Estimating a multivariate normal mean with a bounded signal to noise ratio under scaled squared error loss

Othmane Kortbi, Éric Marchand[1]

Université de Sherbrooke, Département de mathématiques, Sherbrooke, QC, CANADA, J1K 2R1

Summary

For normal models with $X \sim N_p(\theta, \sigma^2 I_p)$, $S^2 \sim \sigma^2 \chi^2_k$, independent, we consider the problem of estimating θ under scale invariant squared error loss $\frac{\|d-\theta\|^2}{\sigma^2}$, when it is known that the signal-to-noise ratio $\frac{\|\theta\|}{\sigma}$ is bounded above by m. Risk analysis is achieved by making use of a conditional risk decomposition and we obtain in particular sufficient conditions for an estimator to dominate either the unbiased estimator $\delta_{UB}(X) = X$, or the maximum likelihood estimator $\delta_{ML}(X, S^2)$, or both of these benchmark procedures. The given developments bring into play the pivotal role of the boundary Bayes estimator $\delta_{BU,0}$ associated with a prior on (θ, σ^2) such that $\theta|\sigma^2$ is uniformly distributed on the (boundary) sphere of radius $m\sigma$ centered at the origin, and a non-informative $\frac{1}{\sigma^2}$ prior measure is placed marginally on σ^2. With a series of technical results related to $\delta_{BU,0}$, which relate to particular ratios of confluent hypergeometric functions, we show that, whenever $m \leq \sqrt{p}$ and $p \geq 2$, $\delta_{BU,0}$ dominates both δ_{UB} and δ_{ML}. The finding can be viewed as both a multivariate extension of $p = 1$ result due to Kubokawa (2005) and a unknown variance extension of a similar dominance finding due to Marchand and Perron (2001). Various other dominance results are obtained, illustrations are provided and commented upon. In particular, for $m \leq \sqrt{p/2}$, a wide class of Bayes estimators, which include priors where $\theta|\sigma^2$ is uniformly distributed on the ball of radius $m\sigma$ centered at the origin, are shown to dominate δ_{UB}.

AMS 2010 subject classifications. 62F10, 62F15, 62F30, 62H12, 62C99.
Keywords and phrases: Bayes estimator, Coefficient of variation, Confluent hypergeometric function, Dominance, Estimation, Maximum likelihood, Multivariate normal, Restricted parameter, Signal-to-noise ratio, Squared error loss.

1. Introduction

1.1. The model

We consider the canonical form of the general linear model (e.g., Lehmann and Casella, 1998, section 3.4)

$$X \sim N_p(\theta, \sigma^2 I_p), \quad S^2 \sim \sigma^2 \chi^2_k, \text{ independent}, \tag{1.1}$$

with $p \geq 1$ and $k \geq 1$, which plays a central role in both statistical theory and practice. The corresponding model density supported on $\mathbb{R}^p \times \mathbb{R}_+$ for (X, S^2) is given by

$$\frac{1}{(2\pi\sigma^2)^{p/2}\,\Gamma(\frac{k}{2})\,2^{k/2}} \left(\frac{s^2}{\sigma^2}\right)^{k/2-1} e^{-\frac{1}{2\sigma^2}(\|x-\theta\|^2+s^2)}\,. \tag{1.2}$$

[1]Corresponding author: eric.marchand@usherbrooke.ca

We wish to estimate θ under scale invariant squared error loss

$$L\left(\left(\theta,\sigma^2\right),d\right) = \frac{\|d-\theta\|^2}{\sigma^2}, \tag{1.3}$$

with (θ,σ^2) belonging to the restricted parameter space

$$\Theta\left(m\right) = \left\{\left(\theta,\sigma^2\right) \in \mathbb{R}^p \times \mathbb{R}_+ : \frac{\|\theta\|}{\sigma} \leq m\,,\sigma^2 > 0\right\}\,, \tag{1.4}$$

and m being a known positive constant.

Model (1.1) and the constraint (1.4) arise in normal full rank linear models $Y \sim N_n(Z\beta,\sigma^2 I_n)$ with orthogonal design matrix $Z(n\times p)$; $n>p$; unknown parameter vector $\beta(p\times 1)$, and with the constraint $\frac{\|\beta\|}{\sigma} \leq m$. Indeed, the sufficient statistic $(\hat{\beta},W) = ((Z'Z)^{-1}Z'Y = Z'Y, \|Y - Z\hat{\beta}\|^2)$ has a distribution which matches (1.1) with $X=\hat{\beta}$, $S^2 = W$, $k=n-p$ accompanied by constraint (1.4) with $\theta=\beta$. More generally, where Z is not necessarily orthogonal, the correspondence also arises as above through the constraint $\frac{\|Z\beta\|}{\sigma} \leq m$ by setting $X=(Z'Z)^{1/2}\hat{\beta}$ and $\theta=(Z'Z)^{1/2}\beta$.

As considered and motivated by Kariya, Giri and Perron (1988), as well as Perron and Giri (1990), model (1.1) arises for the curved model setting

$$Y_1,\ldots,Y_n \text{ independent } N_p(\mu,\sigma^2 I_p)\,,\ \text{with } \sigma^2 = \frac{n\,\mu'\mu}{\lambda^2}. \tag{1.5}$$

The correspondence is achieved by considering the sufficient statistic $(\bar{Y},W) = (\frac{1}{n}\sum_{i=1}^n Y_i, tr(\sum_{i=1}^n (Y_i - \bar{Y})(Y_i-\bar{Y})')$ and setting $X=\sqrt{n}\,\bar{Y}$, $S^2 = W$, $\theta=\sqrt{n}\mu$, and $k=(n-1)p$ in (1.1). Constraint (1.4) arises by assuming $\lambda \in [0,m]$ in (1.5), while the findings of Kariya, Giri and Perron (1988), as well as Perron and Giri (1990), relate to known λ and the best equivariant estimator $\delta_\lambda(X,S)$ for loss $\frac{\|d-\mu\|^2}{\mu'\mu}$ (equivalent to 1.3) under the group of transformations $G=\mathbb{R}_+\times O(p)$, $\mathbb{R}_+$ being the multiplicative group of positive real numbers and $O(p)$ being the group of $p\times p$ orthogonal matrices. We will extract some features of the invariance structure and of the estimator $\delta_\lambda(X,S)$ in Subsection 2.2, but refer to the above mentioned papers for additional details.

Viewing $\frac{\|\theta\|}{\sigma}$ as a multivariate version of a signal to noise ratio, the problem can be described as estimating a normal mean with upper bounded signal to noise ratio. Alternatively, as previously described and analyzed by Kubokawa (2005) for $p=1$, the parametric constraint places an upper bound on the reciprocal of the coefficient of variation in absolute value. As described at the end of Section 4, the above restricted parameter space estimation problem relates to two-sample problems with additional information. We also point out that inference problems for restricted parameter spaces such as $\Theta(m)$ in (1.4); but more so for known variance σ^2; have been extensively studied with a substantial amount of accompanying literature, and arise in non-parametric regression problems (e.g., Efromovich, 1999; Johnstone, 2011).

We will be concerned with Bayesian inference in such restricted parameter space problems, which does not, conceptually, present any difficulties as both the prior and the resulting posterior (if it exists) will be adapted and will adapt to the constraints. Assessing the frequentist performance of Bayesian estimators in such situations is, however, considerably more challenging. Such assessments may include, for instance, testing for minimaxity, an evaluation in comparison to a benchmark procedure such as an unrestricted best equivariant estimator (BEE) or a constrained maximum likelihood (ML) estimator, or a study of the frequentist performance of associated Bayesian credible sets (e.g., Mandelkern, 2002; Marchand and Strawderman, 2006, 2012A; Marchand et al., 2008; Roe and Woodroofe, 2000; Zhang and Woodroofe, 2003).

1.2. The estimators and previous findings

A benchmark procedure, which we will focus on, is the maximum likelihood estimator $\delta_{ML}(X, S^2)$. Although $\delta_{ML}(X, S^2)$ will be shown to be inadmissible for $p \geq 2$ (Corollary 3.1), it is nevertheless of interest to describe the structure of dominating estimators with an emphasis on Bayesian procedures. Another benchmark procedure is the unbiased estimator $\delta_{UB}(X, S^2) = X$. The choice $\delta_{UB}(X, S^2)$ is clearly inefficient for our problem, as seen for instance by a straightforward analysis of the risk of linear estimators of the form $\delta_a(X, S^2) = aX; a \in \mathbb{R}$. Indeed, the choices δ_a with $a \in [\frac{m^2-p}{m^2+p}, 1)$ dominate $\delta_{UB}(X, S^2)$ (strictly for $a \in (\frac{m^2-p}{m^2+p}, 1)$) and the minimax procedure in this class, given by the choice $a = \frac{m^2}{m^2+p}$, always dominates $\delta_{UB}(X, S^2)$. We will establish below that $\delta_{ML}(X, S^2)$, which takes into account the constraint in opposition to $\delta_{UB}(X, S^2)$, also dominates $\delta_{UB}(X, S^2)$ for many combinations of (m, p, k) (Corollary 3.1 a), so that it is more challenging to obtain improvements on the former as opposed to the latter. Hence, comparisons with $\delta_{UB}(X, S^2)$ will remain of secondary interest.

