

On improved predictive density estimation with parametric constraints

Dominique Fourdrinier*

LITIS, EA 4108, Université de Rouen,
Avenue de l'Université, BP 12,
76801 Saint-Étienne-du-Rouvray, France
e-mail: Dominique.Fourdrinier@univ-rouen.fr

Éric Marchand†

Département de mathématiques, Université de Sherbrooke,
Sherbrooke, QC, CANADA, J1K 2R1
e-mail: Eric.Marchand@USherbrooke.ca

Ali Righi

LMRS, UMR 6085, Université de Rouen,
Avenue de l'Université, BP 12,
76801 Saint-Étienne-du-Rouvray, France
e-mail: ali.righi@etu.univ-rouen.fr

and

William E. Strawderman

Department of Statistics, Rutgers University,
561 Hill Center, Busch Campus, Piscataway, N.J., USA 08854-8019
e-mail: straw@stat.rutgers.edu

Abstract: We consider the problem of predictive density estimation for normal models under Kullback-Leibler loss (KL loss) when the parameter space is constrained to a convex set. More particularly, we assume that $X \sim \mathcal{N}_p(\mu, v_x I)$ is observed and that we wish to estimate the density of $Y \sim \mathcal{N}_p(\mu, v_y I)$ under KL loss when μ is restricted to the convex set $C \subset \mathbb{R}^p$. We show that the best unrestricted invariant predictive density estimator \hat{p}_U is dominated by the Bayes estimator \hat{p}_{π_C} associated to the uniform prior π_C on C . We also study so called plug-in estimators, giving conditions under which domination of one estimator of the mean vector μ over another under the usual quadratic loss, translates into a domination result for certain corresponding plug-in density estimators under KL loss. Risk comparisons and domination results are also made for comparisons of plug-in estimators and Bayes predictive density estimators. Additionally, minimaxity and domination results are given for the cases where: (i) C is a cone, and (ii) C is a ball.

AMS 2000 subject classifications: Primary 62C15, 62C20, 62F10, 62H12.

Keywords and phrases: Predictive estimation, risk function, quadratic loss, Kullback-Leibler loss, uniform priors, Bayes estimators, convex sets, cones, multivariate normal.

Received November 2010.

*The support of the ANR grant 08-EMER-002 is gratefully acknowledged.

†The research support of NSERC of Canada is gratefully acknowledged.

Contents

1 Introduction 173
 2 Context and preliminary results 175
 3 Applications for restricted parameter spaces 177
 3.1 Improving on \hat{p}_U 177
 3.2 Improving on a maximum likelihood estimator 179
 3.3 Bayesian improvements on the plug-in $N_p(x, v_y I_p)$ estimator under a constraint 182
 4 Minimax results 182
 4.1 Minimaxity when the mean is restricted to a ball 183
 4.2 Minimaxity of \hat{p}_U when the mean is restricted to a cone 184
 5 Plug-in estimators: Some additional considerations 186
 6 Conclusion 188
 A Appendix 188
 Acknowledgements 190
 References 190

1. Introduction

We consider the problem of predictive density estimation for Gaussian models under Kullback-Leibler loss (KL loss) when the parameter space is constrained to a convex set. More precisely, let $X|\mu \sim \mathcal{N}_p(\mu, v_x I)$ and $Y|\mu \sim \mathcal{N}_p(\mu, v_y I)$ be two independent random vectors having a normal distribution, with common unknown mean μ that we assume to be restricted to a convex set $C \subset \mathbb{R}^p$. The scale parameters v_x and v_y are assumed to be known and we denote by $p(x|\mu, v_x)$ and $p(y|\mu, v_y)$ the conditional densities of X and Y .

Under the above restriction, we seek to determine efficient predictive density estimators $\hat{p}(y|x)$ of the density $p(y|\mu, v_y)$, based on observing only $X = x$, relative to the Kullback-Leibler loss

$$L(\mu, \hat{p}(y|x)) = \int_{\mathbb{R}^p} p(y|\mu, v_y) \log \frac{p(y|\mu, v_y)}{\hat{p}(y|x)} dy \tag{1.1}$$

and the associated Kullback-Leibler(KL) risk

$$\mathcal{R}_{\text{KL}}(\mu, \hat{p}) = \int_{\mathbb{R}^p} p(x|\mu, v_x) L(\mu, \hat{p}(y|x)) dx. \tag{1.2}$$

This model was considered by George, Liang and Xu [6] when the mean μ is unrestricted, that is, $\mu \in \mathbb{R}^p$. The reference density is the generalized Bayes predictive density $\hat{p}_U(y|x)$ associated to the noninformative prior on \mathbb{R}^p , $\pi(\mu) \equiv 1$. Its expression may be derived from a more general result due to Aitchison (see [1]), as the conditional density of Y given $X = x$ associated with prior measure π , and given by

$$p(y|x) = \frac{\int_{\mathbb{R}^p} p(y|\mu, v_y) p(x|\mu, v_x) \pi(\mu) d\mu}{\int_{\mathbb{R}^p} p(x|\mu, v_x) \pi(\mu) d\mu}. \tag{1.3}$$

Komaki [10] noticed that expression (1.3) with $\pi(\mu) \equiv 1$ reduces to

$$\hat{p}_U(y|x) = \frac{1}{[2\pi(v_x + v_y)]^{p/2}} \exp \left[-\frac{1}{2} \frac{\|y - x\|^2}{v_x + v_y} \right], \quad (1.4)$$

and Murray [17] showed that \hat{p}_U is best invariant with constant risk, under translations and non singular linear transformations. For a location family, Ng [18] extended this invariance property and Liang and Barron [12] proved that \hat{p}_U is minimax.

In the normal case, George, Liang and Xu [6] gave a simple direct proof of the minimaxity of \hat{p}_U in (1.4) and showed, among other results, that \hat{p}_U can be improved by Bayes predictive densities \hat{p}_π under superharmonic priors π and for $p \geq 3$, thus adding to previous findings due to Komaki [10]. We refer to a recent review of Bayesian predictive estimation by George and Xu [7] for further exposition and description of recent research in this area.

We will make use of the following key representation for Bayesian estimators given by George, Liang and Xu. Hereafter, we let $W = (v_y X + v_x Y)/(v_x + v_y)$, $v_w = (v_x v_y)/(v_x + v_y)$, and we let $m_\pi(W; v_w)$ and $m_\pi(X; v_x)$ be the marginal densities of W and X respectively under prior π .

Lemma 1.1. ([6] Lemma 2)

We have

$$\hat{p}_\pi(y|X) = \frac{m_\pi(W; v_w)}{m_\pi(X; v_x)} \hat{p}_U(y|X), \quad (1.5)$$

where $\hat{p}_U(\cdot|X)$ is the Bayes estimator associated with the uniform prior on \mathbb{R}^p given by (1.4).

It is also shown that, for any prior π , the difference between the KL risks of $\hat{p}_U(\cdot|x)$ and the Bayesian predictive density $\hat{p}_\pi(\cdot|x)$ is given by

$$\mathcal{R}_{\text{KL}}(\mu, \hat{p}_U) - \mathcal{R}_{\text{KL}}(\mu, \hat{p}_\pi) = E_{\mu, v_w} [\log m_\pi(W; v_w)] - E_{\mu, v_x} [\log m_\pi(X; v_x)] \quad (1.6)$$

where $E_{\mu, v}$ stands for the expectation with respect to the normal distribution $\mathcal{N}_p(\mu, vI)$.

George, Liang and Xu underlined the fact that there exists a parallel between this predictive density estimation problem and the estimation of the mean vector μ under quadratic loss $\|\hat{\mu} - \mu\|^2$ giving rise to the quadratic risk

$$\mathcal{R}_Q^v(\mu, \hat{\mu}) = E_{\mu, v} [\|\hat{\mu} - \mu\|^2]. \quad (1.7)$$

More precisely, they show that the predictive density \hat{p}_U plays a similar role as the standard estimator X of μ which is best invariant and minimax under the quadratic risk (1.7), but inadmissible for $p > 2$.

Our findings here involve the elaboration and the use of similar connections between the risks \mathcal{R}_Q^v and \mathcal{R}_{KL} to draw inferences regarding domination and minimaxity for predictive density estimation problems when the mean μ is restricted to a convex set C . These findings parallel, and rely on several

results applicable to estimating a bounded multivariate mean with risk R_Q ([2, 5, 8, 14, 15, 20], among others). In Section 2, using a result of Brown, George and Xu [3], our first result formalizes a link between the two risks \mathcal{R}_{KL} and \mathcal{R}_Q^v , the quantities involving \mathcal{R}_{KL} being expressed as integrated quantities involving \mathcal{R}_Q^v . This link is used to express the risk differences between Bayes and plug-in estimators and between Bayes estimators as well.

