

On shrinkage estimation of a spherically symmetric distribution for balanced loss functions *

Lahoucine Hobbad[†], Éric Marchand[‡] and Idir Ouassou[§]

Abstract

We consider the problem of estimating the mean vector θ of a d -dimensional spherically symmetric distributed X based on balanced loss functions of the forms: **(i)** $\omega\rho(\|\delta - \delta_0\|^2) + (1 - \omega)\rho(\|\delta - \theta\|^2)$ and **(ii)** $\ell(\omega\|\delta - \delta_0\|^2 + (1 - \omega)\|\delta - \theta\|^2)$, where δ_0 is a target estimator, and where ρ and ℓ are increasing and concave functions. For $d \geq 4$ and the target estimator $\delta_0(X) = X$, we provide Baranchik-type estimators that dominate $\delta_0(X) = X$ and are minimax. The findings represent extensions of those of Marchand & Strawderman ([18]) in two directions: **(a)** from scale mixture of normals to the spherical class of distributions with Lebesgue densities and **(b)** from completely monotone to concave ρ' and ℓ' .

1. Introduction

The balanced loss function (BLF) was introduced and formulated by Zellner (1994) in order to reflect two criteria, namely goodness of fit and precision of estimation. For estimating $\gamma(\theta) \in \mathbb{R}^d$ based on $X \sim f_\theta$, consider loss incurred by estimate δ

$$\omega\rho(\|\delta - \delta_0\|^2) + (1 - \omega)\rho(\|\delta - \gamma(\theta)\|^2), \quad (1)$$

where $\delta_0(X)$ is a target estimator of $\gamma(\theta)$, $\rho(\cdot) \geq 0$, and $\omega \in [0, 1)$. Zellner's original BLF corresponds to $\rho(t) = t$ and δ_0 as a least-squares estimator in a regression framework. The above loss encapsulates a more general choice of the target estimator (e.g., [11]) and a more general choice of ρ (e.g., [9], [10]), with the first term measuring proximity of estimate δ to the target δ_0 in comparison to the second term measuring proximity of δ to the estimand $\gamma(\theta)$, weighted by ω and $1 - \omega$ respectively. Decision making under such a loss will necessarily lead to a compromise modulated by the amplitude of ω , with the case $\omega = 0$ corresponding to the so-called unbalanced loss (denoted L_0) where decisions are not influenced by δ_0 . Alternatively, one may view the first term as a penalty for estimate δ diverging from the target δ_0 . An example of this arises with the choice $\delta_0 = 0$ connecting the balanced loss with a ridge regression or Tikhonov regularization framework. For such reasons, balanced loss functions are appealing for decision making and have interested researchers over the years.

*February 25, 2021

[†]École Nationale des Sciences Appliquées-Marrakech, Université Cadi Ayyad, Morocco (e-mail: la.hobbad@gmail.com)

[‡]Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, Canada (e-mail: eric.marchand@usherbrooke.ca)

[§]École Nationale des Sciences Appliquées-Marrakech, Université Cadi Ayyad, Morocco (e-mail: i.ouassou@uca.ac.ma)

A natural and interesting modification of Zellner's original balanced loss function, introduced in [18] is given by:

$$\ell(\omega \|\delta - \delta_0\|^2 + (1 - \omega) \|\delta - \gamma(\theta)\|^2), \quad (2)$$

where δ_0 is a target estimator of $\gamma(\theta) \in \mathbb{R}^d$, $0 \leq \omega < 1$, and $\ell(\cdot) \geq 0$. Such losses possess similar attractive features as those in (1), and are also especially appealing if ℓ is concave or even bounded.

For the original squared error loss balanced loss function case with $\rho(t) = t$ or $\ell(t) = t$, it is known (e.g., [4], [11], [18]) that frequentist risk performance of estimators is directly related to the frequentist risk performance of associated estimators under unbalanced loss. For instance, we have the following.

Lemma 1.1. *(Corollary 1 of [18]) Let $X \sim f_\theta$ and consider the problem of estimating $\gamma(\theta)$. Then, for $\omega \in (0, 1)$, $\delta_0(X) + (1 - \omega) g_1(X)$ dominates $\delta_0(X) + (1 - \omega) g_2(X)$ under loss (1) with $\rho(t) = t$ if and only if $\delta_0(X) + g_1(X)$ dominates $\delta_0(X) + g_2(X)$ under loss $\|\delta - \gamma(\theta)\|^2$.*

Marchand and Strawderman [18] considered the estimation of the mean of a multivariate normal θ or a scale mixture of normal distribution under losses (1) and (2). They provided, for three dimensions or more, for increasing and concave ρ and ℓ which also satisfy a completely monotone property, Baranchik-type estimators of θ which dominate the benchmark $\delta_0(X) = X$. Their findings apply to a vast collections of ρ 's and ℓ 's, and quite generally for scale mixtures of normal distributions, but they do not cover non completely monotone ρ' and ℓ' , as well as other spherically symmetric distributions.

In this paper, we provide extensions with respect to the choices of ρ and ℓ , as well to the class of spherically symmetric densities. More precisely, for $X \in \mathbb{R}^d$, $d \geq 4$, with spherically symmetric Lebesgue density $f(\|x - \theta\|^2)$, we obtain, for estimating θ for losses of types (1) and (2) with target $\delta_0(X) = X$, Baranchik-type estimators ([1]) that dominate the benchmark estimator X . Our results apply to increasing and concave ρ and ℓ and do not require complete monotonicity. The Baranchik-type estimators studied are of the form:

$$\delta_{a,s}(X) = \left(1 - a(1 - \omega) \frac{s(\|X\|^2)}{\|X\|^2}\right) X, \quad (3)$$

with s twice differentiable a.e.,

$$a > 0, 0 \leq s(\cdot) \leq 1, s(\cdot) \neq 0, s'(\cdot) \geq 0, \text{ and } s''(\cdot) \leq 0. \quad (4)$$

Such estimators include simple choices like $s(t) = \frac{t}{t+b}$ with $b \geq 0$, including James-Stein estimators for $b = 0$. It is worthwhile noting that the findings here can be applied to cases where a sample $X^{(1)}, \dots, X^{(n)}$ is drawn from $f(\|x - \theta\|^2)$ and when inference is then based on an estimator (such as \bar{X}) which is a spherically symmetric and translation invariant function of these data (e.g., [3]).

The paper is organized as follows. In Sections 2 and 3, we present dominance results for balanced losses (1) and (2) respectively. Both sections include a subsection of illustrations and remarks. Several technical results are presented in these sections, and others are relegated to an Appendix. Finally, a detailed example is presented in Section 4, with a comparison of frequentist risks and some details worth sharing about the calculations themselves.

