On estimating a bounded normal mean with applications to predictive density estimation [1]

ÉRIC MARCHAND[a], FRANÇOIS PERRON[b], IRAJ YADEGARI[a]

a Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1
(e-mail: eric.marchand@usherbrooke.ca; iraj.yadegari@usherbrooke.ca)

b Université de Montréal, Département de mathématiques et de statistique, Montréal Qc,
CANADA, H3C 3J7   (e-mail: perronf@dms.umontreal.ca)

Abstract. For a normally distributed $X \sim N(\mu, \sigma^2)$ and for estimating μ when restricted to an interval $[-m, m]$ under general loss $F(|d - \mu|)$ with strictly increasing and absolutely continuous F, we establish the inadmissibility of the restricted maximum likelihood estimator δ_{mle} for a large class of F's and provide explicit improvements. In particular, we give conditions on F and m for which the Bayes estimator δ_{BU} with respect to the boundary uniform prior $\pi(-m) = \pi(m) = 1/2$ dominates δ_{mle}. Specific examples include L^s loss with $s > 1$, as well as reflected normal loss. Connections and implications for predictive density estimation are outlined, and numerical evaluations illustrate the results.

1. Introduction

For a normally distributed observable $X \sim N(\mu, \sigma^2)$ we consider the restricted parameter space point estimation problem with $\mu \in \Theta(m) = \{\mu \in \mathbb{R} : |\mu| \leq m\sigma\}$. Both m and σ^2 are known. We provide analysis for symmetric and strict bowled-shaped loss functions

$$L(\mu, d) = F(|d - \mu|), d, \mu \in \Theta(m), \tag{1.1}$$

with F strictly increasing and absolutely continuous on $[0, 2m\sigma]$, but otherwise general. The simple model addressed here still remains relevant and influential. It is relevant to various situations, namely via limit theorems, where all the uncertainty lies in the mean signal with the noise controlled and X is a normally distributed summary statistic (that may be sufficient as well). The other aspect of the model is the boundedness to an interval assumption on the mean, centered here about 0 without loss of generality. It is indeed often the case that bounds on the parameters can be stipulated and arise from a given practical context.

A central objective is to study the efficiency of estimators that take values on or close to the boundary of the parameter space, such as the benchmark maximum likelihood estimator

$$\delta_{\mathrm{mle}}(X) = (m\sigma \wedge |X|) \operatorname{sgn}(X), \tag{1.2}$$

and to provide dominating estimators $\delta(X)$, including Bayes estimators, whenever δ_{mle} is inadmissible, in terms of frequentist risk

$$R(\mu, \delta) = \mathbb{E}_\mu F(|\delta(X) - \mu|); \mu \in \Theta(m). \tag{1.3}$$

---

[1]Draft of November 21, 2016

With much previous work focussed on squared-error loss (i.e., $F(t) = t^2$), our findings encompass many other alternative convex losses, but alternatively loss function features that may well be more attractive for the decision-maker such as boundedness and non-convexity. These include for instance reflected normal loss (e.g., Spiring, 1993) given by

$$L_\gamma(\mu, d) = 1 - e^{-\frac{|d-\mu|^2}{2\gamma}}, \tag{1.4}$$

γ being a positive constant. It is useful to observe that $\lim_{\gamma \to \infty} 2\gamma L_\gamma(d, \mu) = |d - \mu|^2$, so that we can reasonably anticipate that results relative to L_γ with γ large mimic those for squared-error loss. Furthermore, as described in Section 2, additional sources of motivation and connections with reflected normal loss L_γ arise from a predictive density estimation perspective.

Remark 1.1. *Other classes of loss functions for which our findings apply, and which also arise in a predictive density estimation setting (see Lemma 2.3), are cases where F is generated by a cdf on $\mathbb{R}_+$, such as a half-normal $F(t) = \Phi(t) - 1/2$, Φ being the $N(0,1)$ cdf. Our results also apply to a vast collection of losses F which are reflected densities of the form $F(t) = 1 - \frac{\psi(t)}{\psi(0)}, t > 0$, with ψ a decreasing density on $\mathbb{R}_+$, which include reflected normal loss, as well as losses $F(t) = 1 - e^{-t^\alpha/2\gamma}$, $\alpha, \gamma > 0$, which are reflections of exponential power densities and which we will study. Alternatively, these latter losses are generated by Weibull distribution functions which relate to the other losses of this example.*

Using various conditional risk decompositions, including a technique introduced by Moors (1985), Marchand and Perron (2001) obtained estimators that dominate δ_{mle} under squared-error loss, including Bayesian dominators for small enough m. Namely, one of their results, which duplicates an earlier finding by Casella and Strawderman (1981), establishes that $\delta_{BU}(X)$ dominates $\delta_{\mathrm{mle}}(X)$ for $m \leq \sigma$, where $\delta_{BU}(X) = m\sigma \tanh(\frac{mX}{\sigma})$ is the Bayes estimator with respect to the two-point uniform prior on the boundary $\{-m, m\}$ of $\Theta(m)$. Otherwise for larger m, the projection $\delta_p(X) = (|\delta_{BU}(X)| \wedge |\delta_{\mathrm{mle}}(X)|) \, sgn(X)$ was shown by Moors (1985) to dominate $\delta_{\mathrm{mle}}(X)$ under squared-error loss. For other losses, the situation may differ. Indeed, for absolute value loss (i.e., $F(t) = t$), Iwasa and Moritani (1997), as well as Kucerovsky et al. (2009), show that δ_{mle} is a proper Bayes estimator and is thus admissible. On the other hand, Iwasa and Moritani (1997) establish the inadmissibility of δ_{mle} for a large class of convex losses, including L^s losses ($F(t) = t^s$) with $s > 1$.

As for squared-error loss and as shown below using a conditional risk decomposition, we establish that δ_{mle}, as well as others estimator taking values on the boundary of the parameter space, are inadmissible under reflected normal loss L_γ and for a large class of losses. For such situations, an improvement is always provided by the above-defined projection $\delta_p(X)$, with $\delta_{BU}(X)$ the corresponding boundary uniform Bayes estimator. Moreover, despite not having in general an explicit expression for $\delta_{BU}(X)$, we obtain a simple sufficient condition (Theorem 4.4), applicable to a large class of losses F, for $\delta_{BU}(X)$ to dominate $\delta_{\mathrm{mle}}(X)$, namely $m \leq c_0\sigma$, with c_0 the unique solution in c of $c\frac{f(c)}{f'(c)} = 1$ and $f = F'$ a.e.. For L^s loss $F(t) = t^s$ with $s > 1$, this reduces to $m \leq \sigma\sqrt{s-1}$ (Example 5.1), extending the L^2 result mentioned above. For reflected normal loss, the above condition for $\delta_{BU}(X)$ to dominate $\delta_{\mathrm{mle}}(X)$ reduces to $m \leq \sigma\sqrt{\frac{\gamma}{1+\gamma}}$ (Example 5.2). In contrast to such losses; conditions of which are presented in Section 4 and which include strictly convex losses; we show in Section 4.1 that losses in (1.1) with concave F do not imply the deficiency of boundary taking estimators such as $\delta_{\mathrm{mle}}(X)$. In fact, in such cases, we show that the boundary uniform Bayes estimator is given by $\delta_{BU}(X) = m\sigma \, sgn(X)$.

The organization of the manuscript is as follows. In Section 2, we describe connections with predictive density estimation problems and how the analysis of a collection of such problems is contingent on a point estimation set-up as above with a dual loss, and as analysed in this work. Section 3 deals with preliminary results and definitions, namely those relative to the conditional risk where the conditioning is on $|X| = r$, as well as first properties of the best equivariant estimator $\delta_{g_\lambda^*}$ when $|\mu| = \lambda$. Section 4 begins with a further description of the optimization problem of determining g_λ^* (Theorem 4.1), while Subsection 4.1 deals with concave F, a non-shrinking property, and the lack of inefficiency of estimators that take values on the boundary of the parameter space. Subsections 4.2, 4.3 establish properties of $\delta_{g_\lambda^*}$ and namely a shrinkage property for a wide class of losses, including convex losses. Subsection 4.4 contains dominance findings, while Section 5 is dedicated to examples, illustrations, numerical evaluations, implications for predictive density estimation, and further observations.

2. Connections with predictive density estimation

Additional motivation for the point estimation problem considered here stems from further connections with predictive density estimation. There has been much recent Bayesian and decision theory analysis of predictive density estimators, in particular for multivariate normal or spherically symmetric settings, as witnessed by the work cited in this section, as well as Komaki (2001), George, Liang and Xu (2006), Brown, George and Xu (2008), Kato (2009), Maruyama and Strawderman (2012), Boisbunon and Maruyama (2014), Maruyama and Ohnishi (2016), among others.

We describe here several such problems where the risk evaluation of a subclass of predictive density estimators, including plug-in density estimators, is equivalent to the risk evaluation of a point estimator under a dual loss. Consider spherically symmetric and independently distributed

$$Y_1|\mu \sim p(\|y_1 - \mu\|^2)\,,\;\; Y_2|\mu \sim q(\|y_2 - \mu\|^2)\,,\; y_1, y_2, \mu \in \mathbb{R}^d\,; \tag{2.1}$$

with p and q known Lebesgue densities. [2] For predictive analysis purposes, one wishes to obtain a predictive density $\hat{q}(y_2; y_1)$, based on observed y_1, as an estimate of $q(\|y_2 - \mu\|^2)$, $y_2 \in \mathbb{R}^d$. Several loss functions are at our disposal to measure efficiency and these include the class of $\alpha-$divergence loss functions (e.g., Csiszár, 1967) given by

$$L_{h_\alpha}(\mu, \hat{q}(\cdot, y_1)) = \int_{\mathbb{R}^d} h_\alpha\left(\frac{\hat{q}(y_2; y_1)}{q(\|y_2 - \mu\|^2)}\right) q(\|y_2 - \mu\|^2)\, dy_2\,, \tag{2.2}$$

with

$$h_\alpha(z) = \begin{cases} \frac{4}{1-\alpha^2}(1 - z^{(1+\alpha)/2}) & \text{for } |\alpha| \leq 1 \\ z\log(z) & \text{for } \alpha = 1 \\ -\log(z) & \text{for } \alpha = -1. \end{cases}$$

Notable examples in this class include Kullback-Leibler (h_{-1}), reverse Kullback-Leibler (h_1), and Hellinger ($h_0/4$). Integrated absolute error and squared error losses, referred hereafter as L^1 and

---

[2]The predictive density estimation-point estimation relationship which we exploit in this paper relates to the univariate $d = 1$ case, but we present here the general $d-$variate case for sake of interest.