In Section 3, various applications are given for restricted parameter spaces C . First, referring to Hartigan [8] who showed that the Bayes estimator of μ with respect to the uniform prior π_C on a convex set with a non-empty interior C dominates X under the quadratic risk (1.7), we obtain, via two different paths, a similar result for the domination of \hat{p}_{π_C} over \hat{p}_U . We also show that our proof of dominance for KL-loss implies dominance for quadratic, thus providing an alternative proof of Hartigan’s result. Secondly, we turn our attention to plug-in densities. When an estimator δ_1 for $\mu \in C$ is dominated by a Bayes estimator $\hat{\mu}_{\pi,v}$, associated to a prior π , and to a scale factor v , we give conditions under which the Bayes predictive density estimator \hat{p}_π dominates the plug-in density estimator $\hat{p}_1(X) \sim \mathcal{N}_p(\delta_1(X), v_y I_p)$. In the case when $p = 1$, we apply these results to obtain improvements on the plug-in maximum likelihood estimator $\hat{p}_{mle}(X) \sim \mathcal{N}_p(\delta_{mle}(X), v_y I_p)$.

In Section 4, we deal with the minimaxity of Bayes predictive density estimators when μ is restricted to a ball or to a cone. As a specific result, we show that, when $\|\mu\| \leq m$, the boundary uniform Bayes estimator is minimax for risk \mathcal{R}_{KL} whenever $m \leq c_0(p)\sqrt{v_w}$ where $c_0(p)$ is the constant given by Berry [2] and Marchand and Perron [15]. When μ belongs to a convex cone C with non-empty interior, we prove that the unrestricted predictive density estimator \hat{p}_U in (1.4) remains minimax (as it is when no restriction is assumed) for risk \mathcal{R}_{KL} when $\mu \in C$. This finding parallels the result of Tsukuma and Kubokawa [20] who established that X is still minimax for estimating μ under the restriction to a polyhedral cone.

In Section 5 we expand on some additional considerations concerning plug-in estimators. Section 6 contains concluding remarks, and Section A is an appendix with details on some of the proofs.

2. Context and preliminary results

In this section, we expand upon a link between estimation under risks \mathcal{R}_Q^v and risk \mathcal{R}_{KL} . The following lemma and theorem concern plug-in estimators $\hat{p}_1 \sim \mathcal{N}_p(\delta_1(X), v_y I_p)$ and Bayesian estimators $\hat{p}_\pi(\cdot|X)$. Theorem 2.1 provides a useful expression for a \mathcal{R}_{KL} risk difference in terms of integrated \mathcal{R}_Q^v risk differences.

Lemma 2.1. *For a prior π , and a plug-in estimator $\hat{p}_1 \sim \mathcal{N}_p(\delta_1(X), v_y I_p)$, we have*

$$(a) \quad E^{X,Y} \left[\log \left(\frac{m_\pi(W; v_w)}{m_\pi(X; v_x)} \right) \right] = \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (\mathcal{R}_Q^v(\mu, X) - \mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi,v})) dv$$

and

$$(b) E^{X,Y} \left[\log \left(\frac{\hat{p}_U(Y|X)}{\hat{p}_1(Y|X)} \right) \right] = \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} \left(\mathcal{R}_Q^{v_x}(\mu, \delta_1) - \mathcal{R}_Q^v(\mu, X) \right) dv,$$

where \hat{p}_U is, as above, the generalized Bayes estimator with respect to the flat prior on \mathbb{R}^p .

Proof. Part (a) is taken from Brown, George, and Xu ([3], Theorem 1). For part (b), since

$$\int_{v_w}^{v_x} \frac{1}{v^2} \mathcal{R}_Q^{v_x}(\mu, \delta_1) dv = \mathcal{R}_Q^{v_x}(\mu, \delta_1)(1/v_w - 1/v_x) = \mathcal{R}_Q^{v_x}(\mu, \delta_1)/v_y,$$

and

$$\int_{v_w}^{v_x} \frac{1}{v^2} \mathcal{R}_Q^v(\mu, X) dv = p \int_{v_w}^{v_x} \frac{1}{v} dv = p \log((v_x + v_y)/v_y),$$

we need to show that

$$E^{X,Y} \left[\log \left(\frac{\hat{p}_U(Y|X)}{\hat{p}_1(Y|X)} \right) \right] = -\frac{p}{2} \log \left(\frac{v_x + v_y}{v_y} \right) + \frac{\mathcal{R}_Q^{v_x}(\mu, \delta_1)}{2v_y}. \quad (2.1)$$

A direct expansion yields

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

which indeed matches (2.1). \square

Theorem 2.1. For a plug-in estimator $\hat{p}_1 \sim \mathcal{N}_p(\delta_1(X), v_y I_p)$ and for Bayes estimators \hat{p}_π and $\hat{p}_{\pi'}$, we have

$$(a) \mathcal{R}_{KL}(\mu, \hat{p}_1) - \mathcal{R}_{KL}(\mu, \hat{p}_\pi) = \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} \left(\mathcal{R}_Q^{v_x}(\mu, \delta_1) - \mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi,v}) \right) dv$$

and

$$(b) \mathcal{R}_{KL}(\mu, \hat{p}_{\pi'}) - \mathcal{R}_{KL}(\mu, \hat{p}_\pi) = \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} \left(\mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi',v}) - \mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi,v}) \right) dv.$$

Proof. Part (b) follows immediately from part (a) (or, alternatively, from (1.6) and part (a) of Lemma 2.1). From the definition of the risk \mathcal{R}_{KL} , we have for (a):

$$\mathcal{R}_{KL}(\mu, \hat{p}_1) - \mathcal{R}_{KL}(\mu, \hat{p}_\pi) = E^{X,Y} \left[\log \left(\frac{\hat{p}_\pi(Y|X)}{\hat{p}_1(Y|X)} \right) \right].$$

Upon applying (1.5), we obtain equivalently

$$\begin{aligned} \mathcal{R}_{\text{KL}}(\mu, \hat{p}_1) - \mathcal{R}_{\text{KL}}(\mu, \hat{p}_\pi) &= E^{X,Y} \left[\log \left(\frac{m_\pi(W; v_w)}{m_\pi(X; v_x)} \right) \right] \\ &\quad + E^{X,Y} \left[\log \left(\frac{\hat{p}_U(Y|X)}{\hat{p}_1(Y|X)} \right) \right] \end{aligned}$$

and the result now follows from Lemma 2.1. \square

Note that Corollary 1 of Brown *et al.* [3] is an integrated version of (b).

We will make use of Theorem 2.1 to obtain dominance results applicable for risk \mathcal{R}_{KL} by working inside the above integrals and relying on the associated comparisons for the risks \mathcal{R}_Q^v , $v \in (v_w, v_x)$. As an illustration, take $\delta_1(X) = X$, π to be the flat prior on \mathbb{R}^p . We have $\mathcal{R}_Q^{v_x}(\mu, \delta_1) = pv_x$, $\hat{\mu}_{\pi,v}(X) = X$, $\mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi,v}) = pv$, and, from part (a) of Theorem 2.1:

$$\begin{aligned} \mathcal{R}_{\text{KL}}(\mu, \hat{p}_1) - \mathcal{R}_{\text{KL}}(\mu, \hat{p}_U) &= \frac{p}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (v_x - v) dv \\ &= \frac{p}{2} \left(\frac{v_x}{v_w} - \log \left(\frac{v_x}{v_w} \right) - 1 \right), \end{aligned}$$

which is positive. The finding that \hat{p}_U dominates \hat{p}_1 is not new of course (e.g., Aitchison [1]), but the objective here was rather to illustrate Theorem 2.1 above.

Finally, we expand on some definitions and notations with respect to convex sets and cones in \mathbb{R}^p . A subset C of \mathbb{R}^p will be called a (positively homogeneous) cone if it is closed under positive scalar multiplication, i.e. $\alpha x \in C$ when $x \in C$ and $\alpha > 0$ (e.g., [19][]). In the above, C is a cone with vertex the origin. More generally, for such a C and for any $g \in \mathbb{R}^p$, the set $C_g = C + g$ is an affine cone with vertex g , where we adopt hereafter the notation: for $\alpha \in \mathbb{R}$, $\theta \in \mathbb{R}^p$ and $A \subset \mathbb{R}^p$, $\alpha A + \theta = \{\alpha a + \theta \mid a \in A\}$.

3. Applications for restricted parameter spaces

3.1. Improving on \hat{p}_U

For the problem of estimating, under risk $\mathcal{R}_Q^{v_x}$, with μ restricted to a convex subset C of \mathbb{R}^p with a non-empty interior, but otherwise arbitrary, Hartigan [8] showed quite generally that the Bayes estimator with respect to a uniform prior on C dominates the estimator X . We obtain an analogous result here for the predictive density estimation problem with risk \mathcal{R}_{KL} . A first proof follows from Hartigan's result and Theorem 2.1. A second proof with a much different flavor circumvents the explicit use of Hartigan's result and Theorem 2.1 and follows, surprisingly, a more direct route to establishing the result. The second part of this subsection parallels a result by Kubokawa [11] in providing a class of priors for which we obtain Bayesian improvements to \hat{p}_U in the univariate case with μ bounded to an interval.