2. Risk analysis for loss $\omega\rho(\|\delta - X\|^2) + (1 - \omega)\rho(\|\delta - \theta\|^2)$

2.1. Preliminary results and definitions

We begin with an initial observation. It follows from [13] that $\delta_0(X) = X$ is minimax for estimating θ , whenever X has density $f(\|x - \theta\|^2)$, $\theta \in \mathbb{R}^d$, and loss $\rho(\|\delta - \theta\|^2)$ with $\rho(0) = 0$ and $\rho'(t) > 0$ for $t > 0$, as long as X has finite risk. Furthermore, as put forth in [10], if an estimator $\delta_0(X)$ is minimax for estimating θ under loss $\rho(\|\delta - \theta\|^2)$ then, it is also minimax under balanced loss in (1) with $\gamma(\theta) = \theta$ for all $0 < \omega < 1$. Therefore, the estimators of this section which dominate the benchmark $\delta_0(X) = X$ under balanced loss (1) are minimax in cases of spherically symmetric density models and for the conditions **C1** below in (5) on ρ .

For the function ρ in loss (1), we assume the following throughout

$$\mathbf{C1} : \rho(0) = 0, 0 < \rho'(0) < +\infty, \text{ and } \rho \text{ is concave.} \quad (5)$$

Examples for which ρ satisfies condition **C1**, other than $\rho(t) = t$, include: **(i)** $\rho(t) = 1 - \exp(-\alpha t)$ with $\alpha > 0$, **(ii)** $\rho(t) = \log(1 + t)$, **(iii)** $\rho(t) = (1 + t/\gamma)^\beta$ with $\gamma > 0, \beta \in (0, 1)$, **(iv)** $\rho(t) = r^2 t / (rt + 1)$ with $r > 0$, **(v)** $\rho(t) = \arctan(t)$, and **(vi)** $\rho(t) = \tanh(t)$. Except for **(v)** and **(vi)**, the examples were presented by Marchand & Strawderman in [18] as examples of completely monotone ρ' (i.e., $(-1)^{k+1} \rho^{(k)}(t) \geq 0$ for $t > 0$ and $k \in \mathbb{N}$). We do not require completely monotonicity and condition **C1** is weaker. Another class of loss functions satisfying the above conditions are given by: **(vii)** $\rho(t) = G(t)$ with G a cdf on $(0, \infty)$ with non-increasing density G' . Such a class of loss functions arose recently in a predictive density estimation framework described in [15]. Finally, we point out that condition **C1** implies:

$$\mathbf{C2} : \rho(0) = 0, 0 < \rho'(0) < +\infty, \rho \text{ is concave, and } \rho(t)/t \text{ is non-increasing for } t \in \mathbb{R}_+.$$

The added non-increasing property of $\rho(t)/t$ is thus superfluous, but we will use this on several occasions, as well as the outright concavity of ρ .

We make use throughout of standard definitions and properties of spherically symmetric distributions, as well as shrinkage estimation techniques and properties of superharmonic functions. Some key features are either collected in this subsection or in the Appendix, and further properties and definitions can be found for instance in [8].

A well-known, useful property and characterization of spherically symmetric distributions (e.g., Theorem 4.1 in [8]) is the independence between the radius $R = \|X - \theta\|$ and $\frac{X - \theta}{\|X - \theta\|}$, with the conditional distribution of $X|R = r$ uniformly distributed (noted $U_{r,\theta}$) on the sphere $S_{r,\theta} = \{x \in \mathbb{R}^d \mid \|x - \theta\| = r\}$ of radius r centered at θ . Our findings capitalize on a corresponding conditional on R risk decomposition. Moreover, the distribution of R is independent of θ and its density (called radial) is given by $h(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} f(r^2) \mathbb{I}_{\mathbb{R}_+}(r)$ when X has Lebesgue density $f(\|x - \theta\|^2)$.

Properties of superharmonic functions also play an important role. We recall that a continuous function $g : \mathbb{R}^d \rightarrow \mathbb{R}$ is superharmonic if and only if: for all $t_0 \in \mathbb{R}^d$ and $r > 0$, the average of g over the surface of the sphere S_{r,t_0} is less or equal than $g(t_0)$. For twice-differentiable g , the superharmonicity of g is equivalent to its Laplacian being less or equal to 0, i.e., $\Delta(g) \leq 0$ with $\Delta(g) = \sum_{k=1}^d (\partial^2 g(t) / \partial t_k^2)$. As stated in Lemma 6.2, superharmonicity of g can be used to obtain an

inequality relating the conditional expected value on the ball to the conditional expected value on the sphere. An important fact here is that if g is superharmonic, its average over the ball (“volume”) is greater than its average over the sphere (“surface area”). We conclude this subsection with a pivotal inequality, which appeared in [18] for scale mixtures of normals, but which holds here more generally over the class of spherically symmetric densities.

Lemma 2.1. *Suppose that X is spherically symmetric distributed about θ with density $f(\|x - \theta\|^2)$, that $\mathbb{E}_\theta(\|X\|^{-2}) < \infty$, and that the function ρ satisfies **C1**. For $d \geq 4$, and for $s : \mathbb{R}^d \rightarrow [0, 1]$ a twice-differentiable, non-decreasing and concave function then*

$$E_\theta \left[\rho \left(\frac{s(\|X\|^2)}{\|X\|^2} \right) \right] \leq \rho'(0) E_\theta \left[\frac{s(\|X\|^2)}{\|X\|^2} \right] \leq \rho'(0) E_\theta \left[\frac{s(\|Y\|^2)}{\|Y\|^2} \right], \quad (6)$$

where $Y \sim f^*(\|y - \theta\|^2)$ with $f^*(t) = \frac{\rho'(t)f(t)}{K}$ and $K = \mathbb{E}_0(\rho'(\|X\|^2))$.

Proof. The first inequality of (6) follows, on taking expectations, from the concave function inequality

$$\rho(t) \leq \rho(0) + \rho'(0)t = \rho'(0)t$$

given that ρ is concave with $\rho(0) = 0$. Denoting by h and h^* the densities of $\|X - \theta\|$ and $\|Y - \theta\|$ respectively, \mathbb{E}^* the expectation with respect to h^* , and making use of the equality of the conditional distributions $X \mid \|X - \theta\| = r$ and $Y \mid \|Y - \theta\| = r$, we have

$$\begin{aligned} E_\theta \left[\frac{s(\|X\|^2)}{\|X\|^2} \right] &= \int_{\mathbb{R}_+} E_\theta \left(\frac{s(\|X\|^2)}{\|X\|^2} \mid \|X - \theta\| = r \right) h(r) dr \\ &= \int_{\mathbb{R}_+} E_\theta \left(\frac{s(\|Y\|^2)}{\|Y\|^2} \mid \|Y - \theta\| = r \right) \frac{K}{\rho'(r^2)} h^*(r) dr \\ &= E^* \left[E_\theta \left(\frac{s(\|Y\|^2)}{\|Y\|^2} \mid \|Y - \theta\| = R \right) \frac{K}{\rho'(R^2)} \right]. \end{aligned} \quad (7)$$

Since the function $s(\|t\|^2)/\|t\|^2$ is superharmonic for $d \geq 4$ as shown in Lemma 6.1, and since for such superharmonic functions the sphere mean is non-increasing in the radius (e.g., [8], Theorem A.4, page 304), it follows that $E_\theta \left(\frac{s(\|Y\|^2)}{\|Y\|^2} \mid \|Y - \theta\| = r \right)$ is non-increasing in r . Finally, along with the concavity of ρ which implies that $K/\rho'(r^2)$ is non-decreasing in r , the covariance inequality applied to (7) implies that

$$E_\theta \left[\frac{s(\|X\|^2)}{\|X\|^2} \right] \leq E_R^* \left[\frac{K}{\rho'(R^2)} \right] E^* \left[E_\theta \left(\frac{s(\|Y\|^2)}{\|Y\|^2} \mid \|Y - \theta\| = R \right) \right] = E_\theta \left[\frac{s(\|Y\|^2)}{\|Y\|^2} \right],$$

due to the fact that $E_R^* [K/\rho'(R^2)] = 1$. □

2.2. Dominance finding

We are now ready for the main dominance finding of this section.

