On Bayes estimators with uniform priors on spheres and their comparative performance with maximum likelihood estimators for estimating bounded multivariate normal means

Dominique Fourdrinier[a], Éric Marchand[b,1] ,

a Université de Rouen, LITIS EA 4108, BP 12 76801, Saint Étienne du Rouvray, France

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

Summary

For independently distributed observables: $X_i \sim N(\theta_i, \sigma^2), i = 1, \ldots, p$, we consider estimating the vector $\theta = (\theta_1, \ldots, \theta_p)'$ with loss $\|d - \theta\|^2$ under the constraint $\sum_{i=1}^{p} \frac{(\theta_i - \tau_i)^2}{\sigma^2} \leq m^2$, with known $\tau_1, \ldots, \tau_p, \sigma^2, m$. In comparing the risk performance of Bayesian estimators δ_α associated with uniform priors on spheres of radius α centered at $(\tau_1, \ldots, \tau_p)$ with that of the maximum likelihood estimator δ_{mle}, we make use of Stein's unbiased estimate of risk technique, Karlin's sign change arguments, and a conditional risk analysis to obtain for a fixed (m, p) necessary and sufficient conditions on α for δ_α to dominate δ_{mle}. Large sample determinations of these conditions are provided. Both cases where all such δ_α's, or no such δ_α's dominate δ_{mle} are elicited. As a particular case, we establish that the boundary uniform Bayes estimator δ_m dominates δ_{mle} if and only if $m \leq k(p)$ with $\lim_{p\to\infty} \frac{k(p)}{\sqrt{p}} = \sqrt{2}$, improving on the previously known sufficient condition of Marchand and Perron (2001) for which $k(p) \geq \sqrt{p}$. Finally, we improve upon a universal dominance condition due to Marchand and Perron, by establishing that **all** Bayesian estimators δ_π with π spherically symmetric and supported on parameter space dominate δ_{mle} whenever $m \leq c_1(p)$ with $\lim_{p\to\infty} \frac{c_1(p)}{\sqrt{p}} = \sqrt{\frac{1}{3}}$. *AMS 2000 subject classifications.* 62F10, 62F15, 62F30, 62C10, 62H12. *Keywords and phrases*: Restricted parameters, point estimation, squared error loss, dominance, maximum likelihood, Bayes estimators, multivariate normal, unbiased estimate of risk, sign changes, modified Bessel functions.

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

1. Introduction

Consider independently and normally distributed observables: $X_i \sim N(\theta_i, \sigma^2), i = 1, \ldots, p$, with the additional information that $\sum_{i=1}^{p} \frac{(\theta_i - \tau_i)^2}{\sigma^2} \leq m^2$, with known $\tau_1, \ldots, \tau_p, \sigma^2, m$. From a practical point of view, a constraint as the one above signifies that the squared standardized deviations $(|\frac{\theta_i - \tau_i}{\sigma}|)$ are on average bounded by $\frac{m^2}{p}$. Inferential problems with information or constraints of the above type certainly arise in practical settings. It is thus relevant to analyze the comparative performance of various estimators that may well vary in how efficiently they capitalize on the parametric information. For recent decision-theoretic reviews of such restricted parameter space problems, we refer to the paper of Marchand and Strawderman (2004), as well as the monograph of van Eeden (2006).

We proceed by setting $\sigma^2 = 1$ and $\tau_i = 0, i = 1, \ldots, p$, without loss of generality, and writing our model as $X \sim N_p(\theta, I_p)$ with $X = (X_1, \ldots, X_p)'$ and $\theta = (\theta_1, \ldots, \theta_p)' \in \Theta(m) = \{\theta \in \Re^p : \|\theta\| \leq m\}$. We are concerned here with estimating θ under quadratic loss $L(\theta, \delta) = \|\delta - \theta\|^2$. As a followup to the work of Marchand and Perron (2001), we focus here on the determination of Bayesian estimators that improve upon the benchmark, but inadmissible, maximum likelihood estimator given by $\delta_{\text{mle}}(x) = (m \wedge \|x\|)\frac{x}{\|x\|}$. Although dominating estimators can be provided for any pair (m, p), the specification of priors π that lead to dominating Bayesian estimators δ_π is both of interest, and much more difficult. For the purposes of introducing the findings in this paper, here are a couple of key results from Marchand and Perron (2001).

(i) For sufficiently small m, say $m \leq c_1(p)$, all Bayes estimators δ_π with respect to an orthogonally invariant prior π (supported on $\Theta(m)$) dominate δ_{mle};

(ii) The Bayes estimator δ_{BU} (or δ_m as referred to in this paper) with respect to a uniform prior on the boundary of $\Theta(m)$ (or the sphere $S_m = \{\theta : \|\theta\| = m\}$) dominates δ_{mle} whenever $m \leq \sqrt{p}$.[2]

Various other dominance results, such as those pertaining to a fully uniform prior on $\Theta(m)$ and other absolutely continuous priors are also available from Marchand and Perron (2001), but we will focus here on results **(i)** and **(ii)** above.

With respect to important properties of δ_{BU}, we point out that it is the optimal equivariant estimator for $\theta \in S_m$, and thus necessarily improves upon δ_{mle} on S_m. Furthermore, δ_{BU} also represents the Bayes estimator which expands the greatest, or shrinks the least towards the origin (i.e., $\|\delta_\pi\| \leq \|\delta_{BU}\|$ for all π supported on $\Theta(m)$; Marchand and Perron, 2001). Despite this, as expanded upon in Section 2.3, δ_{BU} still shrinks more than δ_{mle} whenever $m \leq \sqrt{p}$, but not otherwise with the consequence of increased risk at $\theta = 0$ and failure to dominate δ_{mle} for large m. With the view of seeking dominance for a wider range of values of m, for potentially modulating these above effects by introducing more (but not too much) shrinkage, we consider the class of uniform priors supported on spheres S_α of radius α; $0 \leq \alpha \leq m$; about the origin, and their corresponding Bayes estimators δ_α. The choice is particularly interesting since the amount of shrinkage is calibrated by the choice of α (as formalized in Section 2.3), with the two extremes $\delta_m \equiv \delta_{BU}$, and $\delta_0 \equiv 0$ (e.g., prior degenerate at 0). Moreover, knowledge about dominance conditions for the estimators δ_α may well lead, through further analytical risk and unbiased estimates of risk comparisons (e.g., Marchand and Perron, 2001, Lemma 5 and the Remarks that follow), to implications relative to other Bayesian estimators such as the fully uniform on $\Theta(m)$ prior Bayes estimator.

