Truncated linear estimation of a bounded multivariate normal mean

Othmane KORTBI and Éric MARCHAND *

May 24, 2011

Abstract

We consider the problem of estimating the mean θ of a $N_p(\theta, I_p)$ distribution with squared error loss and under the constraint $\|\theta\| \leq m$, for some constant $m > 0$. Using Stein's identity to obtain unbiased estimates of risk, Karlin's sign change arguments, and conditional risk analysis, we compare the risk performance of truncated linear estimators with that of the maximum likelihood estimator δ_{mle}. We obtain for fixed (m, p) sufficient conditions for dominance. An asymptotic framework is developed, where we demonstrate that the truncated linear minimax estimator dominates δ_{mle}, and where we obtain simple and accurate measures of relative improvement in risk. Numerical evaluations illustrate the effectiveness of the asymptotic framework for approximating the risks for moderate or large values of p.

AMS 2010 subject classifications. 62C20, 62C99, 62F10, 62F30, 62H12.

Keywords and phrases: Restricted parameters, point estimation, squared error loss, dominance, truncated linear estimators, truncated linear minimax, maximum likelihood, multivariate normal, asymptotic analysis.

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

*Département de mathématiques, Université de Sherbrooke, Canada.

$\tau_1, \ldots, \tau_p, \sigma^2, m$. 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, and 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 descriptions of such restricted parameter space problems, we refer to the paper of Marchand and Strawderman [13], as well as the monograph of van Eeden [15], among others.

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 $\|\delta - \theta\|^2$. As a follow up to the work of Marchand and Perron [10], as well as Fourdrinier and Marchand [5], we focus here on the determination of estimators that improve upon the benchmark, but inadmissible, maximum likelihood estimator given by $\delta_{\text{mle}}(x) = (m \wedge \|x\|)\frac{x}{\|x\|}$. For the purposes of introducing the findings in this paper, here is a key result synthesized from [5] and [10].

Theorem 1.1. *For sufficiently small m, say $m \leq c_1(p)$, all equivariant estimators $\delta(X)$ with respect to orthogonal transformations taking values in $\Theta(m)$ with $\|\delta(x)\|$ nondecreasing in $\|x\|$, as well as all Bayes estimators δ_π with respect to an orthogonally invariant prior π supported on $\Theta(m)$ dominate δ_{mle}. Furthermore, we have $\lim_{p\to\infty} \frac{c_1(p)}{\sqrt{p}} = \sqrt{\frac{1}{3}}$.*

We refer to [10] and [5] for various other dominance results, such as those pertaining to Bayes estimators with respect to: the boundary uniform prior (δ_{BU}), uniform priors on spheres centered at the origin, the fully uniform prior on $\Theta(m)$ (δ_U), and other absolutely continuous priors on $\Theta(m)$.

Another unresolved, fundamental, and difficult issue concerns the degree to which δ_{mle} can be improved upon. There is certainly numerical evidence, provided in [10] for various pairs (m, p), that suggests substantial gains over a large part of parameter space for small m relative to p achieved by the Bayesian estimators δ_U and δ_{BU}, among others. As well, the universal dominance result in Theorem 1.1 strongly hints at large gains for quite small m. On the other hand, these gains are either much less impressive or vanish for moderate or large m relative to p. Notwithstanding the above indirect evidence, there is a lack of crisp analytical results providing bounds or good approximations for the degree or the percentage improvement.

To meet this challenge, we study the class of truncated linear estimators $\delta_a(x) = (a\,\|x\| \wedge m)\frac{x}{\|x\|}$, with $0 \leq a \leq 1$. The class of shrinking (with

respect to δ_{mle}) estimators are of interest: their form is simple and appealing, the class contains the mle δ_1, the trivial (but Bayes) estimator δ_0, and the truncated linear minimax estimator δ_{tlx} with $a = \frac{m^2}{m^2+p}$. Furthermore, such estimators are perhaps just as, or more, plausible than δ_{mle}. For instance, for $a \in [0,1)$, δ_a truncates an admissible Bayes (with respect to the prior $\theta \sim N_p(0,(a/(1-a))I_p)$ estimator aX onto the parameter space $\Theta(m)$, while δ_{mle} truncates an inadmissible for $p \geq 3$ procedure X onto $\Theta(m)$. Also, [8] provides analytical evidence limiting the loss of efficiency incurred with the linear minimax procedure, and hence with δ_{tlx}, at the boundary of $\Theta(m)$ as opposed to the optimal best equivariant estimator (for $\|\theta\| = m$) δ_{BU}. For small enough m, the universal dominance result tells us that all such δ_a's, $a \in [0,1)$, dominate δ_{mle}, while finer analyses for $p = 1$ were carried in [9] for δ_a and δ_{tlx}, and in [4] for δ_0.

We provide below further analysis for all (finite) dimensions p and we describe precisely in an asymptotic framework the relative gain or loss in efficiency in the risk comparison of δ_a and δ_{mle}. Moreover, we determine exactly which estimators δ_a improve on δ_{mle}; and δ_{tlx} turns out to always be such an improvement. More precisely, we consider a sequence $\{(m,p) = (d\sqrt{p}\,,\,p), p \geq 1\}$ for fixed $d > 0$, and show that there exists, for all $d > 0$, a large enough value $p_0(d)$ and values of a (which depend on d) such that δ_a dominates δ_{mle} for all $p \geq p_0(d)$. With such a technique exploited previously in [5] and [10], we also extend this idea to study sequences $\{(m, \|\theta\|, p) = (d\sqrt{p}\,, c\sqrt{p}\,,\,p), p \geq 1\}$, for $c \in [0,d]$, $d > 0$, obtaining sharp asymptotic analysis concerning the relative difference in risks, and solid approximations for finite p. Observe that Theorem 1.1's cutoff point is proportional to $\sqrt{p}$, which translates to the average squared standardized deviation $(\frac{\theta_i - \tau_i}{\sigma})^2$ being less than or equal to 1/3 for universal dominance. Analogously, the asymptotic setup incorporates growth conditions with m and $\|\theta\|$ proportional to $\sqrt{p}$, with findings hence being amenable to similar interpretations. We also point out as an additional motivating factor that an increasing number of modern statistical inference problems relate to large number of parameters (see for instance [3] and the references therein).

This paper is organized as follows. In Section 2, we make use of Stein's ([14]) technique for estimating unbiasedly the difference in risks between δ_a and δ_{mle} along with sign change arguments, to obtain necessary and sufficient conditions for δ_a to dominate δ_{mle}. Capitalizing on this condition (i.e., Corollary 2.1) and making use of a conditional (on $\|X\|$) risk decomposition introduced in [10], Section 3.1 includes findings for finite p. Section 3.2 exploits the asymptotic framework to produce inferences and these include a precise determination of dominating δ_a's, along exact expressions for the asymptotic relative difference in risks. Throughout, examples, numerical evaluations, and visual representations are provided. Section 4 contains

concluding remarks.

2 An unbiased estimator for the difference of risks

Throughout this manuscript, we shall denote $\|X\|$, $\|x\|$ and $\|\theta\|$ by R, r and λ, respectively. An estimator $\delta(X)$ of θ is evaluated by its risk function

$$R(\theta,\delta) = E_\theta\left[\|\delta(X)-\theta\|^2\right],\ \theta\in\Theta(m), \tag{2.1}$$

where E_θ denotes the expectation with respect to $X\sim\mathcal{N}_p(\theta, I_p)$. We will make use of Stein's identity ([14]) for an unbiased estimator of risk which allows us to write for an estimator $\delta(X)=X+h(X)$, with h any function weakly differentiable and $E_\theta\left[h(\|X\|^2)\right]<\infty$, as

$$R(\theta,\delta) = E_\theta\left[\|\delta(X)-\theta\|^2)\right] = E_\theta\left[p+2divh(X)+\|h(X)\|^2\right], \tag{2.2}$$

with div being the divergence operator. We proceed with the derivation of an unbiased estimator $D_a(\|X\|)$ of the risk difference $R(\theta,\delta_{mle})-R(\theta,\delta_a)$, as well as a sign change analysis of $D_a(\cdot)$.