Theorem 2.1. Suppose that X is spherically symmetric distributed about θ with density $f(\|x - \theta\|^2)$, that both $\mathbb{E}_0(\|X\|^2)$ and $\mathbb{E}_0(\|X\|^{-2})$ are finite, and that the function ρ satisfies **C1**. For $d \geq 4$ and for estimating θ under loss (1) with $\delta_0(X) = X$, the estimator $\delta_{a,s}(X)$ in (3) satisfying conditions (4) dominates δ_0 provided:

$$0 < a < \frac{2K(d-2)/d}{\{\omega\rho'(0) + K(1-\omega)\} E_0^*(\|Y\|^{-2})}, \quad (8)$$

where E_0^* is the expectation taken with respect to $Y \sim f^*(\|y - \theta\|^2)$ with $f^*(t) = \frac{\rho'(t)f(t)}{K}$ and $K = \mathbb{E}_0(\rho'(\|X\|^2))$. An equivalent condition for the above dominance condition is:

$$0 < a < \frac{2K^2(d-2)/d}{\{\omega\rho'(0) + K(1-\omega)\} E_0\left(\frac{\rho'(\|X\|^2)}{\|X\|^2}\right)}. \quad (9)$$

Proof. Set $\delta_g(X) = X + (1-\omega)g(X)$ and consider the difference in risks

$$\Delta\mathcal{R}(\theta) = \mathcal{R}(\theta, \delta_g) - \mathcal{R}(\theta, \delta_0), \quad \theta \in \mathbb{R}^d.$$

We have

$$\begin{aligned} \Delta\mathcal{R}(\theta) &= E_\theta \left\{ \omega\rho(\|(1-\omega)g(X)\|^2) + (1-\omega)\rho(\|X - \theta + (1-\omega)g(X)\|^2) \right. \\ &\quad \left. - (1-\omega)\rho(\|X - \theta\|^2) \right\} \\ &\leq E_\theta \left[\omega\rho'(0)\|(1-\omega)g(X)\|^2 + (1-\omega) \left\{ \rho(\|X - \theta\|^2) + 2(1-\omega)g(X)^\top(X - \theta) \right. \right. \\ &\quad \left. \left. + (1-\omega)^2\|g(X)\|^2 - \rho(\|X - \theta\|^2) \right\} \right] \\ &\leq E_\theta \left[\omega\rho'(0)(1-\omega)^2\|g(X)\|^2 + (1-\omega)K \times \frac{\rho'(\|X - \theta\|^2)}{K} \left\{ 2(1-\omega)g(X)^\top(X - \theta) \right. \right. \\ &\quad \left. \left. + (1-\omega)^2\|g(X)\|^2 \right\} \right] \\ &= (1-\omega)^2 \left[E_\theta \left[\omega\rho'(0)\|g(X)\|^2 \right] + KE_\theta \left\{ 2g(Y)^\top(Y - \theta) + (1-\omega)\|g(Y)\|^2 \right\} \right], \quad (10) \end{aligned}$$

where the two inequalities follow from the concave function inequality: $\rho(t_1) - \rho(t_2) \leq \rho'(t_1)(t_1 - t_2)$, with the condition $\rho(0) = 0$. Now, for the Baranchik $g(x) = -a(s(\|x\|^2)/\|x\|^2)x$ and $a > 0$, it is easy to show that

$$\text{div}(g(x)) = -2as'(\|x\|^2) - a(d-2)\frac{s(\|x\|^2)}{\|x\|^2} \leq -a(d-2)\frac{s(\|x\|^2)}{\|x\|^2}.$$

From (10), using the condition $0 \leq s(\cdot) \leq 1$ as well as Lemma 2.1 and Lemma 6.3, we obtain

$$\begin{aligned} \frac{\Delta\mathcal{R}(\theta)}{(1-\omega)^2} &\leq E_\theta \left[\omega\rho'(0)\frac{a^2s(\|X\|^2)}{\|X\|^2} \right] + KE_\theta \left[2g(Y)^\top(Y - \theta) \right] + KE_\theta \left[(1-\omega)\frac{a^2s(\|Y\|^2)}{\|Y\|^2} \right] \\ &\leq E_\theta \left[\omega\rho'(0)\frac{a^2s(\|Y\|^2)}{\|Y\|^2} \right] - 2Ka\frac{d-2}{d} \int_{\mathbb{R}_+} r^2 \int_{B_{r,\theta}} \frac{s(\|y\|^2)}{\|y\|^2} dV_{r,\theta}(y) h^*(r) dr \quad (11) \\ &\quad + KE_\theta \left[(1-\omega)\frac{a^2s(\|Y\|^2)}{\|Y\|^2} \right], \end{aligned}$$

$$, \quad (12)$$

h^* being the radial density of $\|Y - \theta\|$. As $t \rightarrow s(\|t\|^2)/\|t\|^2$ is superharmonic, then according to Lemma 6.4 expression (11) is bounded above by

$$\begin{aligned} & E_\theta \left[\omega \rho'(0) \frac{a^2 s(\|Y\|^2)}{\|Y\|^2} \right] - 2Ka \frac{d-2}{d} \int_{\mathbb{R}_+} r^2 \int_{S_{r,\theta}} \frac{s(\|y\|^2)}{\|y\|^2} dV_{r,\theta}(y) h^*(r) dr \\ & + K E_\theta \left[(1-\omega) \frac{a^2 s(\|Y\|^2)}{\|Y\|^2} \right], \\ & = a E \left\{ E_\theta \left(R^2 \frac{s(\|Y\|^2)}{\|Y\|^2} \mid \|Y - \theta\| = R \right) \left(a \frac{\omega \rho'(0) + K(1-\omega)}{R^2} - 2K \frac{d-2}{d} \right) \right\}. \end{aligned}$$