[2] With the cutoff point proportional to $\sqrt{p}$, the condition translates to the average squared deviation being less than or equal to 1. Analogously, the dominance conditions of this paper involve a growth condition with the radius m proportional to $\sqrt{p}$ and hence become amenable to similar interpretations.

We thus propose, and arrive at extensions of Marchand and Perron's dominance condition **(ii)** to $\alpha < m$. This is achieved by:

(A) making use of Stein's (1981) technique for estimating unbiasedly the difference in risks between δ_α and δ_{mle} (Section 2.2), as well as Karlin's (1957) sign change arguments, to obtain **necessary and sufficient** conditions (for $p \geq 3$) for estimators δ_α to dominate δ_{mle} (Section 2.3);

(B) making use of various analytical and large sample properties (Sections 2.2 and 2.3) to obtain more explicit and asymptotically precise (in p) conditions on α for δ_α to dominate δ_{mle} (Section 2.3).

In comparison to **(ii)**, we also derive a more incisive analysis (Example 1) for the boundary uniform estimator δ_m, obtaining a necessary and sufficient dominance condition $m \leq k_2(p, m)$, and with the large sample p approximation $k_2(p, m) \approx \sqrt{2p}$. Various other examples and implications are presented as well in Section 2.4. It

Finally in Section 3, with respect to the universal dominance condition in **(i)**, we derive the large sample behaviour of $c_1(p)$ obtaining the simple approximation $c_1(p) \approx \sqrt{\frac{p}{3}}$. We make use of Marchand and Perron's framework, as well as a key asymptotic result (Lemma 6) previously introduced by Marchand and Perron (2002).

2. Main results

2.1. Preliminary results

We begin with analytical properties relative to the Bayes estimator δ_α associated with the uniform prior on the sphere $S_\alpha = \{\, \theta \in \Re^p : \|\theta\| = \alpha \,\}$, $0 \le \alpha \le m$. We denote $\|X\|$, $\|x\|$, and $\|\theta\|$ by R, r, and λ respectively. We define $\rho_\nu(t) = I_{\nu+1}/I_\nu(t)$; $t > 0, \nu > -1/2$; where I_ν represents the modified Bessel function of order ν.

Lemma 1. *(Robert, 1990, Berry 1990) We have for any θ element of S_α,*

$$\delta_\alpha(x) = \frac{1}{r} E_\theta[\frac{\theta' X}{\|X\|} | \|X\| = r]\; x \;=\; \frac{\alpha}{r}\; \rho_{p/2-1}(\alpha r)\; x\;.$$

In Lemma 2, we pursue with known properties of ρ, while a key property given in part (a) of Lemma 3; previously unknown to the best of our knowledge; actually relates to a more general increasing monotone likelihood ratio property of the family of conditional distributions $\{W|R = r : r > 0\}$, with parameter r, $W = \|X\| - \frac{\theta' X}{\|\theta\|}$, and $\theta \in S_\alpha$ (fixed).

Lemma 2. *(a) (Watson, 1983) The function $\rho_\nu(\cdot)$ is increasing and concave on $(0, \infty)$, with $\lim_{t\to 0^+} \rho_\nu(t) = 0$, $\lim_{t\to\infty} \rho_\nu(t) = 1$; and $\frac{\rho_\nu(t)}{t}$ decreasing in t with $\lim_{t\to 0^+} \frac{\rho_\nu(t)}{t} = \frac{1}{2(\nu+1)}$. Also, we have the identity $\frac{d}{dt}\rho_\nu(t) = 1 - (1+2\nu)\frac{\rho_\nu(t)}{t} - \rho_\nu^2(t)$, and the inequality $\frac{d}{dt}\rho_\nu(t) \le \frac{\rho_\nu(t)}{t}$.*

(b) (Amos, 1974) For all $\nu \ge 0$ and $t > 0$, we have

$$L(\frac{2(\nu+1)}{t}, \frac{2(\nu+1)}{t}) \le \rho_\nu^2(t) \le L(\frac{2\nu}{t}, \frac{2(\nu+2)}{t})\,, \tag{1}$$

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

Lemma 3. *(a) For all $p \in \{3, 4, \ldots\}$ and $\alpha \ge 0$, the function given by $r\ \{1 - \rho_{p/2-1}(\alpha r)\}$ is increasing in r; $r \ge 0$;*

(b) For all $p \in \{3, 4, \ldots\}$, we have the inequality $\rho_{p/2-1}(t) + t\,\rho'_{p/2-1}(t) \leq 1$, for all $t > 0$;

(c) For all $p \in \{3, 4, \ldots\}$, we have $\lim_{t\to\infty} t\,\{1 - \rho_{p/2-1}(t)\} = \frac{p-1}{2}$. [3]

Proof. **(c)** The result follows from the fact that $t\rho'_\nu(t) \to 0$ as $t \to \infty$, which must the case for part (b) to hold since $\rho_\nu(t) \to 1$ as $t \to \infty$, as well as the given expression for ρ'_ν given in Lemma 2.

(b) Part (a) tells us (take $\alpha = 1$) that $t\,(1 - \rho_{p/2-1}(t))$ increases in t, in other terms : $\frac{\partial}{\partial t}\,\{t\,(1 - \rho_{p/2-1}(t))\} \geq 0$ which is equivalent to, and establishes, part (b).

