Minimaxity in Predictive Density Estimation with Parametric Constraints

Tatsuya Kubokawa* Éric Marchand† William E. Strawderman ‡
Jean-Philippe Turcotte§
University of Tokyo, Université de Sherbrooke and Rutgers University

September 7, 2012

Abstract

This paper is concerned with estimation of a predictive density with parametric constraints under Kullback-Leibler loss. When an invariance structure is embedded in the problem, general and unified conditions for the minimaxity of the best equivariant predictive density estimator are derived. These conditions are applied to check minimaxity in various restricted parameter spaces in location and/or scale families. Further, it is shown that the generalized Bayes estimator against the uniform prior over the restricted space is minimax and dominates the best equivariant estimator in a location family when the parameter is restricted to an interval of the form $[a_0, \infty)$. Similar findings are obtained for scale parameter families. Finally, the presentation is accompanied by various observations and illustrations, such as normal, exponential location, and gamma model examples.

AMS 2000 subject classifications: 62C20, 62C86, 62F10, 62F15, 62F30.
Key words and phrases: Bayes estimators, decision theory, dominance, Kullback-Leibler loss, invariance, location family, location-scale family, minimaxity, order restriction, predictive density, restricted parameter space, scale family.

1 Introduction

1.1 Preamble

We consider here predictive density estimation for continuous models with

$$X \sim p_\theta(\cdot)\,,\, Y \sim q_\theta(\cdot)\,, \tag{1.1}$$

*Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN E-Mail: tatsuya@e.u-tokyo.ac.jp

†Département de mathématiques, Université de Sherbrooke, Sherbrooke, QC, CANADA, J1K 2R1 E-Mail: eric.marchand@usherbrooke.ca

‡Department of Statistics and Biostatistics, Rutgers University, 561 Hill Center, Busch Campus, Piscataway, N.J., USA 08854-8019 E-Mail: straw@stat.rutgers.edu

§Département de mathématiques, Université de Sherbrooke, Sherbrooke, QC, CANADA, J1K 2R1 E-Mail: jean-philippe.turcotte2@usherbrooke.ca

where the parameter θ is restricted. We seek efficient estimators $\hat{q}(\cdot|X)$ of q_θ based on X under Kullback-Leibler loss

$$L_{KL}(q_\theta, \hat{q}) \;=\; \int q_\theta(y)\;\log \frac{q_\theta(y)}{\hat{q}(y)}\,\mathrm{d}y, \tag{1.2}$$

and as measured by the Kullback-Leibler risk

$$R_{KL}(\theta, \hat{q}) \;=\; E_\theta^{X,Y}\, L_{KL}(q_\theta, \hat{q}(Y|X))\,. \tag{1.3}$$

Such a framework includes normal models with, for instance, $X \sim N_p(\mu, \sigma_X^2 I_p)$, $Y \sim N_p(\mu, \sigma_Y^2 I_p)$ with μ restricted to a convex subset of $\mathbb{R}^p$ as studied recently by Fourdrinier et al. (2011). Our findings will focus on two fundamental questions:

(A) whether the best equivariant procedure $\hat{q}^{BI}$ is minimax for both the unrestricted version and the restricted version of the problem;

(B) whether the Bayes estimator $\hat{q}^{U}$ with respect to the truncated (onto the restricted parameter space) right Haar invariant measure improves upon uniformly on $\hat{q}^{BI}$.

Part **(A)** requires an invariance structure which we will expand on in Section 2. Point estimation *unrestricted* parameter space versions of **(A)**, with affirmative answers in many situations, date back to Girshick and Savage (1951), Kiefer (1957), Hora and Buehler (1966, 1967), among others. Point estimation *restricted parameter* versions of **(A)** and **(B)**, with affirmative answers, date back to Katz (1961) who showed under squared error loss that the Bayes estimator with respect to the flat prior on $[0, \infty)$, for normal models with mean μ and known variance, dominates the best equivariant estimator and is minimax for the restricted parameter space $\mu \in [0, \infty)$. There are several related results in the literature (e.g., Farrell, 1964; Kubokawa, 2004; Marchand and Strawderman, 2005A,B; Tsukuma and Kubokawa, 2008) for restricted (unbounded) parameter spaces, with a quite general minimax result given recently by Marchand and Strawderman (2012). As further illustrated by the work of Casella and Strawderman (1981), Marchand and Perron (2001), Hartigan (2004), Marchand and Strawderman (2004), Kubokawa (2005A,B), and van Eeden (2006) among others, frequentist properties like minimaxity of best equivariant estimators, restricted maximum likelihood estimators or Bayesian estimators depend on the model, the loss, but also intimately on the nature of the parametric restriction.
Predictive density estimation addresses the challenging and ambitious problem of estimating the whole distribution of a future observation Y. This has become a field of active study with early findings due to Aitchison (1975). In particular, for Gaussian models under Kullback-Leibler loss, fascinating connections with *Stein estimation* have been developed, as recently reviewed by George, Liang, and Xu (2012), and as expanded upon below in subsection 1.3.

1.2 Outline of Paper

In this paper, we investigate minimaxity of the best equivariant predictive density estimator in location and/or scale families with parametric constraints under Kullback-Leibler loss. In Section 2, we treat a setup with a general invariance structure given by Hora and Buehler (1966, 67), where the parameter space is restricted to a subset of multi-dimensional Euclidean space. Using similar arguments as in Girshick and Savage (1951), we derive unified conditions under which the best equivariant estimator is minimax. These conditions are available for both restricted and non-restricted cases, and in a sense, the minimaxity result is an extension of findings by Liang and Barron (2004), who showed minimaxity when the parameter space is unrestricted. Minimaxity under parametric constraints for a given type of problem can thus be tested by checking those unified conditions.

Section 3 deals with a location or scale family. In Section 3.1, minimaxity of the best location equivariant estimator is verified under a one-sided restriction of the location parameter in a location family. In section 3.2, we make use of a novel variation of the IERD method introduced by Kubokawa (1994A,B) and Kubokawa and Saleh (1998) to prove that the generalized Bayes estimator against the uniform prior over the restricted space dominates the best location equivariant estimator if the target density to be predicted has a monotone likelihood ratio property. It is interesting to note that the density of the observation does not have to have a monotone likelihood ratio and need not be of the same family as the target density. Analogous findings for scale parameter families are obtained in Section 3.3. Various other observations, detailed examples for normal, exponential and gamma models, and a non-minimaxity result for a compact interval restriction, complement the presentation.

In Section 4, we treat various restrictions in location-scale families and investigate minimaxity of the best location-scale equivariant estimator. Section 4.1 considers the cases that the location and scale parameters are in one-sided open spaces, and Section 4.2 investigates cases with a compact interval restriction for the location parameter and an unknown scale. Through several examples of parametric restrictions given in Sections 4.1 and 4.2, we demonstrate how to use the conditions given in Section 2. Minimaxity in the cases of ordered location or scale parameters in multidimensional distributions is shown in Section 4.3.

1.3 Brief review of previous findings for normal models

We conclude this introduction with a brief review on developments under a multivariate normal distribution with unknown mean vector and known covariance matrices which are multiples of identity, since most decision-theoretic results have been studied in this model and since such a review is helpful for the overall presentation of our findings. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be mutually independent random vectors such that $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, v_x\boldsymbol{I})$ and $\boldsymbol{Y} \sim \mathcal{N}_p(\boldsymbol{\mu}, v_y\boldsymbol{I})$ for known constants v_x and v_y. The density functions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are denoted by $f(\boldsymbol{x}-\boldsymbol{\mu}|v_x)$ and $f(\boldsymbol{y}-\boldsymbol{\mu}|v_y)$. The problem is to predict the density $f(\boldsymbol{y}-\boldsymbol{\mu}|v_y)$ based on $\boldsymbol{X}$ in terms of the following risk relative of the Kullback-Leibler (KL) divergence

$$R_{KL}(\boldsymbol{\mu}, \hat{f}) = E_{\boldsymbol{X}}\Big[\int f(\boldsymbol{y}-\boldsymbol{\mu}|v_y)\log\left(\frac{f(\boldsymbol{y}-\boldsymbol{\mu}|v_y)}{\hat{f}(\boldsymbol{y}|\boldsymbol{X}, v_x, v_y)}\right)\mathrm{d}\boldsymbol{y}\Big],$$

where $\hat{f}(\boldsymbol{y}|\boldsymbol{X}, v_x, v_y)$ is a predictive density estimator of $f(\boldsymbol{y}-\boldsymbol{\mu}|v_y)$. Since this model is invariant under location transformations, the best equivariant estimator of $f(\boldsymbol{y}-\boldsymbol{\mu}|v_y)$

is the generalized Bayes estimator against the uniform prior with respect to Lebesgue measure. As expressed in (2.3), the best equivariant estimator is given by

$$\hat{f}^{BI}(\boldsymbol{y}-\boldsymbol{x}|v_x,v_y)=\frac{\int f(\boldsymbol{s}|v_x)f(\boldsymbol{y}-\boldsymbol{x}+\boldsymbol{s}|v_y)\mathrm{d}\boldsymbol{s}}{\int f(\boldsymbol{s}|v_x)\mathrm{d}\boldsymbol{s}}=f(\boldsymbol{y}-\boldsymbol{x}|v_x+v_y).$$

Liang and Barron (2004) showed that $\hat{f}^{BI}(\boldsymbol{y}-\boldsymbol{x}|v_x,v_y)$ is minimax. Concerning the admissibility of $\hat{f}^{BI}(\boldsymbol{y}-\boldsymbol{x}|v_x,v_y)$ in the case of normal distributions, Komaki (2001) showed that it is inadmissible when $p\geq 3$, namely, it is improved on by a generalized Bayes estimator against a shrinkage prior. Brown, George and Xu (2008) showed that it is admissible when $p=1,2$. These are noteworthy results in the sense that the so called Stein inadmissibility result in point estimation is inherited by the problem of estimation of a predictive normal density function. George, Liang and Xu (2006) extended Komaki's result, and along with Brown, George and Xu (2008), showed that several decision-theoretic results for point estimation of a multivariate normal mean with a known variance still hold for the predictive density estimation problem. Kato (2009) succeeded in deriving a minimax and improved generalized Bayes predictive density estimator in the case of unknown variance.

