g(d) = E[\|Z - d\|^2]$  atteint son minimum en  $d_* = E[Z]$ .*

Par conséquent la fonction  $\mathbb{E}\left[L(\theta, \delta(X)) \middle| x\right] = \mathbb{E}\left[\|\delta(X) - h(\theta)\|^2 \middle| x\right]$  atteint son minimum pour :  $\delta_*(x) = \mathbb{E}\left[h(\theta) \middle| x\right] = \int_{\Theta} h(\theta)\pi(\theta|x) d\nu(\theta)$ . C'est-à-dire, l'estimateur de Bayes est l'espérance *a posteriori* de  $h(\theta)$ .

**Exemple 1.4** *On suppose que  $X_1, X_2, \dots, X_n$  sont des vecteurs aléatoires i.i.d. de loi  $N_d(\theta, \sigma_X^2 I_d)$  où la variance  $\sigma_X^2$  est connue. On suppose de plus que la loi *a priori* de  $\theta$  est la loi  $N_d(\mu, \tau^2 I_d)$ , où la moyenne  $\mu$  et la variance  $\tau^2$  sont connues. La loi *a posteriori* est la loi  $N_d(\mu_* \sigma_*^2 I_d)$  avec  $\mu_* = \frac{\sigma_X^2 \mu + n\tau^2 \bar{X}}{\sigma_X^2 + n\tau^2}$  et  $\sigma_*^2 = \frac{\sigma_X^2 \tau^2}{\sigma_X^2 + n\tau^2}$ . L'estimateur bayésien de  $\theta$  sous la perte  $\|\delta - \theta\|^2$*

$$\delta_\pi(X) = \mathbb{E}[\theta | X] = \mu_* = \frac{\sigma_X^2 \mu + n\tau^2 \bar{X}}{\sigma_X^2 + n\tau^2}.$$

*Il s'agit d'une moyenne pondérée de la moyenne *a priori*  $\mu_0$  et de la moyenne de l'échantillon  $\bar{X}$  avec poids respectifs  $w = \frac{\sigma_X^2}{\sigma_X^2 + n\tau^2}$  et  $1 - w$ .*

Voici d'autres critères d'optimalité fréquentiste et les liens communs avec l'estimation bayésienne.

**Définition 1.5** (*Admissibilité et minimaxité*)

*Pour un problème d'estimation avec risque  $R(\theta; \delta) = E^{X|\theta} [L(\theta, \delta)]$  associé à la perte  $L(\theta, \delta)$ , un estimateur  $\delta_0$  :*

- domine un autre estimateur  $\delta$  si  $R(\theta, \delta_0) \leq R(\theta, \delta) \forall \theta \in \Theta$ , avec inégalité stricte pour au moins un  $\theta \in \Theta$ ;
- est inadmissible s'il existe un estimateur  $\delta_1$  qu'il le domine. Sinon,  $\delta_0$  est dit admissible ;
- est minimax s'il minimise le risque maximal en  $\theta$ , c'est-à-dire si

$$\sup_{\theta} R(\theta, \delta_0) = \inf_{\delta} \sup_{\theta} R(\theta, \delta).$$

**Théorème 1.6** (Lehmann et Casella, 1999) *Si un estimateur a un risque constant (ne dépend pas de  $\theta$ ) et s'il est Bayes ou admissible, alors il est aussi un estimateur minimax.*

**Exemple 1.7** *Pour  $X \sim N_d(\theta, \sigma^2 I_d)$ , avec  $L(d, \theta) = \|d - \theta\|^2$ , l'estimateur  $\delta(X) = X$  est minimax et Bayes pour la loi a priori  $\pi(\theta) = 1$ . Il est admissible pour  $d = 1, 2$ , mais inadmissible pour  $d \geq 3$  (Stein, 1957) et dominé par une grande classe d'estimateurs qui seront minimax, dont l'estimateur de James–Stein  $\delta_{JS}(X) = \left(1 - \frac{(d-2)\sigma^2}{X'X}\right) X$  et sa partie positive  $\delta_{JS}^+(X) = \max \left\{ 0, \left(1 - \frac{(d-2)\sigma^2}{X'X}\right)_+ X \right\}$ .*

### 1.1.2 Lois de probabilité hypergéométrique généralisée

Une grande classe de lois discrètes sur  $\mathbb{N}$  apparaissent comme des exemples de lois hypergéométriques généralisées et telles que présentées par Johnson, Kemp et Kotz (2005). Des lois usuelles telles Poisson, Binomiale Négative, des sous-familles de lois exponentielles y sont incluses. Ces fonctions de masse apparaissent à la section 1.3 de ce présent Chapitre et au Chapitre 4, ainsi que dans Marchand (2001) pour des mélanges de type HyperBêta.

**Définition 1.8** *Une fonction de masse sur  $\mathbb{N}$  est de type Hypergéométrique généralisée, avec paramètres  $a_1, \dots, a_p; b_1, \dots, b_q; \theta$  lorsqu'elle s'exprime comme :*

$$p_{\theta}(k) = \frac{\prod_i (a_i)_k \frac{\theta^k}{k!}}{{}_pF_q(a_1, \dots, q_p; b_1, \dots, b_q; \theta)}, \quad k = 0, 1, \dots$$

où les  $a_i \in \mathbb{R}$ , les  $b_j \in \mathbb{R}_+$  et  $\theta \in \Theta \subseteq (0, \infty)$ .

$(\alpha)_k$  est la fonction factorielle croissante ou symbole de Pochhammer définie par

$$(\alpha)_k = \alpha(\alpha + 1) \dots (\alpha + k - 1) = \frac{(\alpha + k - 1)!}{(\alpha - 1)!} \text{ pour } k \in \mathbb{N}_+, \quad (\alpha)_0 = 1,$$

et  ${}_pF_q$  la fonction hypergéométrique généralisée définie par

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z) = \sum_{k=0}^{\infty} \frac{\prod_{i=1}^p (a_i)_k z^k}{\prod_{j=1}^q (b_j)_k k!}, \quad (1.1)$$

convergente pour  $p \leq q$  et pour  $p = q + 1$  avec argument sur  $(0, 1)$  (voir Rainville, 1960).

On remarque que ces fonctions de masse pour  $a_i, b_j$  fixes forment une famille exponentielle avec paramètre  $\theta$ . De plus, la fonction génératrice d'une telle loi admet la forme simple :

$$\mathbb{E}(s^K) = \frac{{}_pF_q(a_1, \dots, q_p; b_1, \dots, b_q; s\theta)}{{}_pF_q(a_1, \dots, q_p; b_1, \dots, b_q; \theta)}.$$

On réfère à Johnson, Kemp et Kotz (2005) pour des propriétés, exemples et applications en lien avec ces lois de probabilité.

### 1.1.3 Lois de Bêta de type II et Gamma de type II

La thèse fait intervenir plusieurs lois. Il est utile ici de définir quelques unes, notamment les lois Bêta et Gamma de type II.

**Définition 1.9** On dit que  $X$  se distribue selon une loi Bêta de type II à trois paramètres, notée  $Beta_{II}(\alpha, \beta, \gamma)$ , si sa fonction de densité est de la forme

$$f(x) = \frac{\gamma^{-\alpha} x^{\alpha-1}}{B(\alpha, \beta) (1 + \frac{x}{\gamma})^{\alpha+\beta}} \mathbb{I}_{(0, \infty)}(x), \quad \alpha, \beta, \gamma > 0. \quad (1.2)$$

C'est la composition de deux lois Gamma dans le sens où

$$Beta_{II}(x; \alpha, \beta, \gamma) = \int_0^\infty Ga(x|z, \alpha, \frac{1}{z}) Ga(z|\beta, \gamma) dz. \quad (1.3)$$

Il s'agit d'une composition importante, et elle sera utilisée dans plusieurs des démonstrations et développements.

**Définition 1.10** On dit qu'une variable aléatoire  $X$  se distribue selon une loi Gamma de type II à quatre paramètres, que l'on notera  $Ga_{II}(\alpha, \beta, \epsilon, \nu)$ , si sa fonction de densité est de la forme

$$\frac{(x + \epsilon)^{\nu-1} x^{\alpha-1} e^{-\beta x}}{\Psi(\nu, \alpha, \epsilon, \beta)} \mathbb{I}_{(0, \infty)}(x), \quad (1.4)$$

avec  $\alpha, \beta > 0$ ,  $\nu \in \mathbb{R}$ ,  $\epsilon > 0$  et  $\Psi(\nu, \alpha, \epsilon, \beta) = \frac{\Gamma(\alpha)}{\beta^\alpha} E((Z + \epsilon)^{\nu-1})$ , où  $Z \sim Ga(\alpha, \frac{1}{\beta})$ .

**Remarque 1.11** a) Si  $\epsilon = 0$ ,  $Ga_{II}(\alpha, \beta, 0, \nu) = Ga(\alpha + \nu - 1, \beta)$ .

b) Si  $\nu = 1$ , on a  $Ga_{II}(\alpha, \beta, \epsilon, 1) = Ga(\alpha, \beta)$ .

c) Si  $\alpha = 1$ , alors  $X + \epsilon$  suit une loi Gamma  $Ga(\nu, \beta)$  tronquée sur  $[\epsilon, \infty[$ .

d) Si  $\epsilon = 1$  et  $\nu < 1 - \alpha$ , on a la loi de Kummer de type 2 de paramètre  $(\alpha, 1 - \alpha - \nu, \beta)$  (voir Ng et Kotz, 1995), avec  $\Psi(\alpha, 1 - \alpha - \nu, \beta) = \Gamma(\alpha) {}_1F_1(\alpha; \alpha + \nu; \beta)$ .

En général, on ne connaît pas la forme close pour  $\Psi(\nu, \epsilon, \alpha, \beta)$ . Par ailleurs, deux propriétés sont présentées ci-dessous. Les démonstrations s'effectuent par calculs directs.

**Propriété 1.12** a) Si  $\nu \in \mathbb{N}_+$ , un développement par le binôme de Newton mène à

l'expression 
$$\Psi(\nu, \epsilon, \alpha, \beta) = \frac{\epsilon^{\nu-1}}{\beta^\alpha} \sum_{k=0}^{\nu-1} C_k^{\nu-1} \frac{\Gamma(\alpha + k)}{(\beta\epsilon)^k}.$$

b) Le moment d'ordre  $m$  de  $X$ , est donnée par :

$$E(X^m) = \frac{\Psi(\nu, \epsilon, \alpha + m, \beta)}{\Psi(\nu, \epsilon, \alpha, \beta)}.$$

La loi Bêta de type *II* apparaîtra aux Chapitres 1, 2 et 4, tandis que la loi Gamma de type *II* apparaîtra au Chapitre 4.

## 1.2 Estimation de densités prédictives

Supposons que  $X \sim p_\theta(\cdot)$  et que l'on veuille prédire  $Y \sim q_\theta(\cdot)$ , avec  $p_\theta$  et  $q_\theta$  des densités de Lebesgue sur  $\mathbb{R}^d$  connues pour un  $\theta \in \Theta$  inconnu. Nous avons deux grandes approches statistiques dans le cadre paramétrique : l'approche fréquentiste et l'approche bayésienne. Dans ce qui suit, on utilise la notation  $\hat{q}_X(y)$ ,  $y \in \mathbb{R}^d$ , pour représenter une densité prédictive basée sur  $X$ .

Nous devons d'abord définir une fonction de perte à partir d'une divergence entre deux distributions. De préférence, il faudrait qu'elle soit définie de façon intrinsèque (voir la Définition 1.13 en bas).

### 1.2.1 Fonction de perte intrinsèque : Pertes $L^\alpha$ et $\alpha$ -divergence

**Définition 1.13** *Une fonction de perte intrinsèque est une fonction de perte définie à partir d'une divergence entre deux distributions, et invariante par transformation monotone inversible.*

Soit la fonction de perte  $f$ -divergence de la forme

$$L(\theta, \hat{q}) = \int_{\mathbb{R}^d} f\left(\frac{\hat{q}(y)}{q(y|\theta)}\right) g(q(y|\theta)) dy, \quad (1.5)$$

où  $\hat{q}$  est une densité prédictive pour  $q(\cdot|\theta)$ ,  $f$  est une fonction continue décroissante sur  $(0, 1)$  et croissante sur  $(1, \infty)$ , et  $g$  est une fonction positive. Cette classe inclut les pertes  $L^s$  intégrée et les pertes  $\alpha$ -divergence,  $|\alpha| \leq 1$ , qui englobe les pertes Kullback-Leibler, Kullback-Leibler inversée et Hellinger. En effet, les choix  $f(t) = |t - 1|^s$  et  $g(t) = t^{s+1}$ ,  $s > 0$ , mènent aux pertes  $L^s$  intégrées

$$L^s(q, \hat{q}) = \int_{\mathbb{R}^d} |q(y) - \hat{q}(y)|^s dy,$$

dont les cas importants  $L^2$  (coût carré intégrable) et  $L^1$  (coût valeur absolu).

En (1.5), on peut montrer que les choix  $g(t) = t$  mènent à des pertes intrinsèques.

### Perte $\alpha$ -divergence (Csiszar, 1963) avec $f$ définie en (1.6) et $g(t) = t$

La fonction de perte  $\alpha$ -divergence entre deux densités de Lebesgues,  $q$  et  $\hat{q}$ , est définie par

$$L_\alpha(q, \hat{q}) = \int_{\mathbb{R}^d} h_\alpha\left(\frac{\hat{q}(y)}{q(y)}\right) q(y) dy, \quad (1.6)$$

où

$$h_\alpha(z) = \begin{cases} \frac{4}{1-\alpha^2} \left(1 - z^{\frac{1+\alpha}{2}}\right) & |\alpha| < 1 \\ z \log z & \alpha = 1 \\ -\log z & \alpha = -1. \end{cases}$$

Deux cas particuliers sont :

- 1- La perte Kullback-Leibler (dite entropie) pour  $\alpha = -1$ , donnant

$$L_{KL}(q, \hat{q}) = \int_{\mathbb{R}^d} \log\left(\frac{q(y)}{\hat{q}(y)}\right) q(y) dy;$$

- 2- Le coût Hellinger pour  $\alpha = 0$ , donné par

$$L_H(q, \hat{q}) = 2 \int_{\mathbb{R}^d} \left(\sqrt{\frac{\hat{q}(y)}{q(y)}} - 1\right)^2 q(y) dy.$$

Même si les fonctions  $h_\alpha$  de la perte  $\alpha$ -divergence ne sont pas positives, ni décroissantes sur  $(0, 1)$  ni croissantes sur  $(1, \infty)$ , on peut voir que la perte  $L_\alpha$  est tout de même de la forme (1.5) avec

$$h_\alpha(z) = \begin{cases} \frac{4}{1-\alpha^2} \left( \frac{1+\alpha}{2} z - z^{\frac{1+\alpha}{2}} + \frac{1-\alpha}{2} \right) & |\alpha| < 1 \\ 1 - z + z \log(z) & \alpha = 1 \\ z - \log(z) - 1 & \alpha = -1. \end{cases} \quad (1.7)$$

En effet, dans chacun des cas  $h_\alpha$  ci-dessus, on a que  $h_\alpha$  décroît sur  $(0, 1)$  et croît sur  $(1, \infty)$  avec minimum atteint à 1. La Figure 1.1 suivant représente  $h_\alpha$  définie par (1.7) pour différentes valeurs de  $\alpha$ .

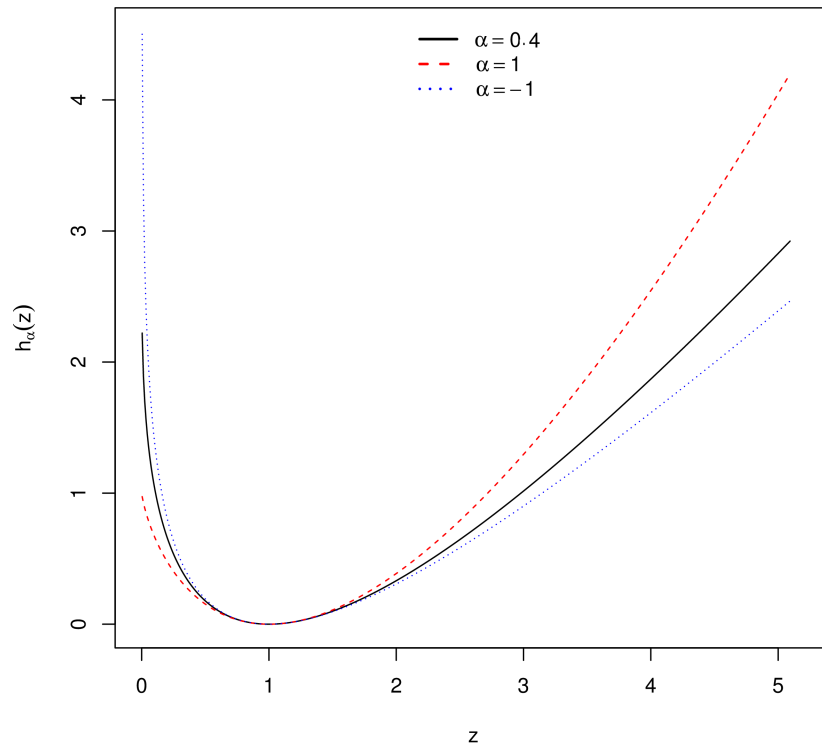


Figure 1.1 – Les fonctions  $h_\alpha$  pour différentes valeurs de  $\alpha$ .

Il existe une propriété d'invariance bayésienne liant les densités prédictives  $\hat{q}_\pi$  pour la densité  $q(\cdot|\theta)$  de  $Y$  aux densités prédictives  $\hat{q}_{g,\pi}$  pour la densité de  $q_g$  de  $g(Y)$  (voir Figure 1.2).

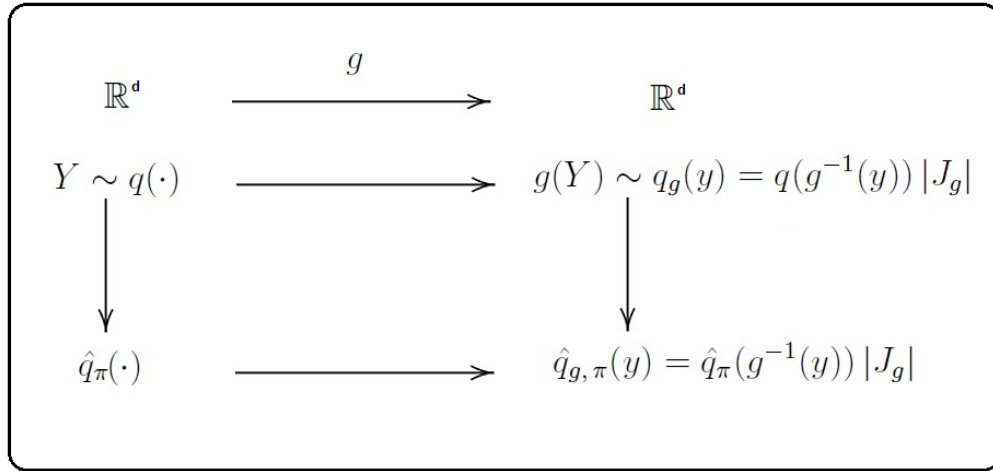


Figure 1.2 – Diagramme représentatif de la propriété d'invariance bayésienne liant les densités prédictives.

Soient  $y \sim q(y|\theta)$ ,  $y \in \mathbb{R}^d$ ,  $g$  une bijection différentiable avec jacobien  $|J_g|$  pour  $g^{-1}$  et une fonction de perte  $f$ -divergence, nous avons alors la propriété suivante.

**Théorème 1.14** *Pour une loi a posteriori  $\pi(\cdot|x)$  quelconque et pour une fonction de coût (1.5), on a*

$$\hat{q}_{g,\pi}(t; x) = \hat{q}_\pi(g^{-1}(t); x) |J_g|, \quad t \in \mathbb{R}^d,$$

*C'est-à-dire que  $\hat{q}_{g,\pi}$  correspond à la densité de  $g(T)$  où  $T \sim \hat{q}_\pi$ .*

**Démonstration.** Puisque  $\hat{q}_\pi$  est la densité prédictive bayésienne pour  $q(\cdot|\theta)$ , alors

$$\mathbb{E}^{\theta|x} \left[ \int_{\mathbb{R}^d} f \left( \frac{\hat{q}_\pi(y)}{q(y|\theta)} \right) q(y|\theta) dy \right] \leq \mathbb{E}^{\theta|x} \left[ \int_{\mathbb{R}^d} f \left( \frac{\hat{q}(y)}{q(y|\theta)} \right) q(y|\theta) dy \right], \quad \text{pour tout } \hat{q}$$



Donc,

$$\mathbb{E}^{\theta|x} \left[ \int_{\mathbb{R}^d} f \left( \frac{\hat{q}_\pi(g^{-1}(y))|J_g|}{q(g^{-1}(y)|\theta)|J_g|} \right) q(g^{-1}(y)|\theta)|J_g| dy \right] \leq \mathbb{E}^{\theta|x} \left[ \int_{\mathbb{R}^d} f \left( \frac{\hat{q}(g^{-1}(y))|J_g|}{q(g^{-1}(y)|\theta)|J_g|} \right) q(g^{-1}(y)|\theta)|J_g| dy \right],$$

ce qui implique que

$$\mathbb{E}^{\theta|x} \left[ \int_{\mathbb{R}^d} f \left( \frac{\hat{q}_{g,\pi}(t)}{q_g(t|\theta)} \right) q_g(t|\theta) dt \right] \leq \mathbb{E}^{\theta|x} \left[ \int_{\mathbb{R}^d} f \left( \frac{\hat{q}_g(t)}{q_g(t|\theta)} \right) q_g(t|\theta) dt \right], \text{ pour tout } \hat{q}_g.$$

Le principe d'invariance est vérifié pour les pertes  $\alpha$ -divergence et la perte  $L^1$ .

### Perte duale : pertes entropie, quadratique, et normale réfléchie

En étudiant le problème de l'estimation des densités prédictives, nous trouvons des résultats duaux entre l'estimation ponctuelle et l'estimation de densités prédictives. Voici un premier exemple.

**Exemple 1.15** Pour  $Y \sim Ga(\alpha_2, \beta)$  et  $q_{\hat{\beta}}(\cdot, X) \sim Ga(\alpha_2, \hat{\beta}(X))$ , on obtient avec la perte de Kullback-Leibler

$$\begin{aligned} L_{KL}(q_\beta, q_{\hat{\beta}}) &= \int_{\mathbb{R}_+} \log \left( \frac{q(y|\beta)}{q_{\hat{\beta}}(y;x)} \right) q(y|\beta) dy \\ &= \int_{\mathbb{R}_+} \left( y \left( \frac{1}{\hat{\beta}(x)} - \frac{1}{\beta} \right) + \alpha_2 \left( \log \left( \frac{\hat{\beta}(x)}{\beta} \right) \right) \right) q(y|\beta) dy \\ &= \alpha_2 \left( \frac{\beta}{\hat{\beta}(x)} - \log \left( \frac{\beta}{\hat{\beta}(x)} \right) - 1 \right) \\ &= L_0(\beta, \hat{\beta}), \end{aligned}$$

la perte entropie. Par conséquent, le risque Kullback-Leibler fréquentiste de l'estimateur de densité de plug-in  $q_{\hat{\beta}}(\cdot; X) \sim Ga(\alpha_2, \hat{\beta}(X))$  est égal au risque fréquentiste de l'estimateur ponctuel  $\hat{\beta}(X)$  de  $\beta$  sous la perte entropie  $L_0$ .

**Exemple 1.16** *La perte normale réfléchie (introduite par Spiring (1993) pour le cas unidimensionnel)*

$$L_\gamma(\theta, \hat{\theta}) = 1 - \exp\left(\frac{-\|\theta - \hat{\theta}\|^2}{2\gamma}\right); \quad \gamma > 0; \quad (1.8)$$

*survient comme perte duale pour les pertes  $\alpha$ -divergence et  $L^2$  pour le cas de la loi multinormale ( voir le Lemme 1.44).*

Maintenant, pour évaluer la performance d'une densité prédictive  $\hat{q}$  pour  $Y \sim q_\theta$  basée sur  $X \sim p_\theta$ , nous calculons le risque fréquentiste

$$R(\theta, \hat{q}) = E^X \left[ L\left(q_\theta(\cdot), \hat{q}(\cdot; X)\right) \right]. \quad (1.9)$$

Par exemple, pour la perte Kullback-Leibler, on obtient le risque

$$R_{KL}(\theta, \hat{q}) = E^{X,Y} \left[ \log \left( \frac{q(Y|\theta)}{\hat{q}(Y; X)} \right) \right]. \quad (1.10)$$

Ensuite, pour comparer deux densités  $\hat{q}_1$  et  $\hat{q}_2$ , la différence des risques est donnée par

$$\Delta(\theta) = R(\theta, \hat{q}_1) - R(\theta, \hat{q}_2);$$

et pour la perte Kullback-Leibler, nous avons

$$\Delta_{KL}(\theta) = R_{KL}(\theta, \hat{q}_1) - R_{KL}(\theta, \hat{q}_2) = E^{X,Y} \left[ \log \left( \frac{\hat{q}_2(Y; X)}{\hat{q}_1(Y; X)} \right) \right]. \quad (1.11)$$

### 1.2.2 Approche fréquentiste : Estimateur par substitution (*plug-in*)

Une méthode pratique d'estimation de densité prédictive est la méthode de substitution (plug-in en anglais) qui consiste à estimer le paramètre  $\theta$  de modèle à partir des observations  $X \sim p_\theta(\cdot)$ , par un estimateur ponctuel  $\hat{\theta}(X)$ , tel l'estimateur à vraisemblance maximum, pour générer  $q_{\hat{\theta}}$  comme estimateur de la densité  $q_\theta$  de  $Y$ .

**Définition 1.17** Si  $\hat{\theta}_{evm}$  est un estimateur du maximum de vraisemblance de  $\theta$ , alors

$$\hat{q}_{evm}(\cdot; X) = q_{\hat{\theta}_{evm}(X)}(\cdot)$$

est la densité prédictive du maximum de vraisemblance.

**Exemple 1.18** Soient  $X \sim Ga(\alpha_1, \beta)$  et  $Y \sim Ga(\alpha_2, \beta)$  avec  $\alpha_1$  et  $\alpha_2$  connus et positifs et  $\beta \in [a, b]$  inconnu. Un estimateur de plug-in pour  $q(y|\beta)$  est la densité prédictive

$$\hat{q}_c(\cdot; x) \sim Ga(\alpha_2, \hat{\beta}_c(x)), \text{ où } \hat{\beta}_c(x) = \begin{cases} a & \text{if } cx \leq a \\ cx & \text{if } a < cx < b \\ b & \text{if } cx \geq b \end{cases}$$

Ces densités prédictives sont intéressantes car elles incluent l'evm pour  $c = \frac{1}{\alpha_1}$ , la troncature du meilleur estimateur équivariant de  $\beta$  pour la perte entropie pour  $c = \frac{1}{\alpha_1 - 1}$ ,  $\alpha_1 > 1$ .

**Exemple 1.19** Pour un modèle de loi normale, on cherche à estimer la fonction de densité de  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$  à partir de  $X \sim N_d(\theta, \sigma_X^2 I_d)$  où les variances  $\sigma_X^2$  et  $\sigma_Y^2$  sont connues. Un estimateur de plug-in est  $q_{\hat{\theta}} \sim N_d(\hat{\theta}(x), \sigma_Y^2 I_d)$ . Par exemple, pour l'espace paramétrique  $\Theta = \mathbb{R}^d$ , on a  $q_{\hat{\theta}_{evm}} \sim N_d(X, \sigma_Y^2 I_d)$ . Plus généralement  $\hat{\theta}(X)$  peut-être un estimateur de Bayes, Bayes généralisé, de type James-Stein, ou un evm sur un espace paramétrique contraint.

## Expansion de la variance

Dans le cadre de l'exemple 1.19, Fourdrinier et coll. (2011) ( resp. Kubokawa, Marchand et Strawderman (2015, 2017)) ont proposé une méthode qui apporte une amélioration du risque par rapport à une densité prédictive *plug-in* pour la perte Kullback-Leibler (resp. pour les pertes  $L^2$  et  $L^1$  intégrées).

Des résultats analogues seront obtenus pour le modèle Gamma (Chapitre 2) et pour des pertes  $\alpha$ -divergence (Chapitre 3).

**Théorème 1.20** (*Fourdrinier et coll. 2011*)

Soit  $\hat{\theta}(X)$  un estimateur de  $\theta \in C \subset \mathbb{R}^d$ , tel que  $\underline{R} = \inf_{\theta \in C} \mathbb{E}^X(\|\hat{\theta}(X) - \theta\|^2) > 0$ . Alors,  $\hat{q}_{\hat{\theta},c} \sim N_d(\hat{\theta}(X), c\sigma_Y^2 I_d)$  améliore la densité plug-in  $\hat{q}_{\hat{\theta},1} \sim N_d(\hat{\theta}(X), \sigma_Y^2 I_d)$  sous la perte Kullback-Leibler si  $1 < c \leq (1 + \frac{\underline{R}}{d\sigma_Y^2})$ , et le domine ssi  $1 < c \leq c_0(1 + \frac{\underline{R}}{d\sigma_Y^2})$ , où  $c_0(z)$  est la racine sur  $]z, \infty[$  de  $h_z(c) = (1 - \frac{1}{c})z - \log(c)$ .

**Exemple 1.21** Pour  $\hat{\theta}_a(X) = aX$  ( $a \in (0, 1]$ ) un estimateur de  $\theta \in \mathbb{R}^d$ . Nous avons  $\mathbb{E}^X(\|aX - \theta\|^2) = a^2 d\sigma_X^2 + (1-a)^2 \|\theta\|^2$  qui atteint son minimum en  $\theta = 0$ , donc  $\underline{R} = a^2 d\sigma_X^2$ . Par conséquent  $\hat{q}_{\hat{\theta}_a,c}$  domine  $\hat{q}_{\hat{\theta}_a,1}$  sous la perte Kullback-Leibler si et seulement si  $c \in (1, c_0(1 + a^2 \frac{\sigma_X^2}{\sigma_Y^2})]$  ( $c_0 > 1$ ), l'évaluation donne  $\underline{R} = 0.219$  et  $c_0 = 1.0294$  pour  $a = 0.27$ ,  $d = 6$ ,  $\sigma_X^2 = 0.5$  et  $\sigma_Y^2 = 2.5$ . Pour  $a = 1$ , parmi les densités  $\hat{q}_{\hat{\theta}_1,c}$  un choix optimal est  $\hat{q}_{\hat{\theta}_1, 1 + \frac{\sigma_X^2}{\sigma_Y^2}} \sim N_d(X, (\sigma_X^2 + \sigma_Y^2) I_d)$  c'est à dire  $c^* = 1 + \frac{\sigma_X^2}{\sigma_Y^2}$ , résultat déjà connu comme meilleur densité équivariante (Murray, 1977).

Pour le cas Gamma, on a des résultats semblables au théorème 1.20 dans le Chapitre 2 avec des améliorations du type  $Ga\left(\frac{\alpha_2}{c}, c\hat{\beta}(X)\right)$  au  $Ga\left(\alpha_2, \hat{\beta}(X)\right)$  avec  $c > 1$ .

### 1.2.3 Approche bayésienne et meilleure densité prédictive équivariante

Pour  $X \sim p(x|\theta)$  et  $\theta \sim \pi(\theta)$ , on peut calculer la loi *a posteriori*  $\pi(\theta|x)$ . Ainsi, la densité de  $Y$  étant donnée  $X$ , lorsque  $X$  et  $Y$  sont conditionnellement indépendants,

$$\hat{q}_\pi(y|x) = \int_{\Theta} q(y|x, \theta) \pi(\theta|x) d\nu(\theta) = \int_{\Theta} q(y|\theta) \pi(\theta|x) d\nu(\theta) \quad (1.12)$$

est une densité prédictive plausible pour  $Y \sim q(y|\theta)$ , qui tient compte de l'a priori et de l'information apportée par  $X$ .

**Remarque 1.22** *La densité définie en (1.12) n'est pas nécessairement une règle de Bayes dépendamment de la fonction perte  $L$ , d'après la Définition 1.2 (voir par exemple le Théorème 1.26).*

Nous donnons deux exemples d'estimation de la densité prédictive à l'aide de formule (1.12) avec des lois a priori non informatives.

**Exemple 1.23** *Soit  $X_1, X_2, \dots, X_n, Y$  un échantillon i.i.d. issu de la loi  $\mathbb{U}_{(\theta, \theta+1)}$ , avec loi a priori  $\pi_0(\theta) = 1$  qui conduit à une loi a posteriori  $\theta | X \sim \mathbb{U}_{(x_{(n)}-1, x_{(1)})}$ . Un calcul donne*

$$\hat{q}_{mde}(y; x) = \begin{cases} y + 1 - x_{(n)} & \text{si } y \in [x_{(n)} - 1, x_{(1)}] \\ x_{(1)} - x_{(n)} + 1 & \text{si } y \in [x_{(1)}, x_{(n)}] \\ x_{(1)} + 1 - y & \text{si } y \in [x_{(n)}, x_{(1)} + 1] \end{cases}$$

qui est une densité prédictive de forme trapézoïdale pour  $Y$ , tel que  $X_{(1)} = \min\{X_i\}$  et  $X_{(n)} = \max\{X_i\}$ .

**Exemple 1.24** *Soient  $X \sim \text{Poisson}(\theta)$ ,  $Y \sim \text{Poisson}(\theta)$  et la loi a priori  $\pi_0(\theta) = \mathbb{I}_{[a, b]}(\theta)$ , avec  $a$  et  $b > 0$ . La densité prédictive par l'équation (1.12) est donnée par :*

$$\begin{aligned} \hat{q}_{\pi_0}(y; x) &= \frac{\int_a^b \frac{\theta^y e^{-\theta}}{y!} \frac{\theta^x e^{-\theta}}{x!} d\theta}{\int_a^b \frac{\theta^x e^{-\theta}}{x!} d\theta} \mathbb{I}_{\{0,1,2,\dots\}}(y) \\ &= \frac{\int_a^b \frac{2^{x+y+1} \theta^{x+y+1-1} e^{-2\theta}}{(x+y)!} d\theta}{\int_a^b \frac{\theta^x e^{-\theta}}{x!} d\theta} \frac{(x+y)!}{x! y!} \frac{1}{2^{x+y+1}} \mathbb{I}_{\{0,1,2,\dots\}}(y) \end{aligned}$$

$$\begin{aligned}
\hat{q}_{\pi_0}(y; x) &= \frac{\int_a^{2b} \frac{t^{x+y+1-1} e^{-t}}{\Gamma(x+y+1)} dt}{\int_a^b \frac{\theta^x e^{-\theta}}{\Gamma(x+1)} d\theta} \frac{(x+y)!}{x! y!} \frac{1}{2^{x+y+1}} \mathbb{I}_{\{0,1,2,\dots\}}(y) \\
&= \frac{F_{y+x+1}(2b) - F_{y+x+1}(2a)}{F_x(b) - F_x(a)} \binom{x+y}{y} \left(\frac{1}{2}\right)^{y+x+1} \mathbb{I}_{\{0,1,2,\dots\}}(y),
\end{aligned}$$

où  $F_\alpha(\cdot)$  est la fonction de répartition de loi  $Ga(\alpha, 1)$ . Si,  $a = 0$  et  $b = \infty$  (l'a priori  $\pi_0(\theta) = 1$ ), nous avons :

$$\hat{q}_{\pi_0}(y; x) = \binom{x+y}{y} \left(\frac{1}{2}\right)^{y+x+1} \mathbb{I}_{\{0,1,2,\dots\}}(y), \quad (1.13)$$

qui est la loi binomiale négative de paramètres  $x$  et  $p = \frac{1}{2}$  notée  $BiNég(x, \frac{1}{2})$ .

**Remarque 1.25** De plus, pour une loi a priori quelconque  $\pi(\theta)$  par rapport à la loi de Lebesgue sur  $\mathbb{R}_+$ , on montre que

$$\hat{q}_\pi(y; x) = \frac{m_\pi(x+y+1, 2)}{m_\pi(x+1, 1)} \hat{q}_{\pi_0}(y; x)$$

$$\text{avec } m_\pi(\alpha, \beta) = \int_{\mathbb{R}_+} \frac{\theta^{\alpha-1} \beta^\alpha}{\Gamma(\alpha)} e^{-\beta\theta} \pi(\theta) d\theta.$$

Maintenant, pour donner une application directe de la Définition 1.2 dans le cas de la perte  $\alpha$ -divergence définie en (1.6) et une loi a priori  $\pi$  absolument continue par rapport à la mesure de Lebesgue, nous avons besoin du résultat suivant.

**Théorème 1.26** (Corcuera et Giunmolè, 1999)

Pour  $X \sim p(x|\theta)$  et une loi a priori  $\pi$ , la densité prédictive de Bayes pour la densité  $q(y|\theta)$  sous le coût  $\alpha$ -divergence est donnée par :

$$\hat{q}_\pi(y; x) \propto \begin{cases} \left[ \int_{\mathbb{R}^d} q^{\frac{1-\alpha}{2}}(y|\theta) \pi(\theta|x) d\nu(\theta) \right]^{\frac{2}{1-\alpha}} & \text{si } \alpha \neq 1 \\ \exp\left(\int_{\mathbb{R}^d} \log[q(y|\theta)] \pi(\theta|x) d\nu(\theta)\right) & \text{si } \alpha = 1, \end{cases} \quad (1.14)$$

où  $\pi(\theta|x)$  est la densité a posteriori par rapport à une mesure  $\sigma$ -finie  $\nu$ .

**Démonstration.** Pour le cas où  $q(\cdot|\theta)$  est une densité de Lebesgue, on cherche  $\hat{q}$  qui minimise la perte espérée *a posteriori* donnée par :

$$\begin{aligned} R_\alpha(\theta, \hat{q}) &= \frac{4}{1-\alpha^2} \left( 1 - \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \hat{q}^{\frac{1+\alpha}{2}}(y; x) q^{\frac{1-\alpha}{2}}(y|\theta) \pi(\theta|x) d\nu(\theta) dy \right) \\ &= \frac{4}{1-\alpha^2} \left( 1 - \int_{\mathbb{R}^d} \hat{q}^{\frac{1+\alpha}{2}}(y; x) \underbrace{\left[ \int_{\mathbb{R}^d} q^{\frac{1-\alpha}{2}}(y|\theta) \pi(\theta|x) d\nu(\theta) \right]}_{h_\pi(y|x)} dy \right), \end{aligned}$$

ce qui revient à maximiser  $\int_{\mathbb{R}^d} \hat{q}^{\frac{1+\alpha}{2}}(y; x) h_\pi(y|x) dy$  par rapport à  $\hat{q}$ .

En vertu de l'inégalité de Hölder  $\int f_1 f_2 \leq (\int f_1^a)^{1/a} (\int f_2^b)^{1/b}$ , avec  $a = \frac{2}{1-\alpha}$ ,  $b = \frac{2}{1+\alpha}$  ( $1/a + 1/b = 1$ ),  $f_1 = h_\pi$  et  $f_2 = \hat{q}^{\frac{1+\alpha}{2}}$ , nous avons

$$\int_{\mathbb{R}^d} \hat{q}^{\frac{1+\alpha}{2}}(y; x) h_\pi(y|x) dy \leq \underbrace{\left[ \int_{\mathbb{R}^d} \hat{q}(y|\theta) dy \right]^{\frac{1+\alpha}{2}}}_{=1} \left[ \int_{\mathbb{R}^d} h_\pi^{\frac{2}{1-\alpha}}(y|x) dy \right]^{\frac{1-\alpha}{2}}.$$

Il est clair maintenant que la perte espérée *a posteriori* est majorée par  $\left[ \int_{\mathbb{R}^d} h_\pi^{\frac{2}{1-\alpha}}(y|x) dy \right]^{\frac{1-\alpha}{2}}$ , et atteint cette borne à  $h_\pi^{\frac{2}{1-\alpha}}(y|x)$ , et donc la densité prédictive bayésienne est

$$\hat{q}_\pi(y|x) \propto \left[ \int_{\mathbb{R}^d} q^{\frac{1-\alpha}{2}}(y|\theta) \pi(\theta|x) d\nu(\theta) \right]^{\frac{2}{1-\alpha}}. \quad (1.15)$$

Pour  $\alpha = 1$ , un passage semblable mène au résultat. Par ailleurs, la formule (1.15) nous

permet d'inférer

$$\begin{aligned}
\lim_{\alpha \rightarrow 1} \hat{q}_\pi(y; x) &\propto \lim_{\alpha \rightarrow 1} \exp\left(-2 \frac{\log \left[ \int_{\mathbb{R}^d} q^{\frac{1-\alpha}{2}}(y|\theta) \pi(\theta|x) d\nu(\theta) \right]}{\alpha - 1}\right) \\
&\propto \exp\left(-2 \frac{\partial \log \left[ \int_{\mathbb{R}^d} e^{\frac{1-\alpha}{2} \log[q(y|\theta)]} \pi(\theta|x) d\nu(\theta) \right]}{\partial \alpha} \Bigg|_{\alpha=1}\right) \\
&\propto \exp\left(\int_{\mathbb{R}^d} \log q[(y|\theta)] \pi(\theta|x) d\nu(\theta)\right).
\end{aligned}$$

**Remarque 1.27** Pour  $\alpha = -1$ , on trouve la densité donnée par l'équation (1.12) pour la perte Kullback-Leibler.

Dans le cas de la loi normale, sous la loi a priori uniforme  $\pi_0(\theta) = 1$ , la densité prédictive de Bayes est présentée dans le Lemme suivant. La démonstration de ce dernier est une application du Théorème 1.26.

**Lemme 1.28** Sous la perte (1.6), la densité prédictive bayésienne associée à la loi a priori uniforme sur  $\mathbb{R}^d$  ( $\pi_0(\theta) = 1$ ) est donnée par  $\hat{q}_{\pi_0, \alpha}(\cdot; X) \sim N_d\left(X, (\sigma_x^2 \left(\frac{1-\alpha}{2}\right) + \sigma_y^2) I_d\right)$ .

La forme exacte de la densité prédictive bayésienne  $\hat{q}_{\pi, \alpha}$  associée à la loi a priori quelconque  $\pi(\theta)$ , avec la normalisation, est réécrite dans le résultat suivant.

**Théorème 1.29** Sous la perte (1.6), la densité prédictive bayésienne associée à la loi a priori  $\pi(\theta)$  est donnée par

$$\hat{q}_{\pi, \alpha}(y; x) = \frac{m_\pi(\gamma x + (1-\gamma)y, \sigma_x^2 \gamma)^{\frac{2}{1-\alpha}}}{\mathbb{E}^Z[m_\pi(x + \zeta Z, \sigma_x^2 \gamma)^{\frac{2}{1-\alpha}}]} \hat{q}_{\pi_0, \alpha}(y; x)$$

$$\text{où } \gamma = \frac{1}{1 + \frac{1-\alpha}{2} \frac{\sigma_x^2}{\sigma_y^2}}, \quad \zeta = (1-\gamma) \left(\frac{\sigma_y^2}{\gamma}\right)^{1/2} \text{ et } Z \sim N(0, I_d).$$



**Démonstration.** Voir Maruyama et Ohnishi (2017).

Ce dernier résultat représente une généralisation du Lemme 2 de George, Liang et Xu, (2006). La différence des risques sous les fonctions coût (1.6) entre l'estimateur équivariant  $\hat{q}_{\pi_0, \alpha}$  et l'estimateur de Bayes  $\hat{q}_{\pi, \alpha}$ , associée à une loi *a priori* quelconque  $\pi$ , donne le résultat suivant.

**Corollaire 1.30** *Sous la perte (1.14), la différence des risques de l'estimateur équivariant  $\hat{q}_{\pi_0, \alpha}$  et l'estimateur de Bayes  $\hat{q}_{\pi, \alpha}$  est donnée par*

$$R_\alpha(\theta, \hat{q}_{\pi_0}) - R_\alpha(\theta, \hat{q}_\pi) = \frac{4}{1 - \alpha^2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left\{ \left( \frac{\hat{q}_{\pi, \alpha}(y; x)}{\phi\left(\frac{y - \theta}{\sigma_y}\right)} \right)^{\frac{1+\alpha}{2}} - \left( \frac{\hat{q}_{\pi_0, \alpha}(y; x)}{\phi\left(\frac{y - \theta}{\sigma_y}\right)} \right)^{\frac{1+\alpha}{2}} \right\} \\ \times \phi\left(\frac{y - \theta}{\sigma_y}\right) \phi\left(\frac{x - \theta}{\sigma_x}\right) dy dx. \quad (1.16)$$

Cependant, en utilisant le Corollaire 1.30 et le Théorème de Bolzano pour les fonctions continues sur un intervalle, on obtient le résultat suivant.