(a) Begin with Lemma 1 which implies that $r\,\{\,1 - \rho_{p/2-1}(\alpha r)\,\} = E_\theta[W\,|\,R = r]$, with $W = \|X\| - \frac{\theta' X}{\|\theta\|}$, and $\theta \in S_\alpha$. It will hence suffice to show that a family of conditional distributions $\{W\,|\,R = r : r > 0\}$ satisfies (for $p \geq 3$) an increasing in W monotone likelihood ratio property, with parameter r. Observe also that the probability distribution of W remains unchanged with orthogonal transformations $X \to \Gamma X$ (and $\theta \to \Gamma\theta$), which permits us, since the actions are transitive on S_α, to set without loss of generality $\theta = \theta_0 = (\alpha, 0, \ldots, 0)'$. Pursue next with the joint density (for $\theta = \theta_0$ and $p > 1$) of $(Y_1 = X_1, Y_2 = X'X - {X_1}^2)$, given by:

$$f_{Y_1,Y_2}(y_1, y_2) \propto e^{-\frac{1}{2}[(y_1-\alpha)^2+y_2]}\; y_2^{\frac{p-1}{2}-1}\; 1_{(0,\infty)}(y_2)\,,$$

to derive the joint density of $(W = \sqrt{Y_1^2 + Y_2} - Y_1, R = \sqrt{Y_1^2 + Y_2})$,

$$f_{W,R}(w, r) \propto r\;\exp\{\,-\frac{r^2}{2} + \alpha(r - w)\,\}\;[w(2r - w)]^{\frac{p-3}{2}}\;1_{(\,0,\,2r\,)}(w)\;1_{(\,0,\infty\,)}(r)\,,$$

[3] Alternatively, the more general result $\lim_{t\to\infty} t(1-\rho_\nu(t)) = \nu + 1/2$ holds for all $\nu > 0$ by bounds given by Amos (1974) for $\rho_\nu(t)$, $t > 0$, which are

$$L(\frac{2(\nu+1/2)}{t}, \frac{2(\nu+3/2)}{t}) \leq \rho_\nu^2(t) \leq L(\frac{2(\nu+1/2)}{t}, \frac{2(\nu+1/2)}{t})\,.$$

This is verified with the evaluation $\lim_{z\to\infty} z(1 - \frac{z}{(\nu+1/2)+\sqrt{z^2+b}}) = \nu + 1/2$.

and the conditional density:[4]

$$f_{W|R=r}(w) \propto \exp\{-\alpha w\} \, [w(2r-w)]^{\frac{p-3}{2}} \, 1_{(0,2r)}(w) \, ; r > 0. \tag{2}$$

To conclude, the result follows by checking that the ratio $\frac{f_{W|R=r_1}(w)}{f_{W|R=r_0}(w)}$ is nondecreasing in w for all $r_1 > r_0 > 0$. □

Finally, we will make use in the next section of Stein's unbiased estimate of risk, which permits us to express the risk of an estimator $\delta_g(X) = X + g(X)$, with $X \sim N_p(\theta, I_p)$, g weakly differentiable, and $E_\theta[g(\|X\|^2)] < \infty$, as:

$$R(\theta, \delta_g) = E_\theta[\|\delta_g(X) - \theta\|^2] = E_\theta[p + 2div\, g(X) + \|g(X)\|^2], \tag{3}$$

with $div\, g$ being the divergence operator.

2.2. An unbiased estimator of the difference of risks

The next result provides an unbiased estimate $D_\alpha(\|X\|)$ of the risk difference $R(\theta, \delta_\alpha) - R(\theta, \delta_{\text{mle}})$, as well as a sign change analysis of $D_\alpha(\cdot)$ for all triplets (p, α, m) for which $p \geq 3$ and $0 \leq \alpha \leq m$.

Lemma 4. (**a**) *An unbiased estimator of the difference in risks $R(\theta, \delta_\alpha) - R(\theta, \delta_{mle})$ is given by $D_\alpha(\|X\|) = D_{\alpha,1}(\|X\|)\, [0 \leq \|X\| \leq m] + D_{\alpha,2}(\|X\|)\, [\|X\| > m]$, with*

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

and

$$D_{\alpha,2}(r) = 2\alpha^2 - m^2 - \alpha^2 \rho^2_{p/2-1}(\alpha r) + 2mr\{1 - \frac{\alpha}{m}\rho_{p/2-1}(\alpha r)\} - 2(p-1)\frac{m}{r}. \tag{4}$$

(**b**) *For $p \geq 3$, and $0 \leq \alpha \leq m$, $D_\alpha(r)$ changes signs as a function of r according to the order:* **(i)** $(-,+)$ *whenever $\alpha \leq \sqrt{p}$, and* **(ii)** $(+,-,+)$ *whenever $\alpha > \sqrt{p}$.*

[4]Interestingly, for $p = 3$, the distribution $W|R = r$ is truncated exponential.

Proof. (a) Since $\delta_{\text{mle}}(x) = x + g_{\text{mle}}(x)$ with $g_{\text{mle}}(x) = (\frac{m}{r} - 1)x\,[r > m]$, we obtain[5]:

$$2div g_{\text{mle}}(x) + \|g_{\text{mle}}(x)\|^2 = \{2(p-1)\frac{m}{r} - 2p + (m-r)^2\}\; 1_{(m\,,\infty)}(r)\,;$$

and, by virtue of Stein's identity in (3), $R(\theta, \delta_{\text{mle}}) = E_\theta[\eta_{\text{mle}}(X)]$ with

$$\eta_{\text{mle}}(x) = p\,1_{[0,\,m]}(r) \;+\; \{2(p-1)\frac{m}{r} - p + (m-r)^2\}\; 1_{(m,\,\infty)}(r). \tag{5}$$

Analogously, as derived by Berry (1990), the representations of δ_α and $\frac{d}{dt}\rho_\nu(t)$ given in Lemmas 1 and 2, along with (3), permit us to write $R(\theta, \delta_\alpha) = E_\theta[\lambda_\alpha(X)]$ with

$$\eta_\alpha(x) = 2\alpha^2 + r^2 - p - 2\alpha\, r\, \rho_{p/2-1}(\alpha\, r) - \alpha^2 \rho^2_{p/2-1}(\alpha\, r). \tag{6}$$

Finally, the given expression for the unbiased estimator $D_\alpha(\|X\|)$ follows directly from (5) and (6).

(b) We begin with three intermediate observations which are proven below.

(I) The sign changes of $D_{\alpha,1}(r)$; $r \in [0, m]$; are ordered according to one of the five following combinations: $(+), (-), (-,+), (+,-), (+,-,+)$;

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

(III) the function $D_{\alpha,2}(r)$; $r \in (m, \infty)$ is either positive, or changes signs once from $-$ to $+$.

From properties (I), (II), and (III), we deduce that the sign changes of $D_\alpha(r)$ $r \in (0,\infty)$; an everywhere continuous function except for the jump discontinuity at m; are ordered according to one of the three following combinations: $(+), (-,+), (+,-,+)$. Now, recall that δ_α is a Bayes and admissible estimator of θ under squared error loss. Therefore, among the combinations above, $(+)$ is not possible since this would imply that δ_α is dominated by δ_{mle} in contradiction with

[5]Notice that g_{mle} is weakly differentiable.

its admissibility. Finally, the two remaining cases are distinguished by observing that, $D_\alpha(0) = 2\alpha^2 - 2p \leq 0$ if and only if $\alpha \leq \sqrt{p}$.