Remark 1.1. *(Minimaxity) The above paragraph implies that $\delta_{UB}(X, S^2)$ is never minimax for our problem, in contrast with the case where θ, rather than $\frac{\theta}{\sigma}$, is constrained to a ball of radius m and the loss is the same as in (1.3). Indeed, as shown by Kubokawa (2005) for $p = 1$ and as a consequence of a general minimax result given by Marchand and Strawderman (2012B) applicable for $p \geq 1$, $\delta_{UB}(X, S^2)$ is minimax (but still inadmissible) for the restricted parameter space with $\|\theta\| \leq m$ and $\sigma^2 > 0$.*

We set $S_h = \{u \in \mathbb{R}^p : \|u\| = h\}$ and $B_h = \{u \in \mathbb{R}^p : \|u\| \leq h\}$, respectively, as the p dimensional sphere and ball of radiuses h centered at the origin. We consider priors on the restricted parameter space $\Theta(m)$ admitting the representations:

$$(\mathbf{i})\ \theta|\sigma^2 \sim \text{Uniform}(S_{m\sigma}), \sigma^2 \sim (\sigma^2)^{\frac{l}{2}-1}, \text{ or } (\mathbf{ii})\ \theta|\sigma^2 \sim \text{Uniform}(B_{m\sigma}), \sigma^2 \sim (\sigma^2)^{\frac{l}{2}-1}, \tag{1.6}$$

with $l < p+k$ so that the associated posterior distributions be well defined. These interesting priors are improper, but they are proper with respect to θ for fixed σ^2. The case $l = 0$ is of particular significance since it corresponds to the right Haar invariant case for σ^2. In **(ii)**, with the volume of the ball B_h equal to $\frac{\pi^{p/2}}{\Gamma(\frac{p}{2}+1)} h^p$, we have prior measure $\pi(\theta, \sigma^2)$ proportional to $(\sigma^2)^{\frac{l-p}{2}-1} \mathbb{I}_{B_{m\sigma}}(\theta)\, \mathbb{I}_{(0,\infty)}(\sigma^2)$ so that the case $l = p$ corresponds to the interesting case of the truncation onto the restricted parameter space $\Theta(m)$ of the usual noninformative prior for (θ, σ^2) set on $\mathbb{R}^p \times \mathbb{R}^+$. The Bayes estimators associated with priors in **(i)** and **(ii)** may hence be described as boundary uniform or fully uniform, and we denote these $\delta_{BU,l}$ and $\delta_{U,l}$ respectively.

The above priors and Bayes estimators are analogous to the boundary uniform and fully uniform prior Bayes estimators δ_{BU} and δ_U studied by Marchand and Perron (2001) for the known σ^2 case in model (1.1), with loss (1.3) and the restriction (1.4), where both δ_{BU} and δ_U were proven to dominate the maximum likelihood estimator $\delta_{ML}(X) = (\frac{m}{\|X\|} \wedge 1)X$ for sufficiently small m; and where namely δ_{BU} dominates $\delta_{ML}(X)$, and hence X, under the simple condition $m \leq \sqrt{p}$. Furthermore, a result by Hartigan (2004) applicable for much more general restricted convex parameter spaces, and reviewed along with analogous findings by Marchand and Strawderman (2004), implies that δ_U dominates X for all (m, p).

Returning to our unknown σ^2 case, Kubokawa (2005) provided various dominating estimators of $\delta_{UB}(X, S^2) = X$ for the univariate case $p = 1$, showing **(a)** dominance of $\delta_{U,l}$ for all $m > 0$ and

$l \in [0, k+1]$, as well as **(b)** the dominance of $\delta_{BU,l}$ for $m \leq 1$ and $l \in [0, k+1]$. Hence, these findings may be viewed as unknown σ^2 univariate extensions of some of the existing results in the literature, including some described in the previous paragraph. However, Kukokawa's results are limited to the univariate case and do not apply to either the maximum likelihood estimator $\delta_{ML}(X, S^2)$, nor to linear estimators aX, $a \in [0, 1)$.

In this paper, we extend Kubokawa's findings concerning $\delta_{BU,l}$ to: **(i)** the multivariate case $p > 1$, and to **(ii)** the benchmark δ_{ML}, providing sufficient conditions for dominance. Namely, we show that $\delta_{BU,0}$ dominates $\delta_{ML}(X, S^2)$ whenever $m \leq \sqrt{p}$ and $p \geq 2$ (Corollary 3.1). This yields a striking parallel with Marchand and Perron's dominance result for the known σ^2 case. We provide various other dominance results, present illustrations and accompanying commentary. Namely, for small enough parameter spaces (precisely $m \leq \sqrt{p/2}$), we obtain a universal dominance result showing that a vast class of Bayes estimators, which includes $\delta_{U,l}$ and $\delta_{BU,l}$ for $l \leq 0$, dominate δ_{UB}.

Our methods also depart from the methods used by Kubokawa in the univariate case applicable to $\delta_{BU,l}$. Key features of Kubokawa's results are techniques previously introduced by Kubokawa himself (Kubokawa, 1994, Integral Expression for Risk Difference IERD), as well by Marchand and Strawderman (2005). With extensions of these techniques challenging to obtain and elusive in the multivariate case, with the absence of results applicable to other estimators such as the $\delta_{ML}(X, S^2)$, we rather exploit a conditional risk decomposition on a maximal invariant $T = \frac{\|X\|^2}{S^2}$ in a similar fashion as Marchand and Perron (2001) and Moors (1985). The deficiency of $\delta_{ML}(X, S^2)$ (and of $\delta_{UB}(X, S^2)$) is revealed as one of providing estimates that are too far from the origin. Even when $m > \sqrt{p}$, we will show that dominating estimators are obtained by projecting towards the Bayes estimator $\delta_{BU,0}$. The Bayes estimator $\delta_{BU,0}$ coincides with the best equivariant estimator δ_m, and analytical results for $\delta_{BU,0}$, which we will describe and make use of, were previously given by Kariya, Giri and Perron (1988), as well as Perron and Giri (1990).

The remainder of this paper is organized as follows. Section 2 contains key features and properties of the invariance structure and the risk function of equivariant estimators, as well as various expressions and key properties relative to the Bayes estimators $\delta_{BU,l}$ and the maximum likelihood estimator $\delta_{ML}(X, S^2)$. The dominance results are presented in Section 3 along with various illustrations and comments. Concluding remarks are given in Section 4. Finally, Section 5 is an Appendix with additional details and proofs or earlier results.

2. Preliminary results

2.1. The invariance structure and risk function

The challenge here is to obtain dominating estimators under loss (1.3) that capitalize on the parametric information in (1.4), as measured by the risk function

$$R((\theta, \sigma^2), \delta) = \frac{1}{\sigma^2} E_{\theta,\sigma^2} \left[\|\delta(X, S) - \theta\|^2 \right], (\theta, \sigma^2) \in \Theta(m) \,. \tag{2.7}$$

With the problem being invariant with respect to the group of transformations $\mathbb{R}_+ \times O(p) = \{(x, s^2) \mapsto (bQx, b^2 s^2), b \in \mathbb{R}_+, Q \in O(p)\}$, where $O(p)$ is the set of all $p \times p$ orthogonal matrices, we focus on (nonrandomized) equivariant estimators, which as shown by Kariya, Giri and Perron

(1988), or Perron and Giri (1990), or Giri (2004), are of the form

$$\delta_h(X, S^2) = h(\frac{\|X\|^2}{S^2})X\,, \tag{2.8}$$

for some measurable function $h(\cdot)$. Equivariant estimators are thus collinear to X and conveniently represented by the corresponding multiplier h which controls the degree of expansion or shrinkage with respect to X, depending on (X, S^2) only through the maximal invariant statistics $T = \frac{\|X\|^2}{S^2}$. This class of estimators include δ_{UB} ($h \equiv 1$), linear estimators aX ($h \equiv a$), δ_{ML} (see Lemma 2.5), generalized Bayes estimators $\delta_{BU,l}$ (see Theorem 2.1), and more generally Bayes estimators with a spherically symmetric structure which includes $\delta_{U,l}$ (see Theorem 3.7). Equivariant estimators have a risk function in (2.7) depending on the unknown parameters only through the maximal invariant $\lambda = \frac{\|\theta\|}{\sigma} \in [0, m]$. By a slight abuse of notation, we will write this risk $R(\lambda, \delta_h)$. Here is a useful representation for the risk $R(\lambda, \delta_h)$ of equivariant estimators achieved by conditioning on the maximal invariant statistic T and highlighting the key role of the best equivariant estimator $\delta_\lambda(X, S^2)$ (for $\frac{\|\theta\|}{\sigma} = \lambda$).

Lemma 2.1. *For model (1.1) and loss (1.3), we have*

$$R(\lambda, \delta_h) = \lambda^2 + E_\lambda \left[a_\lambda(T)\left\{(h(T) - h_\lambda(T))^2 - h_\lambda^2(T)\right\}\right]\,, \tag{2.9}$$

with $a_\lambda(T) = E_{\theta,\sigma^2}\left(\frac{X'X}{\sigma^2}\middle| T\right)$ *and* $h_\lambda(T) = \frac{E_{\theta,\sigma^2}\left(\frac{\theta'X}{\sigma^2}\middle|T\right)}{a_\lambda(T)}$.