Brown, George and Xu (2008) derived an interesting identity which expresses the relationship between point estimation and predictive density estimation. Let $R_Q^v(\boldsymbol{\mu},\widehat{\boldsymbol{\mu}})$ be the risk function of a point estimator $\widehat{\boldsymbol{\mu}}=\widehat{\boldsymbol{\mu}}(\boldsymbol{z})$ under a normal distribution $\mathcal{N}_p(\boldsymbol{\mu},v\boldsymbol{I})$, namely,

$$R_Q^v(\boldsymbol{\mu},\widehat{\boldsymbol{\mu}})=\int\|\widehat{\boldsymbol{\mu}}(\boldsymbol{z})-\boldsymbol{\mu}\|^2f(\boldsymbol{z}-\boldsymbol{\mu}|v)\mathrm{d}\boldsymbol{z},$$

for the Euclidean norm $\|\cdot\|$. Let $\widehat{\boldsymbol{\mu}}_v^\pi$ be the Bayes point estimator of $\boldsymbol{\mu}$ for a prior distribution $\pi(\boldsymbol{\mu})$ in terms of the loss $\|\hat{\mu}-\mu\|^2$. Also, let $\hat{f}^\pi(\boldsymbol{y}|\boldsymbol{x},v_x,v_y)$ be the Bayes estimator of the predictive density. Then, Brown, Brown, George and Xu (2008) showed that

$$R_{KL}(\boldsymbol{\mu},\hat{f}^{BI})-R_{KL}(\boldsymbol{\mu},\hat{f}^\pi)=\frac{1}{2}\int_{v_w}^{v_x}\frac{1}{v^2}\left[R_Q^v(\boldsymbol{\mu},\boldsymbol{X})-R_Q^v(\boldsymbol{\mu},\widehat{\boldsymbol{\mu}}_v^\pi)\right]\mathrm{d}v,\tag{1.4}$$

for $v_w=v_xv_y/(v_x+v_y)$. This implies that dominance properties in point estimation can be automatically inherited by predictive density estimation. An essential point in the above identity is that in the normal distribution, the following representation due to George, Liang and Xu (2006) holds:

$$\hat{f}^\pi(\boldsymbol{y}|\boldsymbol{x},v_x,v_y)=\frac{m_\pi(\boldsymbol{W};v_w)}{m_\pi(\boldsymbol{X};v_x)}\hat{f}^{BI}(\boldsymbol{y}|\boldsymbol{x},v_x,v_y),\tag{1.5}$$

where $m_\pi(\boldsymbol{W};v_w)$ and $m_\pi(\boldsymbol{X};v_x)$ are marginal densities of $\boldsymbol{W}$ and $\boldsymbol{X}$ for $\boldsymbol{W}=(v_y\boldsymbol{X}+v_x\boldsymbol{Y})/(v_x+v_y)$. Using this equality, Fourdrinier, *et al*. (2011) extended identity (1.4) to plug-in estimators of the predictive density.

Identity (1.4) can be applied when the parameter space $\boldsymbol{\theta}$ is restricted to a convex cone C, or more generally to a convex set. In the framework of point estimation under a constraint and squared error loss, Hartigan (2004) proved that $\boldsymbol{X}$ is improved on by the generalized Bayes estimator against the uniform prior over C, and Tsukuma and Kubokawa (2008) showed that $\boldsymbol{X}$ is minimax under the constraint. As developed in Fourdrinier *et al.* (2011), combining these results and the identity (1.4) implies that these properties hold for the estimation of the predictive density.

The inferences are valid for normal distributions where key property (1.4) can be derived from the equality (1.5). The equality (1.5) holds under normality with known variances, but it does not hold in the case of unknown variances. Thus, it is not clear whether a decision-theoretic property in point estimation is inherited by estimation of the predictive density under normality with unknown variances or for another distribution.

2 General conditions for minimaxity

In this section, we treat general parametric distributions in which an invariance structure is embedded, and derive general conditions for minimaxity of the best equivariant estimator. The conditions will be used for checking minimaxity in location and/or scale families.

Let X be an observable random variable and Y be a future random variable. Let $(\mathcal{X} \times \mathcal{Y}, \mathcal{B}_X \times \mathcal{B}_Y)$ be a measurable space of (X, Y) and $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of identifiable probability measures with parameter space Θ. We assume the following conditions.

(A1) There exist a group $\mathcal{G}$ and a measurable space $(\mathcal{G}, \mathcal{B}_G)$ on which there exists a left invariant Haar measure γ satisfying

$$\gamma(gG) = \gamma(G) \quad \text{for all } g \in \mathcal{G} \text{ and all } G \in \mathcal{B}_G.$$

Each $g \in \mathcal{G}$ induces a one-to-one transformation $\overline{g}$ from Θ onto itself defined by $P_{\overline{g}\theta}(gA) = P_\theta(A)$ for any $A \in \mathcal{B}_X \times \mathcal{B}_Y$ and any $\theta \in \Theta$. The induced space $\overline{\mathcal{G}} = \{\overline{g} : \overline{g} \in \mathcal{G}\}$ is measurable.

(A2) There exists a one-to-one correspondence $X \leftrightarrow (t_x, u_x)$ between $\mathcal{X}$ and $\mathcal{G} \times \mathcal{U}_X$ such that gX corresponds to (gt_x, u_x) and $\mathcal{U}_X$ is a measurable space. Also, there exists a one-to-one correspondence $Y \leftrightarrow (t_y, u_y)$ between $\mathcal{Y}$ and $\mathcal{G} \times \mathcal{U}_Y$ such that gY corresponds to (gt_y, u_y) and $\mathcal{U}_Y$ is a measurable space. The statistics u_x and u_y are maximal invariant under the transformation $\mathcal{G}$.

(A3) There exists a one-to-one correspondence $\theta \leftrightarrow \overline{g}_\theta$ between Θ and $\overline{\mathcal{G}}$ such that $\overline{g}\theta$ corresponds to $\overline{g}\,\overline{g}_\theta$ for all $\overline{g} \in \overline{\mathcal{G}}$. The correspondence of $\overline{g}_\theta$ in $\mathcal{G}$ is denoted by g_θ.

(A4) There exist conditional probability density functions $p(g_\theta^{-1} t_x | u_x)$ and $q(g_\theta^{-1} t_y | u_y)$ given u_x and u_y such that for all $A \in \mathcal{B}_X$, $B \in \mathcal{B}_Y$,

$$P_\theta[A] = \int_A p(g_\theta^{-1} t_x | u_x) p_x(u_x) \gamma(\mathrm{d}t_x) \gamma_x(\mathrm{d}u_x),$$

$$P_\theta[B] = \int_B q(g_\theta^{-1} t_y | u_y) q_y(u_y) \gamma(\mathrm{d}t_y) \gamma_y(\mathrm{d}u_y),$$

where $p_x(\cdot)$ is a marginal density function of u_x with respect to a measure $\gamma_x(\cdot)$ on $\mathcal{U}_X$, and $p_y(\cdot)$ and $\gamma_y(\cdot)$ are defined similarly.

We define a measure $\nu(\cdot)$ by

$$\nu(\mathrm{d}g) = \gamma(\mathrm{d}g^{-1}).$$

This is a right invariant Haar measure. Since $\gamma(\cdot)$ is left invariant, it is noted that $\gamma(h\mathrm{d}g) = \gamma(\mathrm{d}g)$ and $\gamma((\mathrm{d}g)h) = \Delta(h)\gamma(\mathrm{d}g)$ for $h, g \in \mathcal{G}$, where $\Delta(\cdot)$ is a modular function.

Now we can set up the problem of estimating the joint predictive density $q(g_\theta^{-1}t_y|u_y)\, q_y(u_y)$ based on (t_x, u_x). When we estimate q_θ by a density $\hat{q}(t_y|u_y, t_x, u_x)q_y(u_y)$, we evaluate the performance using the Kullback-Leibler (KL) divergence in (1.2) and we may write

$$\begin{aligned} L_{KL}(\theta, \hat{q}(\cdot|\cdot, t_x, u_x)) =& \int q(g_\theta^{-1}t_y|u_y)q_y(u_y) \log\left(\frac{q(g_\theta^{-1}t_y|u_y)q_y(u_y)}{\hat{q}(t_y|u_y, t_x, u_x)q_y(u_y)}\right)\gamma(\mathrm{d}t_y)\gamma_y(\mathrm{d}u_y) \\ =& E^{u_y}\Big[\int q(g_\theta^{-1}t_y|u_y)\log\left(\frac{q(g_\theta^{-1}t_y|u_y)}{\hat{q}(t_y|u_y, t_x, u_x)}\right)\gamma(\mathrm{d}t_y)\Big]. \end{aligned}$$

Then, the risk function is

$$R_{KL}(\theta, \hat{q}) = E[L_{KL}(\theta, \hat{q}(\cdot|\cdot, t_x, u_x)] = E^{u_x,u_y}[R_{KL}(\theta, \hat{q}|u_x, u_y)], \tag{2.1}$$

where $E^{u_x,u_y}[\cdot]$ is the expectation with respect to the marginal distribution of (u_x, u_y), and $R_{KL}(\theta, \hat{q}|u_x, u_y)$ is the conditional risk function given (u_x, u_y) equal to

$$\begin{aligned} R_{KL}(\theta, \hat{q}|u_x, u_y) =& \int p(g_\theta^{-1}t_x|u_x)\gamma(\mathrm{d}t_x) \\ &\times \quad \Big\{\int q(g_\theta^{-1}t_y|u_y)\log\left(\frac{q(g_\theta^{-1}t_y|u_y)}{\hat{q}(t_y|u_y, t_x, u_x)}\right)\gamma(\mathrm{d}t_y)\Big\}. \end{aligned} \tag{2.2}$$

This demonstrates that estimation of the joint density function $q(g_\theta^{-1}t_y|u_y)q_y(u_y)$ can be reduced to that of estimating the conditional density function $q(g_\theta^{-1}t_y|u_y)$ as long as estimators of the form $\hat{q}(t_y|u_y, t_x, u_x)q_y(u_y)$ are considered.

Since the problem has an invariance structure, we can derive the best equivariant estimator. Conditional predictive density equivariant estimators under the transformation $\mathcal{G}$ satisfy

$$\hat{q}(gt_y|u_y, gt_x, u_x) = \hat{q}(t_y|u_y, t_x, u_x) \quad \text{for all } g \in \mathcal{G},$$

which implies that a class of (nonrandomized) equivariant estimators is given by

$$\mathcal{Q}_I = \left\{\hat{q}_I(t_x^{-1}t_y|u_y, u_x) \;\Big|\; \int \hat{q}_I(s|u_y, u_x)\gamma(\mathrm{d}s) = 1\right\}.$$

The best equivariant estimator is given in the following proposition.

