Bernoulli **23**(4B), 2017, 3197–3212
DOI: 10.3150/16-BEJ842

On predictive density estimation for location families under integrated absolute error loss

TATSUYA KUBOKAWA[1,*], ÉRIC MARCHAND[1,**] and WILLIAM E. STRAWDERMAN[2]

[1]*Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan and Département de mathématiques, Université de Sherbrooke, Sherbrooke Qc, J1K 2R1, Canada. E-mail: [*]tatsuya@e.u-tokyo.ac.jp; [**]eric.marchand@usherbrooke.ca*
[2]*Department of Statistics and Biostatistics, Rutgers University, 501 Hill Center, Busch Campus, Piscataway, NJ, 08855, USA. E-mail: straw@stat.rutgers.edu*

This paper is concerned with estimating a predictive density under integrated absolute error (L_1) loss. Based on a spherically symmetric observable $X \sim p_X(\|x-\mu\|^2)$, $x, \mu \in \mathbb{R}^d$, we seek to estimate the (unimodal) density of $Y \sim q_Y(\|y-\mu\|^2)$, $y \in \mathbb{R}^d$. We focus on the benchmark (and maximum likelihood for unimodal p) plug-in density estimator $q_Y(\|y-X\|^2)$ and, for $d \geq 4$, we establish its inadmissibility, as well as provide plug-in density improvements, as measured by the frequentist risk taken with respect to X. Sharper results are obtained for the subclass of scale mixtures of normal distributions which include the normal case. The findings rely on the duality between the predictive density estimation problem with a point estimation problem of estimating μ under a loss which is a concave function of $\|\hat{\mu}-\mu\|^2$, Stein estimation results and techniques applicable to such losses, and further properties specific to scale mixtures of normal distributions. Finally, (i) we address univariate implications for cases where there exist parametric restrictions on μ, and (ii) we show quite generally for logconcave q_Y that improvements on the benchmark mle can always be found among the scale expanded predictive densities $\frac{1}{c} q_Y(\frac{(y-x)^2}{c^2})$, with $c-1$ positive but not too large.

Keywords: concave loss; dominance; frequentist risk; inadmissibility; L_1 loss; multivariate normal; plug-in; predictive density; restricted parameter space; scale mixture of normals; Stein estimation

1. Introduction

The developments of this paper relate to spherically symmetric and independently distributed

$$X \mid \mu \sim p_X\big(\|x-\mu\|^2\big), \qquad Y \mid \mu \sim q_Y\big(\|y-\mu\|^2\big); \qquad x, y, \mu \in \mathbb{R}^d; \tag{1}$$

with p and q known Lebesgue densities, not necessarily equal, and μ is unknown. The set-up in (1) includes the normal model with

$$X \mid \mu \sim N_d\big(\mu, \sigma_X^2 I_d\big), \qquad Y \mid \mu \sim N_d\big(\mu, \sigma_Y^2 I_d\big), \tag{2}$$

as well as scale mixtures of normal distributions (Definition 2.1).

For predictive analysis purposes, researchers are interested in specifying a predictive density $\hat{q}(y; x)$, based on observation x, as an estimate of the density $q_Y(\|y-\mu\|^2)$. In turn, such a density may play a surrogate role for generating either future or missing values of Y. Our interest and motivation here lies in assessing the efficiency of such predictive densities with integrated absolute error loss (hereafter referred to as L_1) and corresponding frequentist risk, with

$$L_1(\mu, \hat{q}) = \int_{\mathbb{R}^d} \bigl|q_Y\bigl(\|y-\mu\|^2\bigr) - \hat{q}(y)\bigr|\,dy, \tag{3}$$

$$R(\mu, \hat{q}) = \int_{\mathbb{R}^d} \biggl\{\int_{\mathbb{R}^d} \bigl|q_Y\bigl(\|y-\mu\|^2\bigr) - \hat{q}(y;x)\bigr|\,dy\biggr\} p_X\bigl(\|x-\mu\|^2\bigr)\,dx. \tag{4}$$

L_1 loss is a quite recognizable and an appealing distance. It has played a prominent role in assessing the efficiency of density estimators over the years, both in a nonparametric and a parametric setting (e.g., [9,10]; among many others). It is also intrinsic in the sense that, for one-to-one functions $g : \mathbb{R}^d \to \mathbb{R}^d$ with inverse jacobian J, the L_1 distance between the densities of $g(Y)$ and $g(Y')$ is independent of g; i.e., $\int_{\mathbb{R}^d} |q_Y(\|g^{-1}(y) - \mu\|^2)|J| - \hat{q}(g^{-1}(y))|J||\,dy$ is independent of g.

The tractability of the L_1 distance and its associated risk is another matter, and analytical results relative to the performance of predictive density estimators are lacking. In terms of a benchmark procedure, the evaluation of the minimum risk equivariant predictive density estimator, which is the Bayes procedure associated with the constant prior measure $\pi_0(\mu) = 1$ is quite challenging. And more generally, we have been unable to provide the form of Bayesian predictive density estimators for a given prior π.[1] With such a paucity of results and bearings, we focus on the performance of the plug-in predictive density estimator $q_Y(\|y - X\|^2)$, $y \in \mathbb{R}^d$, which is also for unimodal p_X the maximum likelihood predictive density estimator (mle) of $q_Y(\|y-\mu\|^2)$, $y, \mu \in \mathbb{R}^d$. Our main objective and common theme is to provide dominating predictive density estimators of $q_Y(\|y - X\|^2)$, $y \in \mathbb{R}^d$. This is achieved in Section 2 for $d \geq 4$, quite generally with respect to model (1) by substituting the estimator X by a more efficient $\hat{\mu}(X)$ as a plug-in estimator. In Section 3, we obtain improvements in the univariate case by variance expansion of the predictive density estimator. Here are further details.

In Section 2, we focus on the performance of plug-in predictive density estimators $q_Y(\|y - \hat{\mu}(X)\|^2)$, $y \in \mathbb{R}^d$, with $\hat{\mu}(X)$ an estimator of μ. For $d \geq 4$, we provide dominating estimators of the plug-in $q_Y(\|y - X\|^2)$, $y \in \mathbb{R}^d$. This is achieved by capitalizing on an explicit representation for the L_1 distance (Lemma 2.1) between two densities of the same spherically symmetric family, which implies that our predictive density estimation problem for plug-in estimators is dual to a point estimation problem under a loss which is a concave function of $\|\hat{\mu} - \mu\|^2$ (Corollary 2.1). Using Stein estimation results and techniques applicable to such concave losses (e.g., [5–7]), we establish the inadmissibility of plug-in densities $q_Y(\|y - X\|^2)$ for $d \geq 4$ and obtain dominating predictive density estimators. In Section 2.2, we provide further sharper developments for scale mixtures of normals p_X and q_Y, which include of course the normal case. The dual loss functions that intervene are of interest on their own and our findings also represent contributions from the point estimation perspective. Namely, the dual loss for the normal model turns out to

[1]Two-point priors are an exception with the Bayes estimator given by $q_Y(\|y - \mathrm{med}(\mu \mid x)\|^2)$, with $\mathrm{med}(\mu \mid x)$ being the posterior median.

be the interesting loss $L(\mu, \hat{\mu}) = 4\Phi(\frac{\|\hat{\mu}-\mu\|}{2}) - 2$, where Φ is the standard normal c.d.f. In Section 2.3, for univariate situations where μ is either restricted to an interval (a, b), or restricted to a half-interval (a, ∞), we make use of existing results for strict-bowled shaped losses and our duality results to show that the plug-in density estimator $q_Y(|y - \hat{\mu}_{\pi_U}(X)|^2)$ dominates the plug-in $q_Y(|y - X|^2)$ for log-concave density p_X, where $\hat{\mu}_{\pi_U}(X)$ is the Bayes point estimator of μ associated with a uniform prior on the restricted parameter space and the given dual loss. For Kullback–Leibler loss and normal models as in (2), similar inadmissibility results applicable to the MRE predictive density estimator as well as plug-in predictive density estimators, as well as connections between predictive density estimation and Stein estimation have been obtained by [8,13,15], and [12], among others. Various findings for integrated squared error loss and spherically symmetric distributions are given in [16].