Proof of (I) Begin by making use of Lemma 2 to differentiate $D_{\alpha,1}$ and obtain:

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

Since, the quantities $r^{-1}\rho_{p/2-1}(\alpha r)$ and $\rho'_{p/2-1}(\alpha r)$ are positive and decreasing in r by virtue of Lemma 2, $r^{-1}\, D'_{\alpha,1}(r)$ is necessarily increasing in r, $r \in [0,m]$. Hence, $D'_{\alpha,1}(\cdot)$ has, on $[0,m]$, sign changes ordered as either: $(+), (-)$, or $(-,+)$. Finally, observe as a consequence that $D_{\alpha,1}(\cdot)$ has at most two sign changes on $[0,m]$, and furthermore that, among the six possible combinations, the combination $(-,+,-)$ is not consistent with the sign changes of $D'_{\alpha,1}$.

Proof of (II) Follows by a direct evaluation of $D_{\alpha,1}(m)$ and $D_{\alpha,2}(m)$ which are given in part (a) of this lemma.

Proof of (III) First, one verifies from (4), part (a) of Lemma 2, and part (c) of Lemma 3 that $\lim_{r\to\infty} D_{\alpha,2}(r)$ is $+\infty$, for $\alpha < m$; and equal to $p-1$ if $\alpha = m$. Moreover, part (a) also permits us to express $D_{\alpha,2}(r)$; $r > m$; as $(1 - \frac{\alpha}{m}\,\rho_{p/2-1}(\alpha r))\,\sum_{i=1}^{3} H_i(\alpha, r)$ with

$$\begin{aligned}
H_1(\alpha,r) &= 2rm\{1 - \frac{(1-\frac{\alpha^2}{m^2})m}{r\{\,1-\frac{\alpha}{m}\rho_{p/2-1}(\alpha r)\,\}}\}\,,\\
H_2(\alpha,r) &= \frac{-2(p-1)\,m}{r\,\{\,1-\frac{\alpha}{m}\rho_{p/2-1}(\alpha r)\,\}}\,,\\
\text{and }\ H_3(\alpha,r) &= m^2 + \alpha\, m\, \rho_{p/2-1}(\alpha r).
\end{aligned}$$

Hence, to establish property (III), it will suffice to show that each one of the functions $H_i(\alpha,\cdot)$, $i = 1,2,3$, is increasing on $(m\,,\infty)$ under the given conditions on (p,α,m). The properties of Lemma 2 clearly demonstrate that $H_3(\alpha,\cdot)$ is increasing, and it is the same for $H_2(\alpha,\cdot)$ given also Lemmas 2 and 3 since

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

Finally, for the analysis of $H_1(\alpha, r), r > m$, begin by differentiation and a rearrangement of terms to obtain

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

where, for $r > m \geq \alpha$,

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

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

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

by Lemma 3, part (b), since $r \geq m \geq \alpha$. The above establishes that $T(m) \geq T(\alpha) \geq 0$ for all $m \geq \alpha$, that $H_1(\alpha, r)$ increases in r, and completes the proof of the Theorem.

2.3. Risk comparisons and dominance results

With the estimators δ_α and δ_{mle} being equivariant (with respect to orthogonal transformations), and with their risks depending on θ only through $\|\theta\|$, we denote their difference in risks as

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

Given that $D_\alpha(\|X\|)$ is an unbiased estimator of $\Delta_\alpha(\lambda)$, and in view of the sign change behaviour of $D_\alpha(\cdot)$, we may now deduce the possible sign changes of $\Delta_\alpha(\lambda)$, as a function of $\lambda \in [0, m]$ (or $\lambda \in [0, \infty)$), as well as necessary and sufficient conditions for δ_α to dominate δ_{mle} on $\Theta(m)$.

Corollary 1. *For $p \geq 3$ and $0 \leq \alpha \leq m$, the estimator δ_α dominates δ_{mle} if and only if:*

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

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

Proof. The probability distribution of $\|X\|^2$ is $\chi^2_p(\lambda^2)$, so that the potential sign changes of $\Delta_\alpha(\lambda) = E_\lambda[D_\alpha(\|X\|)]$ are controlled by the variational properties of $D_\alpha(\cdot)$ in terms of sign changes (e.g., Brown, Johnstone et MacGibbon, 1981). Therefore, in situation (i) with $\alpha \leq \sqrt{p}$, it follows from part (b) of Lemma 4 that, as $\Delta_\alpha(\cdot)$ varies on $[0,\infty]$ (or $[0,m]$), the number of sign changes is at most one, and that such a change must be from $-$ to $+$. Therefore, since δ_α is admissible; and that the case $\Delta_\alpha(\lambda) \geq 0$ for all $\lambda \in [0,m]$ is not possible[6]; we must have indeed that $\Delta_\alpha(\cdot) \leq 0$ on $[0,m]$ if and only if $\Delta_\alpha(m) \leq 0$ establishing (i). A similar line of reasoning implies the result in (ii) as well. □

Observe how the necessary and sufficient conditions for δ_α to dominate δ_{mle} have simplified, expressed only in terms of the risks at $\theta = 0$, i.e., the centre of the parameter space $\Theta(m)$, and for $\theta \in S_m$, the boundary of $\Theta(m)$. Our task now consists in translating these necessary and sufficient conditions in terms of (α, m, p). We pursue with this, as well as asymptotics for large p. We will require first the following risk function representations for $\theta = 0$ and $\theta \in S_m$.

Lemma 5. *For an equivariant estimator δ of the form $\delta(X) = g(\|X\|)X$ for some nonnegative function g, we have $R(0,\delta) = E_0[\,\|\delta(X)\|^2\,]$, and, for $\theta \in S_m$,*

$$R(\theta,\delta) = R(\theta,\delta_m) + E_\theta[\{\|\,\delta(X)\| \,-\, \|\,\delta_m(X)\|)\}^2]\ . \tag{7}$$

[6]The risks of δ_α and δ_{mle} cannot match either, since a linear combination of these two distinct estimators would improve on δ_α.