Proposition 2.1 *Assume conditions* **(A1)** *to* **(A4)***. Then, the best equivariant estimator of $q(g_\theta^{-1}t_y|u_y)$ is given by*

$$\begin{aligned} \hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x) =& \int p(t|u_x)\, q(tt_x^{-1}t_y|u_y)\,\gamma(\mathrm{d}t) \\ =& \frac{\int p(g^{-1}t_x|u_x)\, q(g^{-1}t_y|u_y)\,\nu(\mathrm{d}g)}{\int p(g^{-1}t_x|u_x)\,\nu(\mathrm{d}g)}. \end{aligned} \tag{2.3}$$

Proof. Note that the conditional risk function of $\hat{q}_I(t_x^{-1}t_y|u_y, u_x)$ is free from θ, and from (2.2), it is expressed as

$$\begin{aligned} R_{KL}(\hat{q}_I|u_x, u_y) &= \int\int p(t_x|u_x)q(t_y|u_y)\log\left(\frac{q(t_y|u_y)}{\hat{q}_I(t_x^{-1}t_y|u_y, u_x)}\right)\gamma(\mathrm{d}t_x)\gamma(\mathrm{d}t_y) \\ &= \int\int p(t_x|u_x)q(t_xs|u_y)\log\left(\frac{q(t_xs|u_y)}{\hat{q}_I(s|u_y, u_x)}\right)\gamma(\mathrm{d}t_x)\gamma(\mathrm{d}s), \end{aligned} \tag{2.4}$$

where $s = t_x^{-1}t_y$ and $\gamma(t_x\mathrm{d}s) = \gamma(\mathrm{d}s)$ for the left invariant measure $\gamma(\cdot)$. With the alternative rewriting

$$\begin{aligned} R_{KL}(\hat{q}_I|u_x, u_y) &= \int\int p(t_x|u_x)q(t_xs|u_y)\gamma(\mathrm{d}t_x)\log\left(\frac{\int p(t_x|u_x)q(t_xs|u_y)\gamma(\mathrm{d}t_x)}{\hat{q}_I(s|u_y, u_x)}\right)\gamma(\mathrm{d}s) \\ &+ \int\int p(t_x|u_x)q(t_xs|u_y)\log\left(\frac{q(t_xs|u_y)}{\int p(t_x|u_x)q(t_xs|u_y)\gamma(\mathrm{d}t_x)}\right)\gamma(\mathrm{d}t_x)\,\gamma(\mathrm{d}s), \end{aligned}$$

it is seen that the best equivariant predictive density estimator is

$$\hat{q}^{BI}(s|u_y, u_x) = \int p(t|u_x)q(ts|u_y)\gamma(\mathrm{d}t). \tag{2.5}$$

Making the transformation $t = g^{-1}t_x$, we see that

$$\gamma(\mathrm{d}(g^{-1}t_x)) = \gamma((\mathrm{d}g^{-1})t_x) = \Delta(t_x)\gamma(g^{-1}) = \Delta(t_x)\nu(\mathrm{d}g).$$

Since $1 = \int p(t|u_x)\gamma(\mathrm{d}t) = \int p(g^{-1}t_x|u_x)\Delta(t_x)\nu(\mathrm{d}g)$, it is seen that

$$\Delta(t_x) = 1/\int p(g^{-1}t_x|u_x)\nu(\mathrm{d}g).$$

Substituting $s = t_x^{-1}t_y$ into (2.5) and using the above arguments shows that $\hat{q}^{BI}(s|u_y, u_x)$ is expressed as (2.3). ■

As seen from the form in (2.3), the best equivariant estimator is the generalized Bayes predictive density estimator against the right invariant measure $\nu(\mathrm{d}g)$. Liang and Barron (2004) showed that the best equivariant estimator $\hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)$ is *minimax* if the group $\mathcal{G}$ is *amenable*, namely, if there is a sequence of probability measures $\gamma_j(\cdot)$ on $\mathcal{G}$ that is asymptotically invariant in the sense that $\lim_{j\to\infty}\int\{\psi(ag) - \psi(a)\}\gamma_j(\mathrm{d}a) = 0$ for every $g \in \mathcal{G}$ and every bounded measurable function ψ on $\mathcal{G}$. However, the best equivariant estimator is not necessarily minimax when the parameter space is restricted.

We now provide unified conditions for the minimaxity of the best equivariant predictive density estimator. Although the conditions can be applied to both cases that parameters are restricted and non-restricted, they lead to new findings in restricted cases only, since minimaxity in non-restricted cases follows from the result of Liang and Barron (2004).

(A5) Θ is restricted, and this restriction is equivalently expressed as $g_\theta \in P$. Also, it is assumed that $P \subset \mathcal{G} \subset \mathbb{R}^r$; namely, $\mathcal{G}$ is a subset of r dimensional Euclidean space and P is a restricted space of $\mathcal{G}$.

(A6) There exist sequences of subsets $P_k(\subset P)$ and one-to-one functions $h_k(\cdot)$ between $P_k \leftrightarrow \Xi \subset \mathbb{R}^r$ with $\xi = h_k(g_\theta)$ for $g_\theta \in P_k$, where P_k, $h_k(\cdot)$ and Ξ satisfy the following conditions:

(A6-1) $\cup_{k=k_0}^\infty P_k = P$ for some $k_0 \geq 1$.

(A6-2) Let $V(P_k) = \int_{P_k} \nu(\mathrm{d}g_\theta)$. Let $\gamma_k(\cdot)$ be an induced measure defined by $\gamma_k(A) = \nu(h_k^{-1}(A))$ for $A \in \Xi$. Then, $h_k(P_k) = \Xi = \prod_{i=1}^r [-1 + a_{i,k}, 1 + b_{i,k}]$ and

$$\int_{h_k(P_k)} f(\xi_k)\gamma_k(\mathrm{d}\xi_k)/V(P_k) \geq \frac{1}{2^r + c_k} \int I\Big(\xi \in \prod_{i=1}^r [-1 + a_{i,k}, 1 + b_{i,k}]\Big) f(\xi)\mathrm{d}\xi, \qquad (2.6)$$

where $f(\cdot) > 0$, $\xi_k = h_k(g_\theta)$, $I(\cdot)$ is the indicator function, and $\lim_{k\to\infty} a_{i,k} = \lim_{k\to\infty} b_{i,k} = \lim_{k\to\infty} c_k = 0$ for $i = 1, \ldots, r$.

(A6-3) For any small enough $\varepsilon > 0$ and any $\xi \in \prod_{i=1}^r [-1 + a_{i,k} + \varepsilon, 1 + b_{i,k} - \varepsilon]$, there exists a sequence of subsets P_k^* such that P_k^* does not depend on ξ, $\cup_{k=k_1}^\infty P_k^* = \mathcal{G}$ for some $k_1 \geq 1$ and

$$P_k^* \subset \{[h_k^{-1}(\xi)]^{-1} g;\ g \in P_k\}.$$

Theorem 2.1 *Assume conditions* **(A1)** *to* **(A6-3)**. *Then, the best equivariant estimator $\hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)$ is minimax in estimation of the conditional density $q(g_\theta^{-1}t_y|u_y)$ in terms of the conditional risk* (2.2).

Proof. We can show this theorem along the same lines as in Kubokawa (2004) who modified the method of Girshick and Savage (1951). Consider the sequence of prior distributions given by

$$\pi_k(g_\theta)\nu(\mathrm{d}g_\theta) = \begin{cases} \{V(P_k)\}^{-1}\nu(\mathrm{d}g_\theta) & \text{if } g_\theta \in P_k \\ 0 & \text{otherwise.} \end{cases}$$

This yields the Bayesian predictive densities

$$\hat{q}_k^\pi(t_y|u_y, t_x, u_x) = \int_{P_k} p(g^{-1}t_x|u_x)q(g^{-1}t_y|u_y)\nu(\mathrm{d}g) \int_{P_k} p(g^{-1}t_x|u_x)\nu(\mathrm{d}g)$$

with conditional Bayes risks

$$\begin{aligned} r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) =& \frac{1}{V(P_k)} \int_{P_k} \int \int p(g_\theta^{-1}t_x|u_x)q(g_\theta^{-1}t_y|u_y) \\ & \times \log\left(\frac{q(g_\theta^{-1}t_y|u_y)}{\hat{q}_k^\pi(t_y|u_y, t_x, u_x)}\right) \gamma(\mathrm{d}t_x)\gamma(\mathrm{d}t_y)\nu(\mathrm{d}g_\theta). \end{aligned}$$

Since $r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) \leq r_k(\pi_k, \hat{q}^{BI}|u_x, u_y) = R_0(u_x, u_y)$, it is sufficient to show that $\liminf_{k\to\infty} r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) \geq R_0(u_x, u_y)$. Making the transformations $s_x = g_\theta^{-1}t_x$ and $s_y = g_\theta^{-1}t_y$ yields

$$\begin{aligned} r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) =& \frac{1}{V(P_k)} \int_{P_k} \int \int p(s_x|u_x)q(s_y|u_y) \log\left(\frac{q(s_y|u_u)}{\hat{q}_k^\pi(g_\theta s_y|u_y, g_\theta s_x, u_x)}\right) \nu(\mathrm{d}g_\theta) \\ & \times \gamma(\mathrm{d}s_x)\gamma(\mathrm{d}s_y)\nu(\mathrm{d}g_\theta), \end{aligned} \qquad (2.7)$$

where $\hat{q}_k^\pi(g_\theta s_y|u_y, g_\theta s_x, u_x)$ is expressed as

$$\hat{q}_k^\pi(g_\theta s_y|u_y, g_\theta s_x, u_x) = \frac{\int_{P_k} p(g^{-1}g_\theta s_x|u_x)q(g^{-1}g_\theta s_y|u_y)\nu(\mathrm{d}g)}{\int_{P_k} p(g^{-1}g_\theta s_x|u_x)\nu(\mathrm{d}g)}.$$

Now, make the transformation $g_1 = g_\theta^{-1}g$ with $\nu(\mathrm{d}g) = \Delta(g_\theta)\nu(\mathrm{d}g_1)$ in order to rewrite $\hat{q}_k^\pi(g_\theta s_y|u_y, g_\theta s_x, u_x)$ as

$$\hat{q}_k^\pi(g_\theta s_y|u_y, g_\theta s_x, u_x) = \frac{\int_{g_\theta g_1 \in P_k} p(g_1^{-1}s_x|u_x)q(g_1^{-1}s_y|u_y)\nu(\mathrm{d}g_1)}{\int_{g_\theta g_1 \in P_k} p(g_1^{-1}s_x|u_x)\nu(\mathrm{d}g_1)}.$$

In view of the assumptions, there exists a transformation $\xi_k = h_k(g_\theta)$ satisfying the condition **(A6)**. Note that $g_\theta g_1 \in P_k$ is equivalent to $h_k^{-1}(\xi_k)g_1 \in P_k$, or

$$g_1 \in \{[h_k^{-1}(\xi_k)]^{-1}g; g \in P_k\} \equiv \tilde{P}_k(\xi_k).$$

Then, the Bayes estimator $\hat{g}_k^\pi(g_\theta s_y|u_y, g_\theta s_x, u_x)$ is rewritten as

$$\hat{q}_k^\pi(h_k^{-1}(\xi_k)s_y|u_y, h_k^{-1}(\xi_k)s_x, u_x) = \frac{\int_{g_1 \in \tilde{P}_k(\xi_k)} p(g_1^{-1}s_x|u_x)q(g_1^{-1}s_y|u_y)\nu(\mathrm{d}g_1)}{\int_{g_1 \in \tilde{P}_k(\xi_k)} p(g_1^{-1}s_x|u_x)\nu(\mathrm{d}g_1)}, \tag{2.8}$$

and the conditional Bayes risk (2.7) is rewritten as

$$\begin{aligned} r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) =& \frac{1}{V(P_k)} \int_{h_k(P_k)} \int \int p(s_x|u_x)q(s_y|u_y) \\ & \times \log\left(\frac{q(s_y|u_u)}{\hat{q}_k^\pi(h_k^{-1}(\xi_k)s_y|u_y, h_k^{-1}(\xi_k)s_x, u_x)}\right) \gamma_k(\mathrm{d}\xi_k)\gamma(\mathrm{d}s_x)\gamma(\mathrm{d}s_y). \end{aligned}$$

It is noted that from **(A6-2)**, for any small $\varepsilon > 0$,

$$h_k(P_k) = \prod_{i=1}^r [-1 + a_{i,k}, 1 + b_{i,k}] \supset \prod_{i=1}^r [-1 + a_{i,k} + \varepsilon, 1 + b_{i,k} - \varepsilon] \equiv I_{k,\varepsilon}.$$