In Section 3, we focus on the univariate case and scale expansions of the form $\hat{q}_c(y; x) = \frac{1}{c} q_Y(\frac{|y-x|^2}{c^2})$, $y \in \mathbb{R}$, with $c > 1$. We show the plug-in density $q_Y(|y - X|^2)$, $y \in \mathbb{R}$, is dominated by a subclass of such scale expansions $\hat{q}_c$, with $c - 1$ positive but not too large, as long as q is logconcave. This applies to the normal case, as well as models like Logistic and Laplace, among others. This is paradoxical in the sense that the variance associated with the plug-in density $q_Y(|y - X|^2)$ matches the variance of the true density $q_Y(|y - \mu|^2)$, but that improvements can be nevertheless found among the $\hat{q}_c$'s with $c > 1$. Such a result goes back to [1] who showed directly, for Kullback–Leibler loss and univariate normal models, that the plug-in density $\hat{q}_1 \sim N(X, \sigma_Y^2)$ is dominated by the scale expansion (and MRE predictive density estimator) $\hat{q}_{mre} \sim N(X, \sigma_Y^2 + \sigma_X^2)$. This type of phenomenon was generalized, for multivariate normal models, plug-in density estimators $q_Y(\|y - \hat{\mu}(X)\|^2)$ and Kullback–Leibler loss by [12], and addressed recently for integrated squared error loss by [16]. In the former Kullback–Leibler case, the authors showed that, universally for any (non-degenerate) estimator $\hat{\mu}(X)$, any dimension d, any restricted or not parameter space, dominating predictive density estimators $\frac{1}{c^d} q_Y(\|\frac{y-\hat{\mu}(X)}{c}\|^2)$ of $q_Y(\|y - \hat{\mu}(X)\|^2)$ can always be found among choices $c > 1$. In the latter case, the authors provide results for $\hat{\mu}(X) = aX$ with $0 < a \le 1$ with similar scale expansion improvements always available.

2. L_1 loss and plug-in estimators

2.1. An identity for L_1 distance and general dominance results of plug-in predictive density estimators

We begin with a useful L_1 distance identity.

Lemma 2.1. *Let $Y = (Y_1, \dots, Y_d)'$ be a spherically symmetric distributed random vector with unimodal, Lebesgue density $f_\mu(y) = q_Y(\|y - \mu\|^2)$; $y \in \mathbb{R}^d$. Then for any $\mu_1, \mu_2 \in \mathbb{R}^d$, the L_1 distance between f_{μ_1} and f_{μ_2} is given by*

$$\rho_{L_1} = \int_{\mathbb{R}^d} \left| q_Y(\|y - \mu_1\|^2) - q_Y(\|y - \mu_2\|^2) \right| dy = 4F\left(\frac{\|\mu_1 - \mu_2\|}{2}\right) - 2, \tag{5}$$

where $F(t) = P_0(Y_1 \le t)$, $t \in \mathbb{R}$, *is the univariate cumulative distribution function of* Y_1 *when* $\mu_1 = 0$.

Proof. We have $q_Y(\|y-\mu_1\|^2) \ge q_Y(\|y-\mu_2\|^2) \Leftrightarrow \|y-\mu_1\|^2 \le \|y-\mu_2\|^2 \Leftrightarrow L(y) \le 0$, where $L(y) = (\mu_2-\mu_1)'y + \frac{\|\mu_1\|^2 - \|\mu_2\|^2}{2}$. Setting $A = \{y \in \mathbb{R}^d : L(y) \le 0\}$, we obtain splitting the integration on A and its complement A^c

$$\begin{aligned}\rho_{L_1} &= P_{\mu_1}(Y \in A) - P_{\mu_2}(Y \in A) + P_{\mu_2}\big(Y \in A^c\big) - P_{\mu_1}\big(Y \in A^c\big) \\ &= 2\big\{P_{\mu_1}\big(L(Y) \le 0\big) + P_{\mu_2}\big(L(Y) \ge 0\big) - 1\big\}.\end{aligned} \tag{6}$$

Observe that $L(Y)$ is a linear function of the spherically symmetric distributed Y. For such linear functions, we have (e.g., [20])

$$\frac{(l'Y+k) - (l'\mu+k)}{\|l\|} \sim F,$$

for all $l \in \mathbb{R}^d - \{0\}, k \in \mathbb{R}$. We thus obtain $P_{\mu_1}(L(Y) \le 0) = F(-(L(\mu_1))) = F(\frac{\|\mu_1-\mu_2\|}{2})$. Similarly, we obtain $P_{\mu_2}(L(Y) \ge 0) = F(\frac{\|\mu_1-\mu_2\|}{2})$, and the desired expression for ρ_{L_1} follows from (6). □

Remark 2.1. For the multivariate normal case, identity (5) was given and derived in a different manner by [9]. An existing reference for the general case seems likely to us, but we could not find such a reference. Observe that the distance ρ_{L_1} is always a concave function of $\|\mu_1-\mu_2\|$ on $(0,\infty)$ since F' is unimodal, and also of $\|\mu_1-\mu_2\|^2$ given that F is increasing.

Remark 2.2. The L_1 distance formula above also provides an explicit form for the much studied overlap coefficient (e.g., [23]) OVL between two spherically symmetric densities. The latter is defined for densities g_1 and g_2 as

$$OVL(g_1, g_2) = \int_{\mathbb{R}^d} \min\big(g_1(y), g_2(y)\big)\, dy, \tag{7}$$

and is related to the L_1 distance through the identity $OVL(g_1,g_2) = 1 - \frac{1}{2}\rho_{L_1}(g_1,g_2)$ given that $2\min(g_1(y), g_2(y)) = g_1(y) + g_2(y) - |g_1(y) - g_2(y)|$ for all y.

Corollary 2.1. *For estimating a unimodal spherically symmetric Lebesgue density* $q_Y(\|y-\mu\|^2)$, $y \in \mathbb{R}^d$, *under* L_1 *loss and based on* $X \sim p_X(\|x-\mu\|^2)$, *the frequentist risk of the plug-in density estimator* $q_Y(\|y-\hat{\mu}(X)\|^2)$ *is equal to the frequentist risk of the point estimator* $\hat{\mu}(X)$ *of* μ *under loss* $4F(\frac{\|\hat{\mu}-\mu\|}{2}) - 2$, *with* F *being the common marginal c.d.f. associated with* q_Y. *Consequently,* $q_Y(\|y-\hat{\mu}_1(X)\|^2)$ *dominates* $q_Y(\|y-\hat{\mu}_2(X)\|^2)$ *iff* $\hat{\mu}_1(X)$ *dominates* $\hat{\mu}_2(X)$ *under loss* $2F(\frac{\|\hat{\mu}-\mu\|}{2}) - 1$.