Lemma 2.1. **(a)** *For $0\le a\le 1$, an unbiased estimator of the difference in risks $R(\theta,\delta_{mle})-R(\theta,\delta_a)$ is given by $D_a(\|X\|)=D_{a,1}(\|X\|)\ [0\le\|X\|\le m]+D_{a,2}(\|X\|)\ [m\le\|X\|\le m/a]$, with*

$$D_{a,1}(r) = 2p-2ap-(a-1)^2r^2,$$

and

$$D_{a,2}(r) = 2(p-1)\frac{m}{r}+(m-r)^2-2ap-(a-1)^2r^2. \tag{2.3}$$

(b) *For $p\ge 1$, and $0\le a<1$, $D_a(r)$ changes signs as a function of $r\in[0,\infty)$ once from positive to negative.*

Proof. **(a)** Since $\delta_a(x)=x+h_a(x)$ with $h_a(x)=\{(a-1)\ [0\le r\le m/a]$ $+(\frac{m}{r}-1)\ [r>m/a]\}\,x$, we obtain

$$\begin{aligned}2divh_a(x)+\|h_a(x)\|^2 &= \left\{2ap-2p+(a-1)^2r^2\right\}\ 1_{\left[0,\frac{m}{a}\right]}(r)\\ &\quad+\left\{2(p-1)\frac{m}{r}-2p+(m-r)^2\right\}1_{(\frac{m}{a},\infty)}(r);\end{aligned}$$

and, by virtue of Stein's identity (2.2), $R(\theta, \delta_a) = E_\theta[\eta_a(X)]$ with

$$\eta_a(X) = \left\{2ap - p + (a-1)^2 r^2\right\} \; 1_{\left[0, \frac{m}{a}\right]}(r) + \left\{2(p-1)\frac{m}{r} - p + (m-r)^2\right\} 1_{(\frac{m}{a}, \infty)}(r) \,. \quad (2.4)$$

Finally, the given expression for the unbiased estimator $D_a(\|X\|)$ follows directly from (2.4) as $\delta_{\text{mle}} \equiv \delta_1$ with our notation.

(b) First, observe that $D_a(\cdot)$ decreases on $[0, m)$ with $D_a(0) = 2p(1-a) > 0$, and is thus either positive, or changes signs once from positive to negative on $[0, m)$. Secondly, $D_a(\cdot)$ is convex on $(m, m/a]$ with $D_a(\frac{m}{a}) = -2a \leq 0$, which implies that $D_a(\cdot)$ is either negative, or changes signs once from positive to negative on $(m, m/a]$. Thirdly, we have $\lim_{r \to m^+} D_a(r) = \lim_{r \to m^-} D_a(r) - 2$, in other words a drop discontinuity at m. The result follows from the above properties and since $D_a(\cdot)$ is everywhere continuous on $[0, \infty)$ except at m and m/a. $\square$

With the estimators δ_a and δ_{mle} being equivariant (with respect to orthogonal transformations), and hence with their risks constant on spheres centered at the origin, we denote their difference in risks as

$$\Delta_a^p(\lambda) = R(\theta, \delta_{mle}) - R(\theta, \delta_a) \,, \quad (2.5)$$

for $\theta \in S_\lambda = \{\theta \in \Re^p : \|\theta\| = \lambda\}$, $\lambda \in [0, m]$, emphasizing the role of the dimension p. Recall also that, the probability distribution of $\|X\|^2$ is $\chi_p^2(\lambda^2)$, so that the potential sign changes of $\Delta_a(\lambda) = E_\theta\left[D_a(\|X\|)\right]$ are controlled by the variational properties of $D_a(\cdot)$ in terms of sign changes (e.g., [7]; [2]). Therefore, according to Lemma 2.1 and since $\Delta_a^p(0) = E_0[\|\delta_{\text{mle}}(X)\|^2 - \|\delta_a(X)\|^2] > 0$, we deduce that $\Delta_a^p(\lambda)$ is, as function of $\lambda \in [0, m]$, either positive, or changes signs once from positive to negative. So, we conclude this section with the following.

Corollary 2.1. *For any $0 \leq a < 1$, δ_a dominates δ_{mle} if and only if*

$$\Delta_a^p(m) \geq 0 \,.$$

3 Analytical risk comparisons

In this section, we focus on: **(A)** dominance results obtained as an application of Corollary 2.1, achieved by further risk analysis at the boundary S_m of the parameter space both for finite p (subsection 3.1) and within an asymptotic framework when $p \to \infty$ (subsection 3.2); and **(B)** approximations for

the actual difference in risks $\Delta_a^p(\lambda)$; $\lambda \in [0, m]$ (subsection 3.2). In what follows, we write $\|\delta_a(x)\| = g_a(r) = m \wedge ar$, $\left\|\delta_{\text{mle}}(x)\right\| = g_1(r) = m \wedge r$ and, similarly, we note and recall (e.g., [10]) that $\|\delta_{\pi_\lambda}(x)\| = g_{\pi_\lambda}(r) = \lambda\, \rho_{p/2-1}(\lambda r)$, where $\rho_\nu(t) = I_{\nu+1}(t)/I_\nu(t), t > 0, \nu \geq -1/2$, I_ν representing the modified Bessel function of order ν. Some further properties of ρ_ν, presented in Lemma 5.1 of the Appendix, will be used below. We require the following risk and difference in risks representations, achieved by conditioning on $R = \|X\|$, and bringing into play the Bayes estimator δ_{π_λ} associated with a uniform prior on S_λ. These representations were introduced by Marchand and Perron [10], and were further exploited in [12] and [5].

Lemma 3.1. **(a)** *For an equivariant estimator $\delta_g(x) = g(\|x\|)\frac{x}{\|x\|}$, we have for $\theta \in S_\lambda$*

$$R(\theta, \delta_g) = R(\theta, \delta_{\pi_\lambda}) + E_\theta\left[\{\|\delta_g(X)\| - \|\delta_{\pi_\lambda}(X)\|\}^2\right], \tag{3.1}$$

where δ_{π_λ} is the Bayes estimator with respect to uniform prior on the sphere S_λ.

(b) *For $\theta \in S_\lambda$, the difference in risks between the estimators δ_{mle} and δ_a is given by*

$$\Delta_a^p(\lambda) = E_\theta\left\{(g_{mle}(R) - g_a(R))\,(g_{mle}(R) + g_a(R) - 2g_{\pi_\lambda}(R))\right\}. \tag{3.2}$$

Proof. Part (b) follows from part (a). Part (a) is given in [10].

3.1 Finite p

We seek analytical results describing the subset of parameter space values $\{\theta \in \Theta(m) : \Delta_a^p(\|\theta\|) \geq 0\}$, as well as dominance situations (δ_a dominates δ_{mle}) where $\Delta_a^p(\lambda) \geq 0$ for all $\lambda \in [0, m]$. Beyond the universal cutoff point of Theorem 1.1, the following inferences may be derived from (3.2) for cases where $\lambda = \|\theta\| < \sqrt{p}$.

Theorem 3.1. **(a)** *For $a \in (0, 1)$, $\lambda < \sqrt{p}$, we have $\Delta_a^p(\lambda) \geq 0$ whenever $a \geq (\frac{2\lambda^2}{p} - 1) \vee (\frac{\lambda}{m}\rho_{p/2-1}(\lambda m))$;*

(b) *For $a \in (0, 1)$, $m < \sqrt{p}$, δ_a dominates δ_{mle} whenever $a \geq (\frac{2m^2}{p} - 1) \vee \rho_{p/2-1}(m^2)$.*

Proof. It suffices to establish (**a**), since (**b**) follows from (**a**) as $\rho(\cdot)$ is increasing and positive on $(0,\infty)$. Since $g_a(r) \le g_{\text{mle}}(r)$ for all $a \in (0,1), r > 0$, it follows from (3.2) that $\Delta_a^p(\lambda) \ge 0$ as soon as

$$g_{\text{mle}}(r) + g_a(r) \ge 2g_{\pi_\lambda}(r) = 2\lambda\ \rho_{p/2-1}(\lambda r) \ \text{ for all } r > 0. \tag{3.3}$$