Proof. The risk at $\theta = 0$ is obtained directly, while the risk for $\theta \in S_m$ follows by conditioning on $R = \|X\|$, as was presented for instance by Marchand and Perron (2005, Theorem 1). $\square$

Witness the role of $\|\delta(X)\|$, i.e., the amount of shrinkage associated with estimator $\delta(X)$ in these above representations. Indeed, the more $\delta(X)$ is shrunk towards 0, the smaller the risk at 0, with minimal risk when the shrinkage is most extreme, that is $\delta \equiv 0$. On the other hand, because any proper Bayes estimator δ_π with a prior supported on $\Theta(m)$ shrinks with respect to δ_m (Marchand and Perron, 2001, Theorem 4), in other words $\|\delta_\pi\| \leq \|\delta_m\|$, the less δ_π is shrunk, the smaller the risk on the boundary S_m, with minimal risk for the least shrinkage when $\delta_\pi \equiv \delta_m$.

In terms of our estimators δ_α studied here, the above extremes correspond simply to $\alpha = 0$ and $\alpha = m$. Moreover, since $\|\delta_\alpha(x)\| = \alpha\rho_{p/2-1}(\alpha\|x\|)$ (Lemma 1), we see that the amount of shrinkage is controlled directly by α, and is strictly decreasing as α increases (Lemma 2). We obtain thus the following result immediately from the analysis above as well as Corollary 1.

Corollary 2. *For $p \geq 3$ and $0 \leq \alpha \leq m$, the estimator δ_α dominates δ_{mle} if and only if*[7]*:*

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

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

where $k_1(p,m)$ and $k_2(p,m)$ are respectively the (unique) solutions in α of the equations $\Delta_\alpha(m) = 0$ and $\Delta_\alpha(0) = 0$ respectively.

Example 1. *(The boundary uniform Bayes estimator) Observe how Corollary 1 (or Corollary 2) applies to the boundary uniform Bayes estimator δ_m. Since $\Delta_m(m) \leq 0$ (or $k_1(p,m) \leq m$), as*

[7]Stating these conditions for dominance does not imply of course that they can be fulfilled in α for any pair (m,p); as we will see below with m large relative to $\sqrt{p}$. The conditions are still useful nevertheless representing conditions for non-dominance given their necessary and sufficient nature.

pointed out in the argumentation above, it follows immediately that δ_m dominates δ_{mle} whenever $m \leq \sqrt{p}$ (and $p \geq 3$), as previously established by Marchand and Perron (2001) (for all $p \geq 1$ though).

However, our above conditions for dominance will permit us to obtain a more incisive analysis for δ_m, as well as for δ_α, $\alpha < m$. For analyzing the large sample (large p) behaviours of $k_i(p,m); i = 1,2$, and for producing sharp approximations, we will require the following lemma, available from Marchand and Perron (2002), but with a short proof reproduced here nevertheless for sake of completeness.

Lemma 6. *Let $R_p^2 \sim \chi_p^2(d^2 p)$ for $d \geq 0$ and $p = 1, 2, \ldots$. Then, for any $c > 0$, the sequence of random variables $\rho^2_{p/2-1}(c\sqrt{p}\, R_p)$ converges, as $p \to \infty$, in probability to $L(\frac{1}{c\sqrt{1+d^2}}, \frac{1}{c\sqrt{1+d^2}})$. For the particular cases where $d = 0$ and $d = c$, the convergence is to $(\frac{2c}{1+\sqrt{4c^2+1}})^2$ and $\frac{d^2}{1+d^2}$ respectively.*

Proof. It follows from (1) that, for $p \geq 3$,

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

with probability one. The result follows by the convergence in probability, as $p \to \infty$, of $\frac{R_p^2}{p}$ to $1 + d^2$, and the resulting convergence in probability of both the lower and upper bounds to $L(\frac{1}{c\sqrt{1+d^2}}, \frac{1}{c\sqrt{1+d^2}})$. Finally, the given results for the cases $d = 0$ and where $d = c$ are obtained by evaluations of $L(\frac{1}{c}, \frac{1}{c})$ and $L(\frac{1}{d\sqrt{1+d^2}}, \frac{1}{d\sqrt{1+d^2}})$. $\square$

Theorem 1. *For the cutoff point $k_2(p,m)$ given in Corollary 2 above, we have for* **all** *$d \geq 1$*

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

Proof. It suffices to show that

$$\lim_{p\to\infty} \frac{\Delta_{c\sqrt{p}}(0)}{p} = 0 \text{ if and only if } c = \sqrt{2}.$$

First, it follows from Lemma 5 that

$$\frac{\Delta_{c\sqrt{p}}(0)}{d^2 p} = E_0[\, \frac{c^2}{d^2} \rho^2_{p/2-1}(c\sqrt{p}\|X\|) - (1 \wedge \frac{\|X\|^2}{d^2\, p})\,] \,. \tag{8}$$

From Lemma 6, we have

$$\lim_{p\to\infty} E_0[\, \rho^2_{p/2-1}(c\sqrt{p}\|X\|) \;=\; L(\frac{1}{c}, \frac{1}{c}).$$

Moreover, for $\theta = 0$, the sequence of random variables $\frac{\|X\|^2}{p}$ converges in probability to 1 by the weak law of large numbers, and hence $(1 \wedge \frac{\|X\|^2}{d^2\, p})$ converges in probability to $\frac{1}{d^2}$ for $d \geq 1$. Therefore, for $d \geq 1$, we have $\lim_{p\to\infty} E_0[\,(1 \wedge \frac{\|X\|^2}{d^2\, p})\,] = \frac{1}{d^2}$, and from (8):

$$\lim_{p\to\infty} \frac{\Delta_{c\sqrt{p}}(0)}{d^2 p} = \frac{1}{d^2}\{\, c^2\, L(\frac{1}{c}, \frac{1}{c}) - 1\,\}\,.$$

Consequently $\lim_{p\to\infty} \frac{\Delta_{c\sqrt{p}}(0)}{p} = 0$ if and only if

$$c^2\, L(\frac{1}{c}, \frac{1}{c}) = 1 \;\Leftrightarrow\; \frac{2c^2}{1+\sqrt{1+4c^2}} = 1 \Leftrightarrow c = \sqrt{2}\,;$$

using Lemma 6. □

Before moving ahead with a similar analysis for the cutoff point $k_1(p, m)$, we obtain an asymptotic expression for the difference in risks $\Delta_\alpha(m)$.