Theorem 3.1. *Let $C \subset \mathbb{R}^p$ be convex set with a non-empty interior, and let $\pi_C = \mathbb{1}_C(\mu)$ be the noninformative prior restricted to C . Then \hat{p}_{π_C} dominates \hat{p}_U in (1.4) for risk \mathcal{R}_{KL} and $\mu \in C$.*

First proof. We apply part (b) of Theorem 2.1 with $\hat{p}_{\pi'} \equiv \hat{p}_U$, $\hat{p}_{\pi} \equiv \hat{p}_{\pi_C}$, $\hat{\mu}_{\pi',v}(X) = X$ for all v , and $\hat{\mu}_{\pi,v}(X)$ the Bayes estimator of μ for \mathcal{R}_Q^v under prior π_C . We infer from Hartigan [8] that $\mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi',v}) - \mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi,v}) \geq 0$, for all $\mu \in C$, $v > 0$, with equality iff C is a cone and μ is its vertex. Making use of this and Theorem 2.1, the result follows immediately. \square

Second proof. Using the expression of the risk difference in (1.6), the risk difference between \hat{p}_U and \hat{p}_{π_C} equals

$$\begin{aligned} \Delta \mathcal{R}(\mu) &= \mathcal{R}_{KL}(\mu, \hat{p}_U) - \mathcal{R}_{KL}(\mu, \hat{p}_{\pi_C}) \\ &= E_{\mu;v_w} \log m_{\pi_C}(W; v_w) - E_{\mu;v_x} \log m_{\pi_C}(X; v_x) \\ &= \int_{\mathbb{R}^p} p(w|\mu, v_w) \log \left[\int_C p(w|\theta, v_w) d\theta \right] dw \\ &\quad - \int_{\mathbb{R}^p} p(x|\mu, v_x) \log \left[\int_C p(x|\theta, v_x) d\theta \right] dx \end{aligned}$$

by expressing the marginals m_{π_C} . Then, applying the successive changes of variables $z = \frac{1}{\sqrt{v_w}}(w - \mu)$, $s = \frac{1}{\sqrt{v_w}}(\theta - w)$, $z = \frac{1}{\sqrt{v_x}}(x - \mu)$ and $s = \frac{1}{\sqrt{v_x}}(\theta - x)$, we obtain

$$\Delta \mathcal{R}(\mu) = \int_{\mathbb{R}^p} p(z|0, 1) \log \left[\frac{\int_{C_1} p(s|0, 1) ds}{\int_{C_2} p(s|0, 1) ds} \right] dz \quad (3.1)$$

where $C_1 = \frac{1}{\sqrt{v_w}}\{C - \mu\} - z$ and $C_2 = \frac{1}{\sqrt{v_x}}\{C - \mu\} - z$. We need to prove that, for any $\mu \in C$, $\Delta \mathcal{R}(\mu) \geq 0$. Let $\mu \in C$. Clearly it suffices to show that, for any $z \in \mathbb{R}^p$, we have

$$C_2 = \frac{1}{\sqrt{v_x}}\{C - \mu\} - z \subset C_1 = \frac{1}{\sqrt{v_w}}\{C - \mu\} - z. \quad (3.2)$$

This is equivalent to showing that

$$\frac{1}{\sqrt{v_x}}\{C - \mu\} \subset \frac{1}{\sqrt{v_w}}\{C - \mu\}. \quad (3.3)$$

Let $c_1 \in \frac{1}{\sqrt{v_x}}\{C - \mu\}$. Then there exists $c \in C$ such that $c_1 = \frac{1}{\sqrt{v_x}}(c - \mu)$. Hence we can write $c_1 = \frac{1}{\sqrt{v_w}}(c' - \mu)$, with $c' = \frac{\sqrt{v_w}}{\sqrt{v_x}}c + (1 - \frac{\sqrt{v_w}}{\sqrt{v_x}})\mu$. As $v_w < v_x$, the last quantity is a convex combination of c and μ . Since $c \in C$, $\mu \in C$ and C is a convex set then $c' \in C$, which implies that (3.3) is satisfied.

Finally, since $\{C - \mu\}$ is a convex set containing the origin, (3.3) is satisfied with equality iff C is a cone and μ is its vertex which means that $\Delta \mathcal{R}(\mu) > 0$, except when C is a cone and μ is its vertex. This completes the proof. \square

It is worth noting that this second proof of Theorem 3.1 and part (b) of Theorem 2.1 guarantee that the domination of \hat{p}_{π_C} over \hat{p}_U implies Hartigan's result, that is, the domination of $\hat{\mu}_{\pi_C}$ over X when C is convex with a non-empty interior. Indeed, for any $v_x > 0$ and $v_y > 0$ fixed, the domination of \hat{p}_{π_C} over \hat{p}_U indicates that

$$\frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (\mathcal{R}_Q^v(\mu, X) - \mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi_c, v})) dv \geq 0, \quad (3.4)$$

for all $v_x > v_w > 0$. Hence, for any $0 < a < b$, the integral on (a, b) of the integrand term in (3.4) is non negative, which implies that this integrand term is non negative almost everywhere. Finally, by continuity of this function in v , it follows that $\mathcal{R}_Q^v(\mu, X) - \mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi_c, v}) \geq 0$. Hence, our method provides an independent proof of Hartigan's result.

We now turn to the particular case where $p = 1$ and C is a compact interval, say $[-m, m]$ without loss of generality. Kubokawa [11] obtained a class of priors that lead to improvements on the minimum risk equivariant estimator X for risk \mathcal{R}_Q^v . Here, in analogous manner as above, we give a parallel result for our density estimation problem with risk \mathcal{R}_{KL} .

Theorem 3.2. *Let $p = 1$ and $\mu \in [-m, m]$. Let π be a prior with a symmetric density supported on $[-m, m]$, which is also increasing and logconcave on $[0, m]$. Then \hat{p}_π dominates \hat{p}_U for risk \mathcal{R}_{KL} and $\mu \in [-m, m]$.*

Proof. Since Kubokawa [11] showed that $\hat{\mu}_{\pi, v}(X)$ dominates X for $\mu \in [-m, m]$ for all $v > 0$ under the given assumptions on π , the result follows directly from Theorem 2.1 with $\hat{p}_{\pi'} \equiv \hat{p}_U$. \square

3.2. Improving on a maximum likelihood estimator

In this subsection, we give other direct implications of Theorem 2.1 and further applications when the mean μ is restricted to a ball centered at 0 of radius m . As in the above Hartigan type result, we can borrow known or easy to derive results for the analogous estimation problem with risk \mathcal{R}_Q^v . However, in contrast, we seek to dominate here the plug-in density based on the maximum likelihood estimator for a univariate bounded normal mean, rather than the generalized Bayes estimator \hat{p}_U .

Corollary 3.1. *Suppose that the Bayesian estimator $\hat{\mu}_{\pi, v}$ dominates a given estimator δ_1 for $\mu \in C$, and for all $v \in (v_w, v_x)$ under risk \mathcal{R}_Q^v . Suppose further that either*

- (A) $\int_{v_w}^{v_x} \frac{1}{v^2} (\mathcal{R}_Q^v(\mu, \delta_1) - \mathcal{R}_Q^v(\mu, \delta_1)) dv \geq 0$ for all $\mu \in C$
- or
- (B) $\mathcal{R}_Q^v(\mu, \delta_1)$ increases in v for $v \in (v_w, v_x)$, for all $\mu \in C$.

Then $\hat{p}_\pi(X)$ improves upon $\hat{p}_1(X) \sim \mathcal{N}_p(\delta_1(X), v_y I_p)$ under risk \mathcal{R}_{KL} for $\mu \in C$.

Proof. The proof is straightforward. First, (B) implies (A). Secondly, by virtue of Theorem 2.1 and given our assumptions, we have

$$\begin{aligned} \mathcal{R}_{\text{KL}}(\mu, \hat{p}_1) - \mathcal{R}_{\text{KL}}(\mu, \hat{p}_\pi) &= \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (\mathcal{R}_Q^v(\mu, \delta_1) - \mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi, v})) dv \\ &\quad + \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (\mathcal{R}_Q^{v_x}(\mu, \delta_1) - \mathcal{R}_Q^v(\mu, \delta_1)) dv \\ &> 0 \end{aligned}$$

□

Proceeding in analogous manner, we also have the following potentially useful result which focusses on the risk behavior of the Bayes estimators $\hat{\mu}_{\pi, v_x}$, for $v_w < v < v_x$, instead of that of δ_1 .