With $E_\theta \left[R^2 \frac{s(\|Y\|^2)}{\|Y\|^2} \mid \|Y - \theta\| = R \right]$ non-decreasing in R by Lemma 6.5, and $\{a[\omega \rho'(0) + K(1-\omega)]/R^2\}$ non-increasing in R , an application of the covariance inequality leads to

$$\begin{aligned} \frac{\Delta \mathcal{R}(\theta)}{(1-\omega)^2} & \leq a E \left\{ E_\theta \left[R^2 \frac{s(\|Y\|^2)}{\|Y\|^2} \mid \|Y - \theta\| = R \right] \right\} \times E \left[\left(a \frac{\omega \rho'(0) + K(1-\omega)}{R^2} - 2K \frac{d-2}{d} \right) \right] \\ & = a E_\theta \left(\|Y - \theta\|^2 \frac{s(\|Y\|^2)}{\|Y\|^2} \right) \times \left[a \{ \omega \rho'(0) + K(1-\omega) \} E_\theta \left(\frac{1}{\|Y - \theta\|^2} \right) - 2K \frac{d-2}{d} \right], \end{aligned}$$

and the result follows. \square

2.3. Examples and discussion

The dominance finding of Section 2.2 is applicable to many choices of ρ and the underlying spherically symmetric f . The results guarantee the existence of a class of Baranchik estimators which are minimax. For $d \geq 4$, subject to finiteness of risk or the minimaxity of X , the results apply quite generally with respect to the choice of the Baranchik estimator, the weight ω , the choice of ρ , and the underlying model density f . Our findings do expand on existing results, namely the findings of Marchand and Strawderman [18], with both a wider class of loss functions and underlying models. However, they do not duplicate their results when applicable as our conditions turn out to be stronger, i.e., the cut-off points are smaller (see Section 2.3.2). Although the focus of this paper is on the balanced case with $\omega > 0$, Theorem 2.1 applies for the case $\omega = 0$ nevertheless with the cut-off point in (8) simplifying to $2(d-2) \{dE_0^*(\|Y\|^{-2})\}^{-1}$, and matching the one obtained by Brandwein and Strawderman ([3], Theorem 2.1).

Without dwelling too much on the wealth of applicable situations or calculations of the cut-off points, we elaborate here a little bit with some illustrations and remarks.

2.3.1. Choices of ρ and determination of the cut-off points

Cut-off point (9) of Theorem 2.1 is explicitly represented and is conveniently expressible in terms of the expectations $K = \mathbb{E}(\rho'(W))$ and $\mathbb{E}(\rho'(W)/W)$, with respect to the density of $W = \|X\|^2$ when $\theta = 0$ and given by

$$\frac{\pi^{d/2}}{\Gamma(d/2)} w^{d/2-1} f(w) \mathbb{I}_{\mathbb{R}_+}(w). \quad (13)$$

With $\rho(t) \rightarrow \frac{\rho(t)}{\rho'(0)}$; which tells us that we can set $\rho'(0) = 1$ without loss of generality; and by setting

$$I_k = \int_0^\infty w^{k-1} \rho'(w) f(w) dw ,$$

cut-off point (9) can simply be expressed as

$$a_0 = \frac{2(\frac{d-2}{d}) I_{d/2}^2}{\{\omega \Gamma(d/2)/\pi^{d/2} + (1-\omega)I_{d/2}\} I_{d/2-1}} . \quad (14)$$

Marchand & Strawderman provide a similar cut-off point to (9) applicable to completely monotone ρ' and a scale mixture of normals model density, while the findings of Section 2.2 apply to non-completely monotone choices of ρ' such as those referred at the beginning of Section 2.1. As an example, consider the $\rho(t) = 2\Phi(t) - 1$ with Φ the standard normal cdf (this is the cdf of a truncated standard normal distribution). Then, Theorem 2.1 applies and can one simply take $\rho'(t)/\rho'(0) = e^{-t^2/2}$ for evaluating (9) or (14) numerically. Even in the normal case with $X \sim N_d(\theta, I_d)$ and $W \sim \chi_d^2(0)$, the result is new and yields a_0 evaluated with

$$I_k = \int_0^\infty w^{k-1} e^{-(w+w^2)/2} dw , \text{ for } k = d/2, d/2 - 1.$$

2.3.2. Scale mixtures of normal distributions case

Theorem 2.1 applies to scale mixtures of normals admitting representation

$$X|V \sim N_d(\theta, VI_d), V \sim G , \quad (15)$$

where G is a c.d.f. for V , and including many familiar distributions such as normal, logistic, Laplace, exponential power, Student, etc. For choices of ρ satisfying condition **C1** such that also ρ' is completely monotone, the cut-off point given in [18] is greater than (9) by a factor of $d/(d-2)$. So, Theorem 2.1's result is weaker. However, as reviewed above in subsection 2.3.1, Theorem 2.1 applies for more general ρ . Also, the results of Section 2.2 apply for more general models and we pursue with such examples.

2.3.3. Uniform distribution on a ball

Theorem 2.1 applies for the case of a uniform distribution on a ball centered at θ of radius m : $B_{m,\theta} = \{x \in \mathbb{R}^d : \|x - \theta\| \leq m\}$, with $d \geq 4$ and densities

$$f(\|x - \theta\|^2) = \frac{\Gamma(d/2 + 1)}{m^d \pi^{d/2}} \mathbb{I}_{(0,m)}(\|x - \theta\|) . \quad (16)$$

In such cases, $\frac{W}{m^2} = \frac{\|X-\theta\|^2}{m^2}$ has Beta density $\frac{d}{2} t^{d/2-1} \mathbb{I}_{(0,1)}(t)$, and cut-off point (9) is readily available by either analytical or numerical evaluations of $I_k = \frac{\Gamma(d/2+1)}{m^d \pi^{d/2}} \int_0^{m^2} w^{k-1} \rho'(w) dw$ for $k = d/2$ and $d/2 - 1$. As an illustration, the choice $\rho(t) = \log(1+t)$, $d = 4$ yields:

$$a_0 = \frac{\{m^2 - \log(1+m^2)\}^2}{\{\omega m^4/2 + (1-\omega)(m^2 - \log(1+m^2))\} \log(1+m^2)}$$

For the unbalanced squared error loss case with $\omega = 0$ and $\rho(t) = t$, early shrinkage estimation analysis was provided in [2] for such uniform distributions and their mixtures, the latter playing a key role in modelling since their ensemble spans the entire class of unimodal spherically symmetric distributions.