**Théorème 1.31** *Il existe un  $\alpha_0 \in (-1, 1)$ , tel que  $\forall \alpha \in (-1, \alpha_0)$  la densité prédictive bayésienne  $\hat{q}_{\pi_H, \alpha}$  associée à la loi *a priori* uniforme  $\pi_H(\theta) = \|\theta\|^{-(d-2)}$  ( $d > 2$  et  $\theta \in \mathbb{C}$  un compact dans  $\mathbb{R}^d$ ) domine l'estimateur équivariant  $\hat{q}_{\pi_0, \alpha}$  sous la perte (1.6).*

**Démonstration.** On pose  $\Delta(\alpha) = R_\alpha(\theta, \hat{q}_{\pi_0}) - R_\alpha(\theta, \hat{q}_{\pi_H})$ ,

qui est une fonction continue par rapport à la variable  $\alpha \in (-1, 1)$ .

De plus  $\Delta(-1) > 0$ , voir George et coll. (2006). Donc, s'il existe un  $\alpha_1$  tel que  $\Delta(\alpha_1) < 0$ , alors d'après le théorème de Bolzano, il existe au moins un  $\alpha_i \in (-1, \alpha_1)$  tel que  $\Delta(\alpha_i) = 0$ , on pose  $\alpha_0 = \min(\alpha_i)$  par conséquent,  $\Delta(\alpha) > 0$  pour tout  $\alpha \in (-1, \alpha_0)$ .

Les résultats s'étendent aussi à d'autres lois *a priori* données par George, Liang et Xu, (2006).

## 1.2.4 Quelques résultats pour la loi Gamma

Nous considérons le modèle Gamma,  $X \sim Ga(\alpha_1, \beta)$  et  $Y \sim Ga(\alpha_2, \beta)$ , avec paramètres  $\alpha_1, \alpha_2$  connus et  $\beta$  inconnu. Les densités prédictives de Bayes de  $p_Y(\cdot | \beta)$  associées aux lois *a priori*  $\pi_0(\beta) = \frac{1}{\beta} \mathbf{1}_{(0, \infty)}(\beta)$  et  $\pi(\beta)$  une loi *a priori* quelconque, sous le coût  $\delta$ -divergence ( $\delta$  à la place de  $\alpha$  pour ne pas confondre avec les  $\alpha$  de la loi Gamma), sont données par le Théorème 1.32. Ce dernier résultat dans le cas de Kullback-Leibler est donné par L'Moudden et coll. (2017) (voir Chapitre 2).

**Théorème 1.32** *Pour estimer la densité de  $Y \sim Ga(\alpha_2, \beta)$  à partir de  $X \sim Ga(\alpha_1, \beta)$  sous la perte  $\delta$ -divergence ( $\delta \neq 1$ ), les densités prédictives associées aux lois *a priori*  $\pi_0(\beta) = \frac{1}{\beta} \mathbf{1}_{(0, \infty)}(\beta)$  et  $\pi(\beta)$  (absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ ) sont données par :*

$$a) \hat{q}_{\pi_0, \delta}(y; x) = \frac{\Gamma\left(2 \frac{\alpha_\delta}{1-\delta}\right)}{\Gamma\left(2 \frac{\alpha_1}{1-\delta}\right) \Gamma(\alpha_2)} \left(\frac{1-\delta}{2} \frac{1}{x}\right) \left(\frac{1-\delta}{2} \frac{y}{x}\right)^{\alpha_2-1} \left(1 + \frac{1-\delta}{2} \frac{y}{x}\right)^{-2 \frac{\alpha_\delta}{1-\delta}} \mathbf{1}_{(0, \infty)}(y);$$

*c'est-à-dire une loi de Bêta de type II avec paramètres de formes  $\alpha_2$  et  $\frac{2\alpha_1}{1-\delta}$ , et paramètre d'échelle  $\frac{2x}{1-\delta}$ , avec  $\alpha_\delta = \alpha_1 + \frac{1-\delta}{2} \alpha_2$ .*

$$b) \hat{q}_{\pi, \delta}(y; x) = k_{\pi, \delta}(x) \hat{q}_{\pi_0, \delta}(y; x) \left[ \frac{z_\delta m(z_\delta, \pi, \alpha_\delta)}{x m_\pi(x | \alpha_1)} \right]^{\frac{2}{1-\delta}},$$

$$\text{où } z_\delta = x + y \frac{1-\delta}{2}, \quad m(z_\delta, \pi, \alpha_\delta) = \int_0^\infty \frac{z_\delta^{\alpha_\delta-1}}{\Gamma(\alpha_\delta) \beta^{\alpha_\delta}} \exp\left(-\frac{z_\delta}{\beta}\right) \pi(\beta) d\beta \text{ et}$$

$$k_{\pi, \delta}^{-1}(x) = \mathbf{E}^Z \left( \left[ \frac{\left(x + \frac{1-\delta}{2} Z\right) m\left(\left(x + \frac{1-\delta}{2} Z\right), \pi, \alpha_\delta\right)}{x m_\pi(x | \alpha_1)} \right]^{\frac{2}{1-\delta}} \right), \quad \text{avec } Z \sim \text{Beta}_{II}\left(\alpha_2, \frac{2\alpha_1}{1-\delta}, \frac{2x}{1-\delta}\right).$$

**Démonstration.** a) Pour  $\pi_0(\beta) = \frac{1}{\beta} \mathbf{1}_{(0, \infty)}(\beta)$  et  $x, y > 0$ , en se servant du Théorème 1.14 avec  $\delta \neq 1$

$$\text{et de l'identité } \int_0^\infty z^{-(a+1)} \exp\left(-\frac{d}{z}\right) dz = \Gamma(a) d^{-a}, \quad \forall a, d > 0,$$

nous avons

$$\begin{aligned}\hat{q}_{\pi_0, \delta}(y; x) &\propto y^{\alpha_2-1} \left[ \int_0^\infty \beta^{-(\alpha_1+\alpha_2\frac{1-\delta}{2}+1)} \exp\left(-\frac{x+y\frac{1-\delta}{2}}{\beta}\right) d\beta \right]^{\frac{2}{1-\delta}} \\ &\propto \frac{y^{\alpha_2-1}}{\left(x+y\frac{1-\delta}{2}\right)^{2(\alpha_1+\alpha_2\frac{1-\delta}{2})/(1-\delta)}},\end{aligned}$$

ce qui mène au résultat en a).

b) Maintenant, pour une densité a priori quelconque  $\pi(\beta)$  et  $x, y > 0$ , nous obtenons

$$\begin{aligned}\hat{q}_{\pi, \delta}(y|x) &\propto \frac{\Gamma(\alpha_\delta)^{\frac{2}{1-\delta}}}{\Gamma(\alpha_2)\Gamma(\alpha_1)^{\frac{2}{1-\delta}}} \left[ \frac{x^{\alpha_1-1}}{m_\pi(x|\alpha_1)} \right]^{\frac{2}{1-\delta}} y^{\alpha_2-1} z_\delta^{-2\frac{\alpha_\delta}{1-\delta}} \\ &\quad \times \left[ z_\delta \int_0^\infty \frac{z_\delta^{\alpha_\delta-1}}{\Gamma(\alpha_\delta)} \beta^{-\alpha_\delta} \exp\left(-\frac{z_\delta}{\beta}\right) \pi(\beta) d\beta \right]^{\frac{2}{1-\delta}} \\ &\propto \frac{\Gamma(\alpha_\delta)^{\frac{2}{1-\delta}}}{\Gamma(\alpha_1)^{\frac{2}{1-\delta}}} \frac{\Gamma(\alpha_1\frac{2}{1-\delta})}{\Gamma(\alpha_\delta\frac{2}{1-\delta})} \hat{q}_{\pi_0, \delta}(y|x) \left[ \frac{z_\delta m(z_\delta, \pi, \alpha_\delta)}{x m_\pi(x|\alpha_1)} \right]^{\frac{2}{1-\delta}} \\ &\propto \hat{q}_{\pi_0, \delta}(y|x) \left[ \frac{z_\delta m(z_\delta, \pi, \alpha_\delta)}{x m_\pi(x|\alpha_1)} \right]^{\frac{2}{1-\delta}},\end{aligned}$$

ce qui mène au résultat avec la constante de normalisation donnée  $k_{\pi, \delta}(x)$ .

**Exemple 1.33** Considérons la densité a priori  $\pi(\theta) = \frac{\gamma^{\alpha_3}}{\Gamma(\alpha_3)} \beta^{-(\alpha_3+1)} e^{-\frac{\gamma}{\beta}} \mathbb{I}_{\mathbb{R}_+}(\beta)$  (inverse-gamma). Du Théorème 1.32, la densité prédictive de Bayes est donnée par

$$\begin{aligned}\hat{q}_{\pi, \delta}(y|x) &\propto y^{\alpha_2-1} z_\delta^{-2\frac{\alpha_\delta}{1-\delta}} \left[ z_\delta \int_0^\infty \frac{z_\delta^{\alpha_\delta-1}}{\Gamma(\alpha_\delta)} \beta^{-\alpha_\delta} \exp\left(-\frac{z_\delta}{\beta}\right) \frac{\gamma^{\alpha_3}}{\Gamma(\alpha_3)} \beta^{-(\alpha_3+1)} e^{-\frac{\gamma}{\beta}} d\beta \right]^{\frac{2}{1-\delta}} \\ &\propto y^{\alpha_2-1} z_\delta^{-2\frac{\alpha_\delta}{1-\delta}} \left[ z_\delta^{\alpha_\delta} \int_0^\infty \exp\left(-\frac{z_\delta+\gamma}{\beta}\right) \beta^{-(\alpha_3+\alpha_\delta+1)} d\beta \right]^{\frac{2}{1-\delta}} \propto \frac{y^{\alpha_2-1}}{(\gamma+z_\delta)^{2\frac{(\alpha_3+\alpha_\delta)}{1+\delta}}} \\ &\propto \frac{y^{\alpha_2-1}}{(\gamma+x+y\frac{1-\delta}{2})^{2\frac{(\alpha_3+\alpha_1)}{1+\delta}+\alpha_2}} \propto \frac{y^{\alpha_2-1}}{\left(1+\frac{y}{\frac{2(\gamma+x)}{1-\delta}}\right)^{2\frac{(\alpha_3+\alpha_1)}{1+\delta}+\alpha_2}},\end{aligned}$$

c'est-à-dire une loi de Bêta de type II avec paramètres de formes  $\alpha_2$  et  $2\frac{(\alpha_3+\alpha_1)}{1+\delta}$ , et paramètre d'échelle  $\frac{2(\gamma+x)}{1-\delta}$ . La partie a) Théorème 1.32 est un cas particulier avec  $\alpha_3 = \gamma = 0$ .

De plus, pour  $\delta = -1$  nous trouvons l'Exemple 2.3 de Chapitre 2.

**Remarque 1.34** Par identification avec la partie a) en faisant  $\pi = \pi_0$ , on trouve  $k_{\pi, \delta}(x) = 1, \forall \delta$  et  $\forall x$ . De plus, on va montrer dans le Théorème 2.1 de Chapitre 2, que  $k_{\pi, \delta}(x) = 1$  pour  $\delta = -1, \forall x$  et pour toute loi a priori  $\pi$ .

Pour la suite, nous considérons des restrictions sur  $\beta$ , en supposant la loi a priori  $\pi_C(\beta) = \frac{1}{\beta} 1_{(a,b)}(\beta)$  avec  $a = 0, b < \infty, a > 0, b = \infty$  ou  $a > 0, b < \infty$ . Nous allons comparer les estimateurs de Bayes  $\hat{q}_{\pi, \delta}$  et  $\hat{q}_{\pi_0, \delta}$  sous  $L_\delta$  pour  $\delta \neq 1$ . Une application directe du Théorème 1.32, nous permet d'obtenir ce Corollaire.

**Corollaire 1.35** La densité prédictive de Bayes associée à la loi a priori  $\pi_C$  est donnée par

$$\hat{q}_{\pi_C, \delta}(y; x) = k_{\pi, \delta}(x) \hat{q}_{\pi_0, \delta}(y | x) \left[ \frac{F_{\alpha_\delta} \left( \frac{z_\delta}{a} \right) - F_{\alpha_\delta} \left( \frac{z_\delta}{b} \right)}{F_{\alpha_1} \left( \frac{x}{a} \right) - F_{\alpha_1} \left( \frac{x}{b} \right)} \right]^{\frac{2}{1-\delta}},$$

où  $z_\delta = x + y \frac{1-\delta}{2}$ ,  $\alpha_\delta = \alpha_1 + \alpha_2 \frac{1-\delta}{2}$  et  $F_\alpha$  la fonction de répartition de loi  $Ga(\alpha, 1)$ .

Pour les cas avec contrainte  $\beta \in [a, b]$ , il est d'intérêt d'évaluer la performance fréquentiste de différents modèles prédictives et notamment de comparer  $\hat{q}_{\pi_C}$ , qui tient compte de la contrainte. D'ailleurs, il sera démontré avec la densité repère  $\hat{q}_{\pi_0}$  que cette dernière est minimax pour la perte Kullback-Leibler et pour  $a = 0$  ou  $b = \infty$ .

**Conjecture 1.1** La densité prédictive  $\hat{q}_{\pi_C}$  domine  $\hat{q}_{\pi_0}$  sous la perte  $\delta$ -divergence (1.6) avec  $\beta \in C = (a, b)$  et basée sur  $X \sim Ga(\alpha_1, \beta)$ .

Deux observations immédiates sont recevables. Numériquement, la conjecture est vérifiée pour quelques cas. La figure (1.3), par exemple, montre la dominance de  $\hat{q}_{\pi_C}$  par rapport

à  $\hat{q}_{\pi_0}$  sous le coût Hellinger  $\delta = 0$  pour  $\alpha_1 = 1$ ,  $\alpha_2 = 2$  et  $\beta \in [1, 4]$ . Au Chapitre 2, on va montrer que cette conjecture est vraie pour la perte Kullback-Leibler ( $\delta = -1$ ).<sup>1</sup>

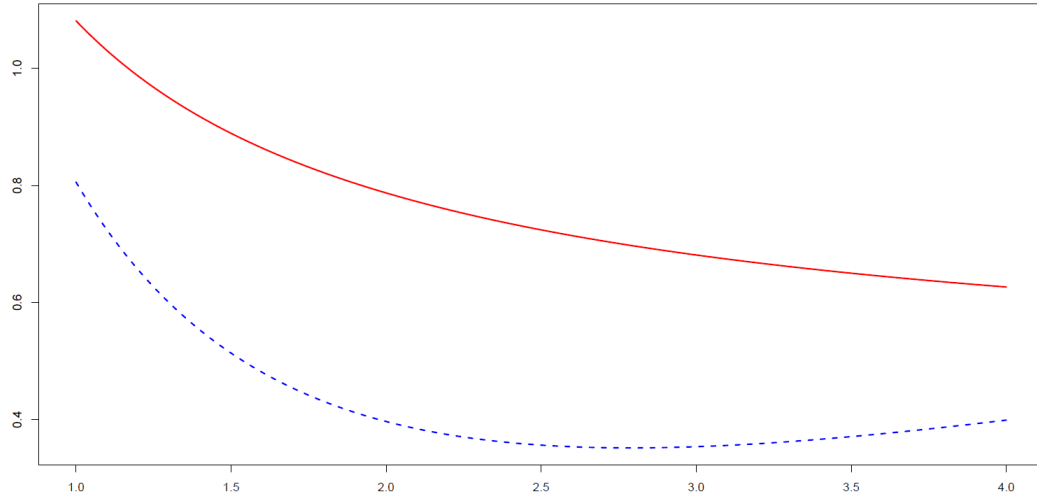


Figure 1.3 – Risques des densité prédictives  $\hat{q}_{\pi_0}$  en rouge et  $\hat{q}_{\pi_C}$  ( $C = [1, 4]$ ) en bleu pour  $\delta = 0$ ,  $\alpha_1 = 1$  et  $\alpha_2 = 2$ .

Dans la section 2.3 du Chapitre 2, on va examiner les graphes des risques (par exemple, la Figure 2.4), on remarque que les risques fréquentistes aux points d’extrémité de l’espace paramétrique  $C = [a, b]$  coïncident, c’est-à-dire

$$R_{KL}(a, \hat{q}_{\pi}) = R_{KL}(b, \hat{q}_{\pi}), \quad (1.17)$$

ce qui est équivalent à

$$\mathbb{E}^T \left( \log \left[ F_{\alpha}(T) - F_{\alpha}\left(\frac{aT}{b}\right) \right] \right) = \mathbb{E}^T \left( \log \left[ F_{\alpha}\left(\frac{bT}{a}\right) - F_{\alpha}(T) \right] \right);$$

voir la démonstration du Théorème 2.9 au Chapitre 2.

Ou, plus généralement  $\mathbf{E}[\log(F(X) - F(cX))] = \mathbf{E}[\log(F(\frac{X}{c}) - F(X))]$  pour  $0 < c < 1$

---

1. Le résultat pour  $\delta = 1$  semble acquis à l’aide de la représentation des densité de Bayes qui sont des plug-in pour les familles exponentielles (Yanagimoto et Ohnishi, (2009)), la dualité avec la perte entropie (inverse) et les résultats de Marchand et Strawderman (2005, Journal of Statistical Planning and Inference). Le problème clé est équivalent à comparer la performance de  $\hat{q}_{\pi_0}$  et  $\hat{q}_{\pi}$  où  $\pi_0(\theta) = \frac{1}{\beta} \mathbb{I}_{(0, \infty)}(\beta)$  et  $\pi(\theta) = \frac{1}{\beta} \mathbb{I}_{(a, b)}(\beta)$ .

et  $X \sim Ga(\alpha; 1)$  tel que démontré par Jones, Marchand et Strawderman (2018). Ces derniers ont obtenu des résultats beaucoup plus généraux que la Proposition 1.36 en bas.

**Proposition 1.36** *Pour une variable aléatoire continue et positive  $X$  avec fonction de répartition  $F$ , les variables aléatoires  $F(X) - F(cX)$  et  $F(\frac{X}{c}) - F(X)$  sont équidistribuées pour  $c > 0$ .*

Voici un résultat plus général qui inclut la Proposition 1.36 et qui confirme l'égalité (1.17) comme cas particuliers.

**Théorème 1.37** *(Jones et coll, (2018)) Soit  $X$  une variable aléatoire de densité continue, avec fonction de répartition  $F$  strictement croissante sur son support  $(a, b)$ . Soit  $g : (a, b) \rightarrow (a, b)$  une fonction monotone croissante et inversible. Soient  $H_1(X) = F(X) - F(g(X))$  et  $H_2(X) = F(g^{-1}(X)) - F(X)$ . Alors  $H_1(X)$  et  $H_2(X)$  ont des distributions identiques.*

**Démonstration.**<sup>2</sup>

Nous montrons que  $\mathbf{P}(H_1(X) \geq t) = \mathbf{P}(H_2(X) \geq t)$  pour tout  $t \in \mathbf{R}$ . Fixons  $t$  et posons

$$A = \{x : H_1(x) \geq t\} = \bigcup_{n \geq 1} [a_n, b_n]$$

où  $a_k = b_k$  pour tout  $k > \frac{1}{2}$  le nombre de fois que  $H_1$  dépasse  $t$ . De même, considérons  $B = \{x : H_2(x) \geq t\}$  et observons que  $x \in A \Leftrightarrow g(x) \in B$  impliquant que

$$B = \bigcup_{n \geq 1} [g(a_n), g(b_n)].$$

Il s'ensuit que

$$\mathbf{P}(H_1(X) \geq t) = \sum_{n \geq 1} F(b_n) - F(a_n) \text{ et } \mathbf{P}(H_2(X) \geq t) = \sum_{n \geq 1} F(g(b_n)) - F(g(a_n)). \quad (1.18)$$

---

2. Alternativement : Comme  $H_1(X)$  et  $H_2(X)$  sont des variables aléatoires bornées, il suffit de voir que  $E(u^{H_1(X)}) = E(u^{H_2(X)})$  pour tout  $u > 0$ .

Enfin, puisque  $H_1(a_n) = H_1(b_n)$  par définition des  $a_n$  et  $b_n$ , on obtient que  $F(b_n) - F(a_n) = F(g(b_n)) - F(g(a_n))$  et l'égalité de  $\mathbf{P}(H_1(X) \geq t)$  et  $\mathbf{P}(H_2(X) \geq t)$  suit de (1.18), pour tout  $t$ .

## 1.2.5 Lien entre l'estimation d'une densité prédictive et l'estimation ponctuelle

Nous passons en revue différents résultats de dualité entre le problème d'estimation de la densité prédictive et le problème d'estimation ponctuelle de  $\theta$ . Par exemple, nous présentons des parallèles et liens très précis entre la performance sous le risque Kullback-Leibler et une collection de problèmes d'estimation ponctuelle.

### Cas de la loi double exponentielle sous la perte d'erreur absolue $L^1$

**Théorème 1.38** *Soit  $Y \sim \mathcal{L}(\theta)$ , une variable aléatoire de loi double exponentielle ou de Laplace avec densité  $q_\theta(y) = \frac{1}{2} e^{-|y-\theta|}$  pour  $\theta, y \in \mathbb{R}$ . La perte Hellinger associée à la densité plug-in  $q_{\hat{\theta}}$  est donnée par  $2\rho(|\hat{\theta} - \theta|)$  avec  $\rho(t) = 1 - e^{-\frac{t}{2}}(1 + \frac{t}{2})$ ,  $t \geq 0$ . Par conséquent pour  $X \sim q_\theta$ , la densité  $q_{\hat{\theta}_1(X)}$  domine  $q_{\hat{\theta}_2(X)}$  sous la perte Hellinger si et seulement si  $\hat{\theta}_1(X)$  domine  $\hat{\theta}_2(X)$  comme estimateur de  $\theta$  sous la perte*

$$L(\theta, \hat{\theta}) = 1 - e^{-\frac{|\hat{\theta} - \theta|}{2}} \left( 1 + \frac{|\hat{\theta} - \theta|}{2} \right).$$

**Démonstration.** Sans perte de généralité, on suppose que  $\theta < \hat{\theta}$ . On a

$$\begin{aligned} \int_{\mathbb{R}} \left( \sqrt{q_\theta(y)} - \sqrt{q_{\hat{\theta}}(y)} \right)^2 dy &= 2 - \int_{\mathbb{R}} e^{-\frac{1}{2}(|y-\theta|+|y-\hat{\theta}|)} dy \\ &= 2 - \int_{-\infty}^{\theta} e^y e^{-\frac{\theta+\hat{\theta}}{2}} dy - \int_{\theta}^{\hat{\theta}} e^{-\frac{\hat{\theta}-\theta}{2}} dy - \int_{\hat{\theta}}^{\infty} e^{-y} e^{\frac{\theta+\hat{\theta}}{2}} dy \\ &= 2 - e^{-\frac{\hat{\theta}+\theta}{2}} e^\theta - (\hat{\theta} - \theta) e^{-\frac{\hat{\theta}-\theta}{2}} - e^{-\hat{\theta}} e^{\frac{\theta+\hat{\theta}}{2}} \\ &= 2 - e^{-\frac{\hat{\theta}-\theta}{2}} (2 + \hat{\theta} - \theta). \end{aligned}$$

## Cas d'une loi normale multivariée sous le risque quadratique

**Théorème 1.39** Soient  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$ ,  $c > 0$ ,  $\hat{\theta}_2(X)$  et  $\hat{\theta}_1(X)$  des estimateurs pour  $\theta \in \mathbb{C} \subseteq R^d$ . Alors, pour estimer la densité de  $Y$ , la densité plug-in  $q_{\hat{\theta}_2, c} \sim N_d(\hat{\theta}_2(X), c\sigma_Y^2 I_d)$  domine la densité plug-in  $q_{\hat{\theta}_1, c} \sim N_d(\hat{\theta}_1(X), c\sigma_Y^2 I_d)$  sous le critère du risque de Kullback-Leibler  $R_{KL}$  ssi  $\hat{\theta}_2(X)$  domine  $\hat{\theta}_1(X)$  comme estimateur de  $\theta$  sous le critère du risque quadratique  $R_Q$ .

**Démonstration.** Le résultat découle directement de la décomposition

$$\begin{aligned} R_{KL}(\theta, q_{\hat{\theta}_2, c}) - R_{KL}(\theta, q_{\hat{\theta}_1, c}) &= \mathbb{E}^{X,Y} \left[ \log \left( \frac{q_{\hat{\theta}_1, c}(Y; X)}{q_{\hat{\theta}_2, c}(Y; X)} \right) \right] = \mathbb{E}^{X,Y} \left[ \log \left( e^{\frac{-\|Y - \hat{\theta}_1(X)\|^2}{2c\sigma_Y^2}} \right) \right] \\ &= \mathbb{E}^{X,Y} \left[ \frac{\|Y - \hat{\theta}_2(X)\|^2 - \|Y - \hat{\theta}_1(X)\|^2}{2c\sigma_Y^2} \right] \end{aligned} \quad (1.19)$$

$$\begin{aligned} &= \mathbb{E}^X \left[ \frac{\|\hat{\theta}_2(X) - \theta\|^2 - \|\hat{\theta}_1(X) - \theta\|^2}{2c\sigma_Y^2} \right] \quad (1.20) \\ &= \frac{R_Q(\theta, \hat{\theta}_2) - R_Q(\theta, \hat{\theta}_1)}{2c\sigma_Y^2}. \end{aligned}$$

On a bien que (1.19) = (1.20), car  $X$  et  $Y$  sont indépendants ce qui implique que

$$\mathbb{E}^{X,Y} \left[ \|Y - \hat{\theta}(X)\|^2 \right] = \mathbb{E}^Y \left[ \|Y - \theta\|^2 \right] + \mathbb{E}^X \left[ \|\hat{\theta}(X) - \theta\|^2 \right].$$

Le prochain résultat est tiré de George, Liang et Xu (2006).

**Lemme 1.40** Soient  $X \sim N_d(\theta, \sigma_X^2 I_d)$  et  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$  deux vecteurs aléatoires indépendants. Alors la densité prédictive de Bayes  $\hat{q}_\pi$  associée à une loi a priori quelconque  $\pi$  s'écrit :

$$\hat{q}_\pi(y; x) = \frac{m_\pi(W, \sigma_W^2)}{m_\pi(X, \sigma_X^2)} \hat{q}_U(y; x)$$



où  $W = \frac{(X\sigma_Y^2 + Y\sigma_X^2)}{\sigma_X^2 + \sigma_Y^2}$ ,  $\sigma_W^2 = \frac{\sigma_X^2\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}$ ,  $m_\pi(W; \sigma_W^2)$  et  $m_\pi(X; \sigma_X^2)$  sont les densités marginales de  $W$  et de  $X$  sous  $\pi$ , et  $\hat{q}_U(y; X)$  est la densité prédictive sous  $\pi(\theta) = 1$  donnée par une densité  $N_d\left(X, (\sigma_X^2 + \sigma_Y^2) I_d\right)$  (George, Liang et Xu (2006)).

**Corollaire 1.41** Nous avons pour les densités prédictives  $\hat{q}_\pi$  et  $\hat{q}_U$  la différence de risques

$$R_{KL}(\theta, \hat{q}_U) - R_{KL}(\theta, \hat{q}_\pi) = \mathbb{E}^{X,Y} \left[ \log \left( \frac{\hat{q}_\pi(Y; X)}{\hat{q}_U(Y; X)} \right) \right] = \mathbb{E}^{X,W} \left[ \log \left( \frac{m_\pi(W, \sigma_W^2)}{m_\pi(X, \sigma_X^2)} \right) \right]$$

De plus, la différence des risques de ces deux densités peut s'écrire sous la forme donnée dans le Lemme 1.42 suivante.

**Lemme 1.42** (Brown, George, Xu, 2008)

$$\mathbb{E}^{X,W} \left[ \log \left( \frac{m_\pi(W, \sigma_W^2)}{m_\pi(X, \sigma_X^2)} \right) \right] = \frac{1}{2} \int_{\sigma_W^2}^{\sigma_X^2} \frac{1}{\sigma^2} (R_Q^{\sigma^2}(\theta, X) - R_Q^{\sigma^2}(\theta, \hat{\theta}_{\pi, \sigma^2}(X))) d\sigma^2 \quad (1.21)$$

où  $\hat{\theta}_{\pi, \sigma^2}(X)$  est Bayes sous le critère du risque quadratique  $R_Q$  et pour la loi a priori  $\pi$ .

On voit que la différence des risques  $R_{KL}$  est associée à une collection de différences de risques  $R_Q$ . On voit notamment que  $\hat{q}_\pi$  domine  $\hat{q}_U$  lorsque  $\hat{\theta}_{\pi, \sigma^2}$  domine  $X$  sous le critère du risque quadratique  $R_Q$  pour tout  $\sigma^2 \in (\sigma_W^2, \sigma_X^2)$ . Des applications sont données par George, Liang et Xu (2006), Brown, George et Xu (2008), ainsi que Fourdrinier et coll. (2011).

**Exemple 1.43** Prenons le cas de la loi normale unidimensionnelle, et considérons la densité a priori  $\pi(\theta) = \mathbb{I}_{\mathbb{R}_+}(\theta)$ . Du Lemme 1.40, la densité prédictive de Bayes est donnée

$$\hat{q}_\pi(y; x) = \frac{\Phi \left( \left( \frac{\sigma_Y^2 x + \sigma_X^2 y}{\sigma_X^2 + \sigma_Y^2} \right) / \sqrt{\frac{\sigma_X^2 \sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}} \right)}{\Phi \left( \frac{x}{\sigma_X} \right)} \phi \left( \frac{y - x}{\sigma_X^2 + \sigma_Y^2} \right)$$

avec  $\phi$  et  $\Phi$  sont les fonctions de densité et de répartition de la loi normale  $N(0, 1)$ .

De tels résultats pour le coût  $\alpha$ -divergence ( $-1 < \alpha < 1$ ) ne sont pas connus. Mais voir quand même un résultat de dualité mettent en lien le coût  $\alpha$ -divergence et la perte normale réfléchie.

**Lemme 1.44** *Soient  $c > 0$ ,  $\hat{\theta}_2(X)$  et  $\hat{\theta}_1(X)$  des estimateurs pour  $\theta \in \mathbb{C} \subseteq \mathbb{R}^d$ . Pour estimer la densité de  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$ , la densité  $q_{\hat{\theta}_2, c} \sim N_d(\hat{\theta}_2(X), c \sigma_Y^2 I_d)$  domine l'estimateur plug-in  $q_{\hat{\theta}_1, c} \sim N_d(\hat{\theta}_1(X), c \sigma_Y^2 I_d)$  sous la perte  $\alpha$ -divergence  $L_\alpha$  ( $-1 < \alpha < 1$ ) ssi  $\hat{\theta}_2(X)$  domine  $\hat{\theta}_1(X)$  comme estimateur de  $\theta$  sous le perte normale réfléchie  $L_{\gamma_0}$  donnée par (1.8), avec  $\gamma_0 = 2(\frac{c^2}{1+\alpha} + \frac{1}{1-\alpha})\sigma_Y^2$ .*

**Démonstration.** On se base sur le Lemme 2.2 de Kubokawa et coll. (2017). On a

$$L_\alpha(\theta, \hat{q}_{\hat{\theta}, c}) = \frac{4}{1-\alpha^2} \left( 1 - \left( \frac{2 c^{1-\alpha}}{(1-\alpha)c^2 + 1 + \alpha} \right)^{\frac{d}{2}} \exp \left( -\frac{(1-\alpha^2)\|\hat{\theta}(X) - \theta\|^2}{4[(1-\alpha)c^2 + 1 + \alpha]\sigma_Y^2} \right) \right) \quad (1.22)$$

$$= \frac{4}{1-\alpha^2} \left( 1 - b(c) + b(c) L_{\gamma_0}(\theta, \hat{\theta}) \right),$$

avec  $b(c) = \left( \frac{2 c^{1-\alpha}}{(1-\alpha)c^2 + 1 + \alpha} \right)^{\frac{d}{2}}$  et  $\gamma_0 = 2\left(\frac{c^2}{1+\alpha} + \frac{1}{1-\alpha}\right)\sigma_Y^2$ . On exploitera cette identité au Chapitre 3.

Puisque  $\hat{\theta}_\pi(X)$  domine  $X$  pour  $\theta \geq 0$  sous le perte normale réfléchie  $L_{\gamma_0}$  (1.8),  $\pi_U(\theta) = \mathbb{I}_{\mathbb{R}_+}(\theta)$  et  $X \sim N_d(\theta, \sigma_X^2)$ , il découle du Lemme 1.42 que  $\hat{q}_{\pi_U}$  domine  $\hat{q}_U \sim N_d(X, (\sigma_X^2 + \sigma_Y^2) I_d)$ . Voir aussi Kubokawa et coll. (2013) pour des modèles plus généraux avec paramètres de position et d'échelle de tels résultats de dominance.

Un lien additionnel, donné par Kubokawa, Marchand et Strawderman (2015) permet plus de résultats communs pour la perte quadratique.

**Lemme 1.45** *Soit  $X \sim N_d(\theta, \sigma_X^2 I_d)$  avec  $\hat{\theta}(X)$  un estimateur de  $\theta$ . Alors,  $\hat{\theta}(X)$  domine  $X$  sous une perte normale réfléchie  $L_\gamma$  définie par (1.8), si  $\hat{\theta}(Z)$  domine  $Z$  pour  $Z \sim N_d(\theta, \frac{\gamma \sigma_X^2}{\gamma+1} I_d)$  sous la perte quadratique.*

**Corollaire 1.46** *Sous la perte  $\alpha$ -divergence, la densité plug-in  $\frac{1}{c\sigma_Y}\phi\left(\frac{y-\hat{\theta}(X)}{c\sigma_Y}\right)$  domine l'estimateur plug-in  $\frac{1}{c\sigma_Y}\phi\left(\frac{y-X}{c\sigma_Y}\right)$ , si  $\hat{\theta}(Z)$  domine  $Z$  pour  $Z \sim N_d(\theta, \frac{\gamma_0 \sigma_X^2}{\gamma_0+1} I_d)$  sous la perte quadratique, avec  $\gamma_0 = \frac{2((1+\alpha)c^2+1-\alpha)\sigma_Y^2}{1-\alpha^2}$ .*

**Démonstration.** Sous la perte  $\alpha$ -divergence, on a du Lemme 1.44

$$\begin{aligned}
\Delta_\alpha(\theta) &= R_{L_\alpha}(\theta, \hat{q}_{\hat{\theta}(X)}) - R_{L_\alpha}(\theta, \hat{q}_X) \\
&= \frac{4}{1-\alpha^2} \left( \frac{2 c^{1+\alpha}}{(1+\alpha)c^2+1-\alpha} \right)^{\frac{d}{2}} \mathbb{E}^X \left( \exp\left(-\frac{\|X-\theta\|^2}{2\sigma_X^2\gamma_0}\right) - \exp\left(-\frac{\|\hat{\theta}(X)-\theta\|^2}{2\sigma_X^2\gamma_0}\right) \right) \\
&= \frac{4}{1-\alpha^2} \left( \frac{2 c^{1+\alpha}}{(1+\alpha)c^2+1-\alpha} \right)^{\frac{d}{2}} \\
&\quad \mathbb{E}^X \left( \exp\left(-\frac{\|X-\theta\|^2}{2\sigma_X^2\gamma_0}\right) \left[ 1 - \exp\left(-\frac{\left(\|\hat{\theta}(X)-\theta\|^2 - \|X-\theta\|^2\right)}{2\sigma_X^2\gamma_0}\right) \right] \right) \\
&= \frac{4}{1-\alpha^2} \left( \frac{2 c^{1+\alpha}}{(1+\alpha)c^2+1-\alpha} \right)^{\frac{d}{2}} \mathbb{E}^Z \left( 1 - \exp\left(-\frac{\left(\|\hat{\theta}(Z)-\theta\|^2 - \|Z-\theta\|^2\right)}{2\sigma_X^2\gamma_0}\right) \right).
\end{aligned}$$

Avec  $Z \sim N_d(\theta, c \frac{\gamma_0 \sigma_X^2}{\gamma_0} I_d)$ . Et puisque  $1 - \exp(-t) < t, \forall t \neq 0$ , on obtient

$$\Delta_\alpha(\theta) < \frac{4}{1-\alpha^2} \left( \frac{2 c^{1+\alpha}}{(1+\alpha)c^2+1-\alpha} \right)^{\frac{d}{2}} \mathbb{E}^Z \left( \frac{\|\hat{\theta}(Z)-\theta\|^2 - \|Z-\theta\|^2}{2\sigma_X^2\gamma_0} \right).$$

### 1.3 Une stratégie d'évaluation des risques

Dans le chapitre 3, nous présentons, des évaluations numériques de fonctions de risques et des graphiques pour illustrer la théorie. Ces évaluations ne sont pas basées sur les méthodes de type Monte-Carlo, mais plutôt sur une décomposition de risque conditionnel permettant l'évaluation du risque avec une simple intégration numérique

sur  $\mathbb{R}_+$ . La stratégie d'évaluation exploite aussi la représentation de la densité d'un loi  $\chi^2$  décentrée à l'aide de la fonction hypergéométrique  ${}_0F_1$ .

Considérons  $X \sim N_d(\theta, \sigma_x^2 I_d)$ , une perte ou une fonction quelconque  $\rho(\|\hat{\theta} - \theta\|^2)$  et des estimateurs équivariants par rapport aux transformations orthogonales de la forme  $\hat{\theta}_h(X) = h(\|X\|^2)X$ , où  $h(\cdot)$  est une fonction quelconque mesurable sur  $\mathbb{R}_+$ . Ces estimateurs incluent ceux de type Baranchik,  $\hat{\theta}(X) = \left(1 - \frac{r(\|X\|^2)}{\|X\|^2}\right) X$ , ici  $h(t) = \left(1 - \frac{r(t)}{t}\right)$ , de type James-Stein  $h(t) = \left(1 - \frac{(d-2)\sigma_x^2}{t}\right)$  et sa partie positive  $h(t) = \max\{0; \left(1 - \frac{(d-2)\sigma_x^2}{t}\right)\}$ . On rappelle que pour  $X \sim N_d(\theta, \sigma_x^2 I_d)$ , la loi  $W = \frac{\|X\|^2}{\sigma_x^2} \sim \chi_d^2(\lambda^2)$ , dite du  $\chi_d^2$  non centrée, a pour densité  $f_W(w; d, \lambda) = \frac{e^{-\lambda^2/2}}{2^{d/2}\Gamma(d/2)} {}_0F_1\left(\cdot; d/2; \lambda^2 w/4\right) e^{-w/2} w^{d/2-1}$ , où  $\lambda = \frac{\|\theta\|}{\sigma_x}$ .

**Lemme 1.47** *Soit  $X \sim N_d(\theta, \sigma_x^2 I_d)$ ,  $d > 1$ ,  $\lambda = \frac{\|\theta\|}{\sigma_x}$ , et  $T = \frac{\theta'X}{\lambda\|X\|}$ , alors,*

a) *La loi conditionnelle de  $T$  sachant  $\|X\| = r$  est donnée par*

$$\frac{e^{\lambda r t} (1 - t^2)^{\frac{d-1}{2}-1}}{B\left(\frac{1}{2}, \frac{d-1}{2}\right) {}_0F_1\left(-, \frac{d}{2}; \frac{(\lambda r)^2}{4}\right)} \mathbb{I}_{(-1,1)}(t), \text{ avec } B(\cdot, \cdot) \text{ la fonction Bêta.}$$

b) *L'espérance conditionnelle de  $e^{aT}$  sachant  $\|X\| = r$  est donnée par*

$$\mathbb{E}[e^{aT} | \|X\| = r] = \frac{{}_0F_1\left(-, \frac{d}{2}; \frac{(a+\lambda r)^2}{4}\right)}{{}_0F_1\left(-, \frac{d}{2}; \frac{(\lambda r)^2}{4}\right)}, \forall a \in \mathbb{R}. \quad (1.23)$$

c) *L'espérance conditionnelle de  $T e^{aT}$  sachant  $\|X\| = r$  est donnée par*

$$\mathbb{E}[T e^{aT} | \|X\| = r] = \frac{(a + \lambda r)}{d} \frac{{}_0F_1\left(-, \frac{d}{2} + 1; \frac{(a+\lambda r)^2}{4}\right)}{{}_0F_1\left(-, \frac{d}{2}; \frac{(\lambda r)^2}{4}\right)}, \forall a \in \mathbb{R}. \quad (1.24)$$

d) *L'espérance conditionnelle de  $T^2 e^{aT}$  sachant  $\|X\| = r$  est donnée par*

$$\mathbb{E}[T^2 e^{aT} | \|X\| = r] = \frac{{}_0F_1\left(-, \frac{d}{2} + 1; \frac{(a+\lambda r)^2}{4}\right)}{d} \frac{{}_0F_1\left(-, \frac{d}{2} + 2; \frac{(a+\lambda r)^2}{4}\right)}{{}_0F_1\left(-, \frac{d}{2}; \frac{(\lambda r)^2}{4}\right)} + \frac{(a + \lambda r)^2}{{}_0F_1\left(-, \frac{d}{2}; \frac{(\lambda r)^2}{4}\right)}. \quad (1.25)$$

Pour la démonstration de ce Lemme, nous avons besoin du prochain lemme.

**Lemme 1.48** *Pour  $a \in \mathbb{R}$ ,  $\alpha > 0$ , nous avons*

$$\int_{-1}^1 e^{at} (1-t^2)^{\alpha-1} dt = B\left(\frac{1}{2}, \alpha\right) {}_0F_1\left(-, \alpha + \frac{1}{2}; \frac{a^2}{4}\right). \quad (1.26)$$

**Démonstration.** Voir Abramowitz et Stegun (1964) page 376.

**Démonstration du Lemme 1.47.**

a) Fixe  $\theta = (\lambda, 0, \dots, 0)$  sans perte de généralité pour que  $T = \frac{X_1}{\|X\|}$ . De la distribution conjointe de  $(X_1, X'X - X_1^2)$ ; qui a des composantes indépendantes  $N(\lambda, 1)$  et  $\chi_{d-1}^2(0)$ ; on obtient facilement la densité de  $(X_1, \|X\|)$  et la densité conditionnelle de  $X_1/\|X\| = r$ .

b) L'expression (1.23) découle du Lemme 1.48.

c) et d) Les formules (1.24) et (1.25) sont obtenues à l'aide de la dérivée première et seconde de l'expression (1.23) par rapport à  $a$ , avec le fait que

$$\frac{d}{dx} {}_0F_1(-, \alpha; f(x)) = \frac{f'(x)}{\alpha} {}_0F_1(-, \alpha + 1; f(x)).$$

qui découle de la dérivation et d'intégration par parties de l'expression (1.26).