Corollary 3.2. *Suppose that the Bayesian estimator $\hat{\mu}_{\pi, v_x}$ dominates a given estimator δ_1 for $\mu \in C$ under risk $\mathcal{R}_Q^{v_x}$. Suppose further that either*

$$(A') \int_{v_w}^{v_x} \frac{1}{v^2} (\mathcal{R}_Q^{v_x}(\mu, \hat{\mu}_{\pi, v_x}) - \mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi, v})) dv \geq 0 \text{ for all } \mu \in C$$

or

$$(B') \mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi, v}) \text{ increases in } v \text{ for } v \in (v_w, v_x), \text{ for all } \mu \in C.$$

Then $\hat{p}_\pi(X)$ improves upon $\hat{p}_1(X) \sim \mathcal{N}_p(\delta_1(X), v_y I_p)$ under risk \mathcal{R}_{KL} for $\mu \in C$.

We now turn to an application of Corollary 3.1 for a univariate normal mean which is bounded to an interval, but, in view of making use of its condition (B), we require the following preliminary result.

Lemma 3.1. *For estimating a univariate normal mean μ bounded to an interval $[a, b]$, the risk $\mathcal{R}_Q^v(\mu, \delta_{mle})$ increases in $v \in (0, \infty)$ for all $\mu \in [a, b]$, where $\delta_{mle}(X)$ is the maximum likelihood estimator of μ .*

Proof. Let $X \sim \mathcal{N}(\mu, v)$ with $a \leq \mu \leq b$, and let $\phi(t) = (1/\sqrt{2\pi}) e^{-t^2/2}$ be the standard normal density function. The maximum likelihood estimator of μ is given by

$$\delta_{mle}(X) = a\mathbb{1}_{(-\infty, a)}(X) + X\mathbb{1}_{[a, b]}(X) + b\mathbb{1}_{(b, \infty)}(X) = \begin{cases} a & \text{if } X \leq a \\ X & \text{if } a \leq X \leq b \\ b & \text{if } X \geq b \end{cases}$$

Let $\sigma = \sqrt{v}$. Then the risk of δ_{mle} at μ , $\mathcal{R}_Q^v(\mu, \delta_{mle})$, equals

$$\begin{aligned} \mathcal{R}(\mu, \sigma, \delta_{mle}) &= E_\mu \left[(\delta_{mle}(X) - \mu)^2 \right] \\ &= (a - \mu)^2 \int_{-\infty}^a \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) dx + (b - \mu)^2 \int_b^{+\infty} \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) dx \\ &\quad + \int_a^b (x - \mu)^2 \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) dx \end{aligned}$$

$$\begin{aligned}
 &= (a - \mu)^2 \int_{-\infty}^{(a-\mu)/\sigma} \phi(y) dy + (b - \mu)^2 \int_{(b-\mu)/\sigma}^{+\infty} \phi(y) dy \\
 &\quad + \sigma^2 \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} y^2 \phi(y) dy.
 \end{aligned}$$

By differentiating under the integral sign with respect to σ we obtain

$$\begin{aligned}
 \frac{\partial}{\partial \sigma} \mathcal{R}(\mu, \sigma, \delta_{mle}) &= \frac{(\mu - a)^3}{\sigma^2} \phi\left(\frac{a - \mu}{\sigma}\right) + \frac{(b - \mu)^3}{\sigma^2} \phi\left(\frac{b - \mu}{\sigma}\right) \\
 &\quad + 2\sigma \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} y^2 \phi(y) dy + \sigma^2 \frac{(b - \mu)^2}{\sigma^2} \phi\left(\frac{b - \mu}{\sigma}\right) \frac{\mu - b}{\sigma^2} \\
 &\quad - \sigma^2 \frac{(a - \mu)^2}{\sigma^2} \phi\left(\frac{a - \mu}{\sigma}\right) \frac{\mu - a}{\sigma} \\
 &= 2\sigma \int_{(a-\mu)/\sigma}^{(b-\mu)/\sigma} y^2 \phi(y) dy \\
 &> 0,
 \end{aligned}$$

after simplification. Hence $\mathcal{R}(\mu, \sigma, \delta_{mle})$ increases in σ and we have the desired result. \square

For risk \mathcal{R}_Q^v , $\mu \in [a, b]$, Marchand and Perron [14], as well as Casella and Strawderman [4] (for δ_{BU} below), showed that:

- $\delta_{BU}(X)$ dominates $\delta_{mle}(X)$ for **(i)** $(b - a)/2 \leq \sqrt{v}$;
- all Bayes estimators $\delta_\pi(X)$ with π a symmetric measure about $(a + b)/2$ supported on $[a, b]$ dominate $\delta_{mle}(X)$ for **(ii)** $(b - a)/2 \leq c_1 \sqrt{v}$;
- and $\delta_{FU}(X)$ dominates $\delta_{mle}(X)$ for **(iii)** $(b - a)/2 \leq c_2 \sqrt{v}$;

with $c_1 \approx 0.4837$, $c_2 \approx 0.5230$, and where $\delta_{BU}(X)$ is the Bayes estimator with respect to the two-point uniform prior on $\{-m, m\}$, $\delta_{FU}(X)$ is the Bayes estimator with respect to the uniform $U(a, b)$ prior. For the density estimating problem, we denote $\hat{p}_{BU}(X)$, $\hat{p}_{FU}(X)$ as the corresponding Bayes estimators for the boundary uniform and fully uniform priors respectively. In view of the above dominance results for risk \mathcal{R}_Q^v , part (B) of Corollary 3.1, and Lemma 3.1, we now derive improvements on the plug-in maximum likelihood estimator $\hat{p}_{mle}(X) \sim \mathcal{N}_p(\delta_{mle}(X), v_y I_p)$.

Corollary 3.3. *For estimating $p_Y(\cdot | \mu; \nu_y)$ under risk \mathcal{R}_{KL} , with $p = 1, \mu \in [a, b]$,*

- (a) $\hat{p}_{BU}(X)$ dominates $\hat{p}_{mle}(X)$ for **(i)** $(b - a)/2 \leq \sqrt{v_w}$;
- (b) all Bayes estimators $\hat{p}_\pi(X)$, with π a symmetric measure about $(a + b)/2$ supported on $[a, b]$, dominate $\hat{p}_{mle}(X)$ for **(ii)** $(b - a)/2 \leq c_1 \sqrt{v_w}$;
- (c) $\hat{p}_{FU}(X)$ dominates $\hat{p}_{mle}(X)$ for **(iii)** $(b - a)/2 \leq c_2 \sqrt{v_w}$.

The Bayesian estimators $\hat{p}_\pi(X)$, $\hat{p}_{BU}(X)$ and $\hat{p}_{FU}(X)$ may be evaluated directly by computing the predictive density $\hat{p}(\cdot | x)$ as in (1.3), or via (1.5). For

instance, in the case of the two-point uniform prior on $\{-m, m\}$, we obtain the predictive density $\hat{p}_{BU}(X)$ as being the density of a mixture of the two normal distributions $N(-m, v_y)$ and $N(m, v_y)$ with respective weights $w(x) = (1 + e^{2mx/v_x})^{-1}$ and $1 - w(x)$.

3.3. Bayesian improvements on the plug-in $N_p(x, v_y I_p)$ estimator under a constraint

Here is an instructive example in relationship with Corollaries 3.1 and 3.2. Consider normal priors $\pi_\tau \sim \mathcal{N}_p(0, \tau I)$ and the corresponding Bayes estimators $\hat{\mu}_{\pi_\tau, v}(x) = \tau/(\tau + v)x$ under squared error loss. Now, for cases where μ is constrained to the ball $B_m = \{\mu \in \mathbb{R}^p : \|\mu\| \leq m\}$, one may verify that aX dominates X under risk \mathcal{R}_Q^v iff $a \in [(m^2 - pv)/(m^2 + pv), 1)$, which tells us that $\hat{\mu}_{\pi_\tau, v}(X)$ lowers the risk of $\delta_1(X)$ on B_m iff $\tau \geq (m^2/2p) - v/2$. Since, $\mathcal{R}_Q^v(\mu, X) = pv$ is increasing in v , part (B) of Corollary 3.1 tells us that $\hat{p}_{\pi_\tau}(X)$ lowers the risk \mathcal{R}_{KL} of the plug-in $N_p(X, v_y I_p)$ whenever $\tau \geq (m^2/2p) - v/2$ for all $v \in (v_w, v_x)$, that is

$$\tau \geq \frac{m^2}{2p} - \frac{v_w}{2}. \quad (3.5)$$

So sufficiently large prior variances lead here to improvements, with the dominance for $\tau \rightarrow \infty$ interpretable as the dominance of \hat{p}_U on B_m for all $m > 0$, which is of course already known.