Then from (2.6), the conditional Bayes risk is evaluated as

$$\begin{aligned} r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) \geq& \frac{1}{2^r} \int I(\xi \in I_{k,\varepsilon}) \int \int p(s_x|u_x)q(s_y|u_y) \\ & \times \log\left(\frac{q(s_y|u_u)}{\hat{q}_k^\pi(h_k^{-1}(\xi)s_y|u_y, h_k^{-1}(\xi)s_x, u_x)}\right) \mathrm{d}\xi\gamma(\mathrm{d}s_x)\gamma(\mathrm{d}s_y). \end{aligned}$$

For $\xi \in I_{k,\varepsilon}$, from **(A6-3)**, it can be seen that $\hat{q}_k^\pi(h_k^{-1}(\xi)s_y|u_y, h_k^{-1}(\xi)s_x, u_x) \to \hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)$

as $k \to \infty$. Hence, Fatou's lemma is used to bound the Bayes risks as

$$\begin{aligned}
\liminf_{k\to\infty} r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) \geq & \frac{1}{2^r}\int \liminf_{k\to\infty} I(\xi \in I_{k,\varepsilon}) \int\int p(s_x|u_x)q(s_y|u_y) \\
& \times \liminf_{k\to\infty} \log\left(\frac{q(s_y|u_y)}{\hat{q}_k^\pi(h_k^{-1}(\xi)s_y|u_y, h_k^{-1}(\xi)s_x, u_x)}\right)\gamma(\mathrm{d}s_x)\gamma(\mathrm{d}s_y)\mathrm{d}\xi \\
= & \frac{1}{2^r}\int_{[-1+\varepsilon, 1-\varepsilon]^r} \mathrm{d}\xi \\
& \times \int\int p(s_x|u_x)q(s_y|u_y)\log\left(\frac{q(s_y|u_y)}{\hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)}\right)\gamma(\mathrm{d}s_x)\gamma(\mathrm{d}s_y) \\
= & (1-\varepsilon)^r R(\theta, \hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)) = (1-\varepsilon)^r R_0(u_x, u_y)
\end{aligned}$$

From the arbitrariness of $\varepsilon > 0$, it follows that $\liminf_{k\to\infty} r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y) \geq R_0(u_x, u_y)$, completing the proof of Theorem 2.1. ■

In the above proof, the Bayes risk is given by $r_k(\pi_k, \hat{q}_k^\pi) = E^{u_x,u_y}[r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y)]$. It is easy to see that $r_k(\pi_k, \hat{q}_k^\pi) \leq E^{u_x,u_y}[r_k(\pi_k, \hat{q}^{BI}|u_x, u_y)] = E^{u_x,u_y}[R_0(u_x, u_y)]$. On the other hand, Fatou's lemma is used to evaluate the Bayes risk as $\liminf_{k\to\infty} r_k(\pi_k, \hat{q}_k^\pi) \geq E^{u_x,u_y}[\liminf_{k\to\infty} r_k(\pi_k, \hat{q}_k^\pi|u_x, u_y)] \geq E^{u_x,u_y}[R_0(u_x, u_y)]$. Thus, we get the following corollary.

Corollary 2.1 *Assume conditions* **(A1)** *to* **(A6-3)***. Then, the best equivariant estimator* $\hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)q_y(u_y)$ *is minimax for the estimation of the joint density* $q(g_\theta^{-1}t_y|u_y)q_y(u_y)$ *in terms of the Kullback-Leibler risk* (2.1)*.*

As we will show in various situations, Theorem 2.1 includes both non-restricted and restricted cases and thus provides a unified result for the minimaxity of the best equivariant estimator.

3 Location and scale families: minimaxity and improvements on $\hat{q}^{BI}$

3.1 Minimaxity for location families

We first deal with the estimation of a density with a restricted location parameter. Let $\boldsymbol{X} = (X_1, \ldots, X_{n_1})$ be a random variable having a density $f(\boldsymbol{x} - \mu)$ for $\boldsymbol{x} - \mu = (x_1 - \mu, \ldots, x_{n_1} - \mu)$, and let $\boldsymbol{Y} = (Y_1, \ldots, Y_{n_2})$ be a random variable having a density $g(\boldsymbol{y} - \mu)$ for $\boldsymbol{y} - \mu = (y_1 - \mu, \ldots, y_{n_2} - \mu)$, where the location parameter is restricted to the one-sided parameter space

$$A = \{\mu \mid \mu \geq a_0\} \quad \text{for known} \ \ a_0.$$

Let $u_x = (x_2 - x_1, \ldots, x_{n_1} - x_1)$ and $u_y = (y_2 - y_1, \ldots, y_{n_2} - y_1)$ be the maximal invariants. The location models are expressed as $p(x_1 - \mu|u_x) = f(x_1 - \mu, u_x + x_1 - \mu)/p_x(u_x)$ and $q(y_1 - \mu|u_y) = g(y_1 - \mu, u_y + y_1 - \mu)/q_y(u_y)$ for $p_x(u_x) = \int f(t, u_x + t)\mathrm{d}t$ and $q_y(u_y) = \int g(t, u_y + t)\mathrm{d}t$, where $u_x + a$ means $u_x + a = (x_2 - x_1 + a, \ldots, x_{n_1} - x_1 + a)$ for a scalar a.

When the parameter μ is not restricted, it follows from (2.3) that the best equivariant estimator for predicting the density $q(y_1-\mu|u_y)q_y(u_y)$ is $\hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x}) = \hat{q}^{BI}(y_1-x_1|u_y,u_x)q_y(u_y)$, where

$$\hat{q}^{BI}(y_1 - x_1|u_y, u_x) = \frac{\int_{-\infty}^{\infty} p(x_1 - a|u_x)q(y_1 - a|u_y)\mathrm{d}a}{\int_{-\infty}^{\infty} p(x_1 - a|u_x)\mathrm{d}a}, \tag{3.1}$$

which is minimax without the restriction A. When μ is restricted to A, we can show the minimaxity of $\hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x})$.

Theorem 3.1 *The best equivariant estimator $\hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x})$ in the location problem is minimax for estimation of the predictive density under the restricted parameter space A relative to L_{KL}-loss; and the minimax risk is given by $R_0 = R(\mu, \hat{q}^{BI})$.*

Proof. It is sufficient to check conditions **(A6)-(A6-3)** in Theorem 2.1. In this case, $P = \{\mu \geq a_0\}$, $\mathcal{G} = \mathbb{R}$, $\gamma(\mathrm{d}\mu) = \nu(\mathrm{d}\mu) = \mathrm{d}\mu$, $P_k = \{\mu | a_0 < \mu < a_0 + k\}$ and $V(P_k) = k$. Take $\xi_k = h_k(\mu) = (2/k)(\mu - a_0) - 1$. Then, $h_k(P_k) = [-1, 1]$, $\gamma_k(\mathrm{d}\xi_k) = (k/2)\mathrm{d}\xi_k$ and $\int_{h_k(P_k)} f(\xi_k)\gamma_k(\mathrm{d}\xi_k)/V(P_k) = (1/2)\int_{[-1,1]} f(\xi)\mathrm{d}\xi$, which satisfies condition **(A6-2)**. For any $\xi \in [-1+\varepsilon, 1-\varepsilon]$, it is noted that $\mu = h_k^{-1}(\xi) = a_0 + (k/2)(\xi+1)$, so that $\{[h_k^{-1}(\xi)]^{-1}g; g \in P_k\} = \{\mu - a_0 - (k/2)(\xi+1); a_0 < \mu < a_0 + k\} = (-(k/2)(\xi+1), (k/2)(1-\xi)) \supset (-(k/2)\varepsilon, (k/2)\varepsilon) \equiv P_k^*$. Since $\lim_{k\to\infty} P_k^* = \mathbb{R}$, condition **(A6-3)** is satisfied, and the minimaxity of $\hat{q}^{BI}$ is established. ■

3.2 Improvements on the best equivariant estimator $\hat{q}^{BI}$

Although the best equivariant predictive density is minimax, it is not reasonable from a Bayesian or optimization perspective because the prior distribution is taken over whole the space of μ. This suggests that the unrestricted uniform prior Bayes predictive density is likely to be inadmissible and may be improved upon by other (necessarily minimax) predictive densities. A reasonable alternative is the generalized Bayes predictive density against the uniform prior over the restricted space A, given by $\hat{q}^{U}(\boldsymbol{y}|\boldsymbol{x}) = \hat{q}^{U}(y_1,|x_1,u_y,u_x)q_y(u_y)$, where

$$\hat{q}^{U}(y_1|x_1, u_y, u_x) = \frac{\int_{a_0}^{\infty} p(x_1 - a|u_x)q(y_1 - a|u_y)\mathrm{d}a}{\int_{a_0}^{\infty} p(x_1 - a|u_x)\mathrm{d}a}. \tag{3.2}$$

We will indeed establish the minimaxity of the uniform prior Bayes predictive density $\hat{q}^{U}(\boldsymbol{y}|\boldsymbol{x})$ under the following logconcavity or increasing monotone likelihood ratio property:

(C1) The density $q(y_1 - \mu|u_y)$ is a continuously differentiable function such that $q(y_1 - \mu|u_y)/q(y_1 - a_0|u_y)$ is nondecreasing in y_1 for $\mu > a_0$.

Lemma 3.1 *Assume that $q(y_1-\mu|u_y)$ satisfies condition* **(C1)**. *Define $A(y_1|x_1,u_x,u_y,\mu)$ by*

$$A(y_1|x_1, u_x, u_y, \mu) = \frac{\int_{-\infty}^{0} p(x_1 + w - \mu|u_x)q(y_1 + w - \mu|u_y)\mathrm{d}w}{\int_{-\infty}^{0} p(x_1 + w|u_x)q(y_1 + w|u_y)\mathrm{d}w}. \tag{3.3}$$

Then, the following properties hold:

(i) *$q'(y_1|u_y)/q(y_1|u_y)$ is nonincreasing in y_1, where $q'(y_1|u_y) = \nabla_{y_1}q(y_1|u_y)$ for $\nabla_{y_1} = \partial/\partial y_1$;*

(ii) *For $\mu > 0$, $A(y_1|x_1, u_x, u_y, \mu)$ is nondecreasing in y_1.*

Proof. Property **(i)** follows from the fact that $\nabla_{y_1}\{q(y_1 - \mu|u_y)/q(y_1|u_y)\} \geq 0$. For establishing **(ii)**, we shall show that $\nabla_{y_1} A(y_1|x_1, u_x, u_y, \mu) \geq 0$ under assumption **(C1)**. Carrying out the differentiation, we see that this inequality is equivalent to

$$\frac{\int_{-\infty}^{0} p(x_1 + w - \mu|u_x) q'(y_1 + w - \mu|u_y) \mathrm{d}w}{\int_{-\infty}^{0} p(x_1 + w - \mu|u_x) q(y_1 + w - \mu|u_y) \mathrm{d}w} \geq \frac{\int_{-\infty}^{0} p(x_1 + w|u_x) q'(y_1 + w|u_y) \mathrm{d}w}{\int_{-\infty}^{0} p(x_1 + w|u_x) q(y_1 + w|u_y) \mathrm{d}w},$$

or

$$\frac{\int_{-\infty}^{-\mu} p(x_1 + w|u_x) q'(y_1 + w|u_y) \mathrm{d}w}{\int_{-\infty}^{-\mu} p(x_1 + w|u_x) q(y_1 + w|u_y) \mathrm{d}w} \geq \frac{\int_{-\infty}^{0} p(x_1 + w|u_x) q'(y_1 + w|u_y) \mathrm{d}w}{\int_{-\infty}^{0} p(x_1 + w|u_x) q(y_1 + w|u_y) \mathrm{d}w}. \tag{3.4}$$