2.3.4. Kotz type distribution

Theorem 2.1 applies to Kotz model densities $f(\|x - \theta\|^2)$ with

$$f(t) = c_d t^{s\nu - \frac{d}{2}} e^{-rt^s} \mathbb{I}_{\mathbb{R}_+}(t), \quad (17)$$

with $r, s, \nu > 0$ and $c_d = \frac{s\Gamma(\frac{d}{2})}{\pi^{\frac{d}{2}}\Gamma(\nu)} r^\nu$.¹ The distribution, originally introduced in [14] for $s = 1$ has generated much interest over the years (e.g. [17]), namely for its flexibility in representing non-unimodal densities (for $2s\nu > d$), non-scale mixture of normal densities (any choice except $0 < s \leq 1$ and $2s\nu = d$), as well as including the normal case ($s = 1, \nu = d/2$), scale mixtures of normal ($0 < s \leq 1$ and $2s\nu = d$), and exponential power densities ($2s\nu = d$). Moreover, it is simple to verify that the distribution of $W^s = \|X - \theta\|^{2s}$ is distributed as $\text{Gamma}(\nu, r)$, which facilitates the expression of cut-off point (9). Resulting integrals will not be available in general in closed forms. One exception arises for $s = 1$ and reflected normal $\rho(t) = 1 - e^{-\alpha t}$. For $\nu > 1$ (we need this for finiteness of risk), an evaluation of (9) or (14) yields:

$$a_0 = \frac{2(\nu - 1)(1 - 2/d)/(r + \alpha)}{\omega(1 + \alpha/r)^\nu + (1 - \omega)}.$$

A numerical illustration with an underlying Kotz density is provided and commented upon in Section 4.3.

3. Risk analysis for loss $\ell(\omega\|\delta - X\|^2 + (1 - \omega)\|\delta - \theta\|^2)$

3.1. Dominance finding

For spherically symmetric model densities, we evaluate the frequentist risk performance of an estimator $\delta(X)$ of θ under the balanced loss (2) which incorporates the target estimator $\delta_0(X) = X$. For the function ℓ , we work with the following conditions throughout this section:

$$\mathbf{C3} : \ell(0) = 0, \ell'(\cdot) > 0, \text{ and } \ell \text{ is twice-differentiable and concave.} \quad (18)$$

As for the choice of ρ in the previous section, the completely monotone requirement on ℓ' in [18] is relaxed here. Examples of ρ that satisfy condition **C1**, including (i) to (vii) following (5), provide examples of ℓ that satisfy above condition **C3**. But, we do not require finiteness of $\ell'(0)$ so that many other losses, such as $L^{q/2}$ losses with (viii) $\ell(t) = t^q, 0 < q < 1$, satisfy **C3** as well. Another interesting choice comes from [15] as: (ix) $\ell(t) = 2Q(\sqrt{t}) - 1$ with Q a c.d.f. on \mathbb{R} with even and unimodal density Q' . This arises as an intrinsic loss in measuring the distance between estimate

¹A more frequent parametrization in the literature has $N = s\gamma + 1 - d/2$.

$\hat{\mu}$ and location parameter μ through the L^1 discrepancy between model $q(\|y - \mu\|^2)$ and plug-in densities $q(\|y - \hat{\mu}\|^2)$ (i.e., $\int_{\mathbb{R}^d} |q(\|y - \mu\|^2) - q(\|y - \hat{\mu}\|^2)| dy$), with q unimodal and Q the common c.d.f. of the univariate components Y_1, \dots, Y_d with joint density $q(\|y - \mu\|^2)$.

We proceed with a preparatory lemma which exploits the concavity of ℓ , and which relates the difference in losses (2) between estimates $\delta_g(X) = \delta_0(X) + (1 - \omega)g(X)$ and $\delta_0(X)$, to the balanced squared-error loss difference. Referring to the loss in (2) as $L_{\omega, \ell}(\theta, \delta)$, we now define:

$$\Delta_{\omega, \ell}(\theta, \delta) = L_{\omega, \ell}(\theta, \delta) - L_{\omega, \ell}(\theta, \delta_0).$$

Lemma 3.1. (Lemma 6 in [18]) *Suppose that X is spherically symmetric distributed about θ with density $f(\|x - \theta\|^2)$. For the problem of estimating θ under (2) with twice-differentiable, increasing, and concave ℓ , we have*

$$\Delta_{\omega, \ell}(\theta, \delta) \leq (1 - \omega)^2 \ell' \{ (1 - \omega) \|\delta_0 - \theta\|^2 \} \Delta_{0, \ell}(\theta, \delta).$$

We now have the following.

Theorem 3.1. *Suppose that X is spherically symmetric distributed about θ with density $f(\|x - \theta\|^2)$ and that the function ℓ satisfies **C3**. For $d \geq 4$, the estimator $\delta_{a, s}(X)$ in (3) with condition (4) dominates $\delta_0(X) = X$ under loss (2) with $\delta_0(X) = X$ provided $0 < a < 2(d - 2) / (d E_0^* [\|Z\|^{-2}])$, where*

$$Z \sim f^*(\|z - \theta\|^2) = \frac{\ell' \{ (1 - \omega) \|z - \theta\|^2 \} f(\|z - \theta\|^2)}{\int_{\mathbb{R}^d} \ell' \{ (1 - \omega) \|z - \theta\|^2 \} f(\|z - \theta\|^2) dz},$$

and provided both $\mathbb{E}_0(\|X\|^2)$ and $\mathbb{E}_0(\|X\|^{-2})$ are finite.

Proof. We show that the difference in risks between $\delta_0(X) = X$ and $\delta_{a, s}(X)$ is non-negative under the given conditions. Let $K_Z = \int_{\mathbb{R}^d} \ell' \{ (1 - \omega) \|z - \theta\|^2 \} f(\|z - \theta\|^2) dz$, and h^* be the radial density for $\|Z - \theta\|$. Then, we have setting $g(X) = -a \frac{s(\|X\|^2)}{\|X\|^2} X$:

$$\begin{aligned} \Delta \mathcal{R}(\theta) &= E_\theta [\Delta_{\theta, \ell}(\theta, \delta_{a, s})] \\ &\leq E_\theta [(1 - \omega)^2 \ell' \{ (1 - \omega) \|X - \theta\|^2 \} \Delta_{0, \ell}(\theta, \delta_{a, s})] \\ &= (1 - \omega)^2 E_\theta [\ell' \{ (1 - \omega) \|X - \theta\|^2 \} (\|\delta_{a, s}(X) - \theta\|^2 - \|X - \theta\|^2)] \\ &= (1 - \omega)^2 E_\theta [\ell' \{ (1 - \omega) \|X - \theta\|^2 \} (\|X + g(X) - \theta\|^2 - \|X - \theta\|^2)] \quad (19) \\ &= K_Z (1 - \omega)^2 \{ E_\theta [\|g(Z)\|^2 + 2(Z - \theta)^\top g(Z)] \} \\ &\leq (1 - \omega)^2 \left\{ E_\theta \left[\frac{as(\|Z\|^2)}{\|Z\|^2} \right] - 2 \frac{d - 2}{d} \int_{\mathbb{R}_+} r^2 \int_{B_{r, \theta}} \frac{s(\|z\|^2)}{\|z\|^2} dV_{r, \theta}(z) h^*(r) dr \right\} \\ &= a K_Z (1 - \omega)^2 E \left\{ \left(a \frac{1}{R^2} - 2 \frac{d - 2}{d} \right) E_\theta \left[R^2 \frac{s(\|Z\|^2)}{\|Z\|^2} \mid \|Z - \theta\| = R \right] \right\}, \quad (20) \end{aligned}$$