Alternatively, in view of applying Corollary 3.2, we begin with the weaker condition $\tau \geq (m^2/2p) - v_x/2$ for $\hat{\mu}_{\pi_\tau, v_x}(X)$ to dominate $\delta_1(X) = X$ under risk $\mathcal{R}_Q^{v_x}$. However, turning to condition (B'), we observe that the quadratic risk of $\hat{\mu}_{\pi_\tau, v}(X)$, given by $\mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi_\tau, v}) = \frac{\tau^2}{(\tau+v)^2}pv + \frac{v^2}{(\tau+v)^2}\|\mu\|^2$, is only increasing in $v \in (v_w, v_x)$ for all $\mu \in B_m$ (or $\mu \in \mathbb{R}^p$) iff $R(0, \hat{\mu}_{\pi_\tau, v})$ is increasing in v for all $v \in (v_w, v_x)$, or equivalently $\tau \geq v_x$. We hence obtain $\tau \geq \max(v_x, (m^2/2p) - v_x/2)$ as an alternative sufficient condition to (3.5), and the improved sufficient condition

$$\tau \geq \min\left(\frac{m^2}{2p} - \frac{v_w}{2}, \max\left(v_x, \frac{m^2}{2p} - \frac{v_x}{2}\right)\right),$$

for $\hat{p}_{\pi_\tau}(X)$ to dominate the plug-in $N_p(X, v_y I_p)$ under risk R_{KL} with $\mu \in B_m$. Finally, we point out that yet a stronger result can be arrived at by working with condition (A') of Corollary 3.2.

4. Minimax results

In this section, we derive minimaxity results for cases where μ is restricted. First, we fully exploit Section 2's relationships between the risks \mathcal{R}_{KL} and \mathcal{R}_Q , as well as known minimaxity results for cases where the mean is restricted to a ball. Secondly, we establish that the estimator \hat{p}_U remains minimax when the mean μ is restricted to a cone yielding a general result analogous to Tsukuma and Kubokawa's result [20] under squared error loss.

4.1. Minimavity when the mean is restricted to a ball

Here is a general framework which we will seek to apply in cases where the mean μ is restricted to a ball of radius m centered around the origin.

Theorem 4.1. *Consider our general problems of estimating μ with risk \mathcal{R}_Q^v and the density $p_Y(\cdot|\mu, v_y)$ with risk \mathcal{R}_{KL} , where $\mu \in C \subseteq \mathbb{R}^p$. Suppose there exists a prior measure π^* and a subset C_0 of C such that $\pi^*(C_0) = 1$; $\mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi^*,v}) = \bar{\mathcal{R}}_v$ for $\mu \in C_0$; and $\sup_{\mu \in C} \{\mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi^*,v})\} = \bar{\mathcal{R}}_v$ for all $v \in (v_w, v_x)$. Then \hat{p}_{π^*} is minimax for risk \mathcal{R}_{KL} among all Bayesian estimators that have constant risk \mathcal{R}_Q^v on C_0 .*

Remark 4.1. Note that these assumptions imply that, for all $v \in (v_w, v_x)$, $\hat{\mu}_{\pi^*,v}$ is minimax under \mathcal{R}_Q^v and π^* is least favorable.

Proof of Theorem 4.1. For any π' such that $\mathcal{R}_Q^v(\cdot, \hat{\mu}_{\pi',v})$ is constant on C_0 , there exists $\mu_0 \in C_0$ (independent of v) such that $\mathcal{R}_Q^v(\mu_0, \hat{\mu}_{\pi',v}) \geq \bar{\mathcal{R}}_v$ for all $v \in (v_w, v_x)$, since $\bar{\mathcal{R}}_v$ is the Bayes risk associated with π^* . Now, making use of Theorem 2.1 with $\hat{p}_{\pi'} \equiv \hat{p}_U$ and the constant \mathcal{R}_{KL} risk of \hat{p}_U , we obtain indeed for any Bayes estimator $\hat{p}_{\pi'}$,

$$\begin{aligned} \sup_{\mu \in C} \mathcal{R}_{KL}(\mu, \hat{p}_{\pi'}) &\geq \mathcal{R}_{KL}(\mu_0, \hat{p}_{\pi'}) \\ &= \mathcal{R}_{KL}(\mu_0, \hat{p}_U) + \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (\mathcal{R}_Q^v(\mu_0, \hat{\mu}_{\pi',v}) - pv) dv \\ &\geq \mathcal{R}_{KL}(\mu_0, \hat{p}_U) + \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (\sup_{\mu \in C} \{\mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi^*,v})\} - pv) dv \\ &\geq \sup_{\mu \in C} \left\{ \mathcal{R}_{KL}(\mu, \hat{p}_U) + \frac{1}{2} \int_{v_w}^{v_x} \frac{1}{v^2} (\mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi^*,v}) - pv) dv \right\} \\ &= \sup_{\mu \in C} \mathcal{R}_{KL}(\mu, \hat{p}_{\pi^*}). \end{aligned}$$

□

As an example, consider a constraint to a ball where $\mu \in C = \{\mu \in \mathbb{R}^p : \|\mu\| \leq m\}$. For this problem, Bayes estimators form a complete class and a minimax estimator can be found among orthogonally invariant estimators. Since such estimators have constant risk on spheres where $\|\mu\| = \lambda$, a minimax estimator can be found among orthogonally invariant Bayes estimators, or equivalently among Bayes estimators with spherically symmetric priors ([9]). Now take $C_0 = \{\mu \in C : \|\mu\| = m\}$ and π^* to be the uniform prior measure on C_0 . As studied in Berry [2], Marchand and Perron [15], or Casella and Strawderman [4] ($p = 1$), the corresponding Bayes estimator is given by $\hat{\mu}_{\pi^*,v}(x) = \frac{I_{p/2}(m\|x\|/v)}{I_{p/2-1}(m\|x\|/v)} \frac{m x}{\|x\|}$, where $I_k(\cdot)$ is the modified Bessel function of order k , and $\hat{\mu}_{\pi^*,v}(X)$ is unique minimax for risk \mathcal{R}_Q^v with $\sup_{\mu \in C} \{\mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi^*,v})\} = \sup_{\mu \in C_0} \{\mathcal{R}_Q^v(\mu, \hat{\mu}_{\pi^*,v})\} = \bar{\mathcal{R}}_v$ iff $m \leq c_0(p)\sqrt{v}$. Hence, to satisfy the assumptions of Theorem 4.1, we require simply $m \leq c_0(p)\sqrt{v}$ for all $v \in (v_w, v_x)$, which yields the following.

Corollary 4.1. *For estimating the density $p(\cdot|\mu, v_y)$ with the constraint $\|\mu\| \leq m$, the boundary uniform Bayes estimator \hat{p}_{π^*} is minimax for risk \mathcal{R}_{KL} whenever $m \leq c_0(p)\sqrt{v_w}$, where $c_0(p)$ is the constant defined and evaluated by Berry [2], and Marchand and Perron [15].*

We conclude this subsection by pointing out that, as at the end of Section 3.2, \hat{p}_{π^*} may be evaluated directly as in (1.3), or by using (1.5), and that Marchand and Perron [15] provide various properties of $c_0(p)$, including the lower bound: $\sqrt{p} \leq c_0(p)$.

4.2. Minimality of \hat{p}_U when the mean is restricted to a cone

In this section, we deal with the minimality of \hat{p}_U in (1.4) when the mean μ is restricted to a cone. We point out that another potential and related approach to the problem is given by the recent work of Marchand and Strawderman [16].

Theorem 4.2. *Let C be a convex cone, with non-empty interior. Then the unrestricted predictive density estimator \hat{p}_U in (1.4) is minimax under the KL loss (1.1), when μ is restricted to C .*

Proof. Let r be the constant risk of \hat{p}_U . We will show that r is a limit of Bayes risks of a sequence of Bayes predictive densities $\hat{p}_{\pi_k}(y|x)$ with respect to a sequence of proper priors π_k lying in C . Let $C_k = C \cap B_k = \{c \in C \mid \|c\| \leq k\}$ where B_k denotes the ball of radius k centered at 0. By the positive homogeneity of C , we can express C_k as

$$kC_1 = \{kc \mid c \in C_1\}. \quad (4.1)$$

Consider, as a sequence of proper priors, the uniform distributions on the convex sets C_k , that is,

$$\pi_k(\mu) = \frac{1}{\lambda(C_k)} \mathbb{1}_{C_k}(\mu) = \frac{1}{k^p} \frac{1}{\lambda(C_1)} \mathbb{1}_{C_k}(\mu) \quad (4.2)$$

where λ is the Lebesgue measure on \mathbb{R}^p .