Proof. The result is immediate by writing the loss as $\frac{\|h(t)x-\theta\|^2}{\sigma^2} = h^2(t)\frac{x'x}{\sigma^2} + \lambda^2 - 2h(t)\frac{\theta'x}{\sigma^2}$ and decomposing the risk as $E[L((\theta, \sigma^2), \delta_h)] = E^T\{E[L((\theta, \sigma^2), \delta_h)|T]\}$. □

Remark 2.2. *As seen by (2.9) above, the risk of δ_h is constant for the restriction $\frac{\|\theta\|}{\sigma} = \lambda$ and the optimal procedure is given by the BEE $\delta_\lambda(X, S^2) = h_\lambda(T)X$. Furthermore, the performance of the estimator $\delta_h(X, S^2)$ is related to the average proximity of h to h_λ with respect to T and as weighted by a_λ. We also point out, as seen in the above proof, that the conditional risk $E[L((\theta, \sigma^2), \delta_h)|T]$ depends on (θ, σ^2) only through the maximal invariant parameter $\lambda = \frac{\|\theta\|}{\sigma}$.*

Kariya, Giri and Perron (1988), as well as Perron and Giri (1990), gave an explicit expression for $\delta_\lambda(X, S^2)$. It is reproduced below in Theorem 2.1 where we derive a general expression for the Bayes estimators $\delta_{BU,l}$. The estimators $\delta_m(X, S^2)$ and $\delta_{BU,0}$ necessarily coincide by virtue of general relationships between best equivariant estimators and Bayes estimators with respect to Haar right invariant priors (e.g., Eaton, 1989).

2.2. Bayes estimators

We begin here with a general expression for Bayes estimators associated with priors of the form

$$\pi\left(\theta, \sigma^2\right) = \pi_{\sigma^2}(\theta)\left(\sigma^2\right)^{l/2-1}, \quad l < k + p\,, \tag{2.10}$$

where $\pi_{\sigma^2}(\cdot)$ is for fixed σ^2 a (proper) density with respect to a finite measure ν_{σ^2} supported on, or a subset of, the ball $B(m\sigma)$.

Lemma 2.2. *For model (1.1) and loss (1.3), Bayes estimators with respect to priors as in (2.10) are equal to*

$$\delta_\pi\left(x, s^2\right) = x + \frac{\int_0^\infty \sigma^{l-k-p-2}\; e^{-\frac{s^2}{2\sigma^2}}\left(\nabla_x\, m^\pi_{\sigma^2}\left(x\right)\right) d\sigma^2}{\int_0^\infty \sigma^{l-k-p-4}\; e^{-\frac{s^2}{2\sigma^2}}\, m^\pi_{\sigma^2}\left(x\right)\, d\sigma^2}\,, \tag{2.11}$$

where ∇_x denotes the gradient vector with respect to x, and $(2\pi\sigma^2)^{-p/2}\, m^\pi_{\sigma^2}$ is the marginal density of $X|\sigma^2$ with

$$m^\pi_{\sigma^2}\left(x\right) = \int_{B_{m\sigma}} e^{-\frac{1}{2\sigma^2}\|x-\theta\|^2}\pi_{\sigma^2}\left(\theta\right)\, d\nu_{\sigma^2}(\theta)\,. \tag{2.12}$$

Proof. With the density in (1.2), we have

$$\begin{aligned}
\delta_\pi\left(x, s^2\right) &= \frac{E\left[\,\theta/\sigma^2\,|x, s^2\right]}{E\left[\,1/\sigma^2\,|x, s^2\right]} && (2.13)\\
&= \frac{\int_0^\infty \int_{B_{m\sigma}} \theta\; e^{-\frac{1}{2\sigma^2}\|x-\theta\|^2} e^{-\frac{s^2}{2\sigma^2}}\;\pi_{\sigma^2}\left(\theta\right)\sigma^{l-k-p-4}\, d\nu_{\sigma^2}(\theta)\, d\sigma^2}{\int_0^\infty \int_{B_{m\sigma}} e^{-\frac{1}{2\sigma^2}\|x-\theta\|^2} e^{-\frac{s^2}{2\sigma^2}}\;\pi_{\sigma^2}\left(\theta\right)\sigma^{l-k-p-4}\, d\nu_{\sigma^2}(\theta)\, d\sigma^2} && (2.14)\\
&= x + \frac{\int_0^\infty \sigma^{l-k-p-4}\; e^{-\frac{s^2}{2\sigma^2}} \int_{B_{m\sigma}} \left(\theta - x\right)\; e^{-\frac{1}{2\sigma^2}\|x-\theta\|^2}\pi_{\sigma^2}\left(\theta\right)\; d\nu_{\sigma^2}(\theta)\, d\sigma^2}{\int_0^\infty \sigma^{l-k-p-4}\; e^{-\frac{s^2}{2\sigma^2}} \int_{B_{m\sigma}} e^{-\frac{1}{2\sigma^2}\|x-\theta\|^2}\pi_{\sigma^2}\left(\theta\right)\; d\nu_{\sigma^2}(\theta)\, d\sigma^2}\,.
\end{aligned}$$

The result follows since

$$\sigma^2\nabla_x m^\pi_\sigma\left(x\right) = \int_{B_{m\sigma}} \left(\theta - x\right)\; e^{-\frac{1}{2\sigma^2}\|x-\theta\|^2}\pi_{\sigma^2}\left(\theta\right)\; d\nu_{\sigma^2}(\theta)\,, \tag{2.15}$$

with an interchange of ∇_x and $\int$. □

Lemma 2.2 applies to the boundary uniform and fully uniform priors in (1.6), among others. We pursue here with useful expressions for the former, which will also serve as a benchmark for other Bayesian estimators (see Lemma 3.7) in Section 3.

Theorem 2.1. *We have $\delta_{BU,l}\left(X, S^2\right) = h(m, l, T)\, X$, with $T = \frac{\|X\|^2}{S^2}$ and*

$$h(m, l, t) = \frac{m^2}{p}\frac{F\left(\frac{k+p-l}{2} + 1, \frac{p}{2} + 1, \frac{m^2 t}{2(1+t)}\right)}{F\left(\frac{k+p-l}{2} + 1, \frac{p}{2}, \frac{m^2 t}{2(1+t)}\right)}\,, \tag{2.16}$$

F being the confluent hypergeometric function given by $F(a, b, z) = \sum_{i=0}^\infty \frac{(a)_i}{(b)_i}\frac{z^i}{i!}$, $z \in \mathbb{R}$, with $(c)_i = \prod_{j=0}^{i-1}(c + j)$ for $i = 1, 2, \ldots,$ and $(c)_0 = 1$.

Proof. See Appendix. □

Ratios of confluent hypergeometric functions and their properties hence play an important role here, as witnessed by Theorem 2.1's representation of $h(m, l, t)$. Furthermore, key analytical properties of $\delta_{BU,l}$ (Lemmas 2.4 and 2.6), of other Bayesian estimators (Lemma 3.7) and corresponding risk function comparisons (Section 3) will hinge on various properties of such ratios, as those given by the following intermediate result.

Lemma 2.3. *For all $a > 0$, $b > 0$, $z > 0$, and $c \in \{0, 1\}$,*

(a) *the function $K(\cdot) = \frac{F(a-c+1,\, b-c+1,\, \cdot)}{F(a+1,\, b,\, \cdot)}$ is strictly decreasing on $[0, \infty)$, with $\lim_{z\to\infty} K(z) = 0$;*

(b) *the function $z\, \frac{F(a+1,\, b+1,\, z)}{F(a+1,\, b,\, z)}$ is strictly increasing in z, $z \in (0, \infty)$;*

(c) *the function $H(\cdot) = \frac{F(\cdot,\, b+1,\, z)}{F(\cdot,\, b,\, z)}$ is strictly decreasing on $(0, \infty)$.*

Proof. See Appendix. □

By making use of these above properties, we derive the following analytical results concerning Bayes estimators $\delta_{BU,l}$ given in Theorem 2.1. More precisely, we now describe how $h(\lambda, l, t)$ varies with respect to (λ, l, t).

Lemma 2.4. **(a)** *The function $h(\lambda, l, \cdot)$ is, for $\lambda > 0$, $l < k + p$, strictly decreasing on $[0, \infty)$ with $\sup_{t\geq 0} h(\lambda, l, t) = h(\lambda, l, 0) = \frac{\lambda^2}{p}$ and $\inf_{t\geq 0} h(\lambda, l, t) = \frac{\lambda^2}{p} \frac{F\left(\frac{k+p-l}{2}+1, \frac{p}{2}+1, \frac{\lambda^2}{2}\right)}{F\left(\frac{k+p-l}{2}+1, \frac{p}{2}, \frac{\lambda^2}{2}\right)}$;*

(b) *the function $h(\cdot, l, t)$ is, for $t > 0$, $l < k + p$, strictly increasing on $[0, \infty)$:*

(c) *the function $h(\lambda, \cdot, t)$ is, for $\lambda > 0$, $t > 0$, strictly increasing on $(-\infty, k+p)$; and consequently $h(\lambda, l, t) > h(\lambda, 0, t)$ for $l \in (0, p + k)$.*

Proof. We use (2.16) throughout. In part **(a)**, the monotonicity follows from part (a) of Lemma 2.3 setting $c = 0$, and implies $\sup_{t\geq 0} h(\lambda, l, t) = h(\lambda, l, 0)$ and $\inf_{t\geq 0} h(\lambda, l, t) = \lim_{t\to\infty} h(\lambda, l, t)$ yielding the results. Part **(b)** follows from part (b) of Lemma 2.3 by setting $z = \frac{\lambda^2 t}{2(1+t)}$, $a = \frac{k-p+l}{2}$, and $b = \frac{p}{2}$. Finally, part **(c)** is a consequence of part (c) of Lemma 2.3 and expression (2.16). □

Remark 2.3. *As shown above in Lemma 2.4, $h(\lambda, l, t)$ increases in l for fixed (λ, t) so that the amount of shrinkage or expansion of the boundary uniform estimators is controlled by the choice of the l in the prior (1.6). Namely, in comparison to the benchmark $\delta_{BU,0}$, $\delta_{BU,l}$ will either expand for $k+p > l > 0$, or shrink for $l < 0$, with the case $l \to -\infty$ leading to $\delta_{BU,l}$ shrinking to the degenerate estimator $\delta \equiv 0$. These properties will lead directly to precise risk comparisons in Section 3.*

2.3. The maximum likelihood estimator

Here, we determine the maximum likelihood estimator δ_{ML} and some of its analytical properties. In particular, we show that $\delta_{ML}(X, S^2)$ is equivariant, hence of the form $h_{ML}(T)X$, and determine a useful lower envelope for $h_{ML}(\cdot)$ (Lemma 2.6). The latter will be indicative of the fact that δ_{ML} provides estimates that are too far from the origin and will be exploited to obtain dominance results in Section 3.