Proof. This is a direct consequence of Lemma 2.1. □

Since the dual problem described above involves loss functions $l(\|d-\mu\|^2)$ with $l(t) = 2F(\frac{\sqrt{t}}{2}) - 1$ being concave (see Remark 2.1), we consider using Stein estimation techniques and results for such concave losses [5–7], along with Corollary 2.1, to obtain dominating estimators of the plug-in density $q_Y(\|y-X\|^2)$, $y \in \mathbb{R}^d$, which we now proceed to do, elaborate on, and illustrate. For what follows, we denote f as the density of $\|X-\mu\|$ under p_X and we recall that $f(t) = \frac{2\pi^{d/2}}{\Gamma(d/2)} t^{d-1} p_x(t^2)$ (e.g., [20]). Here is an adaptation of Theorem 2.1 of [6] applicable to Baranchik type estimators [3], and followed by related inferences for improving on plug-in density estimators under L_1 loss.

Theorem 2.1 (Brandwein and Strawderman [6]). *Let X have a spherically symmetric distribution with density $p_X(\|x-\mu\|^2)$, $x \in \mathbb{R}^d$, with respect to σ-finite measure ν. For $d \geq 4$ and for estimating $\mu \in \mathbb{R}^d$ under loss $l(\|\hat{\mu}-\mu\|^2)$ with l non-decreasing and concave on $\mathbb{R}_+$, estimators $\hat{\mu}_{a,r(\cdot)}(X) = (1 - a\frac{r(X'X)}{X'X})X$ dominate X, and are thus minimax, provided*:

(i) $0 \leq r(\cdot) \leq 1$ *and* $r(\cdot) \neq 0$;
(ii) $r(t)$ *is non-decreasing for* $t > 0$;
(iii) $r(t)/t$ *is non-increasing for* $t > 0$;
(iv) $0 < E_{p_X} l'(\|X-\mu\|^2) < \infty$;
(v) $0 < a \leq \frac{2(d-2)}{d} \frac{1}{E_h(R^{-2})}$, *where the expectation is taken with respect to the density $h(s)$ on $\mathbb{R}_+$ proportional to* $l'(s^2) f(s) = \frac{2\pi^{d/2}}{\Gamma(d/2)} l'(s^2) s^{d-1} p_x(s^2)$.

This following result follows from Corollary 2.1 and Theorem 2.1.

Corollary 2.2. *For estimating a unimodal spherically symmetric Lebesgue density $q_Y(\|y-\mu\|^2)$, $y, \mu \in \mathbb{R}^d$ and $d \geq 4$, under integrated L_1 loss and based on $X \sim p_X(\|x-\mu\|^2)$, a plug-in Baranchik density estimator $q_Y(\|y-\hat{\mu}_{a,r(\cdot)}(X)\|^2)$, with $\hat{\mu}_{a,r(\cdot)}(X) = (1 - a\frac{r(X'X)}{X'X})X$, dominates the plug-in $q_Y(\|y-X\|^2)$ provided conditions* (i), (ii), *and* (iii) *of Theorem 2.1 are satisfied as well as*:

(iv′) $0 < E_{p_X}(\frac{q_Y(\|X-\mu\|^2/16)}{\|X-\mu\|}) < \infty$;
(v′) $0 < a \leq \frac{2(d-2)}{d} \frac{\int_{(0,\infty)} u^{\frac{d-3}{2}} p_X(u) F'(\frac{u}{4})\, d\nu(u)}{\int_{(0,\infty)} u^{\frac{d-5}{2}} p_X(u) F'(\frac{u}{4})\, d\nu(u)}$.

Proof. This follows from Corollary 2.1 and Theorem 2.1 with $l(u) = 2F(\frac{\sqrt{u}}{2}) - 1$ and $l'(u) = \frac{F'(\frac{\sqrt{u}}{2})}{2\sqrt{u}}$, as well as the change of variables $u = s^2$. □

Remark 2.3. In our set-up, the model density q_Y determines the loss l via Lemma 2.1 and is thus taken to be unimodal and Lebesgue. On the other hand, there no restrictions on p_X other than risk-finiteness for the estimators $\hat{\mu}_{a,r(\cdot)}(X)$. Condition (iv′) is weak. For instance, it is satisfied when both the densities q_Y and p_X are bounded. The upper bound for the multiplier a of the estimator $\hat{\mu}_{a,r(\cdot)}(X)$ in (v′) depends on both q_Y and p_X.

Here is an evaluation for the particular case when both p_X and q_Y are normal densities.

Example 2.1 (Normal case). For the normal case (2) with $q_Y(u) = (2\pi\sigma_Y^2)^{-d/2} e^{-u/2\sigma_Y^2}$, $F'(t) = (2\pi\sigma_Y^2)^{-1/2} e^{-t^2/2\sigma_Y^2}$, and $p_X(u) = (2\pi\sigma_X^2)^{-d/2} e^{-u/2\sigma_X^2}$, Corollary 2.2 applies with (iv′) satisfied and (v′) specializing to

$$0 < a \le \frac{2(d-2)}{d} \frac{\int_{0,\infty} u^{\frac{d-3}{2}} e^{-u/2\sigma_X^2} e^{-u/8\sigma_Y^2}\, du}{\int_{0,\infty} u^{\frac{d-5}{2}} e^{-u/2\sigma_X^2} e^{-u/8\sigma_Y^2}\, du} = \frac{(d-2)(d-3)}{d} \frac{8\sigma_X^2\sigma_Y^2}{\sigma_X^2 + 4\sigma_Y^2}. \tag{8}$$

We point out that a simultaneous dominance result is available for a family of p_X models by taking the infimum with respect to p_X on the rhs of (v′). For the normal case, if we have for instance, $X \sim N_d(\mu, \sigma_X^2 I_d)$ with σ_X^2 unknown, but known to bounded below by $a_X > 0$, then simultaneous dominance occurs for all such p_X's by taking $0 < a \le \frac{(d-2)(d-3)}{d} \frac{8a_X\sigma_Y^2}{a_X + 4\sigma_Y^2}$.

2.2. Improvements for scale mixture of normals

Distributions in (1) include the subclass of scale mixture of normals, with examples given by the multivariate Cauchy, Student, Logistic, Laplace, Generalized Hyperbolic and Exponential Power distributions, among others (e.g., [2]).

Definition 2.1. *Model (1) is referred to as a scale mixture of normals model whenever*

$$p_X(t) = \int_{\mathbb{R}_+} (2\pi v)^{-d/2} e^{-\frac{t}{2v}}\, dG(v), \qquad q_Y(t) = \int_{\mathbb{R}_+} (2\pi w)^{-d/2} e^{-\frac{t}{2w}}\, dH(w), \tag{9}$$

for $t \in \mathbb{R}^d$ and $W \sim G$, $V \sim H$ are independently distributed mixing random variables on $\mathbb{R}_+$, for which we further assume that $E(V^{-d/2})$ and $E(W^{-d/2})$ are finite. We denote such models as $X - \mu \sim SN_d(G)$ and $Y - \mu \sim SN_d(H)$.