Hence from (3.4), it is sufficient to show that

$$\frac{\partial}{\partial \mu} \frac{\int_{-\infty}^{-\mu} p(x_1 + w|u_x) q'(y_1 + w|u_y) \mathrm{d}w}{\int_{-\infty}^{-\mu} p(x_1 + w|u_x) q(y_1 + w|u_y) \mathrm{d}w} \geq 0. \tag{3.5}$$

In fact, this derivative is proportional to

$$-p(x_1 - \mu|u_x) q'(y_1 - \mu|u_y) \int_{-\infty}^{-\mu} p(x_1 + w|u_x) q(y_1 + w|u_y) \mathrm{d}w + p(x_1 - \mu|u_x) q(y_1 - \mu|u_y) \int_{-\infty}^{-\mu} p(x_1 + w|u_x) q'(y_1 + w|u_y) \mathrm{d}w,$$

which is rewritten as

$$p(x_1 - \mu|u_x) q(y_1 - \mu|u_y) \int_{-\infty}^{-\mu} p(x_1 + w|u_x) q(y_1 + w|u_y) \times \left\{ \frac{q'(y_1 + w|u_y)}{q(y_1 + w|u_y)} - \frac{q'(y_1 - \mu|u_y)}{q(y_1 - \mu|u_y)} \right\} \mathrm{d}w. \tag{3.6}$$

From property **(i)**, note that $\nabla_{y_1} q(y_1|u_y)/q(y_1|u_y)$ is nonincreasing in y_1. Hence, the integrand in (3.6) is not negative, and the inequality (3.5) holds. This proves Lemma 3.1. ■

Using this lemma, we prove the following theorem.

Theorem 3.2 *Assume condition* **(C1)**. *Then, the uniform prior Bayes predictive density $\hat{q}^U(\boldsymbol{y}|\boldsymbol{x})$ is minimax under the restriction $\mu \geq a_0$. The risks of $\hat{q}^U(\cdot)$ and $\hat{q}^{BI}(\cdot)$ coincide if and only if $\mu = a_0$.*

Proof. Let $a_0 = 0$ without any loss generality. Since $\hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x})$ is a minimax estimator with a constant risk, we shall show that $\hat{q}^{U}(\boldsymbol{y}|\boldsymbol{x})$ improves on $\hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x})$. From (2.1), it is sufficient to show the improvement in terms of the conditional risk (2.2). The IERD method developed by Kubokawa (1994A,B) is useful for the purpose. The conditional risk difference of the two predictive densities $\hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x})$ and $\hat{q}^{U}(\boldsymbol{y}|\boldsymbol{x})$ is written as

$$\begin{aligned}\Delta(\mu) =& R_{KL}(\mu, \hat{q}^{BI}|u_x, u_y) - R_{KL}(\mu, \hat{q}^{U}|u_x, u_y)\\ =& \int\int p(x_1-\mu|u_x)q(y_1-\mu|u_y)\Big\{\log \hat{q}^{U}(\boldsymbol{y}|\boldsymbol{x}) - \log \hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x})\Big\}\mathrm{d}x_1\mathrm{d}y_1.\end{aligned}$$

Observe that

$$\begin{aligned}&\log\hat{q}^{U}(\boldsymbol{y}|\boldsymbol{x}) - \log \hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x})\\ &=\log\frac{\int_0^\infty q(y_1-a|u_y)p(x_1-a|u_x)\mathrm{d}a}{\int_0^\infty p(x_1-a|u_x)\mathrm{d}a} - \log\frac{\int_{-\infty}^\infty q(y_1-a|u_y)p(x_1-a|u_x)\mathrm{d}a}{\int_{-\infty}^\infty p(x_1-a|u_x)\mathrm{d}a}\\ &=\int_{-\infty}^0 \frac{\mathrm{d}}{\mathrm{d}t}\left(\log\frac{\int_t^\infty q(y_1-a|u_y)p(x_1-a|u_x)\mathrm{d}a}{\int_t^\infty p(x_1-a|u_x)\mathrm{d}a}\right)\mathrm{d}t\\ &=\int_{-\infty}^0\left\{\frac{p(x_1-t|u_x)}{\int_t^\infty p(x_1-a|u_x)\mathrm{d}a} - \frac{q(y_1-t|u_y)p(x_1-t|u_x)}{\int_t^\infty q(y_1-a|u_y)p(x_1-a|u_x)\mathrm{d}a}\right\}\mathrm{d}t,\end{aligned}$$

which permits us to write

$$\begin{aligned}\Delta(\mu) =& \int\int p(x_1-\mu|u_x)q(y_1-\mu|u_y)\,\mathrm{d}x_1\,\mathrm{d}y_1\\ &\times\int_{-\infty}^0\left\{\frac{p(x_1-t|u_x)}{\int_t^\infty p(x_1-a|u_x)\mathrm{d}a} - \frac{q(y_1-t|u_y)p(x_1-t|u_x)}{\int_t^\infty q(y_1-a|u_y)p(x_1-a|u_x)\mathrm{d}a}\right\}\mathrm{d}t\,.\end{aligned}$$

Making the transformation $w = -a + t$ with $\mathrm{d}w = -\mathrm{d}a$ gives that $\int_t^\infty p(x_1-a|u_x)\mathrm{d}a = \int_{-\infty}^0 p(x_1-t+w|u_x)\mathrm{d}w$ and $\int_t^\infty q(y_1-a|u_y)p(x_1-a|u_x)\mathrm{d}a = \int_{-\infty}^0 q(y_1-t+w|u_y)p(x_1-t+w|u_x)\mathrm{d}w$. Then, making the transformations $x = x_1 - t$ and $y = y_1 - t$ yields

$$\begin{aligned}\Delta(\mu) =& \int\int\int_{-\infty}^0 p(x+t-\mu|u_x)q(y+t-\mu|u_y)\,\mathrm{d}t\\ &\times\left\{\frac{p(x|u_x)}{\int_{-\infty}^0 p(x+w|u_x)\mathrm{d}w} - \frac{q(y|u_y)p(x|u_x)}{\int_{-\infty}^0 q(y+w|u_y)p(x+w|u_x)\mathrm{d}w}\right\}\mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

Replacing t with w, we can get the expression

$$\begin{aligned}\Delta(\mu) =& \int\int p(x|u_x)\frac{\int_{-\infty}^0 p(x+w-\mu|u_x)q(y+w-\mu|u_y)\mathrm{d}w}{\int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)\mathrm{d}w}\\ &\times\left\{\frac{\int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)\mathrm{d}w}{\int_{-\infty}^0 p(x+w|u_x)\mathrm{d}w} - q(y|u_y)\right\}\mathrm{d}x\mathrm{d}y\\ =& \int\int p(x|u_x)\int A(y|x,u_x,u_y,\mu)\\ &\times\left\{\frac{\int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)\mathrm{d}w}{\int_{-\infty}^0 p(x+w|u_x)\mathrm{d}w q(y|u_y)} - 1\right\}q(y|u_y)\,\mathrm{d}y\,\mathrm{d}x.\end{aligned}$$

Let $B(y|x,u_x,u_y) = \int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)\mathrm{d}w/\{\int_{-\infty}^0 p(x+w|u_x)\mathrm{d}wq(y|u_y)\} - 1$. Denote an expectation with the density $q(y|u_y)$ by $E_q[\cdot]$. From Lemma 3.1, it follows that $A(y|x,u_x,u_y,\mu)$ is nondecreasing in y for $\mu > 0$. Since $q(y+w|u_y)/q(y|u_y)$ is nondecreasing in y, it is seen that $B(y|x,u_x,u_y)$ is nondecreasing in y. Thus, for $\mu > 0$

$$\begin{aligned}
&\int A(y|x,u_x,u_y,\mu)\Big\{\frac{\int_{-\infty}^0 p(x+w|u_x)q(y+w|u_y)\mathrm{d}w}{\int_{-\infty}^0 p(x+w|u_x)\mathrm{d}wq(y|u_y)} - 1\Big\}q(y|u_y)\mathrm{d}y \\
=&E_q[A(Y|x,u_x,u_y,\mu)B(Y|x,u_x,u_y)] \\
\geq&E_q[A(Y|x,u_x,u_y,\mu)]E_q[B(Y|x,u_x,u_y)] \qquad (3.7)\\
=&E_q[A(y|x,u_x,u_y,\mu)]\Big\{\frac{\int_{-\infty}^0 p(x+w|u_x)\int q(y+w|u_y)\mathrm{d}y\mathrm{d}w}{\int_{-\infty}^0 p(x+w|u_x)\mathrm{d}w} - \int q(y|u_y)\mathrm{d}y\Big\},
\end{aligned}$$

where the inequality in (3.7) follows from the well known covariance inequality since both functions $A(y|x,u_x,u_y,\mu)$ and $B(y|x,u_x,u_y)$ are nondecreasing in y (see Wijsman (1984) for example). Since $\int q(y+w|u_y)\mathrm{d}y = \int q(y|u_y)\mathrm{d}y = 1$, it follows that

$$\frac{\int_{-\infty}^0 p(x+w|u_x)\int q(y+w|u_y)\mathrm{d}y\mathrm{d}w}{\int_{-\infty}^0 p(x+w|u_x)\mathrm{d}w} - 1 = 0\,,$$

showing that $\Delta(\mu) \geq 0$ for all $\mu \geq 0$. Observe that $A(y|x,u_x,u_y,0)$ is constant$(=1)$ in y, so that $\Delta(0) = 0$ as seen with the above expansion with an equality replacing the inequality in (3.7). Finally, the covariance inequality in (3.7) is strict when $\mu > 0$ and the proof of Theorem 3.2 is therefore complete. ■

Other improvements on $\hat{q}^{BI}$

Theorem 3.2 establishes a general comparison between the generalized Bayes estimator $\hat{q}^U$ and the best equivariant estimator $\hat{q}^{BI}$, with the former dominating the latter under the simple condition that q be logconcave. It is of interest to seek classes of other dominating procedures. Although we will not explore this issue in depth here, it is nevertheless pertinent to make the following observation which generates many other dominating procedures. The next result follows from the strict concavity of the log function on $(0,\infty)$, or alternatively from the strict convexity with respect to $\hat{q}$ of the loss $L_{KL}(q_\theta, \hat{q})$.