Lemma 2.5. *For model (1.1) and parameter space $\Theta(m)$ in (1.4), the maximum likelihood estimator of θ is given by $\delta_{ML}(X, S^2) = h_{ML}(T)X$, with $T = \frac{\|X\|^2}{S^2}$ and*

$$h_{ML}(t) = \left(\frac{m^2}{2(p+k)} \left(\sqrt{1 + 4\frac{(k+p)}{m^2}\frac{1+t}{t}} - 1 \right) \wedge 1 \right). \tag{2.17}$$

Proof. See Appendix. □

Remark 2.4. *Observe that* $\|\delta_{ML}(x,s^2)\| \leq \|x\|$ *with strict inequality for* $\frac{x'x}{s^2} > \frac{m^2}{p+k}$, *so that the maximum likelihood shrinks the unbiased estimator towards the origin. Moreover, the amplitude of this shrinkage as measured by* $h_{ML}(t)$ *is seen to be increasing with respect to* t *(i.e.,* $h_{ML}(t)$ *decreases in* t*), with* $\lim_{t\to\infty} h_{ML}(t) = \frac{\sqrt{1+4\gamma}-1}{2\gamma}$, $\gamma = \frac{p+k}{m^2}$.

Now, for various risk comparisons involving the estimators δ_{ML} and the Bayes estimators $\delta_{BU,l}$, the inequalities of the following lemma will be pivotal.

Lemma 2.6. **(a)** *We have* $h(m,0,t) < h_{ML}(t)$ *for* $t > \frac{m^2}{p+k}$, *whenever* $p \geq 2, k \geq 2$, *and whenever* $k = 1$ *and* $m \leq \sqrt{p}$;

(b) *For* $p \geq 2$ *and for* $p = 1, k = 1$, *the condition* $m \leq \sqrt{p}$ *is necessary and sufficient so that* $h(m,0,t) \leq h_{ML}(t)$ *for all* $t \geq 0$ *and, in such cases, equality is attained if and only if* $t = 0$ *and* $m = \sqrt{p}$.

Proof. Since $h_{ML}(\cdot)$ is constant on $[0, \frac{m^2}{p+k}]$, and $h(m,0,\cdot)$ is decreasing and bounded above by $\frac{m^2}{p}$ on $[0, \frac{m^2}{p+k}]$ by virtue of Lemma 2.4, part **(a)** implies part **(b)**. For proving part **(a)**, by making use of representations (2.17) and (2.16), it suffices to verify that

$$\frac{F\left(\frac{k+p}{2}+1, \frac{p}{2}+1, z\right)}{F\left(\frac{k+p}{2}+1, \frac{p}{2}, z\right)} < \frac{p}{2(p+k)} \left(\sqrt{1+\frac{2(k+p)}{z}} - 1\right),$$

or equivalently

$$R(z) = \frac{F\left(\frac{k+p}{2}+1, \frac{p}{2}, z\right)}{F\left(\frac{k+p}{2}+1, \frac{p}{2}+1, z\right)} > \frac{z}{p} \left(\sqrt{1+\frac{2(k+p)}{z}} + 1\right), \tag{2.18}$$

for all $p \geq 2, k \geq 2, z > 0$ or for $k = 1$, $m \leq \sqrt{p}$ and $z = \frac{m^2 t}{2(1+t)} \in (\frac{m^4}{2(m^2+p+k)}, \frac{m^2}{2})$. With (2.18) established in the Appendix, the proof is complete. □

3. Dominance results

Following Lemma 2.1's representation of risks, and the analytical properties of $h(m,l,\cdot)$ and $h_{ML}(\cdot)$ worked out in Lemmas 2.4 and 2.6, we obtain various risk comparisons and dominance results which will be consequences of the following sufficient conditions for dominance. These conditions are a follow-up to Lemma 2.1's risk decomposition, make use of the conditional risk with respect to T, and are obtained with conditions on h_1 and h_2 for which δ_{h_2} has conditional risk smaller than or equal to δ_{h_1} a.e. T for all $\lambda \in [0,m]$.

Theorem 3.2. *Consider equivariant estimators* δ_h *as in (2.8) for estimating* θ, *model (1.1), loss (1.3) and the parameter space* $\Theta(m)$.

(a) *A sufficient condition for* δ_{h_2} *to dominate* δ_{h_1} *(*$\delta_{h_2} \neq \delta_{h_1}$*) is* $h_1(t) \geq h_2(t)$ *and* $\frac{h_2(t)+h_1(t)}{2} \geq h(m,0,t)$ *for all* $t > 0$;

(b) *Let $S = \{t \in \mathbb{R}_+ : h_1(t) > h(m,0,t)\}$. Let $\underline{S}$ be a subset (not strict) of S such that $\nu(\underline{S}) > 0$, where ν is the Lebesgue measure on $\mathbb{R}$. Then δ_{h_1} is inadmissible, and dominated by δ_h with $h(t) = h_1(t) \wedge h(m,0,t)$, as well as by any δ_h, $h(t) = h_1(t)\mathbb{I}_{\underline{S}^c}(t) + h^*(t)\mathbb{I}_{\underline{S}}(t)$, $h \neq h_1$, with $2h(m,0,t) - h_1(t) \leq h^*(t) \leq h_1(t)$ for all $t \in \underline{S}$.*

(c) *If $h_1(t) \geq h(m,0,t)$ for all $t \geq 0$, $h_1(\cdot) \neq h(m,0,\cdot)$, then the boundary uniform Bayes estimator $\delta_{BU,0}$ given in (2.16) dominates δ_{h_1}.*

Proof. Parts **(b)** and **(c)** are consequences of part **(a)** by setting $h_2 \equiv h$ (as given) and $h_2(\cdot) \equiv h(m,0,\cdot)$ respectively. There remains part **(a)**. From (2.9), we obtain

$$R(\lambda,\delta_{h_1}) - R(\lambda,\delta_{h_2}) = E_\lambda\left[a_\lambda(T)\{(h_1(T)-h_2(T)) \times (h_1(T)+h_2(T)-2h(\lambda,0,T)\}\right] \quad (3.19)$$
$$= E_\lambda\left[a_\lambda(T)\,\Delta_\lambda(T)\right] \text{ (say)}, \quad (3.20)$$

where $a_\lambda(\cdot)(>0)$ is given in Lemma 2.1. Therefore, the conditions on h_1 and h_2, along with part **(b)** of Lemma 2.4 which implies $h(\lambda,0,\cdot) < h(m,0,\cdot)$ for $\lambda < m$, force $\Delta_\lambda(T) \geq 0$ with probability one for all $\lambda \in [0,m]$. Finally, dominance indeed occurs with $\Delta_\lambda(t) > 0$ for all $\lambda \in [0,m)$ and $t \in \{u : h_1(u) > h_2(u)\}$. □

Corollary 3.1. **(a)** *For (m,p,k) such that (i) $p \geq 2, k \geq 2$, or (ii) $k = 1, m \leq \sqrt{p}$, δ_{ML} dominates the unbiased estimator X.*

(b) *For $m \leq \sqrt{p}$, $\delta_{BU,0}$ dominates δ_{ML} whenever $p \geq 2$ or $(p,k) = (1,1)$.*

(c) *For all (m,p) and $l \in (0,k+p)$, $\delta_{BU,l}$ is inadmissible and dominated by $\delta_{BU,0}$.*

(d) *For all (m,p) with $m > \sqrt{p}$, such that $p \geq 2, k \geq 2$, the estimator δ_{ML} is inadmissible and dominated by the truncation δ_h, with $h(t) = h_{ML}(t) \wedge h(m,0,t)$.* [2]