Lemma 7. *For $\theta \in S_m$, $m = d\sqrt{p}$, $\alpha = c\sqrt{p}$ with $c \in [0, d]$, and the difference in risks $\Delta_\alpha(m) = R(\theta, \delta_\alpha) - R(\theta, \delta_{mle})$, we have*

$$\lim_{p\to\infty} \frac{\Delta_\alpha(m)}{p} = c^2 L(\frac{1}{c\sqrt{1+d^2}}, \frac{1}{c\sqrt{1+d^2}}) - \frac{2c\, d^2}{\sqrt{1+d^2}} \sqrt{L(\frac{1}{c\sqrt{1+d^2}}, \frac{1}{c\sqrt{1+d^2}})} - d^2 + \frac{2\, d^3}{\sqrt{1+d^2}}\,.$$

Proof. With Lemma 1's representation of δ_α and with $\delta_{\text{mle}}(X) = (m \wedge R)\frac{X}{R}$, we have from (7) and for $\theta \in S_m$,

$$\frac{\Delta_\alpha(m)}{p} = E_\theta[\frac{\alpha^2\, \rho^2_{\frac{p}{2}-1}(\alpha\, R)}{p}] - 2E_\theta[\frac{\alpha m}{p}\, \rho_{\frac{p}{2}-1}(\alpha\, R)\, \rho_{\frac{p}{2}-1}(m\, R)] - E_\theta[\frac{(m \wedge R)^2}{p}] + 2\, E_\theta[(\frac{m \wedge R}{\sqrt{p}})\frac{m}{\sqrt{p}}\, \rho_{\frac{p}{2}-1}(m\, R)]\,.$$

With the given assumptions, it follows that $\sqrt{\frac{R^2}{p}}$ converges in probability, as $p \to \infty$, to $\sqrt{1+d^2}$, and hence $(\frac{m \wedge R}{\sqrt{p}})$ converges in probability to d. Now make use as well of Lemma 6 to infer that:

$$\begin{aligned}
\lim_{p\to\infty} E_\theta[\frac{\alpha^2\,\rho^2_{\frac{p}{2}-1}(\alpha\,R)}{p}] &= c^2\,L(\frac{1}{c\sqrt{1+d^2}},\frac{1}{c\sqrt{1+d^2}})\,;\\
\lim_{p\to\infty} E_\theta\{\frac{\alpha\,m}{p}\,\rho_{\frac{p}{2}-1}(\alpha\,R)\,\rho_{\frac{p}{2}-1}(m\,R)]\} &= cd\,\sqrt{L(\frac{1}{c\sqrt{1+d^2}},\frac{1}{c\sqrt{1+d^2}})}\;\frac{d}{\sqrt{1+d^2}}\,;\\
\lim_{p\to\infty} E_\theta[\{\frac{m\wedge R}{\sqrt{p}})\}^2] &= d^2\,;\\
\lim_{p\to\infty} E_\theta\{(\frac{m\wedge R}{\sqrt{p}})\frac{m}{\sqrt{p}}\,\rho_{\frac{p}{2}-1}(m\,R)\} &= \frac{d^3}{\sqrt{1+d^2}}\,,
\end{aligned}$$

given the growth conditions $m = d\sqrt{p}$, $\alpha = c\sqrt{p}$, and the boundedness of $\frac{m\wedge R}{p}$ and $\rho_{p/2-1}$. The result follows then by collecting terms.

Theorem 2. *Let* $\gamma(d) = 3d^2 + \frac{2d^4}{1+d^2} - \frac{d(1+4d^2)}{\sqrt{1+d^2}}$, $d > 0$.

(a) For all $d > 0$, *we have*

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

(b) In particular, we have $\lim_{p\to\infty} \frac{k_1(p,d\sqrt{p})}{\sqrt{p}} = 0$, *for all* $d \leq \sqrt{\frac{1}{3}}$.

Proof. Part (b) follows from (a) as $\gamma(d) \leq 0$ if and only if $d \leq \sqrt{\frac{1}{3}}$ by a direct evaluation. For part (a), we have by Lemma 7 with $m = d\sqrt{p}, \alpha = c\sqrt{p}$, $L = L(\frac{1}{c\sqrt{1+d^2}}, \frac{1}{c\sqrt{1+d^2}})$, $c \in [0, d]$, $d > \sqrt{\frac{1}{3}}$: $\lim_{p->\infty} \frac{\Delta_\alpha(m)}{p} \leq 0$ if and only if $c^2\,L - 2c\frac{d^2}{\sqrt{1+d^2}}\sqrt{L} \leq d^2 - 2\frac{d^3}{\sqrt{1+d^2}}$, or equivalently

$$\lim_{p->\infty} \frac{\Delta_\alpha(m)}{p} \leq 0 \Leftrightarrow (c\,\sqrt{L} - \frac{d^2}{\sqrt{1+d^2}})^2 \leq d^2\,(\frac{d}{\sqrt{1+d^2}} - 1)^2\,. \tag{9}$$

Now, observe that $c\sqrt{L} = \frac{\sqrt{4c^2(1+d^2)+1}-1}{2\,\sqrt{1+d^2}}$ is, as a function of c for $c \in [0, d]$ increasing and bounded above by $d\sqrt{L(\frac{1}{d\sqrt{1+d^2}}, \frac{1}{d\sqrt{1+d^2}})} = \frac{d^2}{\sqrt{1+d^2}}$. Consequently, (9) is equivalent to $\frac{d^2}{\sqrt{1+d^2}} - c\sqrt{L} \leq d\,(1 - \frac{d}{\sqrt{1+d^2}}) \Leftrightarrow \frac{d^2+1/2-\sqrt{c^2(1+d^2)+1/4}}{\sqrt{1+d^2}} \leq \frac{d\,\sqrt{1+d^2}-d^2}{\sqrt{1+d^2}} \Leftrightarrow c^2 \geq 3d^2 + \frac{2d^4}{1+d^2} - \frac{d(1+4d^2)}{\sqrt{1+d^2}} \Leftrightarrow c \geq \sqrt{0 \vee \gamma(d)}\,.$

2.4. Illustrations and implications

We now illustrate how the above results apply (Theorems 1, 2, Corollary 2), and lead to several implications.