**Lemme 1.49** *Soit  $X \sim N_d(\theta, \sigma_x^2 I_d)$ , avec  $d > 1$ ,  $\lambda = \frac{\|\theta\|}{\sigma_x}$ ,  $T = \frac{X'\theta}{\lambda\|X\|}$ ,  $\hat{\theta}(X) = h(\|X\|^2) X$  et  $Z = \|\hat{\theta}(X) - \theta\|^2$ , alors :*

$$\begin{aligned} \mathbb{E}[Z e^{-sZ}] &= \frac{e^{-\frac{\lambda^2}{2}(1+2s)}}{2^{d/2}\Gamma(d/2)} \left\{ \int_0^\infty e^{-\frac{w}{2}(1+2sh^2(w))} h^2(w) w^{\frac{d}{2}} {}_0F_1\left(-, \frac{d}{2}, \frac{\lambda^2 w (1+2sh(w))^2}{4}\right) dw \right. \\ &\quad + \lambda^2 \int_0^\infty e^{-\frac{w}{2}(1+2sh^2(w))} w^{d/2-1} {}_0F_1\left(-, \frac{d}{2}; \frac{\lambda^2 w (1+2sh(w))^2}{4}\right) dw \\ &\quad \left. - \frac{2\lambda^2}{d} \int_0^\infty e^{-\frac{w}{2}(1+2sh^2(w))} w^{\frac{d}{2}} h(w) (1+2sh(w)) {}_0F_1\left(-, \frac{d}{2} + 1; \frac{\lambda^2 w (1+2sh(w))^2}{4}\right) dw \right\} \end{aligned}$$

**Démonstration.** À l'aide du Lemme 1.47, on a

$$\begin{aligned}
\mathbb{E}_\theta[Ze^{-sZ}] &= \mathbb{E}^X(\|\hat{\theta}(X) - \theta\|^2 e^{-s\|\hat{\theta}(X) - \theta\|^2}) = \mathbb{E}^X(\|h(X'X)X - \theta\|^2 e^{-s\|h(X'X)X - \theta\|^2}) \\
&= \mathbb{E}^X\left((h^2(X'X)X'X - 2h(X'X)X'\theta + \|\theta\|^2) e^{-s(h^2(X'X)X'X - 2h(X'X)X'\theta + \|\theta\|^2)}\right) \\
&= \mathbb{E}^W\left(\mathbb{E}^T\left[(h^2(W)W - 2\lambda W^{\frac{1}{2}}h(W)T + \lambda^2)e^{-s(h^2(W)W - 2\lambda W^{\frac{1}{2}}h(W)T + \lambda^2)}\middle|X'X = W\right]\right) \\
&= \mathbb{E}^W\left(e^{-s(h^2(W)W + \lambda^2)}\left[(h^2(W)W + \lambda^2)\mathbb{E}^T\left(e^{2s\lambda W^{\frac{1}{2}}h(W)T}\middle|X'X = W\right)\right.\right. \\
&\quad \left.\left.- 2\lambda W^{\frac{1}{2}}h(W)\mathbb{E}^T\left(T e^{2s\lambda W^{\frac{1}{2}}h(W)T}\middle|X'X = W\right)\right]\right) \\
&= \mathbb{E}^W\left(e^{-s(h^2(W)W + \lambda^2)}\left[\frac{{}_0F_1\left(-, \frac{d}{2}; \frac{(2s\lambda W^{\frac{1}{2}}h(W) + \lambda W^{\frac{1}{2}})^2}{4}\right)}{{}_0F_1\left(-, \frac{d}{2}; \frac{(\lambda W^{\frac{1}{2}})^2}{4}\right)}\right.\right. \\
&\quad \left.\left.- 2\lambda W^{\frac{1}{2}}h(W)\frac{B\left(\frac{3}{2}, \frac{d-1}{2}\right)_0F_1\left(-, \frac{d}{2} + 1; \frac{(2s\lambda W^{\frac{1}{2}}h(W) + \lambda W^{\frac{1}{2}})^2}{4}\right)}{B\left(\frac{1}{2}, \frac{d-1}{2}\right)_0F_1\left(-, \frac{d}{2}; \frac{(\lambda W^{\frac{1}{2}})^2}{4}\right)}\right]\right) \\
&= \frac{e^{-\frac{\lambda^2}{2}(1+2s)}}{2^{d/2}\Gamma(d/2)}\left\{\int_0^\infty e^{-\frac{w}{2}(1+2sh^2(w))}h^2(w)w^{\frac{d}{2}}{}_0F_1\left(-, \frac{d}{2}; \frac{\lambda^2w(1+2sh(w))^2}{4}\right)dw\right. \\
&\quad \left.+ \lambda^2\int_0^\infty e^{-\frac{w}{2}(1+2sh^2(w))}w^{d/2-1}{}_0F_1\left(-, \frac{d}{2}; \frac{\lambda^2w(1+2sh(w))^2}{4}\right)dw - \frac{2\lambda^2 B\left(\frac{1}{2}, \frac{d+1}{2}\right)}{(d-1)B\left(\frac{1}{2}, \frac{d-1}{2}\right)}\right. \\
&\quad \left.\int_0^\infty e^{-\frac{w}{2}(1+2sh^2(w))}w^{\frac{d}{2}}h(w)(1+2sh(w)){}_0F_1\left(-, \frac{d}{2} + 1; \frac{\lambda^2w(1+2sh(w))^2}{4}\right)dw\right\} \\
&= \frac{e^{-\frac{\lambda^2}{2}(1+2s)}}{2^{d/2}\Gamma(d/2)}\left\{\int_0^\infty e^{-\frac{w}{2}(1+2sh^2(w))}h^2(w)w^{\frac{d}{2}}{}_0F_1\left(-, \frac{d}{2}; \frac{\lambda^2w(1+2sh(w))^2}{4}\right)dw\right. \\
&\quad \left.+ \lambda^2\int_0^\infty e^{-\frac{w}{2}(1+2sh^2(w))}w^{d/2-1}{}_0F_1\left(-, \frac{d}{2}; \frac{\lambda^2w(1+2sh(w))^2}{4}\right)dw\right. \\
&\quad \left.- \frac{2\lambda^2}{d}\int_0^\infty e^{-\frac{w}{2}(1+2sh^2(w))}w^{\frac{d}{2}}h(w)(1+2sh(w)){}_0F_1\left(-, \frac{d}{2} + 1; \frac{\lambda^2w(1+2sh(w))^2}{4}\right)dw\right\}.
\end{aligned}$$

**Lemme 1.50** Pour  $X \sim N_d(\theta, \sigma_X^2 I_d)$  et  $\hat{\theta}(X) = h(X'X)X$ , on a

$$\mathbb{E}_\theta \left[ e^{-s \|h(X'X)X - \theta\|^2} \right] = \frac{e^{-\frac{(2\sigma_X^2 s + 1)\|\theta\|^2}{2}}}{2^{\frac{d}{2}} \Gamma(\frac{d}{2})} \int_0^\infty w^{\frac{d}{2}-1} e^{-\frac{w}{2}(1+2\sigma_X^2 s h^2(w))} {}_0F_1\left(-, \frac{d}{2}, \frac{w\|\theta\|^2}{4}(1+2\sigma_X^2 s h(w))^2\right) dw.$$

**Démonstration.** Sans perte de généralité on suppose que  $\sigma_X = 1$ , on a

$$\begin{aligned} \mathbb{E}_\theta [e^{-s \|h(X'X)X - \theta\|^2}] &= \mathbb{E}^X \left( e^{-s(h^2(X'X)X'X - 2h(X'X)X'\theta + \|\theta\|^2)} \right) \\ &= \mathbb{E}^W \left( \mathbb{E}^T \left[ e^{-s(h^2(W)W - 2\|\theta\|W^{\frac{1}{2}}h(W)T + \|\theta\|^2)} \mid X'X = W \right] \right) \\ &= \mathbb{E}^W \left( e^{-s(h^2(W)W + \|\theta\|^2)} \left[ \mathbb{E}^T \left( e^{2s\|\theta\|W^{\frac{1}{2}}h(W)T} \mid X'X = W \right) \right. \right. \\ &= \mathbb{E}^W \left( e^{-s(h^2(W)W + \|\theta\|^2)} \left[ \frac{{}_0F_1\left(-, \frac{d}{2}, \frac{(2s\|\theta\|W^{\frac{1}{2}}h(W) + \|\theta\|W^{\frac{1}{2}})^2}{4}\right)}{{}_0F_1\left(-, \frac{d}{2}, \frac{(\|\theta\|W^{\frac{1}{2}})^2}{4}\right)} \right] \right) \\ &= \frac{e^{-\frac{\|\theta\|^2}{2}(1+2s)}}{2^{d/2}\Gamma(d/2)} \int_0^\infty e^{-\frac{w}{2}(1+2sh^2(w))} w^{\frac{d}{2}-1} {}_0F_1\left(-, \frac{d}{2}, \frac{\|\theta\|^2 w(1+2sh(w))^2}{4}\right) dw. \end{aligned}$$

Enfin, l'expression (1.22) et le Lemme 1.50 nous donnent le Corollaire suivant.

**Corollaire 1.51** Pour l'estimation d'une densité normale multivariée de  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$  sous la perte  $\alpha$ -divergence, le risque de la densité prédictive  $q_{\hat{\theta}, c^2} \sim N_d(\hat{\theta}, c\sigma_Y^2 I_d)$  avec  $\hat{\theta}(X) = h(X'X)X$  est donné par

$$\begin{aligned} R_\alpha(\theta, \hat{q}_{c, \hat{\theta}(X)}) &= \frac{4}{1-\alpha^2} \left\{ 1 - \left( \frac{c^{1-\alpha}}{(1-\alpha)c^2 + 1 + \alpha} \right)^{\frac{d}{2}} e^{-\frac{(2\gamma_0+1)\|\theta\|^2}{2}} \right. \\ &\quad \left. \times \int_0^\infty \frac{w^{\frac{d}{2}-1} e^{-\frac{w}{2}(1+2\gamma_0 h^2(w))}}{\Gamma(\frac{d}{2})} {}_0F_1\left(-, \frac{d}{2}, \frac{w\|\theta\|^2}{4}(1+2\gamma_0 h(w))^2\right) dw \right\}, \\ \text{avec } \gamma_0 &= \frac{(1-\alpha^2)\sigma_X^2}{4((1-\alpha)c^2 + 1 + \alpha)\sigma_Y^2}. \end{aligned}$$

## CHAPITRE 2

# On predictive density estimation for Gamma models with parametric constraints.

Ce Chapitre contient un article, avec les Professeurs Éric Marchand, William E. Strawderman et Othmane Kortbi, publié en 2017 dans : *Journal of Statistical Planning and Inference*, 185, 56-68.



AZIZ L'MOUDDEN<sup>a</sup>, ÉRIC MARCHAND<sup>a</sup>, OTHMANE KORTBI<sup>b</sup>, WILLIAM E.  
STRAWDERMAN<sup>c</sup>

*a Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA,  
J1K 2R1 (e-mails : aziz.lmoudden@usherbrooke.ca ; eric.marchand@usherbrooke.ca)*

*b Department of Statistics, CBE, UAE University (e-mail : okortbi@uaeu.ac.ae)*

*c Rutgers University, Department of Statistics and Biostatistics, 501 Hill Center, Busch  
Campus, Piscataway, N.J., USA, 08855 (e-mail : straw@stat.rutgers.edu)*

## SUMMARY

This paper is concerned with the estimation of a predictive density for  $Y \sim \text{Gamma}(\alpha_2, \beta)$  under Kullback-Leibler loss, based on  $X \sim \text{Gamma}(\alpha_1, \beta)$ , and for situations where there is a parametric constraint of the form  $\beta \in C = (a, b)$ . We obtain representations for Bayes predictive densities and the minimum risk equivariant predictive density in the unconstrained problem. It is shown that the generalized Bayes estimator against the truncation of the non-informative prior onto  $C$  dominates the minimum risk equivariant predictive density and is minimax whenever  $a = 0$  or  $b = \infty$ . Analytical comparisons of plug-in predictive densities  $\text{Gamma}(\alpha_2, \hat{\beta})$ , which include the predictive mle density, are obtained with results applying as well for point estimation under dual entropy loss  $\frac{\beta}{\hat{\beta}} - \log(\frac{\beta}{\hat{\beta}}) - 1$ . Numerical evaluations confirm that such predictive densities are much less efficient than some Bayesian alternatives in exploiting the parametric restriction. Finally, it is shown that variance expansion improvements of the form  $\text{Gamma}(\frac{\alpha_2}{k}, k\hat{\beta})$  of plug-in predictive densities can always be found for a subset of  $k > 1$  values and non-degenerate  $\hat{\beta}$ .

*AMS 2010 subject classifications : 62C20, 62C86, 62F10, 62F15, 62F30*

*Keywords and phrases : Bayes estimator ; Dominance ; Frequentist risk ; Gamma ; Kullback-Leibler loss ; Minimax ; Minimum risk equivariant ; Plug-in ; Predictive density ; Restricted parameter space ; Variance expansion.*

## 2.1 Introduction

Prediction of randomly generated quantities is a central theme in statistics. The richest prediction takes the form of a predictive density over the domain of possible values. Bayesian predictive densities are optimal in response to a given prior and loss function.

Nonnegative random variables arise in a wide variety of applications, such as life-lengths of components or biological organisms typical of reliability and survival analysis, magnitudes related to physical objects (e.g., plant diameters or heights, wind speeds, measures of precipitation), and measurable quantities in economics (e.g., incomes, prices).

For parametric models of positive-valued random variables, it is often the case that bounds on the parameters can be stipulated, especially when these bounds relate to the practical context in which they arise. The perhaps most straightforward way to visualize this is through restrictions on quantiles or means which translate to parametric restrictions.

In this paper, we focus on Gamma models which play an important role for nonnegative data. In particular, we study the estimation of predictive densities with restrictions on the scale parameter. We thus consider independently Gamma distributed

$$X|\beta \sim Ga(\alpha_1, \beta), Y|\beta \sim Ga(\alpha_2, \beta); \quad (2.1)$$

where  $\alpha_1$  and  $\alpha_2$  are known, not necessarily equal, and  $\beta$  is unknown but restricted to  $C = (a, b)$ . Our analysis will address : **(i)** the lower bounded case (i.e.,  $a > 0$  and  $b = \infty$ ), **(ii)** the upper bounded case (i.e.,  $a = 0$  and  $b < \infty$ ), and **(iii)** the doubly-bounded case with  $a > 0$  and  $b < \infty$ . For predictive analysis purposes, researchers are interested in specifying a predictive density  $\hat{q}(\cdot; x)$  as an estimate of the density  $q(\cdot|\beta)$  of  $Y$ . In turn, such a density may play a surrogate role for generating either future or missing values of  $Y$ . To evaluate the performance of such predictive densities, we resort to the familiar Kullback-Leibler loss

$$L_{KL}(\beta, \hat{q}(\cdot; x)) = \int_{R_+} \log \left( \frac{q(y|\beta)}{\hat{q}(y; x)} \right) q(y|\beta) dy, \quad (2.2)$$

with corresponding frequentist risk given by

$$R_{KL}(\beta, \hat{q}) = E^X [L_{KL}(\beta, \hat{q}(\cdot; X))]. \quad (2.3)$$

The main findings and contributions of this paper consist of :

- (A) A general dominance result (Theorem 2.9) establishing that the Bayes predictive density  $\hat{q}_{\pi_C}(\cdot; X)$  generated from the prior  $\pi_C(\beta) = \frac{1}{\beta} \mathbb{I}_C(\beta)$  (in other words the truncation of the non-informative prior  $\pi_0(\beta) = \frac{1}{\beta} \mathbb{I}_{\mathbb{R}_+}(\beta)$ ) onto the restricted parameter space  $C = (a, b)$ ; dominates the minimum risk equivariant predictive density estimator  $\hat{q}_{\pi_0}(\cdot; X)$ , which is also Bayes with respect to  $\pi_0$ . The approach is unified for all  $C = (a, b)$  and exploits results in (B) and a star-shaped property of Gamma distributions. For the doubly-bounded case, the result is new, while for the lower bounded or upper bound case, the result duplicates in a much different manner that of Kubokawa et al. (2013). Moreover, for doubly-bounded parameter spaces  $C = (a, b)$ , we show that the Bayesian predictive estimator  $\hat{q}_{\pi_0, C}(\cdot; X)$  dominates  $\hat{q}_{\pi_0, C'}(\cdot; X)$  for both  $C' = (a, \infty)$  and  $C' = (0, b)$ .
- (B) Various representation and properties of Bayesian predictive densities (Section 2) which are required for (A) and which are of interest on their own.
- (C) An explicit form for the minimax risk for both the lower bounded ( $\beta \geq a$ ) and the upper bounded ( $\beta \leq b$ ) cases.
- (D) An analysis (Section 5) of the risk performance under the parametric restriction  $\beta \in (a, b)$  of plug-in predictive density estimators  $q_{\hat{\beta}(\cdot; X)} \sim \text{Ga}(\alpha_2, \hat{\beta}(X))$  for truncated linear plug-in estimators, and with corresponding point estimation results for the dual entropy loss  $L_0(\beta, \hat{\beta}) = \left( \frac{\beta}{\hat{\beta}} - \log\left(\frac{\beta}{\hat{\beta}}\right) - 1 \right)$ .
- (E) A variance expansion class of improvements (Section 6) on plug-in predictive density estimators. These improvements are obtained as  $\text{Ga}\left(\frac{\alpha_2}{k}, k \hat{\beta}(X)\right)$  densities with  $1 < k < k_0$ , where  $k_0$  depends only on  $\alpha_2$  and the infimum entropy risk of the plug-in estimator  $\hat{\beta}(X)$ .

These above developments are further introduced throughout and constitute valuable additions to both the predictive density estimation body of work and inference problems under paramete-

tric restrictions. The former area of research has been quite active in recent years with key contributions in particular for normal and Poisson models without parametric restrictions, as witnessed by the work of Boisbunon and Maruyama (2014); Kato (2009); Komaki (2006, 2004, 2001); George, Liang and Xu (2006); Brown, George and Xu (2008); Liang and Barron (2004); among others. Predictive density estimation with parametric restrictions has been studied by Fourdrinier et al. (2011) for normal models, as well as by Kubokawa et al. (2013) for location and scale families of distributions, while van Eeden (2006), as well as Marchand and Strawderman (2004) review estimation in restricted parameter spaces.

## 2.2 Bayes estimators

We begin here with a general representation for Bayesian predictive density estimators  $\hat{q}_\pi(\cdot; X)$  where  $\pi$  is a prior for  $\beta$ , and where a key role is played by the predictive density estimator  $\hat{q}_{\pi_0}(\cdot; X)$ , with  $\pi_0(\beta) = \frac{1}{\beta} \mathbb{I}_{(0, \infty)}(\beta)$  being the usual non-informative prior for  $\beta$ .

**Theorem 2.1** *Consider model (2.1), the noninformative prior  $\pi_0(\beta) = \frac{1}{\beta} \mathbb{I}_{(0, \infty)}(\beta)$ , and an arbitrary generalized prior  $\pi(\beta)$  (with respect to  $\sigma$ -finite measure  $\nu$ ) for which the posterior  $\pi(\beta|x)$  exists. Then, we have for the predictive densities  $\hat{q}_{\pi_0}(\cdot; x)$  and  $\hat{q}_\pi(\cdot; x)$*

$$(a) \quad \hat{q}_{\pi_0}(y; x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{1}{x} \left(\frac{y}{x}\right)^{\alpha_2 - 1} \left(1 + \frac{y}{x}\right)^{-(\alpha_1 + \alpha_2)} \mathbb{I}_{(0, \infty)}(y);$$

$$(b) \quad \hat{q}_\pi(y; x) = \hat{q}_{\pi_0}(y; x) \frac{(y+x)}{x} \frac{m_\pi(y+x|\alpha_1 + \alpha_2)}{m_\pi(x|\alpha_1)};$$

where  $m_\pi(z|\alpha) = \int_0^\infty \frac{z^{\alpha-1} e^{-z/\beta}}{\Gamma(\alpha)\beta^\alpha} \pi(\beta) d\nu(\beta)$  is the marginal density of  $Z \sim \Gamma(\alpha, \beta)$  associated with prior  $\pi$ .

**Proof.** Bayes estimators under Kullback-Leibler are posterior expectations of the model density (e.g., Aitchison, 1975). For (b), we thus have

$$\begin{aligned}
\hat{q}_\pi(y; x) &= \int_0^\infty q(y|\beta) \pi(\beta|x) d\nu(\beta) \\
&= \frac{\int_0^\infty q(y|\beta) p(x|\beta) \pi(\beta) d\nu(\beta)}{\int_0^\infty p(x|\beta) \pi(\beta) d\nu(\beta)}
\end{aligned} \tag{2.4}$$

$$\begin{aligned}
&= \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1) \Gamma(\alpha_2)} y^{\alpha_2-1} \frac{x^{\alpha_1}}{(y+x)^{\alpha_1+\alpha_2}} \left(\frac{y+x}{x}\right) \frac{\int_0^\infty \frac{(y+x)^{\alpha_1+\alpha_2-1}}{\beta^{\alpha_1+\alpha_2} \Gamma(\alpha_1+\alpha_2)} \exp\left(-\frac{y+x}{\beta}\right) \pi(\beta) d\nu(\beta)}{\int_0^\infty \frac{x^{\alpha_1-1}}{\beta^{\alpha_1} \Gamma(\alpha_1)} \exp\left(-\frac{x}{\beta}\right) \pi(\beta) d\nu(\beta)} \\
&= \hat{q}_{\pi_0}(y; x) \left(\frac{y+x}{x}\right) \frac{m_\pi(y+x|\alpha_1+\alpha_2)}{m_\pi(x|\alpha_1)}.
\end{aligned} \tag{2.5}$$

Part **(a)** follows from (2.5) with the evaluation  $m_{\pi_0}(z|\alpha) = z^{-1}$  obtained from its definition and the Gamma function identity

$$\int_0^\infty z^{-(c+1)} \exp\left(-\frac{d}{z}\right) dz = \Gamma(c) d^{-c}, \quad \forall c, d > 0. \tag{2.6}$$

Part **(a)** was obtained by Aitchison (1975). Part **(b)** of Theorem 2.1, which represents the Bayes predictive density  $\hat{q}_\pi$  as a weighted version of  $\hat{q}_{\pi_0}$  is general with respect to the choice of prior, will be illustrated and exploited below. The density  $\hat{q}_{\pi_0}(\cdot; x)$  is the density of a Beta2( $c = \alpha_2, d = \alpha_1, \sigma = x$ ) distribution of type II, with shape parameters  $c$  and  $d$ , scale parameter  $\sigma$ , defined as follows (see Section 3 for further interpretation).

**Definition 2.2** We will say that  $U \sim \text{Beta2}(c, d, \sigma)$ , with  $c, d, \sigma > 0$ , whenever  $U$  has density on  $\mathbb{R}_+$  given by

$$f(u) = \frac{\Gamma(c+d)}{\Gamma(c)\Gamma(d)} \frac{1}{\sigma} \frac{(u/\sigma)^{c-1}}{(1+u/\sigma)^{c+d}}. \tag{2.7}$$

Equivalently, such Beta2 distributions are also known as Fisher distributions. They are related to Beta (type I) distributions as follows :  $\frac{U}{\sigma} \sim \text{Beta2}(c, d, 1) =^d \frac{V}{1-V}$  where  $V \sim \text{Beta}(c, d)$  with density  $\frac{\Gamma(c+d)}{\Gamma(c)\Gamma(d)} v^{c-1} (1-v)^{d-1} \mathbb{I}_{(0,1)}(v)$ .

Here are some instructive examples of Theorem 2.1.

**Example 2.3** (Conjugate inverse Gamma priors)

For inverse Gamma priors  $\pi(\beta) \propto \beta^{-\gamma_1} e^{-\gamma_2/\beta}$  with  $\gamma_1 > 1 - \alpha_1$  and  $\gamma_2 > 0$ , we have  $\hat{q}_\pi(y; x) \sim$

$Beta2(\alpha_2, \gamma_1 + \alpha_1 - 1, x + \gamma_2)$ . This was obtained by Amaral and Dunsmore (1980). It can be derived from (2.5) via the evaluation  $m_\pi(z|\alpha) = \frac{\Gamma(\alpha+\gamma_1-1)}{\Gamma(\alpha)} \frac{z^{\alpha-1}}{(z+\gamma_2)^{\alpha+\gamma_1-1}}$ , which makes use of (2.6) again.

**Example 2.4** (Priors truncated to  $C = (a, b)$ ) With the parametric restriction  $\beta \in C = (a, b)$ , it seems natural to consider as priors the truncation onto  $C$  of the noninformative prior  $\pi_0(\beta) = \frac{1}{\beta}$  and, more generally, truncations of inverse Gamma priors given by prior densities

$$\pi_{(a,b),\gamma_1,\gamma_2}(\beta) \propto \beta^{-\gamma_1} e^{-\gamma_2/\beta} \mathbb{I}_{(a,b)}(\beta), \quad (2.8)$$

with  $\gamma_1 > 1 - \alpha_1$  and  $\gamma_2 \geq 0$ . Evaluating Theorem 2.1's marginals  $m_\pi$ , we have

$$\begin{aligned} z m_\pi(z|\alpha) &= \int_a^b \frac{z^\alpha e^{-z/\beta}}{\Gamma(\alpha) \beta^\alpha} \pi(\beta) d\beta \\ &= \frac{z^\alpha}{\Gamma(\alpha)} \int_a^b \beta^{-(\gamma_1+\alpha)} e^{-(z+\gamma_2)/\beta} d\beta \\ &= \frac{z^\alpha}{\Gamma(\alpha)} \int_{(z+\gamma_2)/b}^{(z+\gamma_2)/a} \frac{u^{\alpha+\gamma_1-2} e^{-u}}{(z+\gamma_2)^{\alpha+\gamma_1-1}} du. \\ &= \frac{\Gamma(\alpha + \gamma_1 - 1) z^\alpha}{\Gamma(\alpha) (z + \gamma_2)^{\alpha+\gamma_1-1}} \left( F_{\alpha+\gamma_1-1} \left( \frac{z + \gamma_2}{a} \right) - F_{\alpha+\gamma_1-1} \left( \frac{z + \gamma_2}{b} \right) \right), \end{aligned}$$

where  $F_\gamma(\cdot)$  is the Gamma( $\gamma, 1$ ) cdf. Using the above in (2.5), the Bayes estimator can be expressed as

$$\hat{q}_{\pi_{(a,b),\gamma_1,\gamma_2}}(y; x) = \frac{F_{\alpha_1+\alpha_2+\gamma_1-1} \left( \frac{x+y+\gamma_2}{a} \right) - F_{\alpha_1+\alpha_2+\gamma_1-1} \left( \frac{x+y+\gamma_2}{b} \right)}{F_{\alpha_1+\gamma_1-1} \left( \frac{x+\gamma_2}{a} \right) - F_{\alpha_1+\gamma_1-1} \left( \frac{x+\gamma_2}{b} \right)}, \quad (2.9)$$

with  $\hat{q}_{\pi,\gamma_1,\gamma_2}(\cdot; x)$  being Example 2.3's (unrestricted) Bayes Beta2( $\alpha_2, \gamma_1 + \alpha_1 - 1, x + \gamma_2$ ) predictive density. In the expression above, and throughout the paper, it shall be understood that  $F_\alpha(t/a)$  and  $F_\alpha(t/b)$  evaluated at  $a = 0$  or  $b = +\infty$  are taken to be equal to 1 and 0 respectively. Expression (2.9) encompasses all parameter spaces of the form  $(a, b)$  and prior parameters  $\gamma_1, \gamma_2$ . The expression for the case  $a = 0, b < \infty, \gamma_1 = 1, \gamma_2 = 0$  was given by Kubokawa et al. (2013).

## 2.2.1 A general scale family representation of the MRE predictive density estimator

We briefly expand here on a general representation for Bayes predictive density estimators in scale family problems and associated with the usual noninformative prior. Observe that the predictive density  $\hat{q}_{\pi_0}(\cdot; x)$ , given in Theorem 2.1, is the density of  $x \frac{T_2}{T_1}$ , where  $T_1 \sim \text{Ga}(\alpha_1, \beta)$ ,  $T_2 \sim \text{Ga}(\alpha_2, \beta)$  are independently distributed (as in (2.1)), as such ratios are Beta2( $\alpha_2, \alpha_1, x$ ) distributed. Here is a generalization to other scale parameter families followed by a pair of examples.

**Theorem 2.5** *Let  $X|\sigma \sim \frac{1}{\sigma} p(\frac{x}{\sigma})$ ,  $Y|\sigma \sim \frac{1}{\sigma} q(\frac{y}{\sigma})$  be independently distributed nonnegative random variables. Let  $h$  be the density of the ratio  $\frac{Y}{X}$  (which does not depend on  $\sigma$ ) given by  $h(v) = \int_0^\infty u q(vu) p(u) du$ . Then, the Bayes predictive density estimator under Kullback-Leibler loss in (2.2) of the density of  $Y$  associated with the prior  $\pi_0(\sigma) = \frac{1}{\sigma} \mathbb{I}_{(0,\infty)}(\sigma)$ , is given by  $\hat{q}_{\pi_0}(y; x) = \frac{1}{x} h(\frac{y}{x})$ .*

**Proof.** We have

$$\begin{aligned} \hat{q}_{\pi_0}(y; x) &= \frac{\int_0^\infty \frac{1}{\sigma} q(\frac{y}{\sigma}) \frac{1}{\sigma} p(\frac{x}{\sigma}) \frac{d\sigma}{\sigma}}{\int_0^\infty \frac{1}{\sigma} q(\frac{x}{\sigma}) \frac{d\sigma}{\sigma}} \\ &= \frac{1}{x} \frac{\int_0^\infty u q(\frac{y}{x}u) p(u) du}{\int_0^\infty p(u) du} \\ &= \frac{1}{x} h(\frac{y}{x}). \end{aligned}$$

**Example 2.6** (*Uniform model*)

For  $X|\sigma \sim U(0, \sigma)$   $Y|\sigma \sim U(0, \sigma)$ , we have  $h(v) = \int_0^\infty u \mathbb{I}_{(0,1)}(uv) \mathbb{I}_{(0,1)}(u) du = \frac{1}{2} \mathbb{I}_{(0,1)}(v) + \frac{1}{2v^2} \mathbb{I}_{(1,\infty)}(v)$ ; i.e., the density of the ratio  $\frac{Y}{X}$ , and thus  $\hat{q}_{\pi_0}(y; x) = \frac{1}{x} h(\frac{y}{x}) = \frac{1}{2x} \mathbb{I}_{(0,x]}(y) + \frac{x}{2y^2} \mathbb{I}_{(x,\infty)}(y)$ .

**Example 2.7** (*Beta type 2 model*)

For  $X|\sigma \sim \text{Beta2}(c_1, d_1, \sigma)$ ,  $Y|\sigma \sim \text{Beta2}(c_2, d_2, \sigma)$ , as in Definition 2.2, we obtain  $\hat{q}_{\pi_0}(y; x) =$

$\frac{1}{x} h\left(\frac{y}{x}\right)$ , with the distribution of the ratio  $\frac{Y}{X}$  and its density  $h$  studied in Pham-Gia and Turkkan (2002) for instance.

## 2.3 Dominance results

Let the parameter space be restricted to  $C = (a, b)$ . We focus in this subsection on the Bayesian predictive estimators  $\hat{q}_{\pi_0}(\cdot; X)$  and  $\hat{q}_{\pi_0, C}(\cdot; X)$  associated respectively with the noninformative prior  $\pi_0(\beta) = \frac{1}{\beta}$  and its truncation to the parameter space  $\pi_{0, C}(\beta) = \pi_0(\beta) \mathbb{I}_C(\beta)$ . The former predictive estimator is given in Theorem 2.1, while the latter is given by  $\hat{q}_{\pi_0, 1, 0}(y; x)$  in (2.9) as

$$\hat{q}_{\pi_0, C}(y|x) = \hat{q}_{\pi_0}(y|x) \frac{F_{\alpha_1 + \alpha_2}\left(\frac{x+y}{a}\right) - F_{\alpha_1 + \alpha_2}\left(\frac{x+y}{b}\right)}{F_{\alpha_1}\left(\frac{x}{a}\right) - F_{\alpha_1}\left(\frac{x}{b}\right)}, \quad (2.10)$$

(as  $\pi_{0, C}$  corresponds to (2.8) with  $\gamma_1 = 1, \gamma_2 = 0$ ). We show below in Theorem 2.9 that  $\hat{q}_{\pi_0, C}(\cdot; X)$  dominates  $\hat{q}_{\pi_0}(\cdot; X)$  for all  $C$  under Kullback-Leibler risk  $R_{KL}$  given in (2.3). For both the upper bounded ( $a = 0, b < \infty$ ) and lower-bounded ( $a > 0, b = \infty$ ) cases, the dominance result was established by Kubokawa et al. (2013) using an adaptation of Kubokawa's IERD method (e.g., Kubokawa, 1994), but we provide a direct, alternative and unified in  $C$  route in establishing the dominance. With Kubokawa et al. (2013) having shown that  $\hat{q}_{\pi_0}(\cdot; X)$  is minimax for such one-sided restrictions, dominating predictive density estimators such as  $\hat{q}_{\pi_0, C}(\cdot; X)$  are also minimax. In the doubly-bounded case with  $a > 0, b < \infty$ , our dominance finding is new, but neither  $\hat{q}_{\pi_0}(\cdot; X)$  nor  $\hat{q}_{\pi_0, C}(\cdot; X)$  is minimax. Finally, we also show that  $\hat{q}_{\pi_0, C}(\cdot; X)$  dominates  $\hat{q}_{\pi_0, C'}(\cdot; X)$  for doubly-bounded parameter spaces  $C = (a, b)$  and  $C' = (a, \infty)$  or  $C' = (0, b)$ .

We will require the next two lemmas. The first of these involves inequalities given and exploited by Misra and Arshad (2014), although we are stating these as strict inequalities. It is an immediate consequence of the star-shaped property of the family of Gamma( $\alpha, 1$ ) distributions, i.e.,  $\frac{F_{\alpha}^{-1}(F_{\alpha'}(x))}{x}$  is an increasing function of  $x \in (0, \infty)$  whenever  $\alpha < \alpha'$ , which in turn follows from the convexity of  $F_{\alpha}^{-1} \circ F_{\alpha'}$  (van Zwet, 1964), for  $\alpha < \alpha'$ , and  $F_{\alpha}^{-1}(F_{\alpha'}(0)) = 0$ .



**Lemma 2.8** Let  $F_\alpha$  be the cdf of a Gamma( $\alpha, 1$ ) distribution. Then, for  $0 < \alpha < \alpha'$ ,  $c_1, c_2 \in (0, 1)$ , and for all  $x > 0$ , we have

- (i)  $c_1 F_\alpha^{-1}(F_{\alpha'}(x)) > F_\alpha^{-1}(F_{\alpha'}(c_1 x))$ ;
- (ii)  $\frac{1}{c_2} F_\alpha^{-1}(F_{\alpha'}(x)) < F_\alpha^{-1}\left(F_{\alpha'}\left(\frac{x}{c_2}\right)\right)$ .

Here is our main dominance result.

**Theorem 2.9** Let  $C = (a, b)$  be the restricted parameter space, either lower bounded with  $b = +\infty$ , upper bounded with  $a = 0$ , or doubly-bounded with  $a > 0$  and  $b < \infty$ . Consider estimating the density of  $Y \sim \text{Gamma}(\alpha_2, \beta)$  under Kullback-Leibler loss (2.2) with  $\beta \in C$ , and based on  $X \sim \text{Gamma}(\alpha_1, \beta)$ .

- (a) Then, the predictive density estimator  $\hat{q}_{\pi_0, C}(\cdot; X)$  dominates  $\hat{q}_{\pi_0, C'}(\cdot; X)$  where  $C \subset C' = (a', b')$  with either (i)  $a' = a, b' = +\infty$ ; (ii)  $a' = 0, b' = b$ ; or (iii)  $C' = \mathbb{R}_+$ ,
- (b) with the Kullback-Leibler risks being equal if and only if  $C' = \mathbb{R}_+$  and  $\beta = a (> 0), b = +\infty$  or  $a = 0, \beta = b (< +\infty)$ .

**Proof.** Representation (2.10) allows us to express the difference of risks as

$$\begin{aligned}
\Delta_{KL}(\beta) &= R_{KL}(\beta, \hat{q}_{\pi_0, C'}) - R_{KL}(\beta, \hat{q}_{\pi_0, C}) \\
&= E_\beta^{X, Y} \left[ \log \frac{\hat{q}_{\pi_0, C}(Y; X)}{\hat{q}_{\pi_0, C'}(Y; X)} \right] \\
&= E_\beta^{X, Y} \left[ \log \frac{F_{\alpha_1 + \alpha_2}\left(\frac{X+Y}{a}\right) - F_{\alpha_1 + \alpha_2}\left(\frac{X+Y}{b}\right)}{F_{\alpha_1}\left(\frac{X}{a}\right) - F_{\alpha_1}\left(\frac{X}{b}\right)} - \log \frac{F_{\alpha_1 + \alpha_2}\left(\frac{X+Y}{a'}\right) - F_{\alpha_1 + \alpha_2}\left(\frac{X+Y}{b'}\right)}{F_{\alpha_1}\left(\frac{X}{a'}\right) - F_{\alpha_1}\left(\frac{X}{b'}\right)} \right] \\
&= E_1^{X, Y} \left[ \log \frac{F_{\alpha_1 + \alpha_2}\left(\beta \frac{X+Y}{a}\right) - F_{\alpha_1 + \alpha_2}\left(\beta \frac{X+Y}{b}\right)}{F_{\alpha_1 + \alpha_2}\left(\beta \frac{X+Y}{a'}\right) - F_{\alpha_1 + \alpha_2}\left(\beta \frac{X+Y}{b'}\right)} - \log \frac{F_{\alpha_1}\left(\beta \frac{X}{a}\right) - F_{\alpha_1}\left(\beta \frac{X}{b}\right)}{F_{\alpha_1}\left(\beta \frac{X}{a'}\right) - F_{\alpha_1}\left(\beta \frac{X}{b'}\right)} \right] \\
&= \phi(\alpha_1 + \alpha_2, \beta) - \phi(\alpha_1, \beta),
\end{aligned}$$

with

$$\phi(\alpha, \beta) = E_\alpha^T \left[ \log \left( \frac{F_\alpha\left(\frac{\beta}{a}T\right) - F_\alpha\left(\frac{\beta}{b}T\right)}{F_\alpha\left(\frac{\beta}{a'}T\right) - F_\alpha\left(\frac{\beta}{b'}T\right)} \right) \right],$$

the expectation  $E_\alpha^T$  taken with respect to  $T \sim Ga(\alpha, 1)$ . For establishing the result in **(a)**, it will suffice to show, for cases **(i)**, **(ii)**, **(iii)**, that  $\phi(\alpha', \beta) \geq \phi(\alpha, \beta)$  for all  $\beta \in C = (a, b)$  and  $\alpha' > \alpha$ . We have by definition of  $\phi(\alpha, \beta)$  and with the change of variable  $x = F_{\alpha'}^{-1}(F_\alpha(t))$

$$\phi(\alpha, \beta) = \int_0^\infty \log \left( \frac{F_\alpha\left(\frac{\beta}{a}t\right) - F_\alpha\left(\frac{\beta}{b}t\right)}{F_\alpha\left(\frac{\beta}{a'}t\right) - F_\alpha\left(\frac{\beta}{b'}t\right)} \right) dF_\alpha(t) \quad (2.11)$$

$$= \int_0^\infty \log \left( \frac{F_\alpha\left(\frac{\beta}{a}F_\alpha^{-1}(F_{\alpha'}(x))\right) - F_\alpha\left(\frac{\beta}{b}F_\alpha^{-1}(F_{\alpha'}(x))\right)}{F_\alpha\left(\frac{\beta}{a'}F_\alpha^{-1}(F_{\alpha'}(x))\right) - F_\alpha\left(\frac{\beta}{b'}F_\alpha^{-1}(F_{\alpha'}(x))\right)} \right) dF_{\alpha'}(x). \quad (2.12)$$

We now separate the remainder of the proof into cases **(i)**, **(ii)**, and **(iii)**.

**(i)** For  $a' = a, b' = +\infty, \alpha' > \alpha$ , we obtain from (2.12)

$$\begin{aligned} \phi(\alpha, \beta) &= \int_0^\infty \log \left( 1 - \frac{F_\alpha\left(\frac{\beta}{b}F_\alpha^{-1}(F_{\alpha'}(x))\right)}{F_\alpha\left(\frac{\beta}{a}F_\alpha^{-1}(F_{\alpha'}(x))\right)} \right) dF_{\alpha'}(x) \\ &< \int_0^\infty \log \left( 1 - \frac{F_{\alpha'}\left(\frac{\beta}{b}x\right)}{F_{\alpha'}\left(\frac{\beta}{a}x\right)} \right) = \phi(\alpha', \beta), \end{aligned}$$

by making use of Lemma 2.8 with  $c_2 = \frac{a}{\beta}$  and  $c_1 = \frac{\beta}{b}$ . This now completes the proof required in **(i)**.

**(ii)** Similarly, we have for  $a' = 0, b = b', \alpha' > \alpha$  via (2.12), with  $\bar{F} \equiv 1 - F$ ,

$$\begin{aligned} \phi(\alpha, \beta) &= \int_0^\infty \log \left( \frac{\bar{F}_\alpha\left(\frac{\beta}{a}F_\alpha^{-1}(F_{\alpha'}(x))\right)}{\bar{F}_\alpha\left(\frac{\beta}{b}F_\alpha^{-1}(F_{\alpha'}(x))\right)} - 1 \right) dF_{\alpha'}(x) \\ &< \int_0^\infty \log \left( \frac{\bar{F}_{\alpha'}\left(\frac{\beta}{a}x\right)}{\bar{F}_{\alpha'}\left(\frac{\beta}{b}x\right)} - 1 \right) = \phi(\alpha', \beta), \end{aligned}$$

completing the proof for **(ii)**.

**(iii)** As above for  $C' = \mathbb{R}_+$  and  $\alpha' > \alpha$ , the result follows from (2.12) as

$$\begin{aligned} \phi(\alpha, \beta) &= \int_0^\infty \log \left( F_\alpha\left(\frac{\beta}{a}F_\alpha^{-1}(F_{\alpha'}(x))\right) - F_\alpha\left(\frac{\beta}{b}F_\alpha^{-1}(F_{\alpha'}(x))\right) \right) dF_{\alpha'}(x) \\ &\leq \int_0^\infty \log \left( F_{\alpha'}\left(\frac{\beta}{a}x\right) - F_{\alpha'}\left(\frac{\beta}{b}x\right) \right) dF_{\alpha'}(x) = \phi(\alpha', \beta), \end{aligned}$$

with equality if and only if  $\beta = a, b = +\infty$  or  $a = 0, \beta = b$ .

With this last observation and the strict inequalities in **(i)** and **(ii)**, part **(b)** is established and the proof is complete.

**Remark 2.10** (A) *The dominance findings of Theorem 2.9 are illustrated below in Section 7. The general theme concerns the plausible frequentist risk improvement when comparing the predictive density estimators  $\hat{q}_{\pi_0, C}(\cdot; X)$  and  $\hat{q}_{\pi_0, C'}(\cdot; X)$  for  $\beta \in C$ , and  $C$  a strict subset of  $C'$ . Others have come across and established similar results in the point estimation literature when  $C'$  corresponds to the unrestricted parameter space. For instance, Hartigan's result (Hartigan, 2003) applies for  $p$ -variate normal models, squared error loss and  $C' = \mathbb{R}^p$ . In this sense, Theorem 2.9's dominance result applied to  $C = (a, b)$ , and  $C' = (a, \infty)$  or  $C' = (0, b)$ , is a departure on known restricted parameter space results.*

(B) *Along the same theme, and as an example of non-dominance when one truncates the prior, we point out that  $\hat{q}_{\pi_0, C}(\cdot; X)$  **does not** dominate  $\hat{q}_{\pi_0, C'}(\cdot; X)$  when  $C = (a, \infty)$  and  $C' = (a', \infty)$  with  $a' > a$ . Indeed, Theorem 2.9 tells us that  $R_{KL}(a, \hat{q}_{\pi_0, C}) = R_{KL}(a, \hat{q}_{\pi_0}) > R_{KL}(a, \hat{q}_{\pi_0, C'})$ . The same is true for  $C = (0, b)$  and  $C' = (0, b')$  with  $b < b'$ .*

## 2.4 Minimax estimators and minimax risk

The prior  $\pi_0(\beta) = \frac{1}{\beta} \mathbb{I}_{(0, \infty)}(\beta)$  is not adapted to the constraint  $\beta \in C = (a, b)$ . And, we have just seen that its truncation  $\pi_{0, C}$  is efficient in the sense that the predictive density estimator  $\hat{q}_{\pi_0, C}(\cdot; X)$  dominates, under KL loss,  $\hat{q}_{\pi_0}(\cdot; X)$  as an estimator of  $q \sim \text{Ga}(\alpha_2, \beta)$  based on  $X \sim \text{Ga}(\alpha_1, \beta)$ . However, as shown by Kubokawa et al. (2013),  $\hat{q}_{\pi_0}(\cdot; X)$  remains minimax whenever  $a = 0$  or  $b = +\infty$ . For such parametric restrictions, it thus provides a useful benchmark with dominating estimators, such as  $\hat{q}_{\pi_0, C}$ , being necessarily minimax, and with its constant risk yielding the minimax risk.

We evaluate this constant risk below and we first recall and establish some facts about the

digamma function given by  $\psi(t) = \frac{\Gamma'(t)}{\Gamma(t)} = \frac{d}{dt} \log \Gamma(t)$ , where  $\Gamma(t)$  is the Gamma function given by  $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$  and  $\log$  is hereafter used in base  $e$ .

**Lemma 2.11** (a) *We have  $\psi(\alpha) = E_\alpha[\log Z]$ , where  $Z \sim \text{Ga}(\alpha, 1)$ .*

(b) *Also  $\psi(1) = -\gamma$ , the recursive property  $\psi(t+1) = \psi(t) + \frac{1}{t}$  holds for  $t > 0$  which is equivalent to  $\psi(t+1) = -\gamma + \sum_{k=1}^t \frac{1}{k}$  for positive integer  $t$ , where  $\gamma$  is Euler's constant given by  $\gamma = \lim_{m \rightarrow \infty} (\sum_{k=1}^m \frac{1}{k} - \log m) \approx 0.577216$ .*

(c) *The function  $\psi(t) - \log t$  increases in  $t$ , for  $t > 0$ .*

**Proof.** Part (a) is verified by differentiating  $\Gamma(\alpha) = \int_{\mathbb{R}_+} t^{\alpha-1} e^{-t} dt$  under the integral sign, while the properties in (b) are well documented (see for instance Andrews, Askey and Roy, 1999). For (c), the stronger result  $\psi'(t) - \frac{1}{t} - \frac{1}{2t^2} > 0$ , for all  $t \in \mathbb{R}_+$  is given in Muldoon (1978).