Proof. **(a)** We apply part **(b)** of Theorem 3.2 with $h_1 \equiv 1$, $\underline{S} = (\frac{m^2}{p+k}, \infty)$, and $h^* \equiv h_{ML}$. Recall that $h_{ML}(t) < 1$ for $t > \frac{m^2}{p+k}$ as pointed out in Remark 2.4. Under the given conditions on (m,p,k), part **(a)** of Lemma 2.6 tells us that $S = \{t \in \mathbb{R}_+ : 1 > h(m,0,t)\}$ contains $\underline{S}$ and that $h(m,0,t) \leq h_{ML}(t) < h_1(t)$ for all $t \in \underline{S}$, from which the result follows.
(b) Under the assumptions on (m,p,k), part **(b)** of Lemma 2.6 tells us that $h_{ML}(\cdot) \geq h(m,0,\cdot)$ and the result is hence an immediate consequence of part **(c)** of Theorem 3.2.
(c) Given part **(c)** of Lemma 2.4, the result follows as an application of part **(c)** of Theorem 3.2.
(d) Follows from part **(b)** of Theorem 3.2 with $h_1(\cdot) = h_{ML}(\cdot)$, $h^*(\cdot) = h(m,0,\cdot)$ and $\underline{S} = S = (\frac{m^2}{p+k}, \infty)$, and by making use of part **(a)** of Lemma 2.6. □

Remark 3.5. *The key finding above is the dominance of the Bayes estimator $\delta_{BU,0}$ over δ_{ML} for $m \leq \sqrt{p}$ and $p \geq 2$.* [3] *For cases where $m > \sqrt{p}$, $p,k \geq 2$, δ_{ML} is still inadmissible and explicit dominating estimators are available from part* **(d)** *of Corollary 3.1. We believe and conjecture that these results extend to the univariate case, but we have been unable to establish the inequality in part* **(a)** *of Lemma 2.6, which is a critical element of the analysis. Insofar as δ_{ML} dominates δ_{UB} for $p \geq 2, k \geq 2$, and $\delta_{BU,0}$ dominates δ_{ML} for $p \geq 2, m \leq \sqrt{p}$ (as shown above in Corollary 3.1), it is immediate that $\delta_{BU,0}$ dominates δ_{UB} for $p \geq 2, k \geq 2, m \leq \sqrt{p}$ which represents an extension of Kubokawa's (2005) univariate result showing that $\delta_{BU,0}$ dominates δ_{UB} for $p = 1$ and $m \leq 1$.*

[2] In both **(c)** and **(d)**, further dominating procedures can be derived using part **(b)** of Theorem 3.2.

[3] For completeness, part (b) of Corollary 3.1 handles the very special case $p = k = 1$.

Example 3.1. *Part* **(b)** *of Theorem 3.2 indicates clearly that equivariant estimators* δ_h *with* h *taking values that are too large on a subset of* $\mathbb{R}_+$*, such as* δ_{ML}*, are inefficient and can be improved upon by projecting towards the benchmark* $h(m, 0, \cdot)$*. Otherwise said, the function* $h(m, 0, \cdot)$ *provides an upper envelope for a complete class of estimators. As an example, taking* $(p, k, m) = (5, 10, 2)$*, Figure 1 shows the multipliers and risk functions of* δ_{UB}*,* δ_{ML} *and* $\delta_{BU,0}$ *as a function of* $\lambda = \frac{\|\theta\|}{\sigma} \in [0, 2]$*. Here* $m \leq \sqrt{p}$*, so that the ordering of the multipliers and risks is clearly dictated, namely by Lemma 2.6 and Corollary 3.1. We see that* $h_{UB} \equiv 1$ *is much too large, with very poor risk performance. The estimator* δ_{ML} *fares better but the gains provided by the Bayes estimator* $\delta_{BU,0}$ *are nevertheless important; in relative terms ranging from a maximum of over 50% at* $\lambda = 0$ *to a minimum of around 12% at the boundary* $\lambda = 2$*.*

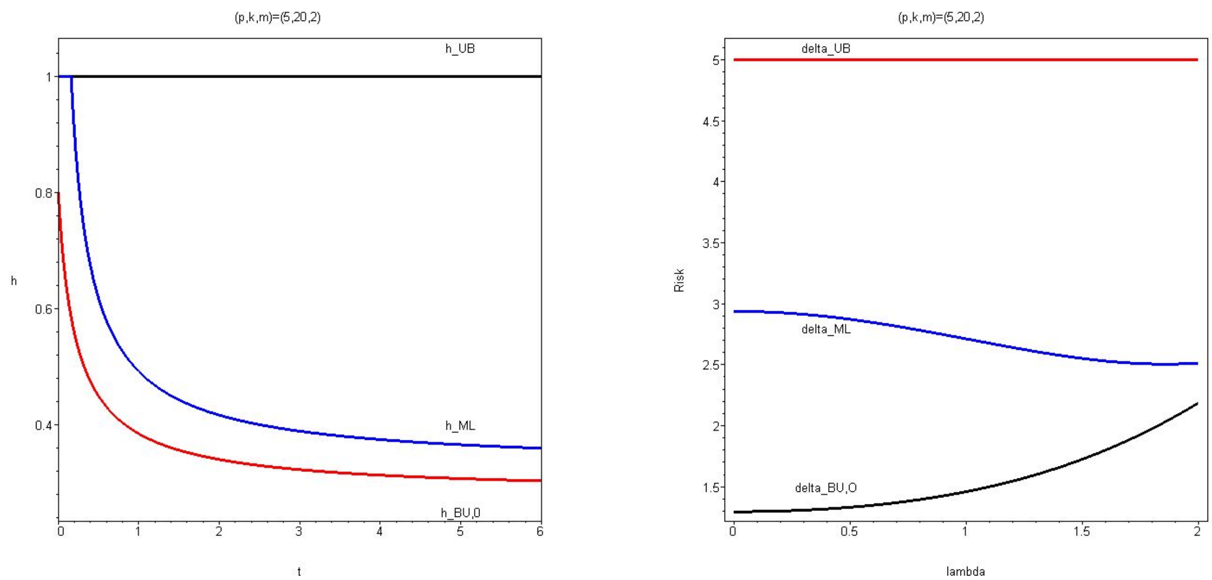

Figure 1: Multipliers and risks of δ_{UB}, δ_{ML} and $\delta_{BU,0}$ for $(p, k, m) = (5, 20, 2)$.

Example 3.2. *Further risk function comparisons are presented in Figure 2 for other combinations of* (p, k, m)*. In both cases, we do not have* $m \leq \sqrt{p}$ *and the dominance findings for the boundary Bayes estimator* $\delta_{BU,0}$ *do not apply, and there is also no guarantee that it performs satisfactorily from a frequentist risk point of view, even in comparison to the unbiased estimator. In one of the cases (with* $p = 5$ *and* $m = 3$*), the numerical results indicate that* $\delta_{BU,0}$ *performs quite well in comparison to* δ_{ML} *with dominance and relative gains between 5% and 20%. But notice how the gains are less impressive than in Figure 1 where* m *is smaller and the dimension* $p = 5$ *is the same. In the other case* ($p = 3, m = 3$)*, where the ratio of* m *relative to* $\sqrt{p}$ *is larger, not only do the findings not apply, notwithstanding the dominance result applicable to the truncation given in part* **(d)** *of Corollary 3.1, but the risk performance of* $\delta_{BU,0}$ *is arguably quite poor. More research is thus required on alternative Bayes estimators, especially when* $m > \sqrt{p}$*. The deficiency of* $\delta_{BU,0}$ *lies in the fact that it expands too much. Here, as an example, we have* $m^2/p = (3^2)/3 = 3$ *so that* $\delta_{BU,0}(x, s^2) \approx 3x$ *when* x *is in a neighbourhood of* 0*, which leads to poor estimates when* $\|\theta\|/\sigma$ *is small. Other priors in (2.10) with* $l = 0$ *will shrink* $\delta_{BU,0}$ *towards the origin. One such class of choices, studied in the known variance case by Fourdrinier and Marchand (2010), are obtained by taking* π_{σ^2} *in (2.10) to be a uniform density on the sphere of radius* $\alpha\sigma$*, for all* $\sigma > 0$ *with* $\alpha \in [0, m)$*. Of course, for such priors, Theorem 2.16 provides an expression for the Bayes estimator by replacing* m *by* α*. Here, we illustrate one such choice with* $\alpha = 2.5$ *and the numerical evaluation indicates* *quite satisfactory performance with a significant improvement on $\delta_{BU,0}$ near the centre and on a large part of the parameter space, with a slightly worse performance near or on the boundary.*