Now consider cases: **(i)** $r \in (0,m]$, **(ii)** $r \in (m, m/a)$, and **(iii)** $r \ge m/a$. In **(iii)**, (3.3) is satisfied since $g_{\text{mle}}(r) + g_a(r) = 2m$ for $r \ge m/a$, while $\rho_{p/2-1}(\cdot)$ is bounded above by 1 (see Lemma 5.1 of the Appendix). In **(i)**, condition (3.3) is equivalent to:

$$\begin{aligned} & r(1+a) && \ge 2\lambda\rho_{p/2-1}(\lambda r) \ \text{ for all } r \in (0,m] \\ \Leftrightarrow\quad & (1+a) && \ge 2\lambda^2 \sup_{r\in(0,m]} \{\frac{\rho_{p/2-1}(\lambda r))}{\lambda r}\} \\ \Leftrightarrow\quad & a && \ge \frac{2\lambda^2}{p} - 1, \end{aligned}$$

using the properties of Lemma 5.1. Finally in **(ii)**, condition (3.3) becomes

$$\frac{m}{r} + a \ge 2\lambda^2 \frac{\rho_{p/2-1}(\lambda r))}{\lambda r}, \ \text{ for all } \ r \in (m, m/a),$$

which is satisfied whenever

$$\begin{aligned} & \inf_{r\in(m,m/a)}\{\tfrac{m}{r} + a\} \ge 2\lambda^2 \sup_{r\in(m,m/a)}\{\tfrac{\rho_{p/2-1}(\lambda r))}{\lambda r}\} \\ \Leftrightarrow\quad & a \ge \tfrac{\lambda}{m}\rho_{p/2-1}(\lambda m), \end{aligned}$$

using Lemma 5.1 again. This establishes the lemma. □

Before proceeding with an illustration, we point out that the actual comparison of risks is stronger than stated in Theorem 3.1 as it applies to all conditional on $\|X\| = r$ risk differences.

Example 1. *The performances of δ_{tlx} and δ_{mle} are compared in Figure 1 for two combinations of (m,p) with δ_{tlx} seen to improve on δ_{mle} for a large part of the parameter space $\Theta(m)$. These represent situations where part (b) of Theorem 3.1 does not apply as $m \ge \sqrt{p}$. However, part (a) of the Theorem applies for $a = \frac{m^2}{m^2+p}$, and we find by evaluation that δ_{tlx} has a smaller risk than δ_{mle} when $\lambda \le \sqrt{11/13} \approx 0.9199$ for $(m,p) = (1.5,1)$, and when $\lambda \le \sqrt{21/8} \approx 1.6202$ for $(m,p) = (3,3)$. These are sufficient conditions of course and the actual regions where δ_{tlx} improves on δ_{mle} stretches beyond these limits, to about $\lambda \le 1.18$ and $\lambda \le 2.72$ respectively. Observe as well that the graphs illustrate Section 2's sign change analysis for $\Delta_a^p(\|\theta\|)$, as well as Corollary 2.1.*

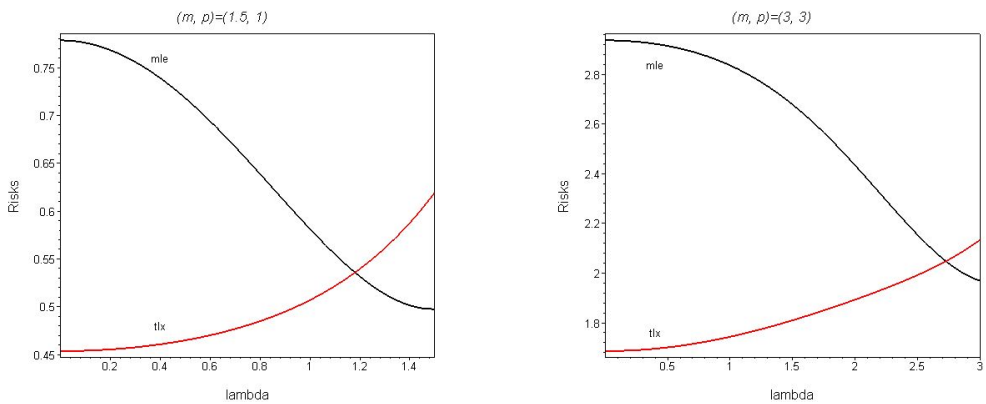

Figure 1: Risks of δ_{mle} and δ_{tlx} for $(m,p) = (1.5, 1)$ and $(m,p) = (3,3)$.

Now consider the set $C(m,p) = \{a \in [0,1) : \Delta_a^p(m) \geq 0\}$, representing the subclass of estimators δ_a that dominate δ_{mle} for a given (m,p) in view of Corollary 2.1. Part (b) of Theorem 3.1 produces a clear result when $m < \sqrt{p}$: $C(m,p)$ includes an interval of the form $[a(m,p), 1)$. For instance, we have for $(m,p) = (1,3)$: $a(1,3) = \rho_{1/2}(1) = \coth(1) - 1 = 2/(e^2 - 1) \approx 0.3130$ (see Lemma 5.1). This is illustrated in Figure 2 where the actual set $C(1,3)$ is, of course, larger and coincides almost with the full interval $[0,1)$. For $m \geq \sqrt{p}$, as Theorem 3.1 suggests, it seems plausible that $C(m,p)$ also includes, and is given by an interval of the form $[a(m,p), 1)$ when it is not empty. Such a situation is shown in Figure 2 and where we have evaluated $C(6,9) \approx [0.833, 1)$. We pursue in this direction by identifying instances where $\Delta_a^p(m)$ decreases in a locally at $a = 1$, and thus guarantees that δ_a dominates δ_{mle} for values of a such that $1 - a$ is sufficiently small.

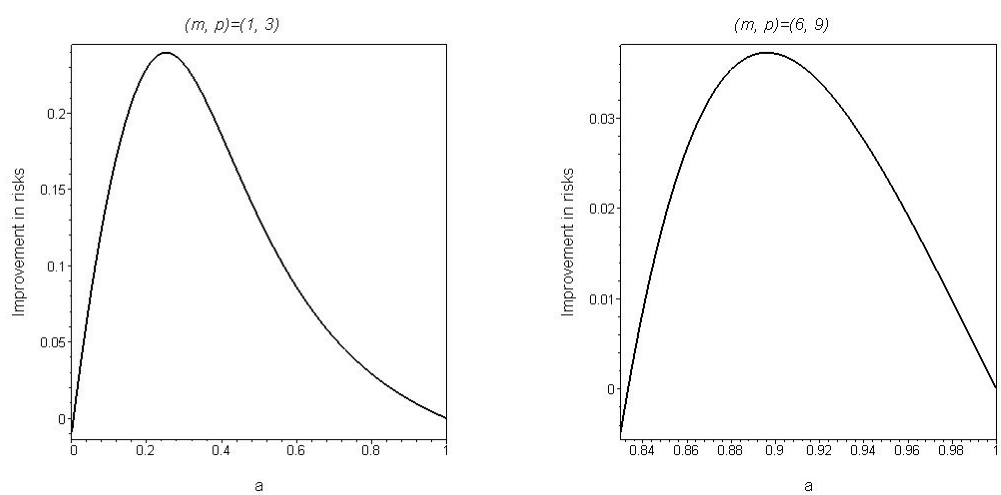

Figure 2: Boundary risk difference $\Delta_a^p(m)$ between δ_{mle} and δ_a for $(m,p) = (6,9)$ and $(m,p) = (1,3)$.