**Corollary 2.12** *The Kullback-Leibler risk of  $\hat{q}_{\pi_0}(\cdot; X)$  for estimating the density of  $Y \sim \text{Gamma}(\alpha_2, \beta)$  based on  $X \sim \text{Gamma}(\alpha_1, \beta)$ , is constant and given by*

$$R_{KL}(\beta, \hat{q}_{\pi_0}) = (\alpha_1 + \alpha_2)\psi(\alpha_1 + \alpha_2) + \log \Gamma(\alpha_1) - \log \Gamma(\alpha_1 + \alpha_2) - \alpha_2 - \alpha_1\psi(\alpha_1). \quad (2.13)$$

*For the unrestricted parameter space  $(0, \infty)$  and for restricted parameter spaces of the form  $\beta \in (a, b)$  with  $a = 0$  or  $b = +\infty$ , the above is equal to the minimax risk, which is moreover independent of the parameter space endpoints.*

**Proof.** The minimaxity follows from Kubokawa et al. (2013). For the evaluation of the risk, we have from (2.3) and Theorem 2.1

$$\begin{aligned} R_{KL}(\beta, \hat{q}_{\pi_0}) &= E^{X,Y} \left[ \log \left( \frac{q(Y|\beta)}{\hat{q}_{\pi_0}(Y; X)} \right) \right] \\ &= E^{X,Y} \left[ \log \left( \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_1 + \alpha_2)} \right) + \alpha_2 \log \left( \frac{X}{\beta} \right) - \frac{Y}{\beta} + (\alpha_1 + \alpha_2) \{ \log(X + Y) - \log X \} \right] \\ &= \log \left( \frac{\Gamma(\alpha_1)}{\Gamma(\alpha_1 + \alpha_2)} \right) - \alpha_1\psi(\alpha_1) - \alpha_2 + (\alpha_1 + \alpha_2)\psi(\alpha_1 + \alpha_2), \end{aligned}$$

with the property  $X + Y \sim \text{Ga}(\alpha_1 + \alpha_2, \beta)$  and Lemma 2.11.

We will make use of the above constant risk for comparisons later on (see Example 2.17). With evaluations of the digamma and the gamma functions, we obtain explicit values for integer  $\alpha_1$  and  $\alpha_2$ . As an example, for  $\alpha_1 = 2, \alpha_2 = 1$ , we obtain  $R_{KL}(\beta, \hat{q}_{\pi_0}) = \frac{3}{2} - \log 2 - \gamma \approx 0.22916$  by making use of Lemma 2.11.

## 2.5 Plug-in predictive density estimators

Plug-in density estimators are ubiquitous in statistical theory and practice. It may well be appealing to reduce the action space to plug-in density estimators. For our model, they are of the form  $q_{\hat{\beta}}(\cdot; X) \sim \text{Ga}(\alpha_2, \hat{\beta}(X))$ , where  $\hat{\beta}(X)$  is a point estimator of  $\beta$ . Such estimators include the maximum likelihood predictive density estimator  $\hat{q}_{mle} = q_{\hat{\beta}_{mle}} \sim \text{Ga}(\alpha_2, \hat{\beta}_{mle}(X))$ , where  $\hat{\beta}_{mle}(X)$  is the point estimator mle of  $\beta$ . As the next result describes, the Kullback-Leibler loss and frequentist risk incurred by plug-in density estimators are connected to a dual entropy loss and risk.

**Lemma 2.13** *The Kullback-Leibler loss associated with the plug-in density estimate  $q_{\hat{\beta}}(\cdot; x) \sim \text{Ga}(\alpha_2, \hat{\beta}(x))$  is given by the entropy loss*

$$L_0(\beta, \hat{\beta}) = \alpha_2 \left( \frac{\beta}{\hat{\beta}} - \log\left(\frac{\beta}{\hat{\beta}}\right) - 1 \right) \quad (2.14)$$

*in the corresponding point estimation problem. Consequently, the frequentist Kullback-Leibler risk of the plug-in density estimator  $q_{\hat{\beta}}(\cdot; X) \sim \text{Ga}(\alpha_2, \hat{\beta}(X))$  is equal to the frequentist risk of the point estimator  $\hat{\beta}(X)$  of  $\beta$  under the above loss  $L_0$ .*

**Proof.** We have by a direct evaluation

$$\begin{aligned} L_{KL}(\beta, q_{\hat{\beta}}(\cdot; x)) &= \int_{\mathbb{R}_+} \log\left(\frac{q(y|\beta)}{q_{\hat{\beta}}(y|x)}\right) q(y|\beta) dy \\ &= \int_{\mathbb{R}_+} \left( y \left( \frac{1}{\hat{\beta}(x)} - \frac{1}{\beta} \right) + \alpha_2 \left( \log\left(\frac{\hat{\beta}(x)}{\beta}\right) \right) \right) q(y|\beta) dy \\ &= \alpha_2 \left( \frac{\beta}{\hat{\beta}(x)} - \log\left(\frac{\beta}{\hat{\beta}(x)}\right) - 1 \right), \end{aligned}$$

since  $E_\beta(Y) = \alpha_2 \beta$ .

**Remark 2.14** *For the problem of estimating  $\beta$  under entropy loss  $L_0$ , equivariant estimators (under changes of scale) are multiples of  $X$ , i.e., of the form  $\hat{\beta}_c(X) = cX$ . For  $\alpha_1 > 1$ , such estimators have constant entropy risk, and the optimal or best equivariant estimator is easily shown to be given by  $\hat{\beta}_{mre}(X) = \frac{X}{\alpha_1 - 1}$ . This result is well known and it is also the case that the minimum risk equivariant (mre) estimator is minimax for the unrestricted parameter space, and Generalized Bayes with respect to the prior  $\frac{1}{\beta} \mathbb{I}_{(0, \infty)}(\beta)$  (e.g., Lehmann and Casella, 1998). In view of Lemma 2.13's duality and these known results for entropy loss  $L_0$ , it is immediate that the subclass of plug-in predictive density estimators  $\hat{q} \sim \text{Ga}(\alpha_2, cX)$  have a constant Kullback-Leibler risk, and the optimal choice is given by  $c = \frac{1}{\alpha_1 - 1}$ .*

*More generally, comparisons at the entropy risk level carry-over to the predictive Kullback-Leibler risk for assessing the performance of plug-in predictive density estimators. As a further illustration, Marchand and Strawderman (2005) showed for the restricted parameter space  $\beta \in C = (a, b)$  that the generalized Bayes estimator  $\hat{\beta}_{\pi_C}(X)$  with respect to the prior  $\pi_C(\beta) = \frac{1}{\beta} \mathbb{I}_C(\beta)$ , dominates  $\hat{\beta}_{mre}(X)$ . It thus follows that the plug-in predictive density estimator  $q_{\hat{\beta}_{\pi_C}}(\cdot; X)$  dominates the plug-in  $q_{\hat{\beta}_{mre}}(\cdot; X)$  in terms of Kullback-Leibler risk for the restricted parameter space  $\beta \in C = (a, b)$ .*

We further focus in this section on both **(i)** the lower bound constraint  $\beta \geq a$  and **(ii)** the upper bound constraint  $\beta \leq b$ , in the study of the performance of plug-in estimators  $\hat{q}_c(\cdot; x) \sim \text{Ga}(\alpha_2, \hat{\beta}_c)$ ,  $c > 0$ , where  $\hat{\beta}_c$  is a truncated linear estimator of  $\beta$  given by

$$\hat{\beta}_c(x) = \begin{cases} a & \text{if } cx \leq a \\ cx & \text{if } a < cx < b \\ b & \text{if } cx \geq b \end{cases} . \quad (2.15)$$

These predictive density or point estimators are interesting since they include the mle for  $c = \frac{1}{\alpha_1}$ , as well as the truncation of  $\hat{\beta}_{mre}$  onto the parameter space for  $c = \frac{1}{\alpha_1 - 1}$ . Furthermore, for estimating  $\beta$  under entropy loss  $L_0$ , the truncated MRE estimator is minimax for both the

restrictions  $\beta \geq a$  and  $\beta \leq b$ , since it dominates the MRE estimator, with the latter being minimax despite the fact that it is not adapted to the restricted parameter space (e.g., Marchand and Strawderman, 2005). However, for the predictive density estimation problem, the situation is quite different and we will show that even the minimax choice among the  $\hat{q}_c$ 's is “far” from being minimax.

**Lemma 2.15** *Let  $F_{\alpha_1}$  and  $\bar{F}_{\alpha_1} \equiv 1 - F_{\alpha_1}$  be the  $Ga(\alpha_1, 1)$  cumulative and survivor functions respectively. For estimating the density of  $Y \sim \text{Gamma}(\alpha_2, \beta)$  with  $\beta \in C$  based on  $X \sim \text{Gamma}(\alpha_1, \beta)$ , the Kullback-Leibler risk of the plug-in predictive  $\hat{q}_c \sim Ga(\alpha_2, \hat{\beta}_c)$ , where  $\hat{\beta}_c$  is given in (2.15), is equal to*

$$(i) \quad \alpha_2 \left( \left( \frac{\beta}{a} - \log\left(\frac{\beta}{a}\right) \right) F_{\alpha_1}\left(\frac{a}{c\beta}\right) - 1 + \int_{\frac{a}{c\beta}}^{\infty} \left( \frac{1}{cx} + \log c + \log x \right) p(x|1) dx \right), \text{ for } \beta \geq a, \quad (2.16)$$

$$(ii) \quad \alpha_2 \left( \left( \frac{\beta}{b} - \log\left(\frac{\beta}{b}\right) \right) \bar{F}_{\alpha_1}\left(\frac{b}{c\beta}\right) - 1 + \int_0^{\frac{b}{c\beta}} \left( \frac{1}{cx} + \log c + \log x \right) p(x|1) dx \right), \text{ for } \beta \leq b. \quad (2.17)$$

**Proof.** By virtue of Lemma 2.13, we have in **(i)**

$$\begin{aligned} R_{KL}(\beta, \hat{q}_c) &= \alpha_2 E \left( \frac{\beta}{\hat{\beta}_c(X)} - \log\left(\frac{\beta}{\hat{\beta}_c(X)}\right) - 1 \right) \\ &= -\alpha_2 + \alpha_2 \left( \int_0^{a/c} \left( \frac{\beta}{a} - \log\left(\frac{\beta}{a}\right) \right) p(x|\beta) dx + \int_{a/c}^{\infty} \left( \frac{\beta}{cx} - \log\left(\frac{\beta}{cx}\right) \right) p(x|\beta) dx \right), \end{aligned}$$

which leads to (2.16) with a change of variables  $x \rightarrow x/\beta$ . Finally, the Kullback-Leibler risk in (2.17) for situation **(ii)**  $\beta \leq b$  is obtained in a similar fashion.

The following is an analysis of the performance of the predictive density estimators  $\hat{q}_c$  in terms of a maximum risk or minimax criterion.

**Theorem 2.16** Consider the subclass of predictive density estimators  $\hat{q}_c \sim Ga(\alpha_2, \hat{\beta}_c)$ , with  $\hat{\beta}_c$  a truncated linear estimator as given in (2.15), for estimating the density of  $Y \sim \text{Gamma}(\alpha_2, \beta)$ , with  $\beta \in C$ , based on  $X \sim \text{Gamma}(\alpha_1, \beta)$  with  $\alpha_1 > 1$ . Then,

(a) The Kullback-Leibler risk of  $\hat{q}_c$  is increasing in  $\beta \in C$  for situation : (i)  $\beta \geq a$ , and decreasing in  $\beta \in C$  for situation : (ii)  $\beta \leq b$ ;

(b) The minimax procedure and risk among the  $\hat{q}_c$ 's are given by the choice  $c = \frac{1}{\alpha_1 - 1}$ , and the value

$$\inf_{c > 0} \{ \sup_{\beta \in C} R_{KL}(\beta, \hat{q}_c) \} = \alpha_2 (\psi(\alpha_1) - \log(\alpha_1 - 1)), \quad (2.18)$$

independently of  $C = [a, \infty)$  or  $C = (0, b]$ ;

(c) The difference in risks  $\Delta_{c_1, c_2}(\beta) = R_{KL}(\beta, \hat{q}_{c_1}) - R_{KL}(\beta, \hat{q}_{c_2})$ , where  $0 < c_1 < c_2$  increases in  $\beta$  for both situations (i)  $\beta \geq a$ , and situation (ii)  $\beta \leq b$ . Consequently,  $\hat{q}_{(\alpha_1 - 1)^{-1}}$  dominates all  $\hat{q}_c$  with  $c > \frac{1}{\alpha_1 - 1}$  in situation (i)  $\beta \geq a$ , and  $\hat{q}_{(\alpha_1 - 1)^{-1}}$  dominates all  $\hat{q}_c$  with  $c < \frac{1}{\alpha_1 - 1}$ , including  $\hat{q}_{mle}$ , in situation (ii)  $\beta \leq b$ .

**Proof.** (a) A calculation using (2.16) yields the expression  $\frac{d}{d\beta} R_{KL}(\beta, \hat{q}_c) = \frac{\alpha_2}{\beta} (\frac{\beta}{a} - 1) F_{\alpha_1}(\frac{a}{c\beta}) > 0$  for  $\beta > a$ , which implies the result for (i). Similarly, we obtain  $\frac{d}{d\beta} R_{KL}(\beta, \hat{q}_c) = \frac{\alpha_2}{\beta} (\frac{\beta}{b} - 1) \bar{F}_{\alpha_1}(\frac{b}{c\beta}) < 0$  for  $\beta < b$ .

(b) From part (a), we obtain for case (i)  $\beta \geq a$

$$\begin{aligned} \sup_{\beta \geq a} R_{KL}(\beta, \hat{q}_c) &= \lim_{\beta \rightarrow \infty} R_{KL}(\beta, \hat{q}_c) \\ &= -\alpha_2 + \alpha_2 \left( \int_0^\infty \left( \frac{1}{cx} + \log c + \log x \right) p(x|1) dx \right) \\ &= \alpha_2 \left( \frac{1}{c(\alpha_1 - 1)} + \log c + \psi(\alpha_1) - 1 \right), \end{aligned} \quad (2.19)$$

making use of  $E_1(\frac{1}{X}) = \frac{1}{\alpha_1 - 1}$ , as well as part (a) of Lemma 2.11, and observing that  $\lim_{z \rightarrow \infty} z F_{\alpha_1}(1/z) = \lim_{z \rightarrow \infty} p(\frac{1}{z} | \alpha_1) = 0$  for  $\alpha_1 > 1$ . Similarly, for case (ii)  $\beta \leq b$ , the supremum risk is attained as  $\beta \rightarrow 0$  yielding (2.19) as well. In both cases, minimization of (2.19) leads to the stated result.

(c) Again here, both the lower bounded and upper bounded cases (i) and (ii) are analogous,



so we only tackle the upper bounded case  $\beta \leq b$ . It follows from the derivative in **(a)** that  $\frac{\partial}{\partial \beta} \Delta_{c_1, c_2}(\beta) = \frac{\alpha_2}{\beta} \left( \frac{\beta}{b} - 1 \right) \left( \bar{F}_{\alpha_1} \left( \frac{b}{c_1 \beta} \right) - \bar{F}_{\alpha_1} \left( \frac{b}{c_2 \beta} \right) \right) > 0$  for  $0 < \beta < b$  and  $c_1 < c_2$ . Furthermore, for  $c_1 < \frac{1}{\alpha_1 - 1}$  and  $0 < \beta \leq b$ , we have  $\Delta_{c_1, \frac{1}{\alpha_1 - 1}}(\beta) > \Delta_{c_1, \frac{1}{\alpha_1 - 1}}(0^+) > 0$  given that  $\hat{q}_{\frac{1}{\alpha_1 - 1}}$  minimizes the risk at  $\beta \rightarrow 0^+$  among estimators  $\hat{q}_c$ .

Both plug-in predictive density estimators  $\hat{q}_c$  and the Bayes predictive density estimator  $\hat{q}_{\pi_0, C}$  exploit the restricted parameter information  $\beta \in C$ . As expanded on with the next example, and in opposition to a large collection of various point estimation problems (e.g., Marchand and Strawderman, 2012), they differ in efficiency and the difference can be significant. Actually, the maximum risk of all plug-in's  $\hat{q}_c$  even exceeds the constant risk of the predictive MRE density  $\hat{q}_{\pi_0}$  which ignores the lower bound restriction.

**Example 2.17** *We illustrate here some of the above findings as applied to restricted parameter spaces  $C = (a, b)$  with either  $a = 0$  or  $b = +\infty$ . Denote  $\bar{R}_{KL}(\hat{q})$  the supremum risk of a predictive density estimator  $\hat{q}$ . Consider the plug-in predictive density estimators  $\hat{q}_{mle}$  and  $\hat{q}_{1/(\alpha_1 - 1)}$ , as well as the minimax estimator  $\hat{q}_{\pi_0, C}$ . Our findings above tell us that  $\bar{R}_{KL}(\hat{q}_{mle}) > \bar{R}_{KL}(\hat{q}_{1/(\alpha_1 - 1)}) > \bar{R}_{KL}(\hat{q}_{\pi_0, C})$ . From (2.19), we have  $\bar{R}_{KL}(\hat{q}_{mle}) = \alpha_2 \left( \psi(\alpha_1) + \frac{\alpha_1}{\alpha_1 - 1} - \log \alpha_1 - 1 \right)$ , while  $\bar{R}_{KL}(\hat{q}_{1/(\alpha_1 - 1)})$  and  $\bar{R}_{KL}(\hat{q}_{\pi_0, C})$  are given by (2.18) and (2.13) respectively. As a numerical illustration, consider  $\alpha_1 = 2$  and  $\alpha_2 = 1$ . Evaluations yield  $\bar{R}_{KL}(\hat{q}_{mle}) = 2 - \log 2 - \gamma \approx 0.729627$ ,  $\bar{R}_{KL}(\hat{q}_{1/2}) = 1 - \gamma \approx 0.423784$ , and  $\bar{R}_{KL}(\hat{q}_{\pi_0, C}) = 2 - \log 2 - \gamma \approx 0.22916$ . The differences are quite important here in relative terms, but are attenuated for larger  $\alpha_1$ . This is further illustrated in Figure 1, where we have exhibited the ratio of minimax risks*

$$\frac{\inf_{c > 0} \{ \sup_{\beta \geq a} R_{KL}(\beta, \hat{q}_c) \}}{\inf_{\hat{q}} \{ \sup_{\beta \geq a} R_{KL}(\beta, \hat{q}) \}} = \frac{\alpha_2 (\psi(\alpha_1) - \log(\alpha_1 - 1))}{(\alpha_1 + \alpha_2) \psi(\alpha_1 + \alpha_2) + \log \Gamma(\alpha_1) - \log \Gamma(\alpha_1 + \alpha_2) - \alpha_2 - \alpha_1 \psi(\alpha_1)},$$

as a function of  $\alpha_1$ , for a selection of  $\alpha_2$  values. Finally, we point out the elegant difference

$$\bar{R}_{KL}(\hat{q}_{mle}) - \bar{R}_{KL}(\hat{q}_{1/(\alpha_1 - 1)}) = \alpha_2 \left( \frac{\alpha_1}{\alpha_1 - 1} - \log \left( \frac{\alpha_1}{\alpha_1 - 1} \right) - 1 \right),$$

and reiterate that this expression, as well as the findings of this section relative to the performance

of the predictive density estimators  $\hat{q}_c$ , also represent point estimation results under entropy loss  $L_0$  for truncated linear estimators of the form  $\hat{\beta}_c$ .

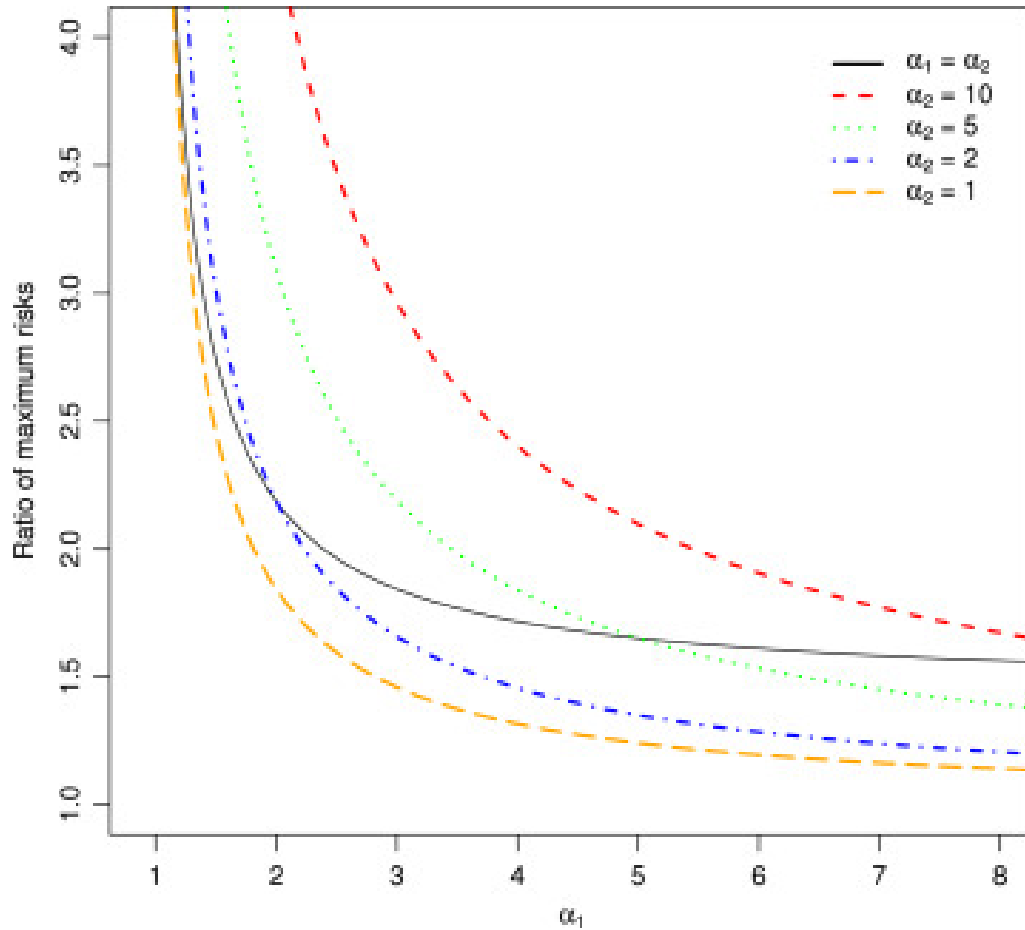


Figure 2.1 – Ratios  $\frac{\inf_c \bar{R}(\hat{q}_c)}{\inf_{\hat{q}} \bar{R}(\hat{q})}$  of minimax risks for  $\beta \geq a$ .

## 2.6 Variance expansion improvements on plug-in predictive density estimators

Previous predictive density estimation findings for location models have shown that improvements on plug-in predictive density estimators can be found by expanding the variance up to a

certain level (Fourdrinier et al. 2011; Kubokawa, Marchand and Strawderman, 2015A, 2015B). In the spirit of searching for plausible improvements, we investigate below predictive density estimators of the form  $\text{Ga}(\frac{\alpha_2}{k}, k\hat{\beta}(X))$  for  $k \geq 1$ , a given plug-in estimator  $\hat{\beta}(X)$  of  $\beta$ . For fixed  $x$ , the related expectation is constant as a function of  $k$ , but the variance  $k\alpha_2\hat{\beta}(x)$  increases in  $k$ . We now show that improvements on the plug-in predictive density estimator (i.e.,  $k = 1$ ) can always be found, regardless of the parameter space and of  $\hat{\beta}(X)$ .

**Theorem 2.18** *Let  $q_{\hat{\beta}}(\cdot; X) \sim \text{Ga}(\alpha_2, \hat{\beta}(X))$  be a plug-in estimator, with non-degenerate  $\hat{\beta}(X)$ , for estimating the density of  $Y \sim \text{Gamma}(\alpha_2, \beta)$  under Kullback-Leibler loss with  $\beta \in C = (a, b)$ , and based on  $X \sim \text{Gamma}(\alpha_1, \beta)$ . Denote  $R(\beta, \hat{\beta}) = E(\frac{\beta}{\hat{\beta}(X)} - \log(\frac{\beta}{\hat{\beta}(X)}) - 1)$  the entropy risk of  $\hat{\beta}(X)$  and denote  $\underline{R}(\hat{\beta}) = \inf_{\beta \in C} R(\beta, \hat{\beta})$ . Then,  $q_{\hat{\beta}}(\cdot; X)$  is inadmissible and dominated by the subclass of predictive density estimators  $q_{\hat{\beta},k}(\cdot; X) \sim \text{Ga}(\frac{\alpha_2}{k}, k\hat{\beta}(X))$  with  $1 < k \leq k_0(\underline{R}(\hat{\beta}))$ ,  $k_0(\underline{R}(\hat{\beta}))$  being the unique solution in  $k \in (1, \infty)$  of the equation  $G_{\underline{R}(\hat{\beta})}(k) = 0$ , with*

$$G_s(k) = \alpha_2 \left( \frac{1}{k} - 1 \right) (s + 1 - \psi(\alpha_2)) + \frac{\alpha_2}{k} \log k + \log \frac{\Gamma(\frac{\alpha_2}{k})}{\Gamma(\alpha_2)}.$$

**Proof.** First fix  $\beta \in (a, b)$  and set

$$h(k, \beta) = R_{KL}(\beta, q_{\hat{\beta},k}).$$

From (2.3), we obtain

$$\begin{aligned} h(k, \beta) &= E \left\{ \alpha_2 \left( 1 - \frac{1}{k} \right) \log \left( \frac{Y}{\beta} \right) + \frac{Y}{\beta} \left( \frac{\beta}{k\hat{\beta}(X)} - 1 \right) + \frac{\alpha_2}{k} \log \left( \frac{k\hat{\beta}(X)}{\beta} \right) + \log \Gamma \left( \frac{\alpha_2}{k} \right) - \log \Gamma(\alpha_2) \right\} \\ &= \alpha_2 (\psi(\alpha_2) - 1) - \log \Gamma(\alpha_2) + \frac{\alpha_2}{k} \left( \log k - \psi(\alpha_2) + 1 + R(\beta, \hat{\beta}) \right) + \log \Gamma \left( \frac{\alpha_2}{k} \right), \end{aligned} \quad (2.20)$$

by the independence of  $X$  and  $Y$ , and by using the identities  $E_{\alpha_2}(\frac{Y}{\beta}) = \alpha_2$  and  $E_{\alpha_2}(\log(\frac{Y}{\beta})) = \psi(\alpha_2)$ . By differentiation, we obtain  $\frac{d}{dk} h(k, \beta) = \frac{\alpha_2}{k^2} T(k)$ , with

$$T(k) = \psi(\alpha_2) - R(\beta, \hat{\beta}) - \psi \left( \frac{\alpha_2}{k} \right) - \log k. \quad (2.21)$$

Now, observe that  $T(k)$  increases in  $k \geq 1$  as an application of part (c) of Lemma 2.11. Observe as well that  $\lim_{k \rightarrow \infty} h(k, \beta) = +\infty$ . Since  $\frac{d}{dk} h(k, \beta)|_{k=1+} = -\alpha_2 R(\beta, \hat{\beta}) < 0$ , it follows that

that  $\frac{\partial}{\partial k} h(k, \beta)$  changes signs from  $-$  to  $+$  as  $k$  increases from 1 to  $\infty$ , which implies that  $h(k, \beta) - h(1, \beta) < 0$  iff  $1 < k < k_0(\beta)$ , where  $G_{R(\beta, \hat{\beta})}(k_0(\beta)) = 0$ . With  $k_0(\beta)$  increasing in  $R(\beta, \hat{\beta})$ , the result follows.

**Remark 2.19** *Theorem 2.18 is appealing since it is universal in our context, applying for all parameter spaces  $C = (a, b)$  and all plug-in predictive density estimators  $q_{\hat{\beta}}$  with non-degenerate  $\hat{\beta}(X)$ . We also point out that, while the Gamma distributed assumption for  $Y$  is necessary, the same is not true for the distribution of  $X$  and the theorem also holds for other scale family models  $X \sim \frac{1}{\beta} f_1(\frac{x}{\beta})$ , which only impacts the risk  $R(\beta, \hat{\beta})$ , the corresponding minimum risk  $\underline{R}(\hat{\beta})$ , and the upper cut-off point  $k_0(\underline{R}(\hat{\beta}))$ . Furthermore, Theorem 2.18 does provide explicit predictive density estimators that dominate plug-in predictive density estimators, such as the predictive mle  $\hat{q}_{1/\alpha_1}$ , the minimax  $\hat{q}_c$  procedure (e.g., Theorem 2.16) given by the choice  $c = \frac{1}{\alpha_1 - 1}$  for  $\alpha_1 > 1$ , and  $q_{\hat{\beta}_{\pi_C}}$  with  $\hat{\beta}_{\pi_C}(X)$  the Bayes estimator of  $\beta$  with respect to the prior  $\pi_C(\beta) = \frac{1}{\beta} \mathbb{I}_C(\beta)$  (see Remark 2.14).*

**Remark 2.20** *Theorem 2.18 applies for the case of the unrestricted parameter space  $C = (0, \infty)$ . In particular, it applies for the plug-in MRE estimator  $\hat{\beta}(X) = \frac{X}{\alpha_1 - 1}$  when  $\alpha_1 > 1$ . However, a stronger inference is available as there exists an optimal predictive density estimator in the class of  $Ga(\frac{\alpha_2}{k}, \frac{kX}{\alpha_1 - 1})$  predictive densities with  $k > 1$ . This follows as the risk of such estimators are, as seen by (2.20), constant as a function of  $\beta$  with  $R(\beta, \hat{\beta}_{mre}) = \psi(\alpha_1) - \log(\alpha_1 - 1)$  (see 2.18), and uniquely minimized at  $k_0 = T^{-1}(0)$  where, from (2.21), we have*

$$T(k) = \psi(\alpha_2) + \log(\alpha_1 - 1) - \psi(\alpha_1) - \psi\left(\frac{\alpha_2}{k}\right) - \log k.$$

*Notwithstanding the above illustrative discussion, all these predictive density estimators remain inadmissible and, being invariant with respect to changes of scale, they are dominated by the minimum risk equivariant predictive density estimator  $\hat{q}_{\pi_0}$ .*

## 2.7 Illustrations and final comments

Numerical risk evaluations are presented in Figures 2, 3 and 4 and commented on in this section. These serve to illustrate most of the findings of this paper, are useful in assessing the strengths and weaknesses of different predictive density estimators, and are particularly representative of the degree of improvement obtained when a theoretical dominance result applies.

- Figure 2 relates to  $\alpha_1 = 3, \alpha_2 = 5, C = [1, \infty)$ . First, the risk function of the Bayes predictive density estimator  $\hat{q}_{\pi_0, C}$  is represented. It performs well overall and its minimaxity is seen to be quite beneficial. In accordance to Theorem 2.9, it dominates the Bayes predictive density  $\hat{q}_{\pi_0}$ , which is also minimax (Corollary 2.12), with matching risks at  $\beta = 1$  and at  $\beta \rightarrow \infty$  (Theorem 2.9). The exact minimax risk is given by (2.13) and is equal to  $8\psi(8) + \log \Gamma(3) - \log \Gamma(8) - 5 - 3\psi(3) = -5\gamma - \log 2520 + \frac{787}{70} \approx 0.524763$ .
- Looking again at Figure 2, the risk functions of the plug-in predictive mle  $\hat{q}_{1/3} \sim \text{Gamma}(5, \hat{\beta}_{1/3}(X))$  and  $\hat{q}_{1/2} \sim \text{Gamma}(5, \hat{\beta}_{1/2}(X))$ , as well as variance expansion improvements derived from Theorem 2.18, are also represented. The improvements are  $\hat{q}_{c, k_1} \sim \text{Gamma}(\frac{5}{k_1}, k_1 \hat{\beta}_c(X))$ ,  $c = 1/3, 1/2$ , predictive densities, with  $k_1 = k_0(R(1, \hat{\beta}_c)) \approx 1.88, 5.14$  (respectively) taken as the upper cutoff point of Theorem 2.18, and with  $\underline{R}(\hat{\beta}_c) = \inf_{\beta \geq 1} R(\beta, \hat{\beta}_c) = R(1, \hat{\beta}_c)$  by virtue of part (a) of Theorem 2.16 and Lemma 2.13. The degree of improvement is striking with  $\hat{q}_{1/2, k_1}$  even dominating  $\hat{q}_{1/3}$ . Compared to the Bayes predictive density estimator  $\hat{q}_{\pi_0, C}$ , the predictive density mle  $\hat{q}_{1/2, k_1}$  is efficient for small  $\beta$  with smaller risk for  $\beta \leq \beta_0$ , but with a less favorable performance otherwise and supremum risk well above the minimax risk. The variance inflated predictive density  $\hat{q}_{1/3, k_1}$  is more efficient in terms of supremum risk, but performs poorly being dominated by both  $\hat{q}_{\pi_0}$  and  $\hat{q}_{\pi_0, C}$ , with the former not even exploiting the parametric restriction.
- Other cases  $\alpha_1, \alpha_2$  and choices of  $k$ , many of which we looked at, yield similar comparisons with the Bayes predictive density estimator performing well in comparison to the plug-in predictive density estimators  $\hat{q}_c$  and their variance inflated improvements.

- Figure 3 is such an illustration with  $\alpha_1 = 2$ ,  $\alpha_2 = 1$ ,  $\beta \leq 1.0$ . The plug-in MRE estimator is an  $\text{Exponential}(X)$  density, while the chosen variance expansion improvement is a  $\text{Gamma}(\frac{1}{k_1}, k_1 X)$  density with  $k_1 \approx 1.745$ . The inefficiency of  $\hat{q}_{mle}(\cdot; x) \sim \text{Exponential}(\frac{x}{2})$  is more striking here, with  $\hat{q}_{1/(\alpha_1-1)}$  offering much lower risk in accordance with Theorem 2.16, and with

$$\begin{aligned} R_{KL}(\beta, \hat{q}_{mle}) - R_{KL}(\beta, \hat{q}_{1/(\alpha_1-1)}) &\geq \lim_{\beta \rightarrow 0^+} (R_{KL}(\beta, \hat{q}_{mle}) - R_{KL}(\beta, \hat{q}_{1/(\alpha_1-1)})) \\ &= 1 - \log 2, \end{aligned}$$

for all  $\beta \leq 1.0$ , as a consequence of Theorem 2.16 and (2.19).

- Finally, Figure 4 illustrates a doubly-bounded case with  $\alpha_1 = 3$ ,  $\alpha_2 = 3$  and  $C = [1, 8]$ . In accordance with Theorem 2.9, the Bayes predictive density estimator  $\hat{q}_{\pi_0, C}$  dominates both  $\hat{q}_{\pi_0, C'}$  and the MRE predictive density estimator  $\hat{q}_{\pi_0}$  with  $C' = [1, \infty)$ . It also performs quite well in comparison to the plug-in predictive density estimators  $\hat{q}_{mle}$  and  $\hat{q}_{\frac{1}{\alpha_1-1}}$  with risk improvement on a very large part of the parameter space.

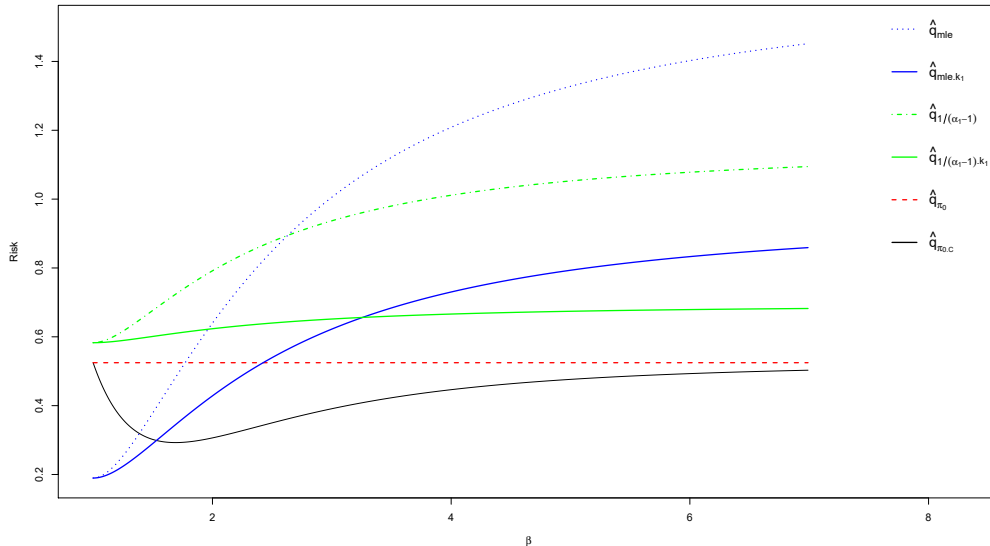


Figure 2.2 – Risks of predictive density estimators :  $\alpha_1 = 3$ ,  $\alpha_2 = 5$ ,  $\beta \geq 1$ .

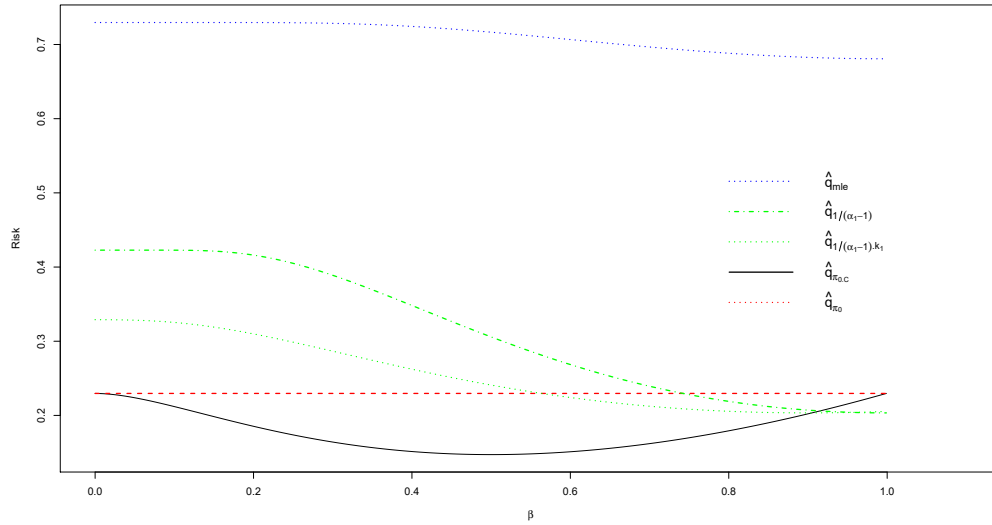


Figure 2.3 – Risks of predictive density estimators :  $\alpha_1 = 2$ ,  $\alpha_2 = 1$ ,  $\beta \leq 1.0$ .

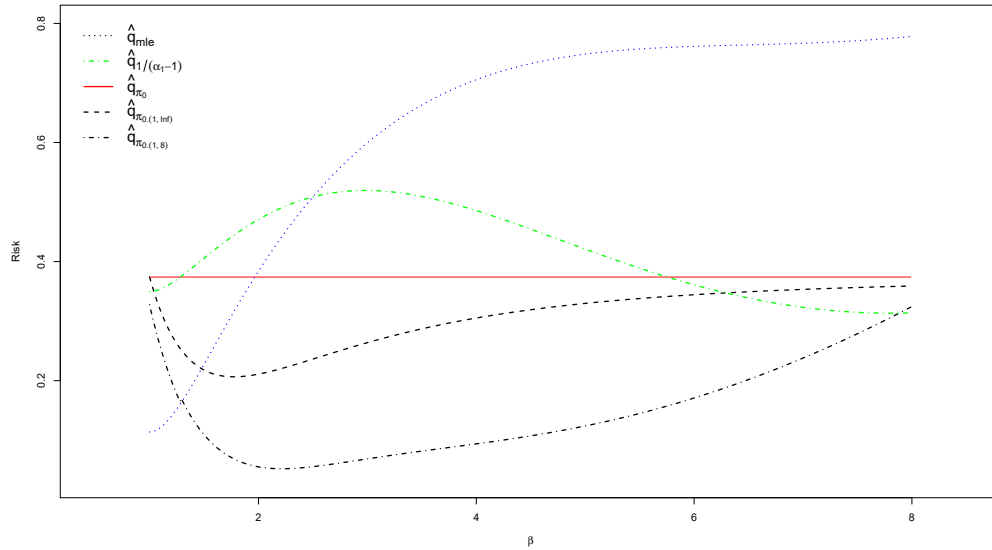


Figure 2.4 – Risks of predictive density estimators :  $\alpha_1 = 3$ ,  $\alpha_2 = 3$ ,  $\beta \in [1, 8]$ .

## Acknowledgements

Eric Marchand's research is supported in part by a grant from 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).

## References

- Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, **62**, 547-554.
- Amaral, M.A. and Dunsmore, I.R. (1980). Optimal estimates of predictive distributions. *Biometrika*, **67**, 685-689.
- Andrews, G.E., Askey, R. and Roy, R. (1999). *Special functions*. Encyclopedia of Mathematics and its Applications, **71**, Cambridge University Press.
- Boisbunon, A. and Maruyama, Y. (2014). Inadmissibility of the best equivariant predictive density in the unknown variance case. *Biometrika*, **101**, 733-740.
- Brown, L.D., George, E.I., and Xu, X. (2008). Admissible predictive density estimation. *Annals of Statistics*, **36**, 1156-1170.
- 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.
- George, E. I., Liang, F. and Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Annals of Statistics*, **34**, 78-91.
- Hartigan, J. (2004). Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, **67**, 285-288.
- 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.
- Komaki, T. (2006). A class of proper priors for Bayesian simultaneous prediction of independent Poisson observables. *Journal of Multivariate Analysis*, **97**, 1815-1828.
- Komaki, T. (2004). Simultaneous prediction of Poisson observables. *Annals of Statistics*, **32**, 1744-1769.



- Komaki, F. (2001). A shrinkage predictive distribution for multivariate normal observables. *Biometrika*, **88**, 859-864.
- Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Annals of Statistics*, **22**, 290-299.
- Kubokawa, T., Marchand, É & Strawderman, W.E. (2015A). On predictive density estimation for location families under integrated absolute error loss. To appear in *Bernoulli*.
- Kubokawa, T., Marchand, É & Strawderman, W.E. (2015B). On predictive density estimation for location families under integrated squared error loss. *Journal of Multivariate Analysis*, **142**, 57-74.
- Kubokawa, T., Marchand, É., Strawderman, W.E. and Turcotte, J-P. (2013). Minimavity in predictive density estimation with parametric constraints. *Journal of Multivariate Analysis*, **116**, 382-397.
- Lehmann, E. L., Casella, G. (1998). *Theory of Point Estimation*. Springer-Verlag, New York, 2nd edition.
- 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.
- Marchand, É., and Strawderman, W.E. (2012). A unified minimax result for restricted parameter spaces. *Bernoulli*, **18**, 635-643.
- Marchand, É., and Strawderman, W. E. (2005). 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.
- 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.
- Misra, N. and Arshad, M. (2014). Monotonicity of certain integrals involving gamma distributions and their applications in multiple comparisons. *Statistics & Probability Letters*, **85**, 144-152.

- Muldoon, M.E. (1978). Some monotonicity properties and characterizations of the gamma function. *Aequationes Math.*, **18**, 54-63.
- Pham-Gia, T. and Turkkan, N. (2002). Operations on Generalized-F variables and applications. *Statistics*, **36**, 195-209.
- van Eeden, C. (2006). *Restricted parameter space problems : Admissibility and minimaxity properties*. Lecture Notes in Statistics, **188**, Springer.
- van Zwet, W.R. (1964). Convex transformations of random variables. *Mathematical Centre Tracts*, 7 *Mathematisch Centrum, Amsterdam vi+116 pp*

# CHAPITRE 3

## On Predictive Density Estimation under $\alpha$ -divergence Loss.

Ce Chapitre contient un manuscrit écrit conjointement avec mon directeur de recherche le professeur Éric Marchand. Il comprend des résultats originaux pour l'estimation de la densité prédictive de la loi normale multivariée sous la fonction de perte de  $\alpha$ -divergence.

AZIZ L'MOUDDEN<sup>a</sup>, ÉRIC MARCHAND<sup>a</sup>,

*a Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K*

*2R1 (e-mails : aziz.lmoudden@usherbrooke.ca ; eric.marchand@usherbrooke.ca)*

## SUMMARY

Based on  $X \sim N_d(\theta, \sigma_X^2 I_d)$ , we study the efficiency of predictive densities under  $\alpha$ -divergence loss  $L_\alpha$  for estimating the density of  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$ . We identify a large number of cases where improvement on a plug-in density are obtainable by expanding the variance, thus extending earlier findings applicable to Kullback-Leibler loss. The results and proofs are unified with respect to the dimension  $d$ , the variances  $\sigma_X^2$  and  $\sigma_Y^2$ , the choice of loss  $L_\alpha$ ;  $\alpha \in (-1, 1)$ . The findings also apply to a large number of plug-in densities, as well as for restricted parameter spaces with  $\theta \in \Theta \subset \mathbb{R}^d$ . The theoretical findings are accompanied by various observations, illustrations, and implications dealing for instance with robustness with respect to the model variances and simultaneous dominance with respect to the loss.