Example 1 (continued) *For making inferences for the cases where $m \geq \sqrt{p}$, our above results tell us that the Bayes estimator δ_m dominates δ_{mle}* **iff** *$m \leq k_2(p,m)$, with $k_2(p,m) \approx \sqrt{2p}$ for large p, independently of m.* In comparison to the sufficient condition $m \leq \sqrt{p}$ for dominance given by Marchand and Perron (2001), the new conditions here are necessary and sufficient and asymptotically precise. Moreover, the precise region for dominance (asymptotically) is far more vast, as seen when comparing the required information for application : $\frac{\sum_{i=1}^{p} \theta_i^2}{p} \leq g$; $g = 1, 2$. [8] In conjunction with the frequentist properties of the Bayes estimator δ_m, it is worth recalling earlier work by Berry (1990) and Marchand and Perron (2002) investigating the minimaxity of δ_m. In particular, Berry showed that δ_m was unique minimax iff $m \leq c_0(p)$, while Marchand and Perron established that $c_0(p) \geq \sqrt{p}$ for all $p \geq 1$, and $\lim_{p\to\infty} \frac{c_0(p)}{\sqrt{p}} = \kappa$, with $\kappa \approx 1.150964$.

Example 2. *(Small parameter spaces) As analyzed by Marchand and Perron (2001), and as expanded upon below in Section 3, the maximum likelihood estimator can be very inefficient for small parameter spaces, in the sense of being dominated by any Bayes estimator with respect to an orthogonally invariant prior. For our subclass of Bayesian estimators δ_α, $\alpha \in [0,m]$, Theorem 2, part b, tells us that* **all** *these estimators δ_α will necessarily dominate δ_{mle} for large enough p whenever $\frac{m}{\sqrt{p}} \leq \frac{1}{3}$. The result applies for the particular case of the "zero" estimator $\delta_0 \equiv 0$, with the above asymptotic (sufficient) condition being also necessary by virtue of part (a) of Theorem 2, as $\gamma(d) > 0$ for $d > \sqrt{\frac{1}{3}}$.*

Example 3. *To further visualize the conditions for dominance, as well as interpretations, consider*

[8] or the average squared signal to noise ratio is less or equal than g.

the case $p = 9, m = 4.5$*, as an illustration. Here we have* $d = \frac{m}{\sqrt{p}} = \frac{4.5}{\sqrt{9}} = 1.5$*,* $\gamma(1.5) = 3(3/2)^2 + \frac{2(3/2)^4}{1+(3/2)^2} - \frac{(3/2)(1+4(3/2)^2)}{\sqrt{1+(3/2)^2}} \approx 1.5449$*, so that*

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

from Theorem 2. On the other hand, Theorem 1 tells us that

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

We thus infer from Corollary 2 that, for $p = 9, m = 4.5, \alpha \in [0, 4.5]$*:* δ_α *dominates* δ_{mle} *if and only if* $\alpha \in [k_1(9, 4.5), k_2(9, 4.5)] \approx [3.7288, 4.2426]$*. Observe that both too small and too large values of* α *do not lead to dominance failing to better the risk of* δ_{mle} *on the boundary* S_m *and at* $\theta = 0$ *respectively. Of course, we can evaluate* $k_1(p, m)$ *and* $k_2(p, m)$ *numerically for any pair* (p, m)*. Here, for instance and for comparison's sake, we obtain* $k_1(9, 4.5) \approx 3.835$ *and* $k_2(9, 4.5) \approx 4.223$*, which are values quite close indeed to the approximations above given the sample size* $p = 9$*.*

Example 4. *(Larger parameter spaces) As pointed out by Marchand and Perron (2001), and as indicated by several of their analytical and numerical results, it becomes difficult to exhibit dominators of* δ_{mle} *for large parameter spaces. Focussing on this, we return to Example 3, still with* $p = 9$ *but with the objective of studying the impact of larger* m *on the necessary and sufficient condition for* δ_α *to dominate* δ_{mle}*. On one hand, dominance requires* $\alpha \leq k_2(9, m)$*, with* $k_2(9, m) \approx 3\sqrt{2}$ *by Theorem 1. On the other hand, we require* $\alpha \geq k_1(9, m)$*, with* $k_1(p, m) \approx 3\sqrt{\gamma(d)}$*,* $d = m/3$*. Now, it is easy to show that* $\gamma(d)$ *increases in* d*. Therefore, to avoid incompatibility and to extract a non-empty subclass of dominators* δ_α*, we require (approximately)* $\gamma(d) \leq 2 \Leftrightarrow d \leq d_0 = \gamma^{-1}(2) \approx 1.656$*.*[9] *Therefore, for large enough parameter spaces with* $m > d_0\sqrt{p}$*, where the constraint is much less informative,* δ_{mle} *is rather efficient in the sense that no dominator can be found in the class of*

[9]Still, dominance is possible for an average squared signal to noise ratio of less than $(\gamma^{-1}(2))^2 \approx 2.74$ (approx.).

estimators δ_α. Of course, δ_{mle} remains inadmissible for any (m, p), but perhaps not seriously so for larger m. This contrasts certainly a great deal with the situation for small parameter spaces as illustrated in Example 2 and further expanded upon Section 3.

Remark 1. *One may wish to consider fixed α, and obtain conditions on m for dominance. In particular, the choice $\alpha = \sqrt{p}$ is interesting since the necessary and sufficient condition for dominance only involves the risks on the boundary of $\Theta(m)$, as seen by Corollary 1. Proceeding as for the previous example, we obtain that dominance occurs only for small enough m, i.e., $\sqrt{p} \le m \le \gamma^{-1}(1)\sqrt{p}$ with $\gamma^{-1}(1) \approx 1.281$.*

3. A universal dominance result

Here, we expand on a previously unknown result applicable to a larger class of Bayes estimators and derived as a corollary of results given by Marchand and Perron (2001). Indeed, as described in the Introduction, they show that for sufficiently small parameter space, i.e., $m \le c_1(p)$ where $c_1(p)$ is defined in Definition 1, a very large class of equivariant shrinkage estimators, which includes **all** Bayesian estimators with respect to orthogonally invariant priors supported on the parameter space, dominate δ_{mle}. Here, the new finding consists of a more explicit representation for $c_1(p)$, given in Corollary 3, and showing that the universal dominance condition requires asymptotically in p that $m \le \sqrt{\frac{p}{3}}$. Note that, notwithstanding its asymptotic nature, this sufficient condition is also necessary since the cutoff $m \le \sqrt{\frac{p}{3}}$ is necessary for the specific case of the prior with a point mass at 0 (Example 2). We pursue with a formal presentation and derivation of these results.