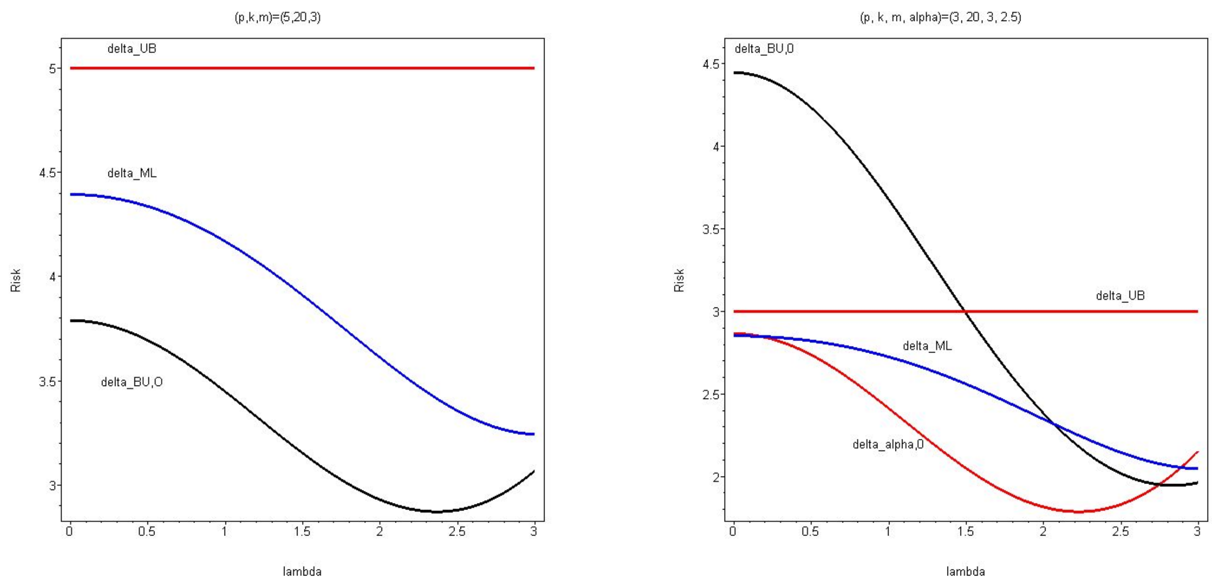

Figure 2: Risks of various estimators : unbiased (δ_{UB}), maximum likelihood (δ_{ML}), boundary uniform ($\delta_{BU,0}$) and Bayes with $\theta|\sigma^2$ uniform on sphere of radius $\alpha\sigma$ ($\delta_{\alpha,0}$), for $(p,k,m) = (5,20,3)$ and $(p,k,m,\alpha) = (3,20,3,2.5)$

We conclude with a general representation for Bayesian estimators and a universal dominance result applicable to quite small parameter spaces (with $\frac{\|\theta\|}{\sigma} \leq m \leq \sqrt{p/2}$) which focuses quite clearly on the inadequacy of the unbiased estimator and lies in continuity with its inefficiency among linear estimators described in the introduction and the above risk comparisons.

Lemma 3.7. *For priors as in (2.10) with spherically symmetric densities π_{σ^2} for all σ^2, Bayes estimators are equivariant, of the form $\delta_\pi(x,s^2) = h_\pi(t)\,x$, with* **(i)** $0 \leq h_\pi(t) \leq h(m,l,t)$ *for all $t > 0$, and with* **(ii)** $h_\pi(\cdot) \leq h(m,0,\cdot)$ *for all $l < 0$.*

Proof. Part **(ii)** is immediate from part **(i)** and part **(c)** of Lemma 2.4. For **(i)**, first observe that Bayes estimators in (2.13) are also posterior expectations $E_{\pi^*}(\theta|x,s^2)$ for prior measures $\pi^*(\theta,\sigma^2) = \pi_{\sigma^2}(\theta)(\sigma^2)^{l/2-2}$ (as expressed in equation 2.14). Now, proceeding as in Marchand and Perron (2001, Theorem 4), with spherically symmetric choices, the prior densities π_{σ^2} in (2.10) admit the representation $\theta|\sigma^2 =^d RU$, where R is supported on $[0,m]$, U is uniformly distributed on the sphere S_σ, and R and U are independent (conditional on σ^2). Hence, by making use of the developments in Theorem 2.1, we have $\delta_\pi(x,s^2) = E_{\pi^*}(\theta|x,s^2) = E[RE(U|R,x,s^2)] = E[h(R,l,t)|x,s^2]\,x$, where $h(r,l,t)$ is given in (2.16) and $R|x,s^2$ is the posterior distribution of $\|\theta\|$ associated with π^*. Since this posterior distribution is supported on $[0,m]$, the result follows since $h(\cdot,l,t)$ is nonnegative and increasing for all (l,t) by virtue of part **(b)** of Lemma 2.4. □

Theorem 3.3. *Whenever $m \leq \sqrt{p/2}$, all equivariant estimators $\delta_h(x,s^2) = h(t)x$ with $0 \leq h(t) \leq h(m,0,t)$ for all $t > 0$ dominate the unbiased estimator X. These include all Bayes estimators with respect to a prior π as in (2.10) with spherically symmetric densities π_{σ^2} for all σ^2 and $l \leq 0$.*

Proof. Applying part **(a)** of Theorem 3.2 with $h_1 \equiv 1$ and $h_2 \equiv h$ with the given assumptions on h, m and p, we have

$$\frac{h(t)+1}{2} \geq \frac{1}{2} \geq \frac{m^2}{p} \geq h(m,0,t)\,,$$

for all $t > 0$, with the rightmost inequality a consequence of part **(a)** of Lemma 2.4. The proof is complete by observing that the inclusion of the Bayesian estimators among the dominating estimators is a consequence of Lemma 3.7. □

Remark 3.6. *A similar result can be derived for the same loss and the known variance model* $X \sim N_p(\theta, \sigma^2 I_p)$, $\frac{\|\theta\|}{\sigma} \leq m$ *(known* m, σ^2*). This corresponds to the setting studied by Marchand and Perron (2001) but they actually establish a universal dominance result (also see Fourdrinier and Marchand, 2010) applicable for the more challenging problem of dominating the corresponding maximum likelihood estimator given by* $(m\sigma \wedge \|x\|)\frac{x}{\|x\|}$*. The above dominance results involving the unbiased estimator may not be that surprising given that even the trivial estimator* $\delta \equiv 0$ *dominates* X *whenever* $m \leq \sqrt{p}$*. But, we have still provided a method of proof applicable to a very large class of Bayesian estimators. And, our findings elsewhere relate as well to* δ_{ML}*.*

4. Concluding Remarks

For estimating a multivariate normal mean with an upper bounded signal to noise ratio $\frac{\|\theta\|}{\sigma}$, we have provided dominance results which can viewed as both multivariate extensions of results obtained by Kubokawa (2005), and unknown variance extensions of results obtained by Marchand and Perron (2001). In opposition to similar extensions for Stein estimation, the presence of the unknown scale here leads to challenges in describing Bayes estimators and some of their analytical properties which ultimately relate to frequentist risk performance.

We have focused mostly on boundary Bayes estimators and the benchmark maximum likelihood estimator deriving the elegant result that $\delta_{BU,0}$ dominates δ_{ML} for $p \geq 2$ and $m \leq \sqrt{p}$. As illustrated theoretically and numerically, and analogously to the known σ^2 case, the relative merits of the boundary Bayes procedure seem to fairly well correlate with the ratio of the radius m relative to $\sqrt{p}$ (see as well Marchand and Perron, 2001; Fourdrinier and Marchand, 2010; Kortbi and Marchand, 2012).

More research is required to assess the performance of other Bayes estimators and namely to propose more attractive choices when m is larger relative to $\sqrt{p}$. Such alternatives include fully uniform Bayes estimators. In this regard, an interesting question is whether the estimator $\delta_{U,0}$ dominates for all (m,p) on X with an affirmative answer representing an unknown variance extension of Hartigan's (2004) result in the particular case of balls.

Several other related problems or issues are also of interest. As an example, our findings do not address directly the issues of minimaxity and admissibility of Bayesian estimators (notwithstanding part (c) of Corollary 3.1). But it seems plausible and we conjecture that $\delta_{BU,0}$ is minimax for small enough m as suggested by its risk in Figure 1 with the maximum risk attained on the boundary and since $\delta_{BU,0}$ is quite likely a candidate to be an extended Bayes procedure.

Finally, we point out that the results obtained here are applicable to two-sample problems with additional information as described by Marchand and Strawderman (2004) and Marchand, Jafari Jozani and Tripathi (2012). These involve independently distributed observables $X_1 \sim N_p(\theta_1, \sigma'^2), X_2 \sim$

$N_p(\theta_2, \sigma'^2), S'^2 \sim \sigma'^2 \chi^2_k$, and the objective of estimating θ_1 with the additional information that $\frac{\|\theta_1-\theta_2\|}{\sigma'} \leq m'$. To achieve this, one "rotates" X_1 and X_2 to the independent coordinates $X = (X_1 - X_2)/2$ and $W = (X_1 + X_2)/2$ and considers estimators of θ_1 of the form $\delta_\psi(X, W, S'^2) = W + \psi(X, S'^2)$ showing that δ_{ψ_1} dominates δ_{ψ_2} for estimating θ_1 under loss $\frac{\|\delta-\theta_1\|^2}{\sigma^2}$ with the additional information that $\frac{\|\theta_1-\theta_2\|}{\sigma} \leq m'$ **if and only if** $\psi_1(X, S^2)$ dominates $\psi_2(X, S^2)$ for estimating $\theta = E(X) = \frac{\theta_1-\theta_2}{2}$ under loss $\frac{\|\psi-\theta\|^2}{\sigma^2}$ and constraint $\frac{\|\theta\|}{\sigma} \leq m$, for model (1.1) with $X \sim N_p(\theta, \sigma^2 = \frac{\sigma'^2}{2})$, $S^2 = \frac{S'^2}{2}$ and $m = \frac{m'}{\sqrt{2}}$.

5. Appendix

5.1. Proof of Theorem 2.1

Here is a familiar identity (e.g., Watson, 1983), related to the normalization constant of a Langevin distribution and useful in the proof of Theorem 2.1.