Theorem 3.2. *For all $p \geq 1$, there exists $m_p > \sqrt{p}$ such that for any $m \leq m_p$, there exists $a(m,p)$ for which δ_a dominates δ_{mle} for all $a \in [a(m,p), 1)$. The cutoff point m_p is given as the smallest positive root of $H_p(\cdot)$, with $H_p(m) = E[(m\,Y\,\rho_{p/2-1}(mY) - Y^2)\,1_{(0,m)}(Y)]$, and $Y \sim \sqrt{\chi^2_p(m^2)}$.*

Proof. By virtue of Corollary 2.1, it suffices to focus on the behavior of $\Delta^p_a(m)$ for fixed (m,p) and in a neighbourhood of $a = 1$. From (3.2), we obtain

$$\frac{\partial}{\partial a}\Delta^p_a(m) = 2\int_0^{m/a} r f_{p,m}(r)\left[m\rho_{p/2-1}(mr) - ar\right]\,dr\,,$$

where $f_{p,m}(\cdot)$ represents the density of R (i.e., of Y) when $\|\theta\| = m$. Hence, at $a = 1$, we have

$$\begin{aligned}\frac{\partial}{\partial a}\,\Delta^p_a(m)|_{a=1^-} &= 2\int_0^m r^2 f_{p,m}(r)\left[m\frac{\rho_{p/2-1}(mr)}{r} - 1\right]\,dr \\ &= 2H_p\,(m)\ \ .\end{aligned}$$

Now, using Lemma 5.1 of the Appendix, we have for all $p \geq 1$ and any $m \leq \sqrt{p}$

$$m\frac{\rho_{p/2-1}(mr)}{r} - 1 < \frac{m^2}{p} - 1 \leq 0\,,$$

for all $0 < r \leq m$, which in turn implies that $H_p\,(m) < 0$ for $0 \leq m \leq \sqrt{p}$. The result follows from continuity of $H_p(\cdot)$. □

Example 2. *For $p = 9$, we obtain numerically $m_9 \approx 7.408$ as the smallest positive root of $H_9(\cdot)$. Theorem 3.2 tells us that for any $m \leq 7.408$ (approx.), there exists dominating estimators δ_a of δ_{mle} for small enough positive $1-a$. For instance, if we consider $m = 5$, we obtain numerically $a(5,9) \approx 0.6979$, and hence the estimators δ_a with $a \in [0.6979, 1)$ (approx.) dominate δ_{mle}. These include δ_{tlx} since $a_{tlx} = 25/34 \geq a(5,9)$.*

We conjecture but have been unable to prove that, for all p, $C(m,p)$ is empty for $m > m_p$. For the univariate case $p = 1$, we go further into the analysis with regards to Theorem 3.2's cutoff point m_1, as well as properties of $R(m,\delta_a)$ and $C(m,1)$. We will require the following expressions concerning the risk $R(m,\delta_a)$ and its derivative $\frac{d}{da}R\,(m,\delta_a)$. We denote hereafter ϕ and Φ as the standard normal pdf and cdf respectively.

Lemma 3.2. *For $p = 1$ and $a \in (0,1)$, we have*

$$\begin{aligned}R\,(m,\delta_a) = &\left(a^2\left(1+m^2\right) - 2am^2\right)(B - A) \\ &+ m\left(a^2 - 3a\right)D - m\left(a^2 - a\right)C + m^2\,(4 - A - 3B)\ ,\end{aligned} \quad (3.4)$$

and

$$\frac{d}{da}R\left(m,\delta_a\right)=\left(2a+2m^2\left(a-1\right)\right)\left(B-A\right)+2m\left(a-2\right)D-2maC\,,$$

where $A=\Phi\left(m-m/a\right)$, $B=\Phi\left(m+m/a\right)$, $C=\phi\left(m-m/a\right)$, *and* $D=\phi\left(m+m/a\right)$.

Proof. See the appendix. □

Here is a follow up to Theorem 3.2 for the univariate case.

Theorem 3.3. *For* $p=1$, *Theorem 3.2 applies with* $m_1\approx 1.1404$ *being the unique root in* $(0,\infty)$ *of* $H_1(m)=-m\phi\left(0\right)+m\phi\left(2m\right)-\frac{1}{2}+\Phi\left(2m\right)$. *Furthermore if* $m>m_1$, *then the set* $C(m,1)=\{a\in[0,1):\Delta_a^1(m)\geq 0\}$ *does not contain an interval of the form* $[a(m,1),1)$.

Proof. From Lemma 3.2, we obtain $\frac{\partial}{\partial a}\Delta_a^1(m)|_{a=1^-}=2H_1(m)$. Observe now that $H_1(0)=0$, $H_1'\left(m\right)=\phi\left(2m\right)-\phi\left(0\right)+4m^2\phi\left(2m\right)$, and $H_1''\left(m\right)=4m\,\phi\left(2m\right)\left(1-4m^2\right)$. Since $H_1'\left(0\right)=0$, and $H_1'(m)\to\phi(0)>0$ when $m\to\infty$, we infer that $H_1'\left(m\right)$ changes signs once from positive to negative on $(0,\infty)$, and we conclude indeed that $H_1\left(m\right)$ has a unique positive root m_1, which implies that $\frac{\partial}{\partial a}\Delta_a^1(m)|_{a=1^-}\leq 0$ if and only if $m\leq m_1$, establishing the result. □

Lemma 3.3. *The function* $R\left(m,\delta_a\right)$ *is decreasing in* a *if* $0\leq a\leq\frac{m^2}{m^2+1}$ *and increasing when* $m<m_1\approx 1.1404$ *and* $b(m)\leq a<1$, *with* $b(m)=\frac{m^2(\Phi(2m)-1/2)+2m\phi(2m)}{(1+m^2)(\Phi(2m)-1/2)+m\phi(2m)-m\phi(0)}$.

Proof. See the Appendix.

As an illustration of the above univariate results, we take $m=0.95$. Theorems 3.2 and 3.3 tell us that there are necessarily some δ_a's, with $1-a$ small enough and positive, that dominate δ_{mle}. An evaluation leads to dominance for $a\in[a(0.95,1),1)\approx[0.5180,1)$, but this does not include $\delta_{\text{tlx}}=\delta_{361/761}$. The value $b(m)$ provides, for $m\leq m_1$, an explicit upper bound for $a(m,1)$. Here, for instance, we obtain $b(0.95)\approx 0.733$.

Previous findings related to the univariate case were given by Dou and van Eeden [4] and Marchand et al. [9]. Dou and van Eeden have shown that the trivial estimator $\delta_0=0$ dominates δ_{mle} if and only if $0<m\leq c_0\approx 0.52044$. By virtue of Lemma 3.3's decreasing property and Corollary 2.1, Dou and van Eeden's result implies automatically that δ_a dominates δ_{mle} for $p=1$, $m\leq c_0$, $a\in(0,a_0]$ with $a_0>m^2/(m^2+1)$. This includes the truncated linear minimax δ_{tlx}, but [9] have shown by a different route the stronger result that δ_{tlx} dominates δ_{mle} whenever $m\leq c_1\approx 0.60936$.

3.2 An asymptotic framework and inferences for large p

In this section, we not only investigate Corollary 2.1's dominance condition in an asymptotic framework, but we also provide accurate approximations for the risk differences $\Delta_a^p(\lambda)$ in (2.5), as well as for the relative risk differences

$$w_a^p(m,\lambda) = \frac{\Delta_a^p(\lambda)}{R(\theta,\delta_{\text{mle}})}\,, \text{ with } \theta \in S_\lambda. \tag{3.5}$$

To achieve this, we consider a sequence of problems $\{(m,\lambda,p) = (d\sqrt{p}, c\sqrt{p}, p)\,; p \geq 1\}$, with $d > 0$ and $0 \leq c \leq d$, and study the corresponding limiting value of $w_a^p(d\sqrt{p}, c\sqrt{p})$ as p grows to $+\infty$. We will denote this value as

$$h(a,d,c) \;=\; \lim_{p\to\infty} w_a^p(d\sqrt{p}, c\sqrt{p})\,. \tag{3.6}$$

For assessing whether or not δ_a; $a \in [0,1)$; dominates δ_{mle}, it is true, by virtue of Corollary 2.1, that we only need to consider $w_a^p(d\sqrt{p}, c\sqrt{p})$, or $h(a,d,c)$, for $c = d$, but our objective here is also to quantify explicitly the relative gains in risk. In our findings below, δ_a is taken to be independent of p. For instance, observe for the truncated linear estimator, that $\delta_{\text{tlx}}(x) = \delta_{\frac{d^2}{d^2+1}}(x)$ with the setting $m = d\sqrt{p}$. We now require the following technical result.

Lemma 3.4. *For $c \geq 0$, $p = 1, 2, \ldots$, $R_p^2 \sim \chi_p^2(c^2 p)$, and $m = d\sqrt{p}$ with $d > 0$,*

(a) The sequence of random variables $(m \wedge aR_p)/\sqrt{p}$, $a \geq 0$, converges as $p \to \infty$, in probability to $d \wedge a\sqrt{c^2+1}$;

(b) The sequence of random variables $\rho_{p/2-1}(c\sqrt{p}R_p)$ converges, as $p \to \infty$, in probability to $c/\sqrt{c^2+1}$.