Further developments for scale mixtures of normals are provided in this section and lead to wider classes of dominating estimators than those given by Corollary 2.2. We revisit this latter corollary for situations in (1) where

$$X - \mu \sim SN_d(G), \qquad Y - \mu \sim SN_d(H). \tag{10}$$

We define Z as a random variable, F_Z as its c.d.f., and τ as a bivariate c.d.f. such that

$$Z =^d \frac{4Z_1Z_2}{Z_1 + 4Z_2}, \qquad \text{with } (Z_1, Z_2) \sim d\tau(z_1, z_2) \propto \frac{z_2^{(d-1)/2}}{(z_1 + 4z_2)^{d/2}}\, dG(z_1)\, dH(z_2). \tag{11}$$

Theorem 2.2. *Let $X \sim p_X(\|x-\mu\|^2)$ and $Y \sim q_Y(\|y-\mu\|^2)$, $x, y, \mu \in \mathbb{R}^d$, be scale mixtures of normals as in (10) and consider estimating $q_Y(\|y-\mu\|^2)$ under L_1 loss based on X.*

(a) *For $d > 1$,*[2] *the plug-in density estimator $q_Y(\|y - \hat{\mu}(X)\|^2)$ dominates the plug-in density $q_Y(\|y - X\|^2)$ provided $\hat{\mu}(X')$ dominates X' under loss $\|\hat{\mu} - \mu\|^2$ and for $X' \sim p^*(\|x - \mu\|^2)$, with*

$$p^*\big(\|s\|^2\big) = \frac{K}{\|s\|} \int_{(0,\infty)} (2\pi z)^{-d/2} e^{-\frac{\|s\|^2}{2z}} \, dF_Z(z), \qquad s \in \mathbb{R}^d, \tag{12}$$

where K is a normalization constant.

(b) *In particular, a plug-in Baranchik density estimator $q_Y(\|y - \hat{\mu}_{a,r(\cdot)}(X)\|^2)$, with $\hat{\mu}_{a,r(\cdot)}(X) = (1 - a\frac{r(X'X)}{X'X})X$, dominates the plug-in density $q_Y(\|y - X\|^2)$ provided conditions* (i), (ii), *and* (iii) *of Theorem 2.1 are satisfied, $d \geq 4$, the expectations $E(Z^{-1/2})$ and $E(Z^{-3/2})$ are finite, and $0 < a \leq 2(d-3)\frac{E(Z^{-1/2})}{E(Z^{-3/2})}$.*

Proof. (a) We apply Corollary 2.1. We thus seek conditions for which the difference in risks $\Delta(\mu, \hat{\mu}) = E_\mu[l(\|\hat{\mu}(X) - \mu\|^2) - l(\|X - \mu\|^2)]$ is less than 0, where $l(\|\hat{\mu} - \mu\|^2) = 2F(\frac{\|\hat{\mu}-\mu\|}{2}) - 1$. We apply the inequality $l(s) - l(t) < l'(t)(s - t)$ for strictly concave l and $s \neq t$, which implies for the difference in losses that

$$l\big(\big\|\hat{\mu}(x) - \mu\big\|^2\big) - l\big(\|x - \mu\|^2\big) < l'\big(\|x - \mu\|^2\big)\big(\big\|\hat{\mu}(x) - \mu\big\|^2 - \|x - \mu\|^2\big), \tag{13}$$

for all $x, \mu \in \mathbb{R}^d$ such that $x \neq \hat{\mu}(x)$. Observe that

$$\begin{aligned} l'\big(\|x - \mu\|^2\big) &= \frac{1}{2\|x - \mu\|} F'\left(\frac{\|x - \mu\|}{2}\right) \\ &= \frac{1}{2\|x - \mu\|} \int_0^\infty (2\pi w)^{-1/2} e^{-\frac{\|x-\mu\|^2}{8w}} \, dH(w), \end{aligned} \tag{14}$$

since the marginal distributions associated with a scale mixture of normals as in (10) are themselves univariate scale mixtures of normals with the same mixing distribution.[3] Now, using (13) and (14), it follows that

$$\begin{aligned} \Delta(\mu, \hat{\mu}) &< E_\mu^X\left[\frac{(\|\hat{\mu}(X) - \mu\|^2 - \|X - \mu\|^2)}{2\|X - \mu\|} \int_0^\infty (2\pi w)^{-1/2} e^{-\frac{\|X-\mu\|^2}{8w}} \, dH(w)\right] \\ &= \int_{\mathbb{R}^d} \frac{(\|\hat{\mu}(x) - \mu\|^2 - \|x - \mu\|^2)}{2\|x - \mu\|} \int_0^\infty \int_0^\infty \frac{(2\pi)^{-\frac{d+1}{2}}}{(wv^d)^{\frac{1}{2}}} e^{-\frac{\|x-\mu\|^2}{(\frac{8wv}{v+4w})}} \, dG(v) \, dH(w) \, dx \\ &\propto \int_{\mathbb{R}^d} \frac{(\|\hat{\mu}(x) - \mu\|^2 - \|x - \mu\|^2)}{2\|x - \mu\|} \int_0^\infty \int_0^\infty \left(\frac{8\pi wv}{v + 4w}\right)^{-\frac{d}{2}} e^{-\frac{\|x-\mu\|^2}{(\frac{8wv}{v+4w})}} \, d\tau(v, w) \, dx \\ &\propto \int_{\mathbb{R}^d} \frac{(\|\hat{\mu}(x) - \mu\|^2 - \|x - \mu\|^2)}{2\|x - \mu\|} \int_0^\infty (2\pi z)^{-d/2} e^{-\frac{\|x-\mu\|^2}{2z}} \, dF_Z(z) \, dx, \end{aligned}$$

[2] For $d = 1$, the density in (12) is not well defined.

[3] It is not the case that spherically symmetric distributions share a similar consistency property, but it is true indeed for scale mixtures of normals distributions (e.g., [14]).

which establishes part (a).

(b) We apply part (a). We show below that the density $p^*(\|s\|^2)$ is a scale mixture of normals. This permits us to apply the dominance result of [22] for Baranchik estimators satisfying conditions (i), (ii), (iii) of Theorem 2.1, in cases where both $E_0^{p^*}\|X\|^2$ and $E_0^{p^*}\|X\|^{-2}$ are finite, and for $0 < a \le 2/(E_0^{p^*}\|X\|^{-2})$. The finiteness conditions are satisfied for $d \ge 4$ and with the finiteness of $E(Z^{1/2})$ and $E(Z^{-3/2})$, and a calculation yields