Lemma 5.8. *For $y \in \mathbb{R}^p$, and U uniformly distributed on the sphere of radius r, we have*

$$E_r[e^{y'U}] = \Gamma(\frac{p}{2}) \, \frac{I_{p/2-1}\left(\|y\| \, r\right)}{(\|y\| \, r)^{(p/2-1)}} \, , \tag{5.21}$$

where $I_\nu(\cdot)$ is the modified Bessel function of order ν given by

$$I_\nu(z) = \sum_{k\geq 0} \frac{(\frac{z}{2})^{\nu+2k}}{k! \, \Gamma(\nu+k+1)} \, .$$

Proof of Theorem 2.1. It suffices to calculate $m^{BU}_{\sigma^2}(x) = m^{\pi}_{\sigma^2}(x)$ for a boundary uniform prior π as in (1.6, **i**). We have from (2.12) and (5.21), with U uniformly distributed on S_m,

$$\begin{aligned} m^{BU}_{\sigma^2}(x) &= E_m\left[e^{-\frac{1}{2}\|\frac{x}{\sigma^2}-U\|^2}\right] \\ &= e^{-\frac{m^2}{2}} e^{-\frac{\|x\|^2}{2\sigma^2}} E_m\left[e^{\frac{x'}{\sigma}U}\right] \\ &= e^{-\frac{m^2}{2}} e^{-\frac{\|x\|^2}{2\sigma^2}} \, \Gamma(\frac{p}{2}) \, (\frac{m\,\|x\|}{\sigma})^{(1-p/2)} I_{p/2-1}\left(m\,\|x\|/\,\sigma\right). \end{aligned} \tag{5.22}$$

Now, using the identity $\frac{d}{dt}\left(t^{1-\nu} I_\nu(t)\right) = t^{1-\nu} I_{\nu+1}(t)$, calculations lead to

$$\nabla_x m^{BU}_{\sigma^2}(x) = -\frac{x}{\sigma^2} m^{BU}_{\sigma^2}(x) + \Gamma(\frac{p}{2}) \, e^{-\frac{m^2}{2}} e^{-\frac{\|x\|^2}{2\sigma^2}} \, m^{2-p/2} \, \sigma^{p/2-2} \frac{x}{\|x\|^{p/2-2}} \, I_{p/2}\left(m\,\|x\|/\,\sigma\right). \tag{5.23}$$

Substituting this and (5.22) into (2.11) yields

$$\delta_{BU,l}\left(x, s^2\right) = m\frac{x}{\|x\|} \frac{\int_0^\infty \sigma^{l-k-p/2-4} \; e^{-\frac{\|x\|^2+s^2}{2\sigma^2}} I_{p/2}\left(m\,\|x\|/\,\sigma\right) d\sigma^2}{\int_0^\infty \sigma^{l-k-p/2-5} \; e^{-\frac{\|x\|^2+s^2}{2\sigma^2}} I_{p/2-1}\left(m\,\|x\|/\,\sigma\right) d\sigma^2} \, . \tag{5.24}$$

Finally, the result follows by substituting the series expression for the Bessel functions above, interchanging sums and integrals, integrating out with respect to σ^2, and some simplification. [4] $\square$

5.2. Proof of Lemma 2.3

(a) We have for $z \geq 0$

$$K(z) = (ca + (1-c)b)\, E_z \left[\frac{1}{(ca + (1-c)b) + I} \right],$$

where I is a discrete random variable with density proportional to $\frac{(a+1)_i}{(b)_i} \frac{z^i}{i!} \mathbb{I}_{\{0,1,\ldots\}}(i)$. Since these densities form a family with strictly increasing monotone likelihood ratio in I with parameter z, the result follows since $(ca + (1-c)b > 0$, and $\frac{1}{(ca+(1-c)b)+i}$ decreases in i, $i \in \{0, 1, \ldots\}$. The limiting value is established by exploiting the representation (e.g., Abramowitz and Stegun, 1964) $F(\alpha, \beta, z) = \frac{\Gamma(\beta)}{\Gamma(\alpha)} \exp(z) z^{\alpha-\beta} \left(1 + O\left(\frac{1}{z}\right)\right)$, where $O\left(\frac{1}{z}\right)$ is a bounded function by $\frac{1}{z}$ in a neighborhood of infinity.

(b) With the recurrence relation (e.g., Abramowitz and Stegun, 1964)

$$zF(\alpha+1, \beta+1, z) = \beta F(\alpha+1, \beta, z) - \beta F(\alpha, \beta, z),$$

we obtain

$$z \frac{F(a+1, b+1, z)}{F(a+1, b, z)} = \beta \left(1 - \frac{F(a, b, z)}{F(a+1, b, z)}\right),$$

so that the desired increasing property is seen as a direct consequence of part **(a)** by setting $c = 1$.

(c) Similarly as in **(a)**, write $\frac{1}{H(a)} = 1 + \frac{1}{b} E_a[J]$, where J is a discrete random variable with mass function

$$p_a(j) \propto \frac{(a)_j}{(b+1)_j} \frac{z^j}{j!} \mathbb{I}_{\{0,1,\ldots\}}(j),$$

and observe a strictly increasing monotone likelihood ratio property in J with parameter a. This implies the desired monotonicity of $H(\cdot)$. $\square$

5.3. Proof of Lemma 2.5

According to (1.2), the loglikelihood is given by

$$\ln L\left(\theta, \sigma^2\right) \;=\; \text{constant} - \frac{1}{2\sigma^2}\left(\|x-\theta\|^2 + s^2\right) - \frac{p+k}{2} \ln \sigma^2 .$$

[4] Equivalently, one can proceed with an intermediate identity of the form

$$\int_0^\infty A^{-\alpha} e^{-T/2A} I_\nu\left(\frac{\mu}{\sqrt{A}}\right) dA = \frac{\Gamma(\alpha+\nu/2-1)}{\Gamma(\nu+1)} \left(\frac{\mu^2}{2T}\right)^{\nu/2} \left(\frac{2}{T}\right)^{\alpha-1} F\left(\alpha+\nu/2-1, \nu+1, \frac{\mu^2}{2T}\right),$$

for positive μ, α, T.

For fixed σ^2, the likelihood with respect to θ is maximized for

$$\hat{\theta}_{\sigma^2} = \begin{cases} x & \text{if} \quad \|x\| \leq m\sigma \\ m\sigma \frac{x}{\|x\|} & \text{if} \quad \|x\| \geq m\sigma \,. \end{cases}$$

Therefore, we have $\sup_{(\theta,\sigma^2)\in\Theta(m)} L\left(\theta, \sigma^2\right) = \sup_{\sigma^2>0} L\left(\hat{\theta}_{\sigma^2}, \sigma^2\right)$. Next, we can see that

$$\begin{aligned} -\ln L\left(\hat{\theta}_{\sigma^2}, \sigma^2\right) &= \text{constant } + \frac{1}{2\sigma^2}\left(\left\|x - \hat{\theta}_{\sigma^2}\right\|^2 + s^2\right) + \frac{p+k}{2}\ln\sigma^2 \\ &= \begin{cases} \frac{s^2}{2\sigma^2} + \frac{p+k}{2}\ln\sigma^2 & \text{if} \quad \|x\| \leq m\sigma \\ \frac{s^2}{2\sigma^2} + \frac{p+k}{2}\ln\sigma^2 + \frac{1}{2}\left(m^2 + \frac{\|x\|^2}{\sigma^2} - 2m\frac{\|x\|}{\sigma}\right) & \text{if} \quad \|x\| \geq m\sigma \,. \end{cases} \end{aligned}$$

Now, the minimum of $-\ln L\left(\hat{\theta}_{\sigma^2}, \sigma^2\right)$ is attained on $\left(\frac{\|x\|}{m}, \infty\right)$ whenever $\frac{s^2}{p+k} \geq \frac{x'x}{m^2}$. Therefore, $\hat{\theta}_{ML}(x,s) = x$ whenever $t = \frac{x'x}{s^2} \leq \frac{m^2}{p+k}$. On the other hand, when $t > \frac{m^2}{p+k}$, the minimum of $-\ln L\left(\hat{\theta}_{\sigma^2}, \sigma^2\right)$ is attained at $\hat{\sigma_0}^2$ with

$$\hat{\sigma_0} = \frac{m\ \|x\|}{2(p+k)}\left(\sqrt{1 + 4\frac{(k+p)}{m^2}\left(\frac{s^2}{\|x\|^2} + 1\right)} - 1\right).$$

Finally, since $\hat{\theta}_{ML}(x,s^2) = m\frac{x}{\|x\|}\hat{\sigma_0}$ for such values of (x,s^2), the result follows. $\square$