L^2 losses, provide other choices. These are L^s losses, $s = 1, 2$, given by

$$L(\mu, \hat{q}(\cdot, y_1)) = \int_{\mathbb{R}^d} |\hat{q}(y_2; y_1) - q(\|y_2 - \mu\|^2)|^s \, dy_2. \tag{2.3}$$

For all of the above losses, the performance of predictive densities $\hat{q}(\cdot; Y_1)$ may be measured by the frequentist risk

$$R(\mu, \hat{q}) = \int_{\mathbb{R}^d} L(\mu, \hat{q}(\cdot; y_1)) \, p(\|y_1 - \mu\|^2) \, dy_1. \tag{2.4}$$

The dual relationships which we describe apply to predictive densities of the form

$$\frac{1}{c^d} q\left(\frac{\|y_2 - \hat{\mu}(Y_1)\|^2}{c}\right), y_2 \in \mathbb{R}^d, \tag{2.5}$$

where $\hat{\mu}(Y_1)$ is an estimator of μ. Cases $c = 1$ correspond to plug-in predictive density estimators, while cases $c > 1$ correspond to scale expanded variants. As discussed in Fourdrinier et al. (2011) for Kullback-Leibler loss, and Kubokawa, Marchand & Strawderman (2015, 2016) for L^2 and L^1 losses, such scale expansions are interesting to consider and can provide significant risk improvement on plug-in procedures. As an illustration, consider the following multivariate normal version of (2.1):

$$Y_1|\mu \sim N_d(\mu, \sigma_{Y_1}^2 I_d), \ Y_2|\mu \sim N_d(\mu, \sigma_{Y_2}^2 I_d), \tag{2.6}$$

for which the minimum risk equivariant MRE (under changes of location) predictive density estimator, or equivalently the generalized Bayes predictive density estimator with respect to the prior $\pi(\mu) = 1$, is given by a $N_d(Y_1, ((\frac{1-\alpha}{2})\sigma_{Y_1}^2 + \sigma_{Y_2}^2)I_d)$ density for α-divergence loss L_{h_α} (e.g., Ghosh, Mergel and Datta, 2008). With the exception of reverse Kullback-Leibler loss, all the above MRE predictive density estimators are indeed scale expansion variants and dominate the plug-in $N_d(Y_1, \sigma_{Y_2}^2)$ density under the corresponding loss.

Here is an adapted version of Kubokawa, Marchand and Strawderman (2015, Theorem 3.1, part a).

Lemma 2.1. *(Duality between integrated L^2 and reflected normal losses)*
For normal model (2.6), the frequentist risk of a predictive density estimator $\hat{q}_{c,\hat{\mu}} \sim N_d(\hat{\mu}(Y_1), c^2\sigma_{Y_2}^2 I_d)$ of the density of $Y_2 \sim N_d(\mu, \sigma_{Y_2}^2 I_d)$ under integrated L^2 loss, is dual to the frequentist risk of $\hat{\mu}(Y_1)$ for estimating μ under reflected normal loss L_{γ_0} with $\gamma_0 = (c^2 + 1)\sigma_{Y_2}^2$. Namely, $\hat{q}_{c,\hat{\mu}_1} \sim N_d(\hat{\mu}_1(Y_1), c^2\sigma_{Y_2} I_d)$ dominates $\hat{q}_{c,\hat{\mu}_2} \sim N_d(\hat{\mu}_2(Y_1), c^2\sigma_{Y_2}^2 I_d)$ under integrated L^2 loss iff $\hat{\mu}_1(Y_1)$ dominates $\hat{\mu}_2(Y_1)$ under loss L_{γ_0}.

In the context of Lemma 2.1, it turns out that both Kullback-Leibler and reverse Kullback-Leibler losses have squared error as a dual point estimation loss (e.g., Fourdrinier et al., 2011 for KL loss). For other $\alpha-$divergence losses, it is again reflected normal loss which is dual for plug-in predictive density estimators and for scale expansion variants as in (2.5) (also see Ghosh, Mergel and Datta, 2008 for related work).

Lemma 2.2. *(Duality between $\alpha-$divergence and reflected normal losses)*
For normal model (2.6), the frequentist risk of a predictive density estimator $\hat{q}_{c,\hat{\mu}} \sim N_d(\hat{\mu}(Y_1), c^2\sigma_{Y_2}^2 I_d)$ of the density of $Y_2 \sim N_d(\mu, \sigma_{Y_2}^2 I_d)$ under $\alpha-$divergence loss (2.2), with $|\alpha| < 1$, is dual to the frequentist risk of $\hat{\mu}(Y_1)$ for estimating μ under reflected normal loss L_{γ_0} with $\gamma_0 = (\frac{c^2}{1+\alpha} + \frac{1}{1-\alpha})\sigma_{Y_2}^2$. Namely, $\hat{q}_{c,\hat{\mu}_1} \sim N_d(\hat{\mu}_1(Y_1), c^2\sigma_{Y_2}^2 I_d)$ dominates $\hat{q}_{c,\hat{\mu}_2} \sim N_d(\hat{\mu}_2(Y_1), c^2\sigma_{Y_2}^2 I_d)$ under $\alpha-$divergence loss iff $\hat{\mu}_1(Y_1)$ dominates $\hat{\mu}_2(Y_1)$ under loss L_{γ_0}.

Proof. See the Appendix. □

The last duality result for integrated L^1 loss, due to Kubokawa, Marchand and Strawderman (2016), is limited to plug-in predictive density estimators, but applies not only for normal models, but also quite generally for unimodal spherically symmetric models.

Lemma 2.3. *(General spherically symmetric model duality for integrated L^1 losses)*
For estimating a unimodal spherically symmetric Lebesgue density of $Y_2 \sim q(\|y_2 - \mu\|^2)$, $y_2 \in \mathbb{R}^d$, under integrated L^1 loss and based on $Y_1 \sim p(\|y_1 - \mu\|^2)$, the frequentist risk of the plug-in density estimator $q(\|y_2 - \hat{\mu}(X)\|^2)$ is equal to the frequentist risk of the point estimator $\hat{\mu}(X)$ of μ under loss $4Q(\frac{\|\hat{\mu} - \mu\|}{2}) - 2$, with Q being the common univariate marginal cdf associated with q. Consequently, $q(\|y_2 - \hat{\mu}_1(Y_1)\|^2)$ dominates $q(\|y_2 - \hat{\mu}_2(Y_1)\|^2)$ iff $\hat{\mu}_1(Y_1)$ dominates $\hat{\mu}_2(Y_1)$ under loss $2Q(\frac{\|\hat{\mu} - \mu\|}{2}) - 1$.

With regards to the last result, the findings of this paper will apply for $d = 1$, $X = Y_1 \sim N(\mu, \sigma^2)$, but general cdf Q.

3. Notations, definitions, and preliminary results

For our model $X \sim N(\mu, \sigma^2)$ with $|\mu| \leq m\sigma$, we assume $\sigma = 1$ without loss of generality and we set $\lambda = |\mu|$, $R = |X|$, $r = |x|$, and ϕ as the $N(0,1)$ pdf. With F absolutely continuous, we can write $F(t) = \int_0^t f(x)\, dx$, $t \geq 0$, f representing (throughout) the Radon-Nikodym derivative of F and with $f = F'$ a.e. The problem is invariant under sign changes and equivariant estimators are even functions of X. Equivariant estimators will thus possess constant risk on orbits $S_\lambda = \{-\lambda, \lambda\}$, and we can therefore extract a best equivariant estimator for (what we will refer to as) the local problem where $|\mu| = \lambda$. This can be done by conditioning on the maximal invariant statistic R (e.g., Eaton, 1989). We hence study the conditional risks

$$\rho(\lambda, g(r), r) = E_\mu\left(F(|\delta_g(X) - \mu|)|R = r\right), \, r > 0, \tag{3.1}$$

for equivariant estimators δ_g of the form

$$\delta_g(x) = g(r)\, \text{sgn}(x), \tag{3.2}$$

with defining multiplier function $g : [0, \infty) \to \mathbb{R}$.

The decomposition leads to the evaluation of the unconditional frequentist risk

$$R(\mu, \delta_g) = E_\mu\left(F(|\delta_g(X) - \mu|)\right) = E_\lambda(\rho(\lambda, g(R), R)), \tag{3.3}$$

where $R^2 \sim \chi_1^2(\lambda^2)$ and $\lambda \in [0, m]$. Such inferences are summarized with the following Lemma and will be further applied in Section 4.4. The main idea used here to obtain dominance results, which goes back to the work of Moors (e.g., Moors, 1985), is that estimator δ_g will dominate estimator δ_{g_1} whenever it lowers the conditional risk for all $r > 0$ and for all λ (also see Marchand and Strawderman, 2004, for a presentation).

Lemma 3.1. **(a)** *For the local problem $|\mu| = \lambda$, an equivariant estimator $\delta_{g_\lambda^*}$ such that $g_\lambda^*(r) = argmin_{y \in \mathbb{R}} \{\rho(\lambda, y, r)\}$ is best equivariant;*

5

(b) *For the local problem $|\mu| = \lambda$, a sufficient condition for δ_g to dominate δ_{g_1} is*

$$\nu\{r > 0 : \rho(\lambda, g(r), r) > \rho(\lambda, g_1(r), r)\} = 0\,, \ \ and \ \nu\{r > 0 : \rho(\lambda, g(r), r) < \rho(\lambda, g_1(r), r)\} > 0\,, \tag{3.4}$$

with ν the Lebesgue measure on $\mathbb{R}$;

(c) *For the full restricted parameter space $|\mu| \leq m$, a sufficient condition for δ_g to dominate δ_{g_1} is*

$$\nu\{r > 0 : \rho(\lambda, g(r), r) > \rho(\lambda, g_1(r), r)\} = 0 \ \ for \ all \ \lambda \in [0, m]\,,$$
$$and \ \nu\{r > 0 : \rho(\lambda_0, g(r), r) < \rho(\lambda_0, g_1(r), r)\} > 0\,, \ \ for \ some \ \lambda_0 \in [0, m]\,. \tag{3.5}$$

Here is a useful expression for the conditional risk, accompanied by a further specification of a best equivariant estimator.

Lemma 3.2. **(a)** *For $\lambda, r > 0$, the conditional risk in (3.1) is given by*

$$\rho(\lambda, y, r) \ = \ \alpha_\lambda(r)\, F(|y - \lambda|) + (1 - \alpha_\lambda(r))\, F(y + \lambda)\,, \tag{3.6}$$

$$= \ F(\lambda) + \int_0^y \psi(\lambda, t, r)\, dt\,, \ \ \ if \ \ \ 0 \leq y \leq \lambda\,, \tag{3.7}$$

with $\alpha_\lambda(r) = \frac{e^{\lambda r}}{e^{\lambda r} + e^{-\lambda r}}$ and $\psi(\lambda, t, r) = (1 - \alpha_\lambda(r))\, f(\lambda + t) - \alpha_\lambda(r)\, f(\lambda - t)$.