Proof. Part (a) follows since R_p^2/p converges, as $p \to \infty$, in probability to c^2+1, while part (b) was given by [11]. □

Applying Lemma 3.4 allows us to calculate $lim_{p\to\infty}\frac{\Delta_a^p(c\sqrt{p})}{d^2 p}$ in terms of c, d and a, and to infer dominance results in the corresponding asymptotic setting.

Theorem 3.4. *For $d > 0$, $0 \leq c \leq d$, and $a \in [0,1)$, we have*

$$\lim_{p\to\infty} \frac{\Delta_a^p(c\sqrt{p})}{d^2 p} = \begin{cases} (1-a)\,\frac{c^2+1}{d^2}\,(a - \frac{c^2-1}{c^2+1}) & \text{if } c < \sqrt{(d^2-1)_+} \\ \left(1 - a\frac{\sqrt{c^2+1}}{d}\right)_+ (1 + a\frac{\sqrt{c^2+1}}{d} - \frac{2c^2}{d\sqrt{c^2+1}}) & \text{if } c \geq \sqrt{(d^2-1)_+}; \end{cases}$$

where $y_+ = y \vee 0$, $y \in \Re$.

Proof. Applying (3.2), we have for $\theta \in S_\lambda$

$$\lim_{p\to\infty} \frac{\Delta_a^p(c\sqrt{p})}{d^2p} = \lim_{p\to\infty} E_\theta \left[\left(\frac{m \wedge R}{d\sqrt{p}} - \frac{m \wedge aR}{d\sqrt{p}} \right) \right.$$
$$\left. \times \left(\frac{m \wedge R}{d\sqrt{p}} + \frac{m \wedge aR}{d\sqrt{p}} - 2\frac{c}{d}\rho_{p/2-1}(c\sqrt{p}R) \right) \right] .$$

Now, from Lemma 3.4, we obtain

$$\lim_{p\to\infty} \frac{\Delta_a^p(c\sqrt{p})}{d^2p} = \left[\left((1 \wedge \frac{\sqrt{c^2+1}}{d}) - (1 \wedge \frac{a\sqrt{c^2+1}}{d}) \right) \right.$$
$$\left. \times \left((1 \wedge \frac{\sqrt{c^2+1}}{d}) + (1 \wedge \frac{a\sqrt{c^2+1}}{d}) - 2\frac{c}{d}\frac{c}{\sqrt{c^2+1}} \right) \right] ,$$

which leads to the results. □

Corollary 3.1. *For $a \in [0,1)$, we have $\lim_{p\to\infty} \frac{\Delta_a^p(c\sqrt{p})}{p} \geq 0$, for all $c \in [0,d]$, if and only if $a \in \left[\frac{2d^2}{d^2+1} - \frac{d}{\sqrt{d^2+1}}, 1\right)$. Furthermore, the inequality is strict for all $c \in [0,d]$ if and only if $a \in \left(\frac{2d^2}{d^2+1} - \frac{d}{\sqrt{d^2+1}}, \frac{d}{\sqrt{d^2+1}}\right)$, and this includes δ_{tlx}. Finally, for $d < 1$ and $a \in [d,1)$, the estimators δ_a and δ_{mle} are asymptotically equivalent with $\lim_{p\to\infty} \frac{\Delta_a^p(c\sqrt{p})}{p} = 0$ for all $c \in [0,d]$.*

Proof. The results follow by working directly with Theorem 3.4's expression for $\lim_{p\to\infty} \frac{\Delta_a^p(c\sqrt{p})}{p}$. The inferences specific to δ_{tlx} are a direct consequence of the more general result since $\frac{d^2}{d^2+1} \in \left(\frac{2d^2}{d^2+1} - \frac{d}{\sqrt{d^2+1}}, \frac{d}{\sqrt{d^2+1}}\right)$ for all $d > 0$. □

Corollary 3.1 exhibits, through Theorem 3.4's expression for the exact asymptotics risk difference, the asymptotic dominance, as well as strict asymptotic dominance, relative to a sequence $(m = d\sqrt{p}, p)$ with $p \to \infty$, of a subclass of truncated linear estimators δ_a over δ_{mle}. This subclass of improvements includes the truncated linear minimax estimator δ_{tlx} for all d. Coupled with the fixed p analytical and numerical results in Section 3.1 and below as well, this finding clearly indicates that δ_{tlx} and other asymptotically dominating δ_a's will necessarily outperform δ_{mle} for large enough p. Observe as well that Corollary 3.1's lower bound $\frac{2d^2}{d^2+1} - \frac{d}{\sqrt{d^2+1}}$ is less or equal to 0 for $d \leq \sqrt{1/3}$, so that the dominance condition is universal applying to all δ_a's with $a \in [0,d)$, in accordance with Theorem 1.1's universal dominance cutoff point.

We now turn to the relative degree of improvement attained by such δ_a's as measured by the quantity $h(a,d,c)$ given in (3.6).

Theorem 3.5. *For $d > 0$, $0 \le c \le d$, and $a \in [0,1)$, we have*

$$h(a,d,c) = \begin{cases} (1-a)\,(c^2+1)\,(a - \frac{c^2-1}{c^2+1}) & \text{if } c < \sqrt{(d^2-1)_+}\,; \\ 1 - (\,\frac{c^2(a-1)^2 + a^2}{c^2+d^2-\frac{2c^2 d}{\sqrt{c^2+1}}}\,) & \text{if } c \ge \sqrt{(d^2-1)_+} \text{ and } c < \sqrt{(\frac{d^2}{a^2}-1)_+}\,; \\ 0 & \text{if } c \ge \sqrt{(\frac{d^2}{a^2}-1)_+}\,. \end{cases}$$

Furthermore, $h(a,d,\cdot)$ decreases on $[0,d]$ whenever $d \ge d_0 = (\frac{\sqrt{5}-1}{2})^{1/2} \approx 0.7862$, or $d < d_0$ and $a \le d/(d+1)$.

Proof. The monotonicity property is established in the Appendix. Since $R(\theta,\delta_{mle}) = R(\theta,X) + \{R(\theta,\delta_{mle}) - R(\theta,X)\}$, Lemma 3.1 permits us to write

$$R(\theta,\delta_{mle}) = p + E_\theta\left[(g_{mle}(R) - R)\,(g_{mle}(R) + R - 2g_{\pi_\lambda}(R))\right]\,, \qquad (3.7)$$

for all $\theta \in S_\lambda$. Making use of Lemma 3.4, as in the proof of Theorem 3.4, we obtain from (3.7)

$$\lim_{p\to\infty} \frac{R(c\sqrt{p},\delta_{\text{mle}})}{p} = \begin{cases} 1 & \text{if } \sqrt{c^2+1} < d \\ 1 + (d - \sqrt{c^2+1})(d + \sqrt{c^2+1} - \frac{2c^2}{\sqrt{c^2+1}}) & \text{if } \sqrt{c^2+1} \ge d. \end{cases}$$

Finally, the result follows by combining Theorem 3.4 with the above, and rearranging terms. □

For the particular case of the truncated linear minimax estimator δ_{tlx}, we obtain the following.

Corollary 3.2. *For $d > 0$ and $0 \le c \le d$, let $\kappa(d,c) = h(\frac{d^2}{d^2+1}, d, c)$ represent the asymptotic relative risk difference between the estimators δ_{tlx} and δ_{mle}. Then, we have*

$$\kappa(d,c) = \begin{cases} \frac{2d^2-c^2+1}{(d^2+1)^2} & \text{if } c < \sqrt{(d^2-1)_+}\,; \\ 1 - \frac{1}{(d^2+1)^2}\,\frac{(c^2+d^4)\sqrt{c^2+1}}{(c^2+d^2)\sqrt{c^2+1}-2c^2 d} & \text{if } c \ge \sqrt{(d^2-1)_+}\,. \end{cases}$$

Furthermore, $\kappa(d,c)$ is, for fixed $d > 0$, decreasing in $c \in [0,d]$ with minimum value $k(d,0) = \frac{d^2+1+(d^2\wedge d^4)}{(d^2+1)^2}$, and maximum value $k(d,d) = \frac{1}{2}(1 - \frac{d}{\sqrt{d^2+1}})$.