Lemma 3.2 *Let $\alpha \in (0,1)$. Let $\hat{q}_i$, $i = 0,1,2$ be estimators such that $\hat{q}_1 \neq \hat{q}_2$. If $R_{KL}(\theta,\hat{q}_i) \leq R_{KL}(\theta,\hat{q}_0)$ for $i = 1,2$ and for all $\theta \in \Theta$, then $R_{KL}(\theta, \alpha\hat{q}_1 + (1-\alpha)\hat{q}_2) \leq R_{KL}(\theta,\hat{q}_0)$, with equality at a given θ_0 if and only if $R_{KL}(\theta_0,\hat{q}_i) = R_{KL}(\theta_0,\hat{q}_0)$ for $i = 1,2$.*

The above result implies directly that convex linear combinations of $\hat{q}_{BI}$ and $\hat{q}_U$ dominate $\hat{q}_{BI}$ in the context of Theorem 3.2 by taking $\hat{q}_0 = \hat{q}_1 = \hat{q}^{BI}$ and $\hat{q}_2 = \hat{q}^U$. Finally, since Theorem 3.2 applies for the conditional risks, the weights can be made to depend on the maximal invariants u_x and u_y and it thus follows that estimators $\alpha(u_x,u_y)\hat{q}^U(y|x_1,u_y,u_x)q_y(u_y) + (1-\alpha(u_x,u_y))\hat{q}^{BI}(y|x_1,u_y,u_x)q_y(u_y)$ with $\alpha(\cdot,\cdot) \in (0,1)$ are also minimax.

Examples

We proceed with instructive examples and illustrations.

Example 3.1 (normal models) The results above apply to the particular setup:

$$X|\mu \sim \mathcal{N}(\mu, \sigma_X^2), Y|\mu \sim \mathcal{N}(\mu, \sigma_Y^2), \tag{3.8}$$

with the restriction $\mu \geq a_0$. Namely, Theorem 3.1 tells us that $\hat{q}^{BI}(\cdot|X) \sim \mathcal{N}(X, \sigma_X^2 + \sigma_Y^2)$ remains minimax under the restriction $\mu \geq a_0$, while Theorem 3.2 implies that the generalized Bayes estimator $\hat{q}^U$ is also minimax, and dominates $\hat{q}^{BI}$ under the restriction $\mu \geq a_0$. Figure 1 compares the risks of these two estimators for $a_0 = 0, \sigma_X^2 = 1, \sigma_Y^2 = 1$. The curve measures the relative difference in risks (i.e., $\frac{R_{KL}(\mu,\hat{q}^{BI}) - R_{KL}(\mu,\hat{q}^U)}{R_{KL}(\mu,\hat{q}^{BI})}$). Observe that the risks coincide indeed at the lower boundary of the parameter space and at $\mu = \infty$ and that the gains are appreciable, particularly around one standard deviation from the boundary where they fluctuate around 40%.

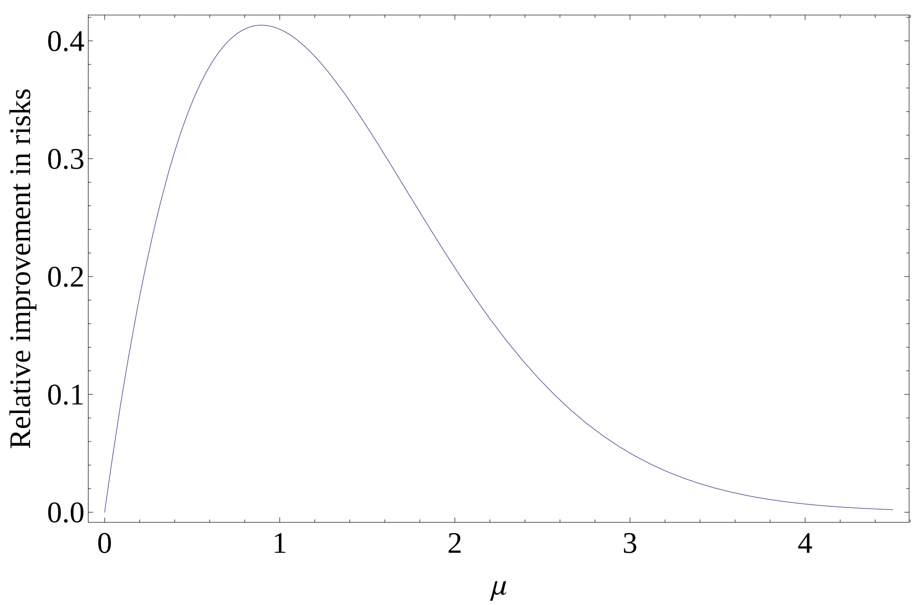

Figure 1: Relative difference in risks between $\hat{q}^{BI}$ and $\hat{q}^U$ (normal model with $\mu \geq 0, \sigma_X^2 = \sigma_Y^2 = 1$)

For the specific normal case illustrated here, the above dominance and minimax results are not new and were previously obtained through a different route by Fourdrinier et al. (2011) by methods which are also applicable for the multivariate case. Interestingly, yet another proof of the dominance result can be derived by a more direct and instructive approach. We now expand on this, considering the more general problem $\mu \in [a_0, a_0 + m)$, with $m = \infty$ corresponding to the lower bounded case and setting hereafter $a_0 = 0$ without loss of generality. Making use of (1.5), the uniform Bayes estimator $\hat{q}^U$ with respect to the flat prior on $[0,m)$ is given by

$$\begin{aligned}\hat{q}^U(Y|X) &= \frac{m_U(W;\sigma_W^2)}{m_U(X;\sigma_X^2)}\,\hat{q}^{BI}(Y|X) \\ &= \left\{\frac{\Phi(\frac{W}{\sigma_W})-\Phi(\frac{W-m}{\sigma_W})}{\Phi(\frac{X}{\sigma_X})-\Phi(\frac{X-m}{\sigma_X})}\right\}\,\hat{q}^{BI}(Y|X)\,,\end{aligned}$$

with $W =^d (\sigma_Y^2 X + \sigma_X^2 Y)/(\sigma_X^2+\sigma_Y^2) \sim \mathcal{N}(\mu,\sigma_W^2)$ for $\sigma_W^2 = (\sigma_X^2\sigma_Y^2)/(\sigma_X^2+\sigma_Y^2)$. Consequently, the difference in risks may be expressed as

$$\begin{aligned}\Delta(\mu) &= R_{KL}(\mu,\hat{q}^{BI}) - R_{KL}(\mu,\hat{q}^U) = E^{X,Y}\log\left(\frac{\hat{q}^U(Y|X)}{\hat{q}^{BI}(Y|X)}\right) \\ &= E^{X,Y}\left\{\log\left(\Phi(\frac{W}{\sigma_W})-\Phi(\frac{W-m}{\sigma_W})\right)-\log\left(\Phi(\frac{X}{\sigma_X})-\Phi(\frac{X-m}{\sigma_X})\right)\right\}.\end{aligned}$$

Here, set $W' = W/\sigma_W \sim \mathcal{N}(\mu/\sigma_W,1)$, $X' = X/\sigma_X \sim \mathcal{N}(\mu/\sigma_X,1)$ and observe that $W' =^d X'+\delta$, with $\delta = \mu(1/\sigma_W - 1/\sigma_X) \geq 0$ for $\mu \geq 0$ with equality iff $\mu = 0$, given that $\sigma_W < \sigma_X$. Hence,

$$\Delta(\mu) = E^{X'}\left\{\log\left(\Phi(X'+\delta)-\Phi(X'+\delta-\frac{m}{\sigma_W})\right)-\log\left(\Phi(X')-\Phi(X'-\frac{m}{\sigma_X})\right)\right\} \geq 0\,,$$

for all $\mu \in [0,m]$, since $\Phi(\cdot)$ is strictly increasing on $\mathbb{R}$ and $x'+\delta \geq x'$ and $x'+\delta-m/\sigma_W \leq x'-m/\sigma_X$ for all $x' \in \mathbb{R}$, and with equality occurring only if $\mu = 0$ and $m = \infty$. We have thus shown directly that the uniform Bayes procedure $\hat{q}^U$ dominates $\hat{q}^{BI}$ for the normal model in (3.8) with the restriction $\mu \in [a_0, a_0+m)$. This offers an alternative to Fourdrinier et al.'s proof. Notwithstanding this development (as well as the next Remark), the search for efficient Bayesian procedures under a compact interval constraint which merits further study will not be pursued here. Recent advances for point estimation versions of this problem were obtained by Kubokawa (2005B), as well as Marchand and Payandeh (2011).

Remark 3.1 (non-minimaxity of $\hat{q}^{BI}$ in the compact interval case)
In the previous example for the compact interval case with $m < \infty$, observe that $\Delta(\mu) > 0$ for all $\mu \in [a_0, a_0+m]$, which implies in turn that $\inf_{\mu\in[a_0,a_0+m]}\Delta(\mu) > 0$ and that $\hat{q}^{BI}$ is not minimax, in contrast to the unbounded lower bounded case. This provides an analog of a familiar point estimation version of this argument (e.g., Lehmann and Casella, 1998, page 327). Moreover, the non-minimaxity argument is more general under condition **(C1)** in the context of Theorem 3.2 as seen by the following elements of proof:

- Theorem 3.2 implies that $\hat{q}^{U1}$ dominates $\hat{q}^{BI}$ for the restriction $\mu \in [a_0, a_0+m]$ where $\hat{q}^{U1}$ is the generalized Bayes predictive density with respect to the flat prior on $[a_0,\infty)$ with equality in risks iff $\mu = a_0$;

- Theorem 3.2 implies that $\hat{q}^{U2}$ dominates $\hat{q}^{BI}$ for the restriction $\mu \in [a_0, a_0+m]$ where $\hat{q}^{U2}$ is the generalized Bayes predictive density with respect to the flat prior on $[-\infty, a_0+m]$ with equality in risks iff $\mu = a_0+m$;

- Paired with the above, Lemma 3.2 implies that the predictive density estimator $\frac{1}{2}\hat{q}^{U1} + \frac{1}{2}\hat{q}^{U2}$ dominates $\hat{q}^{BI}$ strictly for $\mu \in [a_0, a_0 + m]$;
- Consequently, as in the first paragraph of this Remark, $\hat{q}^{BI}$ cannot be minimax for $\mu \in [a_0, a_0 + m]$ when q satisfies condition **(C1)**.

Example 3.2 The results of this section also apply to Exponential location models with $X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ i.i.d. $\text{Exp}(\mu, \sigma)$, $\mu \geq 0$ and known σ, with density $\sigma^{-1} \exp\{-(t - \mu)/\sigma\} 1_{(\mu,\infty)}(t)$. Here the order statistics $X_{(1)}$ and $Y_{(1)}$ form a sufficient statistic, and we can take $\sigma = 1$ without loss of generality, so that it suffices to consider the setup

$$X \sim \text{Exp}(\mu, {n_1}^{-1}),\ Y \sim \text{Exp}(\mu, {n_2}^{-1})\,. \tag{3.9}$$

Evaluating (3.1) and (3.2), we obtain with a little bit of manipulation

$$\hat{q}^{BI}(y|x) = \frac{n_1 n_2}{n_1 + n_2} \left\{ e^{-n_2|x-y|} 1_{[x,\infty)}(y) + e^{-n_1|x-y|} 1_{(-\infty,x)}(y) \right\} ,$$

and

$$\hat{q}^{U}(y|x) = \hat{q}^{BI}(y|x) \left\{ \frac{e^{(n_1+n_2)x} - 1}{e^{(n_1+n_2)x} - e^{n_2 x}} 1_{[x,\infty)}(y) + \frac{1 - e^{-(n_1+n_2)y}}{1 - e^{-(n_1)x}} 1_{(0,x)}(y) \right\} ,$$

Observe that $\hat{q}^{BI}$ is an asymmetric Laplace distribution (and symmetric Laplace for $n_1 = n_2$), while $\hat{q}^{U}$ is a skewed version of $\hat{q}^{BI}$. Theorems 3.1 and 3.2 apply and tell us that both $\hat{q}^{BI}$ and $\hat{q}^{U}$ are minimax under the restriction $\mu \geq 0$, with $\hat{q}^{U}$ dominating $\hat{q}^{BI}$.