*AMS 2010 subject classifications* : 62C20, 62C86, 62F10, 62F15, 62F30

*Keywords and phrases* : Alpha-divergence ; Dominance ; Frequentist risk ; Hellinger loss ; Multivariate normal ; Plug-in ; Predictive density ; Restricted parameter space ; Variance expansion.

## 3.1 Introduction

Consider normally and independently distributed  $X|\theta \sim N_d(\theta, \sigma_X^2 I_d)$  and  $Y|\theta \sim N_d(\theta, \sigma_Y^2 I_d)$  and the objective of predicting  $Y$  having observed  $X$ . For predictive analysis purposes, researchers are interested in finding a predictive density  $\hat{q}(\cdot; X)$  of the density of  $Y$ . In turn, such a density may play a surrogate role for generating either future or missing values of  $Y$ . To evaluate the performance of such predictive densities, attractive choices are given by the family of  $\alpha$ -divergence losses (e.g., Csiszàr, 1967)

$$L_\alpha(\theta, \hat{q}(\cdot|x)) = \int_{\mathbb{R}^d} h_\alpha \left( \frac{\hat{q}(y|x)}{q(y|\theta)} \right) q(y|\theta) dy, \quad (3.1)$$

where

$$h_\alpha(z) = \begin{cases} \frac{4}{1-\alpha^2} \left(1 - z^{\frac{1+\alpha}{2}}\right) & |\alpha| < 1 \\ z \log z & \alpha = 1 \\ -\log z & \alpha = -1, \end{cases} \quad (3.2)$$

and the corresponding frequentist risk of  $\hat{q}(\cdot; X)$  given by

$$R_\alpha(\theta, \hat{q}) = E^X [L_\alpha(\theta, \hat{q}(\cdot | X))]. \quad (3.3)$$

Notable examples of  $L_\alpha$  include Kullback-Leibler ( $L_{-1}$ ), reverse Kullback-Leibler ( $L_1$ ), and Hellinger ( $L_0/4$ ). We point out that it is equivalent to consider

$$h_\alpha(z) = \begin{cases} \frac{4}{1-\alpha^2} \left(\frac{1+\alpha}{2}z - z^{\frac{1+\alpha}{2}} + \frac{1+\alpha}{2}\right) & |\alpha| < 1 \\ 1 - z + z \log z & \alpha = 1 \\ z - \log z - 1 & \alpha = -1. \end{cases} \quad (3.4)$$

With these functions being nonnegative, equal to 0 iff  $z = 1$ , decreasing for  $z \in (0, 1)$ , increasing for  $z > 1$ , they connect in a more straightforward manner to desirable features for losses in (3.1). The cases  $|\alpha| < 1$  stand apart, and merit study, as they typically lead to finite loss, unlike the cases of Kullback-Leibler and reverse Kullback-Leibler losses.

This paper is concerned with improvements on plug-in predictive densities of the form

$$q_{\hat{\theta},1} \sim N_d(\hat{\theta}(X), \sigma_Y^2 I_d), \quad (3.5)$$

where  $\hat{\theta}(X)$  is a non-degenerate point estimator of  $\theta$ , and where  $\theta \in C \subset \mathbb{R}^d$ . Whereas, such plug-in densities were shown in Fourdrinier et al. (2011) to be universally deficient for Kullback-Leibler risk, and improved upon by a subclass of scale expansion variants

$$q_{\hat{\theta},c} \sim N_d(\hat{\theta}(X), c^2 \sigma_Y^2 I_d), \quad (3.6)$$

with  $1 < c < c_0$ , whereas **all** Bayesian predictive densities are plug-in densities for reverse Kullback-Leibler loss (Yanigomoto and Ohnishi, 2009; Marchand and Sadeghkhani, 2017), analogous analytical results are lacking for the cases  $|\alpha| < 1$ . This manuscript fills this gap. Improvements are obtained, interpreted, and illustrated. The main result (Theorem 3.9) reproduces for a large class of choices of  $\hat{\theta}$  the Kullback-Leibler result with predictive density estimators  $q_{\hat{\theta},c}$  dominating the plug-in  $q_{\hat{\theta},1}$  for expansions  $c \in (1, c_0)$ . The cut-off point  $c_0$  depends on  $\alpha, \hat{\theta}$ , the parameter space  $C$ , the dimension  $d$ ,  $\sigma_X^2$ , and  $\sigma_Y^2$ , but the phenomenon of improvement by variance expansion is otherwise quite general, subject to conditions, namely on  $\hat{\theta}$  which we address. Further inferences become available as well. On one hand, we deduce simultaneous dominance results with representatives  $q_{\hat{\theta},c}$  dominating a given plug-in density  $q_{\hat{\theta},1}$  for many  $\alpha$ -divergence losses, including Kullback-Leibler. On

the other hand, we expand on the fact that dominating predictive densities are not necessarily restricted to normal densities in (3.6), but also include variance mixture of normals with the variance mixing variable taking values on  $(1, c_0)$ . The scope of our results is also enlarged in view of applications to restricted parameter spaces  $C$ , including univariate means constrained to an interval or a half-interval,  $L^2$  balls  $C = \{\theta \in \mathbb{R}^d : \|\theta\| \leq m\}$ , order constraints such as  $C = \{\theta \in \mathbb{R}^d : \theta_{i+1} \geq \theta_i \text{ for } i = 1, \dots, d-1\}$  among many other types (see Remark 3.11).

The predictive density estimation framework considered here was put forth for Kullback-Leibler divergence loss in the pioneering work of Aitchison and Dunsmore (1975), as well as Aitchison (1975), and has found applications in information theory, econometrics, machine learning, image processing, and mathematical finance, among others. For multivariate normal observables, as considered here, much interest was generated following Komaki (2001) where Bayesian improvements on the minimum risk equivariant predictive density; corresponding to  $\hat{\theta}(X) = X$ ,  $c^2 = 1 + \sigma_X^2/\sigma_Y^2$  in (3.6); were obtained for  $d \geq 3$  and Kullback-Leibler loss. For Kullback-Leibler loss, further findings with respect to minimaxity, admissibility, dominance, parametric space restrictions, were obtained by George, Liang and Xu (2006), Brown, George and Xu (2008), Fourdrinier et al. (2011), among others, while findings for  $\alpha$ -divergence include the work of Ghosh, Mergel and Datta (2008), Maruyama and Strawderman (2012), Maruyama and Ohnishi (2017), and Marchand and Sadeghkhani (2017). The theme of improvements on plug-in densities by variance expansion arises as well for Gamma models (LMoudden et al., 2017) under Kullback-Leibler divergence loss, as well for spherically symmetric and normal models under integrated  $L^2$  and  $L^1$  losses (Kubokawa, Marchand and Strawderman; 2017, 2015). Finally, several researchers have studied the asymptotic efficiency of predictive densities, with  $\alpha$ -divergence findings for exponential families obtained by Corcuera and Giummolè (1999).

The rest of the paper is organized as follows. At the outset of Section 2, we consider the instructive case  $\hat{\theta}(X) = X$  with properties that resonate throughout the manuscript. In Subsections 2.1 and 2.2, we first consider two specific and instructive cases: **(i)** the affine linear case  $\hat{\theta}_a(X) = aX$ , and **(ii)** the case of a univariate nonnegative normal mean and the maximum likelihood estimator  $\hat{\theta}_{mle}(X) = \max\{0, X\}$ . In both cases, we obtain necessary and sufficient conditions for a variance expansion  $q_{\hat{\theta},c}$  to dominate the plug-in density  $q_{\hat{\theta},1}$  and provide numerical illustrations. We proceed in Subsection 2.3 with a more general dominance finding with implications. The result is applicable to a large class of plugged-in estimators  $\hat{\theta}$  in (3.6), and represents an otherwise unified finding with respect to the dimension  $d$ , the parameter space  $C$ , the loss  $L_\alpha$  for  $\alpha \in (-1, 1)$ , the variances  $\sigma_X^2$  and  $\sigma_Y^2$ . In Subsection 2.4, we expand on with further analysis and observations relative to the allowable degree of expansion maintaining dominance, as well as simultaneous dominance for several choices of loss  $L_\alpha$ . We conclude with an example in Subsection 2.5.

## 3.2 Main results

To begin with, it is instructive to review the case  $\hat{\theta}_1(X) = X$  for the densities in (3.6), and for which the  $\alpha$ -divergence risk is constant as a function of  $\theta \in \mathbb{R}^d$ . and given by

$$R_\alpha(\theta, q_{\hat{\theta}_1, c}) = \frac{4}{1 - \alpha^2} \left( 1 - \left( \frac{4c^{1-\alpha}}{2c^2(1-\alpha) + 2(1+\alpha) + (1-\alpha^2)r} \right)^{d/2} \right), \quad (3.7)$$

with  $r = \sigma_X^2/\sigma_Y^2$  (see Ghosh, Mergel and Datta; 2008; or equation (3.11) below that applies for  $\hat{\theta}_a(X) = aX$  and  $a = 1$ ). With frequentist risk constant as a function of  $\theta$ , there exists an optimal choice  $c_{opt}^2 = 1 + \frac{r(1-\alpha)}{2}$  which minimizes (3.7) in  $c^2$ . Observe that this optimal degree of expansion increases in  $r$ , and decreases in  $\alpha$  ranging from a Kullback-Leibler ( $\alpha \rightarrow -1$ ) expansion of  $1 + r$  to an absence of expansion for reverse Kullback-Leibler ( $\alpha \rightarrow 1$ ). Moreover, the risk in (3.7) is for fixed  $(d, \alpha, r)$  decreasing in  $c$  for  $c \in (1, c_{opt})$ , and increasing in  $c$  for  $c > c_{opt}$  (see Theorem 3.2). This implies that an optimal choice  $c_{opt}^2 = 1 + \frac{r(1-\alpha_0)}{2}$  for loss  $L_{\alpha_0}$  also leads to dominance for Kullback-Leibler loss, as well as losses  $L_\alpha$  with  $\alpha \in (-1, \alpha_0)$ . We thus also have simultaneous dominance with respect to a class of loss functions. Such features recur throughout the paper in the study of predictive densities (3.6). On the other hand, the above value  $c_{opt}$  does not depend on  $d$ , which will not be the case for other choices of the plugged-in estimator  $\hat{\theta}$ . We point out that  $q_{\hat{\theta}_1, c_{opt}}$  is the (generalized) Bayes predictive density with respect to the prior  $\pi(\theta) = 1$ , as well as the minimum risk equivariant predictive density with respect to changes of location (e.g., Ghosh, Mergel and Datta, 2008).

The amount gained by expanding the variance to the optimal level is reflected by the ratio

$$\frac{R_\alpha(\theta, q_{\hat{\theta}_1, 1})}{R_\alpha(\theta, q_{\hat{\theta}_1, c_{opt}})} = \frac{1 - \{1 + (r(1 - \alpha^2)/4)\}^{-d/2}}{1 - \{(1 + (r(1 - \alpha)/2))\}^{-d(1+\alpha)/4}}. \quad (3.8)$$

Numerical and analytical evaluations (see Figure 1) suggest that this ratio decreases in  $\alpha \in (-1, 1)$ , as well as in  $d$ . As a function of  $r$ , the ratio approaches 1 as  $r \rightarrow 0$  and  $r \rightarrow \infty$ , increasing up to a maximal  $r = r_0$ , and decreasing for  $r > r_0$ . As an exemplar, for  $\alpha = 0, d = 2$ , the ratio above reduces to  $\frac{2+r+\sqrt{4+2r}}{4+r}$ , and behaves as above with a maximum value of around 1.2071 attained at  $r_0 = 4(1 + \sqrt{2}) \approx 9.6568$ . Finally, further numerical evaluations suggest that such maximal gains increase as  $\alpha$  decreases and are attenuated with increasing dimension  $d$ .

As mentioned above, it seems plausible that such aspects of this benchmark case may

recur for other choices  $\hat{\theta}$  in predictive densities (3.6).

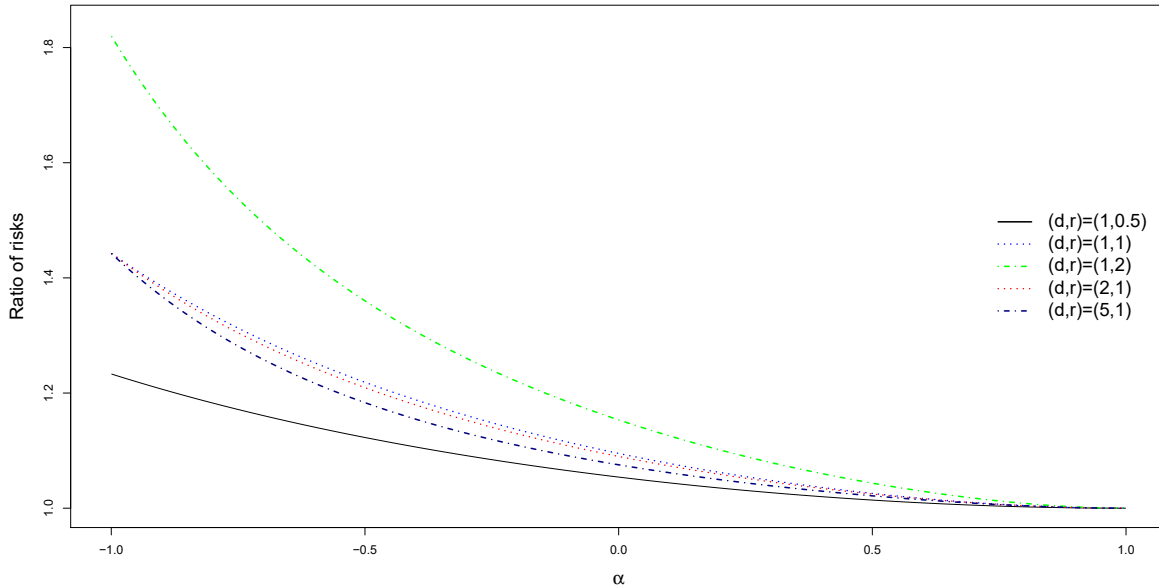


Figure 3.1 – Risk ratios (3.8) as functions of  $\alpha$  for various pairs  $(d, r)$ .

### 3.2.1 Case $\hat{\theta}_a(X) = aX$

Before analyzing the risk performance of predictive densities in (3.6) associated with affine linear estimators, we will first require the following result, which highlights the relationship between  $\alpha$ -divergence loss and reflected normal loss. The result is known (e.g., Ghosh, Mergel and Datta; 2008) and we will expand on its significance below in Remark 3.3. Hereafter, we denote  $\phi$  as the  $N_d(0, I_d)$  p.d.f., and  $\Phi$  as the  $N(0, 1)$  c.d.f.

**Lemma 3.1** *For  $|\alpha| < 1$ , the loss  $L_\alpha$  incurred by the predictive density estimate  $q_{\hat{\theta},c}$  as in (3.6) for estimating the density  $q(\cdot|\theta)$  of  $Y$  is given by :*

$$L_\alpha(\theta, q_{\hat{\theta},c}) = \frac{4}{1-\alpha^2} \left( 1 - \left( \frac{2c^{1-\alpha}}{(1-\alpha)c^2 + 1 + \alpha} \right)^{\frac{d}{2}} \exp \left( -\frac{\|\hat{\theta}(x) - \theta\|^2}{2\gamma_0} \right) \right), \quad (3.9)$$

with  $\gamma_0 = 2 \left( \frac{c^2}{1+\alpha} + \frac{1}{1-\alpha} \right) \sigma_Y^2$ .



**Proof.** The result follows by a development of (3.1) for  $\hat{q}(y; x) = \frac{1}{(c\sigma_Y)^d} \phi\left(\frac{y-\hat{\theta}(x)}{c\sigma_Y}\right)$  and  $q(y|\theta) = \frac{1}{(c\sigma_Y)^d} \phi\left(\frac{y-\theta}{\sigma_Y}\right)$ ,  $y \in \mathbb{R}^d$ .

**Theorem 3.2** Consider  $X \sim N_d(\theta, \sigma_X^2 I_d)$  independent of  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$  and the problem of estimating the density of  $Y$  under  $\alpha$ -divergence loss  $L_\alpha$  as in (3.1) with  $|\alpha| < 1$  and  $\theta \in \mathbb{R}^d$ . Let  $\hat{\theta}_a(X) = aX$  for  $0 < a < 1$ , set  $r = \sigma_X^2/\sigma_Y^2$ , and consider predictive density estimators  $q_{\hat{\theta}_a, c}(\cdot; X) \sim N_d(aX, c^2 \sigma_Y^2 I_d)$  for  $c \geq 1$ . Then  $q_{\hat{\theta}_a, c}$  dominates the plug-in density  $q_{\hat{\theta}_a, 1}$  if and only if  $1 < c^2 \leq k^2(\alpha, a, r)$ , where  $k(\alpha, a, r)$  is the solution in  $c \in (1, \infty)$  of

$$2(1-\alpha)c^2 - (4 + (1-\alpha^2)a^2r)c^{1-\alpha} + (1+\alpha)(2 + (1-\alpha)a^2r) = 0. \quad (3.10)$$

**Proof.** It follows from Lemma 3.1 that

$$R_\alpha(\theta, q_{\hat{\theta}_a, c}) = \frac{4}{1-\alpha^2} \left( 1 - \left( \frac{4\sigma_Y^2 c^{1-\alpha}}{\gamma_0(1-\alpha^2)} \right)^{d/2} \mathbb{E} \left( e^{-\frac{\|aX-\theta\|^2}{2\gamma_0}} \right) \right),$$

with  $\gamma_0 = 2\sigma_Y^2 \frac{c^2(1-\alpha)+1+\alpha}{1-\alpha^2}$ . Since  $\|aX - \theta\|^2 \sim a^2 \sigma_X^2 \chi_d^2\left(\frac{(a-1)^2 \|\theta\|^2}{a^2 \sigma_X^2}\right)$ , one obtains

$$\mathbb{E} \left( e^{-\frac{\|aX-\theta\|^2}{2\gamma_0}} \right) = \left( \frac{\gamma_0}{\gamma_0 + a^2 \sigma_X^2} \right)^{d/2} e^{-\frac{(a-1)^2 \|\theta\|^2}{2(\gamma_0 + a^2 \sigma_X^2)}},$$

whence the expression

$$R_\alpha(\theta, q_{\hat{\theta}_a, c}) = \frac{4}{1-\alpha^2} \left( 1 - \left( \frac{4c^{1-\alpha}}{2c^2(1-\alpha) + 2(1+\alpha) + a^2(1-\alpha^2)r} \right)^{d/2} e^{-\frac{(a-1)^2 \|\theta\|^2}{2(\gamma_0 + a^2 \sigma_X^2)}} \right). \quad (3.11)$$

Now, observe for the difference in risks that, for  $c > 1$ ,

$$\begin{aligned} & \frac{1-\alpha^2}{4} \{R_\alpha(\theta, q_{\hat{\theta}_a, c}) - R_\alpha(\theta, q_{\hat{\theta}_a, 1})\} \\ & \leq \left\{ \left( \frac{4}{4 + a^2(1-\alpha^2)r} \right)^{d/2} - (H(c))^{d/2} \right\} e^{-\frac{(a-1)^2 \|\theta\|^2 / 2\sigma_Y^2}{a^2 r + 4/(1-\alpha^2)}}, \end{aligned} \quad (3.12)$$

with

$$H(c) = \left( \frac{4c^{1-\alpha}}{2c^2(1-\alpha) + 2(1+\alpha) + a^2(1-\alpha^2)r} \right), \quad (3.13)$$

and with equality if and only if  $\theta = 0$ . The result follows by verifying that, for  $c > 1$ ,  $H(c) \geq \frac{4}{4+a^2(1-\alpha^2)r}$  if and only if  $1 < c^2 \leq k^2(\alpha, a, r)$ .

**Remark 3.3** *This paper focuses on the effect of variance expansion, that is the role of  $c^2$  on the frequentist risk performance of  $q_{\hat{\theta},c}$ . Alternatively, it is natural and of interest to study the role of the plugged-in estimator  $\hat{\theta}$ . In view of expression (3.9), it is apparent that the frequentist risk under  $\alpha$ -divergence loss of the predictive density estimator  $q_{\hat{\theta},c}$  relates to the point estimation risk performance of  $\hat{\theta}$  as an estimator of  $\theta$  under reflected normal loss  $L(\theta, \hat{\theta}) = 1 - e^{-\|\hat{\theta}-\theta\|^2/2\gamma_0}$ . Ghosh, Mergel and Datta (2008) capitalized on such a dual relationship to derive predictive densities  $q_{\hat{\theta},c}$  dominating the minimum risk equivariant, and minimax, predictive density  $\hat{q}_{mre}(\cdot; X) \sim N_d(X, (\frac{1-\alpha}{2}\sigma_X^2 + \sigma_Y^2)I_d)$  for  $d \geq 3$ . Further applications were recently obtained by Marchand, Perron and Yadegari (2017), as well as Marchand and Sadeghkhani (2017).*

Here are some further observations and implications of Theorem 3.2.

**Remark 3.4** — *For Hellinger loss (i.e.,  $\alpha = 0$ ), Theorem 3.2's cut-off point simplifies to  $k^2(0, a, r) = (1 + \frac{a^2r}{2})^2$ . In this case, and more generally for other choices of  $\alpha$ , it is easy to show that  $k(\alpha, a, r)$  increases in  $r = \sigma_X^2/\sigma_Y^2$ , converging to 1 as  $r \rightarrow 0$ . Smaller values of  $r$ ; which may also translate to larger samples sizes from  $X$ ; correlate with greater efficiency of  $aX$  for estimating  $\theta$  and less of a need to expand on the plug-in density. Larger values of  $r$  have the opposite effect. One can also infer a robustness result : if the ratio  $r$  of variances is misspecified and that the actual ratio is equal to  $r'$  dominance of  $q_{\theta_{a,c}}$  over  $q_{\theta_{a,1}}$  persists for  $1 < c^2 \leq k^2(\alpha, a, r)$  as long as  $r' > r$ , i.e., one has underestimated the ratio of variances.*

- *An explicit lower bound for Theorem 3.2's cut-off point is given by the inequality  $k^2(\alpha, a, r) \geq 1 + \frac{a^2r(1-\alpha)}{2}$ . This follows from (3.12) and observing that  $H(c)$  in (3.13) increases in  $c$  for  $1 < c^2 \leq 1 + \frac{a^2r(1-\alpha)}{2}$ .*
- *Theorem 3.2 applies for  $a = 1$  with the  $1 < c^2 < k^2(\alpha, 1, r)$  as the necessary and sufficient condition for dominance.*
- *Numerical evidence suggests that, for fixed  $a \in (0, 1)$ ,  $r > 0$ ,  $k(\alpha, a, r)$  decreases in  $\alpha$ , which is quite plausible (and undoubtedly true for the lower bound in the previous paragraph). If true, choices  $c^2 = k^2(\alpha_0, a, r)$  would not only lead to domination for loss  $L_{\alpha_0}$ , but also for all other  $\alpha$ -divergence losses with  $-1 \leq \alpha \leq \alpha_0$ .*

We conclude this section with a numerical illustration.

Figure 2 represents for Hellinger loss (i.e.,  $\alpha = 0$ ),  $d = 3$ ,  $\sigma_X^2 = 1$ ,  $a = 0.75$ , ratios of risks  $\frac{R_0(\theta, q_{\hat{\theta}_{0.75}, c})}{R_0(\theta, q_{\hat{\theta}_{0.75}, 1})}$ , as a function of  $\|\theta\|$ , for  $r = 0.5, 1, 2$  and  $c = k(\alpha, a, r)$  and  $c = (1 + k(\alpha, a, r))/2$ . The graphs illustrate the dominance result given by Theorem 3.2 and

permit us to focus here on the effect of the ratio of variances  $r = \frac{\sigma_X^2}{\sigma_Y^2}$ , as well as the degree of variance expansion  $c^2$ , in relationship to the gains that are attainable by variance expansion as opposed to the plug-in density. Overall, the gains can be significant, as illustrated here for a specific setting of  $\alpha, d, \sigma_X^2$ , and  $a$ . Here are some observations based on Figure 2 and several other numerical evaluations.

- (i) For larger  $r$ , maximal gains are more important, as well as gains for small or moderate  $\|\theta\|$ . This is a recurrent feature below in other situations, for other plug-in choices  $\hat{\theta}(X)$ . Given the decreasing relative reliability of the information provided by  $X$  for making inferences about  $\theta$ , such behaviour is somewhat anticipated and relates in this example to the allowable degree of expansion to maintain dominance which increases in  $r$  (Remark 3.4). It is somewhat delicate here as the choice itself of  $a$  would typically depend on  $\sigma_X^2$ , as is the case of the posterior expectation for a  $N_d(0, I_d)$  prior corresponding to the multiple  $\frac{d}{d+\sigma_X^2}$ ;
- (ii) Expansions to  $c = k(\alpha, a, r)$  lead to more important maximal gains, while the compromise choice  $c = (1 + k(\alpha, a, r))/2$  flattens out the gains with better performance for small  $\|\theta\|$ . Observe as well the ratios equal to 1 for  $\theta = 0$  as established within the proof of Theorem 3.2, and that more important expansions will not lead to dominance in view of the necessity and sufficiency of Theorem 3.2.

### 3.2.2 Case of a non-negative mean with $\hat{\theta}_+(X) = \max\{X, 0\}$

As in the previous section, we proceed with an instructive example bringing into play a nonnegativity constraint and the choice of the maximum likelihood estimator (mle)  $\hat{\theta}_+(X) = \max\{X, 0\}$ . The relative tractability of the  $\alpha$ -divergence risk, which arises with a convenient expression for the expected reflected normal loss, leads to a necessary and sufficient condition for an variance expansion  $q_{\hat{\theta}_+, c}$  to dominate the predictive mle.

**Theorem 3.5** *Consider  $X \sim N(\theta, \sigma_X^2)$  independent of  $Y \sim N(\theta, \sigma_Y^2)$  and the problem of estimating the density of  $Y$  under  $\alpha$ -divergence loss  $L_\alpha$  as in (3.1) with  $|\alpha| < 1$  and  $\theta \geq 0$ . Let  $\hat{\theta}_+(X) = \max\{X, 0\}$ , and consider predictive density estimators  $q_{\hat{\theta}_+, c}(\cdot; X) \sim N(\hat{\theta}_+(X), c^2\sigma_Y^2)$  for  $c \geq 1$ . Set  $r = \sigma_X^2/\sigma_Y^2$ ,  $\gamma_0(c) = 2\sigma_Y^2 \left( \frac{c^2(1-\alpha)+1+\alpha}{1-\alpha^2} \right)$ ,  $\gamma_1(c) = \sqrt{\frac{\gamma_0(c)}{\gamma_0(c)+\sigma_X^2}}$ ,  $A_1(c) = \sqrt{\frac{2c^{1-\alpha}}{c^2(1-\alpha)+1+\alpha}}$ , and  $A_2(c) = -1 + A_1(c)(1 + \gamma_1(c))$ . Then  $q_{\hat{\theta}_+, c}$  dominates the plug-in m.l.e. density  $q_{\hat{\theta}_+, 1}$  if and only if  $1 < c \leq \kappa(\alpha, r)$ , where  $\kappa(\alpha, r)$  is*

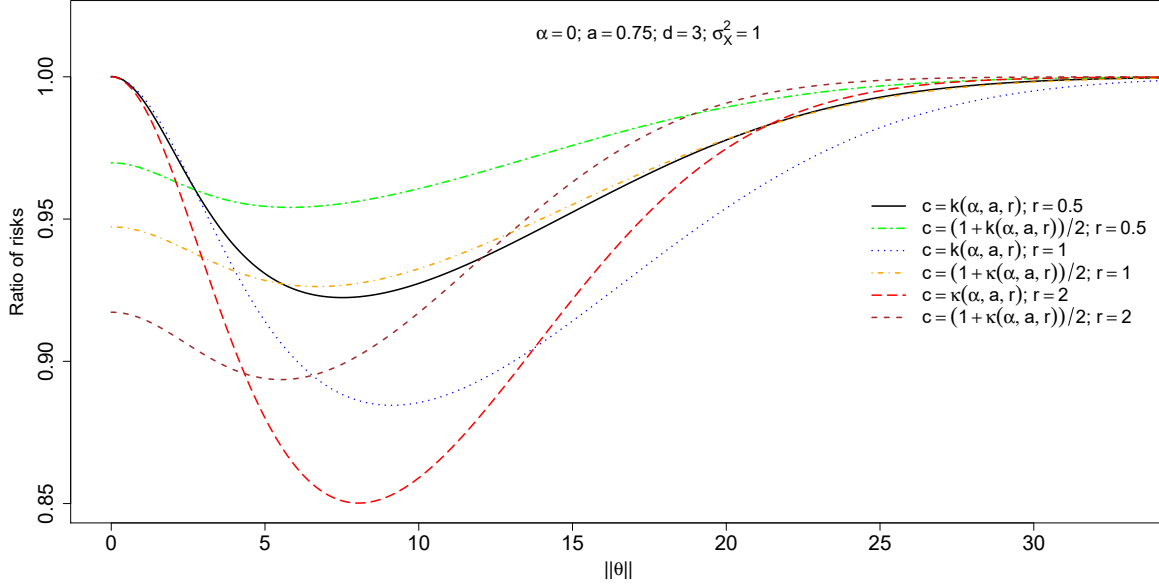


Figure 3.2 – Hellinger risk ratios  $\frac{R_0(\theta, q_{\hat{\theta}_a, c})}{R_0(\theta, q_{\hat{\theta}_a, 1})}$  for various  $c$  and  $r$

the solution in  $c \in (1, \infty)$  of

$$A_2(c) = \left(1 + \frac{r(1 - \alpha^2)}{4}\right)^{-1/2}. \quad (3.14)$$

The dominance is strict for  $\theta > 0$  or  $1 < c < \kappa(\alpha, r)$ , with equality of risks if and only if  $\theta = 0$  and  $c = \kappa(\alpha, r)$ .

**Proof. (I)** It follows from Lemma 3.1 that

$$R_\alpha(\theta, q_{\hat{\theta}_+, c}) = \frac{4}{1 - \alpha^2} [1 - A_1(c) G(\theta, c)], \quad (3.15)$$

with  $G(\theta, c) = \mathbb{E} \left( e^{-\frac{|\hat{\theta}_+(X) - \theta|^2}{2\gamma_0(c)}} \right)$  and the given notation. Calculations yield the expression  $G(\theta, c) = G_1(\theta, c) + G_2(\theta, c)$ , with  $G_1(\theta, c) = e^{-\frac{\theta^2}{2\gamma_0(c)}} \Phi\left(-\frac{\theta}{\sigma_X}\right)$  and  $G_2(\theta, c) = \gamma_1(c) \Phi\left(\frac{\theta}{\gamma_1(c)\sigma_X}\right)$ , as well as

$$\frac{\partial}{\partial \theta} G(\theta, c) = -\frac{\theta}{\gamma_0(c)} G_1(\theta, c).$$

(II) For the difference in risks, we thus obtain from the above

$$\begin{aligned}\Delta(\theta, c) &= \frac{1 - \alpha^2}{4} \left\{ R_\alpha(\theta, q_{\hat{\theta}_{+,c}}) - R_\alpha(\theta, q_{\hat{\theta}_{+,1}}) \right\} \\ &= G(\theta, 1) - A_1(c) G(\theta, c),\end{aligned}\tag{3.16}$$

and

$$\frac{\partial}{\partial \theta} \Delta(\theta, c) = \frac{\theta e^{-\frac{\theta^2}{2\gamma_0(1)}} \Phi\left(-\frac{\theta}{\sigma_X}\right)}{\gamma_0(1)} \left\{ -1 + A_1(c) \frac{\gamma_0(1)}{\gamma_0(c)} e^{\frac{\theta^2}{2}\left(\frac{1}{\gamma_0(1)} - \frac{1}{\gamma_0(c)}\right)} \right\}.\tag{3.17}$$

Now, since  $\gamma_0(c)$  increases in  $c$ ,  $c \geq 1$ , and  $A_1(c) \leq 1$  for all  $c \geq 1$  with equality iff  $c = 1$ , we infer, for fixed  $c \in (1, \infty)$ , that  $\frac{\partial}{\partial \theta} \Delta(\theta, c)$  changes signs from  $-$  to  $+$  as  $\theta$  varies on  $[0, \infty)$ . Therefore, for a given  $c > 1$ ,  $q_{\hat{\theta}_{+,c}}$  will dominate  $q_{\hat{\theta}_{+,1}}$  if and only if  $\Delta(0, c) \leq 0$  and  $\lim_{\theta \rightarrow \infty} \Delta(\theta, c) \leq 0$ . Furthermore, from (3.16) and the earlier expression for  $G(\theta, c)$ , we obtain

$$\Delta(0, c) - \lim_{\theta \rightarrow \infty} \Delta(\theta, c) = \frac{1 - \gamma_1(1)}{2} - A_1(c) \left( \frac{1 - \gamma_1(c)}{2} \right) \geq 0,$$

since  $A_1(c) \leq 1$  and  $1 > \gamma_1(c) \geq \gamma_1(1)$  for all  $c \geq 1$ .

(III) We thus have that  $q_{\hat{\theta}_{+,c}}$  will dominate  $q_{\hat{\theta}_{+,1}}$  if and only if  $\Delta(0, c) \leq 0$ , and there remains to show that this inequality is equivalent to the stated condition, and with equality if and only if  $c = \kappa(\alpha, r)$ . To justify this last step, it suffices to show that, as a function of  $c \in (1, \infty)$ ,  $\Delta(0, c)$  decreases and then increases since  $\lim_{c \rightarrow \infty} \Delta(0, c) = G(0, 1) > 0$  from (3.16).

In turn, by virtue of (3.16), it will suffice to show that

$$\text{sgn} \left( \frac{\partial}{\partial c} A_1(c) G(0, c) \right) \text{ varies from } + \text{ to } - \text{ for } c \in (1, \infty).\tag{3.18}$$

Setting  $H(c) = \sqrt{c^2(1 - \alpha) + (1 + \alpha)}$ , we may express

$$A_1(c) G(0, c) = \frac{1}{\sqrt{2}} c^{\frac{1-\alpha}{2}} \left( \frac{1}{H(c)} + \frac{1}{\sqrt{H^2(c) + \frac{r(1-\alpha^2)}{2}}} \right).$$

With  $H'(c) = \frac{(1-\alpha)c}{H(c)}$ , some calculations and manipulations permit us to write

$$\frac{\partial}{\partial c} A_1(c) G(0, c) = -\frac{(1 - \alpha^2) c^{-\frac{1+\alpha}{2}}}{(\sqrt{2} H(c))^3} (c^2 - 1 + W(c)),$$

with

$$W(c) = \left( \frac{c^2 - 1 - r(1 - \alpha)/2}{(H^2(c) + r(1 - \alpha^2)/2)^{3/2}} \right) H^3(c).$$

From this, we see that  $\lim_{(c \rightarrow 1)^+} c^2 - 1 + W(c) < 0$ , implying that  $\frac{\partial}{\partial c} A_1(c)G(0, c)$  is positive for small enough  $c$ . On the other hand, since  $c^2 - 1 + W(c) \geq 0$  for  $c^2 \geq 1 + r(1 - \alpha)/2$ , it will suffice to complete the proof that  $W(c)$  be increasing in  $c$  for  $c^2 < 1 + r(1 - \alpha)/2$ . Finally, a calculation yields the expression

$$\begin{aligned} \frac{\partial}{\partial c} \log W(c) &= \frac{\partial}{\partial c} \left\{ 3 \log B(c) + \log \left( \frac{r(1 - \alpha)}{2} + 1 - c^2 \right) - \frac{3}{2} \log \left( B^2(c) + \frac{r(1 - \alpha^2)}{2} \right) \right\} \\ &= 3c(1 - \alpha) + \frac{2c}{\frac{r(1 - \alpha)}{2} + 1 - c^2} - \frac{3c(1 - \alpha)/2}{H^2(c) + \frac{r(1 - \alpha^2)}{2}} \\ &\geq 0, \end{aligned}$$

for  $c^2 < 1 + r(1 - \alpha)/2$ , which establishes the result.

**Remark 3.6** *As a function of  $\theta$ ,  $\theta \geq 0$ , the frequentist risks  $R_\alpha(\theta, q_{\hat{\theta}_+, c})$  are increasing with a limiting value at  $\theta \rightarrow \infty$  equal to  $\frac{4}{1 - \alpha^2} (1 - A_1(c)\gamma_1(c))$ , and with a value of  $\frac{4}{1 - \alpha^2} \left( 1 - \frac{A_1(c)}{2} (1 + \gamma_1(c)) \right)$  at  $\theta = 0$ . These properties are obtained from (3.15). As in Remark 3.4, the cut-off point  $\kappa(\alpha, r)$  can be shown to be increasing as a function of  $r$ , and is decreasing as a function of  $\alpha$  according to numerical evaluations. The former is obtained in continuity with the arguments of the proof of Theorem 3.5 and with the r.h.s. of (3.14) decreasing in  $r$ .*

**Remark 3.7** *A surprising robustness result is also available from the analysis above in the proof of Theorem 3.5. Indeed, it is also the case that the difference in risks  $\Delta(\theta, c)$  is negative for all  $\theta < 0$  and  $1 < c \leq \kappa(\alpha, r)$ . In other words, the expansions  $q_{\hat{\theta}_+, c}$  that dominate  $q_{\hat{\theta}_+, 1}$  with lower risk on  $[0, \infty)$  continue providing lower  $\alpha$ -divergence frequentist risk for negative values of  $\theta$ . This is relevant to cases where it is believed that the constraint  $\theta \geq 0$  holds true, but, unbeknownst to the investigator, the actual value of  $\theta$  is negative. In such cases, the choice of the plug-in estimate  $\max\{x, 0\}$  is, of course, not desirable, but the expansion offers better protection against the misspecification. Finally, to see why the difference in risks remains negative for negative  $\theta$ ; for all  $r > 0$  and choice of loss  $L_\alpha$  for  $|\alpha| < 1$ ; it suffices to observe from (3.17) that the sign of  $\frac{\partial}{\partial \theta} \Delta(\theta, c)$  varies from  $-$  to  $+$  as  $\theta$  increases from  $-\infty$  to 0 and that  $\lim_{\theta \rightarrow -\infty} \Delta(\theta, c) = 0$  for all  $c \geq 1$  as seen directly by working with (3.15). Accordingly, analogous inferences with respect to a persistent dominance result when underestimating  $r$ , as well as a simultaneous dominance result for various choices of  $L_\alpha$ , apply.*

We conclude this section with a numerical illustration. Figure 3 exhibits the relative frequentist risk performance of the maximum likelihood plug-in density  $q_{\hat{\theta}_{+,1}}$  and the variance expansion  $q_{\hat{\theta}_{+,c}}$  with Theorem 3.5's cut-off point  $c = \kappa(\alpha, r)$ . More specifically, risk ratios for  $\sigma_X^2 = 1$  are drawn for various combinations of  $r$  and  $\alpha$ . Theoretically, the ratios are bounded by 1, and we point out the equality of risks at  $\theta = 0$  in accordance with Theorem 3.5. As shown by the graphs, the gains can be significant, tend to be more important for smaller values of  $\alpha$ , and large values of  $r$ . Other choices of variance expansion, such as  $c^2 = (1 + \kappa(\alpha, r))/2$  have the same effect as in Figure 3.2.1.

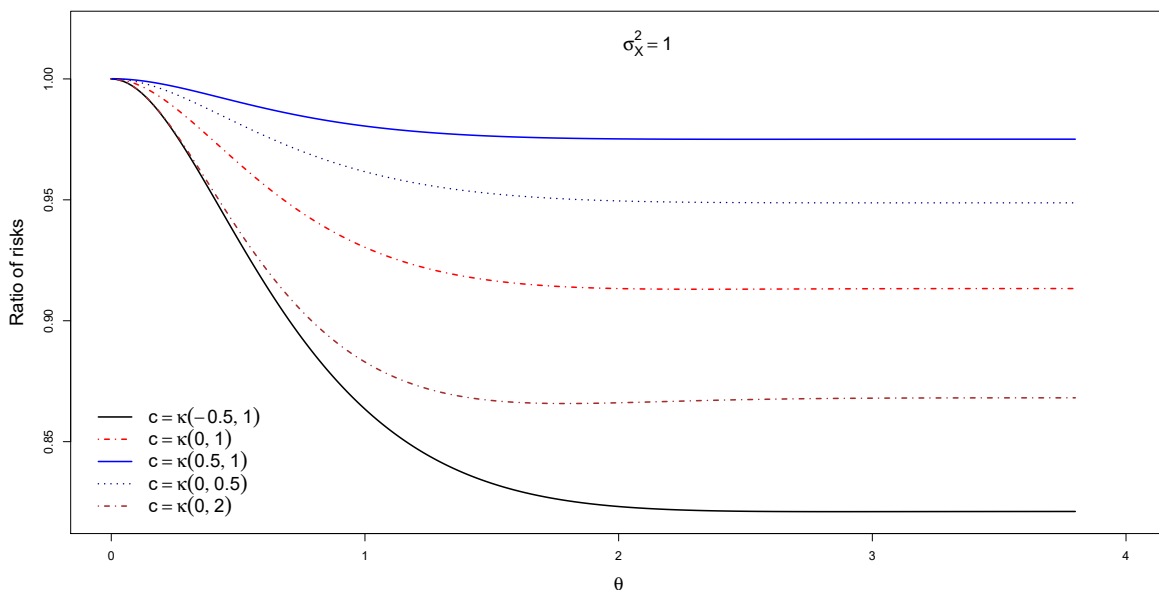


Figure 3.3 – Risk ratios  $\frac{R_{\alpha}(\theta, q_{\hat{\theta}_{+,c}})}{R_{\alpha}(\theta, q_{\hat{\theta}_{+,1}})}$  for  $c = \kappa(\alpha, r)$ , and various  $c$  and  $r$

### 3.2.3 General $\hat{\theta}(X)$

**Lemma 3.8** *Let  $T$  be a nonnegative and continuous random variable such that  $\mathbb{E}(T^2) < \infty$ . Let  $s$  be a positive constant. Then, we have*

$$\mathbb{E}(Te^{-sT}) \geq \mathbb{E}(T) e^{-s\mathbb{E}(T^2)/\mathbb{E}(T)}. \quad (3.19)$$

*Moreover, if the distribution of  $T$  depends on a parameter  $\theta \in C$ , and if there exist positive constants  $b_0, b_1, b_2$  such that  $b_0 \leq \mathbb{E}_{\theta}(T) \leq b_1$  and  $\mathbb{E}_{\theta}(T^2) \leq b_2$  for all  $\theta \in C$ , then*

we have

$$\mathbb{E}_\theta(Te^{-sT}) \geq b_0 e^{-s \frac{b_2}{b_0}}, \quad (3.20)$$

for all  $\theta \in C$ .

**Proof.** Let  $f_T$  be the density of  $T$  and let  $W$  be a random variable with density  $wf_T(w)/\mathbb{E}(T)$ ,  $w > 0$ . We then have

$$\mathbb{E}(Te^{-sT}) = \mathbb{E}(T) \mathbb{E}(e^{-sW}) \geq \mathbb{E}(T) (e^{-s\mathbb{E}W}) = \mathbb{E}(T) e^{-s \frac{\mathbb{E}(T^2)}{\mathbb{E}(T)}},$$

by using Jensen's inequality. This establishes (3.19), and (3.20) is a direct consequence of (3.19).

We now are ready for our main finding.