Proof. The given expression for $\kappa(d,c)$ and its monotonicity are obtained by substituting $a = d^2/(d^2+1)$ in Theorem 3.5. □

Example 3. *The expressions above for the asymptotic relative degree of improvement $h(a,d,c)$ or $\kappa(d,c)$ are remarkably simple. As an illustration,*

take $d = 1$, *so that* $m^2 = p$ *and* $\delta_{tlx}(x) = (\frac{1}{2}\|x\| \wedge \sqrt{p})\frac{x}{\|x\|}$. *Corollary 3.1 indicates that* δ_{tlx} *dominates* δ_{mle} *asymptotically, and Corollary 3.2 tells us that the relative degree of improvement is asymptotically equal to* $\kappa(1,c) = 1 - \frac{1}{4}\frac{(c^2+1)^{3/2}}{(c^2+1)^{3/2}-2c^2}$, *decreases in* c *on* $(0,1)$, *with a maximum of* $\kappa(1,0) = 3/4$, *and a minimum of* $\kappa(1,1) = \frac{\frac{3}{4}\sqrt{2}-1}{\sqrt{2}-1} \approx 0.146$. *In other words, the truncated linear minimax estimator improves on the maximum likelihood estimator asymptotically, as* $m^2 = p$ *and* $p \to \infty$, *by as much as* 75%, *and by at least* 14.6%. *Observe that such gains are attenuated for larger parameter spaces, and augmented for smaller parameter spaces, as seen by the decreasing monotonicty properties of* $k(d,0)$ *and* $k(d,d)$ *as functions of* d. *In comparison to the case* $d = 1$, *we have for instance with* $d = 2$: $k(2,0) = 9/25$ *and* $k(2,2) = \frac{1}{2} - \frac{\sqrt{5}}{5} \approx 0.0528$. *Finally, the curves* $k(d,c)$, $c \in [0,d]$, *are presented in Figures 5 and 6 for* $d = 1, 2$.

Remark 1. *Assessing the comparative performance of the estimators* δ_a, *it is easy to infer from Theorem 3.5 that, for fixed* c, d, $h(\cdot, d, c)$ *increases on* $[0, \frac{c^2}{c^2+1}]$ *and decreases on* $[\frac{c^2}{c^2+1}, 1]$ *with a unique maximum at* $a^*(c) = \frac{c^2}{c^2+1}$. *From this, we infer that, asymptotically, the subclass* $\{\delta_a : 0 \le a \le \frac{d^2}{d^2+1}\}$ *is minimal complete among the class* $\{\delta_a : 0 \le a \le 1\}$, *and furthermore that* δ_{tlx} *dominates* δ_a *(asymptotically) for* $a \in (\frac{d^2}{d^2+1}, 1]$. *These features are illustrated in both Figure 3 and 4 where, for* $d = 2$ *and* $d = 0.5$ *respectively, the degrees of improvement in risk on* δ_{mle} *are compared. For instance in Figure 4 where* $d = 0.5$, *we see that* $\delta_{tlx} = \delta_{0.2}$ *dominates both* $\delta_{0.30}$ *and* $\delta_{0.45}$, *but not* $\delta_{0.10}$, *the latter being a member of the admissible subclass with* $a \in [0, 0.2]$. *Notice as well, for* $d = 0.5$, *that the strict asymptotic dominance is achieved by all* δ_a *with* $a < 0.5$ *by virtue of Corollary 3.1.*

Remark 2. *As a follow up to the previous remark, we expand on a further "asymptotic minimax type" property of* δ_{tlx}. *To do so, consider our estimation problem with weighted squared error loss* $L^*(\theta, \delta) = \frac{(\delta-\theta)^2}{q(\theta)}$, *with* $q(\theta) = R(\theta, \delta_{mle})$ *and associated risk* $R^*(\theta, \delta) = E_\theta(L^*(\theta, \delta(X))$. *By virtue of Theorem 3.5, we have for the sequence* $\{(m, \lambda, p) = (d\sqrt{p},\, c\sqrt{p},\, p); p \ge 1\}$, $\lim_{p\to\infty} R^*(c\sqrt{p}, \delta_a) = 1 - h(a, d, c)$ *and the worst case scenario*

$$\sup_{c\in[0,d]} \{\lim_{p\to\infty} R^*(c\sqrt{p}, \delta_a)\} = 1 - h(a, d, d),$$

for $a \in [0, d^2/(d^2+1)]$, *i.e. for the complete class members given in Remark 1. Finally, as seen in Remark 1, the choice* $a = d^2/(d^2+1)$ *minimizes* $1 - h(a, d, d)$ *in* a, *so that we obtain the optimality property*

$$\inf_{a\in[0,1)} \{\sup_{c\in[0,d]} \{\lim_{p\to\infty} R^*(c\sqrt{p}, \delta_a)\}\} = \sup_{c\in[0,d]} \{\lim_{p\to\infty} R^*(c\sqrt{p}, \delta_{tlx})\}$$

for the truncated linear minimax estimator δ_{tlx}.

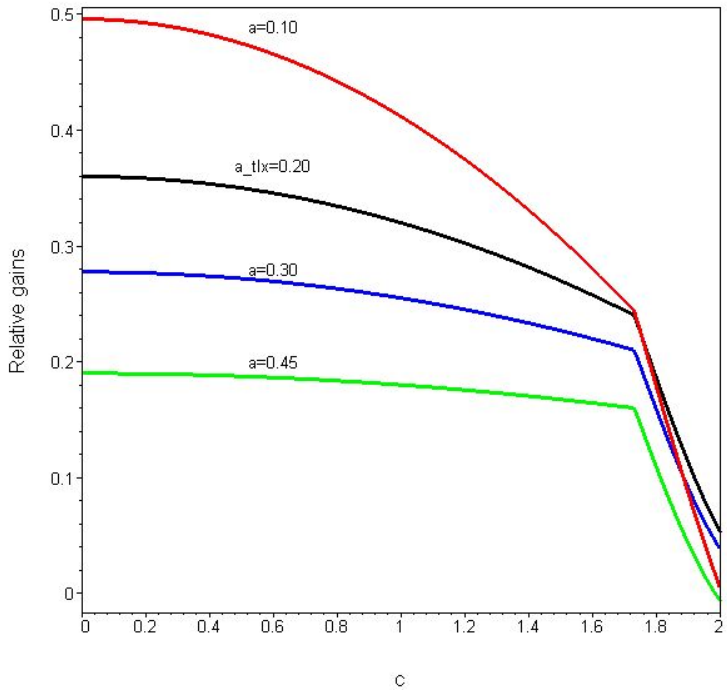

Figure 3: Graphs of the ratio of related improvement in risks for $d = 2$ with $a = 0.71, a_{\text{tlx}} = 0.8, a = 0.85$ and $a = 0.90$.

Remark 3. *In practice for fixed (m, p) and a given estimator δ_a, we are concerned with the behaviour of the relative risk difference $w_a^p(m, \lambda)$, $\lambda \in [0, m]$. To this end, numerical evidence suggests that $w_a^p(d\sqrt{p}, c\sqrt{p})$ is either fairly or very well approximated by $h(a, d, c)$ for large p, and even for moderate p. From a theoretical point of view, we point out that the exhibited quality of the approximation reflects the speed of Lemma 3.4's convergence, or more precisely through the expectations arising in the risk functions of δ_{tlx} and δ_{mle}. Illustrations are provided in Figures 5 and 6, for $\delta_a = \delta_{tlx}$, $p = 9, 16$ and $d = 1, 2$. The approximation $\kappa(d, c)$ produces slight overestimates which are attenuated for larger p. For instance with $d = 1, c = 0.6$, we have relative gains of 0.5028 for $(m, p, \lambda) = (3, 9, 1.8)$, of 0.5223 for $(m, p, \lambda) = (4, 16, 2.4)$, of 0.5367 for $(m, p, \lambda) = (6, 36, 3.6)$, in comparison with $\kappa(1, 0.6) = 0.5422$. Please note however that δ_{tlx} does not always dominate δ_{mle} so that the asymptotics, if misinterpreted, will necessarily produce a contrary dominance conclusion for some pairs (m, p). Such an occurrence arises for $(m, p) = (6, 9)$ as illustrated in Figure 6, and as previously seen in Section 3.1 where we obtained that δ_a dominates δ_{mle} iff $a \in C(6, 9) \approx [0.833, 1)$ with $\delta_{tlx} = \delta_{0.8}$.*