(b) *An equivariant estimator $\delta_{g_\lambda^*}$ with $g_\lambda^*(r) = argmin_{0 \leq y \leq \lambda}\, \{\rho(\lambda, y, r)\}$ is best equivariant.*

Proof. Equation (3.6) follows since, conditional on $|X| = r$, X takes values r and $-r$ with probabilities $\alpha_\lambda(r)$ and $1 - \alpha_\lambda(r)$ respectively where $\alpha_\lambda(r) = \frac{(2\pi)^{-1/2}\, \phi(r - \lambda)}{(2\pi)^{-1/2}(\phi(r - \lambda) + \phi(r + \lambda))} = \frac{e^{\lambda r}}{e^{\lambda r} + e^{-\lambda r}}$. Expression (3.7) is a rewriting of (3.6) making use of the representation $F(t) = \int_0^t f(x) dx$. For part **(b)**, given part **(a)** of Lemma 3.1 and the continuity of ρ, it suffices to show, for a fixed (r, λ), that : (i) $\rho(\lambda, y, r) \geq \rho(\lambda, \lambda, r)$ for $y \geq \lambda$, and (ii) $\rho(\lambda, y, r) \geq \rho(\lambda, -y, r)$ for $y < 0$. These inequalities are readily established with the nondecreasing property of F, with the additional observation for (ii) that: $\rho(\lambda, y, r) - \rho(\lambda, -y, r) = (2\alpha_\lambda(r) - 1)\, \{F(|\lambda - y|) - F(|\lambda + y|)\} \geq 0$, given that $\alpha_\lambda(r) \geq 1/2$ and that $|\lambda - y| \geq |\lambda + y|$ for $y < 0, \lambda \geq 0$. $\qquad \square$

Remark 3.1. *Alternatively and equivalently, the estimator $\delta_{g_\lambda^*}$ is Bayes with respect to the two-point uniform prior $P(\mu = \lambda) = P(\mu = -\lambda) = 1/2$, which we denote as π_λ. This can be checked directly[3], but it also must be the case based on the fact that the best equivariant estimator can be represented as the Bayes estimator associated with the corresponding Haar measure, which is here the two-point uniform prior on $\{-\lambda, \lambda\}$ (e.g., Eaton, 1989).*

---

[3]Expanding on this, first observe that the expected posterior loss of the estimate δ is given by $\rho(\lambda, \delta, x)$ and that

$$\delta_{\pi_\lambda}(-x) = argmin_\delta\{\rho(\lambda, \delta, -x)\} = argmin_\delta\{\rho(\lambda, -\delta, x)\} = -argmin_\delta\{\rho(\lambda, \delta, x)\} = -\delta_{\pi_\lambda}(x)\,,$$

working with (3.6). We thus infer that $\delta_{\pi_\lambda}(x)$ is an odd function of x and that $\delta_{\pi_\lambda}(x) = argmin_y\{\rho(\lambda, y, |x|)\}\, sgn(x) = \delta_{g_\lambda^*}(x)$ for all x.

4. Main results

The main objective here is to study the efficiency of the maximum likelihood estimator δ_{mle}, as well as other boundary-valued estimators for losses $F(|d - \mu|)$ as measured by the corresponding risk in (1.3). In particular, there is the issue of inadmissibility and the question of providing dominating estimators of δ_{mle}, whenever possible. In this regard, we will particularly interested in the performance of the boundary uniform Bayes estimator $\delta_{BU}(X)$ and conditions on (m, F) for which $\delta_{BU}(X)$ dominates δ_{mle}. Other inferences, not only are useful, but also turn out to be of interest. These include monotonicity conditions, as well as conditions for which the best equivariant estimator for the local problem $|\mu| = \lambda$ takes values on $\{-\lambda, \lambda\}$.

Subsection 4.1 deals with losses, which include concave F, where the best equivariant estimator takes values on the boundary of the parameter space, and where improvement by shrinkage is not possible. Afterwards, we study losses where improvement by shrinkage **is** possible. Subsection 4.2 provides conditions and examples, while subsection 4.3 establishes key monotonicity conditions. These lead to dominance results presented in subsection 4.4, and examples and illustrations then follow in Section 5.

Before proceeding, we present the following general result which addresses the minimization problem of the conditional risks $\rho(\lambda, y, r)$ as a function of y for fixed (λ, r).

Theorem 4.1. *Let $\lambda, r > 0$ be fixed.*

(i) *Suppose that*

$$\min\left(e^{-2\lambda r}, \frac{f(\lambda - y)}{f(\lambda + y)} \right) \quad \text{is nonincreasing in } y \text{ almost everywhere on } [0, \lambda]. \qquad (4.1)$$

If $\operatorname{ess\,inf}\{f(\lambda - y)/f(\lambda + y) : y \in [0, \lambda]\} < \exp(-2\lambda r)$, then $\rho(\lambda, \cdot, r)$ is nonincreasing on $[0, y_0(\lambda, r)]$, nondecreasing on $[y_0(\lambda, r), \lambda]$ where $y_0(\lambda, r)$ is such that $\frac{f(\lambda - y)}{f(\lambda + y)} \leq e^{-2\lambda r}$ for almost all $y \geq y_0(\lambda, r)$, and $\frac{f(\lambda - y)}{f(\lambda + y)} \geq e^{-2\lambda r}$ for almost all $y \leq y_0(\lambda, r)$ and $0 \leq y_0(\lambda, r) < \lambda$;

(ii) *If $f(\lambda - y) \geq \exp(-2\lambda r) f(\lambda + y)$ for almost all $y \in [0, \lambda]$, then $\rho(\lambda, \cdot, r)$ in nonincreasing on $[0, \lambda]$.*

Consequently, a best equivariant estimator for the local problem $|\mu| = \lambda$ is given by $\delta_{g_\lambda^*}(X) = g_\lambda^*(|X|)\operatorname{sgn}(X)$, with $g_\lambda^*(r) = y_0(\lambda, r)$ for case (i), and $g_\lambda^*(r) = \lambda$ for case (ii). Also,

(iii) *If F is concave on $[0, 2\lambda]$, then $\rho(\lambda, \lambda, r) \leq \rho(\lambda, t, r)$ for all $t \in [0, \lambda]$, and $g_\lambda^*(r) = \lambda$.*

(iv) *If F is convex on $[0, 2\lambda]$ and $F'(0) < \exp(-2\lambda r)F'(2\lambda)$, then $\rho(\lambda, \cdot, r)$ is nonincreasing on $[0, y_0(\lambda, r)]$, nondecreasing on $[y_0(\lambda, r), \lambda]$, with $0 \leq y_0(\lambda, r) < \lambda$ and $y_0(\lambda, r)$ as given in part (i) above;*

(v) *If F is convex on $[0, 2\lambda]$ and $F'(0) \geq \exp(-2\lambda r)F'(2\lambda)$, then $\rho(\lambda, \cdot, r)$ is nonincreasing on $[0, \lambda]$ and $g_\lambda^*(r) = \lambda$.*

Proof. The proof of parts (i) and (ii) is based on equation (3.7). Under condition (i), there exists $0 < \delta \leq \lambda$ such that $\psi(\lambda, \cdot, r) > 0$ almost everywhere on $(\lambda - \delta, \lambda]$ showing that $\rho(\lambda, \cdot, r)$ is increasing

7

on $[\lambda - \delta, \lambda]$. Assumption (4.1) is then used to show that $y_0(\lambda, r) < \lambda$ exists. Under condition **(ii)**, we obtain that $\psi(\lambda, \cdot, r) \leq 0$ almost everywhere on $[0, \lambda]$ showing that $\rho(\lambda, \cdot, r)$ is nondecreasing on $[0, \lambda]$.

The proof of part **(iii)**, **(iv)**, and **(v)** is based on Equation (3.6). For part **(iii)** the concavity of F on $[0, 2\lambda]$ implies the concavity of $\rho(\lambda, \cdot, r)$ on $[0, \lambda]$. Since $\rho(\lambda, \cdot, r)$ is concave on $[0, \lambda]$ the function $\rho(\lambda, \cdot, r)$ reaches its minimum at 0 or λ. Since

$$
\begin{aligned}
\rho(\lambda, 0, r) &= F(\lambda) \\
&> F(2(1 - \alpha_\lambda(r))\lambda), \quad \text{(as F is increasing and $1/2 < \alpha_\lambda(r) < 1$)} \\
&\geq \alpha_\lambda(r)F(0) + (1 - \alpha_\lambda(r))F(2\lambda), \quad \text{(F is concave)} \\
&= \rho(\lambda, \lambda, r),
\end{aligned}
$$

we obtain that the function $\rho(\lambda, \cdot, r)$ reaches its minimum at λ. For parts **(iv)** and **(v)**, the convexity of F on $[0, 2\lambda]$ implies the convexity of $\rho(\lambda, \cdot, r)$ on $[0, \lambda]$. For $\rho'(\lambda, \lambda, r) = \frac{d}{dy}\rho(\lambda, y, r)|_{y=\lambda}$, we have $\rho'(\lambda, \lambda, r) = (1 - \alpha_\lambda(r))F'(2\lambda) - \alpha_\lambda(r)F'(0)$. If $\rho'(\lambda, \lambda, r) \leq 0$, then the function $\rho(\lambda, \cdot, r)$ is nonincreasing on $[0, \lambda]$, which is **(v)**. Finally, if $\rho'(\lambda, \lambda, r) > 0$, then the function $\rho(\lambda, \cdot, r)$ is increasing in a neighbourhood of λ, which leads to **(iv)**. $\qquad\square$

Example 4.1. *Theorem 4.1 covers a wide class of loss functions F identifying the best point estimate of μ for fixed λ, r. Several developments below deal with rather smooth choices of F, but Theorem 4.1 covers many more cases. Suppose, for the sake of an illustration that $F(t) = \max(t, t^2)$ for $t \in [0, 2\lambda]$. We consider separately cases:* **(A)** $\lambda > 1$, **(B)** $1/2 < \lambda \leq 1$, **(C)** $0 < \lambda \leq 1/2$.