$$\begin{aligned} E_0^{p^*}\|X\|^{-2} &= \frac{\int_{\mathbb{R}^d} \frac{1}{\|x\|^3} \int_0^\infty (2\pi z)^{-d/2} e^{-\frac{\|x\|^2}{2z}}\, dF_Z(z)\, dx}{\int_{\mathbb{R}^d} \frac{1}{\|x\|} \int_0^\infty (2\pi z)^{-d/2} e^{-\frac{\|x\|^2}{2z}}\, dF_Z(z)\, dx} \\ &= \frac{\int_0^\infty \int_{\mathbb{R}^d} \frac{1}{\|x\|^3} (2\pi z)^{-d/2} e^{-\frac{\|x\|^2}{2z}}\, dx\, dF_Z(z)}{\int_0^\infty \int_{\mathbb{R}^d} \frac{1}{\|x\|} (2\pi z)^{-d/2} e^{-\frac{\|x\|^2}{2z}}\, dx\, dF_Z(z)} \\ &= \frac{1}{(d-3)} \frac{E(Z^{-3/2})}{E(Z^{-1/2})}, \end{aligned}$$

using expectation expressions for a central χ_d^2 distribution. This yields the desired result. It remains to show that $p^*(\|s\|^2)$, $s \in \mathbb{R}^d$, is a scale mixture of normals density. Recall that, in general, a spherically symmetric density $f(\|t-\mu\|^2)$ is a scale mixture of normals if and only if f is completely monotone on $(0,\infty)$, i.e., $(-1)^k f^{(k)}(t) \ge 0$ for $t > 0$ and $k = 0, 1, 2, \ldots$ (e.g., [4]). Since both $t^{-1/2}$ and $\int_{(0,\infty)} (2\pi z)^{-d/2} e^{-\frac{t}{2z}}\, d\tau(z)$ are completely monotone, it follows that their product is completely monotone (e.g., [11], page 417) and that the density in (12) is indeed a scale mixture of normals. □

Remark 2.4. Written in terms of the mixing variance cdfs (H, G) in (10), Theorem 2.2(b)'s bound on a is, using (11), equal to

$$8(d-3) \frac{\int_{\mathbb{R}_+} \frac{z_1^{-1/2} z_2^{d/2-1}}{(z_1+4z_2)^{(d-1)/2}}\, dG(z_1)\, dH(z_2)}{\int_{\mathbb{R}_+} \frac{z_1^{-3/2} z_2^{d/2-2}}{(z_1+4z_2)^{(d-3)/2}}\, dG(z_1)\, dH(z_2)}.$$

Example 2.2 (Normal case). In the normal case (2) which arises as a particular case of (10) for degenerate V, W, we obtain that Z in (11) is also degenerate with $P(Z = z_0) = 1$, with $z_0 = \frac{4\sigma_X^2 \sigma_Y^2}{\sigma_X^2 + 4\sigma_Y^2}$. In this case as well, we obtain

$$p^*(\|s\|^2) \propto \frac{1}{\|s\|} (2\pi z_0)^{-d/2} e^{-\frac{-\|s\|^2}{2z_0}},$$

which is the density of a Kotz distribution (see for instance [21]). By virtue of part (a) of Theorem 2.2, minimax or dominance results applicable to this particular Kotz distribution generate plug-in $N_d(\hat{\mu}(X), \sigma_Y^2 I_d)$ density estimators (such as those in part (b)) which dominate the plug-in density of a $N_d(X, \sigma_Y^2 I_d)$ under L_1 loss. The cut-off point in part (b) reduces to $2(d-3)z_0 = \frac{8(d-3)\sigma_X^2\sigma_Y^2}{\sigma_X^2+4\sigma_Y^2}$. In comparison to Corollary 2.1's cutoff point given in (8), the cut-off point here is larger by a multiple of $d/(d-2)$.

Proceeding with a numerical illustration, we set $\sigma_X^2 = 1$ (without loss of generality) and consider the (smooth) Baranchik estimator $\hat{\mu}_a(X) = (\frac{X'X}{X'X+a})X$, corresponding to $r(t) = \frac{t}{t+a}$. which satisfies the dominance condition as long as $0 < a \le a_0(d, \sigma_Y) = \frac{8(d-3)\sigma_Y^2}{1+\sigma_Y^2}$. We select the upper cut-off point $a_0(d, \sigma_Y)$ and compare the risks of the plug-in predictive density estimators $q_Y(\|y-X\|^2)$ and $q_Y(\|y-\hat{\mu}_{a_0}(X)\|^2)$. Figure 1(left) shows the ratio of these risks for $d = 6, \sigma_Y = 0.5, 1, 2$. Our theoretical results tells us that the ratio is less than one. In accordance with the traditional performance of shrinkage estimators, empirical findings show that (relative) improvement is most important when $\lambda = \|\mu\|$ is close to 0 and such an improvement dissipates for large λ. Furthermore, these gains are amplified when σ_Y^2 increases (i.e., larger uncertainty in Y is better mitigated by the shrinkage procedure). Similarly, these gains are amplified as the dimension d rises. With the maximum gain at 0, Figure 1(right) illustrates these above features of the ratio of risks evaluated at $\mu = 0$ as a function of σ_Y, and illustrates how important the gains can be. In practice, one may shrink to any prior plausible value μ_0 of μ (by using $\hat{\mu}(X) = \mu_0 + (1 - \frac{r(\|X-\mu_0\|^2)}{\|X-\mu_0\|^2})(X-\mu_0)$, with improvement expected to be most important for small $\|\mu-\mu_0\|^2$.

Example 2.3 (Cases where the mixing distributions are lower bounded). Suppose in (10) that the mixing variances are lower bounded by positive values, in the sense that there exists known positive constants a_X and a_Y such that $G^-(a_X) = 0$ and $H^-(a_Y) = 0$ (G and H need not be known). In such cases, Theorem 2.2's cutoff point $2(d-3)\frac{E(Z^{-3/2})}{E(Z^{-1/2})}$ for the constant a of the

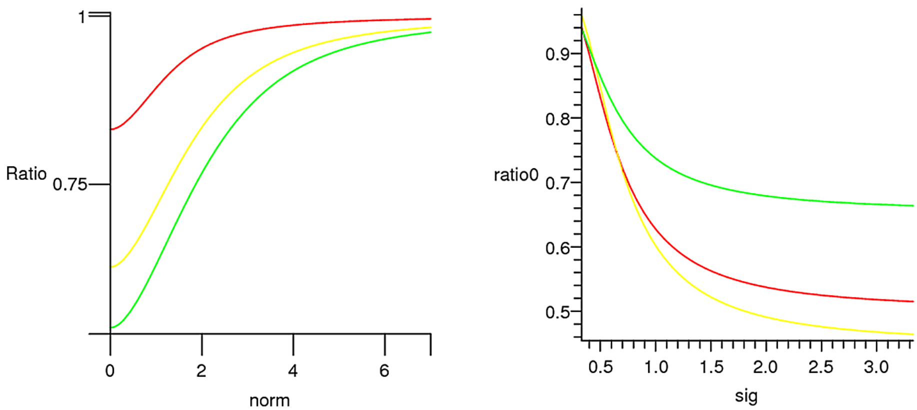

Figure 1. Ratios $\frac{R(\mu, q_Y(\|y-\hat{\mu}_{a_0}(X)\|^2)}{R(\mu, q_Y(\|y-X\|^2)}$ of risks for $d = 6, \sigma_X^2 = 1, \sigma_Y^2 = 2, 1, 0.5$, as functions of $\lambda = \|\mu\|$ (on the left). For fixed λ, ratios increase in σ_Y^2. Same ratios at $\mu = 0$ for $d = 4, 6, 8$ (on the right). These ratios decrease in d.