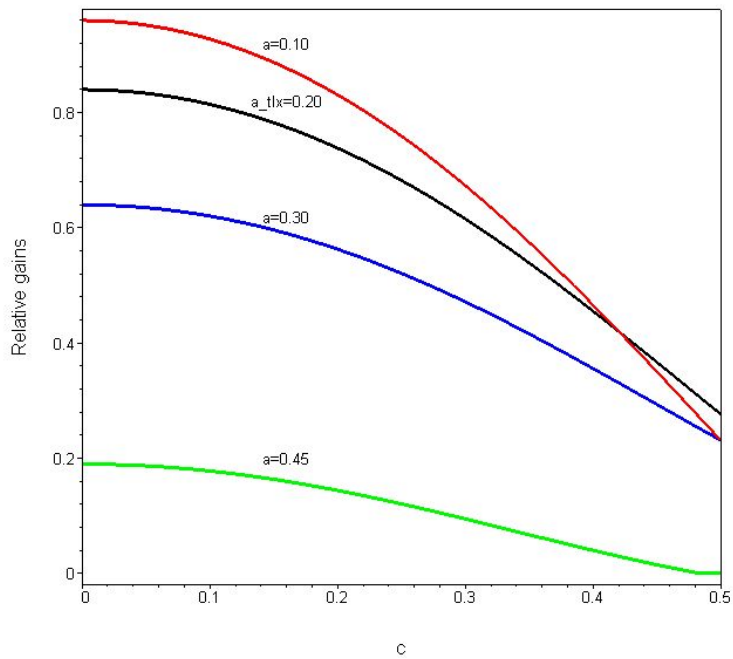

Figure 4: Graphs of the ratio of related improvement in risks for $d = 0.5$ with $a = 0.10, a_{\text{tlx}} = 0.20, a = 0.30$ and $a = 0.45$.

4 Concluding remarks

Section 2 and section 3.1 provides finite dimension necessary and sufficient analysis (including the univariate case) in comparing truncated linear estimators δ_a with δ_{mle}. The asymptotic setting and techniques, which we have made use of, were previously introduced and exploited by Marchand and Perron [11], as well as Fourdrinier and Marchand [5]. The former studied conditions for which the boundary uniform Bayes estimator δ_{BU} is minimax, while the latter give conditions for which δ_{BU}, and other Bayes estimators with respect to uniform priors on spheres, dominate δ_{mle}. Here, our findings go further with the class of truncated linear estimators as we obtained, not only dominance results in the asymptotic setting, but also precise expressions for the relative gains in asymptotic risk. This contrasts with a seemingly paucity of explicit analytical findings quantifying risk differences. It would be most interesting to proceed with a similar asymptotic setting to assess the performance of the fully uniform Bayes estimator δ_U, or other Bayes estimators. Partial results were given in [6], namely the asymptotic dominance in our setting for $d < 1$. We conjecture that further and stronger dominance results hold, and we feel that similar quantification of the relative gains in risk are possible, and that the asymptotic setting holds much promise.

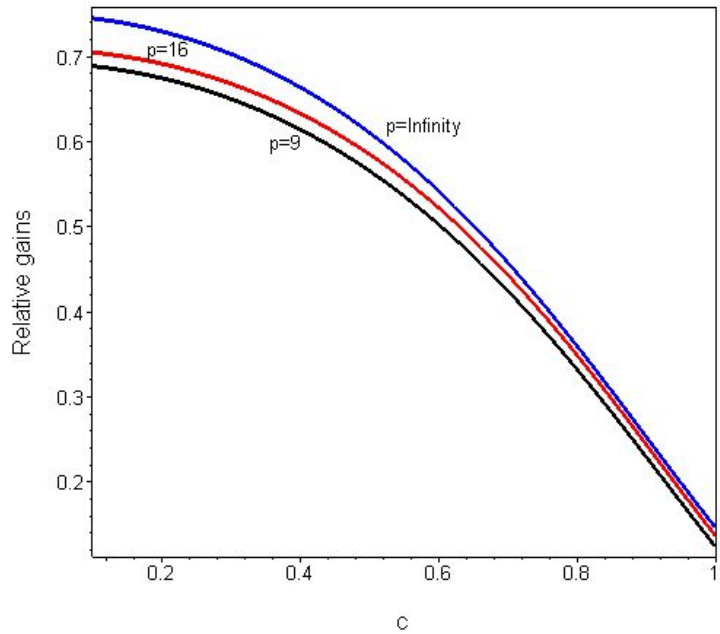

Figure 5: The relative gains in risk $w_a^p(d\sqrt{p}, c\sqrt{p})$ for $d = 1$, $\delta_a = \delta_{\text{tlx}}$, $p = 9, 16, \infty$.

Finally, departing from the asymptotic framework and the subclass of truncated linear estimators, we point out that several challenges remain for the general problem considered in the introduction. Indeed, although dominating estimators can be provided for any pair (m, p), the problem of the specification of priors π that lead to dominating Bayesian estimators δ_π remains unsolved in general with solutions provided for m relatively small enough (see [10]). Moreover, to our knowledge, there exists little or no findings with regards to extensions with unknown σ^2.

Acknowledgments

The research work of Éric Marchand is partially supported by NSERC of Canada. Othmane Kortbi is a Ph.D. student at the Université de Sherbrooke. During his studies, 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).

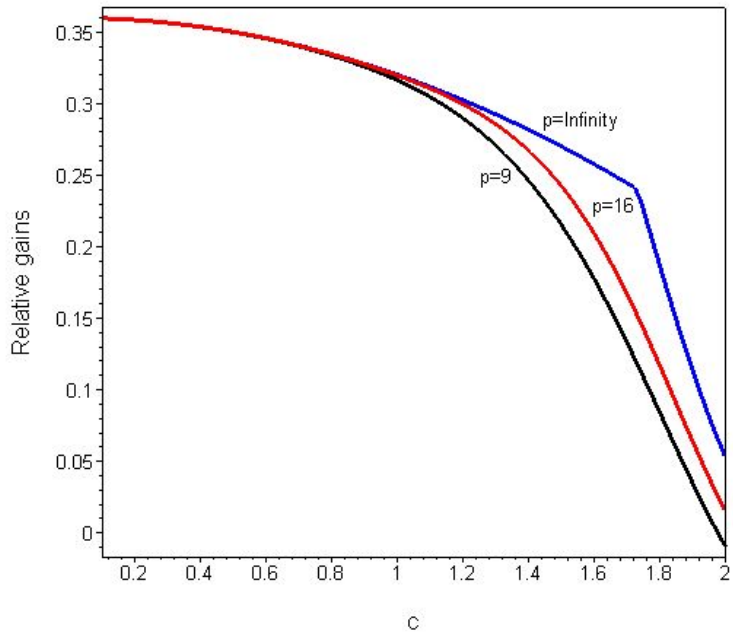

Figure 6: The relative gains in risk $w_a^p(d\sqrt{p}, c\sqrt{p})$ for $d = 2$, $\delta_a = \delta_{\text{tlx}}$, $p = 9, 16, \infty$.

5 Appendix

The function $\rho_{(p/2)-1}$ plays a key role and the next lemma recalls some useful properties, given in [1], and [16] for p $>$ 1 and readily verified otherwise for $p = 1$ by using the representation $\rho_{-1/2} \equiv tanh$.