(A) *Observe that F is not differentiable at 1, but convex on $[0, 2\lambda]$, so that we can focus on parts* **(iv)** *and* **(v)** *of Theorem 4.1 and its discriminating inequality. For $\lambda > 1$, we obtain*

$$
F'(0) \geq \exp(-2\lambda r)F'(2\lambda) \iff 1 \geq 4\lambda\exp(-2\lambda r) \iff r \geq r_0 = \frac{\log(4\lambda)}{2\lambda} .
$$

Therefore, we have that $g_\lambda^*(r) = \lambda$ for $r \geq r_0$ and $g_\lambda^*(r) = y_0(\lambda, r)$ as defined in part **(i)** *for $r < r_0$. With the choice $f(t) = 1 + (2t - 1)\mathbb{I}_{[1,2\lambda]}(t)$, we obtain*

$$
\frac{f(\lambda - y)}{f(\lambda + y)} = \begin{cases} \frac{\lambda - y}{\lambda + y} & \text{for } 0 \leq y \leq \lambda - 1 \\ \frac{1}{2(\lambda + y)} & \text{for } \lambda - 1 < y \leq 2\lambda. \end{cases}
$$

From this, we obtain for $0 < r < r_0$:

$$
y_0(\lambda, r) = \begin{cases} \lambda\tanh(\lambda r) & \text{for } r \leq \frac{\log(2\lambda - 1)}{2\lambda} \\ \frac{1}{2}\exp(2\lambda r) - \lambda & \text{for } r \geq \frac{\log(4\lambda - 2)}{2\lambda} \\ \lambda - 1 & \text{for } r \in \left(\frac{\log(2\lambda - 1)}{2\lambda}, \frac{\log(4\lambda - 2)}{2\lambda}\right). \end{cases}
$$

(B) *For $1/2 < \lambda \leq 1$, we obtain, as in part* **(A)**, *$g_\lambda^*(r) = \lambda$ for $r \geq r_0 = \frac{\log(4\lambda)}{2\lambda}$. But here, with the choice $f(t) = 1 + (2t - 1)\mathbb{I}_{[1,2\lambda]}(t)$, we have*

$$
\frac{f(\lambda - y)}{f(\lambda + y)} = \begin{cases} 1 & \text{for } 0 \leq y \leq 1 - \lambda \\ \frac{1}{2(\lambda + y)} & \text{for } 1 - \lambda < y \leq 2\lambda. \end{cases}
$$

8

From this, we obtain from part **(iv)** *of Theorem 4.1, for $0 < r < r_0$:*

$$g_\lambda^* = y_0(\lambda, r) = \begin{cases} 1 - \lambda & \text{for } r \leq \frac{\log 2}{2\lambda} \\ \frac{1}{2}\exp(2\lambda r) - \lambda & \text{for } r \geq \frac{\log 2}{2\lambda} . \end{cases}$$

We point out that parts **(i)** *and* **(ii)** *of Theorem 4.1 can also be applied directly for the above analysis.*

(C) *Finally, for $0 < \lambda \leq 1/2$, F is concave on $[0, 2\lambda]$, so that $g_\lambda^*(r) = \lambda$ for all $r > 0$ as indicated by part* **(iii)** *of Theorem 4.1. Alternatively, F is still convex as well on $[0, 2\lambda]$ and an application of part* **(iv)** *of Theorem 4.1 leads to the same evaluation of g_λ.*

4.1. Situations where the best equivariant estimator is not a shrinker

Building on Theorem 4.1, we expand here on situations where the best equivariant estimator for $|\mu| = \lambda$ is not a shrinker. These include cases where F is a concave function on $[0, 2\lambda]$.

Theorem 4.2. *Assume that $f(\lambda - y) \geq f(\lambda + y)$ for almost all $y \in [0, \lambda]$. Then,*

(a) *the function $\rho(\lambda, \cdot, r)$ is nonincreasing on $[0, \lambda]$ for all fixed $r > 0$, $\lambda > 0$ and, consequently, $g_\lambda^*(r) = \lambda$; for all fixed (r, λ) ;*

(b) *The estimator $\delta_{g_\lambda}^*(x) = \lambda \, sgn(x)$ with $0 \leq \lambda \leq m$ is, for the full problem $\mu \in \Theta(m)$, a Bayes and admissible estimator of μ.*

Proof. In view of Remark 3.1 and part **(a)**, estimators $\delta_{g_\lambda}^*(x)$ are (essentially) unique Bayes with finite Bayes risk, and thus admissible. This establishes part **(b)**. Part **(a)** is a corollary of Theorem 4.1. Indeed, given that $f(\lambda - y) \geq f(\lambda + y)$ for almost all $y \in [0, \lambda]$, part **(ii)** of Theorem 4.1 yields the result. $\qquad\square$

The above result tells us that estimators taking values on the boundary $\{-m, m\}$ of the parameter space $[-m, m]$, such as δ_{mle}, cannot be dominated by making use of the conditional risk method of Moors (1985), or Marchand and Perron (2001), for a large class of losses F, which include concave F on $[0, 2\lambda]$. Of course, for the specific case of absolute-value loss, we have the stronger result that δ_{mle} is admissible (Iwasa and Moritani, 1997; Kucerovksy et al. 2009), but the result here applies to a whole range of losses.

Remark 4.1. *Examples of losses with concave F in (1.1) for which Theorem 4.2 include L^s loss $F(t) = t^s$ with $0 < s \leq 1$; $F(t) = 1 - e^{-t/\gamma}$ with $\gamma > 0$; $F(t) = \Phi(t/\gamma)$ with $\gamma > 0$; $F(t) = \left(\frac{t}{t+a}\right)^c$ with $a > 0, 0 < c \leq 1$; and more generally cases where F is the cdf associated with a decreasing density F' on $\mathbb{R}_+$, such as the additional examples $F(t) = \frac{2}{\pi}\arctan(\frac{t}{\gamma})$, $\gamma > 0$, (i.e, half-Cauchy), and $F(t) = \frac{e^{\frac{t}{\gamma}} - 1}{e^{\frac{t}{\gamma}} + 1}$, $\gamma > 0$, (i.e., half-Logistic).*

In terms of the predictive density estimation set-up of Section 2, we obtain the following as a consequence of Theorem 4.2.

Corollary 4.1. *For univariate model (2.1) with $Y_1 \sim N(\mu, \sigma_{Y_1}^2)$, unimodal q and parametric constraint $\mu \in [-m, m]$, and for estimating the density $q(|y_2 - \mu|^2)$, $y_2 \in \mathbb{R}$, under integrated L^1 loss, plug-in predictive density estimators $q(|y - \lambda_0 sgn(Y_1)|^2)$, $y \in \mathbb{R}$, with $0 \le \lambda_0 \le m$, are admissible among plug-in density estimators $q(|y - \hat{\mu}(Y_1)|^2)$, $y \in \mathbb{R}$.*

Proof. From Lemma 2.3, we have that $q(|y - \hat{\mu}(Y_1)|^2)$ dominates $q(|y - \lambda_0 sgn(Y_1)|^2)$ under L^1 loss for $\mu \in [-m, m]$ iff $\hat{\mu}(Y_1)$ dominates $\lambda_0 \, sgn(Y_1)$ with loss $F(t) = Q(\frac{t}{2}) - 1$, Q being the cdf of $Y_2 \sim q(|y_2 - \mu|^2)$. The unimodality assumption implies that Q, and thus F, are concave on $\mathbb{R}_+$. Finally, Theorem 4.2 (see as well discussion that follows) implies that the estimator $\lambda_0 \, sgn(Y_1)$ is admissible for the point estimation problem which implies the result. $\square$

Remark 4.2. *In the predictive density estimation context of Corollary 4.1, we can also establish the stronger result that $q(|y - \lambda_0 sgn(Y_1)|^2)$, $y \in \mathbb{R}$, is admissible among all predictive density estimators. Indeed, as pointed out by Kubokawa, Marchand and Strawderman (2016), the predictive density estimator $q(|y - med(\mu|y_1)|^2)$, where $med(\mu|y_1)$ is the posterior median, is Bayes for two-point priors. Since $med(\mu|y_1) = \lambda_0 sgn(y_1)$ for all $y_1 \ne 0$ and the prior $\pi(\lambda_0) = \pi(-\lambda_0) = 1/2$, we conclude that the predictive density estimator $q(|y - \lambda_0 sgn(Y_1)|^2)$ is Bayes (essentially unique), and admissible in view of the finiteness of the Bayes risk.*

4.2. Situations where the best equivariant estimator is a shrinker

We now turn to situations where the best equivariant estimator $\delta_{g_\lambda^*}$ for $|\mu| = \lambda$ takes values on the interior of the interval $[-\lambda, \lambda]$. As a consequence of Theorem 4.1, we obtain the following description of $\delta_{g_\lambda^*}$.

Corollary 4.2. *Let $\lambda > 0, r > 0$ be fixed. Assume that condition (4.1) holds, that F is differentiable at 0 and 2λ, with $F'(0) = 0$ and $F'(2\lambda) > 0$. Then part **(i)** of Theorem 4.1 applies and the best equivariant estimator for $|\mu| = \lambda$ is given by $\delta_{g_\lambda^*}(X) = g_\lambda^*(|X|) \, sgn(X)$, with $g_\lambda^*(r) = y_0(\lambda, r) = y_0$, with y_0 satisfying $\frac{f(\lambda - y)}{f(\lambda + y)} \le e^{-2\lambda r}$ for $y \ge y_0$, and $\frac{f(\lambda - y)}{f(\lambda + y)} \ge e^{-2\lambda r}$ for $y \le y_0$. In cases where f is continuous, y_0 is a solution in $y \in [0, \lambda]$ of*

$$e^{2\lambda r} \frac{f(\lambda - y)}{f(\lambda + y)} = 1. \tag{4.2}$$

Proof. Equation (3.6) is used to show that $\rho'(\lambda, \lambda, r) > 0$ so there exists $0 < \epsilon < \lambda$ such that $\rho(\lambda, \lambda - \epsilon, r) < \rho(\lambda, \lambda, r)$. To justify that part **(i)** of Theorem 4.1 applies, we use equation (3.7) to derive

$$0 < \rho(\lambda, \lambda, r) - \rho(\lambda, \lambda - \epsilon, r) = \int_{\lambda - \epsilon}^{\lambda} \psi(\lambda, t, r) dt,$$

which shows that the set $\{t \colon \psi(\lambda, t, r) > 0\}$ has positive Lebesgue measure, so that $\operatorname{ess\,inf}\{f(\lambda - y)/f(\lambda + y) \colon y \in [0, \lambda]\} < \exp(-2\lambda r)$. $\square$

Specific instances where condition (4.1) holds, and Corollary 4.2 is applicable, includes cases where f is continuous on $[0, 2\lambda]$ and where either: **(I)** F is convex on $[0, 2\lambda]$, or **(II)** $\frac{f(\lambda - y)}{f(\lambda + y)}$ is nonincreasing in y for $y \in [0, \lambda]$, or **(III)** $\frac{f(\lambda - y)}{f(\lambda + y)}$ is nondecreasing in y for $y \ge y_1$, and is nonincreasing in y for $y < y_1$ for some $y_1 \in (0, \lambda)$. In such cases, in contrast to concave F cases, best equivariant estimators for

the local problem $|\mu| = \lambda$ are shrinkers, i.e., $|\delta_{g^*_\lambda}(x)| < \lambda$ for almost all x. This finding, along with further monotonicity properties still to establish, will be exploited below to identify complete classes of estimators and obtain dominating estimators of boundary-taking estimators such as $\delta_{\text{mle}}(X)$ for the global problem with $|\mu| \leq m$. Our formulation highlights namely conditions **(I)** and **(II)**, as these will play a key role later on for monotonicity and dominance results. We pursue with illustrations.