plug-in Baranchik density estimator $\hat{\mu}_{a,r(\cdot)}(X)$ can be lower bounded as follows. We have by the covariance inequality $\mathrm{Cov}(g_1(Z), g_2(Z)) \geq 0$ for decreasing $g_1(z) = z^{-3/2}$, increasing $g_2(z) = z$, and by making use of (11):

$$\frac{E(Z^{-1/2})}{E(Z^{-3/2})} \geq E(Z) \geq 4E\left(\left(\frac{1}{Z_2} + \frac{4}{Z_1}\right)\right)^{-1} \geq 4\left(\frac{1}{a_Y} + \frac{4}{a_X}\right)^{-1} = \frac{4a_X a_Y}{4a_Y + a_X}.$$

Theorem 2.2's cutoff point is thus bounded below by $2(d-3)E(Z)$ which in turn is bounded below by $\frac{8(d-3)a_X a_Y}{4a_Y + a_X}$. In the normal case, these bounds are exact and take us back to the bound given in Example 2.2. The range of predictive density Baranchik estimators which dominate the plug-in $q_Y(\|y - X\|^2)$ is thus narrower with the lower bound, but the bound is simple, and dominance applies simultaneously for all scale mixture of normals in (10) such that the lower bounds a_X and a_Y on the mixing variance apply.

2.3. Improvements in the case of univariate parametric restrictions

We briefly expand on dominance results applicable to univariate ($d = 1$) cases where μ is either restricted to an interval (a, b), or to a half-interval (a, ∞). Combining Corollary 2.1's duality with point estimation loss $2F(\frac{|\hat{\mu}-\mu|}{2}) - 1$, which is a strictly bowled shaped function of $|\hat{\mu} - \mu|$ on $\mathbb{R}$, with findings of [19], or again [18], we derive the following dominance result for estimating an univariate density $q_Y(|y-\mu|^2)$ based on $X \sim p_X(|x-\mu|^2)$ for cases (such as the normal case) where the family of densities for X has an increasing monotone likelihood ratio (or equivalently $p_X(t^2)$ is logconcave in $t \in \mathbb{R}_+$).

Corollary 2.3. *For estimating an unimodal and univariate symmetric Lebesgue density $q_Y(|y - \mu|^2)$, $y \in \mathbb{R}$, $\mu \in (a, b)$ (or $\mu \in (a, \infty)$) under L_1 loss and based on $X \sim p_X(|x - \mu|^2)$ with $p_X(t^2)$ logconcave, the plug-in density estimator $q_Y(|y - \hat{\mu}_U(X)|^2)$ with $\hat{\mu}_U(X)$ the Bayes estimator of μ with respect to the uniform prior on (a, b) (or on (a, ∞)) dominates the plug-in density estimator $q_Y(|y - X|^2)$.*

Proof. Since $\hat{\mu}_{\pi_U}(X)$ dominates the MRE estimator X as shown by [19] for loss functions $\rho(d - \mu)$ with ρ strict bowled shaped, $\rho(t) > \rho(0) = 0$ for all $t \neq 0$, and logconcave densities, the result follows from part (a) of Corollary 2.1. □

3. Scale expansion improvements on plug-in predictive density estimators

We consider here model (1) in the univariate case and simplify the notation for convenience writing $X \sim p(x - \mu)$, $Y \sim q(y - \mu)$, $x, y, \mu \in \mathbb{R}$, with p, q known and even. With normal p.d.f.'s p and q representing the key example for further reference, we investigate the performance of predictive density estimators $\frac{1}{c}q(\frac{y-x}{c})$ of $q(y - \mu)$, for $c > 1$, under L_1 loss as in (3)

$$\int_{\mathbb{R}} \left| q(y - \mu) - \frac{1}{c} q\left(\frac{y - x}{c}\right) \right| dy, \tag{15}$$

and associated frequentist risk

$$R_c(\mu) = \int_{\mathbb{R}} \left\{ \int_{\mathbb{R}} \left| q(y-\mu) - \frac{1}{c} q\left(\frac{y-x}{c}\right) \right| dy \right\} p(x-\mu)\, dx, \qquad \mu \in \mathbb{R}.$$

By the change of variables $(x, y) \to (x-\mu, y-\mu)$ and by exploiting the assumption that p and q are even, the above risk may be expressed as

$$R_c(\mu) = 2 \int_{\mathbb{R}_+} \left\{ \int_{\mathbb{R}} \left| q(y) - \frac{1}{c} q\left(\frac{y-x}{c}\right) \right| dy \right\} p(x)\, dx. \tag{16}$$

Observe that the risk is thus constant as a function of μ and that an optimal choice of c exists for any given (p, q). We are particularly interested in seeking improvements in terms of risk of the maximum likelihood plug-in estimator $q(y-x)$, $y \in \mathbb{R}$, by scale expansions $\frac{1}{c} q(\frac{y-x}{c})$ with $c > 1$.

The crossings of the densities $q(y)$ and $\frac{1}{c} q(\frac{y-x}{c})$ will be critical in decomposing the loss in (15). One may verify without much difficulty that such densities cross once on $\mathbb{R}_-$ and once on $\mathbb{R}_+$ for cases such as normal (also see Example 3.1) and Laplace (with $q(y) = \frac{1}{2\sigma} e^{-\frac{|y|}{\sigma}}$). The next result establishes such behaviour quite generally under the assumption that q is logconcave.

Lemma 3.1. *Suppose q is an even, differentiable a.e., and logconcave density on $\mathbb{R}$. Then, for any fixed (x, c) with $c > 1$ and $x \neq 0$,* (i) *there is exactly one positive (y_+) and one negative (y_-) solution of $q(y) - \frac{1}{c} q(\frac{y-x}{c}) = 0$. Furthermore,* (ii) *$q(y) > \frac{1}{c} q(\frac{y-x}{c})$ iff $y_- < y < y_+$.*

Proof. With the assumptions, we may write $q(y) = Ae^{-h(y)}$ with h even, $h(y)$ increasing in $|y|$, and $h'(y)$ increasing in y. We break up the proof into separate parts: (A) $y > 0$ and (B) $y < 0$ and we assume $x > 0$ without loss of generality. Note that

$$\operatorname{sgn}\left\{ q(y) - \frac{1}{c} q\left(\frac{y-x}{c}\right) \right\} = -\operatorname{sgn}\{D(y)\}, \tag{17}$$

with $D(y) = h(y) - h(\frac{y-x}{c}) - \log(c)$. If (i) holds, part (ii) follows as $D(0) = h(0) - h(\frac{-x}{c}) - \log(c) < 0$, given the properties of h and since $c > 1$.

(A) Case $y > 0$. For $y \geq x$, we have

$$D'(y) = h'(y) - \frac{1}{c} h'\left(\frac{y-x}{c}\right) \geq h'(y)\left(1 - \frac{1}{c}\right) \geq h'(x)\left(1 - \frac{1}{c}\right) > 0. \tag{18}$$

Note also that $D'(y) \geq 0$ for $0 < y < x$ since $h(y)$ is increasing in y for $y > 0$, and $h(\frac{y-x}{c})$ is decreasing in y for $y < x$. Hence, $D(\cdot)$ is increasing on $\mathbb{R}_+$. Since $D(0) < 0$, we conclude that there exists exactly one positive solution of the equation $q(y) - \frac{1}{c} q(\frac{y-x}{c}) = 0$.