**Theorem 3.9** Consider  $X \sim N_d(\theta, \sigma_X^2 I_d)$  independent of  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$  and the problem of estimating the density of  $Y$  under  $\alpha$ -divergence loss  $L_\alpha$  as in (3.1) with  $|\alpha| < 1$  with  $\theta \in C$ . Let  $q_{\hat{\theta},1}$  be a plug-in predictive density based on a non-degenerate  $\hat{\theta}(X)$ , let  $Z = \frac{\|\hat{\theta}(X) - \theta\|^2}{\sigma_Y^2}$  define  $\tau = (1 - \alpha)\epsilon - 2\alpha d$ , with

$$\epsilon \equiv \epsilon(\alpha) = \inf_{\theta \in C} \mathbb{E}_\theta(Z e^{-(1-\alpha^2)Z/8}). \quad (3.21)$$

(a) Then, assuming  $\epsilon > 0$ , the expansion  $q_{\hat{\theta},c} \sim N_d(\hat{\theta}(X), c^2 \sigma_Y^2 I_d)$  dominates  $q_{\hat{\theta},1}$  whenever

$$1 < c^2 \leq k(d, \alpha, \sigma_X^2, \sigma_Y^2) = \frac{\tau + \sqrt{\tau^2 + 4d^2(1 - \alpha^2)}}{2d(1 - \alpha)}.$$

(b) Furthermore, if  $b_0, b_1, b_2$  are positive numbers such that  $b_0 \sigma_Y^2 \leq \mathbb{E}_\theta \|\hat{\theta}(X) - \theta\|^2 \leq b_1 \sigma_Y^2$  and  $\mathbb{E}_\theta \|\hat{\theta}(X) - \theta\|^4 \leq b_2 \sigma_Y^4$  for all  $\theta \in C$ , then the assumption  $\epsilon > 0$  is satisfied and

$$k(d, \alpha, \sigma_X^2, \sigma_Y^2) \geq \frac{\underline{\tau} + \sqrt{\underline{\tau}^2 + 4d^2(1 - \alpha^2)}}{2d(1 - \alpha)}, \quad (3.22)$$

with  $\underline{\tau} = (1 - \alpha) b_0 e^{-\frac{1-\alpha^2}{8} \frac{b_2}{b_0}} - 2\alpha d$ .

**Proof.** For part (a), setting  $B(c) = (1 - \alpha)c^2 + (1 + \alpha)$ ,  $H_\theta(c) = \mathbb{E}_\theta(e^{-\frac{(1-\alpha^2)}{4B(c)}Z})$ , and  $h_\theta(c) = \frac{c^{1-\alpha}}{B(c)} H_\theta^{\frac{2}{d}}(c)$ , we obtain from Lemma 3.1 the risk expression :

$$R_\alpha(\theta, q_{\hat{\theta},c}) = \frac{4}{1 - \alpha^2} \left(1 - 2^{\frac{d}{2}} h_\theta^{\frac{d}{2}}(c)\right). \quad (3.23)$$



To establish the result, it will suffice to show that, for all  $\theta \in C$ ,  $\frac{d}{dc}h_\theta(c) > 0$  for  $c \in (1, k(d, \alpha, \sigma_X^2, \sigma_Y^2)]$ ; i.e., the risk  $R_\alpha(\theta, q_{\hat{\theta}, c})$  decreases, for all  $\theta \in C$ , as a function of  $c$ , for  $1 < c^2 < k(d, \alpha, \sigma_X^2, \sigma_Y^2)$ . We have

$$\frac{d}{dc}h_\theta(c) = \frac{H_\theta^{\frac{2}{d}-1}(1-\alpha^2)c^{2-\alpha}}{B^3(c)} \left[ \frac{(1-c^2)B(c)H_\theta(c)}{c^2} + \frac{(1-\alpha)}{d} \mathbb{E}_\theta(Ze^{-\frac{(1-\alpha^2)}{4B(c)}Z}) \right]. \quad (3.24)$$

Focussing on the sign of the above expression, we have for  $c > 1$  and  $T(c) = \frac{(c^2-1)B(c)}{c^2}$

$$\frac{(1-c^2)B(c)H_\theta(c)}{c^2} + \frac{(1-\alpha)}{d} \mathbb{E}_\theta(Ze^{-\frac{(1-\alpha^2)}{4B(c)}Z}) > \frac{(1-\alpha)\epsilon}{d} - T(c),$$

since  $B(c) > 2$  and  $H_\theta(c) < 1$ . Finally, since  $\epsilon > 0$  by assumption and since  $T(c)$  increases in  $c$ , expression (3.24) is indeed positive for  $\theta \in C$  whenever  $1 < c \leq T^{-1}\left(\frac{(1-\alpha)\epsilon}{d}\right) = (k(d, \alpha, \sigma_X^2, \sigma_Y^2))^{1/2}$ .

For part **(b)**, given the boundedness assumptions on the first and second moments of  $T = Z$ , it follows from Lemma 3.8 that  $\epsilon \geq \underline{\epsilon} = b_0 e^{-\frac{1-\alpha^2}{8} \frac{b_2}{b_1}} > 0$ . The given lower bound on the cut-off point  $k(d, \alpha, \sigma_X^2, \sigma_Y^2)$  follows with the lower bound  $\underline{\tau}$  for  $\tau$  and since  $k(d, \alpha, \sigma_X^2, \sigma_Y^2)$  is increases as  $\tau$  increases.

**Remark 3.10** — *Theorem 3.9 holds as stated for spherically symmetric model  $X \sim f_0(\|x - \theta\|^2)$  and  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$  with known  $f_0$ . The given proof applies throughout with the observation that Lemma 3.1 is an attribute of the normal density assumption on  $Y$  only.*

— *Theorem 3.9 represents a unified finding with respect to : (i) the loss  $L_\alpha$  for  $|\alpha| < 1$ , (ii) the choice of the plugged-in estimator  $\hat{\theta}(X)$ , (iii) the dimension  $d$ , and (iv) the parameter space  $C$ . Moreover, the proof is unified. As mentioned in the Introduction, the result adds to Fourdrinier et al. (2011)'s finding for Kullback-Leibler divergence loss. Interestingly, taking  $\alpha = -1$  in part **(a)** of Theorem 3.9 leads to the cut-off point*

$$k(d, -1, \sigma_X^2, \sigma_Y^2) = 1 + \frac{\inf_{\theta \in C} \mathbb{E}(\|\hat{\theta}(X) - \theta\|^2)}{d\sigma_Y^2},$$

*which matches a sufficient condition given by Fourdrinier et al. (2011) and confirms further unification. We expand further in Subsection 3.2.4 on the behaviour of the cut-off points  $k(d, \alpha, \sigma_X^2, \sigma_Y^2)$ .*

- On the other hand, Theorem 3.9 does not provide a necessary and sufficient condition, as is the case for Theorems 3.2 and 3.5, as well as the Kullback-Leibler finding of Fourdrinier et al. (2011). An illustration below in Subsection 2.5 will further address this issue.

**Remark 3.11** A large number of estimators  $\hat{\theta}$  satisfy either the condition  $\epsilon > 0$  of part (a) of Theorem 3.9, or part (b)'s boundedness conditions for  $\mathbb{E}_\theta \|\hat{\theta}(X) - \theta\|^2$  and  $\mathbb{E}_\theta \|\hat{\theta}(X) - \theta\|^4$ . Since  $\mathbb{E}_\theta(Z e^{-(1-\alpha^2)Z/8}) > 0$  for all  $\theta \in C$ , the compactness of  $C$  will suffice for the condition  $\epsilon > 0$  to be satisfied. For restricted but unbounded parameter spaces, and specifically for polyhedral cones  $C$ , which include orthant restrictions on some or all of the  $\theta_i$ 's, order constraints of the form  $\theta_1 \leq \dots \leq \theta_d$ , tree-order restrictions with  $\theta_i \leq \theta_1$  for  $i = 2, \dots, d$ , and umbrella order restrictions of the form  $\theta_1 \leq \dots \leq \theta_m \geq \theta_{m+1} \geq \dots \geq \theta_q$ , and others, it follows from Marchand and Strawderman (2012) that  $X$  is minimax under loss  $\|\hat{\theta} - \theta\|^4$  with finite and constant minimax risk given by  $\mathbb{E}_\theta \|X - \theta\|^4 = \mathbb{E}_\theta (\|X - \theta\|^2)^2 = (d^2 + 2d)\sigma_X^4$ . Consequently, estimators  $\hat{\theta}(X)$  that dominate  $X$ , such as projections onto  $C$ , will satisfy the conditions of Theorem 3.9. We refer to Marchand and Strawderman (2012) for details and a list of further references.

Otherwise, we point out the following :

- (i) The existence of a value  $b_0 > 0$  arising with the lower bound condition on  $\mathbb{E}_\theta \|\hat{\theta}(X) - \theta\|^2$  is guaranteed with the condition that  $\hat{\theta}(X)$  be non-degenerate.
- (ii) The existence of  $b_1 \in (0, \infty)$  related to the upper bound condition on  $\mathbb{E}_\theta \|\hat{\theta}(X) - \theta\|^2$  will be satisfied, for instance, by estimators  $\hat{\theta}(X)$  that are minimax. Many such choices are available in  $d \geq 3$  dimensions or more. On the other hand, estimators with unbounded squared error loss, such as affine linear estimators  $\hat{\theta}_a(X) = aX$  studied in Section 2.1 will not satisfy the conditions. Moreover, the corresponding value of  $\epsilon$  can be shown to be equal to 0 for all  $\sigma_X^2, \sigma_Y^2, \hat{\theta}_a$  with  $0 < a < 1$ , making Theorem 3.9 inapplicable for such cases. Of course, the analysis provided by Theorem 3.2 is stronger anyway.
- (iii) For the existence of  $b_2 \in (0, \infty)$  such that  $\mathbb{E}_\theta \|\hat{\theta}(X) - \theta\|^4 \leq b_2 \sigma_Y^4$ , it will suffice that  $\hat{\theta}(X)$  dominate  $X$  as a point estimator of  $\theta$  under loss  $\|\hat{\theta} - \theta\|^4$ , as  $\mathbb{E}_\theta \|X - \theta\|^4 = (d^2 + 2d)\sigma_X^4$ , and one can thus choose  $b_2 = (d^2 + 2d)r^2$ . For  $d \geq 3$ , such estimators were obtained by Berger (1978) and include (for  $\sigma_X^2 = 1$ ) James-Stein type estimators of the form  $\hat{\theta}(X) = (1 - \frac{f_1}{f_2 + \|X\|^2})X$  with  $0 < f_1 \leq d - 2$  and  $f_2 \geq 2 + \frac{d+1}{d+2}f_1$ .

Moreover, one can directly verify that Baranchik-type estimators (Baranchik, 1970) of the form  $\hat{\theta}_{s(\cdot)}(X) = \left(1 - \frac{s(\|X\|^2)}{\|X\|^2}\right) X$ , with  $s(\cdot) \geq 0$  have bounded  $\mathbb{E}_\theta (\|\hat{\theta}_{s(\cdot)}(X) -$

$\theta\|^4), \theta \in \mathbb{R}^d, d \geq 3$ , as long as both  $s(t)$  and  $s(t)/t$  are bounded for  $t > 0$ . Such estimators include the positive-part James-Stein estimator  $\hat{\theta}_{JS+}$  obtained with the choice  $s(t) = \sigma_X^2 \max\{t, d - 2\}$ .

- (iv) In related work for the model  $X \sim N_d(\theta, \sigma_X^4 I_d)$ , Fourdrinier, Ouassou and Strawderman (2008) provide various point estimators  $\hat{\theta}$  that dominate  $X$  under quartic loss  $\sum_i (\hat{\theta}_i - \theta_i)^4$ . These include James-Stein type estimators for  $d \geq 5$  and Baranchik-type estimators (Baranchik, 1970)  $\hat{\theta}_{s(\cdot)}(X)$  for  $d \geq 3$  and certain conditions on  $s(\cdot)$ . Specifically, their estimators  $\hat{\theta}$  are such that :

$$\mathbb{E} \left( \sum_i (\hat{\theta}_i(X) - \theta_i)^4 \right) \leq \mathbb{E} \left( \sum_i (X_i - \theta_i)^4 \right) = 3d\sigma_X^4.$$

Lemma 3.12 which follows, permits us to use such boundedness, coupled boundedness of quadratic loss, to guarantee that Theorem 3.9 can be applied.

**Lemma 3.12** If  $\hat{\theta}(X) \in \mathbb{R}^d$  is an estimator of  $\theta \in \mathbb{R}^d$  such that

$$\mathbb{E} \left( \sum_i (\hat{\theta}_i(X) - \theta_i)^4 \right) \leq M_1, \text{ and } \mathbb{E} \left( \|\hat{\theta}(X) - \theta\|^2 \right) \leq M_2, \text{ for all } \theta,$$

then we have  $\mathbb{E} \left( \|\hat{\theta}(X) - \theta\|^4 \right) \leq dM_1$  for all  $\theta$ .

**Proof.** See Appendix.

It thus follows that estimators  $\hat{\theta}(X)$  that dominate  $X$  under quartic loss, that are also minimax under quadratic loss, are such that  $\mathbb{E} \left( \|\hat{\theta}(X) - \theta\|^4 \right) \leq 4d^2\sigma_X^4$  for all  $\theta \in \mathbb{R}^d$  by making use of Lemma 3.12 with  $M_1 = 3d\sigma_X^4$  and  $M_2 = d\sigma_X^4$ . For such estimators  $\hat{\theta}$ , Theorem 3.9's  $\epsilon$  is greater than 0 and thus applies.

We conclude this subsection by pointing out that the dominating predictive density improvements that arise as a consequence of the above theorems, which are normal densities with an expanded variance, can be mixed to generate many other scale mixture of normals predictive densities which dominate the targeted plug-in density. This is a consequence of Jensen's inequality, as laid out by the following.<sup>1</sup>

---

1. A more general result appears in Yadegari, I. (2017). Prédiction, inférence sélective et quelques problèmes connexes. Ph.D. thesis. Université de Sherbrooke (<http://savoirs.usherbrooke.ca/handle/11143/10167>).

**Lemma 3.13** *Under the assumptions of Theorem 3.9, suppose that the predictive density  $q_{\hat{\theta},c} \sim N_d(\theta(\hat{X}), c^2 \sigma_Y^2 I_d)$  dominates  $q_{\hat{\theta},1}$  for  $c \in (1, k^*)$  and under loss  $L_\alpha$ . Let  $H$  be a cdf such that  $H(1) = 0$  and  $H(k^*) = 1$ . Then, the mixture*

$$\hat{q}_H(y; X) = \int_1^{k^*} q_{\hat{\theta},z}(y; X) dH(z)$$

*also dominates  $q_{\hat{\theta},1}$  under loss  $L_\alpha$ .*

**Proof.** By Jensen's inequality, since  $h_\alpha$  in (3.2) is convex, we have with a change in order of integration

$$\begin{aligned} R_\alpha(\theta, \hat{q}_H) &= \mathbb{E} \left( \int_{\mathbb{R}^d} h_\alpha \left( \int_1^{k^*} \frac{\hat{q}_z(y; X)}{q(y|\theta)} dH(z) \right) q(y|\theta) dy \right) \\ &\leq \mathbb{E} \left( \int_1^{k^*} \int_{\mathbb{R}^d} h_\alpha \left( \frac{\hat{q}_z(y; X)}{q(y|\theta)} \right) q(y|\theta) dH(z) dy \right). \end{aligned}$$

Now, use the assumed dominance results to infer that

$$R_\alpha(\theta, \hat{q}_H) \leq \mathbb{E} \left( \int_1^{k^*} \int_{\mathbb{R}^d} h_\alpha \left( \frac{\hat{q}(y; X)}{q(y|\theta)} \right) q(y|\theta) dH(z) dy \right) = R_\alpha(\theta, \hat{q})$$

with strict inequality for at least one  $\theta$ , thus establishing the result.

### 3.2.4 Behaviour of the cut-off point $k(d, \alpha, \sigma_X^2, \sigma_Y^2)$ and simultaneous dominance

We further expand here on the behaviour of Theorem 3.9's cut-off point and implications for simultaneous dominance with respect to several losses  $L_\alpha$ . In accordance with earlier examples, it seems plausible that  $k(d, \alpha, \sigma_X^2, \sigma_Y^2)$  decreases in  $\alpha$ ,  $\alpha \in [-1, 1)$ . As an illustration, Figure 4 represents the cut-off points  $k(d, \alpha, \sigma_X^2, \sigma_Y^2)$  for  $d = 3$ ,  $\sigma_X^2 = 1$ ,  $\sigma_Y^2 = 1, 2, 4$ ,  $\alpha \in (-1, 1)$ , and the James-Stein estimator  $\hat{\theta}_{JS}(X) = (1 - \frac{d-2}{X'})X$ .

It is of particular interest to focus on the benchmark Kullback-Leibler cut-off point for dominance. Fourdrinier et al. (2011) show, for non-degenerate  $\hat{\theta}$  other than  $\hat{\theta}(X) = X$ , that  $q_{\hat{\theta},c}$  dominates the plug-in density  $q_{\hat{\theta},1}$  for  $1 < c \leq 1 + \underline{R}$  (as in Remark 3.10), and if and only if  $1 < c \leq c_0(1 + \underline{R})$ , with

$$\underline{R} = \frac{\inf_{\theta \in C} \mathbb{E}(\|\hat{\theta}(X) - \theta\|^2)}{d\sigma_Y^2},$$

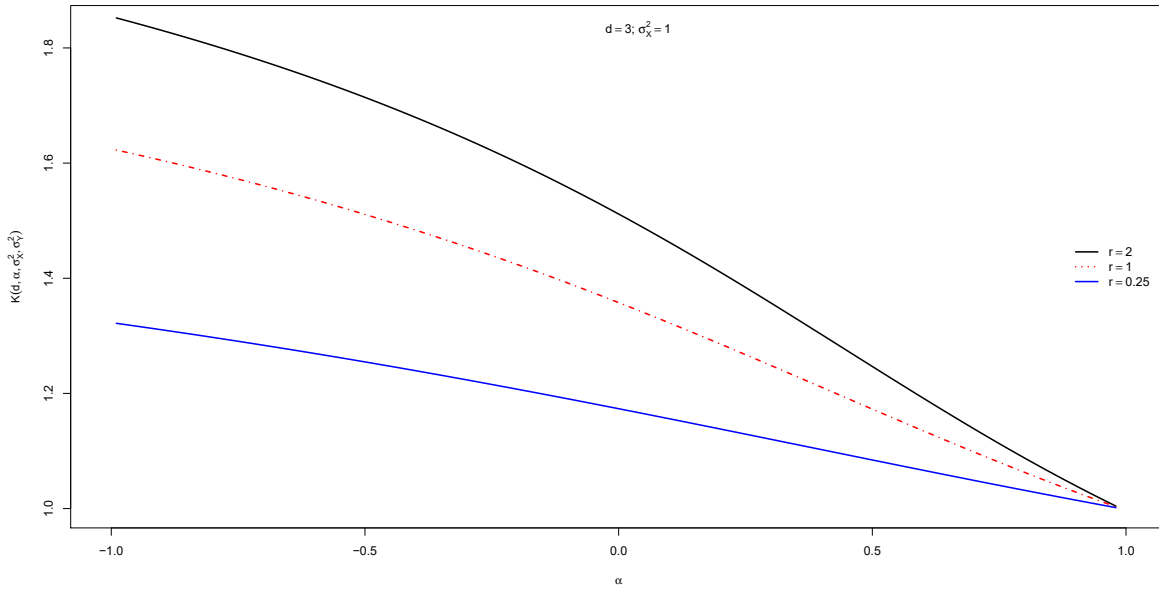


Figure 3.4 – Cut-off points  $k(3, \alpha, 1, \sigma_Y^2 = 1/r)$  as functions of  $\alpha$ , for the James-Stein estimator

and  $c_0(t)$  the solution in  $c \in (t, \infty)$  of the equation  $(1 - \frac{1}{c})t - \log c = 0$ . We pursue with an ordering between Hellinger and Kullback-Leibler cut-off points, as well as a monotonicity property, with implications for simultaneous dominance with respect to  $L_\alpha$ 's stated in the Corollary that follows. The second part implies the first, but the alternative route for the first proof merits exposition.

**Theorem 3.14** *Consider the context of Theorem 3.9 with a given  $\hat{\theta}$  such that  $\epsilon > 0$ , fixed  $d, \sigma_X^2, \sigma_Y^2$  and consider the cut-off points  $k(d, \alpha, \sigma_X^2, \sigma_Y^2)$ . Then, we have*

- (a)  $k(d, -1, \sigma_X^2, \sigma_Y^2) \geq k(d, 0, \sigma_X^2, \sigma_Y^2)$  ;
- (b)  $k(d, \alpha, \sigma_X^2, \sigma_Y^2)$  is a non-increasing function of  $\alpha \in [-1, 0]$ .

**Proof.** See the Appendix for part (b). For part (a), setting  $\alpha = 0$ , we have  $\tau = \epsilon(0) = d\underline{R}$ , and

$$k(d, 0, \sigma_X^2, \sigma_Y^2) = \frac{R}{2} + \sqrt{\left(\frac{R}{2}\right)^2 + 1} \leq 1 + \underline{R} = k(d, -1, \sigma_X^2, \sigma_Y^2),$$

as  $x + \sqrt{x^2 + 1} \leq 1 + 2x$  for all  $x > 0$ .

As a consequence of the above, the following simultaneous dominance result is immediate.

**Corollary 3.15** Consider the context of Theorem 3.9 with loss  $L_{\alpha_0}$ ,  $\alpha_0 \leq 0$ , a given  $\hat{\theta}$  such that  $\epsilon > 0$ , and a value of  $c^2 \in (1, k(d, \alpha_0, \sigma_X^2, \sigma_Y^2))$ . Then, the predictive density  $q_{\hat{\theta}, c}$  dominates the plug-in density  $q_{\hat{\theta}, 1}$  for Kullback-Leibler loss as well as other losses  $L_\alpha$  with  $\alpha \leq \alpha_0$ .

### 3.2.5 Example

**Example 3.16** As seen above, Theorem 3.9 is quite general and applies to many situations and many choices of the plug-in estimator  $\hat{\theta}(X)$ . As an illustration, we focus on the positive-part James-Stein estimator given by  $\hat{\theta}_{JS^+}(X)$  (see part (iii) of Remark 3.11).

Theorem 3.9 applies to the predictive densities  $q_{\hat{\theta}_{JS^+}, c}$  for  $d \geq 3$ ,  $\alpha \in (-1, 1)$ ,  $\sigma_X^2, \sigma_Y^2 > 0$ , but we focus for the illustration on the roles of  $d$  and  $c$ , and set  $\alpha = 0$ ,  $\sigma_X^2 = \sigma_Y^2 = 1$ . For implementing Theorem 3.9, we evaluate  $\epsilon$  numerically, which yields  $\tau$  and the cut-off point  $k(d, 0, 1, 1)$ . For  $d = 3$ , we obtain for instance  $\epsilon \approx 1.2009$  and thus  $k(3, 0, 1, 1) \approx 1.2200$ .

Figure 5 compares the plug-in density with the variance expansion matching the cut-off points  $k(d, 0, 1, 1)$  for  $d = 3, 5, 7, 9$ . The gains are moderate, or minimal, depending on  $\|\theta\|$ , and are further attenuated for larger dimension  $d$  in accordance with the phenomenon exhibited at the outset of this section for the plug-in  $\hat{\theta}(X) = X$ .

In opposition to earlier results, Theorem 3.9 condition on the degree of variance expansion is not necessary and sufficient, so there is in theory room for improvement. We proceeded with a numerical evaluation for  $d = 3$  giving dominance if and only if  $1 < c^2 \leq k^*$  with  $k^* \approx 1.4883$ . An otherwise possible choice is given by the expansion  $c^2 = 1 + (1 - \alpha)r/2$ , which is optimal for  $\hat{\theta}(X) = X$ , and equal to  $c^2 = 3/2$  in our case. For  $d = 3$ , in accordance with the numerical evaluation, this does not lead to dominance, although gains are noticeable and more significant on a large part of the parameter space. However, further numerical illustrations suggest worsened performance for larger  $d$ . Finally, as a consequence of Corollary 3.15 we point out that the dominance illustrated here with the Hellinger cut-off points will hold for Kullback-Leibler loss, as well as all other  $\alpha$ -divergence choices with  $\alpha \in (-1, 0)$ .

## 3.3 Concluding remarks

For estimating the density of  $Y \sim N_d(\theta, \sigma_Y^2 I_d)$  based on  $X \sim N_d(\theta, \sigma_X^2 I_d)$ , we establish the inadmissibility of plug-in densities  $q_{\hat{\theta}, 1} \sim N_d(\hat{\theta}, \sigma_Y^2 I_d)$  with respect to  $\alpha$ -divergence

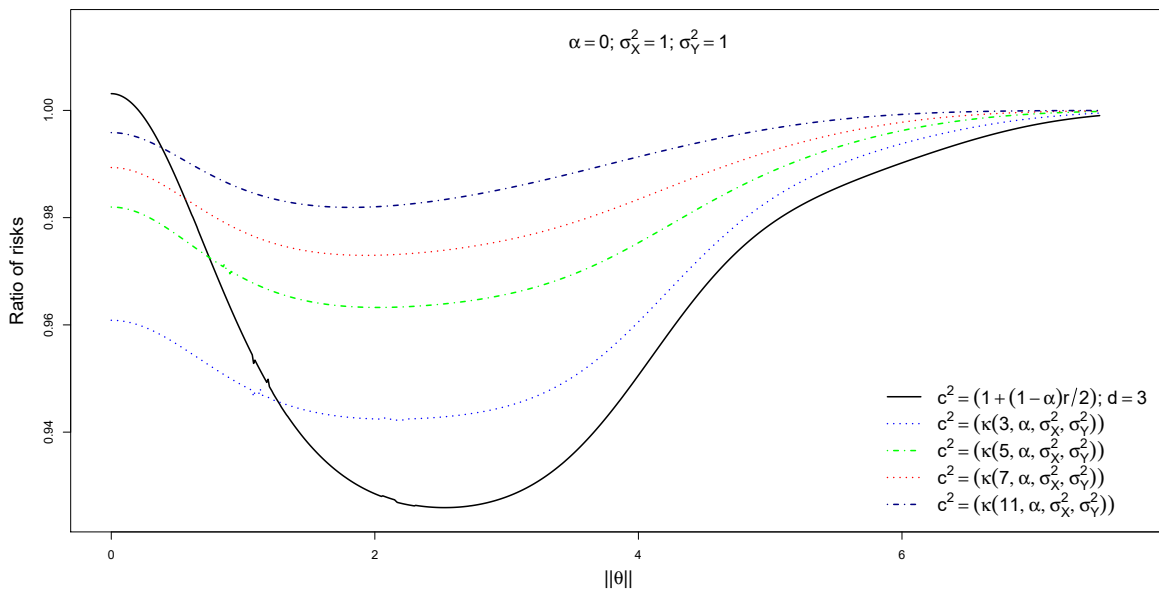


Figure 3.5 – Risk ratios  $\frac{R_{\alpha}(\theta, q_{\hat{\theta}_{JS+}, c})}{R_{\alpha}(\theta, q_{\hat{\theta}_{JS+}, 1})}$  for  $c^2 = k(d, \alpha = 0, \sigma_X^2 = 1, \sigma_Y^2 = 1)$  as functions of  $\|\theta\|$

frequentist risk, and provide explicit variance expansion improvements of the form  $q_{\hat{\theta}, c} \sim N_d(\hat{\theta}, c^2 \sigma_Y^2 I_d)$  with  $c \in (1, c_0]$ . The results are quite general and apply to a large class of plugged-in estimators  $\hat{\theta}(X)$ . Various implications arise, such as those with respect to robustness, as well as simultaneous dominance attained for a class of  $\alpha$ -divergence loss functions, including Kullback-Leibler. Numerical illustrations complement the theory and are quite useful for instance in assessing the degree of improvement.

The findings are applicable in linear models, as well in the presence of normally distributed, or approximatively normally distributed, summary statistics that arise through sufficiency or in asymptotic settings. The theoretical results in this paper highlight deficiencies present in the performance of plug-in densities and give credence to strategies to use alternatives. It would be of interest, for instance, to develop Bayesian improvements and we feel the results here may serve such an objective.

### 3.4 Appendix

#### Proof of Lemma 3.12

Let  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_d)'$ ,  $\theta = (\theta_1, \dots, \theta_d)'$ , and  $Z_i = (\hat{\theta}_i - \theta_i)^2$  for  $i = 1, \dots, d$ . Using the inequality  $\mathbb{E}(Z_i Z_j) \leq \sqrt{\mathbb{E}(Z_i^2) \mathbb{E}(Z_j^2)} + \mathbb{E}(Z_i) \mathbb{E}(Z_j)$ , we have

$$\begin{aligned}
\mathbb{E} \left( \|\hat{\theta}(X) - \theta\|^4 \right) &= \sum_{i,j} \mathbb{E}(Z_i Z_j) \\
&\leq \sum_{i,j} \sqrt{\mathbb{E}(Z_i^2) \mathbb{E}(Z_j^2)} + \sum_{i,j} \mathbb{E}(Z_i) \mathbb{E}(Z_j) \\
&= d^2 \sum_{i,j} \frac{\sqrt{\mathbb{E}(Z_i^2) \mathbb{E}(Z_j^2)}}{d^2} + \left\{ \mathbb{E} \left( \|\hat{\theta}(X) - \theta\|^2 \right) \right\}^2 \\
&\leq d^2 \sqrt{\frac{\sum_{i,j} \mathbb{E}(Z_i^2) \mathbb{E}(Z_j^2)}{d^2}} + M_2^2 \\
&= dM_1 + M_2^2,
\end{aligned}$$

the second inequality a consequence of Jensen's inequality applied to  $g(t) = \sqrt{t}$  on  $(0, \infty)$ .

#### Proof of part (b) of Theorem 3.14

From the definitions of  $k(d, \alpha, \sigma_X^2, \sigma_Y^2)$ ,  $\tau$ , and  $\epsilon$ , we may write

$$k(d, \alpha, \sigma_X^2, \sigma_Y^2) = \left( \beta(\alpha_1, \alpha_2) + \sqrt{\beta^2(\alpha_1, \alpha_2) + \frac{1 + \alpha_1}{1 - \alpha_1}} \right)_{\alpha_1 = \alpha_2 = \alpha}, \quad (3.25)$$

with  $\beta(\alpha_1, \alpha_2) = \frac{\epsilon(\alpha_2)}{2d} - \frac{\alpha_1}{1 - \alpha_1}$  for  $\alpha_1, \alpha_2 \in [-1, 0]$ . With : **(i)**  $\beta(\alpha_1, \alpha_2) \geq 0$  since  $\epsilon(\alpha_2) > 0$  and  $\alpha_1 \leq 0$ , **(ii)**  $\epsilon(\alpha_2)$  is non-increasing in  $\alpha_2 \in [-1, 0]$ , **(iii)**  $\beta + \sqrt{\beta^2 + (1 + \alpha_1)/(1 - \alpha_1)}$  is non-decreasing in  $\beta \geq 0$ , it follows that expression (3.25) is, for fixed  $\alpha_1 \in [-1, 0]$  non-increasing in  $\alpha_2 \in [-1, 0]$ .

To continue, it will thus suffice to show that (3.25) is, for fixed  $\alpha_2 \in [-1, 0]$ , non-increasing in  $\alpha_1 \in [-1, 0]$ . To this end, write (3.25) as

$$T_{\alpha_2}(\alpha_1) = \frac{\epsilon(\alpha_2)}{2d} - \frac{\alpha_1}{1 - \alpha_1} + \sqrt{\frac{\epsilon^2(\alpha_2)}{4d^2} + w(\alpha_1)},$$



with  $w(\alpha_1) = \left(\frac{\alpha_1}{1-\alpha_1}\right)^2 - \frac{\alpha_1 \epsilon(\alpha_2)}{d(1-\alpha_1)} + \frac{1+\alpha_1}{1-\alpha_1}$ ;  $\alpha_1 \in (-1, 0]$ . Finally, we obtain

$$\begin{aligned} \frac{\partial}{\partial \alpha_1} T_{\alpha_2}(\alpha_1) &= -\frac{1}{(1-\alpha_1)^2} + \frac{w'(\alpha_1)}{2\sqrt{\frac{\epsilon^2(\alpha_2)}{4d^2} + w(\alpha_1)}} \\ &\leq \frac{1}{(1-\alpha_1)^3} \left( -(1-\alpha_1) + \frac{1}{\sqrt{\frac{\epsilon^2(\alpha_2)}{4d^2} + w(\alpha_1)}} \right) \\ &\leq 0, \end{aligned}$$

since  $w'(\alpha_1) \leq \frac{\partial}{\partial \alpha_1} \left( \left(\frac{\alpha_1}{1-\alpha_1}\right)^2 + \frac{1+\alpha_1}{1-\alpha_1} \right) = \frac{2}{(1-\alpha_1)^3}$ , and  $w(\alpha_1) \geq \left(\frac{\alpha_1}{1-\alpha_1}\right)^2 + \frac{1+\alpha_1}{1-\alpha_1} = \frac{1}{(1-\alpha_1)^2} \geq 1$ , for  $\alpha_1 \in [-1, 0]$ .

## Acknowledgements

Eric Marchand's research is supported in part by the Natural Sciences and Engineering Research Council of Canada. We thank Bill Strawderman who provided the lower bound in (3.19). Finally, we are grateful to Othmane Kortbi, Iraj Yadegari and Nasser Sadeghkhanian for useful discussions on predictive density estimation under  $\alpha$ -divergence.

# Bibliographie

- [1] Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, **62**, 547-554.
- [2] Aitchison, J. & Dunsmore, I.R. (1975). *Statistical Prediction Analysis*. Cambridge University Press.
- [3] Baranchik, A.J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics*, **41**, 642-645.
- [4] Berger, J.O. (1978). Minimax estimation of a multivariate normal mean under polynomial loss. *Journal of Multivariate Analysis*, **8**, 173-180.
- [5] Brown, L.D., George, E.I., & Xu, X. (2008). Admissible predictive density estimation. *Annals of Statistics*, **36**, 1156-1170.
- [6] Corcuera, J. M. & Giummolè, F. (1999). On the relationship between  $\alpha$  connections and the asymptotic properties of predictive distributions. *Bernoulli*, **5**, 163-176.
- [7] Csiszàr, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2**, 299-318.
- [8] Fourdrinier, D., Marchand, É., Righi, A. & Strawderman, W.E. (2011). On improved predictive density estimation with parametric constraints. *Electronic Journal of Statistics*, **5**, 172-191.
- [9] Fourdrinier, F., D., Ouassou, I. & Strawderman, W.E. (2008). Estimation of a mean vector under quartic loss. *Journal of Statistical Planning and Inference*, **138**, 3841-3857.
- [10] George, E. I., Liang, F. & Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Annals of Statistics*, **34**, 78-91.
- [11] Ghosh, M., Mergel, V. & Datta, G. S. (2008). Estimation, prediction and the Stein phenomenon under divergence loss. *Journal of Multivariate Analysis*, **99**, 1941-1961.
- [12] Kubokawa, T., Marchand, É. & Strawderman, W.E. (2017). On predictive density estimation for location families under integrated absolute value loss. *Bernoulli*, **23**, 3197-3212.

- [13] Kubokawa, T., Marchand, É., & Strawderman, W.E. (2015A). On predictive density estimation for location families under integrated squared error loss. *Journal of Multivariate Analysis*, **142**, 57-74.
- [14] Kubokawa, T., Marchand, É. & Strawderman, W.E. (2015B). On improved shrinkage estimators under concave loss. *Statistics & Probability Letters*, **96**, 241-246.
- [15] LMoudden, A., Marchand, É., O. Kortbi & W. E. Strawderman (2017). On predictive density estimation for Gamma models with parametric constraints. *Journal of Statistical Planning and Inference*, **185**, 56-68.
- [16] Marchand, É., Perron, F., & Yadegari, I. (2017). On estimating a bounded normal mean with applications to predictive density estimation. *Electronic Journal of Statistics*, **11**, 2002-2025.
- [17] Marchand, É. & Sadeghkhan, N. (2017). On predictive density estimation with additional information. arXiv :1709.07778.
- [18] Marchand, É. & Strawderman, W.E. (2012). A unified minimax result for restricted parameter spaces. *Bernoulli*, **18**, 635-643.
- [19] Maruyama, Y. & Ohnishi, T. (2017). Harmonic Bayesian prediction under  $\alpha$ -divergence. arXiv :1605.05899v4
- [20] Maruyama, Y. & Strawderman, W.E. (2012). Bayesian predictive densities for linear regression models under  $\alpha$ -divergence loss : Some results and open problems. *Contemporary Developments in Bayesian analysis and Statistical Decision Theory : A Festschrift for William E. Strawderman*, IMS Collections, **8**, 42-56.
- [21] Yanagimoto, T. & Ohnishi, T. (2009). Bayesian prediction of a density function in terms of e-mixture. *Journal of Statistical Planning and Inference*, **139**, 3064-3075.

# CHAPITRE 4

## Sur les densités prédictives bayésiennes de mélanges

### 4.1 Introduction

Les modèles de mélanges sont omniprésents dans le domaine de statistique. Ces modèles ; qu'ils s'agissent de mélanges fini ou infini, de mélanges d'échelle de distributions normales, des distributions de khi-deux décentrées et de Fisher décentrées ou doublement décentrées ; apparaissent dans diverses situations typiquement liées aux formes quadratiques dans les modèles linéaires normaux et aussi dans divers tests statistiques multivariés, comme le test  $T^2$  de Hotelling pour comparer deux groupes, le test de Pearson pour quantifier la relation entre 2 groupes, test du khi-deux pour la qualité de l'ajustement et bien d'autres. De plus, ces modèles sont très utiles et attrayants pour représenter au mieux les données et les environnements hétérogènes. Ils prennent une place importante dans plusieurs domaines, notamment l'assurance, pour la modélisation des files d'attente, en imagerie médicale et classification, etc.

Cependant, il est difficile d'étudier ces modèles de mélanges. Par exemple, ils représentent un défi pour le calcul bayésien et l'estimation des paramètres par la méthode du maximum de vraisemblance. L'analyse bayésienne et l'inférence statistique des paramètres non-centraux deviennent donc des problèmes à la fois intéressants et difficiles à étudier. Certains de ces problèmes ont été abordés avec des approches bayésienne et

fréquentiste par plusieurs chercheurs [11, 24, 36, 43, 45, 49, 58]. De plus, il est bien connu que plusieurs distributions familières sont représentables en termes de mélanges qui facilite la dérivation de diverses propriétés et approches de l'inférence statistique.

Ce chapitre a pour objectif de décrire des lois a posteriori pour  $\theta$  et des densités prédictives bayésiennes pour  $Y$ , dans le cas de modèles de mélanges pour  $X \sim p_\theta$  et  $Y \sim q_\theta$ . Nous pouvons répertorier les différents types de mélanges dans le schéma suivant

$$\text{Mélanges} \left\{ \begin{array}{l} \text{continus :} \left\{ \begin{array}{l} \text{Type I :} \quad p_\theta(x) = \int f_t(x) dH_\theta(t), \\ \text{Type II :} \quad p_\theta(x) = \int f_{\theta,t}(x) dH(t), \\ \text{Type III :} \quad p_\theta(x) = \int f_{\theta,t}(x) dH_\theta(t), \end{array} \right. \\ \\ \text{discrets :} \left\{ \begin{array}{l} \text{Type I :} \quad p_\theta(x) = \sum_k q_\theta(k) f_k(x), \\ \text{Type II :} \quad p_\theta(x) = \sum_k q(k) f_{\theta,k}(x), \\ \text{Type III :} \quad p_\theta(x) = \sum_k q_\theta(k) f_{\theta,k}(x). \end{array} \right. \end{array} \right. \quad (4.1)$$

*c-à-d*  $X|T = t \sim f_t(\cdot)$  (ou  $X|K = k \sim f_k(\cdot)$ ) avec  $T(\in \mathbb{R}^p) \sim dH(\cdot)$  (ou  $K(\in \mathbb{N}^p) \sim q(\cdot)$ ),

où  $\theta$  est un vecteur de paramètres.

Ici,  $X$  est un vecteur aléatoire de densité  $p_\theta(x)$  continue si  $f$  est une densité, et discrète si  $f$  est une fonction de masse. De plus,  $\theta$  est un paramètre inconnu associé à  $f$  ou  $dH$  (ou  $q$ ) ou les deux selon les trois types de mélanges.

Voici des exemples commentés des situations en (4.1). D'autres exemples, dont la loi de coefficient de détermination  $R^2$  en régression multiples seront abordés plus loin.

**Exemple 4.1** *Mélange d'échelle des lois normales multivariées (Kelker (1970))*

$$p_\theta(x) = \int_{\mathbb{R}_+} \frac{e^{-\frac{\|x-\theta\|^2}{2\nu}}}{(2\pi\nu)^{p/2}} dH(\nu), \quad (4.2)$$

où  $X|\nu \sim N_d(\theta, \nu I_d)$  et  $\nu \sim dH(\cdot)$ , pour  $x$  et  $\theta \in \mathbb{R}^d$  et  $dH(\cdot)$  est une fonction de densité sur  $\mathbb{R}_+$ . Les lois définies par (4.2) sont des mélanges continues de type II.

**Exemple 4.2** *Mélange de loi de Poisson et binomiale négative (Ong et Lee (1979))*

$$p_{\theta}(x) = \sum_{k \geq 0} \frac{\lambda^k e^{-\lambda}}{k!} \binom{k+\nu+x-1}{x} (1-p)^{k+\nu} p^x, \quad x \in \mathbb{N}.$$

*C'est-à-dire  $X|_{K=k} \sim \text{BiNég}(k + \nu, p)$  et  $K \sim \text{Poisson}(\lambda)$ . Dans cet exemple de mélange discret, il est de type I si  $\nu$  ou  $p$  sont inconnus (avec par exemple  $\theta = (\nu, p)$  si les deux sont inconnus), et de type II si  $\lambda$  est inconnu avec  $\theta = \lambda$ .*

**Exemple 4.3** *La classe de distribution mixte d'Erlang (Tijms, 1994) est définie dans le cas univarié par*

$$X \sim p(x|\theta) = \sum_k p(k) f_{\theta,k}(X), \quad x > 0, \quad (4.3)$$

*avec  $p(k)$  est une densité discrète et  $f_{\theta,k}(X) = \text{Ga}(k, \theta)$ . Ce mélange discret est de type II.*

La classe de ces lois est très grande et approchent un grand nombre de distributions continues non négatives.

**Exemple 4.4** *La distribution Gamma bivariée de Kibble a été introduite par Wicksell en 1933 et Kibble en 1941. Tel que présenté en Annexe A, elle apparaît comme distribution d'échantillonnage obtenue à partir d'une loi normale bivariée. La distribution de Kibble, notée  $\text{Kibble}(\nu, \lambda_1, \lambda_2, \rho^2)$  avec  $\nu, \lambda_1, \lambda_2 > 0$  et  $\rho^2 \in (0, 1)$  est définie par la densité*

$$p_{\theta}(x) = \sum_{k \geq 0} \frac{(\nu)_k}{k!} (1 - \rho^2)^{\nu} \rho^{2k} \text{Ga}(x_1 | \nu + k, \frac{1 - \rho^2}{\lambda_1}) \text{Ga}(x_2 | \nu + k, \frac{1 - \rho^2}{\lambda_2}), \quad (x_1, x_2) \in \mathbb{R}_+^2. \quad (4.4)$$

*Ici, on retrouve*

- (a) *Pour le cas  $\nu, \rho^2$  connus, un mélange discret de type II avec  $\theta = (\lambda_1, \lambda_2)$ ,  $q(k)$  la fonction de masse d'une  $\text{BiNég}(\nu, 1 - \rho^2)$  et  $f_{\theta,k}$  la densité conjointe de deux lois Gamma indépendantes avec paramètres  $(\nu + k, \frac{\lambda_1}{1 - \rho^2})$  et  $(\nu + k, \frac{\lambda_2}{1 - \rho^2})$ .*
- (b) *Pour le cas  $\lambda_1, \lambda_2, \rho^2$  inconnus (et même  $\nu, \lambda_1, \lambda_2, \rho^2$  inconnus), on se retrouve avec un mélange discret de type III.*

*L'inférence bayésienne et fréquentiste dans le cas (a) ont été abordées par Liopoulos et coll. (2005).*

Bien sûr, il est facilement envisageable de généraliser la représentation (4.4) de plusieurs manières, par exemple en posant  $K \sim q_\theta$  soit une autre loi discrète sur  $\mathbb{N}$ , comme la loi de Poisson, la loi géométrique, la loi logarithmique etc, obtenant une classe de densités bivariées :

$$p_\theta(x_1, x_2) = \sum_k q_\theta(k) Ga(x_1 | k + \nu_1, \lambda_1 \theta) Ga(x_2 | k + \nu_2, \lambda_2 \theta). \quad (4.5)$$

**Exemple 4.5** *Un autre exemple de type III est la loi de Lindley (Lindley, 1958) avec densité*

$$p_\theta(x) = \frac{\theta^2}{1 + \theta} (x + 1) e^{-\theta x} \mathbb{I}_{(0, \infty)}(x).$$

*C'est un mélange discret, avec  $X|_{K=k} \sim Ga(1 + k, \theta)$  et  $K \sim Ber\left(\frac{\theta}{1 + \theta}\right)$ .*

**Exemple 4.6** *La loi du  $\chi_\nu^2(\theta)$  décentrée est omniprésente dans la théorie distributionnelle des formes quadratiques (par exemple Muirhead (1985)). Sa densité de probabilité est donnée par :*

$$p_\theta(x) = \sum_{k \geq 0} \frac{\left(\frac{\theta}{2}\right)^k e^{-\frac{\theta}{2}}}{k!} Ga\left(x | \nu + k, \frac{1}{2}\right), \quad x \in \mathbb{R}_+.$$

*C'est un mélange discret de type I pour  $\theta$  inconnu.*

**Exemple 4.7** *Une variable aléatoire Fisher doublement décentrée (Scheffe (1959))  $X$  avec paramètres  $\nu_1 > 0, \nu_2 > 0, \theta = (\theta_1, \theta_2) \in (0, \infty)^2$  a une fonction de densité*

$$p_\theta(x) = \sum_{k_1, k_2 \geq 0} \frac{\frac{e^{-\frac{\theta_1}{2}} \left(\frac{\theta_1}{2}\right)^{k_1}}{k_1!} \frac{e^{-\frac{\theta_2}{2}} \left(\frac{\theta_2}{2}\right)^{k_2}}{k_2!} \nu_1^{\frac{\nu_1}{2} + k_1} \nu_2^{\frac{\nu_2}{2} + k_2} x^{\frac{\nu_1}{2} + k_1 - 1} (\nu_2 + \nu_1 x)^{-\frac{\nu_1 + \nu_2}{2} - k_1 - k_2}}{B\left(\frac{\nu_1}{2} + k_1; \frac{\nu_2}{2} + k_2\right)}. \quad (4.6)$$

*Il s'agit d'une mélange de type I discret avec  $k = (k_1, k_2)$  a pour distribution d'une Poisson bivariée indépendante avec paramètres  $(\theta_1/2, \theta_2/2)$ , et une densité Bêta de type II avec paramètres  $(\frac{\nu_1}{2} + k_1, \frac{\nu_2}{2} + k_2, \frac{\nu_2}{\nu_1})$ , telle que présentée à la Définition 2.7.*

*Si  $\theta_2 = 0$ , la densité défini par l'équation (4.6) devient la densité de Fisher décentrée.*

*Par ailleurs la densité en (4.6) est celle du rapport  $X = \frac{X_1}{X_2}$  où  $X_i$  indép  $\sim \chi_{\nu_i}^2(\theta_i)$ .*

**Remarque 4.8** *Si*

$$X|_{T=t} \sim Ga(a, \frac{1}{t}), \quad \text{où } T \sim Ga(b, \frac{1}{\sigma}), \quad \text{alors } X \sim Beta_{II}(a, b, \sigma),$$

*ce qui nous indique que la loi  $Beta_{II}(a, b, \sigma)$  peut être également représentée par un mélange continu de type I ou II.*

Maintenant, nous allons étudier l'estimation de la densité prédictive  $\hat{q}(y; x)$  pour  $Y \sim q_\theta(y)$  basé sur l'observation de  $X \sim p_\theta(x)$  où  $p_\theta(\cdot)$  et  $q_\theta(\cdot)$  sont des densités de mélanges de Lebesgue avec  $\theta$  inconnu, via la fonction de perte de Kullback-Leibler. Parmi les méthodes à notre disposition, on retrouve celles par substitution et la méthode bayésienne. Or, la détermination de l'estimateur du maximum de vraisemblance est complexe et nécessite des techniques et algorithmes spéciaux, dont l'algorithme lié à l'espérance-maximisation EM (Expectation Maximisation algorithm) de Dempster et coll. (1977). On a aussi vu que les densités obtenues par maximum de vraisemblance peuvent être inefficaces en comparaison avec d'autres choix. Pour remédier à ce problème et aussi de cataloguer des solutions bayésiennes, nous adoptons la méthode bayésienne et nous présentons de nouveaux résultats pour les deux cas de mélange (continu et discret).