Definition 1. *Let $m_1(p)$ represent the unique positive value of m such that $\bar{\alpha}(m) = \frac{1}{2}$, where $\bar{\alpha}(m) = E_\theta[\rho_{p/2-1}(m\,R) \mid R > m)]$, with $\theta \in S_m$. Let $c_1(p) = m_1(p) \wedge \sqrt{\frac{p}{2}}$.*

Lemma 8. *(Marchand and Perron, 2001, Corollaries 2 and 3) Let δ be any equivariant estimator with respect to orthogonal transformations such that, for all $x \in \Re^p$, :* **(i)** $\|\delta(x)\| \leq \|\delta_{mle}(x)\|$ *(shrinkage), and* **(ii)** $\frac{\|\delta(x)\|}{\|x\|}$ *increases as $\|x\|$ increases; then δ dominates δ_{mle} as soon as $m \leq c_1(p)$. In particular, the class of such dominating δ's includes the subclass of all Bayes estimators with respect to orthogonally invariant priors on $\Theta(m)$.*[10]

Note that the priors of the above lemma need not be absolutely continuous with respect to the uniform measure (or Lebesgue measure truncated) on $\Theta(m)$ and hence include uniform priors on spheres S_α, $\alpha \in [0, m]$. Also, note that the above lemma is applicable for all $p \geq 1$, but our asymptotic analysis informs more precisely us on cases where the dimension p is large.

Corollary 3. *In the context of Lemma 8, where the subclass of all Bayesian estimators with respect to orthogonally invariant priors dominate δ_{mle} as soon as $m \leq c_1(p)$, we have*

$$\lim_{p\to\infty} \frac{c_1(p)}{\sqrt{p}} = \sqrt{\frac{1}{3}} \ . \tag{10}$$

Proof. It suffices to prove that, for all $d > 0$,

$$\lim_{p\to\infty} \bar{\alpha}(d\sqrt{p}) = \frac{d}{\sqrt{1+d^2}} \,, \tag{11}$$

since along with the definition $m_1(p) = \bar{\alpha}^{-1}(\frac{1}{2})$, we obtain (10). To establish (11), observe that for $R_p^2 \sim \chi_p^2(d^2 p)$, with $d > 0$ and $p \to \infty$, the sequence of random variables $\frac{R_p^2}{p}|R_p > d\sqrt{p}$ converges in probability to $1 + d^2$ since $\frac{R_p^2}{p}$ converges in probability to $1 + d^2$ and $P(\frac{R_p^2}{p} \geq d^2) \to 1$ (as $p \to \infty$). Consequently, the sequence of random variables $\rho_{p/2-1}(d\sqrt{p}R_p)|R_p > d\sqrt{p})$ converges in probability, as $p \to \infty$, to $\frac{d}{\sqrt{1+d^2}}$, as in Lemma 6. Finally, this convergence in probability and the boundedness of ρ imply (11).[11]

[10]For such Bayesian estimators, Marchand and Perron (2001) show that the shrinkage condition is implied by the condition $m \leq \sqrt{p}$ and hence by $m \leq c_1(p)$.

[11]Result (11) previously appeared in Gueye (2003).

Remark 2. *Numerical evaluations of* $c_1(p)$ *were given by Marchand and Perron (2001, Table 1) for* $1 \leq p \leq 10$. *An inspection of these values shows that the quantities* $\frac{c_1(p)}{\sqrt{p}}$ *fluctuate very closely*[12] *about* $\sqrt{\frac{1}{3}}$, *suggesting that the large sample approximation for* $\frac{c_1(p)}{\sqrt{p}}$ *is quite accurate, even for small* p. *Therefore, the rule of thumb: "Average squared signal to noise ratio of less than or equal to* $1/3$*" for universal dominance seems to be quite accurate even for small sample sizes* p.

Acknowledgements

The research work of Éric Marchand is partially supported by NSERC of Canada. The support of the ANR grant 08-EMER-002 is gratefully acknowledged by Dominique Fourdrinier. Many thanks to a referee for useful comments and for pointing out an error in a previous version of Theorem 2. Finally, we are also thankful to François Perron for useful discussions, and to Othmane Kortbi for numerical evaluations.

References

Amos, D.E. (1974). Computation of modified Bessel functions and their ratios. *Mathematics of Computation*, **28**, 239-251.

Berry, C. (1990). Minimax estimation of a bounded normal mean vector. *Journal of Multivariate Analysis*, **35**, 130-139.

Brown, L., Johnstone, I., & MacGibbon, B. (1981). Variation diminishing transformations: a direct approach to total positivity and its statistical applications, *JASA*, **376**, 824-832.

[12]They obtain for instance $c_1(9) = 1.7367$ in comparison to our above approximation $c_1(9) \approx \sqrt{3} = 1.73205$.. Or again, their results satisfy for instance $|\frac{c_1(p)}{\sqrt{p}} - \sqrt{\frac{1}{3}}| \leq 0.01$ for $5 \leq p \leq 10$.

Gueye, N.R. (2003). *Problèmes d'estimation de paramètres avec restriction sur l'espace des paramètres.* Ph.D. thesis, Université de Montréal.

Karlin, S. (1957). Pólya type distributions, II. Annals of Mathematical Statistics, **28**, 281-309.

Marchand, É. & Perron, F. (2005). Improving on the mle of a bounded mean for spherical distributions, *Journal of Multivariate Analysis*, **92**, 227-238.

Marchand, É. & Perron, F. (2002). On the minimax estimator of a bounded normal mean. *Statistics & Probability Letters*, **58**, 327-333.

Marchand, É. & Perron, F. (2001). Improving on the MLE of a bounded normal mean. *Annals of Statistics*, **29**, 1066-1081.

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

Robert, C. (1990). Modified Bessel functions and their application in probability and statistics. *Statistics & Probability Letters*, **9**, 155-161.

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

van Eeden, C. (2006). *Restricted parameter space problems: Admissibility and minimaxity properties.* Lecture Notes in Statistics, **188**, Springer.

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