Using (4.2) and the expression of the risk difference in (1.6), the difference of Bayes risks between \hat{p}_U and \hat{p}_{π_k} is

$$\begin{aligned} r - r(\pi_k, \hat{p}_{\pi_k}) &= \int_{\mathbb{R}^p} [\mathcal{R}_{KL}(\hat{p}_U, \mu) - \mathcal{R}_{KL}(\hat{p}_{\pi_k}, \mu)] \pi_k(\mu) d\mu \\ &= \frac{1}{\lambda(C_k)} \int_{C_k} [E_{\mu, v_w} [\log m_{\pi_k}(W; v_w)] - E_{\mu; v_x} [\log m_{\pi_k}(X, v_x)]] d\mu. \end{aligned} \quad (4.3)$$

Now, using the second expression of $\pi_k(\mu)$ in (4.2) and expressing the marginal m_{π_k} in the bracketed term in (4.3), we have

$$r - r(\pi_k, \hat{p}_{\pi_k}) = \frac{1}{\lambda(C_1)} [L(v_w, k) - L(v_x, k)] \quad (4.4)$$

where

$$\begin{aligned} L(v, k) &= \frac{1}{k^p} \int_{C_k} E_{\mu, v} \left[\log \int_{C_k} p(Z|a, v) da \right] d\mu \\ &= \frac{1}{k^p} \int_{C_k} \left[\int_{\mathbb{R}^p} p(z|\mu, v) \log \left[\int_{C_k} p(z|a, v) da \right] dz \right] d\mu. \end{aligned} \quad (4.5)$$

Then with the changes of variables $s = \frac{1}{\sqrt{v}}(a - z)$, $z' = \frac{1}{\sqrt{v}}(z - \mu)$, and $r = \frac{1}{k} \mu$, $L(v, k)$ becomes

$$\begin{aligned} L(v, k) &= \frac{1}{k^p} \int_{C_k} \left[\int_{\mathbb{R}^p} p(z|\mu, v) \log \left[\int_{\frac{1}{\sqrt{v}}C_k - \frac{1}{\sqrt{v}}z} p(s|0, 1) ds \right] dz \right] d\mu \\ &= \frac{1}{k^p} \int_{C_k} \left[\int_{\mathbb{R}^p} p(z'|0, 1) \log \left[\int_{\frac{1}{\sqrt{v}}C_k - \frac{1}{\sqrt{v}}\mu - z'} p(s|0, 1) ds \right] dz' \right] d\mu \end{aligned} \quad (4.6)$$

$$= \int_{C_1} \left[\int_{\mathbb{R}^p} p(z|0, 1) \log \int_{\frac{k}{\sqrt{v}}C_1 - \frac{k}{\sqrt{v}}r - z} p(s|0, 1) ds dz \right] dr. \quad (4.7)$$

It remains to show that the limit of $L(v, k) = L^*(k/\sqrt{v})$ when k goes to infinity exists and hence does not depend on v , which will imply that $\lim_{k \rightarrow \infty} (r - r(\pi_k, \hat{p}_{\pi_k})) = 0$. Let $A_{k,r} = \frac{k}{\sqrt{v}}\{C_1 - r\}$. As the interior of C is non-empty, for any interior point r of C_1 , there exists a ball B_ϵ in \mathbb{R}^p of radius $\epsilon > 0$, centered at 0, such that

$$B_\epsilon \subset C_1 - r.$$

Then it follows that

$$\mathbb{R}^p \supseteq \liminf_{k \rightarrow \infty} A_{k,r} \supseteq \liminf_{k \rightarrow \infty} B_{\frac{k}{\sqrt{v}}\epsilon} = \mathbb{R}^p \quad (4.8)$$

and hence

$$\lim_{k \rightarrow \infty} \int_{\frac{k}{\sqrt{v}}C_1 - \frac{k}{\sqrt{v}}r - z} p(s|0, 1) ds = 1. \quad (4.9)$$

Now the convexity of C , and hence of C_1 , allows, according to Lemma A.1, to apply the Lebesgue dominated convergence theorem to the expression in (4.7). Thus we obtain $\lim_{k \rightarrow \infty} L(k, v) = 0$ for any fixed v . Therefore $\lim_{k \rightarrow \infty} (r - r(\pi_k, \hat{p}_{\pi_k})) = 0$, which establishes the minimaxity of \hat{p}_U .¹ \square

As an immediate consequence of Theorem 3.1 and Theorem 4.2, we have the following corollary.

Corollary 4.2. *Let $C \subset \mathbb{R}^p$ be a convex cone with non-empty interior, and let $\pi_C = \mathbb{1}_C(\mu)$ be the flat prior restricted to C . Then \hat{p}_{π_C} is minimax for risk \mathcal{R}_{KL} and $\mu \in C$.*

¹Note that we implicitly use the fact that the boundary of C_1 is negligible with respect to the Lebesgue measure as a boundary of a convex subset in \mathbb{R}^p (cf. [13]).

Note that Theorem 4.2 holds also for affine convex cones with non-empty interiors. Indeed, it suffices to consider the sequence of proper priors $\pi_k(\mu) = \frac{1}{\lambda(C_k)} \mathbb{1}_{C_k}(\mu - g)$. More generally, we can generalize Theorem 4.2 to cones which are not necessarily convex. The proof is analogous to the one of Theorem 4.2 and is relegated to the appendix.

Theorem 4.3. *Suppose C is a finite disjoint union of (affine) convex cones, in the sense that the restriction is given by*

$$C = \bigcup_{i=1}^n (C^i + g_i) \quad (4.10)$$

where C^1, \dots, C^n are convex cones with non-empty interiors, and g_1, \dots, g_n are n fixed points in \mathbb{R}^p where, for any $i \neq j$,

$$\lambda((C^i + g_i) \cap (C^j + g_j)) = 0. \quad (4.11)$$

Then \hat{p}_U is still minimax when μ is restricted to C .

5. Plug-in estimators: Some additional considerations

We now expand on a general phenomenon concerning plug-in estimators and how they can be improved upon within the class of normal density estimators. Below, we set $G_m(c) = (1 - \frac{1}{c})m - \log c$ for $c > 1, m > 1$, and we let $c_0(m)$ denote the root of $G_m(c)$ that lies in (m, ∞) . It is easy to show that $G_m(c)$ is positive for $c \in (1, m]$, is maximized on $(1, \infty)$ at $c = m$, and does indeed have a single root on (m, ∞) . As above, we denote $\mathcal{R}_Q^{\nu_x}(\mu, \delta)$ as the risk $E_{\mu, \nu_x}[\|\delta(X) - \mu\|^2]$.

Theorem 5.1. *Suppose $\delta(X)$ is an estimator of μ , $\mu \in C$, such that $\inf_{\mu \in C} \mathcal{R}_Q^{\nu_x}(\mu, \delta) > 0$. Consider the estimator $\hat{p}_c \sim \mathcal{N}_p(\delta(X), cv_y I_p)$ and the plug-in version $\hat{p}_1 \sim \mathcal{N}_p(\delta(X), v_y I_p)$. We then have the following.*

(a) *For any given μ , the optimal value of c is*

$$m(\mu) = 1 + \frac{\mathcal{R}_Q^{\nu_x}(\mu, \delta)}{pv_y},$$

and \hat{p}_c improves upon \hat{p}_1 iff $c \in (1, c_0(m(\mu)))$;

(b) *For $\mu \in C$, \hat{p}_c dominates \hat{p}_1 iff $c \in (1, c_0(\inf_{\mu \in C} m(\mu)))$ or $c = c_0(\inf_{\mu \in C} m(\mu))$ with $m(\cdot)$ not constant on C .*

Proof. We have

$$\begin{aligned} \mathcal{R}_{\text{KL}}(\mu, \hat{p}_1) - \mathcal{R}_{\text{KL}}(\mu, \hat{p}_c) &= E^{X,Y} \left[\log \left(\frac{\hat{p}_c(Y; X)}{\hat{p}_1(Y; X)} \right) \right] \\ &= E^{X,Y} \left[\log \left(\frac{(2\pi cv_y)^{-p/2} e^{-\|Y - \delta(X)\|^2 / 2cv_y}}{(2\pi v_y)^{-p/2} e^{-\|Y - \delta(X)\|^2 / 2v_y}} \right) \right] \end{aligned}$$

$$\begin{aligned}
 &= -\frac{p}{2} \log c + E^{X,Y} \left[\frac{(1-1/c)}{2v_y} \|Y - \delta(X)\|^2 \right] \\
 &= \frac{p}{2} (-\log c + (1-1/c) m(\mu)),
 \end{aligned}$$

since $E^{X,Y}[\|Y - \delta(X)\|^2] = E^{X,Y}[\|Y - \mu\|^2] + E^{X,Y}[\|\mu - \delta(X)\|^2] = pv_y + \mathcal{R}_Q^{v_x}(\mu, \delta)$. The result then follows from the above expression, the properties of $G_m(\cdot)$ and c_0 , and since $m(\mu) \geq \inf_{\mu \in C} m(\mu) > 1$ for all $\mu \in C$. \square

Remark 5.1. In the setup of part (b) of Theorem 5.1, it may be easily derived from a similar analysis to the one above that, among estimators \hat{p}_c , those with $c \in [\inf_{\mu \in C} m(\mu), \sup_{\mu \in C} m(\mu)]$ form a complete subclass, while the estimators \hat{p}_c with $c \in [\inf_{\mu \in C} m(\mu), c_0(\inf_{\mu \in C} m(\mu))]$ form a complete subclass among those that dominate \hat{p}_1 .