Dans le cas de mélange continu, on obtient des formes fermées pour les distributions *a posteriori* et prédictives. Un cas particulier des lois mélanges incluant le paramètre de position  $\theta$  sera traité de manière approfondie. Dans le cas de mélange discret, on présente les lois *a posteriori* et les densités prédictives avec des exemples. De plus, on présente de nouvelles lois intéressantes.

## 4.2 Mélanges continus

Dans cette section, nous traitons le cas des variables aléatoires  $X$  et  $Y$  continues multivariées de même type I ou II. On considère alors le modèle

$$X|\theta \sim \int f_t(x) dH_\theta(t), \quad Y|\theta \sim \int g_r(x) dK_\theta(r), \quad (\text{type I}) \quad (4.7)$$

ainsi que

$$X|\theta \sim \int f_{\theta,t}(x) dH(t), \quad Y|\theta \sim \int g_{\theta,r}(x) dK(r), \quad (\text{type II}) \quad (4.8)$$

avec  $x, y \in \mathbb{R}^d$ .



On aborde en premier les mélanges de type II pour les modèles à paramètre de position ou d'échelle avec loi a priori non-informative  $\pi_0(\theta) = 1$  et  $\pi_0(\theta) = \frac{1}{\theta}$  respectivement. À titre d'exemple, les résultats ci-dessus s'appliqueront aux mélanges de loi normale de l'Exemple 4.2, ainsi qu'autres mélanges de la forme

$$Y|\theta \sim \int_{\mathbb{R}_+} \frac{1}{r^d} g\left(\frac{y-\theta}{r}\right) dH(r),$$

où  $g$  est une densité sur  $\mathbb{R}^d$  (par exemple Logistique, Cauchy pour  $d = 1$  et  $d > 1$ ). On rappelle la définition de convolution qui nous sera utile.

**Définition 4.9** *Le produit de convolution de deux fonctions,  $f$  et  $g$ , est une fonction qui se note généralement  $f * g$  et qui est définie par*

$$f * g(t) = \int_{\mathbb{R}^d} f(t-s)g(s)ds, \quad t \in \mathbb{R}^d.$$

**Lemme 4.10** *Pour estimer la densité de  $Y \sim q_\theta(y) = \int g_r(y-\theta) dK(r)$  à partir de  $X \sim p_\theta(x) = \int f_t(x-\theta) dH(t)$  sous la perte Kullback-Leibler, la densité prédictive  $\hat{q}_\pi$  associée aux lois a priori  $\pi_0(\theta) = 1$  est donnée par :*

$$\hat{q}_{\pi_0}(y; x) = \int \int g_r * \bar{f}_t(y-x) dH(t) dK(r), \quad \text{avec } \bar{f}_t(x) = f_t(-x).$$

**Remarque 4.11** *La densité  $\hat{q}_{\pi_0}$  est la meilleure densité équivariante sous les changements de position.*

**Démonstration.** On a  $X \sim \int f_t(x-\theta) dH(t)$  et  $\pi_0(\theta) = 1$ , alors

$$\pi(\theta|x) = \int f_t(x-\theta) dH(t),$$

et la densité prédictive est donnée par

$$\begin{aligned} \hat{q}_{\pi_0}(y; x) &= \int q_\theta(y) \pi(\theta|x) d\theta = \int \int g_r(y-\theta) dK(r) \int f_t(x-\theta) dH(t) d\theta. \\ &= \int \int \left\{ \int g_r(y-\theta) f_t(x-\theta) d\theta \right\} dH(t) dK(r) \\ &= \int \int \left\{ \int g_r(y-x-(\theta-x)) f_t(-(\theta-x)) d\theta \right\} dH(t) dK(r) \end{aligned}$$

$$\begin{aligned}
&= \int \int \left\{ \int g_r((y-x) - s) f_t(-s) ds \right\} dH(t) dK(r), \text{ avec } s = (\theta - x) \\
&= \int \int g_r * \bar{f}_t(y-x) dH(t) dK(r),
\end{aligned}$$

ce qui est le résultat annoncé.

**Exemple 4.12** Pour les mélanges de loi normale  $X|\sigma_X^2 \sim N_d(\theta, \sigma_X^2 I_d)$  avec  $\sigma_X^2 \sim dH(\sigma_X^2)$ ,  $Y|\sigma_Y^2 \sim N_d(\theta, \sigma_Y^2 I_d)$  avec  $\sigma_Y^2 \sim dK(\sigma_Y^2)$  et pour la loi a priori  $\pi_0(\theta) = 1$ , la densité prédictive  $\hat{q}_{\pi_0}$  est une densité mélange admettant la représentation

$$Y|\sigma^2 \sim N_d(X, \sigma^2 I_d), \text{ avec } \sigma^2 = \sigma_X^2 + \sigma_Y^2 \sim dL(\cdot).$$

La représentation découle du Lemme 4.10 puisque  $g_{\sigma_Y} * \bar{f}_{\sigma_X}$  est celui du produit de convolution de loi  $N_d(0, \sigma_X^2 I_d)$  et  $N_d(0, \sigma_Y^2 I_d)$ , qui est rien d'autre qu'une loi  $N_d(0, (\sigma_Y^2 + \sigma_X^2) I_d)$ .

Dans un autre cas, si  $\theta$  est un paramètre d'échelle et  $d = 1$ , c'est-à-dire  $f_{\theta,t}(x) = \frac{1}{\theta} f_t(\frac{x}{\theta})$  et  $g_{\theta,r}(y) = \frac{1}{\theta} g_r(\frac{y}{\theta})$ , (par exemple loi normale, Gamma, Cauchy...), nous avons le Lemme suivant.

**Lemme 4.13** Pour estimer la densité de  $Y \sim q_\theta(y) = \int \frac{1}{\theta} g_r(\frac{y}{\theta}) dK(r)$  à partir de  $X \sim p_\theta(x) = \int \frac{1}{\theta} f_t(\frac{x}{\theta}) dH(t)$  sous la perte Kullback-Leibler, la densité prédictive  $\hat{q}_\pi$  associée aux lois a priori  $\pi_0(\theta) = \frac{1}{\theta} \mathbb{I}_{(0, \infty)}(\theta)$  est donnée par :

$$\hat{q}_{\pi_0}(y; x) = \int \int \frac{1}{x} h_{t,r}\left(\frac{x}{y}\right) dH(t) dK(r),$$

$$\text{avec } h_{t,r}(v) = \int_0^\infty u g_r(uv) f_t(u) du.$$

**Démonstration.** Nous avons

$$\begin{aligned}
\hat{q}_{\pi_0}(y; x) &= \frac{\int \int \int_0^\infty \frac{1}{\theta} g_r(\frac{y}{\theta}) \frac{1}{\theta} f_t(\frac{x}{\theta}) \frac{d\theta}{\theta} dH(t) dK(r)}{\int \int_0^\infty \frac{1}{\theta} f_t(\frac{x}{\theta}) \frac{d\theta}{\theta} dH(t)} \\
&= \frac{1}{x} \frac{\int \int \int_0^\infty u g_r(\frac{y}{x}u) f_t(u) du dH(t) dK(r)}{\int \int_0^\infty f_t(u) du dH(t)}, \text{ avec } u = \frac{x}{\theta} \\
&= \frac{1}{x} \int \int \int_0^\infty u g_r\left(\frac{y}{x}u\right) f_t(u) du dH(t) dK(r)
\end{aligned}$$

$$= \int \int \frac{1}{x} h_{t,r} \left( \frac{y}{x} \right) dH(t) dK(r),$$

ce qui est le résultat annoncé.

**Exemple 4.14** Pour les mélanges de loi Gamma  $X|\beta \sim Ga(\alpha_1, \beta)$  avec  $\alpha_1 \sim Exp(\gamma_1)$  et  $Y|\beta \sim Ga(\alpha_2, \beta)$  avec  $\alpha_2 \sim Exp(\gamma_2)$ , et pour la loi a priori  $\pi_0(\beta) = \frac{1}{\beta} \mathbb{I}_{(0, \infty)}(\beta)$ , la densité prédictive est une densité mélange donnée par

$$\hat{q}_{\pi_0}(y; x) = \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \frac{1}{x} h_{\alpha_1, \alpha_2} \left( \frac{y}{x} \right) \gamma_1 \gamma_2 e^{-(\gamma_1 \alpha_1 + \gamma_2 \alpha_2)} d\alpha_1 d\alpha_2,$$

avec  $h_{\alpha_1, \alpha_2}$  étant la densité d'une loi Bêta de type II avec paramètres de formes  $\alpha_2$  et  $\alpha_1$ , telle que présentée au Définition 2.2 de Chapitre 2.

**Remarque 4.15** Le Lemme 4.10 étend un résultat de Kubokawa et coll. (2015), tandis que le Lemme 4.13 étend le Théorème 2.4 de Lmoudden et coll. (2017).

Un cas plus général pour les deux type de mélange I et II est donné par la proposition suivante.

**Proposition 4.16** Soient  $X|\theta$  et  $Y|\theta$  des variables aléatoires avec densités de la même forme (4.7) ou (4.8), en remplaçant les différentielles  $dH(t)$  par  $h(t)dt$  et  $dK(r)$  par  $k(r)dr$ , et  $\pi$  une loi a priori quelconque de densité absolument continue par rapport à la mesure  $\sigma$ -finie  $\nu$ . Nous avons

a) **Mélange de type I** : La densité a posteriori de  $\theta$  est donnée par

$$\pi(\theta|x) = \int \pi_t(\theta) f_x^*(t) dt, \quad (4.9)$$

où  $\pi_t$  et  $f_x^*$  sont des fonctions de densités données par  $\pi_t(\theta) = \frac{h_\theta(t) \pi(\theta)}{m(t)}$  et

$$f_x^*(t) = \frac{m(t) f_t(x)}{\int m(t) f_t(x) dt} \text{ avec } m(t) = \int_{\Theta} h_\theta(t) \pi(\theta) d\nu(\theta).$$

La densité prédictive pour  $Y$  sous la perte Kullback-Leibler est donnée par

$$\hat{q}_\pi(y; x) = \int \int g_r(y) f_x^*(t) \hat{q}_\pi^*(r; t) dt dr, \quad (4.10)$$

où  $\hat{q}_\pi^*$  est donnée par

$$\hat{q}_\pi^*(r; t) = \int_{\Theta} k_\theta(r) \pi_t(\theta) d\nu(\theta). \quad (4.11)$$

b) **Mélange de type II** : La densité a posteriori de  $\theta$  est donnée par

$$\pi(\theta|x) = \int \pi_{x,t}(\theta) h_x^*(t) dt, \quad (4.12)$$

où  $\pi_{x,t}$  et  $h_x^*$  sont des fonctions de densités données par  $\pi_{x,t}(\theta) = \frac{f_{\theta,t}(x) \pi(\theta)}{m_t(x)}$  et

$$h_x^*(t) = \frac{m_t(x) h(t)}{\int m_t(x) h(t) dt}, \text{ avec } m_t(x) = \int_{\Theta} f_{\theta,t}(x) \pi(\theta) d\nu(\theta).$$

La densité prédictive pour  $Y$  sous la perte Kullback-Leibler est donnée par

$$\hat{q}_\pi(y; x) = \int \int \hat{q}_{\pi,t,r}(y; x) h_x^*(t) k(r) dr dt, \quad (4.13)$$

où  $\hat{q}_{\pi,t,r}$  est donnée par

$$\hat{q}_{\pi,t,r}(y; x) = \int_{\Theta} g_{\theta,r}(y) \pi_{x,t}(\theta) d\nu(\theta). \quad (4.14)$$

**Démonstrations.** On va faire une démonstration pour le mélange de type II, une démonstration analogue prévaut pour le mélange de type I. Nous avons pour

$$X|\theta \sim \int f_{\theta,t}(x) h(t) dt \text{ et } Y|\theta \sim \int g_{\theta,r}(x) k(r) dr$$

la densité a posteriori

$$\begin{aligned} \pi(\theta|x) &= \frac{\int f_{\theta,t}(x) \pi(\theta) h(t) dt}{\int \int f_{\theta,t}(x) \pi(\theta) h(t) dt d\nu(\theta)} = \frac{\int f_{\theta,t}(x) \pi(\theta) \int f_{\theta,t}(x) \pi(\theta) d\nu(\theta) h(t) dt}{\int \int f_{\theta,t}(x) \pi(\theta) d\nu(\theta)} \\ &= \int \underbrace{\frac{f_{\theta,t}(x) \pi(\theta)}{\int f_{\theta,t}(x) \pi(\theta) d\nu(\theta)}}_{=\pi_{x,t}(\theta)} \underbrace{\frac{\int f_{\theta,t}(x) \pi(\theta) d\nu(\theta) h(t)}{\int \int f_{\theta,t}(x) \pi(\theta) d\nu(\theta) h(t) dt}}_{=h_x^*(t)} dt. \end{aligned}$$

Pour la densité prédictive sous la perte Kullback-Leibler, nous avons

$$\begin{aligned} \hat{q}_\pi(y; x) &= \int \int g_{\theta,r}(y) k(r) dr \pi(\theta|x) d\nu(\theta) = \int \int g_{\theta,r}(y) k(r) dr \int \pi_{x,t}(\theta) h_x^*(t) dt d\nu(\theta) \\ &= \int \int \underbrace{\int g_{\theta,r}(y) \pi_{x,t}(\theta) d\nu(\theta)}_{\hat{q}_{\pi,t,r}(y; x)} h_x^*(t) k(r) dt dr. \end{aligned}$$

Il est utile d'interpréter l'expression de la Proposition 4.16. Par exemple, la densité  $\hat{q}_{\pi, t, r}$  en (4.14) est tout simplement la densité prédictive pour la loi de  $Y|r$  à partir de l'observé  $X$  tiré de la loi de  $X|t$ . Et après, la densité  $\hat{q}_{\pi}$  en (4.13) intègre cette densité par rapport aux densités de  $T|x$  et de  $R$  où  $T \sim h$  et  $R \sim k$ .

**Remarque 4.17** *Nous remarquons que les distributions a posteriori et prédictive sont aussi des modèles de mélanges du même type que  $Y|\theta$ .*

Voici une illustration de la Proposition 4.16.

**Exemple 4.18** *Soient  $X|\theta$  et  $Y|\theta$  deux variables aléatoires indépendantes de lois  $Beta_{II}(a, b, \theta)$  et  $Beta_{II}(c, d, \theta)$  respectivement, qui sont des mélanges de type I de deux lois Gamma d'après la Remarque 4.8, et  $\theta \sim Ga(\alpha, \frac{1}{\beta})$ . Pour la loi de  $X|\theta$ , on a  $f_t(x) = Ga(a, \frac{1}{t})$  et  $h_{\theta}(t) = Ga(b, \frac{1}{\theta})$ , alors  $m(t) = \frac{t^{b-1} \beta^{\alpha}}{B(b, \alpha) (\beta+t)^{b+\alpha}}$  ce qui implique que  $\pi_t(\theta) \propto \theta^{b+\alpha-1} e^{-(\beta+t)\theta}$  et  $f_x^*(t) \propto \frac{t^{a+b-1}}{(t+\beta)^{\alpha+b}} e^{-tx}$ , c-à-d  $\pi_t(\theta) = Ga(\alpha + b, \frac{1}{t+\beta})$  et  $f_x^*(t) = Ga_{II}(a + b, x, \beta, 1 - (\alpha + b))$ , où  $Ga_{II}$  est une loi Gamma de type II (voir la Définition 1.10 de la Section 1.1.3), la densité a posteriori est donnée par*

$$\pi(\theta|x) = \int_{\mathbb{R}_+} Ga(\theta|b + \alpha, \frac{1}{t + \beta}) Ga_{II}(t|a + b, x, \beta, 1 - (\alpha + b)) dt.$$

Également pour  $Y|\theta$ , on a  $g_r(y) = Ga(c, \frac{1}{r})$  et  $k_{\theta}(r) = Ga(d, \frac{1}{\theta})$ , alors

$$\hat{q}_{\pi}(r; t) = \int_{\mathbb{R}_+} Ga(r|d, \frac{1}{\theta}) Ga(\theta|\alpha+b, \frac{1}{t + \beta}) d\theta = Beta_{II}(d, \alpha+b, \frac{1}{t + \beta}), \text{ et la densité}$$

prédictive sous la perte Kullback-Leibler

$$\hat{q}_{\pi}(y; x) = \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} Ga(y|c, \frac{1}{r}) Ga_{II}(t|a+b, x, \beta, 1 - (\alpha+b)) Beta_{II}(r|d, \alpha+b, t+\beta) dt dr.$$

### 4.3 Mélanges discrets

Dans cette section, nous allons étudier les densités a posteriori et prédictive pour des mélanges discrets. Nous visons également des représentations utiles pour des études ultérieures, des connexions informatives et diverses illustrations et applications. Ces derniers

comprennent des mélanges de Gamma multivariés, le mélange bivarié comme la distribution de Kibble, la loi du Chi-carré décentrée, les distributions F décentrée et une distribution d'échantillonnage du coefficient de détermination  $R^2$  apparaissant en régression multiple. Dans ces cas, les densités de mélange sont Poisson ou Binomiale Négative lorsque les distributions mélangées sont des distributions Gamma, Bivariées Gamma, Bêta de type I et II.

Dans cette section, nous traitons le cas des variables aléatoires  $X$  et  $Y$  sont de même type I ou II. Nous présentons d'abord le premier cas de mélanges.

### 4.3.1 $X|\theta$ et $Y|\theta$ de type I

Considérons les modèles de mélanges de la forme

$$X \sim p(x|\theta) = \sum_k p_\theta(k) f_k(x), \quad Y \sim q(y|\theta) = \sum_k q_\theta(k) g_k(y), \quad (4.15)$$

avec  $x$  et  $y \in \mathbb{R}^d$ , où  $f_k$  et  $g_k$  sont des densités de Lebesgue et les fonctions  $p_\theta$  et  $q_\theta$  sont des fonctions de masse. Nous fournissons une analyse bayésienne pour  $\theta$  basée sur  $X$ , ainsi que des densités prédictives bayésiennes pour  $Y$  basées sur  $X$ . La plupart du développement est fait avec des distributions familières. Nous sommes particulièrement intéressés par les structures générales de ces solutions bayésiennes.

#### La distribution a posteriori

**Lemme 4.19** *Soit  $X|\theta$  une variable aléatoire distribuée comme dans le modèle (4.15) et  $\pi$  une densité a priori absolument continue de  $\theta$  par rapport à une mesure  $\sigma$ -finie  $\nu$ . La densité a posteriori de  $\theta$  est donnée par*

$$\pi(\theta|x) = \sum_k f_x^*(k) \pi_k(\theta), \quad (4.16)$$

où  $\pi_k(\theta) = \frac{\pi(\theta)p_\theta(k)}{\int_{\Theta} \pi(\theta)p_\theta(k) d\nu(\theta)}$  et  $f_x^*$  est une fonction de masse donnée par,

$$f_x^*(k) = \frac{f_k(x) m_\pi(k)}{\sum_k f_k(x) m_\pi(k)} \text{ avec } m_\pi(k) = \int_{\Theta} p_\theta(k) \pi(\theta) d\nu(\theta) .$$

**Remarque 4.20** Comme la distribution de  $X|\theta$ ,  $\pi(\theta|x)$  est aussi un modèle de mélange ( les  $f_k$  's et  $p_\theta(k)$ , dans la distribution de  $X|\theta$ , sont respectivement remplacées par  $\pi_k$  et  $f_x^*(k)$ ). De plus, on remarque que  $\pi_k$  correspond à la densité a posteriori de  $\theta$  basée sur l'observation de  $K = k$ , alors que  $m_\pi(\cdot)$  correspond à la fonction de masse de probabilité marginale de  $K \sim p_\theta$ .

**Démonstration du Lemme 4.19.**

Pour  $X \sim p(x|\theta) = \sum_k p_\theta(k) f_k(x)$  et  $\theta \sim \pi(\theta)$ , nous avons

$$\begin{aligned} \pi(\theta|x) &= \frac{p(x|\theta) \pi(\theta)}{\int_{\Theta} p(x|\theta) \pi(\theta) \nu(\theta)} = \frac{\sum_k f_k(x) p_\theta(k) \pi(\theta)}{\sum_k f_k(x) \int_{\Theta} p_\theta(k) \pi(\theta) d\nu(\theta)} \\ &= \sum_k \frac{f_k(x) m_\pi(k)}{\sum_k f_k(x) m_\pi(k)} \frac{p_\theta(k) \pi(\theta)}{m_\pi(k)} = \sum_k f_x^*(k) \pi_k(\theta). \end{aligned}$$

**Exemple 4.21** Considérons un paramètre de mélange de loi de Poisson où  $p_\theta(k) = \frac{e^{-\theta} \theta^k}{k!} \mathbb{I}_{\mathbb{N}}(k)$  mélangé avec une densité  $f_k$  et pour l'a priori  $\pi(\theta) \sim Ga(a, \frac{1}{b})$ ;  $a, b > 0$ . À partir de ceci, nous obtenons  $\pi_k$  comme une densité  $Ga(a+k, \frac{1}{1+b})$  et  $m_\pi(k)$  la fonction de masse d'un BiNég( $a, b/(1+b)$ ). D'après le Lemme 4.19, la distribution a posteriori  $\pi(\theta|x)$  est un mélange des  $\pi_k$  's avec la fonction de masse sur  $\mathbb{N}$  donnée par

$$f_x^*(k) \propto \frac{(a)_k}{k!} \frac{b^a}{(1+b)^{a+k}} f_k(x).$$

Maintenant, considérons les cas spécifiques **(i)**  $f_k \sim Ga(p/2+k, 2)$ , **(ii)**  $f_k \sim Beta(p/2+k, q)$  et **(iii)**  $f_k \sim Beta_{II}(p/2+k, q/2)$ . Le deuxième cas est une densité Bêta de type I sur  $(0, 1)$  proportionnelle à  $x^{p/2+k}(1-x)^{q/2}$ . Le troisième est une densité bêta de type II sur  $\mathbb{R}_+$  proportionnelle à  $\frac{x^{p/2+k-1}}{(1+x)^{p/2+q/2+k}}$ . Nous soulignons que le cas **(i)** correspond à une distribution  $\chi^2$  non centrée avec  $p$  degrés de liberté et paramètre de non-centralité  $2\theta$ . Tandis que le cas **(ii)** correspond à une distribution Bêta non-centrée avec paramètres de formes  $p/2, q/2$  et paramètre de non-centralité  $2\theta$ . Finalement, le cas **(iii)** est une distribution Fisher décentrée avec des degrés de liberté  $p, q$  et paramètre de non-centralité  $2\theta$ .

Pour **(i)**, on obtient

$$f_x^*(k) = \frac{1}{{}_1F_1(a; p/2; \frac{x}{2(1+b)})} \frac{(a)_k}{k! (p/2)_k} \left( \frac{x}{2(1+b)} \right)^k. \quad (4.17)$$

Dans le cas particulier  $2a = p$ , la fonction de masse ci-dessus se réduit à une fonction de masse de Poisson de paramètre  $\frac{x}{2(1+b)}$ . Dans le cas contraire, il s'agit d'une distribution de séries de puissance, également appelée hypergéométrique généralisée (voire la formule (1.1)).

Pour **(ii)**, nous obtenons

$$f_x^*(k) = C \frac{(a)_k \left(\frac{p+q}{2}\right)_k}{k! \left(\frac{p}{2}\right)_k} \left(\frac{x}{1+b}\right)^k, \text{ avec } C^{-1} = {}_2F_1\left(a, \frac{p+q}{2}; p/2; \frac{x}{(1+b)}\right)$$

et pour **(iii)**, nous obtenons

$$f_x^*(k) = C \frac{(a)_k \left(\frac{p+q}{2}\right)_k}{k! \left(\frac{p}{2}\right)_k} \left(\frac{x}{(1+x)(1+b)}\right)^k,$$

$$\text{avec } C^{-1} = {}_2F_1\left(a, (p+q)/2; p/2; \frac{x}{(1+x)(1+b)}\right).$$

Nous observons que le cas  $2a = p$  se réduit à une fonction de masse binomiale négative pour (ii) et (iii).

## Densités prédictives

Nous considérons maintenant les densités prédictives de Bayes  $\hat{q}_\pi$  associées à  $\pi$ , pour  $q_\theta(\cdot)$ , sous la perte de Kullback-Leibler.

**Lemme 4.22** Basée sur  $X \sim p(x|\theta) = \sum_k p_\theta(k) f_k(x)$  et une loi a priori absolument continue de  $\theta$  par rapport à une mesure  $\sigma$ -finie  $\nu$ , la densité prédictive de Bayes pour  $Y \sim q(y|\theta) = \sum_{k^*} q_\theta(k^*) g_{k^*}(y)$  sous la perte de Kullback-Leibler est donnée par

$$\hat{q}_\pi(y; x) = \sum_{k^*} q_x^*(k^*) g_{k^*}(y), \quad (4.18)$$

où  $q_x^*$  est une fonction de masse donnée par

$$q_x^*(k^*) = \sum_k \hat{q}_\pi(k^*|k) f_x^*(k), \quad (4.19)$$

avec  $\hat{q}_\pi(k^*|k) = \int_{\Theta} q_\theta(k^*) \pi_k(\theta) d\nu(\theta)$  et  $\pi_k(\cdot)$  et  $f_x^*(\cdot)$  telles que définies dans le Lemme 4.19.



**Remarque 4.23**

- a) La densité prédictive de Bayes  $\hat{q}_\pi$  est un mélange de même type que  $q(\cdot|\theta)$ , mais avec les poids a posteriori  $q_x^*(k^*)$  remplaçant les poids a priori  $q_\theta(k^*)$ . Ce dernier est aussi un mélange des lois discrètes  $f_x^*$  et  $\hat{q}_\pi$ .
- b) De plus, on remarque que  $\hat{q}_\pi(k^*|k)$  est la densité prédictive pour  $K^* \sim q_\theta$  basée sur l'a posteriori  $\theta|k \sim \pi_k$ .

**Démonstration de lemme 4.22.**

Sous la perte de Kullback-Leibler et pour  $\pi(\theta|x)$  la densité a posteriori (4.16), la densité prédictive pour  $Y \sim q(y|\theta) = \sum_{k^*} q_\theta(k^*)g_{k^*}(\theta)$  est donnée par

$$\begin{aligned} \hat{q}_\pi(y; x) &= \int_{\Theta} q(y|\theta)\pi(\theta|x)d\nu(\theta) = \int_{\Theta} \sum_{k^*} q_\theta(k^*) g_{k^*}(y) \sum_k f_x^*(k)\pi_k(\theta)d\nu(\theta) \\ &= \sum_{k^*} g_{k^*}(y) \sum_k f_x^*(k) \int_{\Theta} q_\theta(k^*)\pi_k(\theta)d\nu(\theta) = \sum_{k^*} q_x^*(k^*) g_{k^*}(y). \end{aligned}$$

Nous poursuivons avec diverses illustrations de Lemme 4.22.

**Exemple 4.24** Suite à l'exemple 4.21, prenons  $p_\theta(k) = \frac{e^{-\theta} \theta^k}{k!} \mathbb{I}_{\mathbb{N}}(k)$ , avec une densité a priori  $\pi(\theta) = \frac{\theta^{\alpha-1} b^\alpha e^{-\theta b}}{\Gamma(\alpha)}$  pour  $f_x^*$  donnée en (4.17). Nous considérons le problème de prédiction avec  $Y$  ayant la même distribution que  $X$  (c.-à-d.  $q(y|\theta) \equiv p(x|\theta)$ ). Alors, d'après l'équation (4.19), nous avons

$$q_x^*(k^*) = \frac{(a)_{k^*} \rho^\alpha (1-\rho)^{k^*}}{k^*!} \frac{{}_1F_1(a+k^*; \frac{\rho}{2}; \frac{\rho x}{2(1+b)})}{{}_1F_1(a; \frac{\rho}{2}; \frac{x}{2(1+b)})}, \quad (4.20)$$

avec  $\rho = \frac{1}{2+b}$ , et on obtient une belle représentation de la densité prédictive

$$\hat{q}_\pi(y; x) = \sum_{k^*} q_x^*(k^*)g_{k^*}(y) = \sum_{k^*} \frac{(a)_{k^*} \rho^\alpha (1-\rho)^{k^*} y^{\frac{\rho}{2}+k^*-1} e^{-\frac{y}{2}} {}_1F_1(a+k^*; \frac{\rho}{2}; \frac{\rho x}{2(1+b)})}{k^*! \Gamma(\frac{\rho}{2} + k^*) 2^{\frac{\rho}{2}+k^*}} \frac{{}_1F_1(a; \frac{\rho}{2}; \frac{x}{2(1+b)})}{{}_1F_1(a; \frac{\rho}{2}; \frac{x}{2(1+b)})}. \quad (4.21)$$

Notons que la formule (4.20) peut être écrite comme

$$q_x^*(k^*) = \sum_j \frac{(a+j)_{k^*} \rho^{a+j} (1-\rho)^{k^*}}{k^*!} \frac{\left(\frac{a}{2}\right)_j \left(\frac{\rho x}{2(1+b)}\right)^j}{\left(\frac{\rho}{2}\right)_j j! \left(\frac{x}{2(1+b)}\right)^j},$$

ce qui nous dit que la fonction de mélange discret  $q_x^*$  admet la représentation :

$$K^* | J = j \sim \text{BiNég}(a+j, 1-\rho) \text{ avec } J \sim \frac{\left(\frac{a}{2}\right)_j \left(\frac{\rho x}{2(1+b)}\right)^j}{\left(\frac{\rho}{2}\right)_j j! \left(\frac{x}{2(1+b)}\right)^j}.$$

C'est-à-dire,  $J$  a une distribution de type *Hypergéométrique généralisée* (Définition 1.8). De plus, si  $a = \frac{\rho}{2}$ , la distribution de  $J$  est *Poisson* $\left(\frac{\rho x}{2(1+b)}\right)$ , et  $q_x^*(k^*)$  est un cas particulier de distribution *binomiale négative non centrée* (voir l'Exemple 4.2), ceci illustre bien la partie b) de la Remarque 4.23. Les densités a posteriori et prédictive sont illustrées aux Figures 4.1 et 4.2, pour  $a = 3$ ,  $b = 5$  et quelques valeurs de  $x$ .

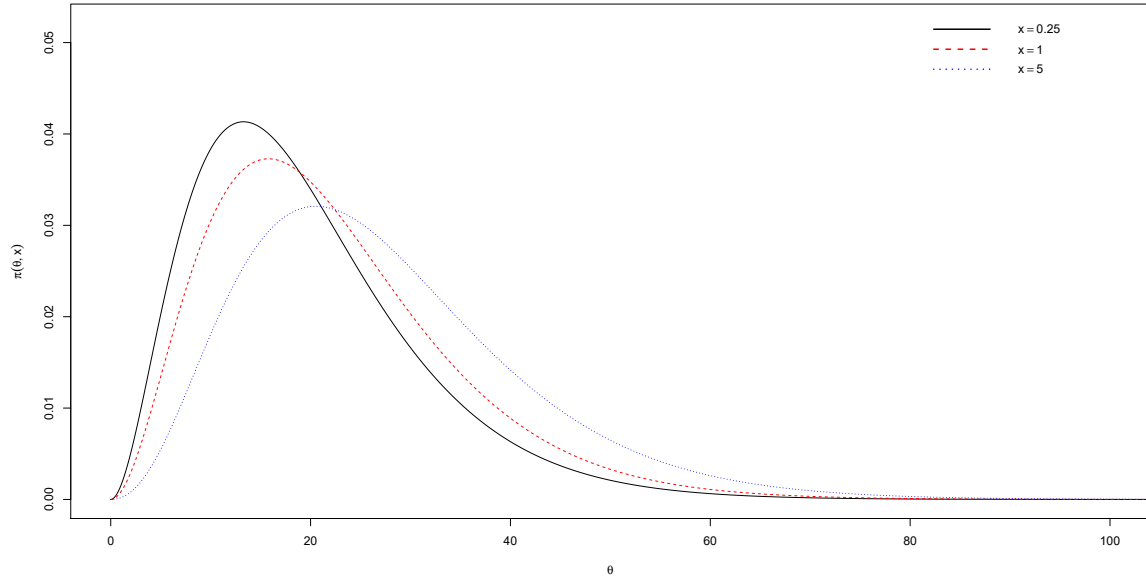


Figure 4.1 – La loi a posteriori  $\pi(\theta|x)$

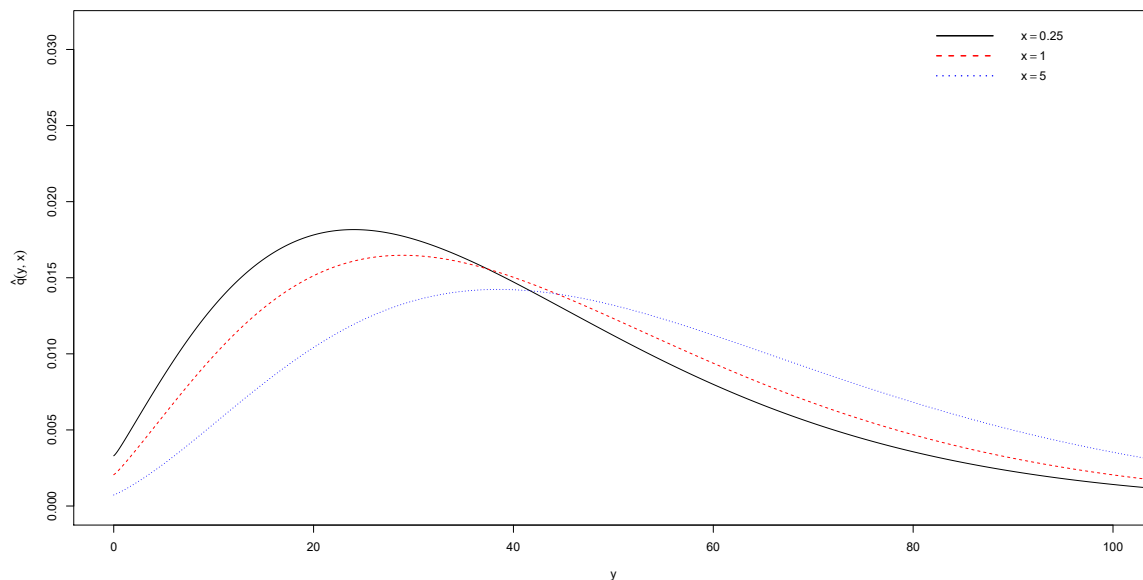


Figure 4.2 – La densité prédictive  $\hat{q}_\pi(y; x)$

Le prochain exemple traite le coefficient de détermination expérimental,  $R^2$ , obtenue d'une régression multiple d'une variable dépendante en fonction de  $m - 1$  variables explicatives. Basé sur un échantillon de taille  $n$  et sous l'hypothèse de normalité des  $m$  variables, la loi du  $R^2$  admet une représentation de type I, comme en (4.14), avec

$$R^2|K \sim \text{Beta} \left( \frac{m-1}{2} + K, \frac{n-m}{2} \right), \quad K \sim \text{BiNég} \left( \frac{n-1}{2}, 1 - \rho^2 \right), \quad (4.22)$$

où  $\rho^2$  est le carré du coefficient de corrélation multiple du modèle sous-jacent (par exemple, voir Muirhead, 1985)<sup>1</sup>. Dans le cadre de l'estimation de  $\rho^2$ , il est d'intérêt d'avoir en plus, à notre portée une méthodologie bayésienne telle qu'abordée par Marchand (2001). Et la représentation s'apprête bien pour appliquer nos résultats ci-dessus. Ainsi, la détermination d'une densité prédictive bayésienne pour une future réalisation de  $R^2$  est d'intérêt. De plus, elle pourrait s'avérer pertinente pour mieux comprendre, ou pour mieux illustrer le comportement attendu pour de futures études, la thématique liée aux difficultés souvent rencontrées à reproduire un résultat significatif obtenu lors d'une étude initiale (par exemple, voir Baker, 2016).

---

1. Il est à noter que la loi de probabilité pour un coefficient de détermination partiel admet la même type de représentation.

**Exemple 4.25** Nous supposons que la loi de  $R_1^2$  (résp  $R_2^2$ ), le coefficient de corrélation basé sur un échantillon de taille  $n_1$  (résp  $n_2$ ), admet la représentation (4.22) avec

$$R_1^2|k \sim \text{Beta}\left(\frac{m-1}{2} + k, \frac{n_1 - m}{2}\right), \quad k \sim \text{BiNeg}\left(\frac{n_1 - 1}{2}, \frac{1}{1 + \theta}\right),$$

$$R_2^2|k \sim \text{Beta}\left(\frac{m-1}{2} + k, \frac{n_2 - m}{2}\right), \quad k \sim \text{BiNeg}\left(\frac{n_2 - 1}{2}, \frac{1}{1 + \theta}\right),$$

avec  $\theta = \frac{\rho^2}{1-\rho^2}$ . Basée sur l'a priori  $\theta \sim \text{Beta}_{II}(a, b, 1)$  et une observation  $r_1^2$ , la loi a posteriori  $\pi(\theta|r_1^2)$  admet la représentation (4.16) par le Lemme 4.19 avec

$$\pi_k(\theta) = \text{Beta}_{II}\left(a+k, \frac{n_1 - 1}{2} + b, 1\right) \text{ et } f_{r_1^2}^*(k) = \frac{\frac{(\frac{n_1-1}{2})_k (\frac{n_1-1}{2})_k (a)_k}{k! (a+b+\frac{n_1-1}{2})_k (\frac{m-1}{2})_k} (r_1^2)^k \mathbb{I}_{\mathbb{N}}(k)}{{}_3F_2\left(\frac{n_1-1}{2}, \frac{n_1-1}{2}, a; a+b+\frac{n_1-1}{2}, \frac{m-1}{2}; r_1^2\right)}.$$

De plus, par le Lemme 4.22, la densité prédictive bayésienne pour  $R_1^2$  admet la représentation (4.18), avec

$$\hat{q}_\pi(r_2; r_1) = \sum_{k^*} q_{r_1}^*(k^*) \text{Beta}\left(r_2^2 \left| \frac{m-1}{2} + k^*, \frac{n_2 - m}{2}\right.\right)$$

où  $q_{r_1}^*(k^*)$  est la fonction de masse de probabilité telle que

$$K^*|_{K=k} \sim \frac{\frac{(a+k)_{k^*} (\frac{n_2-1}{2})_{k^*}}{k^*! (a+b+\frac{n_1+n_2}{2}+k-1)_{k^*}}}{{}_2F_1\left(a+k, \frac{n_2-1}{2}; a+b+\frac{n_1+n_2}{2}+k-1; 1\right)} \mathbb{I}_{\mathbb{N}}(k^*)$$

et

$$K \sim \frac{\frac{(\frac{n_1-1}{2})_k (\frac{n_1-1}{2})_k (a)_k}{k! (a+b+\frac{n_1-1}{2})_k (\frac{m-1}{2})_k} (r_1^2)^{2k}}{{}_3F_2\left(\frac{n_1-1}{2}, \frac{n_1-1}{2}, a; a+b+\frac{n_1-1}{2}, \frac{m-1}{2}; r_1^2\right)} \mathbb{I}_{\mathbb{N}}(k).$$

on remarque que la loi de  $K^*$  est exprimé de façon commune comme un mélange, que les lois de  $K^*|_{K=k}$  et celle de  $K$  sont de type hypergéométrique généralisée<sup>2</sup> (Section 1.1.2) et que la densité prédictive pour  $R_2^2$  est un mélange de densité  $\text{Beta}_{II}$  (Remarque 4.23a).

Les Figures 4.3 et 4.4 présentent la loi a posteriori ainsi que la densité prédictive pour l'exemple 4.25, avec  $a = 3$ ,  $b = 5$ ,  $m = 3$ ,  $n_1 = 27$ ,  $n_2 = 31$  et quelques valeurs de  $r_1^2$ .

---

2. Marchand (2001) appelle ces distributions les lois HyperBeta.

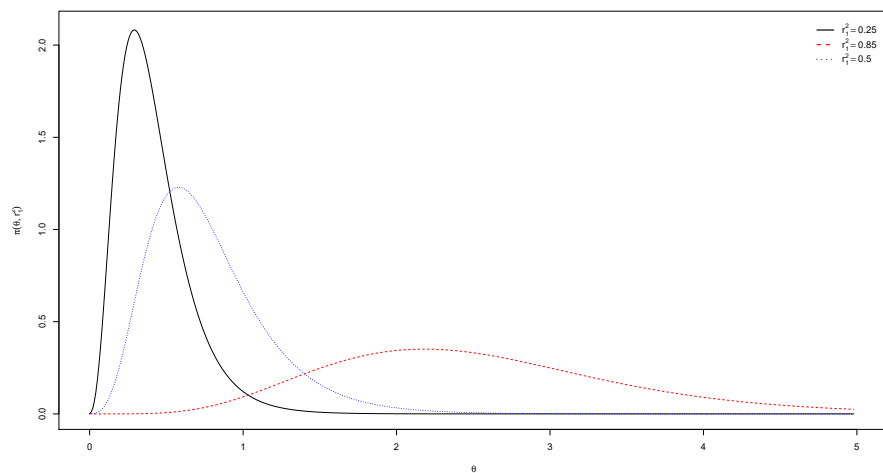


Figure 4.3 – La loi a posteriori  $\pi(\theta | r_1^2)$  pour plusieurs valeurs de  $r_1^2$ .