(B) Case $y < 0$. Set $z = |y|$. By symmetry, for $y < 0$,

$$D(y) = h(z) - h\left(\frac{z+x}{c}\right) - \log(c) = T(z) \qquad \text{(say).}$$

Set $z_0 = \frac{x}{c-1}$. For $z \le z_0$, we have $z \le \frac{z+x}{c}$ and $T(z) = h(z) - h(\frac{z+x}{c}) - \log(c) \le -\log(c) < 0$ given that h is increasing on $\mathbb{R}_+$. There is hence no root of D on $[0, z_0]$. For $z > z_0$, we have

$$T'(z) = h'(z) - \frac{1}{c}h'\left(\frac{z+x}{c}\right) \ge h'(z)\left(1 - \frac{1}{c}\right) \ge h'(z_0)\left(1 - \frac{1}{c}\right) > 0. \tag{19}$$

Therefore T is strictly monotone on (z_0, ∞) with $T(z) \to \infty$ as $z \to \infty$, and we infer that there exists a unique root of T in the interval (z_0, ∞) and none in the interval $(0, z_0)$, which establishes the result.

□

Theorem 3.1. *Suppose q is an even, differentiable a.e., and logconcave density on $\mathbb{R}$. Suppose p is an even density. Then, the risk $R_c(\mu)$ of the predictive density $\frac{1}{c}q(\frac{y-x}{c})$ in estimating $q(y-\mu)$, $y \in \mathbb{R}$, is given by $R_c(\mu) = 4(A_1 + A_2)$, with*

$$A_1 = \int_{\mathbb{R}_+} \left(F\left(\frac{y_- - x}{c}\right) - F(y_-)\right)p(x)\,dx,$$

$$A_2 = \int_{\mathbb{R}_+} \left(F(y_+) - F\left(\frac{y_+ - x}{c}\right)\right)p(x)\,dx,$$

where F is the c.d.f. associated with q, and y_- and y_+ are quantities depending on x representing the negative and positive crossing respectively of $\frac{1}{c}q(\frac{y-x}{c})$ and $q(y)$.

Proof. With $q(y) \ge \frac{1}{c}q(\frac{y-x}{c})$ iff $y_- \le y \le y_+$ by virtue of Lemma 3.1, the given expression for the risk $R_c(\mu)$ follows by evaluating the inner integral in (16) separately on the domains $(-\infty, y_-)$, $[y_-, y_+]$, (y_+, ∞), and by collecting terms. □

Remark 3.1. Taking $c \to 1$ in Theorem 3.1, we obtain $y_- \to -\infty$ and $y_+ \to x/2$. This yields $A_1 \to 0$, $A_2 \to \int_{\mathbb{R}_+}(2F(x/2) - 1)p(x)\,dx$ and

$$R_1(\mu) = \int_{\mathbb{R}_+} \big(8F(x/2) - 4\big)p(x)\,dx = 8\int_{\mathbb{R}_+} F(x/2)p(x)\,dx - 2. \tag{20}$$

The above may be written as $4\int_{\mathbb{R}} F(|x|/2)p(x)\,dx - 2$, and we point out that this also follows from Corollary 2.1.

Example 3.1 (Normal case). For a normal p.d.f. $q(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$, the conclusions of Lemma 3.1 may be directly verified with crossings y_- and y_+ given explicitly by

$$y_- = \frac{-x}{c^2-1} - \frac{c}{c^2-1}D, \qquad y_+ = \frac{-x}{c^2-1} + \frac{c}{c^2-1}D, \tag{21}$$

where $D = \sqrt{x^2 + (c^2-1)\log(c^2)}$. This is so as $q(y) \ge \frac{1}{c}q(\frac{y-x}{c}) \iff y^2(1 - \frac{1}{c^2}) - \log(c^2) - \frac{x^2}{c^2} + \frac{2xy}{c^2} \le 0 \iff y_- \le y \le y_+$ by taking logarithms. Theorem 3.1 thus applies in the normal case with the above values of y_- and y_+. The risk $R_1(\mu)$ of the plug-in predictive density estimator, given by (20) actually reduces to $\frac{2}{\pi}\arcsin(\frac{\sqrt{5}}{5}) \approx 0.590334$. To justify this, write

$$\begin{aligned}\int_{\mathbb{R}_+} F(x/2)p(x)\,dx &= \int_{\mathbb{R}_+}\int_{-\infty}^{x/2} p(y)p(x)\,dy\,dx \\ &= \int_{\mathbb{R}_-}\int_{-\infty}^{-x/2} p(y)p(x)\,dy\,dx = \mathbb{P}(Z_1 \le 0, Z_2 \le 0),\end{aligned}$$

with (Z_1, Z_2) distributed as bivariate normal with means 0, variances 1 and correlation $\rho = \frac{\sqrt{5}}{5}$. The result for $R_1(\mu)$ follows with the quadrant probability identity $\mathbb{P}(Z_1 \le 0, Z_2 \le 0) = \frac{1}{4} + \frac{\arcsin(\rho)}{2\pi}$ for bivariate normal vectors (e.g., [20], page 44).

We now proceed with the main result of this section.

Theorem 3.2. *Suppose q is an even, differentiable a.e., and logconcave density on $\mathbb{R}$. Suppose p is an even density. Then, for estimating the density $q(y-\mu)$, $y \in \mathbb{R}$ based on $X \sim p(x-\mu)$, $x \in \mathbb{R}$, under L_1 loss, the predictive density estimator $q(y-x)$ is inadmissible and dominated by a subclass of predictive density estimators $\frac{1}{c}q(\frac{y-x}{c})$ for $c > 1$ and small enough $c-1$.*

Remark 3.2. Observe that the above result is quite general as long as q is logconcave. It applies namely for the normal case with $Y \sim N(\mu, \sigma_Y^2)$. Little, not even unimodality, is required of p except the evenness.

Proof of Theorem 3.2. We show that $\frac{d}{dc}R_c(\mu)\ |_{c=1+} < 0$, which will suffice. We have from Theorem 3.1 by differentiating under the integral sign

$$\begin{aligned}\frac{d}{dc}A_1 &= \int_{\mathbb{R}_+} \frac{d}{dc}\left(F\left(\frac{y_- - x}{c}\right) - F(y_-)\right)p(x)\,dx \\ &= \int_{\mathbb{R}_+}\left(q\left(\frac{y_- - x}{c}\right)\frac{(cy'_- - (y_- - x))}{c^2} - q(y_-)y'_-\right)p(x)\,dx \\ &= \int_{\mathbb{R}_+} \frac{1}{c}q(y_-)(x - y_-)p(x)\,dx,\end{aligned}$$

where $y'_- = \frac{d}{dc}y_-$, and where we have made use of the property $q(y_-) = \frac{1}{c}q(\frac{y_- - x}{c})$ which y_- satisfies by definition. Similarly, we obtain

$$\frac{d}{dc}A_2 = \int_{\mathbb{R}_+} \frac{1}{c}q(y_+)(y_+ - x)p(x)\,dx.$$

Taking $c \to 1^+$, we have $\lim_{c\to 1^+} y_- = -\infty$, $\lim_{c\to 1^+} y_+ = \frac{x}{2}$ (Remark 3.1), and hence $\lim_{c\to 1^+} y_- q(y_-) = \lim_{t\to -\infty} tq(t) = 0$. We thus have