5.4. Proof of (2.18)

With the decomposition

$$\begin{aligned} F\left(\frac{k+p}{2}+1, \frac{p}{2}, z\right) &= \sum_{j\geq 0} \frac{(\frac{k+p}{2}+1)_j}{(\frac{p}{2}+1)_j}\,\frac{(\frac{p}{2}+1)_j}{(\frac{p}{2})_j}\,\frac{z^j}{j!} = \sum_{j\geq 0} \frac{(\frac{k+p}{2}+1)_j}{(\frac{p}{2}+1)_j}\,(1+\frac{2j}{p})\,\frac{z^j}{j!} \\ &= F\left(\frac{k+p}{2}+1, \frac{p}{2}+1, z\right) + \frac{2z}{p}\sum_{j\geq 0}\frac{(\frac{k+p}{2}+1)_{j+1}}{(\frac{p}{2}+1)_{j+1}}\,\frac{z^j}{j!}\,, \end{aligned}$$

one obtains the representation

$$R(z) = 1 + \frac{2z}{p}\beta(z) = 1 + \frac{2z}{p}(1 + kE_z[\frac{1}{p+2J+2}]), \tag{5.25}$$

with E_z representing expected value with respect to a discrete random variable J with probability mass function proportional to $\frac{(\frac{k+p}{2}+1)_j}{(\frac{p}{2}+1)_j}\frac{z^j}{j!}\,\mathbb{I}_{\{0,1,\ldots\}}(j)$. With a similar expansion, we also have

$$z(\beta(z)-1) = kE_z[\frac{J}{k+p+2J}]\,. \tag{5.26}$$

Using (5.25) and (5.26), (2.18) becomes equivalent to

$$\begin{aligned} & 1 + \tfrac{2z}{p}\beta(z) > \tfrac{z}{p}\left(\sqrt{1+\tfrac{2(k+p)}{z}}+1\right) \\ \iff\quad & A(z) = \tfrac{1}{2k}\left\{4\beta(z)(\beta(z)-1)z^2 + 4p(\beta(z)-1)z + p^2\right\} > z & (5.27) \\ \iff\quad & \tfrac{p^2}{2kz} + T(z) > 1\,, & (5.28) \end{aligned}$$

where

$$T(z) = \frac{2}{k}\{\beta(z)(\beta(z)-1)z + p(\beta(z)-1)\}\,. \tag{5.29}$$

Now, by making use of (5.25) and (5.26), expand $T(\cdot)$ in terms of two independent copies J_1 and J_2 of $2J$ as

$$\begin{aligned}
T(z) &= \left(1 + kE_z(\frac{1}{p+2+J_1})\right)\left(E_z(\frac{J_2}{k+p+J_2})\right) + E_z(\frac{2p}{p+2+J_1}) \\
&= E_z\left[\frac{2p(k+p) + J_1J_2 + (p+2)J_2 + (k+2p)J_2}{(p+2)(k+p) + J_1J_2 + (p+2)J_2 + (k+p)J_1}\right].
\end{aligned}$$

In the above, we have expressed the product of expectations as the expectation of a product given that the J_i's are independent. Given (5.28), it will suffice to show that $T(z)$ is lower bounded by 1 for all $z > 0$. Since $p \geq 2$ by assumption, we obtain with an expansion where we denote p the joint probability function of (J_1, J_2) and $d(j_1, j_2) = (p+2)(k+p) + j_1j_2 + (p+2)j_2 + (k+p)j_1$,

$$\begin{aligned}
T(z) &\geq E_z\left[\frac{(p+2)(k+p) + J_1J_2 + (p+2)J_2 + (k+p)J_2}{(p+2)(k+p) + J_1J_2 + (p+2)J_2 + (k+p)J_1}\right] \\
&= 1 + (k+p)E_z\left[\frac{J_2 - J_1}{(p+2)(k+p) + J_1J_2 + (p+2)J_2 + (k+p)J_1}\right] \\
&= 1 + (k+p)(\sum_{j_2<j_1} + \sum_{j_2>j_1})\left(\frac{(j_2-j_1)\,p(j_1,j_2)}{d(j_1,j_2)}\right) \\
&= 1 + (k+p)(\sum_{j_2<j_1}(j_2-j_1)\left(\frac{p(j_1,j_2)}{d(j_1,j_2)} - \frac{p(j_2,j_1)}{d(j_2,j_1)}\right) \\
&= 1 + (k+p)(k-2)\sum_{j_2<j_1} p(j_1,j_2)\frac{(j_2-j_1)^2}{d(j_2,j_1)\,d(j_1,j_2)} \\
&\geq 1\,,
\end{aligned}$$

since J_1 and J_2 are independent and for $k \geq 2$. This establishes the result for all (m, p, k) such that $p \geq 2$ and $k \geq 2$. Finally, for $k = 1$ and $m \leq \sqrt{p}$, it is clear that (5.27) is verified for $z \leq m^2/2$. This completes the proof. □[5]

Acknowledgments

The authors are grateful to two reviewers, an associate editor, and the editor for useful comments and suggestions which led to a more accessible manuscript. The research work of Éric Marchand is partially supported by NSERC of Canada. During Othmane Kortbi's Ph.D. studies at the Université de Sherbrooke, he benefited from financial support from several sources but he wishes to thank especially the ISM (Institut de sciences mathématiques) and the CRM (Centre de recherches mathématiques). Finally, the authors are grateful to Bill Strawderman and Dominique Fourdrinier for useful discussions and encouraging us to pursue work on this problem.

[5]We point out that we have shown inequality (2.18) for all $p \geq 2, k \geq 2, z \geq 0$. We could not find a similar or as sharp as inequality in the literature. Such an inequality, and the technique used to establish it, may well be of independent interest as various bounds for ratios of special functions like the hypergeometric here have been previously the subject of study (e.g., Joshi and Bissu, 1996; Kokologiannaki, 2012.)

References

Abramowitz, M. and Stegun, I. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover, New York.

Eaton, M. L. (1989). *Group invariance applications in statistics.* Regional Conference Series in Probability and Statistics, vol. 1, Institute of Mathematical Statistics, Hayward, California.

Efromovich, S.Y. (1999). *Nonparametric Curve Estimation: Methods, Theory and Applications.* Springer, New York.

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.

Giri, N.C. (2004). *Multivariate Statistical Analysis.* Marcel Dekker, 2nd edition, revised and expanded.

Hartigan, J. (2004). Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, **67**, 285-288.

Kariya, T., Giri, N., and Perron, F. (1988). Invariant estimation of a mean vector μ of $N(\mu, \Sigma)$ with $\mu'\Sigma^{-1}\mu = 1$, $\Sigma^{-1/2}\mu = C$ or $\Sigma = \delta^2\mu'\mu I$. *Journal of Multivariate Analysis*, **27**, 270-283.

Kokologiannaki, C.G. (2012). Bounds for functions involving ratios of Bessel functions. *Journal of Mathematical Analysis and Applications*, **385**, 737-742.

Kortbi, O. and Marchand, É. (2012). Truncated linear estimation of a bounded multivariate normal mean, *Journal of Statistical Planning and Inference.* **142**, 2607-2618.

Kubokawa, T. (2005). Estimation of a mean of a normal distribution with a bounded coefficient of variation. *Sankhyā: The Indian Journal of Statistics*, **67**, 499-525.

Kubokawa, T. (1994). A unified approach to improving on equivariant estimators. *Annals of Statistics*, **22**, 290-299.

Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation.* Springer-Verlag, 2nd edition.

Johnstone, I. (2011). *Gaussian Estimation: Sequence and Wavelet Models*, Unpublished manuscript.

Joshi, C.M. and Bissu, S.K. (1996). Inequalities for some special functions. *Journal of Computational and Applied Mathematics*, **69**, 251-259.

Mandelkern, M. (2002). Setting Confidence Intervals for Bounded Parameters with discussion. *Statistical Science*, **17**, 149-172.

Marchand, É., Jafari Jozani, M. and Tripathi, Y. M. (2012). On the inadmissibility of various estimators of normal quantiles and on applications to two-sample problems with additional information. *Contemporary Developments in Bayesian analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman*, Institute of Mathematical Statistics Volume Series, **8**, 104-116.

Marchand, É. and Perron, F. (2001). Improving on the mle of a bounded normal mean. *Annals of Statistics*, **29**, 1078-1093.

Marchand, É., and Strawderman, W.E. (2012A). On Bayesian credible sets in restricted parameter space problems and lower bounds for frequentist coverage. (arXiv: 1208.002802)

Marchand, É., and Strawderman, W.E. (2012B). A unified minimax result for restricted parameter spaces. *Bernoulli*, **18**, 635-643.

Marchand, É., Strawderman, W. E., Bosa, K., and Lmoudden, A. (2008). On the frequentist coverage of Bayesian credible intervals for lower bounded means. *Electronic Journal of Statistics*, **2**, 1028-1042.

Marchand, É. and Strawderman, W. E. (2006). On the behaviour of Bayesian credible intervals for some restricted parameter space problems. *Recent Developments in Nonparametric Inference and Probability : A Festschrift for Micheal Woodroofe*, IMS Lecture Notes-Monograph Series, **50**, pp. 112-126.

Marchand, É., and Strawderman, W.E. (2005). On improving on the minimum risk equivariant estimator of a location parameter which is constrained to an interval or a half-interval. *Annals of the Institute of Statistical Mathematics*, **57**, 129-143.

Marchand, É. and Strawderman, W. E. (2004). Estimation in restricted parameter spaces: A review. *Festschrift for Herman Rubin*, IMS Lecture Notes-Monograph Series, **45**, pp. 21-44.

Moors, J.J.A. (1985). Estimation in truncated parameter spaces. Ph.D. thesis, Tilburg University.

Perron, F. and Giri, N.C. (1990). On the best equivariant estimator of mean of a multivariate normal population. *Journal of Multivariate Analysis*, **32**, 1-16.

Roe, B. and Woodroofe, M. (2000). Setting confidence belts. *Physical Review D*, **63**, 013009/01-09.

Watson, G.S. (1983). *Statistics on spheres.* John Wiley, New York.

Zhang, T. and Woodroofe, M. (2003). Credible and confidence sets for restricted parameter spaces. *Journal of Statistical Planning and Inference*, **115**, 479-490.