where the first inequality follows from Lemma 3.1, and the second inequality follows from a calculation using Lemma 6.3 along with the defining inequalities $0 \leq s(\cdot) \leq 1$ and $s'(\cdot) \geq 0$. Finally, an application of Lemma 6.5 and the covariance inequality imply that

$$\Delta \mathcal{R}(\theta) \leq a K_Z (1 - \omega)^3 \left(a E_0 [\|Z\|^{-2}] - 2 \frac{d - 2}{d} \right) E \left\{ R^2 E_\theta \left[\frac{s(\|Z\|^2)}{\|Z\|^2} \mid \|Z - \theta\| = R \right] \right\},$$

establishing the result. \square

Remark 3.1. *A more direct proof of the above result is achieved by applying Theorem 2.1 of [3] immediately after (19). We have given here details following (19) thus providing a more self-contained proof. We also point out, as expected, that Theorems 3.1 and 2.1 match for $\ell(t) = \rho(t) = t$, and that Theorem 3.1 reduces to Theorem 2.1 of [3] for the unbalanced case $\omega = 0$, as was the case for Theorem 2.1 in Section 2.*

3.2. Examples and discussion

As in Section 2, the dominance result above applies for many choices of ℓ and spherically symmetric densities f guaranteeing the existence of Baranchik estimators that dominate the benchmark $\delta_0(X) = X$ under balanced loss $L_{\omega, \ell}$. As in Section 2, Theorem 3.1's cut-off point on a for scale mixture of normals is less than that of [18] (see subsection 3.2.1), but our dominance finding here is more generally applicable to all spherically symmetric densities subject to risk finiteness and to non completely monotone ℓ' .

Theorem 3.1's cut-off point for dominance is representable in terms of $W = \|X - \theta\|^2$ as

$$a_0 = \frac{2(d-2)}{d} \frac{\mathbb{E}[\ell'\{(1-\omega)W\}]}{\mathbb{E}\left[\frac{\ell'\{(1-\omega)W\}}{W}\right]}. \quad (21)$$

An interesting case arises for $\ell(t) = t^q$ with $0 < q < 1$, with the above yielding

$$a_0 = a_0(q) = \frac{2(d-2)}{d} \frac{\{\mathbb{E}(W^{q-1})\}}{\{\mathbb{E}(W^{q-2})\}}, \quad (22)$$

independently of ω as observed in [18]. As expanded on in [18], such a simple form of a_0 gives rise to simultaneous dominance with respect to both choices of ℓ and density f . For instance, with $a_0(q)$ increasing for $q \in (0, 1)$; which may be justified by writing $a_0(q) = \mathbb{E}_q(Z)$ with Z having density to $z^{q-2} g_W(z)$ on \mathbb{R}_+ and observing that the family of such densities with parameter $q \in (0, 1]$ has an increasing monotone likelihood ration in Z ; we have that $a_0 \geq a_0(q_0)$ for all $q \in [q_0, 1]$, so that Theorem 3.1's Baranchik estimators $\delta_{a,s}(X)$ dominate X for $a \leq a_0(q_0)$ simultaneously for all losses in (2) with fixed $w \in [0, 1)$ and such that $\ell(t) = t^q$ with $q \in [q_0, 1]$. We refer to [18] for such further examples with varying f or ℓ .

3.2.1. Scale mixtures of normal distributions

Theorem 3.1 applies for scale mixture of normals as in (15) for both completely monotone and non-completely monotone ℓ' . In the former case, Theorem 3.1's cut-off point, or equivalently (21), on a for the Baranchik estimator $\delta_{a,s}(X)$ to dominate X is smaller by a factor of $\frac{d-2}{d}$ that the one obtained in [18], and it thus weaker. For non-completely monotone ℓ' , applications of Theorem 2.1 are novel however. As an illustration, with $\ell(t) = 2\Phi(t) - 1$, the cut-off point reduces to:

$$a_0 = \frac{2(d-2)}{d} \frac{\mathbb{E}\left[e^{-\frac{(1-\omega)^2 W^2}{2}}\right]}{\mathbb{E}\left[W^{-1} e^{-\frac{(1-\omega)^2 W^2}{2}}\right]},$$

for scale mixtures of normals, as well as for all spherically symmetric densities subject to the risk finiteness conditions.

3.2.2. Example: Uniform distribution on a ball

Theorem 3.1 applies for the uniform distribution on the ball $B_{m,\theta}$ (i.e., $X \sim U(B_{m,\theta})$) with density in (16). Hence, the cut-off point a_0 can be evaluated using the density $\frac{d}{2m^d} w^{d/2-1} \mathbb{I}_{(0,m^2)}(w)$ for W . As an illustration for $\ell(t) = t^q$ with $0 < q < 1$, we obtain from (22):

$$a_0 = a_0(m) = \frac{2(d-2)m^2}{d} \frac{2q+d-4}{2q+d-2}.$$

Observe that the dominance finding has implications even in cases where m is unknown, but bounded below by a positive value \underline{m} , yielding that the dominance result $a \leq a_0(\underline{m})$, applicable for $X \sim U(B_{m,\theta})$, is robust to any discrepancy such that $X \sim U(B_{m,\theta})$ with $m \geq \underline{m}$. Moreover, the argument goes over when m is random yielding a mixture of uniform distributions on balls as long as $\mathbb{P}(M \geq \underline{m}) = 1$, M being the mixing parameter.

3.2.3. Kotz type distribution

As in Section 2, Theorem 3.1 applies to Kotz densities $f(\|x - \theta\|^2)$ with f given in (17). Cut-off point a_0 in (21) is conveniently represented with $W = \|X - \theta\|^{2s} \sim \text{Gamma}(\nu, r)$. For the case $\ell(t) = t^q$, a direct calculation of (22) yields for $\frac{q-2}{s} + \nu > 0$

$$a_0 = \frac{2(d-2)}{d} \frac{\Gamma(\frac{q-1}{s} + \nu)}{\Gamma(\frac{q-2}{s} + \nu)} r^{-1/s}.$$

4. Numerical illustration

4.1. Introduction

We provide here frequentist risk evaluations illustrative of Theorem 2.1, as well as provide details on the calculations themselves which we believe are useful for replication. The results of Section 3 can be illustrated in a similar fashion (but also see [18]).

4.2. Calculation of risk under loss (1)

The numerical evaluation of expectations with respect to a spherically symmetric density $f(\|x - \theta\|^2)$ may be expressed as a d dimensional integral, but can be reduced to a two dimensional integral in our case of balanced risk function evaluations, and for estimators of θ that are equivariant under orthogonal transformations. In this regard, the following lemma given in [12] will turn out to be most useful.