Example 5.1. Take $\delta(X) = aX$ with $0 < a \leq 1$, $C = \mathbb{R}^p$. Since $E_{\mu, v_x}[\|aX - \mu\|^2] = a^2pv_x + (1-a)^2\mu'\mu$, we have $m(\mu) = 1 + \frac{a^2pv_x + (1-a)^2\mu'\mu}{pv_y}$ and $\inf_{\mu \in \mathbb{R}^p} m(\mu) = 1 + a^2\frac{v_x}{v_y}$. Part (b) of Theorem 5.1 tells us that \hat{p}_1 is dominated by \hat{p}_c under \mathcal{R}_{KL} iff $c \in (1, c_0(1 + a^2\frac{v_x}{v_y})]$ (with $c_0(1 + a^2\frac{v_x}{v_y}) > 1 + a^2\frac{v_x}{v_y}$). For the specific case $a = 1$, the optimal value of c reduces to $1 + \frac{v_x}{v_y}$ regardless of μ , yielding the optimal choice $\hat{p}_{1+\frac{v_x}{v_y}} \sim \mathcal{N}_p(X, (v_x + v_y)I_p)$ and recovering its well known optimality property among the estimators $\hat{p}_c \sim \mathcal{N}_p(X, cv_yI_p)$.

Example 5.2. Take the James-Stein estimator $\delta(X) = (1 - \frac{(p-2)v_x}{X'X})X$ for $p \geq 3$ and $C = \mathbb{R}^p$. $\delta(X)$ is minimax with $\inf_{\mu \in \mathbb{R}^p} \mathcal{R}_Q^{v_x}(\mu, \delta) = 2v_x$ and $\sup_{\mu \in \mathbb{R}^p} \mathcal{R}_Q^{v_x}(\mu, \delta) = pv_x$. Both Theorem 5.1 and Remark 5.1 then be applied with $\inf_{\mu \in \mathbb{R}^p} m(\mu) = 1 + \frac{2v_x}{pv_y}$ and $\sup_{\mu \in \mathbb{R}^p} m(\mu) = 1 + \frac{v_x}{v_y}$. Notice that, here and for any case where $\mu \in \mathbb{R}^p$, the supremum value of $m(\mu)$ is bounded below by $1 + \frac{v_x}{v_y}$, since the supremum risk of any estimator $\delta(X)$ surpasses the minimax risk which is pv_x and, hence, the estimator $\hat{p}_{1+\frac{v_x}{v_y}} \sim \mathcal{N}_p(\delta(X), (v_x + v_y)I_p)$ is admissible among the estimators \hat{p}_c regardless of $\delta(X)$. This provides an interesting extension of its outright optimality among the \hat{p}_c 's when $\delta(X) = X$.

Theorem 5.1 and Remark 5.1 encompass the general phenomenon that plug-in estimators $N_p(\delta(X), v_yI_p)$ can generally be improved upon by inflating the variance by a certain explicit factor, regardless of the efficiency (or inefficiency) of $\delta(X)$ for estimating μ . Moreover, the prescribed range of inflation is directly proportional to the risk $\mathcal{R}_Q^{v_x}(\mu, \delta)$, for all δ , and p . Alternatively, estimators \hat{p}_c for which $\delta(X)$ is inadmissible for estimating μ (e.g., Example 5.2) can be improved upon as follows.

Theorem 5.2. Let $c > 0$, $\mu \in C$, and $\delta_1(X)$ and $\delta_2(X)$ be two estimators of μ based on X . Then, the plug-in estimator $\hat{p}_2 \sim \mathcal{N}_p(\delta_2(X), cv_yI_p)$ dominates the plug-in estimator $\hat{p}_1 \sim \mathcal{N}_p(\delta_1(X), cv_yI_p)$ under \mathcal{R}_{KL} for $\mu \in C$ iff $\delta_2(X)$ dominates $\delta_1(X)$ as an estimator of μ , for $\mu \in C$.

Proof. The result follows directly with the decomposition

$$\begin{aligned} \mathcal{R}_{\text{KL}}(\mu, \hat{p}_2) - \mathcal{R}_{\text{KL}}(\mu, \hat{p}_1) &= E^{X,Y} \left[\log \left(\frac{\hat{p}_1(Y; X)}{\hat{p}_2(Y; X)} \right) \right] \\ &= E^{X,Y} \left[\frac{\|\delta_2(X) - \mu\|^2 - \|\delta_1(X) - \mu\|^2}{2cv_y} \right]. \end{aligned}$$

□

6. Conclusion

In [6], George *et al.* put forth an interesting parallel between minimax conditions for estimating a multivariate normal mean vector under quadratic risks \mathcal{R}_Q^v and minimax conditions for a predictive density estimator under Kullback-Leibler risk \mathcal{R}_{KL} . Similar connections with regards to admissibility were also developed by Brown *et al.* ([3]). In this paper, we obtain similar connections for dominance and minimaxity results when the mean is restricted to a convex set or to a cone. For instance, in the case where the mean is restricted to a convex set C , we prove domination of the Bayes predictive density estimator \hat{p}_{π_C} with respect to the uniform prior π_C over the generalized Bayes predictive density $\hat{p}_U(y|x)$ associated with the flat prior on \mathbb{R}^p .

An essential use is made of an explicit link between a collection of risks \mathcal{R}_Q^v and the Kullback-Leibler risk \mathcal{R}_{KL} . It allows us to create settings where a Bayesian predictive density estimator \hat{p}_π dominates a plug-in density estimator $\hat{p}_1(X) \sim \mathcal{N}_p(\delta_1(X), v_y I_p)$. Examples including improvements on the plug-in maximum likelihood estimator $\hat{p}_{mle}(X) \sim \mathcal{N}_p(\delta_{mle}(X), v_y I_p)$ are derived when the dimension p is 1.

Minimaxity results are also obtained when the mean is restricted to a ball as we derive conditions for which the boundary uniform Bayes estimator is minimax for risk \mathcal{R}_{KL} . Also, when the restricted parameter space is a convex cone C with non-empty interior, we prove that the unrestricted predictive density estimator \hat{p}_U remains minimax (as it is when no restriction is assumed).

Appendix A: Appendix

We expand first here on a technical lemma useful in the proof of Theorem 4.2, and we conclude with a proof of Theorem 4.3. For convenience, we denote by $\phi(s)$ the normal density $p(s|0, 1)$.

Let $C_1 \subset B_1$ be a convex set in \mathbb{R}^p with non-empty interior. For $t > 0$, write

$$\begin{aligned} L^*(t) &= \int_{C_1} \int_{\mathbb{R}^p} \phi(z) \log \left[\int_{tC_1 - tr - z} \phi(s) ds \right] dz dr \\ &= \int_{C_1 \times \mathbb{R}^p} \phi(z) g_t(r, z) d(r, z) \end{aligned}$$

with

$$g_t(r, z) = \log \left[\int_{tC_1 - tr - z} \phi(s) ds \right]. \quad (\text{A.1})$$

We showed in the proof of Theorem 4.2 that, for any $z \in \mathbb{R}^p$ and for any $r \in \overset{\circ}{C}_1$, the interior of C_1 , we have $\lim_{t \rightarrow \infty} g_t(r, z) = 0$, implying $\lim_{t \rightarrow \infty} L^*(t) = 0$ as long as dominated convergence can be applied. We now justify this.

Lemma A.1. *The function $|g_t(\cdot, \cdot)|$ given in (A.1) is bounded by an integrable function.*

Proof. Let $0 < t_1 < t_2$. Then, similarly to (3.3) in the proof of Theorem 3.1, we can see that $t_1(C_1 - r) \subset t_2(C_1 - r)$. Hence $g_t(r, z)$ is non decreasing in $t > 0$, for any $z \in \mathbb{R}^p$ and any $r \in C_1$.