3.3 Case of a scale family

We next consider estimation of the predictive density with a restricted scale parameter. Let $\boldsymbol{X} = (X_1, \ldots, X_{n_1})$ be a positive random variable having a density $\sigma^{-n_1} f(\sigma^{-1}\boldsymbol{x})$ for $\sigma^{-1}\boldsymbol{x} = (\sigma^{-1}x_1, \ldots, \sigma^{-1}x_{n_1})$, and let $\boldsymbol{Y} = (Y_1, \ldots, Y_{n_2})$ be a random variable having a density $\sigma^{-n_2} g(\sigma^{-1}\boldsymbol{y})$ for $\sigma^{-1}\boldsymbol{y} = (\sigma^{-1}y_1, \ldots, \sigma^{-1}y_{n_2})$, where the scale parameter is lower bounded belonging to the restricted parameter space

$$B = \{\sigma \mid \sigma \geq b_0\}\,, \quad \text{for known positive } b_0.$$

Let $t_x = |x_1|$, $u_x = (x_1/|x_1|, x_2/|x_1|, \ldots, x_{n_1}/|x_1|)$ and t_y and u_y are defined similarly. The joint densities $\sigma^{-n_1} f(\sigma^{-1}\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ and $\sigma^{-n_2} g(\sigma^{-1}\boldsymbol{y})\mathrm{d}\boldsymbol{y}$ are expressed as, respectively, $p(\sigma^{-1}t_x|u_x)p_x(u_x)\gamma(\mathrm{d}t_x)\gamma_x(\mathrm{d}u_x)$ and $q(\sigma^{-1}t_y|u_y)q_y(u_y)\gamma(\mathrm{d}t_y)\gamma_y(\mathrm{d}u_y)$, where $\gamma(\mathrm{d}\sigma) = \mathrm{d}\sigma/\sigma$, and $p_x(u_x)$ and $q_y(u_y)$ are marginal densities of u_x and u_y.

Note that $\sigma^{-1}t_x = \exp\{\log t_x - \log \sigma\}$ and $\mathrm{d}\log t_x = \mathrm{d}t_x/t_x$. Since the restriction B is written as $\log \sigma > \log b_0$, all the results given in the previous subsection hold for the restricted scale problem. The results corresponding to Theorems 3.1 and 3.2 are described below.

When the parameter σ is not restricted, it follows from (2.3) that the best equivariant estimator for predicting the density $q(\sigma^{-1}t_y|u_y)q_y(u_y)$ is $\hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x}) = \hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x)q_y(u_y)$, where

$$\hat{q}^{BI}(t_x^{-1}t_y|u_y, u_x) = \frac{\int_0^\infty p(b^{-1}t_x|u_x)q(b^{-1}t_y|u_y)b^{-1}\mathrm{d}b}{\int_0^\infty p(b^{-1}t_x|u_x)b^{-1}\mathrm{d}b}, \tag{3.10}$$

which is minimax without the restriction B. Even if σ is restricted on B, the minimaxity of $\hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x})$ still holds.

Theorem 3.3 *The best equivariant estimator $\hat{q}^{BI}(\boldsymbol{y}|\boldsymbol{x})$ is minimax for estimation of the predictive density under the restricted parameter space B relative to the L_{KL}-loss, and the minimax risk is given by $R_0 = R(\sigma, \hat{q}^{BI})$.*

Although the best equivariant predictive density is minimax, it is not reasonable because the prior distribution is taken over whole the space of σ. This suggests that $\hat{q}^{BI}$ is likely to be inadmissible and to be improved upon by other (minimax) predictive densities. A reasonable choice is the generalized Bayes predictive density against the invariant prior over the restricted space B, given by $\hat{q}^U(\boldsymbol{y}|\boldsymbol{x}) = \hat{q}^U(t_y|t_x, u_y, u_x)q_y(u_y)$, where

$$\hat{q}^U(t_y|t_x, u_y, u_x) = \frac{\int_{b_0}^{\infty} p(b^{-1}t_x|u_x)q(b^{-1}t_y|u_y)b^{-1}\mathrm{d}b}{\int_{b_0}^{\infty} p(b^{-1}t_x|u_x)b^{-1}\mathrm{d}b}. \tag{3.11}$$

To establish the minimaxity of the invariant prior Bayes predictive density $\hat{q}^U(\boldsymbol{y}|\boldsymbol{x})$, we assume the following condition analogous to **(C1)**:

(C2) The density $q(\sigma^{-1}t_y|u_y)$ is a continuously differentiable function such that the ratio of the densities $q(\sigma^{-1}t_y|u_y)/q(b_0^{-1}t_y|u_y)$ is nondecreasing in t_y for $\sigma > b_0$.

Theorem 3.4 *Assume condition* **(C2)**. *Then, the Bayes predictive density $\hat{q}^U(\boldsymbol{y}|\boldsymbol{x})$ is minimax under the restriction $\sigma \geq b_0$, and the risks of $\hat{q}^U$ and $\hat{q}^{BI}$ coincide if and only if $\sigma = b_0$.*

Lemma 3.1 used for proving Theorem 3.2 is expressed in the scale case as follows:

Lemma 3.3 *Assume that $q(\sigma^{-1}t_y|u_y)$ satisfies the condition* **(C2)**. *Then, the following properties hold:*

(i) $t_y\{\nabla_{t_y} q(t_y|u_y)\}/q(t_y|u_y)$ is nonincreasing in t_y, where $\nabla_{t_y} = \partial/\partial t_y$.

(ii) Define $B(t_y|t_x, u_x, u_y, \sigma)$ by

$$B(t_y|t_x, u_x, u_y, \sigma) = \frac{\int_0^1 w^{-1}p(\sigma^{-1}wt_x|u_x)q(\sigma^{-1}wt_y|u_y)\mathrm{d}w}{\int_0^1 w^{-1}p(wt_x|u_x)q(wt_y|u_y)\mathrm{d}w}. \tag{3.12}$$

Then for $\sigma > b_0$, $B(t_y|t_x, u_x, u_y, \mu)$ is nondecreasing in t_y.

We can show Theorem 3.4 directly using Lemma 3.3, though we have here applied Theorem 3.2 to the scale case. We conclude this section with an application to Gamma models.

Example 3.3 An interesting application consists of Gamma distributions for X and Y with

$$X|\sigma \sim \mathrm{Gamma}(\alpha_1, \sigma),\ Y|\sigma \sim \mathrm{Gamma}(\alpha_2, \sigma), \tag{3.13}$$

with α_1, α_2 known, and the lower bound restriction $\sigma \geq b_0(> 0)$. We have assumed without loss of generality that the samples for X and Y are of size one by sufficiency of the sums in such Gamma models. Evaluating (3.10) and (3.11), we obtain the elegant representations

$$\hat{q}^{BI}(y|x) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\,\Gamma(\alpha_2)}\,\frac{1}{x}\,(\frac{y}{x})^{\alpha_2-1}\,(1+\frac{y}{x})^{-(\alpha_1+\alpha_2)}\,1_{(0,\infty)}(y)\,,$$

and

$$\hat{q}^{U}(y|x) = \hat{q}^{BI}(y|x)\,\frac{\bar{F}_{\alpha_1+\alpha_2}(\frac{x+y}{b_0})}{\bar{F}_{\alpha_1}(\frac{x}{b_0})}\,,$$

where $\bar{F}_\gamma(\cdot)$ is the survival function of a Gamma$(\gamma,1)$ distribution. Observe that $\hat{q}^{BI}$ is the density of a Fisher distribution with scale parameter $\frac{\alpha_2}{\alpha_1}x$, and shape parameters $2\alpha_2$ (d.f. numerator) and $2\alpha_1$ (d.f. denominator), while $\hat{q}^U$ is a skewed version of $\hat{q}^{BI}$.

The findings of this section apply. First, $\hat{q}^{BI}$ is minimax for the unrestricted parameter space and remains minimax in presence of the lower bound b_0 on the scale parameter. Second, since Gamma densities form a family with an increasing monotone likelihood ratio, condition **(C2)** is satisfied and the Bayes procedure $\hat{q}^U$ dominates $\hat{q}^{BI}$ by virtue of Theorem 3.4. Finally, we point out that analogous results hold here for the case where the scale parameter σ is upper bounded, say $\sigma \in (0, c_0)$. In such cases, we consider the transformed problem with $X' = X$ and $Y' = \frac{1}{Y}$ and consider the setup of Theorem 3.4 with $b_0 = \frac{1}{c_0}$, p_θ being the density of X' and q_θ being the density of Y'. Since inverse Gamma distributions have logconcave densities as well, and the Kullback-Leibler loss is intrinsic, Theorem 3.4 indeed applies.

4 Estimation in location-scale families

In this section, we treat location-scale families with location and/or scale parameters constrained, and investigate minimaxity of the best equivariant estimators using Theorem 2.1.

4.1 Non-bounded case

We begin with the univariate case. Let $\boldsymbol{X} = (X_1, \ldots, X_{n_1})$ be a random variable having a density $\sigma^{-n_1} f((\boldsymbol{x}-\mu)/\sigma)$ for $(\boldsymbol{x}-\mu)/\sigma = ((x_1-\mu)/\sigma, \ldots, (x_{n_1}-\mu)/\sigma)$, and let $\boldsymbol{Y} = (Y_1, \ldots, Y_{n_2})$ be a random variable having a density $\sigma^{-n_2} g((\boldsymbol{y}-\mu)/\sigma)$ for $(\boldsymbol{y}-\mu)/\sigma = ((y_1-\mu)/\sigma, \ldots, (y_{n_2}-\mu)/\sigma)$, where the location and scale parameters are restricted to the space

$$C = \{(\mu,\sigma)|\mu > c_0\sigma + a_0,\ \sigma > b_0\}, \tag{4.1}$$

where a_0, b_0 and c_0 are constants such that $b_0 \geq 0$ and $-\infty \leq a_0, c_0 < \infty$. The unrestricted case is described by $b_0 = c_0 = 0$ and $a_0 = -\infty$. Let $t_x = (|x_2 - x_1|, x_1)$, $u_x = ((x_2 - x_1)/|x_2 - x_1|, \ldots, (x_{n_1} - x_1)/|x_2 - x_1|)$ and let t_y and u_y be defined similarly. Let $\mathcal{G} = \mathbb{R}^+ \times \mathbb{R}$ and define the product by $(a,b)(\sigma,\mu) = (a\sigma, a\mu + b)$. This implies that $(\sigma,\mu)^{-1} = (1/\sigma, -\mu/\sigma)$ and $(\sigma,\mu)^{-1}(|x_2 - x_1|, x_1) = (|x_2 - x_1|/\sigma, (x_1 - \mu)/\sigma)$. Then, $\sigma^{-n_1} f((\boldsymbol{x} - \mu)/\sigma)\mathrm{d}\boldsymbol{x}$ and $\sigma^{-n_2} g((\boldsymbol{y}-\mu)/\sigma)\mathrm{d}\boldsymbol{y}$ are expressed as $p((\sigma,\mu)^{-1}t_x|u_x)p_x(u_x)\gamma(\mathrm{d}t_x)\gamma_x(\mathrm{d}u_x)$ and $q((\sigma,\mu)^{-1}t_y|u_y)q_y(u_y)\gamma(\mathrm{d}t_y)\gamma_y(\mathrm{d}u_y)$, respectively, where $\gamma(\mathrm{d}(\sigma,\mu)) = (\mathrm{d}\mu\mathrm{d}\sigma)/\sigma^2$.