Lemma 4.1. *Let $X \sim f(\|x - \theta\|^2)$ with $X, \theta \in \mathbb{R}^d$, $\theta \neq 0$, $d > 1$. Let $\lambda = \|\theta\|$, $W = \|X\|^2$, and $T = \frac{\theta^\top X}{\lambda \|X\|}$. Then, the joint density of (T, W) is given by:*

$$\psi(t, w) = \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} w^{\frac{d}{2}-1} (1-t^2)^{\frac{d-3}{2}} f(w + \lambda^2 - 2\lambda t w^{1/2}), \quad (23)$$

for $t \in (-1, 1)$ and $w > 0$.

Now, for equivariant estimators of θ , which are of the form $h(\|X\|^2)X$ (e.g., [6]), the loss in (1) becomes

$$\begin{aligned} \omega \rho (\|h(\|X\|^2)X - X\|^2) &+ (1 - \omega) \rho (\|h(\|X\|^2)X - \theta\|^2) \\ &= \omega \rho ((h(W) - 1)^2 W) + (1 - \omega) \rho (\lambda^2 + h^2(W)W - 2\lambda T W^{1/2} h(W)) . \end{aligned}$$

Therefore, the associated frequentist risk for $\theta \in \mathbb{R}^d$ reduces to a function of λ and can be evaluated as

$$\omega \mathbb{E} \{ \rho ((h(W) - 1)^2 W) \} + (1 - \omega) \mathbb{E} \{ \rho (\lambda^2 + h^2(W)W - 2\lambda T W^{1/2} h(W)) \} , \quad (24)$$

with the expectations taken with respect to density (23). This illustrates the dimensional reduction to two dimensions and can be used to numerically evaluate the frequentist risk of Baranchik type estimators with h of the form $h(w) = (1 - b \frac{s(w)}{w})$.

4.3. Illustration of Theorem 2.1

The dominance finding of Theorem 2.1 applies to Baranchik estimators $(1 - b \frac{s(\|X\|^2)}{\|X\|^2})X$ and expression (24) can serve as a numerical illustration or comparison with inputs: **(i)** $\omega \in [0, 1)$ and $d \geq 4$; **(ii)** a model density f ; **(iii)** a choice of ρ satisfying condition **C2**, taking $\rho'(0) = 1$ without loss of generality; **(iv)** a choice of $s(\cdot)$ satisfying (4); and **(v)** a choice of $b \leq a_0(1 - \omega)$ with a_0 given in (14).

We pursue by setting: **(i)** $\omega = 1/2$ and $d = 6$; **(ii)** a Kotz density as in (17) with $r = s = 1$, $\nu = 4$ yielding the model density $X \sim f(\|x - \theta\|^2)$ with $f(t) = \frac{te^{-t}}{3\pi^3} \mathbb{I}_{(0, \infty)}(t)$; **(iii)** $\rho(t) = \log(1 + t)$; **(iv)** a choice of $s(\cdot)$ satisfying (4); and **(v)** a choice of $b \leq a_0/2$ with expression (14) yielding

$$\frac{a_0}{2} = \frac{1}{2} \frac{\frac{8}{3} I_3^2}{(\frac{2}{\pi^3} + I_3) I_2} \approx 0.595 ,$$

with $I_k = \frac{1}{3\pi^3} \int_0^\infty \frac{w^{k-1} e^{-w}}{1+w} dw$, $I_3 \approx 0.01509$ and $I_2 \approx 0.00641$.

Figure 1 compares the risk of: **(i)** $\delta_0(X) = X$, the Baranchik estimators **(ii)** $s(t) = \frac{t}{1+t}$, $b = 1/2$ (red in Figure 1), and **(iii)** $s(t) = \frac{t}{1+t}$, $b = 1$ (green in Figure 1), and the James-Stein estimator **(iv)** with $s(t) = 1$, and $b = 0.5$ (blue in Figure 1). The benchmark estimator δ_0 is minimax (see first paragraph of Section 2) with constant risk $R(\theta, \delta_0) = \frac{1}{2} \mathbb{E}_0(\log(1 + \|X\|^2)) \approx 0.76606$, obtainable with $\|X - \theta\|^2 \sim \text{Gamma}(4, 1)$ (see Subsection 2.3.4). Both the first Baranchik and James-Stein estimators are minimax with sufficiently small cut-off point $b = 0.5 \leq a_0/2$ as a consequence of Theorem 2.1. The maximal gains are attained at $\theta = 0$ and are about 8.58% and 10.43%. The second Baranchik estimator has lower minimum risk, but its cut-off point is too large to satisfy the dominance condition of Theorem 2.1. Furthermore, the numerical evidence suggests that it is (barely) not minimax.

5. Concluding remarks

For spherically symmetric distributed X with densities $f(\|x - \theta\|^2)$, we have provided frequentist risk improvements on the benchmark estimator $\delta_0(X)$ associated with balanced loss functions (1)

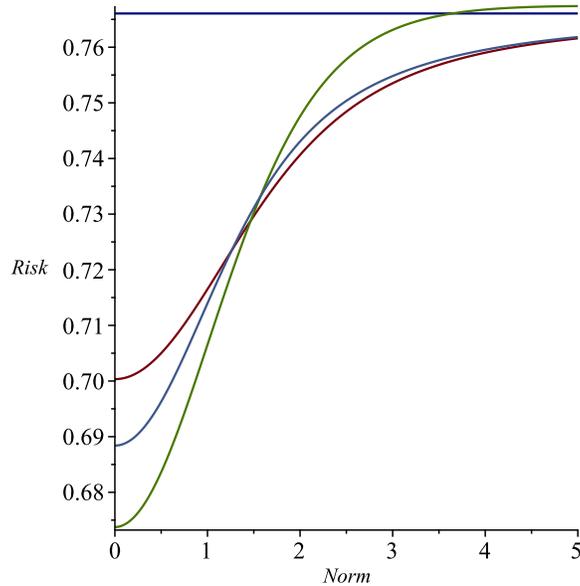


Figure 1: Frequentist risks as functions of $\|\theta\|$ of $\delta_0(X) = X$ (horizontal blue) and $\delta_{b,c}(X) = \left(1 - \frac{b}{\|X\|^2+c}\right) X$ for $(b, c) = (0.5, 1.0)$ (red), $(b, c) = (1.0, 1.0)$ (green), and $(b, c) = (0.5, 0.0)$ (blue); $X \sim \frac{1}{3\pi^3} \|x - \theta\|^2 e^{-\|x-\theta\|^2}$ (a Kotz density), balanced loss (1) with $\rho(t) = \log(1 + t)$, $\omega = 1/2$.

and (2). The findings, which apply to a wide class of Baranchik-type estimators, are unified with respect to the model f , with respect to the choices of ρ in (1) and ℓ in (2), as well as the weight ω in these losses. The findings extend earlier results of Marchand and Strawderman ([18]) which apply to scale mixtures of normals f and completely monotone ρ' or ℓ' , while the results here apply to the whole class of spherically symmetric densities subject to risk finiteness, as well as to choices of ρ and ℓ that are monotone increasing and concave.

The findings testify to the ubiquitous nature of the effectiveness of shrinkage procedures as seen by the applicability of the dominance findings for a large class of densities f and choices of ρ, ℓ , and ω .