We have $|g_t(r, z)| \leq |g_1(r, z)|$ for all $t \geq 1$ since $g_t(r, z) \leq 0$. By applying the Jensen inequality to the logarithmic function, we obtain

$$\begin{aligned} |g_1(r, z)| &= -\log \left\{ \lambda(C_1 - r - z) \int_{C_1 - r - z} \frac{1}{\lambda(C_1 - r - z)} \phi(s) ds \right\} \\ &\leq -\log \lambda(C_1) - \int_{C_1 - r - z} \frac{1}{\lambda(C_1)} \log \phi(s) ds \\ &= -\log \lambda(C_1) + \frac{1}{\lambda(C_1)} \int_{C_1 - r - z} \left[\frac{\|s\|^2}{2} + \frac{p}{2} \log(2\pi) \right] ds. \end{aligned}$$

Using the inequality

$$\|a + b + c\|^2 \leq 3(\|a\|^2 + \|b\|^2 + \|c\|^2)$$

for any a, b, c in \mathbb{R}^p , it follows that, for any $s \in C_1 - r - z$, that is $s = c_1 - r - z$ with $c_1 \in C_1$,

$$\|s\|^2 \leq 3(\|c_1\|^2 + \|r\|^2 + \|z\|^2) \leq 3(2 + \|z\|^2)$$

and hence

$$\begin{aligned} |g_1(r, z)| &\leq \frac{p}{2} \log(2\pi) - \log \lambda(C_1) + \frac{1}{\lambda(C_1)} \int_{C_1 - r - z} \left(3 + 3 \frac{\|z\|^2}{2} \right) ds \\ &= \frac{p}{2} \log(2\pi) - \log \lambda(C_1) + 3 + \frac{3}{2} \|z\|^2. \end{aligned}$$

Therefore

$$\begin{aligned} \int_{C_1 \times \mathbb{R}^p} \phi(z) |g_1(r, z)| d(r, z) &\leq \int_{C_1 \times \mathbb{R}^p} \phi(z) \left[\frac{p}{2} \log 2\pi - \log \lambda(C_1) + 3 + \frac{3}{2} \|z\|^2 \right] d(r, z) \\ &= \lambda(C_1) \left[\frac{p}{2} \log 2\pi - \log \lambda(C_1) + 3 + \frac{3p}{2} \right] \\ &< \infty. \end{aligned}$$

□

Proof of Theorem 4.3. Define for $k \geq 1$, and $i = 1, \dots, n$, $C_k^i = k C_1^i$ where

$$C_1^i = \{c \in C^i \mid \|c\| \leq 1\}.$$

Then define

$$C_k = \cup_{i=1}^n (C_k^i + g_i) = \cup_{i=1}^n (kC_1^i + g_i).$$

Note that $\lambda(C_k) = k^p \lambda(C_1)$ and, for any $k_1 \leq k_2$, we have $C_{k_1} \subset C_{k_2}$. Using the sequence of proper priors $\pi_k(\mu) = \frac{1}{k^p} \frac{1}{\lambda(C_1)} \mathbb{1}_{C_k}(\mu)$, we have, as in the proof of Theorem 4.2,

$$r - r(\pi_k, \hat{p}_{\pi_k}) = \frac{1}{\lambda(C_1)} \sum_{i=1}^n L^i(v_w, k) - L^i(v_x, k)$$

with, for $i = 1, \dots, n$,

$$\begin{aligned} L^i(v, k) &= \frac{1}{k^p} \int_{C_k^i + g_i} \int_{\mathbb{R}^p} \phi(z) \log \left[\int_{\frac{1}{\sqrt{v}} C_k - \frac{1}{\sqrt{v}} \mu - z} \phi(s) ds \right] dz d\mu \\ &= \frac{1}{k^p} \int_{C_k^i + g_i} \int_{\mathbb{R}^p} \phi(z) \log \left[\sum_{j=1}^n \int_{\frac{1}{\sqrt{v}} C_k^j + \frac{1}{\sqrt{v}} g_j - \frac{1}{\sqrt{v}} \mu - z} \phi(s) ds \right] dz d\mu \end{aligned}$$

thanks to (4.6). Then

$$L^i(v, k) \geq \bar{L}^i(v, k) \tag{A.2}$$

where

$$\begin{aligned} \bar{L}^i(v, k) &= \frac{1}{k^p} \int_{C_k^i + g_i} \int_{\mathbb{R}^p} \phi(z) \log \left[\int_{\frac{1}{\sqrt{v}} C_k^i + \frac{1}{\sqrt{v}} g_i - \frac{1}{\sqrt{v}} \mu - z} \phi(s) ds \right] dz d\mu \\ &= \frac{1}{k^p} \int_{C_k^i} \int_{\mathbb{R}^p} \phi(z) \log \left[\int_{\frac{1}{\sqrt{v}} C_k^i - \frac{1}{\sqrt{v}} \mu' - z} \phi(s) ds \right] dz d\mu', \end{aligned}$$

by the change of variables $\mu = \mu' + g_i$.

Note that the expression of $\bar{L}^i(v, k)$ corresponds to the expression of $L(v, k)$ in (4.6) which has been shown to satisfy $\lim_{k \rightarrow \infty} L(v, k) = 0$. Hence $\lim_{k \rightarrow \infty} \bar{L}^i(v, k) = 0$ and from (A.2) we get $\lim_{k \rightarrow \infty} L^i(v, k) \geq 0$. Since $L^i(v, k) \leq 0$, it follows that $\lim_{k \rightarrow \infty} L^i(v, k) = 0$. Therefore $\lim_{k \rightarrow \infty} (r - r(\pi_k, \hat{p}_{\pi_k})) = 0$, which completes the proof. \square

Acknowledgements

The authors are grateful to a referee for a careful reading of the paper and for his/her constructive suggestions.

References

- [1] AITCHISON, J. (1975) Goodness of prediction fit. *Biometrika*, **62** 547–554. [MR0391353](#)

- [2] BERRY, C. (1990) Minimax estimation of a bounded normal mean vector. *Journal of Multivariate Analysis*, **35** 130–139. [MR1084946](#)
- [3] BROWN, L. D., GEORGE, E. I. AND XU, X. (2008) Admissible predictive density estimation. *Annals of Statistics*, **36** 1156–1170. [MR2418653](#)
- [4] CASELLA, G. AND STRAWDERMAN, W. E. (1981) Estimating a bounded normal mean. *Annals of Statistics*, **9** 870–878. [MR0619290](#)
- [5] FOURDRINIER, D. AND MARCHAND, É. (2010) On Bayes estimators with uniform priors on spheres and their comparative performance with maximum likelihood estimators for estimating bounded multivariate normal means. *Journal of Multivariate Analysis*, **101** 1390–1399. [MR2609500](#)
- [6] GEORGE, E. I., FENG, L. AND XU, X. (2006) Improved minimax predictive densities under Kullback-Leibler loss. *Annals of Statistics*, **34** 78–91. [MR2275235](#)
- [7] GEORGE, E. I. AND XU, X. (2010) Bayesian predictive density estimation. *Frontiers of Statistical Decision Making and Bayesian Analysis. In honor of James O. Berger*, 83–95. Springer.
- [8] HARTIGAN, J. A. (2004) Uniform priors on convex sets improve risk. *Statistics and Probability Letters*, **67** 285–288. [MR2060127](#)
- [9] KIEFER, J. (1957) Invariance, minimax sequential estimation, and continuous time processes. *Annals of Mathematical Statistics*, **28** 573–601. [MR0092325](#)
- [10] KOMAKI, F. (2001) A shrinkage predictive distribution for multivariate normal observables. *Biometrika*, **88** 859–864. [MR1859415](#)
- [11] KUBOKAWA, T. (2005) Estimation of bounded location and scale parameters. *Journal of the Japanese Statistical Society*, **35** 221–249. [MR2328426](#)
- [12] LIANG, F. AND BARRON, A. (2004) Exact minimax strategies for predictive density estimation, data compression and model selection. *IEEE Transactions on Information Theory*, **50** 2708–2726. [MR2096988](#)
- [13] LIANG, R. (1986) A note on the measurability of convex sets. *Arch. Math.*, **47** 90–92. [MR0855142](#)
- [14] MARCHAND, É AND PERRON, F. (2001) Improving on the mle of a bounded normal mean. *Annals of Statistics*, **29** 1066–1081. [MR1869241](#)
- [15] MARCHAND, É AND PERRON, F. (2002) On the minimax estimator of a bounded normal mean. *Statistics and Probability Letters*, **58** 327–333. [MR1923455](#)
- [16] MARCHAND, É AND STRAWDERMAN, W. E. (2010) A unified minimax result for restricted parameter spaces. *Bernoulli*, to appear.
- [17] MURRAY, G. D. (1977) A note on the estimation of probability density functions. *Biometrika*, **64** 150–152. [MR0448690](#)
- [18] NG, V. M. (1980) On the estimation of parametric density functions. *Biometrika*, **67** 505–506. [MR0581751](#)
- [19] ROCKAFELLAR, R. T. (1996) *Convex Analysis*. Princeton University Press.
- [20] TSUKUMA, H. AND KUBOKAWA, T. (2008) Stein’s phenomenon in estimation of means restricted to a polyhedral convex cone. *Journal of Multivariate Analysis*, **99** 141–164. [MR2432325](#)