Example 4.2. *As mentioned in the previous paragraph, the shrinkage conclusion of Corollary 4.2 applies to loss functions in (1.1) with F convex on $[0, 2\lambda]$ (condition **I**), such as L^s loss $F(t) = t^s$, with $s > 1$, and such as $F(t) = e^{\beta t} - \beta t - 1$ with $\beta \neq 0$ (i.e., a symmetrized Linex loss function), among many other examples. For L^s loss with $F(t) = t^s$, $s > 1$, the best equivariant multiplier $g^*_\lambda(r)$ is obtained from (4.2) as the solution in $y \in [0, \lambda)$ of $e^{2\lambda r} \left(\frac{\lambda - y}{\lambda + y}\right)^{s-1} = 1$. For L^2 loss, we recover the familiar $g^*_\lambda(r) = \lambda \tanh(\lambda r)$ and $\delta_{g^*_\lambda}(x) = \lambda \tanh(\lambda x)$ (e.g., Casella and Strawderman, 1981).*

Example 4.3. *Investigating the conditions of Corollary 4.2 for reflected normal loss and more generally to loss functions (1.1) with $F(t) = 1 - e^{-\frac{t^\alpha}{2\gamma}}$, $\gamma > 0$, $\alpha > 1$, we have that F is differentiable, $F'(0) = 0$, $F'(2\lambda) > 0$, and convexity of F on $[0, 2\lambda]$ if and only if $\lambda \leq \frac{1}{2} \left(\frac{2\gamma(\alpha - 1)}{\alpha}\right)^{1/\alpha}$. However, we can show that condition **(III)** above, and thus condition (4.1) required for Corollary 4.2 to apply, holds for all $\gamma > 0, \alpha \in (1, 3], \lambda > 0$, telling us that the best equivariant estimator for the local problem $|\mu| = \lambda$ is a shrinker in such cases. Indeed, a calculation permits us to write*

$$\frac{\partial}{\partial y} \log\left(\frac{f(\lambda - y)}{f(\lambda + y)}\right) = \frac{1}{\lambda^2 - y^2} \left(-2\lambda(\alpha - 1) + \frac{\alpha}{2\gamma} T_\alpha(y)\right),$$

*with $T_\alpha(y) = (\lambda - y)(\lambda + y)^\alpha + (\lambda + y)(\lambda - y)^\alpha$. Since $\lim_{y \to \lambda^-} T_\alpha(y) = -2\lambda(\alpha - 1) < 0$, condition **(III)** above is satisfied as long as $T_\alpha(y)$ is nonincreasing in y for $y \in [0, \lambda]$. We can then write*

$$\frac{\partial}{\partial y} T_\alpha(y) = (\lambda + y)^\alpha G(\rho, \alpha),$$

with $G(\rho, \alpha) = \rho^\alpha - \alpha \rho^{\alpha - 1} + \alpha \rho - 1$, and $\rho = \frac{\lambda - y}{\lambda + y} \in [0, 1]$.

*We conclude by showing that $G(\rho, \alpha) \leq 0$ for all $\rho \in [0, 1], \alpha \in (1, 3]$. Observe that: **(a)** $G(0, \alpha) = -1$, **(b)** $G(1, \alpha) = 0$, **(c)** $\frac{\partial}{\partial \rho} G(\rho, \alpha)|_{\rho=1} = \alpha(3 - \alpha) \geq 0$ for all $\alpha \in (1, 3]$. Furthermore, it is easy to see that: **(d)** $G(\cdot, \alpha)$ is convex on $[0, 1]$ whenever $\alpha \leq 2$ and **(e)** $G(\cdot, \alpha)$ is convex on $[0, \alpha - 2]$, then concave on $(\alpha - 2, 3)$ whenever $\alpha \in (2, 3]$. Finally, observe that both possibilities **(a)**, **(b)** and **(d)**, which is the case for $\alpha \in (1, 2]$, and **(a)**, **(b)**, **(c)** and **(e)**, which is the case for $\alpha \in (2, 3]$, imply $G(\rho, \alpha) \leq 0$ for all $\rho \in [0, 1]$. [4] Condition **(III)** is thus satisfied for all losses $F(t) = 1 - e^{-\frac{t^\alpha}{2\gamma}}$, for all $\lambda > 0$, $\gamma > 0$, and $\alpha \in (1, 3]$. Finally, we point out that stronger condition **(II)**, that the ratio $\frac{f(\lambda - y)}{f(\lambda + y)}$ is nonincreasing in $y \in [0, \lambda]$, is seen to be satisfied with the above analysis virtue if and only if $\frac{\partial}{\partial y} \log\left(\frac{f(\lambda - y)}{f(\lambda + y)}\right)$ is negative at $y = 0^+$, which is equivalent to $\lambda \leq \left(\frac{2\gamma(\alpha - 1)}{\alpha}\right)^{1/\alpha}$.*

As an example of the best equivariant estimator $\delta_{g^*_\lambda}$, consider reflected normal loss L_γ (with $\alpha = 2$ above), for which we obtain from (4.2) that $g^*_\lambda(r)$ is the solution in $y \in [0, \lambda)$ of

$$e^{2\lambda(r + \frac{y}{\gamma})} \left(\frac{\lambda - y}{\lambda + y}\right) = 1. \tag{4.3}$$

---

[4] For $\alpha > 3$, we see that $G(\rho, \alpha)$ is positive for ρ close to 1 which implies that $T_\alpha(y)$ is nondecreasing for small y in such cases, and that condition (4.1) required for Corollary 4.2 does not hold.

11

We conclude this example by pointing out that the Bayes estimator $\delta_{g_\lambda^*}$ and its defining equation (4.2) were studied by Towhidi and Behboodian (2002) with the objective of determining conditions for minimaxity (also see van Eeden, 2006; pp. 48-49). We will comment on this below in Example 5.3.

4.3. Monotonicity properties of $g_\lambda^*(r)$

We require the following monotonicity properties of the best equivariant multiplier $g_\lambda^*(r)$.

Lemma 4.1. *Consider the context of Corollary 4.2, the best equivariant estimator $\delta_{g_\lambda^*}$ under loss $F(|d - \mu|)$ for $|\mu| = \lambda$, $g_\lambda^*(r)$ being the solution in y of equation (4.2). Suppose that f is continuous on $[0, 2m]$, $f(0) = 0$, and $f > 0$ on $(0, 2m]$.*

(a) *Let $r > 0$ be fixed. A sufficient condition for $g_\lambda^*(r)$ to be nondecreasing as a function of $\lambda \in (0, m]$ is the logconcavity of f on $[0, 2m]$ and that the ratio $\frac{f(\lambda - y)}{f(\lambda + y)}$ be nonincreasing in $y \in [0, \lambda]$ for all $\lambda \in (0, m]$.*

(b) *Let $\lambda > 0$ be fixed. A sufficient condition for $g_\lambda^*(r)$ to be nondecreasing in r for $r > 0$ is that the ratio $\frac{f(\lambda - y)}{f(\lambda + y)}$ be nonincreasing in $y \in [0, \lambda]$.*

Proof. For part **(a)**, the logconcavity assumption on f implies that $\frac{f(\lambda - y)}{f(\lambda + y)}$ is nondecreasing in λ for $\lambda > y$ and fixed $y > 0$. Threeofore the left-hand side of (4.2) also is nondecreasing in λ for $\lambda > y$ and fixed $y > 0$. Now, in view of this and the monotone decreasing condition for the ratio $\frac{f(\lambda - y)}{f(\lambda + y)}$, it must be the case that $g_\lambda^*(r)$ is nondecreasing in λ for $\lambda \in (0, m)$ given the equilibrium in (4.2). Part **(b)** is established analogously as equation (4.2), along with the given monotonicity condition on $\frac{f(\lambda - y)}{f(\lambda + y)}$, imply that $\frac{f(\lambda - g_\lambda^*(r))}{f(\lambda + g_\lambda^*(r))}$ is nonincreasing in both r and $g_\lambda^*(r)$, whence the monotone increasing property of $g_\lambda^*(r)$ in r. $\square$

Example 4.4. *(L^s losses and other convex losses - continuation of Example 4.2)*
*The logconcavity condition on f in part **(a)** of Lemma 4.1 is quite weak and satisfied by all the losses studied up to now, as well as many more. It may be interpreted as a condition stipulating that the loss does not rise too steeply from the origin as a function of the error of estimation $d - \mu$. An example of a loss in (1.1) that does not satisfy such a condition, and penalizes severely error of estimation, is $F(t) = e^{t^2} - 1$. Among loss functions satisfying the condition, are included L^s loss with $s > 1$ and symmetrized Linex loss given in Example 4.2. Furthermore, losses in (1.1) with convex F have ratios $\frac{f(\lambda - y)}{f(\lambda + y)}$ which decrease in y since $f = F'$ is increasing. We thus conclude that the monotonicity properties of Lemma 4.1 hold for all L^s losses with $s > 1$, symmetrized Linex losses, and more generally all strictly convex losses in (1.1) with log concave F'.*

Example 4.5. *(Reflected normal loss and extensions-continuation of Example 4.3)*
As seen in Example 4.3, for loss functions in (1.1) with $F(t) = 1 - e^{-\frac{t^\alpha}{2\gamma}}$, $\gamma > 0$ and $\alpha \in (1, 3]$, the ratio $\frac{f(\lambda - y)}{f(\lambda + y)}$ decreases in $y \in [0, \lambda]$ for all $\lambda \in (0, m]$, if and only if $m \leq (\frac{2\gamma(\alpha - 1)}{\alpha})^{1/\alpha}$. For reflected normal loss ($\alpha = 2$), this simplifies to $m \leq \sqrt{\gamma}$. Since F' is logconcave on $\mathbb{R}_+$, these conditions are sufficient for the monotonicity properties of Lemma 4.1 to hold.

Remark 4.3. *The monotonicity properties of Lemma 4.1 need not hold in the absence of the given conditions. An illustration is given by Example 4.1 where, for $1/2 < \lambda \leq 1$ and small enough values of r ($r < r_0$ as given), we have $g_\lambda^*(r) = 1 - \lambda$ which is strictly decreasing in λ for $1/2 < \lambda \leq 1$.*

4.4. Dominance results

With the conditional risk properties of Corollary 4.2 and Lemma 4.1's conditions for monotonicity, Lemma 3.1 specializes to the following dominance results applicable to estimators that that take values on or too close to boundary of the parameter space.