Acknowledgements

Éric Marchand's research is supported in part by the Natural Sciences and Engineering Research Council of Canada.

6. Appendix

Here are some technical results used in the paper.

Lemma 6.1. *If $s : \mathbb{R}^d \rightarrow \mathbb{R}_+$ is a twice differentiable and concave function, then the function $x \rightarrow s(\|x\|^2)/\|x\|^2$ is superharmonic for $d \geq 4$.*

Proof. Observe that $s(\cdot)$ must be non-decreasing. A calculation of the Laplacian yields

$$\Delta [s(\|x\|^2)/\|x\|^2] = \frac{2}{\|x\|^4} [2\|x\|^4 s''(\|x\|^2) + (d-4) (\|x\|^2 s'(\|x\|^2) - s(\|x\|^2))] .$$

Since $s(0) \geq 0$ and s is concave, we have $s(0) + ts'(t) \leq s(t)$ and the result follows. \square

The next result, referred to as a covariance inequality, is quite well known (e.g., Lemma 6.6, page 370 in [16]).

Lemma 6.2. *Let Y be a random variable, and g and h be functions for which $\mathbb{E}[g(Y)]$, $\mathbb{E}[h(Y)]$, and $\mathbb{E}[g(Y)h(Y)]$ exist.*

- (a) *If one of the functions g and h is non-increasing and the other is non-decreasing, then $\mathbb{E}[g(Y)h(Y)] \leq \mathbb{E}[g(Y)]\mathbb{E}[h(Y)]$;*
- (b) *If both g and h are either non-decreasing or non-increasing, then $\mathbb{E}[g(Y)h(Y)] \geq \mathbb{E}[g(Y)]\mathbb{E}[h(Y)]$.*

The following is taken from [8] and provides a useful decomposition for an expectation in terms of the radial distribution and uniform measures on balls.

Lemma 6.3. *Let X have a spherically symmetric density about θ , and $g(X)$ be a weakly differentiable function such that $\mathbb{E}_\theta [|(X - \theta)^\top g(X)|] < \infty$. Then, assuming expectations exist, we have*

$$\mathbb{E}_\theta [(X - \theta)^\top g(X)] = \frac{1}{d} \mathbb{E} \left[R^2 \int_{B_{R,\theta}} \operatorname{div}(g(x)) dV_{R,\theta}(x) \right]$$

where \mathbb{E} denotes the expectation with respect to the radial distribution, and where $V_{r,\theta}(\cdot)$ is the uniform distribution on $B_{R,\theta}$, the ball of radius r centered at θ .

Proof. See [8], Lemma 5.4, page 235.

Lemma 6.4. *([5], page 54) Let g be a superharmonic function, Z_1 have a uniform distribution on the sphere $S_{r,\theta}$ centered at θ with radius r , and Z_2 have a uniform distribution on the ball $B_{r,\theta}$ centered at θ with radius r , then $\mathbb{E}g(Z_1) \leq \mathbb{E}g(Z_2)$.*

Lemma 6.5. *Let W be spherically symmetric Lebesgue density $g(\|w - \theta\|^2)$, $w, \theta \in \mathbb{R}^d$. Let $\beta : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a non-decreasing function. Then, $\mathbb{E} \left(R^2 \frac{\beta(\|W\|^2)}{\|W\|^2} \mid \|W - \theta\| = R \right)$ is a non-decreasing function of $R > 0$.*

Proof. See [2], pages 394-395 within the proof of their Theorem 3.3.1. \square

References

- [1] BARANCHIK, A. J. A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics* 41 (1970) 22–42.
- [2] BRANDWEIN, A.C. & STRAWDERMAN, W.E. Minimax estimation of location parameters for spherically symmetric unimodal distributions under quadratic loss. *Annals of Statistics* 6 (1978) 377–416.
- [3] BRANDWEIN, A.C. & STRAWDERMAN, W.E. Minimax estimation of location parameters for spherically symmetric distributions with concave loss. *Annals of Statistics* 8 (1980) 279–284.
- [4] DEY, D., GHOSH, M. & STRAWDERMAN, W.E. On estimation with balanced loss functions. *Statistics & Probability Letters* 45 (1999) 97-101.

- [5] DU PLESSIS, N. *An introduction to potential theory*. Oliver and Boyd, Edinburgh, 1970.
- [6] EATON, M.L. Robust tests for spherical symmetry. Group Invariance Applications in Statistics. Regional Conference Series in Probability and Statistics, 1. Institute of Mathematical Statistics, Beachwood, Ohio.
- [7] FELLER, W. *An introduction to probability theory and its applications, volume II. Second edition*. Wiley & Sons, New York, 1966.
- [8] FOURDRINIER, D., STRAWDERMAN, W.E. & WELLS, M. T. Shrinkage estimation. Springer series in statistics. Springer. New York, Dordrecht, Heidelberg, London, 2018.
- [9] JAFARI JOZANI, M., LEBLANC, A. & MARCHAND. On continuous distribution functions, minimax and best invariant estimators, and integrated balanced loss functions. *Canadian Journal of Statistics* 42 (2014) 470–486.
- [10] JAFARI JOZANI, M., MARCHAND, É. & PARSIAN, A. Bayesian and robust Bayesian analysis under a general class of balanced loss functions. *Statistical Papers* 53 (2012) 51–60.
- [11] JAFARI JOZANI, M., MARCHAND, É. & PARSIAN, A. On estimation with weighted balanced-type loss function. *Statistics & Probability Letters* 76 (2006) 773–780.
- [12] KARIYA, T., EATON, M.L. Robust tests for spherical symmetry. *Annals of Statistics* 1 (1977) 206–215.
- [13] KIEFER, J. Invariance, minimax sequential estimation, and continuous time processes. *Annals of Mathematical Statistics* 28 (1957) 573–601.
- [14] KOTZ, S. Multivariate distributions at a cross-road. In: Patil, G. P., Kotz, S. and Ord, J. K. (Eds.), *Statistical Distributions in Scientific Work*, Vol. 1. D. Reidel Publishing Company, Dordrecht: pp. 247–270, 1975.
- [15] KUBOKAWA, T., MARCHAND, É. & STRAWDERMAN, W.E. On predictive density estimation for location families under integrated absolute error loss. *Bernoulli* 23 (2017) 3197–3212.
- [16] LEHMANN, E.L. & CASELLA, G. *Theory of point estimation*, 2nd edition. Springer, New York, 1998.
- [17] NADARAJAH, S. The Kotz-type distribution with applications *Statistics* 37 (1993) 341–358.
- [18] MARCHAND, É. & STRAWDERMAN, W. E. On shrinkage estimation for balanced loss functions. *Journal of Multivariate Analysis* 175 (2020) 104558.
- [19] ZELLNER, A. Bayesian and non-Bayesian estimation using balanced loss functions. *Statistical Decision Theory and Methods V*, (J.O. Berger and S.S. Gupta Eds). New York: Springer-Verlag, 337–390, 1994.