Lemma 5.1. *For all $p \geq 1$, $\rho_{(p/2)-1}(t)$ is increasing in t, with $\rho_{(p/2)-1}(0) = 0$ and $\rho_{(p/2)-1}(t) \to 1$ as $t \to \infty$. Also, $\rho_{(p/2)-1}(t)/t$ is decreasing in t, with $\rho_{(p/2)-1}(t)/t \to 1/p$ as $t \to 0$. We have as well $\rho_{p/2}(t) = \frac{1}{\rho_{(p/2)-1}(t)} - \frac{p}{t}$, and in particular $\rho_{1/2}(t) = \coth(t) - 1/t$.*

Proof of Lemma 3.2

With $\delta_a(X) = sgn(X)(a\|X\| \wedge m)$ for $p = 1$, the risk function can be directly expressed as

$$\begin{aligned} R(m,\delta_a) &= \int_{-\frac{m}{a}}^{\frac{m}{a}} |ax-m|^2 \phi(x-m)\, dx + \int_{-\infty}^{-\frac{m}{a}} 4m^2\phi(x-m)\, dx \\ &= a^2 \int_{-\frac{m}{a}}^{\frac{m}{a}} x^2\phi(x-m)\, dx + m^2\left(\Phi\left(\frac{m}{a}-m\right) - \Phi\left(-\frac{m}{a}-m\right)\right) \\ &\quad -2am\int_{-\frac{m}{a}}^{\frac{m}{a}} x\phi(x-m)\, dx + 4m^2\Phi\left(-\frac{m}{a}-m\right) \\ &= a^2\int_{-\frac{m}{a}}^{\frac{m}{a}} x^2\phi(x-m)\, dx - 2am\int_{-\frac{m}{a}}^{\frac{m}{a}} x\phi(x-m)\, dx + m^2(4-A-3B). \end{aligned}$$

Lemma 3.2's expression for the risk $R(m,\delta_a)$ follows with the evaluations

$$\int_{-\frac{m}{a}}^{\frac{m}{a}} x^2\phi(x-m)\, dx = \left(1+m^2\right)(B-A) + D\left(m-\frac{m}{a}\right) - C\left(m+\frac{m}{a}\right),$$

$$\int_{-\frac{m}{a}}^{\frac{m}{a}} x\phi(x-m)\, dx = m(B-A) + D - C,$$

and by collecting terms. Finally, the given expression for the derivative arises from straightforward computations and by using the relationships $A' = \frac{m}{a^2}C$, $B' = \frac{-m}{a^2}D$, $C' = \frac{-m^2}{a^2}\left(1-\frac{1}{a}\right)C$ and $D' = \frac{m^2}{a^2}\left(1+\frac{1}{a}\right)D$.

Proof of Lemma 3.3

At first, we can deduce from Lemma 3.2 that $\frac{d}{da}R(m,\delta_a) \le 0$ if $(2a+2m^2(a-1)) \le 0$ which is equivalent to $a \le \frac{m^2}{m^2+1}$. Secondly, after assuming that $a \ge \frac{m^2}{m^2+1}$, the same theorem lets us to deduce that

$$\begin{aligned} \frac{d}{da}R(m,\delta_a) \ge \left(2a+2m^2(a-1)\right)(\Phi(2m)-1/2) \\ + 2m(a-2)\phi(2m) - 2ma\phi(0), \end{aligned}$$

since $B - A = \Phi(m+m/a) - \Phi(m-m/a)$ is a decreasing function in a, and both $D = \phi(m+m/a)$ and $C = \phi(m-m/a)$ are increasing in a. So, $\frac{d}{da}R(m,\delta_a) \ge 0$ if $(2a+2m^2(a-1))(\Phi(2m)-1/2) + 2m(a-2)\phi(2m) - 2am\phi(0) \ge 0$ which is equivalent to $a \ge b(m)$, since $(1+m^2)(\Phi(2m)-1/2) + m\phi(2m) - m\phi(0) \ge -H_1(m) \ge 0$ for all $m \le m_1$. This establishes the desired result. □

Proof of Theorem 3.5's monotonicity property

Note that, when $d > 1$ and c lies in $\left(0, \sqrt{d^2-1}\right)$, the decreasing property is obvious. Otherwise, we consider separately the cases: **(i)** $a \leq \frac{d}{d+1}$ and **(ii)** $a > \frac{d}{d+1}$.

(i) We may write

$$h(a,d,c) = 1 - \left[\frac{\frac{(a-1)^2c^2+a^2}{c^2+d^2}}{1 - \frac{2c^2d}{(c^2+d^2)\sqrt{c^2+1}}}\right],$$

and the monotonicity follows since $\frac{(a-1)^2c^2+a^2}{c^2+d^2}$ is increasing for $c > 0$, given that $a \leq \frac{d}{d+1}$, and $\frac{2c^2d}{(c^2+d^2)\sqrt{c^2+1}}$ is increasing for $c \in \left(0, \sqrt{\frac{d^2}{2} + \frac{d}{2}\sqrt{d^2+8}}\right)$, with $\sqrt{\frac{d^2}{2} + \frac{d}{2}\sqrt{d^2+8}} > d$.

(ii) For $a > \frac{d}{d+1}$, we write

$$h(a,d,c) = 1 - \frac{(a-1)^2c^2+a^2}{c^2+d^2-\frac{2c^2d}{\sqrt{c^2+1}}},$$

and observe that the monotone decreasing property follows if $c^2+d^2-\frac{2c^2d}{\sqrt{c^2+1}}$ is decreasing for $c \in (0,d)$. This in turn is equivalent to

$$\begin{aligned}
\left(c^2+1\right)^{3/2} - d\left(c^2+2\right) \leq 0 \quad &\Leftrightarrow \quad d \geq \frac{\left(c^2+1\right)^{3/2}}{\left(c^2+2\right)}, \text{ for all } c \in (0,d), \\
&\Leftrightarrow \quad d \geq \sup_{c\in(0,d)} \frac{\left(c^2+1\right)^{3/2}}{\left(c^2+2\right)} = \frac{\left(d^2+1\right)^{3/2}}{\left(d^2+2\right)} \\
&\Leftrightarrow \quad d^4+d^2-1 \geq 0,
\end{aligned}$$

which yields the result. □

References

[1] D. E. Amos. Computation of modified bessel functions and their ratios. *Mathematics of computation*, 28(125):239–251, 1974.

[2] L. Brown, I. Johnstone, and B. MacGibbon. Variation diminishing transformations: a direct approach to total positivity and its statistical applications. *Journal of the American Statistical Association*, 376:824–832, 1981.

[3] A. DasGupta and S.N. Lahiri. Density in high and ultra high dimensions, regularization, and the L_1 asymptotics (to appear). *A Festschrift for William E. Strawderman. IMS Collections*, 2011.

[4] Y. Dou and C. van Eeden. Comparisons of the performances of estimators of a bounded normal mean under squared-error loss. *REVSTAT*, 7:203–226, 2009.

[5] D. Fourdrinier and É. Marchand. 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, 2010.

[6] N.R. Gueye. Problèmes d'estimation de paramètres avec restriction sur l'espace des paramètres. *Ph.D. Thesis, Université de Montréal*, 2003.

[7] S. Karlin. Pólya type distributions, II. *Annals of Mathematical Statistics*, 28:281–309, 1957.

[8] É. Marchand. Estimation of a multivariate mean with constraints on the norm. *The Canadian Journal of Statistics*, 21:359–366, 1993.

[9] É.. Marchand, I. Ouassou, A. T. Payandeh, and F. Perron. On the estimation of a restricted location parameter for symmetric distributions. *Journal of the Japanese Statistical Society*, 38:293–309, 2008.

[10] É. Marchand and F. Perron. Improving on the mle of a bounded normal mean. *Annals of Statistics*, 29:1066–1081, 2001.

[11] É. Marchand and F. Perron. On the minimax estimator of a bounded normal mean. *Statistics & Probability Letters*, 58:327–333, 2002.

[12] É. Marchand and F. Perron. Improving on the mle of a bounded location parameter for spherical distributions. *Journal of Multivariate Analysis*, 92:227–238, 2005.

[13] É. Marchand and W. E. Strawderman. Estimation in restricted parameter spaces: A review. *A Festschrift for Herman Rubin, IMS Lecture Note Series*, 45:21–44, 2004.

[14] C. Stein. Estimation of the mean of multivariate normal distribution. *Annals of Statistics*, 9:1135–1151, 1981.

[15] C. van Eeden. *Restricted Parameter Space Estimation Problems*, volume 188. Lecture Notes in Statistics, Springer, 2006.

[16] G. S. Watson. *Statistics on Spheres.* Wiley, New York, 1983.