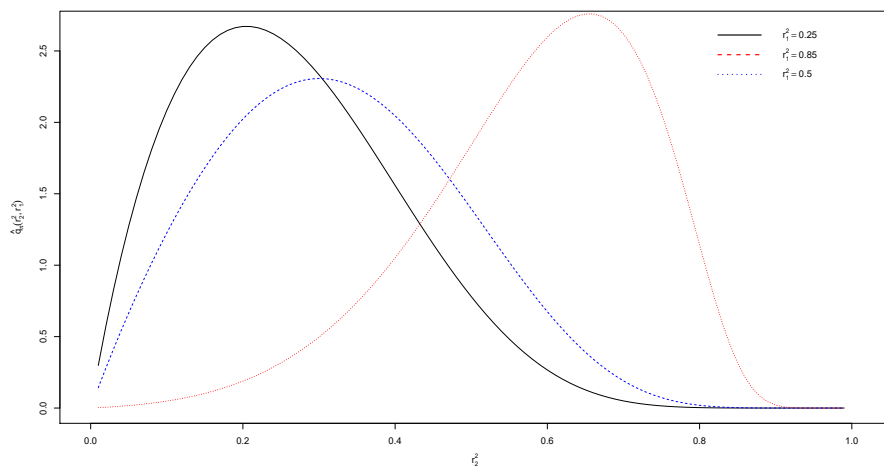


Figure 4.4 – La densité prédictive  $\hat{q}_\pi(r_2^2; r_1^2)$  pour plusieurs valeurs de  $r_1^2$ .

Si  $r_1^2 = 0.25$ , un intervalle bayésien de crédibilité 0.95 est donnée par  $I_\pi(\theta) = [0, 1.947]$ . Par ailleurs, à partir de la densité prédictive bayésienne pour  $R_2^2 | r_1^2 = 0.25$ , on peut évaluer la probabilité  $P(R_2^2 > c | r_1^2)$  où  $c$  est un seuil de signification pour un test de  $H_0 : \rho^2 = 0$ , par exemple pour  $c=0.2$ ,  $P(R_2^2 > 0.2 | r_1^2 = 0.25) = 0.626$ .

### 4.3.2 $X|\theta$ et $Y|\theta$ de type II

Considérons les modèles de mélange de la forme

$$X \sim p(x|\theta) = \sum_k p(k) f_{\theta,k}(x), \text{ et } Y \sim q(y|\theta) = \sum_k q(k) g_{\theta,k}(y) \text{ avec } x, y \in \mathbb{R}^d, \quad (4.23)$$

Le prochain résultat est général pour le modèle (4.23) et se démontre de façon similaire aux résultats de la section précédente.

**Lemme 4.26** *Pour  $X|\theta$  distribuée de la forme (4.23) et une densité a priori  $\pi$  absolument continue pour  $\theta$  par rapport à  $\sigma$ -mesure finie  $\nu$ , nous avons*

a) *la densité a posteriori de  $\theta$  est donnée par*

$$\pi(\theta|x) = \sum_k p^*(k) \pi_k(\theta), \quad (4.24)$$

où  $\pi_k(\theta) = \frac{\pi(\theta) f_{\theta,k}(x)}{\int_{\Theta} \pi(\theta) f_{\theta,k}(x) d\nu(\theta)}$ , et  $p^*(\cdot)$  une fonction de masse donnée par

$$p^*(k) = \frac{p(k) m_{\pi}(k)}{\sum_k p(k) m_{\pi}(k)}, \text{ avec } m_{\pi}(k) = \int_{\Theta} f_{\theta,k}(x) \pi(\theta) d\nu(\theta).$$

b) *la densité prédictive de  $Y$  en (4.23) sous la perte Kullback-Leibler est donnée par*

$$\hat{q}_{\pi}(y; x) = \sum_{k^*} q(k^*) g_{x,k^*}^*(y), \quad (4.25)$$

où  $g_{x,k^*}^*$  est la densité de probabilité

$$g_{x,k^*}^*(y) = \sum_k p^*(k) \hat{g}_{\pi,k,k^*}(y; x), \quad (4.26)$$

avec  $\hat{g}_{\pi,k,k^*}(y; x) = \int_{\Theta} g_{\theta,k^*}(y) \pi_k(\theta) d\nu(\theta)$ .

**Remarque 4.27** *Observons que*

a) *La distribution a posteriori est également un modèle de mélange des lois  $\pi_k$  et  $p^*(k)$  comme pour la distribution de  $X|\theta$  donné dans (4.23).*

b) La densité prédictive  $\hat{q}_\pi$  est un mélange du même type que la densité de  $Y$ , avec paramètre de mélange  $q$ , mais  $g_{\theta, k^*}$  est remplacé par  $g_{x, k^*}^*$ , qui est à son tour un mélange de  $p^*(k)$  et la densité prédictive  $\hat{g}_{\pi, k^*, k}$  de  $g_{\theta, k^*}$  basée sur l'a posteriori  $\theta|k \sim \pi_k$ . De plus, la densité prédictive  $\hat{q}_\pi(y; x)$  est de même type que  $q(y|\theta)$  quel que soit  $p(x|\theta)$  et la loi a priori.

Continuons avec une application du Lemme 4.26 à des mélanges de loi gamma multivarié, qui a reçu une attention particulière en hydrologie et qui est aussi fréquemment utilisé dans le cadre de la théorie de la fiabilité et de l'analyse de survie.

**Exemple 4.28** Soient  $X = (X_1, \dots, X_d)$  et  $Y = (Y_1, \dots, Y_d)$  des vecteurs aléatoires tels

que  $X_i|K=k \stackrel{ind}{\sim} Ga(\alpha_i+k; \frac{1}{\theta_i})$  et  $Y_i|K^*=k^* \stackrel{ind}{\sim} Ga(\beta_i+k; \frac{1}{\theta_i})$ ,  $i = 1, 2, \dots, d$ , avec les  $\alpha_i$  et  $\beta_i$

sont connus,  $K \sim p(\cdot)$  et  $K^* \sim q(\cdot)$  sont connus comme fonctions de masse sur  $\mathbb{N}$ . Supposons que  $\theta = (\theta_1, \dots, \theta_d)$  a comme loi a priori

$$\pi(\theta) = \prod_{i=1}^d \frac{\theta_i^{a_i-1} b^{a_i} e^{-b\theta_i}}{\Gamma(a_i)}.$$

Alors, la loi a posteriori est donnée par

$$\pi(\theta|x) = \sum_{k \geq 0} p^*(k) \prod_{i=1}^d Ga(\theta_i | a_i + \alpha_i + k; x_i + b_i),$$

où  $p^*(k)$  est la fonction de masse proportionnelle à :  $p(k) \prod_{i=1}^d \frac{(\alpha_i + a_i)_k}{(\alpha_i)_k} \left( \frac{x_i}{x_i + b_i} \right)^k$ . De plus, la densité prédictive est donnée par

$$\hat{q}_\pi(y; x) = \sum_{k \geq 0} \sum_{k^* \geq 0} p^*(k) q(k^*) \prod_{i=1}^d Beta_{II}(y_i | \beta_i + k^*, \alpha_i + a_i + k, x_i + b_i).$$

**Remarque 4.29** Si  $p(k) \sim Poisson(\lambda)$ , on obtient  $p^*(k)$  comme la fonction de masse d'une hypergéométrie généralisée  ${}_dF_d(\alpha_1 + a_1, \dots, \alpha_d + a_d; \alpha_1, \dots, \alpha_d; \frac{\lambda \prod_{i=1}^d x_i}{\prod_{i=1}^d (x_i + b_i)})$  définie en (1.1).

### 4.3.3 Inférence bayésienne pour le loi Kibble

Le cas de la loi Kibble avec paramètres inconnus  $\theta_1$  et  $\theta_2$  de lois Gamma indépendantes (a priori) est un cas particulier de l'Exemple précédent avec  $d = 2$  et  $p(k)$  la fonction de masse d'une loi binomiale négative (voir l'Exemple 4.4).

Dans cette section, nous modélisons des cas de dépendance pour  $\theta_1$  et  $\theta_2$  en étudiant les cas

$$(i) \quad \theta \sim Kibble \quad \text{et} \quad (ii) \quad \theta \sim Mackay,$$

cette dernière étant une loi a priori adaptée à une contrainte d'ordre  $\theta_2 \geq \theta_1$ .

#### Cas où $\pi(\theta)$ est de loi de Kibble

Les théorèmes suivants donnent la distribution a posteriori pour  $\theta$  et la densité prédictive sous la perte de Kullback-Leibler pour  $Y$  basé sur  $X$ , où

$$X|\theta \sim Kibble(\nu_x, \theta_1, \theta_2, \rho_x), \quad Y|\theta \sim Kibble(\nu_y, \theta_1, \theta_2, \rho_y). \quad (4.27)$$

**Théorème 4.30** *Soit  $X|\theta$  une variable aléatoire avec densité de la forme (4.27). Pour la loi a priori  $\theta \sim Kibble(\nu_\theta, a_1, a_2, \rho_\theta)$ , la loi a posteriori de  $\theta = (\theta_1, \theta_2)$  admet la représentation suivante*

$$\theta_i|x, \kappa^+ \stackrel{ind}{\sim} Ga\left(\kappa^+ + \nu_x + \nu_\theta, \frac{x_i}{1 - \rho_x^2} + \frac{a_i}{1 - \rho_\theta^2}\right), \quad i = 1, 2,$$

où  $\kappa^+ = \kappa + \kappa^*$  et  $(\kappa, \kappa^*)$  est un vecteur aléatoire sur  $\mathbb{N}^2$  dont la fonction de masse est donnée par

$$p(\kappa, \kappa^*) \propto \frac{\Gamma(\nu_x + \kappa) \Gamma(\nu_\theta + \kappa^*) \gamma_1^\kappa \gamma_1^{\kappa^*}}{\kappa! \kappa^*! B(\nu_x + \nu_\theta + \kappa + \kappa^*)}, \quad (4.28)$$

$$\text{avec } \gamma_1 = \left( \frac{\frac{x_1 x_2 \rho_x^2}{(1 - \rho_x^2)^2}}{\left(\frac{x_1}{1 - \rho_x^2} + \frac{a_1}{1 - \rho_\theta^2}\right) \left(\frac{x_2}{1 - \rho_x^2} + \frac{a_2}{1 - \rho_\theta^2}\right)} \right) \text{ et } \gamma_2 = \left( \frac{\frac{a_1 a_2 \rho_\theta^2}{(1 - \rho_\theta^2)^2}}{\left(\frac{x_1}{1 - \rho_x^2} + \frac{a_1}{1 - \rho_\theta^2}\right) \left(\frac{x_2}{1 - \rho_x^2} + \frac{a_2}{1 - \rho_\theta^2}\right)} \right).$$

**Démonstration.** On a

$$\begin{aligned} \pi(\theta|x) \propto \sum_{\kappa \geq 0} \sum_{\kappa^* \geq 0} BiNég(\kappa|\nu_x, \rho_x) \cdot BiNég(\kappa^*|\nu_\theta, \rho_\theta) Ga(x_1|\nu_x + k, \frac{1 - \rho_x^2}{\theta_1}) \cdot Ga(x_2|\nu_x + k, \frac{1 - \rho_x^2}{\theta_2}) \\ \cdot Ga(\theta_1|\nu_\theta + k, \frac{1 - \rho_\theta^2}{a_1}) \cdot Ga(\theta_2|\nu_\theta + k, \frac{1 - \rho_\theta^2}{a_2}). \end{aligned}$$



Or,  $Ga(x_i|\nu_x + k, \frac{1-\rho_x^2}{\theta_i}) \cdot Ga(\theta_i|\nu_\theta + k, \frac{1-\rho_\theta^2}{a_i}) \propto \frac{x_i^\kappa a_i^{\kappa^*} (\frac{x_i}{1-\rho_x^2} + \frac{a_i}{1-\rho_\theta^2})^{-(\kappa+\kappa^*)}}{B(\nu_x+\nu_\theta+\kappa+\kappa^*) (1-\rho_x^2)^\kappa (1-\rho_\theta^2)^{\kappa^*}} Ga(\theta_i|\nu_x + \nu_\theta + \kappa + \kappa^*, \frac{x_i}{1-\rho_x^2} + \frac{a_i}{1-\rho_\theta^2})$ .

Ainsi,  $BiN\acute{e}g(\kappa|\nu_x, \rho_x) \cdot BiN\acute{e}g(\kappa^*|\nu_\theta, \rho_\theta) \propto \frac{\Gamma(\nu_x+\kappa)\Gamma(\nu_\theta+\kappa^*)(\rho_x^2)^\kappa(\rho_\theta^2)^{\kappa^*}}{\kappa!\kappa^*!}$ .

Donc, on obtient alors

$$\begin{aligned} \pi(\theta|x) \propto & \sum_{\kappa \geq 0} \sum_{\kappa^* \geq 0} \frac{\Gamma(\nu_x + \kappa) \Gamma(\nu_\theta + \kappa^*)}{\kappa! \kappa^*! B(\nu_x + \nu_\theta + \kappa + \kappa^*)} \cdot \left( \frac{\frac{x_1 x_2 \rho_x^2}{(1-\rho_x^2)^2}}{\left(\frac{x_1}{1-\rho_x^2} + \frac{a_1}{1-\rho_\theta^2}\right) \left(\frac{x_2}{1-\rho_x^2} + \frac{a_2}{1-\rho_\theta^2}\right)} \right)^\kappa \\ & \cdot \left( \frac{\frac{a_1 a_2 \rho_\theta^2}{(1-\rho_\theta^2)^2}}{\left(\frac{x_1}{1-\rho_x^2} + \frac{a_1}{1-\rho_\theta^2}\right) \left(\frac{x_2}{1-\rho_x^2} + \frac{a_2}{1-\rho_\theta^2}\right)} \right)^{\kappa^*} \cdot \prod_{i=1}^d Ga\left(\theta_i|\nu_x+\nu_\theta+\kappa+\kappa^*, \frac{x_i}{1-\rho_x^2} + \frac{a_i}{1-\rho_\theta^2}\right), \end{aligned}$$

d'où le résultat.

**Définition 4.31** La première fonction hypergéométrique d'Appell (voir Wilfrid, 1935) est une fonction en double séries qui s'écrit sous la forme

$$F_1(a; b, c; d; \alpha, \beta) = \sum_{\kappa \geq 0} \sum_{\kappa^* \geq 0} \frac{(a)_{\kappa+\kappa^*} (b)_\kappa (c)_{\kappa^*} \alpha^\kappa \beta^{\kappa^*}}{(d)_{\kappa+\kappa^*} \kappa! \kappa^*!} \quad (4.29)$$

avec  $a, b, c, \alpha, \beta \in \mathbb{R}$  et  $d \in \mathbb{R}_+$ . Elle converge absolument pour  $|\alpha| < 1$  et  $|\beta| < 1$ .

**Remarque 4.32** La fonction de masse sur  $\mathbb{N}^2$  définie par l'équation (4.28), qu'on peut dénommer l'Hypergéométrique d'Appell, est de la forme

$$p(\kappa, \kappa^*) = \frac{\frac{(a)_{\kappa+\kappa^*} (b)_\kappa (c)_{\kappa^*} \alpha^\kappa \beta^{\kappa^*}}{(d)_{\kappa+\kappa^*} \kappa! \kappa^*!}}{F_1(a; b, c; d; \alpha, \beta)}, \quad (4.30)$$

avec  $a, b, c \geq 0$ ,  $d > 0$ ,  $|\alpha| < 1$ ,  $|\beta| < 1$ .

La fonction de masse donnée par (4.28) est de ce type avec  $a = 0$ ,  $b = \nu_x$ ,  $c = \nu_\theta$  et  $d = \nu_x + \nu_\theta$ ,

$$\alpha = \frac{\frac{x_1 x_2 \rho_x^2}{(1-\rho_x^2)^2}}{\left(\frac{x_1}{1-\rho_x^2} + \frac{a_1}{1-\rho_\theta^2}\right) \left(\frac{x_2}{1-\rho_x^2} + \frac{a_2}{1-\rho_\theta^2}\right)} \text{ et } \beta = \frac{\frac{a_1 a_2 \rho_\theta^2}{(1-\rho_\theta^2)^2}}{\left(\frac{x_1}{1-\rho_x^2} + \frac{a_1}{1-\rho_\theta^2}\right) \left(\frac{x_2}{1-\rho_x^2} + \frac{a_2}{1-\rho_\theta^2}\right)}.$$

Voici une application direct du Lemme 4.26 (démonstration omise) pour la densité prédictive d'une loi de Kibble.

**Théorème 4.33** *Sous les conditions du Théorème 4.30, la densité prédictive pour  $Y$  (avec densité de la forme (4.27)) est donnée par*

$$\hat{q}_\pi(y; x) = \sum_{\kappa} \sum_{\kappa^*} \sum_{\kappa^{**}} \frac{(\nu_x)_\kappa (\nu_\theta)_{\kappa^*} (\nu_y)_{\kappa^{**}} (1-\rho_y^2)}{(\nu_x + \nu_\theta)_{\kappa + \kappa^*}} \frac{\alpha_1^\kappa \alpha_2^{\kappa^*} (\rho_y^2)^{\kappa^{**}}}{\kappa! \kappa^*! \kappa^{**}!}$$

$$\text{Beta}_{II} \left( y_1 | \nu_y + \kappa^{**}, \nu_x + \nu_\theta + \kappa^* + \kappa, \left( \frac{a_1}{1-\rho_\theta^2} + \frac{x_1}{1-\rho_x^2} \right) (1-\rho_y^2) \right)$$

$$\text{Beta}_{II} \left( y_2 | \nu_y + \kappa^{**}, \nu_x + \nu_\theta + \kappa^* + \kappa, \left( \frac{a_2}{1-\rho_\theta^2} + \frac{x_2}{1-\rho_x^2} \right) (1-\rho_y^2) \right)$$

avec  $\alpha_1 = \frac{\frac{x_1 x_2 \rho_x^2}{(1-\rho_x^2)^2}}{\left( \frac{x_1}{1-\rho_x^2} + \frac{a_1}{1-\rho_\theta^2} \right) \left( \frac{x_2}{1-\rho_x^2} + \frac{a_2}{1-\rho_\theta^2} \right)}$ ,  $\alpha_2 = \frac{\frac{a_1 a_2 \rho_\theta^2}{(1-\rho_\theta^2)^2}}{\left( \frac{x_1}{1-\rho_x^2} + \frac{a_1}{1-\rho_\theta^2} \right) \left( \frac{x_2}{1-\rho_x^2} + \frac{a_2}{1-\rho_\theta^2} \right)}$ , et  $\text{Beta}_{II}$  est la densité (en  $y$ ) Bêta de type II.

C'est un mélange de loi Bêta de type II indépendants (a priori) et  $p(\kappa, \kappa^*, \kappa^{**})$  la fonction de masse d'une loi de type Hypergéométrique d'Appell et une loi binomiale négative indépendantes.

### Cas où $\pi(\theta)$ est de loi de MacKay

Supposons que les paramètres  $(\theta_1, \theta_2) \in \mathbb{R}_+^2$  obéissent la contrainte de la forme  $\theta_2 \geq \theta_1$ . La loi a priori la loi de MacKay de type Gamma bivariée est adaptée pour une telle contrainte d'ordre. Cette distribution a été introduite par MacKay (1934) pour caractériser la loi jointe du couple formée par la variance empirique d'un échantillon et la variance empirique associée à un sous-échantillon de variables gaussiennes centrées i.i.d.

**Définition 4.34** *La densité d'une loi gamma bivariée de MacKay est donnée par*

$$f(u, v) = \frac{c^{a+b}}{\Gamma(a)\Gamma(b)} u^{a-1} (v-u)^{b-1} e^{-cv} \mathbb{I}_{(0 < u < v)}(u, v) \quad (4.31)$$

où,  $c > 0$ ,  $a > 0$  et  $b > 0$ .

Les lois marginales sont  $U \sim Ga(a, \frac{1}{c})$  et  $V \sim Ga(a+b, \frac{1}{c})$ . On peut l'obtenir aussi à partir de deux variables aléatoires indépendantes  $X_1 \sim Ga(a, \frac{1}{c})$  et  $X_2 \sim Ga(b, \frac{1}{c})$ , ainsi  $(U, V) = (X_1, X_1 + X_2)$  admet la densité (4.31).

La définition suivante présente une généralisation de cette distribution.

**Définition 4.35** La loi de Mackay généralisée de paramètres  $a, b, c > 0$ ,  $f \geq 1$  et  $c > d \geq 0$ , notée  $MKG(a, b, c, d, f)$ , a pour densité

$$f_{U,V}(u, v) = Cu^{a-1}(v-u)^{b-1}v^{f-1}e^{-(du+cv)}\mathbb{I}_{(0 < u < v)}(u, v), \quad (4.32)$$

avec la constante de normalisation donnée par

$$C^{-1} = \frac{B(b, c)\Gamma(a+b+f-1)}{c^{a+b+f-1}} {}_2F_1(a, a+b+f-1; a+b; -\frac{d}{c}).$$

Quelques propriétés de loi de Mackay généralisée sont présentés en Annexe B.

**Corollaire 4.36** La loi de Mackay avec paramètres  $a, b, c$  notée  $MK(a, b, c)$  est un cas particulier de la loi  $MKG(a, b, c, d, f)$  avec  $d = 0$  et  $f = 1$ .

Le théorème suivant donne la distribution a posteriori pour  $\theta$  pour un modèle Kibble et une loi a priori de Mackay.

**Théorème 4.37** Soit  $X|\theta$  une variable aléatoire avec une densité de la forme (4.27). Pour la loi a priori  $\theta \sim MK(a, b, c)$ , la loi a posteriori de  $\theta = (\theta_1, \theta_2)$  admet la représentation suivante

$$\theta|_{X=x, K=\kappa} \sim MKG\left(a + \nu_x + \kappa, b, \left(c + \frac{x_2}{1 - \rho_x^2}\right), \frac{x_1}{1 - \rho_x^2}, \nu_x + \kappa + 1\right),$$

où  $\kappa|_{X=x}$  est une variable aléatoire sur  $\mathbb{N}$  dont la fonction de masse est proportionnelle à

$$\frac{\Gamma(a+b+2\nu_x+2\kappa)(x_1x_2\rho_x^2)^\kappa}{\kappa!(\nu_x)_\kappa(c(1-\rho_x^2)^2+x_2)^{2\kappa}} {}_2F_1(a+\nu_x+\kappa, a+b+2\nu_x+2\kappa; a+b+\nu_x+\kappa; -\frac{x_1}{c(1-\rho_x^2)+x_2}).$$

Si  $\rho_x = 0$  alors  $\theta|x$  a pour loi gamma bivariées de Mackay généralisée.

**Démonstration.** On a

$$\begin{aligned} \pi(\theta|x) &\propto \sum_{\kappa=0}^{\infty} \text{BiNég}(\kappa|\rho_x) \text{Ga}(x_1|\nu_x+k, \frac{1-\rho_x^2}{\theta_1}) \cdot \text{Ga}(x_2|\nu_x+k, \frac{1-\rho_x^2}{\theta_2}) \cdot MK(a, b, c) \\ &\propto \sum_{\kappa=0}^{\infty} \frac{1}{(\nu_x)_\kappa \kappa!} \left(\frac{x_1 x_2 \rho_x^2}{(1-\rho_x^2)^2}\right)^\kappa \theta_1^{a+\nu_x+\kappa-1} (\theta_2-\theta_1)^{b-1} \theta_2^{\nu_x+\kappa} e^{-\left(\frac{x_1\theta_1}{1-\rho_x^2} + (c+\frac{x_2}{1-\rho_x^2})\theta_2\right)}. \end{aligned}$$

$$\begin{aligned} &\propto \sum_{\kappa=0} \frac{\Gamma(a+b+2\nu_x+2\kappa)}{(\nu_x)_\kappa \kappa!} \left( \frac{x_1 x_2 \rho_x^2}{(c(1-\rho_x^2)^2 + x_2)^2} \right)^\kappa \\ &\quad {}_2F_1\left(a+\nu_x+\kappa, a+b+2\nu_x+2\kappa; a+b+\nu_x+\kappa; -\frac{x_1}{c(1-\rho_x^2)+x_2}\right) \\ &\quad MKG\left(a+\nu_x+\kappa, b, \left(c+\frac{x_2}{1-\rho_x^2}\right), \frac{x_1}{1-\rho_x^2}, \nu_x+\kappa+1\right), \end{aligned}$$

d'où le résultat.

Nous terminons cette section par la présentation de la densité prédictive associée à la loi *a priori* de Mackay.

**Théorème 4.38** *Soient  $X|\theta$  et  $Y|\theta$  des variables aléatoires de lois kibble données en (4.27), avec la loi a priori de MacKay  $MK(a, b, c)$ , alors*

$$\begin{aligned} \hat{q}_\pi(y; x) &\propto \sum_{\kappa \geq 0} \sum_{\kappa^* \geq 0} \frac{\Gamma(a+b+2\nu_x+2\nu_y+2\kappa+2\kappa^*) A(x_1, x_2, \rho_x^2)^\kappa}{\kappa! \kappa^* (\nu_x)_\kappa (\nu_y)_{\kappa^*}!} \frac{(y_1 y_2)^{\nu_y+\kappa^*-1}}{\left(c+\frac{x_1}{1-\rho_x^2}+\frac{y_2}{1-\rho_y^2}\right)^{a+b+2\nu_x+2\nu_y+2\kappa+2\kappa^*}} \\ &\quad {}_2F_1\left(a+\nu_x+\nu_y+\kappa+\kappa^*, a+b+2\nu_x+2\nu_y+2\kappa+2\kappa^*; a+b+\nu_x+\nu_y+\kappa+\kappa^*; -\frac{\left(\frac{x_1}{1-\rho_x^2}+\frac{y_1}{1-\rho_y^2}\right)}{\left(c+\frac{x_2}{1-\rho_x^2}+\frac{y_2}{1-\rho_y^2}\right)}\right), \end{aligned}$$

$$\text{avec } A(x_1, x_2, \rho_x^2) = \left( \frac{x_1 x_2 \rho_x^2}{(1-\rho_x^2)^2} \right).$$

**Démonstration.** La démonstration est laissée au lecteur.

## 4.4 Discussion

Nous avons étudié, en utilisant l'approche bayésienne, les lois a posteriori et la densité prédictive pour deux cas de lois mélanges (discrets et continues). On a obtenu de nouvelles structures de ces lois. De plus, le modèle proposé englobe plusieurs de famille de distributions.

Une proposition de recherche future serait une analyse comparative de la performance fréquentiste des distributions prédictives, y compris celles basées sur les méthodes bayésiennes et de type plug-in.

En effet, dans leur article, Kubokawa et coll. (2017) ont proposé les estimateurs  $\hat{\theta}$  de  $\theta$  pour le modèle (4.15). Par conséquent, il serait intéressant de comparer les densités prédictive  $q_{\hat{\theta}}$  avec densités bayésienne pour le cas de l'exemple 4.25.

# CONCLUSION

Le but poursuivi dans cette thèse est de traiter l'estimation de la densité prédictive et ses mesures d'efficacité basées sur la perte de type  $\alpha$ -divergence, notamment, le coût Kullback-Leibler. Notre travail est basé sur l'estimation bayésienne et de type plug-in.

Dans un premier temps, nous avons obtenu plusieurs résultats de dominance en exploitant des liens de dualité avec l'estimation ponctuelle et avec l'estimation de Stein pour des pertes concaves. Dans un second temps, nous avons pu améliorer les estimateurs de plug-in à l'aide de la méthode d'expansion d'échelle. Enfin, nous avons obtenu des formes closes de lois a posteriori et les densités prédictives bayésiennes applicables à divers types de mélanges.

Les modèles étudiés sont ceux avec ou sans restriction sur les paramètres. Ces modèles incluent la loi Gamma, la loi normale multivariée de structures de covariance connue, mélange de loi normale, les lois bivariées Gamma (loi de Kibble), les distributions  $\chi^2$  et  $F$  décentrées, la distribution du coefficient détermination expérimental  $R^2$ , etc.

De plus, les sujets et les méthodes développées dans cette thèse ont apporté des nouvelles idées de recherche sur l'estimation de la densité prédictive. Les techniques développées pourront s'ajuster pour traiter d'autres problèmes. Voici quelques-uns de ces problèmes et d'autres questions à considérer :

- Premièrement, étendre les résultats obtenus pour la loi Gamma avec le coût Kullback-Leibler à d'autres familles avec paramètre d'échelle dans la classe des pertes  $\alpha$ -divergence.
- Deuxièmement, extension des résultats obtenus au Chapitre 3 pour le modèle gaussien à des familles plus générales, la loi normale avec paramètre d'échelle inconnu, les lois à symétrie sphériques ou les familles exponentielles.
- Finalement, approfondir les résultats obtenus au Chapitre 4, notamment l'étude de l'efficacité des estimations de densités prédictives bayésienne et de type plug-in.

## 4.5 Annexe A

### La genèse de loi de Kibble

Voici la genèse de loi de Kibble à partir d'un échantillon tiré d'une loi normale bivariee.

Soient les couples  $(X_{1,i}; X_{2,i})$  des vecteurs aléatoires i.i.d. d'un échantillon de lois normales.

$$(X_{1,i}, X_{2,i}) \sim N_2 \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right), \text{ pour } i = 1, \dots, n, |\rho| < 1.$$

On note par  $W_{1,i} = \frac{(X_{1,i} - \mu_1)}{\sigma_1} \sim N(0, 1)$  et  $W_{2,i} = \frac{(X_{2,i} - \mu_2)}{\sigma_2} \sim N(0, 1)$ , on a bien que

$$(W_{1,i}, W_{2,i}) \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), \text{ pour } i = 1, \dots, n.$$

On cherche ici la loi conjointe de  $(T_1, T_2) = (\sum_i W_{1,i}^2, \sum_i W_{2,i}^2)$ .

Or,  $W_{1,i} \Big| w_{2,i} \sim N(\rho w_{2,i}, 1 - \rho^2)$ , donc  $W_{1,i} \Big| W_{2,1} = w_{2,1}, \dots, w_{2,n} \sim N(\rho w_{2,i}, 1 - \rho^2), \forall i$ .

Par conséquence,

$$W_{1,i}^2 \Big| w_{2,1}^2, \dots, w_{2,n}^2 \sim (1 - \rho^2) \chi_1^2 \left( \frac{\rho^2 w_{2,i}^2}{1 - \rho^2} \right), \text{ alors } \sum_i W_{1,i}^2 \Big| w_{2,1}^2, \dots, w_{2,n}^2 \sim (1 - \rho^2) \chi_n^2 \left( \frac{\rho^2 \sum_i w_{2,i}^2}{1 - \rho^2} \right),$$

ce qui implique que  $T_1 = \sum_i W_{1,i}^2 \Big| t_2 = \sum_i w_{2,i}^2 \sim (1 - \rho^2) \chi_n^2 \left( \frac{\rho^2 t_2}{1 - \rho^2} \right)$ , avec  $T_2 = \sum_i W_{2,i}^2 \sim \chi_n^2(0)$

En notant,  $f_{\nu, \lambda}$  la densité d'une loi  $\chi_\nu^2(\lambda)$  décentrée en  $\lambda$ , avec la décomposition suivante

$$Y \Big| K = k \sim \chi_{\nu+2k}^2(0), \quad K \sim \text{Poisson} \left( \frac{\lambda}{2} \right),$$

on obtient que la densité conjointe de  $(T_1, T_2)$  est donnée par

$$f_{(T_1, T_2)}(t_1, t_2) = \frac{1}{1 - \rho^2} f_{n, \left( \frac{\rho^2 t_2}{1 - \rho^2} \right)} \left( \frac{t_1}{1 - \rho^2} \right) f_{n, 0}(t_2)$$

$$\begin{aligned}
&= \frac{1}{1-\rho^2} \sum_k \left( \frac{\rho^2 t_2}{2(1-\rho^2)} \right)^k \frac{e^{-\frac{\rho^2 t_2}{2(1-\rho^2)}}}{k!} \frac{\left( \frac{t_1}{1-\rho^2} \right)^{\frac{n}{2}+k-1} e^{-\frac{t_1}{2(1-\rho^2)}}}{2^{\frac{n}{2}+k} \Gamma\left(\frac{n}{2}+k\right)} \frac{t_2^{\frac{n}{2}-1} e^{-t_2/2}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \\
&= \sum_k \frac{\Gamma\left(\frac{n}{2}+k\right) (1-\rho^2)^{\frac{n}{2}} (\rho^2)^k}{k! \Gamma\left(\frac{n}{2}\right)} \frac{t_1^{\frac{n}{2}+k-1} e^{-\frac{t_1}{2(1-\rho^2)}}}{(2(1-\rho^2))^{\frac{n}{2}+k} \Gamma\left(\frac{n}{2}+k\right)} \frac{t_2^{\frac{n}{2}+k-1} e^{-\frac{t_2}{2(1-\rho^2)}}}{(2(1-\rho^2))^{\frac{n}{2}+k} \Gamma\left(\frac{n}{2}+k\right)} \\
&= \sum_k \text{BiNég}\left(k \middle| \frac{n}{2}, 1-\rho^2\right) \text{Ga}\left(t_1 \middle| \frac{n}{2}+k, \frac{1}{2(1-\rho^2)}\right) \text{Ga}\left(t_2 \middle| \frac{n}{2}+k, \frac{1}{2(1-\rho^2)}\right).
\end{aligned}$$

On conclut alors que la loi conjointe de  $(T_1, T_2)$  est  $Kibble\left(\frac{n}{2}, \frac{1}{2}, \frac{1}{2}, \rho^2\right)$ .



## 4.6 Annexe B

### Propriétés de loi de Mackay généralisée

**Proposition 4.1** *Pour  $(U, V) \sim MKG(a, b, c, d, f)$ , on a :*

a) *La loi conjointe de  $Z = V - U$  et  $V$  est donnée par*

$$f_{Z,V}(z, v) = C(z + v)^{a-1} z^{b-1} e^{-(dz+(e+d)v)}; \quad z, v > 0.$$

*Si  $a = 1$  les variables aléatoires  $Z = V - U$  et  $V$  sont indépendantes de  $Ga(b, \frac{1}{d})$  et  $Ga(1, \frac{1}{c+d})$  respectivement.*

b) *La densité conditionnelle de  $Z = V - U$  étant donnée  $V = v$  est donnée par*

$$f_v(z) \propto (z + v)^{a-1} z^{b-1} e^{-dz}; \quad z' > 0,$$

*il s'agit de la loi Gamma de type II de paramètres  $(b, d, v, a)$  qui a été introduite au Définition 1.10 de Chapitre 1.*

c) *Si  $b = 1$ , les variables aléatoires  $U$  et  $V$  sont indépendantes de  $Ga(a, \frac{1}{d})$  et  $Ga(f, \frac{1}{c})$  respectivement.*

d) *La densité conditionnelle de  $T = \frac{V}{U}$  étant donnée  $U = u$  est égale à*

$$T|u \sim \frac{t^{c-1}(1-t)^{b-1}e^{-fut}}{B(c, b) {}_1F_1(c, c+b; eu)} \mathbb{I}_{(0,1)}(t).$$

*De plus, si  $f = 0$ , alors  $T|u \sim \text{Beta}(c, b)$ .*

e) *La densité marginale de  $U$  est  $\propto u^{a+b+f-2} e^{-du} {}_1F_1(b; f+b; cu) \mathbb{I}_{\mathbb{R}_+}(u)$ .*

f) *La densité marginale de  $V$  est  $\propto v^{a+b+f-2} e^{-cv} {}_1F_1(a; a+b; -dv) \mathbb{I}_{\mathbb{R}_+}(v)$ .*

# Bibliographie

- [1] Abramowitz, M., and I .A. Stegun, eds. (1964). Handbook of Mathematical Functions, U.S.National Bureau of Standards, Dover, New York (1965).
- [2] Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, 62, 547-554.
- [3] Amaral, M.A. and Dunsmore, I.R. (1980). Optimal estimates of predictive distributions. *Biometrika*, 67, 685-689.
- [4] Andrews, G.E., Askey, R. and Roy, R. (1999). Special functions. Encyclopedia of Mathematics and its Applications, 71, Cambridge University Press.
- [5] Balakrishnan N., C. Diew Lai. (2009). Continuous Bivariate Distributions. *Springer-Verlag New York Inc. ; Édition : 2nd ed.*
- [6] Baker, M. (2016). 1500 scientists lift the lid on reproducibility. *Nature*, 533, 452-454.
- [7] Barberis, N. (2000). Investing for the Long Run when Returns are Predictable. *Journal of Finance*, Vol 55, 225–64.
- [8] Berger J. O. (1985). Decision Theory and Bayesian Analysis. *Springer*, New York.
- [9] Brown, L.D., George, E.I., and Xu, X. (2008). Admissible predictive density estimation. *Annals of Statistics*, 36, 1156-1170.
- [10] Brown, L.D., Johnstone, I.M., and MacGibbon, K.B. (1981). Variation diminishing transformations : A direct approach to total positivity and its statistical applications. *Journal of the American Statistical Association*, 76, 824-832.

- [11] Bulgren, W. G. On Representations of the Doubly Non-Central F Distribution. *J. Amer. Stat. Assoc.* 66, 184, 1971.
- [12] Csiszár, I. (1963). Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar. Tud. Akad. Mat. Kutato Int. Kozl.* 8 : 85–108.
- [13] Corcuera J. M. and F. Giummolè (1999). A generalized Bayes rule for prediction. *Scand. J. Statist.* 26 no. 2, 265-279.
- [14] Dasgupta, A. and Lahiri, S.N. (2012). Density estimation in high and ultra dimensions, regularization, and the  $L_1$  asymptotics. *Contemporary Developments in Bayesian analysis and Statistical Decision Theory : A Festschrift for William E. Strawderman*, IMS Collections, 8, 1-23.
- [15] Dempster, A. P. Laird, N. M. Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. With discussion. *J. Roy. Statist. Soc. Ser. B* 39, no. 1, 1–38.
- [16] Eaton, M.L. (1989). Group Invariance Applications in Statistics. *Regional conference series in Probability and Statistics*, vol. 1, Institute of Mathematical Statistics and the American Statistical Association.
- [17] 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.
- [18] Fourdrinier, D., Strawderman, W.E., and Wells, M. (1998). On the construction of Bayes minimax estimators. *Annals of Statistics*, 26, 660-671.
- [19] Geisser, S. (1971). The inferential use of predictive distributions. In *Foundations of statistical inference* (eds V. P. Godambe and D. A. Sprott), 456–469. Holt, Rinehart, and Winston, Toronto. MR0381054.
- [20] Geisser. S. (1993). Predictive Inference. *CRC Press*. vol. 55.
- [21] George, E. I., Liang, F. and Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Annals of Statistics*, 34, 78-91.

- [22] 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.
- [23] Hartigan, J. (2004). Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, 67, 285-288.
- [24] Iliopoulos, G., Karlis, D. & Ntzoufras, I. (2005). Bayesian estimation in Kibble's bivariate gamma distribution. *Canadian Journal of Statistics*, 33, 571-589.
- [25] Jones M.C, Marchand, É and Strawderman, W.E. (2018). On an intriguing distributional identity. *The American Statistician*. pp. 1-6.
- [26] Kemp, A. (1968). A Wide Class of Discrete Distributions and the Associated Differential Equations. *Sankhya : Series A*, 401-410.
- [27] Khintchine A. (1937). A new derivation of one formula by Lévy P. *Bull. Moscow State Univ.*, 1 :1-5.
- [28] Kibble, W. F. (1941). A two-variate gamma-type distribution. *Sankhya*, 5 :137- 150.
- [29] Kotz, S. , N. Balakrishnan and N.L. Johnson, (2000). Continuous Multivariate Distributions, *Volume 1 : Models and Applications*, 2nd edition, John Wiley and Sons, New York.
- [30] Kubokawa, T. (1994). A unified approach to improving equivariant estimators. *Annals of Statistics*, 22, 290-299.
- [31] Kubokawa, T. and Saleh, A.K.M.D.E. (1998). Estimation of location and scale parameters under order restrictions. *Journal of Statistical Research*, 28, 41-51.
- [32] Kubokawa, T., Marchand, É and Strawderman, W.E. (2017). On predictive density estimation for location families under integrated absolute error loss. *Bernoulli*, Volume 23, Number 4B, 3197-3212.
- [33] Kubokawa, T., Marchand, É. and Strawderman, W.E. (2014). On improved shrinkage estimators under concave loss. *Statistics & Probability Letters*.

- [34] Kubokawa, T., Marchand, É., Strawderman, W.E. and Turcotte, J-P. (2013). Minimaxity in predictive density estimation with parametric constraints. *Journal of Multivariate Analysis*, 116, 382-397.
- [35] Kubokawa, T., Marchand, É & Strawderman, W.E. (2017). A unified approach to estimation of noncentrality parameters, the multiple correlation coefficient, and mixture models. *Mathematical Methods of Statistics*, 26, 134-148.
- [36] Landriault, D; Moutanabbir, K; Willmot, G. E A, (2015). Note on order statistics in the mixed Erlang case. *Statist. Probab. Lett.* 13–18.
- [37] Lawless, J. F. Fredette, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika*. 92, no. 3, 529–542.
- [38] Lehmann, E. L., Casella, G. (1999). Theory of Point Estimation. *Springer-Verlag*, New York, 2nd edition.
- [39] Liang, F. and Barron, A. (2004). Exact minimax strategies for predictive density estimation, data compression, and model selection. *IEEE Trans. Inform. Theory*. 50, no. 11, 2708–2726.
- [40] A. T. McKay. (1934). Sampling from batches. *Supplement to the Journal of the Royal Statistical Society*, 207–216.
- [41] 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.
- [42] 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.
- [43] Marchand, É., and Strawderman, W. E. (2005). 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.
- [44] Marchand, É., Point Estimation of the Coefficient of Determination. (2001). *Statistics and Decisions*, 19, 137-154.
- [45] Marchand, É., and Strawderman, W.E. (2012). A unified minimax result for restricted parameter spaces. *Bernoulli*, 18 635-643.

- [46] Marchand, Éric, (2001). Point estimation of the coefficient of determination. *Statist. Decisions* 19 , no. 2, 137–154.
- [47] Misra, N. and Arshad, M. (2014). Monotonicity of certain integrals involving gamma distributions and their applications in multiple comparisons. *Statistics & Probability Letters*, 85, 144-152.
- [48] Muldoon, M.E. (1978). Some monotonicity properties and characterizations of the gamma function. *Aequationes Math.* 18, 54-63.
- [49] Nadarajah, Saralees ; Kotz, (2006). Samuel Bivariate gamma distributions, sums and ratios. *Bull. Braz. Math. Soc. (N.S.)* 37, no. 2, 241–274.
- [50] Ng, K. W. and Kotz, S. (1995). Kummer-Gamma and Kummer-Beta univariate and multivariate distributions. *Research report*, Department of Statistics, The University of Hong Kong, Hong Kong.
- [51] Ong, S. H. ; Lee, P. A, (1979). The noncentral negative binomial distribution. *Biometrical J.* no. 7, 611–627.
- [52] Pham-Gia, T. and Turkkan, N. (2002). Operations on Generalized-F variables and applications. *Statistics*, 36 195-209.
- [53] Rainville, E. D. (1960). Special functions. *The Macmillan Co.*, New York.
- [54] Shao, J. (1999). Mathematical statistics. *Springer Texts in Statistics. Springer-Verlag*, New York.
- [55] Robert, C. P. (1996). Intrinsic loss functions. *Theory and Decision*, 40 :192–214.
- [56] Spiring, Fred A.(1993). The reflected normal loss function. *Canad. J. Statist.* 21 , no. 3, 321–330.
- [57] Tanaka, F. (2006). Generalized Bayesian predictive density operator. *In : Proceedings of 14th Quantum Information Technology Symposium (QIT14), Tokyo, Japan, May 29–30, pp. 107–110.*
- [58] Tanaka, F. and Komaki, F. (2005). Bayesian predictive density operators for exchangeable quantum-statistical models. *Phys. Rev. A* 71, 052323.

- [59] Tijms, H.C., (1994). Stochastic Models : An Algorithmic Approach. *John Wiley, Chichester*.
- [60] Tiwari, R., Chib, S. and Jammalamadaka, S.Ra. (1989). Bayes estimation of the multiple correlation coefficient. *Comm. Statist. Theory Methods*, 18 , no. 4, 1401-1413.
- [61] van Eeden, C. (2006). Restricted parameter space problems : Admissibility and minimaxity properties. *Lecture Notes in Statistics*, 188, Springer.
- [62] van Zwet W. R. (1964). Convex transformations of random variables. *Mathematical Centre Tracts, 7 Mathematisch Centrum, Amsterdam vi+116 pp*
- [63] Wicksell S. D. (1933). On correlation functions of type III. *Biometrika*, 25 :121-133.
- [64] Wilfrid N B. (1935). Generalized Hypergeometric Series. *Cambridge University Press*.
- [65] Yanagimoto, T., Ohnishi, T. (2009). Bayesian prediction of a density function in terms of e-mixture. *Journal of Statistical Planning and Inference*, 139, 3064–3075.