$$\frac{d}{dc} R_c(\mu)\,|_{c=1+} = \lim_{c\to 1_+} \frac{4}{c} \int_{\mathbb{R}_+} \big(q(y_-)(x-y_-) + q(y_+)(y_+ - x)\big) p(x)\,dx$$
$$= -2 \int_{\mathbb{R}_+} xq\left(\frac{x}{2}\right) p(x)\,dx < 0,$$

completing the proof. □

Example 3.2 (Normal case continued). As an illustration, for $\sigma_Y^2 = \sigma_X^2 = 1$, we obtain numerically that $R_c(\mu)$ decreases for $1 < c < c_0$ and increases for $c \geq c_0$, with $c_0 \approx 1.128$ representing the optimal choice of c among the estimators $\hat{q}_c(y;x) = \frac{1}{c} q(\frac{y-x}{c})$. We also obtain that $\hat{q}_c$ dominates $\hat{q}_1$ iff $1 < c < c_1$ with $c_1 \approx 1.2936$. However, the gains are small. For instance $R_{c_0}(\mu) \approx 0.57690$ while $R_1(\mu) \approx 0.590334$ (see Example 3.1), representing a improvement of around 2.27%.

4. Concluding remarks

We have shown that, for estimating a d-dimensional unimodal spherically symmetric density $q_Y(\|y-\mu\|^2)$ based on $X \sim p_X(\|x-\mu\|^2)$, and under L_1 loss, the benchmark plug-in $q_Y(\|y-X\|^2)$ predictive density estimator is quite generally inadmissible for $d \geq 4$ in terms of frequentist risk, and dominated by a class of plug-in predictive density estimators $q_Y(\|y-\hat{\mu}(X)\|^2)$, with the $\hat{\mu}(X)$ being James–Stein and more generally Baranchick-type estimators of μ. We have capitalized on a L_1 distance formula (Lemma 2.1) to establish the link between the predictive density estimation problem and a point estimation of μ problem based on X under a dual loss of the form $\rho(\|d-\mu\|^2)$ with concave ρ. Our inadmissibility and dominance results are obtained by making use of techniques in Stein estimation for such concave losses, and by working with the specific form of ρ. The findings also represent multivariate mean point estimation contributions on their own, and further recent work in this regard appears in [17].

We have also shown in the univariate normal case, and more generally for log-concave density q, that a scale expansion, induced by the predictive density estimator $q_c(y;x) = \frac{1}{c} q(\frac{y-x}{c})$ with $c > 1$, dominates the plug-in $\hat{q}_1$ for c's slightly larger than 1.

Although L_1 distance arises in many varied theoretical and practical situations, its analytical treatment appears to be quite difficult. The techniques and results presented here address such a difficulty and, we believe, pave the way for further findings.

Acknowledgements

The authors are grateful to an Associate Editor and referee for helpful suggestions. Tatsuya Kubokawa's research is supported in part by Grant-in-Aid for Scientific Research Nos.

15H01943 and 26330036 from the Japan Society for the Promotion of Science, Eric Marchand's research is supported in part by the Natural Sciences and Engineering Research Council of Canada, and William Strawderman's research is partially supported by a grant from the Simons Foundation (#209035). Finally, thanks to the Center for International Research on the Japanese Economy which provided financial support for 2012 and 2015 visits by Marchand to the University of Tokyo.

References

[1] Aitchison, J. (1975). Goodness of prediction fit. *Biometrika* **62** 547–554. MR0391353

[2] Andrews, D.F. and Mallows, C.L. (1974). Scale mixtures of normal distributions. *J. Roy. Statist. Soc. Ser. B* **36** 99–102. MR0359122

[3] Baranchik, A.J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Stat.* **41** 642–645. MR0253461

[4] Berger, J. (1975). Minimax estimation of location vectors for a wide class of densities. *Ann. Statist.* **3** 1318–1328. MR0386080

[5] Brandwein, A.C., Ralescu, S. and Strawderman, W.E. (1993). Shrinkage estimators of the location parameter for certain spherically symmetric distributions. *Ann. Inst. Statist. Math.* **45** 551–565. MR1240354

[6] Brandwein, A.C. and Strawderman, W.E. (1980). Minimax estimation of location parameters for spherically symmetric distributions with concave loss. *Ann. Statist.* **8** 279–284. MR0560729

[7] Brandwein, A.C. and Strawderman, W.E. (1991). Generalizations of James–Stein estimators under spherical symmetry. *Ann. Statist.* **19** 1639–1650. MR1126343

[8] Brown, L.D., George, E.I. and Xu, X. (2008). Admissible predictive density estimation. *Ann. Statist.* **36** 1156–1170. MR2418653

[9] DasGupta, A. and Lahiri, S.N. (2012). Density estimation in high and ultra high dimensions, regularization, and the L_1 asymptotics. In *Contemporary Developments in Bayesian Analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman. Inst. Math. Stat. (IMS) Collect.* **8** 1–23. Beachwood, OH: IMS. MR3202499

[10] Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation. The L_1 View*. New York: Wiley. MR0780746

[11] Feller, W. (1966). *An Introduction to Probability Theory and Its Application*, 2nd ed. New York: Wiley.

[12] Fourdrinier, D., Marchand, É., Righi, A. and Strawderman, W.E. (2011). On improved predictive density estimation with parametric constraints. *Electron. J. Stat.* **5** 172–191. MR2792550

[13] George, E.I., Liang, F. and Xu, X. (2006). Improved minimax predictive densities under Kullback–Leibler loss. *Ann. Statist.* **34** 78–91. MR2275235

[14] Kano, Y. (1994). Consistency property of elliptical probability density functions. *J. Multivariate Anal.* **51** 139–147. MR1309373

[15] Komaki, F. (2001). A shrinkage predictive distribution for multivariate normal observables. *Biometrika* **88** 859–864. MR1859415

[16] Kubokawa, T., Marchand, É. and Strawderman, W.E. (2015). On predictive density estimation for location families under integrated squared error loss. *J. Multivariate Anal.* **142** 57–74. MR3412739

[17] Kubokawa, T., Marchand, É. and Strawderman, W.E. (2015). On improved shrinkage estimators for concave loss. *Statist. Probab. Lett.* **96** 241–246. MR3281771

[18] Kubokawa, T. and Saleh, A.K.M.D.E. (1994). Estimation of location and scale parameters under order restrictions. *J. Statist. Res.* **28** 41–51. MR1370413

[19] Marchand, É. and Strawderman, W.E. (2005). Improving on the minimum risk equivariant estimator for a location parameter which is constrained to an interval or a half-interval. *Ann. Inst. Statist. Math.* **57** 129–143. MR2165612

[20] Muirhead, R.J. (2005). *Aspects of Multivariate Statistical Theory*. 2nd ed. New York: Wiley.

[21] Nadarajah, S. (2003). The Kotz-type distribution with applications. *Statistics* **37** 341–358. MR1997185

[22] Strawderman, W.E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. *J. Multivariate Anal.* **4** 255–264. MR0362597

[23] Weizman, M.S. (1970). Measures of overlap of income distributions of white and negro families in the United States. Technical Report 22, US Department of Commerce.

Received March 2015 and revised September 2015