Theorem 4.3. *Consider estimating the mean μ of $X \sim N(\mu, 1)$ under loss $F(|d - \mu|)$ as in (1.1).*

(a) *Let $\lambda \geq 0, r > 0$ be fixed and suppose that the conditions of Corollary 4.2 hold. Suppose g_1 is such that $g_1(r) > g_\lambda^*(r)$. Define $a(\lambda, g_1(r), r) = 0$ if $\rho(\lambda, 0, r) < \rho(\lambda, g_1(r), r)$, and otherwise define $a(\lambda, g_1(r), r)$ as the solution in $a \in [0, g_\lambda^*(r))$ of $\rho(\lambda, a, r) = \rho(\lambda, g_1(r), r)$. Then, for $g(r) \neq g_1(r)$, we have $\rho(\lambda, g(r), r) \leq \rho(\lambda, g_1(r), r)$ if and only if $a(\lambda, g_1(r), r) \leq g(r) < g_1(r)$, with equality occurring only for $g(r) = a(\lambda, g_1(r), r) > 0$, or $g(r) = 0$ and $\rho(\lambda, 0, r) = \rho(\lambda, g_1(r), r)$.*

(b) *Consider the parameter space $\mu \in [-m, m]$, set $\bar{g}_m(r) = \sup_{\lambda \in [0, m]} \{g_\lambda^*(r)\}$, and suppose that g_1 is such that $A_{g_1} = \{r \geq 0 : g_1(r) > \bar{g}_m(r)\} \neq \emptyset$. Then, if the conditions of Corollary 4.2 are satisfied for all $\lambda \in (0, m]$, δ_g dominates δ_{g_1} whenever $g(r) = g_1(r) \mathbb{I}_{A_{g_1}^c}(r) + g_0(r) \mathbb{I}_{A_{g_1}}(r)$ with $g_0(r) \in \left[\sup_{\lambda \in [0, m]} a(\lambda, g_1(r), r)), g_1(r) \right)$.*

(c) *Further, assume that f is logconcave on $[0, 2m]$ and that the ratio $\frac{f(\lambda - y)}{f(\lambda + y)}$ is nonincreasing in $y \in [0, \lambda]$ for all $\lambda \in (0, m]$. Then, part (b) applies with $\bar{g}_m \equiv g_m^*$.*

Proof. *Part (a) follows as a direct application of part (b) of Lemma 3.1 and part (a) of Theorem 4.1 which tells us that the conditional local risk $\rho(\lambda, y, r)$ is nonincreasing in y for $0 < y < g_\lambda^*(r)$, and nondecreasing for $y > g_\lambda^*(r)$. Part (c) follows from part (b) since the additional assumptions and an application of Lemma 4.1 imply that $\bar{g}_m \equiv g_m^*$. For part (b), Theorem 4.1 applies and tells us that $g_1(r) > g_\lambda^*(r)$ for all $\lambda \in [0, m]$ whenever $r \in A_{g_1}$. An application of part (a) implies for the conditional risks that*

$$\rho(\lambda, y, r) \leq \rho(\lambda, g_1(r), r) \quad \text{for all } \lambda \in [0, m], r \in A_{g_1}, \text{ and } y \in \left[\sup_{\lambda \in [0, m]} a(\lambda, g_1, r)), g_1(r) \right),$$

and the dominance result follows. ☐

The previous result is as a complete class result in the spirit as those obtained by Moors (1985), as well as Marchand and Perron (2001, 2005, 2009), applicable to squared error loss. The novelty here is that the results are presented for a class of loss functions, which include strictly convex loss functions, as well as for reflected normal loss L_γ, the latter having motivated our study through the correspondence with predictive density estimation. Practically speaking, equivariant estimators δ_{g_1} that take values on or too close to the boundary $\{-m, m\}$, are inadmissible and improved by projecting towards δ_{g_m}, and even beyond δ_{g_m}, as specified in the previous Theorem.

Remark 4.4. *Among the dominating estimators δ_g given in Theorem 4.3, a complete subclass of estimators is obtained by taking $g_0(r) \in \left[\sup_{\lambda \in [0, m]} a(\lambda, g_1, r)), \bar{g}_m(r) \right]$.*

We now apply the dominance results to the restricted maximum likelihood estimator of μ given in (1.2). For the loss functions satisfying the required conditions, dominating estimators can be determined by applying part (b) of Theorem 4.3.

Corollary 4.3. *Consider estimating the mean μ of $X \sim N(\mu, 1)$ under loss $F(|d - \mu|)$ as in (1.1) under the parametric restriction $\mu \in [-m, m]$. Suppose that the conditions of Corollary 4.2 are satisfied for all $\lambda \in (0, m]$. Then δ_{mle} is inadmissible. Dominating estimators include δ_g with $g(r) = \bar{g}_m(r) \wedge g_{mle}(r)$, as well as other estimators given by part **(b)** of Theorem 4.3.*

We seek more and search for conditions under which the Bayes boundary estimator δ_{BU} satisfies the dominance conditions. For this, we will require that δ_{BU} shrinks δ_{mle} towards 0, but we obtain a quite simple condition on m for such a shrinkage and dominance to occur.

Theorem 4.4. *Consider estimating the mean μ of $X \sim N(\mu, 1)$ under loss $F(|d - \mu|)$ as in (1.1) under the parametric restriction $\mu \in [-m, m]$. Suppose that f is twice-differentiable and $f(0) = 0$. Suppose $f = F'$ is logconcave on $[0, 2m]$ **(C1)**, $\frac{f(\lambda - y)}{f(\lambda + y)}$ is nonincreasing in $y \in [0, \lambda]$ for all $\lambda \in (0, m]$ **(C2)**, and $(\log f)'$ is convex on $(0, m]$ **(C3)**. Then δ_{BU} dominates δ_{mle} for $m \leq c_0$ where $c_0 \frac{f(c_0)}{f'(c_0)} = 1$.*

Proof. With the given assumptions, we can apply Corollary 4.3, as well as part **(c)** of Theorem 4.3 for $\delta_g \equiv \delta_{g_m^*}$ (i.e., $\delta_g \equiv \delta_{BU}$). There remains to further show, under the given condition $m \leq c_0$, that δ_{BU} shrinks δ_{mle} towards 0, which is equivalent to $\frac{g_m^*(r)}{r} \leq 1$ for all $0 < r < m$. Set $h(t) = \log f(t)$, so that defining equation (4.2) for $m = \lambda$ is written as

$$2mr + h(m - g_m^*(r)) - h(m + g_m^*(r)) = 1. \tag{4.4}$$

Differentiate the above twice with respect to r to obtain

$$2m = \frac{\partial g_m^*(r)}{\partial r} \left(h'(m - g_m^*(r)) + h'(m + g_m^*(r)) \right), \tag{4.5}$$

$$\frac{\partial^2 g_m^*(r)}{\partial^2 r} \left(h'(m - g_m^*(r)) + h'(m + g_m^*(r)) \right) = \left(\frac{\partial g_m^*(r)}{\partial r} \right)^2 \left(h''(m - g_m^*(r)) - h''(m + g_m^*(r)) \right). \tag{4.6}$$

Now, observe that the right-hand side of (4.6) is negative by virtue of convexity condition **C3** on $h' = (\log f)'$. With $h'(m - g_m^*(r)) + h'(m + g_m^*(r)) \geq 0$ following from condition **C2** and the monotone increasing in r property of $g_m^*(r)$ (Lemma 4.1), we conclude that $g_m^*(r)$ must be a concave function of $r > 0$ under the given assumptions.

The concavity of $g_m^*(r)$ implies that $\frac{g_m^*(r)}{r}$ is decreasing in $r > 0$ since $g_m^*(r)$ is increasing in r and since $g_m^*(0) = 0$ (equation (4.4)). Therefore a necessary and sufficient condition for the shrinkage property $g_m^*(r) \leq r$, for $r > 0$, to occur is

$$\frac{\partial g_m^*(r)}{\partial r} \Big|_{r=0^+} \leq 1$$

$$\iff \frac{2m}{h'(m - g_m^*(0)) + h'(m + g_m^*(0))} \leq 1$$

$$\iff m \frac{f(m)}{f'(m)} \leq 1,$$

where we have further made use of (4.5). To conclude the proof, there remains to justify that $c \frac{f(c)}{f'(c)} = c \frac{1}{h'(c)}$ increases in c, for $0 < c \leq m$. But, this is indeed the case since $\frac{1}{h'(c)}$ increases in c, $c \in (0, m]$, given condition **C1**, and since $h'(c) > 0$ for $0 < c \leq m$ given condition **C2**, which tells us that $h'(c - y) + h'(c + y) \geq 0$ for all $y \in (0, c)$, and namely for $y \to 0^+$. $\qquad\square$

5. Examples and implications for predictive density estimation

We illustrate here how our dominance results apply for various loss functions, providing also implications for predictive density estimation as well as numerical risk function comparisons for reflected normal loss.

Example 5.1. *(L^s loss with $s > 1$)*
*As seen in Example 4.4, conditions (**C1**) and (**C2**) of Theorem 4.4 hold for loss $|d - \mu|^s$ with $s > 1$. Condition (**C3**) also holds as $(\log f(t))' = \frac{(s-1)}{t}$ is convex on $(0, m]$. With $m\frac{f(m)}{f'(m)} = \frac{m^2}{s-1}$, we conclude that the boundary uniform Bayes estimator δ_{BU} dominates δ_{mle} for $m \leq \sqrt{s-1}$. For $s = 2$ (i.e., squared error loss), this becomes $m \leq 1$ thus recovering a dominance result first obtained by Casella and Strawderman (1981), and with the entire proof an L^s loss extension of Marchand and Perron's (2001) univariate squared error loss result.*

Example 5.2. *(Reflected normal loss and extensions; continuation of Examples 4.3 and 4.5)*
Consider again the class of loss functions $F(|d - \mu|)$ with $F(t) = 1 - e^{-\frac{t^\alpha}{2\gamma}}$, $\gamma > 0$ and $\alpha > 1$ and label these as $F_{\alpha,\gamma}$.

- *As seen in Example 4.3, the conditions of Corollary 4.2 are satisfied for all $\gamma > 0, 1 < \alpha \leq 3$. It thus follows (Corollary 4.3) that the estimator δ_{mle} is inadmissible for all such losses with dominating estimators provided by part (**b**) of Theorem 4.3, and including δ_g with $g(r) = \bar{g}_m(r) \wedge g_{mle}(r)$.*

- *As seen in Example 4.5, conditions **C1** and **C2** of Theorem 4.4 are satisfied, and equivalently Theorem 4.3(c)'s conditions, as long as $m \leq c_1 = (\frac{2\gamma(\alpha-1)}{\alpha})^{1/\alpha}$, $1 < \alpha \leq 3$. Consequently, part (**c**) of Theorem 4.3, as well as Corollary 4.3, apply with $\bar{g}_m = g_m^*$ whenever $m \leq c_1$.*

- *For applying Theorem 4.4, consider $m \leq c_1$ (as above) so that conditions **C1** and **C2** are satisfied. For condition **C3**, we have that $h'(t) = (\log f(t))' = \frac{\alpha-1}{t} - \frac{\alpha t^{\alpha-1}}{2\gamma}$ is convex on $(0, m)$ iff $1 < \alpha \leq 2$, or $\alpha > 2$ and $m \leq c_2 = (\frac{4\gamma}{\alpha(\alpha-2)})^{1/\alpha}$. For δ_{BU} to dominate δ_{mle}, we further require $m \leq c_0$, where*

$$\frac{c_0}{h'(c_0)} = 1 \iff \frac{c_0}{\frac{\alpha-1}{c_0} - \frac{\alpha c_0^{\alpha-1}}{2\gamma}} = 1 \iff 2\gamma c_0^2 + \alpha c_0^\alpha = 2\gamma(\alpha-1).$$

  Observe that $2\gamma c_1^2 + \alpha c_1^\alpha > 2\gamma(\alpha-1)$, which implies that $c_0 \leq c_1$. Theorem 4.4 thus implies that δ_{BU} dominates δ_{mle}: (i) under loss $F_{\alpha,\gamma}$ with $1 < \alpha \leq 2$ whenever $m \leq c_0$; and (ii) under loss $F_{\alpha,\gamma}$ with $2 < \alpha \leq 3$, whenever $m \leq c_0 \wedge c_2$. Since $c_0 \leq c_2 \iff c_2^2 + \frac{2}{\alpha-2} \geq \alpha - 1 \impliedby 2 < \alpha \leq 3$, this last condition simply reduces to $m \leq c_0$.

- *For reflected normal loss with $\alpha = 2$, we obtain explictly $c_0 = \sqrt{\frac{\gamma}{1+\gamma}}$, and thus the dominance of δ_{BU} over δ_{mle} whenever $m \leq \sqrt{\frac{\gamma}{1+\gamma}}$. Observe that this cut-off point approaches 1 for large γ, which corresponds to the cut-off point for dominance under squared error loss (see Example 5.1) and which is plausible given the heuristic argument following (1.4).*

Remark 5.1. *It is of interest to assess how a Bayes estimator relative to a loss function performs for other loss functions. Some of the theoretical developments are useful for addressing such a question. As an illustration, consider reflected normal loss L_γ in (1.4) and the corresponding boundary uniform Bayes estimator $\delta_{BU,\gamma}(x) = g_{BU,\gamma}(|x|)sgn(x)$, where $g_{BU,\gamma}(r)$ solves (4.3) for $\lambda = m$. Taking logarithms on both sides and since $\frac{1}{2}\log\frac{1+u}{1-u} = \tanh^{-1}(u)$ for $u \in (0,1)$, we obtain alternatively*

$$g_{BU,\gamma}(r) = m\tanh\left(m(r + \frac{g_{BU,\gamma}(r)}{\gamma})\right), \quad \text{for all} \quad m, r, \gamma > 0.$$

Clearly then, we have $g_{BU,\gamma}(r) \geq m\tanh(mr)$ for all $m, r, \gamma > 0$ which, in other words, tells us that $\delta_{BU,\gamma}$ always expands on the squared error loss boundary uniform Bayes estimator $\delta_{BU}(x) = m\tanh(m|x|)sgn(x)$. It thus follows that

$$m\tanh(mr) < g_{BU,\gamma}(r) < g_{mle}(r)$$

for all $r > 0$ whenever $m \leq \sqrt{\frac{\gamma}{1+\gamma}}$. An application of Corollary 4.3 and Corollary 4.2 for squared error loss $(d-\mu)^2$ therefore implies that $\delta_{BU,\gamma}$ dominates δ_{mle} under squared error loss when $m \leq \sqrt{\frac{\gamma}{1+\gamma}}$. A summary of inferences under reflected normal loss is as follows.

Theorem 5.1. *Consider estimating the mean μ of $X \sim N(\mu, 1)$ under the parametric restriction $\mu \in [-m, m]$. Under reflected normal loss L_γ in (1.4), the estimator $\delta_{BU,\gamma}$ dominates δ_{mle} under the condition $m \leq \sqrt{\frac{\gamma}{1+\gamma}}$. Otherwise for $m > \sqrt{\frac{\gamma}{1+\gamma}}$, δ_g with $g(r) = g_{BU,\gamma} \wedge g_{mle}(r)$ dominates δ_{mle}. Finally, these dominance results persist for squared error loss.*

Example 5.3. *As a continuation of Example 5.2, Figure 1 represents the comparative performance of δ_{BU} and δ_{mle} in terms of ratio of reflected normal risks (loss L_γ in 1.4) as a function of $\lambda = |\mu| \in [0, m]$ for $m = 0.5, 0.75$ and $\gamma = 1, 2, 4$. For $m = 0.5$, the sufficient condition $m \leq \sqrt{\frac{\gamma}{1+\gamma}}$ for dominance is satisfied for the γ's here and the gains in risk are quite important, and even more so for larger the value of γ. For $m = 0.75$, the gains in risk are significant, but more modest. For $m = 1.0$ (see Figure 2) comparison is more nuanced with relatively poor performance of δ_{BU} for small values of $|\mu|$. This is due to the fact to large values of $|\delta_{BU}|$ (i.e., it does not shrink much).*

In such cases where δ_{BU} does not dominate δ_{mle}, its truncation onto δ_{mle} given by $\delta_{g_1}(x) = (g_{BU}(r) \wedge g_{mle}(r))sgn(x)$ necessarily dominates δ_{mle} (Corollary 4.3). However, gains in risk as witnessed by Figure 2 are quite slim. The Bayes estimator under L^2 loss, none other than the posterior mean given by $\delta_{g_2}(x) = m\tanh(mx)$, remains an interesting benchmark. As seen above in Remark 5.1, it shrinks less than δ_{BU} with corresponding good relative risk performance for small of moderate values of $|\mu|$ as suggested by numerical evaluations such as those illustrated by Figure 2.

To conclude, Figure 2 suggests that the risk of δ_{BU} attains its maximum on the boundary of the parameter space $[-m, m]$ and that, consequently, it is minimax for $m = 0.75$ and $\gamma = 1$ (among other values of (m, γ) where m is small). This is not predicted by Towhidi and Behboodian's (2002) minimaxity condition $m \leq (\frac{\sqrt{\gamma}}{2} \wedge \sqrt{\frac{\gamma}{1+\gamma}}) = 1/2$ (for $\gamma = 1$), but still plausible nevertheless.

Example 5.4. *(Gamma cdf loss function) Let us illustrate how the results above apply for loss functions $F(|d-\mu|)$ with F given by a Gamma(α, β) cdf on $\mathbb{R}_+$ such that $f(t) = F'(t) = \frac{t^{\alpha-1}e^{-t/\beta}}{\Gamma(\alpha)\beta^\alpha}, \alpha, \beta > 0$. For $\alpha \leq 1$, Theorem 4.2, or part (iii) of Theorem 4.1, apply and estimators such as $\lambda_0 sgn(x)$*
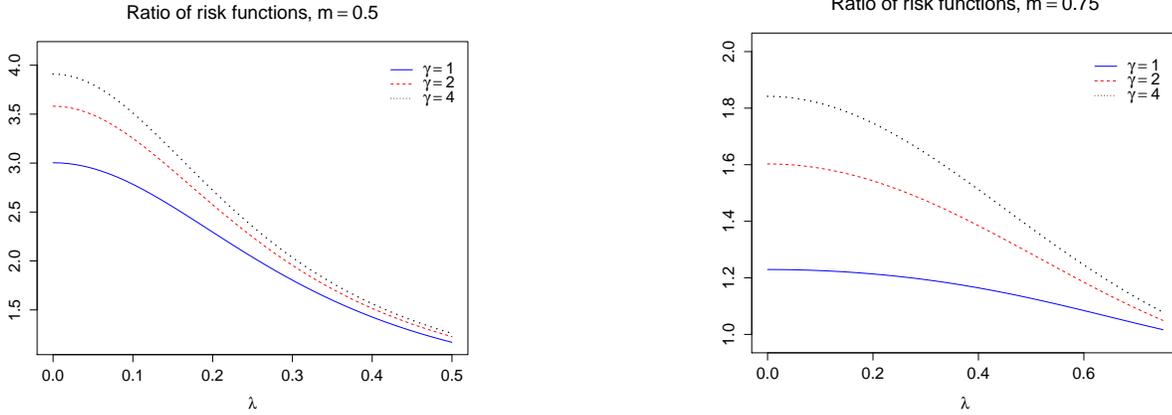
Figure 1: Ratio $\frac{R(\mu,\delta_{mle})}{R(\mu,\delta_{BU,\gamma})}$ of risks for reflected normal loss L_γ

*with $\lambda_0 \in [0, m]$, such as the boundary valued $msgn(x)$ are admissible Bayes estimators. For $\alpha > 1$, the situation is reversed and the best equivariant estimator is a shrinker as either condition (**II**) or (**III**), that follow Corollary 4.2, is satisfied. This is confirmed by the calculation*

$$\frac{\partial}{\partial y} \log \frac{f(\lambda - y)}{f(\lambda + y)} = 2 \left(\frac{1}{\beta} - \frac{\lambda(\alpha - 1)}{\lambda^2 - y^2} \right) ,$$

*which is less or equal than 0 for all $y \in [0, \lambda]$ whenever $\lambda \leq (\alpha - 1)\beta$ (i.e., condition (**II**)), and otherwise varies from positive to negative as y varies on $[0, \lambda]$ (i.e., condition (**III**)). The conditions of Theorem 4.3 are thus verified and estimators that take values on the boundary $\{-m, m\}$ of the parameter space (with positive probability for all μ), such as δ_{mle}, are inadmissible.*

*Turning to Theorem 4.4, take $m \leq (\alpha - 1)\beta$ so that condition **C2** is verified, observe that f is logconcave (i.e., **C1** holds) and that $(\log f(t))' = \frac{\alpha - 1}{t} - \frac{1}{\beta}$ is convex on $(0, m)$ for all $m > 0$ (i.e., condition **C3** is satisfied). We thus have that the Bayes estimator δ_{BU} dominates δ_{mle} whenever we further have $m \leq c_0$ with*

$$\frac{c_0}{h'(c_0)} = 1 \iff \frac{c_0}{\frac{\alpha-1}{c_0} - \frac{1}{\beta}} = 1 \iff c_0 = \frac{1}{2\beta} \left(\sqrt{1 + 4\beta^2(\alpha - 1)^2} - 1 \right) .$$

Finally, since $c_0 \leq (\alpha - 1)\beta$, the condition $m \leq c_0$ is sufficient for δ_{BU} to dominate δ_{mle} under the corresponding Gamma cdf loss and $\mu \in [-m, m]$.

Example 5.5. *As a follow-up to the duality results of Section 2, we illustrate how our dominance findings lead to predictive density estimation results for $\alpha-$divergence losses with $|\alpha| < 1$. Consider normal model (2.6), the restriction $\mu \in [-m, m]$, and predictive density estimators $\hat{q}_{c^2, \delta_{mle}} \sim N(\delta_{mle}(Y_1), c^2 \sigma_{Y_2}^2)$ of the density of $Y_2 \sim N(\mu, \sigma_{Y_2}^2)$, where $\delta_{mle}(Y_1)$ is the restricted mle of μ and c^2 is a fixed constant. The case $c^2 = 1$ corresponds to the predictive mle density. It follows from Lemma 2.2 and the inadmissibility of $\delta_{mle}(Y_1)$ under reflected normal loss (Example 5.2) that $\hat{q}_{c^2, \delta_{mle}}$ is inadmissible for all $c^2, m > 0$, with predictive density improvements given by $\hat{q}_{c^2, \delta_g} \sim N(\delta_g(Y_1), c^2 \sigma_{Y_2}^2)$ and $\delta_g(Y_1)$ a dominating estimator given in Corollary 4.3 for reflected normal loss L_{γ_0} with $\gamma_0 = (\frac{c^2}{1+\alpha} + \frac{1}{1-\alpha})\sigma_{Y_2}^2$. In particular, if one chooses δ_g as the boundary uniform*
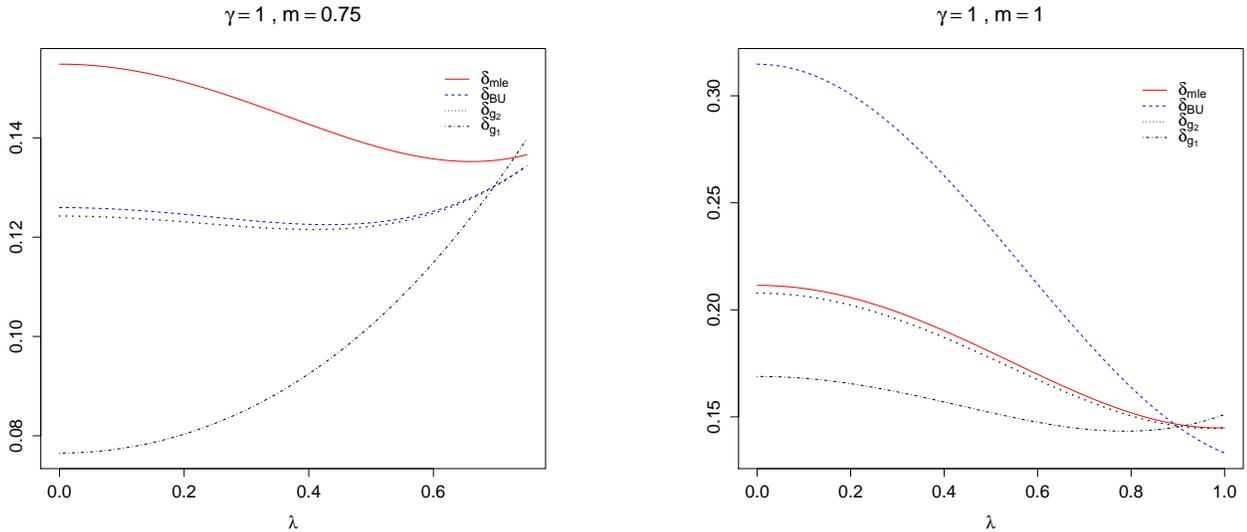
17

Figure 2: Risk comparisons for reflected normal loss L_γ of various estimators: δ_{mle}, $\delta_{BU,\gamma}$, $\delta_{g_1}(x) = m \tanh(mx)$, δ_{g_2} with $g_2 = g_{\mathrm{mle}} \wedge g_{BU,\gamma}$

Bayes point estimator of μ, dominance follows for $m \leq \sqrt{\frac{\gamma_0}{1+\gamma_0}} \sigma_{Y_2}$ (using the result of Example 5.2). Finally, we point out that a similar development for integrated L^2 loss, via Lemma 2.1, holds as well.

6. Concluding remarks

This paper relates to the estimation of a bounded normal mean with a distinctive feature being the analysis provided for a large class of strict bowled-shaped losses, including convex, non-convex or concave choices, L^q and reflected normal losses. Non-convex penalties, as well as various strict bowled-shaped losses, are appealing and relevant to a wide class of statistical problems (e.g., penalized regression, penalized variable selection). The latter arise precisely in predictive density estimation problems as described in Section 2.

In assessing the frequentist risk efficiency of the restricted mle, as well as other estimators taking values on the boundary of the parameter space, our findings clearly highlight the role of the loss, as well as the width of the parameter space. In many cases, which include convex choices of loss like L^q loss, as well as others like reflected normal loss, we establish the inadmissibility of the mle and provide improvements. Namely, we establish using conditional risk techniques conditions for which the Bayes boundary uniform prior δ_{BU} dominates the mle, despite the absence of explicit expressions for δ_{BU} and the risks of the estimators. Further progress relative to the risk performance of other Bayes estimators, such as the fully uniform Bayes estimator, as well as to multivariate versions of the problem, are challenging and pertinent propositions for future work.

Acknowledgments

7. Appendix

Proof of Lemma 2.2. From (2.2), we obtain that the α-divergence loss incurred by the predictive density estimate $\hat{q}_{c,\hat{\mu}} \sim N_d(\hat{\mu}, c^2 \sigma_{Y_2}^2 I_d)$ of the $N_d(\mu, \sigma_{Y_2}^2 I_d)$ density is equal to

$$
\frac{4}{1-\alpha^2} \left(1 - \int_{\mathbb{R}^d} \left(\frac{\phi(\frac{t-\hat{\mu}}{c\sigma_{Y_2}})}{(c\sigma_{Y_2})^d} \right)^{\beta} \left(\frac{\phi(\frac{t-\mu}{\sigma_{Y_2}})}{\sigma_{Y_2}^d} \right)^{1-\beta} dt \right), \tag{7.1}
$$

where we have set $\beta = \frac{1+\alpha}{2}$. With the identity $\phi^k(z) = (2\pi)^{\frac{d}{2}(1-k)} \phi(z\sqrt{k})$, $k > 0$, the above loss reduces to

$$
\frac{1}{\beta(1-\beta)} \left(1 - (2\pi)^{d/2} \frac{\int_{\mathbb{R}^d} \phi(\frac{(t-\hat{\mu})\sqrt{\beta}}{c\sigma_{Y_2}}) \phi(\frac{(t-\mu)\sqrt{1-\beta}}{\sigma_{Y_2}}) dt}{(c^{\beta} \sigma_{Y_2})^d} \right). \tag{7.2}
$$

Finally, using the identity

$$
\int_{\mathbb{R}^d} \phi(\frac{t-\mu_1}{\sigma_1}) \phi(\frac{t-\mu_2}{\sigma_2}) dt = (\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2})^{d/2} \phi(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}), \mu_i \in \mathbb{R}^d, \sigma_i \in \mathbb{R}_+, i = 1, 2, \tag{7.3}
$$

for $\mu_1 = \hat{\mu}$, $\mu_2 = \mu$, $\sigma_1 = \frac{c\sigma_{Y_2}}{\sqrt{\beta}}$, $\sigma_2 = \frac{\sigma_{Y_2}}{\sqrt{1-\beta}}$, the above loss reduces to $1 - b + bL_{\gamma_0}(\mu, \hat{\mu})$ with $b = \frac{c^{d(1-\beta)}}{c^2(1-\beta)+\beta}$ and $\gamma_0 = \frac{c^2(1-\beta)+\beta}{\beta(1-\beta)} \sigma_{Y_2}^2$, thus establishing the result. $\qquad\square$

References

Boisbunon, A. et Maruyama, Y. (2014). Inadmissibility of the best equivariant density in the unknown variance case. *Biometrika*, **101**, 733-740.

Brown, L.D., George, E.I., and Xu, X. (2008). Admissible predictive density estimation. *Annals of Statistics*, **36**, 1156-1170.

Casella, G. & Strawderman, W. E. (1981). Estimating a bounded normal mean. *The Annals of Statistics*, **9**, 870-878.

Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar., **2**, 299-318.

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

Fourdrinier, D., Marchand, É., Righi, A. and Strawderman, W.E. (2011). On improved predictive density estimation with parametric constraints. *Electronic Journal of Statistics*, **5**, 172-191.

George, E. I., Liang, F. and Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Annals of Statistics*, **34**, 78-91.

Ghosh, M., Mergel, V. & Datta, G. S. (2008). Estimation, prediction and the Stein phenomenon under divergence loss. *Journal of Multivariate Analysis*, **99**, 1941-1961.

Iwasa, M. & Moritani, Y. (1997). A note on the admissibility of the maximum likelihood estimator for a bounded normal mean. *Statistics & Probability Letters*, **32**, 99-105.

Kato, K. (2009). Improved prediction for a multivariate normal distribution with unknown mean and variance. *Annals of the Institute of Statistical Mathematics*, **61**, 531-542.

Komaki, F. (2001). A shrinkage predictive distribution for multivariate normal observables. *Biometrika*, **88**, 859-864.

Kubokawa, T., Marchand, É., and Strawderman, W.E. (2015). On predictive density estimation for location families under integrated squared error loss. *Journal of Multivariate Analysis*, **142**, 57-74.

Kubokawa, T., Marchand, É., and Strawderman, W.E. (2016). On Predictive Density Estimation for Location Families under Integrated Absolute Error Loss. *Bernoulli*, to appear.

Kucerovsky, D., Marchand, É., Payandeh, Najafabadi, A. & Strawderman, W. E. (2009). On the bayesianity of maximum likelihood estimators of restricted location parameters under absolute value error loss, *Statistics & Decisions*, **27**, 145-168.

Marchand, É. & Perron, F. (2009). Estimating a bounded parameter for symmetric distributions, *Annals of the Institute of Statistical Mathematics*, **61**, 215-234.

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

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

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.

Maruyama, Y. & Ohnishi, T. (2016). Harmonic Bayesian prediction under α-divergence. (arxiv:1605.05899)

Maruyama, Y. & Strawderman, W.E. (2012). Bayesian predictive densities for linear regression models under $\alpha-$divergence loss: Some results and open problems. *Contemporary Developments in Bayesian analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman*, Institute of Mathematical Statistics Volume Series, **8**, 42-56.

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

Spiring, F.A. (1993). The reflected normal loss function. *Canadian Journal of Statistics*. **21**, 321-330.

Towhidi, M. & Behboodian, J. (2002). Minimax estimation of a bounded parameter under some bounded loss functions *Far East Journal of Theoretical Statistics*, **6**, 39-48.

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