When the parameters are not restricted, it follows from (2.3) that the best equivariant predictive density estimator of $q((\sigma,\mu)^{-1}t_x|u_y)q_y(u_y)$ is given by $\hat{q}^{BI}(t_x^{-1}t_y, u_y|u_x) = \hat{q}^{BI}(t_x^{-1}t_y|u_y,u_x)q_y(u_y)$, where

$$\hat{q}^{BI}(t_x^{-1}t_y|u_y,u_x) = \frac{\int p((b,a)^{-1}t_x|u_x)q((b,a)^{-1}t_y|u_y)\nu(\mathrm{d}(b,a))}{\int p((b,a)^{-1}t_x|u_x)\nu(\mathrm{d}(b,a)}, \tag{4.2}$$

and where $\nu(\mathrm{d}(b,a)) = (\mathrm{d}a\mathrm{d}b)/b^2$. Using Theorem 2.1, we analyze the question of minimaxity of the best equivariant estimator under the restriction C.

[1] Case of $a_0 > -\infty$ and $b_0 > 0$. This case implies that both μ and σ are restricted from one side.

Theorem 4.1 *Assume that a_0 and b_0 satisfy that $a_0 > -\infty$ and $b_0 > 0$. Then, the best equivariant estimator $\hat{q}^{BI}(t_x^{-1}t_y, u_y|u_x)$ is minimax in the estimation of the predictive density under the restricted parameter space C relative to the L_{KL}-loss, and the minimax risk is given by $R_0 = R((\sigma,\mu),\hat{q}^{BI})$.*

Proof. For $c_0 = 0$, we define the sequence $d_k = k$, while for $c_0 \neq 0$ we take $d_k = \log k$. Such a sequence admits the following behaviour when $k \to \infty$,

(a) $(k/d_k)d_k^{\varepsilon/2} \to \infty$ for any $\varepsilon > 0$ when $c_0 = 0$,

(b) $d_k/k \to 0$ and $d_k \to \infty$ when $c_0 \neq 0$.

We proceed by verifying conditions **(A6)-(A6-3)** in Theorem 2.1. In this case, $P = \{(\sigma,\mu)|a_0 + c_0\sigma < \mu, b_0 < \sigma\}$, $\mathcal{G} = \mathbb{R}^+ \times \mathbb{R}$, we set $P_k = \{(\sigma,\mu)|a_0 + c_0\sigma < \mu < a_0 + c_0\sigma + k, b_0 < \sigma < b_0 d_k\}$ and $V(P_k) = k\log d_k$ where d_k is defined above. Take $\xi_1 = (2/\log d_k)\log(\sigma/b_0) - 1$ and $\xi_2 = (2/k)(\mu - a_0 - c_0\sigma) - 1$. Letting $\boldsymbol{\xi} = (\xi_1,\xi_2) = h_k((\sigma,\mu))$, we see that $h_k(P_k) = [-1,1]^2$, $\gamma_k(\mathrm{d}\boldsymbol{\xi}) = \{(k\log d_k)/4\}\mathrm{d}\boldsymbol{\xi}$ and $\int_{h_k(P_k)} f(\xi_k)\gamma_k(\mathrm{d}\boldsymbol{\xi})/V(P_k) = (1/4)\int_{[-1,1]^2} f(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$, which satisfies condition **(A6-2)**. For any $\boldsymbol{\xi} \in [-1+\varepsilon, 1-\varepsilon]^2$, let $(b,a) = h_k^{-1}(\boldsymbol{\xi})$. Then, $b = b_0 d_k^{(1+\xi_1)/2}$ and $a = (k/2)(1+\xi_2) + a_0 + c_0 b_0 d_k^{(1+\xi_1)/2}$ so that $\{[h_k^{-1}(\xi)]^{-1}(\sigma,\mu); (\sigma,\mu) \in P_k\} = \{(\sigma/b, (\mu-a)/b); (\sigma,\mu) \in P_k\}$ and σ/b, $(\mu-a)/b$ satisfy the inequalities

$$\begin{aligned} & d_k^{-(1+\xi_1)/2} < \frac{\sigma}{b} < d_k^{(1-\xi_1)/2}, \\ c_0\frac{\sigma}{b} - \frac{d_k^{-(1+\xi_1)/2}}{b_0}\{\frac{k}{2}(1+\xi_2) + c_0 b_0 d_k^{(1+\xi_1)/2}\} & \\ & < \frac{\mu - a}{b} < c_0\frac{\sigma}{b} + \frac{d_k^{-(1+\xi_1)/2}}{b_0}\{\frac{k}{2}(1-\xi_2) - c_0 b_0 d_k^{(1+\xi_1)/2}\}. \end{aligned}$$

Note that $1 - \xi_i > \varepsilon$ and $1 + \xi_i > \varepsilon$ for $i = 1,2$. The first inequality is satisfied by $d_k^{-\varepsilon/2} < \sigma/b < d_k^{\varepsilon/2}$, which can be expanded to $(0,\infty)$ as $k \to \infty$ if $d_k \to \infty$ as $k \to \infty$. Also, the second inequality is satisfied by

$$\frac{k}{d_k^{(1+\xi_1)/2}}\Big\{c_0\frac{\sigma}{b}\frac{d_k^{(1+\xi_1)/2}}{k} - \frac{\varepsilon}{2b_0} - \frac{c_0}{k}\Big\} < \frac{\mu-a}{b} < \frac{k}{d_k^{(1+\xi_1)/2}}\Big\{c_0\frac{\sigma}{b}\frac{d_k^{(1+\xi_1)/2}}{k} + \frac{\varepsilon}{2b_0} - \frac{c_0}{k}\Big\}.$$

Here, it is noted that

$$\begin{aligned}\frac{\sigma}{b}\frac{d_k^{(1+\xi_1)/2}}{k} &< d_k^{(1-\xi_1)/2}\frac{d_k^{(1+\xi_1)/2}}{k} < \frac{d_k}{k},\\ \text{and}\quad \frac{k}{d_k^{(1+\xi_1)/2}} &= \frac{k}{d_k}d_k^{(1-\xi_1)/2} > \frac{k}{d_k}d_k^{\varepsilon/2}.\end{aligned}$$

Since d_k satisfies the condition (a) or (b), it can be seen that the lower end point of $(\mu - a)/b$ goes to $-\infty$, and the upper point goes to ∞. This verifies condition **(A6-3)**, and the minimaxity of $\hat{q}^{BI}$ is established. ■

[2] Case of $a_0 = -\infty$ and $b_0 > 0$. Although we can show the minimaxity directly by the same arguments as in the proof of Theorem 4.1, we here give a simple proof based on Theorem 3.3. Since μ is not restricted and the problem is invariant under a location transformation, we can consider location equivariant estimators, which depend on x_1 and y_1 through $y_1 - x_1$. Thus, the risk function of the location equivariant estimator does not depend on μ. Then, the problem can be reduced to the estimation in the scale family with the restriction $\sigma > b_0$. Hence from Theorem 3.3, it follows that best equivariant estimator is minimax. This is summarized as follows.

Theorem 4.2 *Assume that μ is not restricted, but σ is restricted to $\sigma > b_0$. Then, the best equivariant estimator $\hat{q}^{BI}(t_x^{-1}t_y, u_y|u_x)$ is minimax in the estimation of the predictive density under the restricted parameter space.*

[3] Case of $a_0 > -\infty$ and $b_0 = 0$. This case implies that μ is restricted as $\mu > a_0$ and σ is not restricted. By considering $x' = x - a_0$, we can set $a_0 = 0$ without loss of generality and the problem becomes invariant (as in the previous case) under a scale transformation. We are thus led to the following.

Theorem 4.3 *Assume that σ is not restricted, but μ is such that $\mu \geq a_0$. Then, the best equivariant estimator $\hat{q}^{BI}(t_x^{-1}t_y, u_y|u_x)$ is minimax in the estimation of the predictive density under the restricted parameter space.*

4.2 Bounded case

Concerning the estimation of the predictive density, we have already seen that the best location equivariant estimator $\hat{q}^{BI}$ (Example 3.1 and Remark 3.1) is generally not minimax for estimating a location parameter bounded to a compact interval. However, the result of Kubokawa (2005) suggests minimaxity in the case of an unknown scale, and the following theorem shows that this suggestion is correct.

Let us consider the following restriction under the same location-scale families as treated in the previous subsection:

$$D = \{(\sigma, \mu) | a_1 < \mu < a_2,\ 0 < \sigma < b_0\},$$

where a_1 and a_2 are bounded constants and b_0 is a positive constant.

Theorem 4.4 *Assume that (μ, σ) is restricted to D. Then, the best equivariant estimator $\hat{q}^{BI}(t_x^{-1}t_y, u_y|u_x)$ is minimax for the estimation of the predictive density under the restricted parameter space.*

Proof. We shall check conditions **(A6)-(A6-3)** in Theorem 2.1. In this case, $P = \{(\sigma, \mu)|a_1 < \mu < a_2, 0 < \sigma < b_0\}$, $P_k = \{(\sigma, \mu)|a_1 < \mu < a_2, b_0/k < \sigma < b_0\}$ for $kb_0 > 1$, and $V(P_k) = (a_2 - a_1)\log k$. Take $\xi_1 = (2/\log k)\log(\sigma/b_0) + 1$ and $\xi_2 = \{2/(a_2 - a_1)\}\{\mu - (a_1 + a_2)/2\}$. Letting $\boldsymbol{\xi} = (\xi_1, \xi_2) = h_k((\sigma, \mu))$, we see that $h_k(P_k) = [-1, 1]^2$, $\gamma_k(\mathrm{d}\boldsymbol{\xi}) = \{(a_2 - a_1)\log k)/4\}\mathrm{d}\boldsymbol{\xi}$ and $\int_{h_k(P_k)} f(\xi_k)\gamma_k(\mathrm{d}\boldsymbol{\xi})/V(P_k) = (1/4)\int_{[-1,1]^2} f(\boldsymbol{\xi})\mathrm{d}\boldsymbol{\xi}$, which satisfies condition (A6-2). For any $\boldsymbol{\xi} \in [-1 + \varepsilon, 1 - \varepsilon]^2$, let $(b, a) = h_k^{-1}(\boldsymbol{\xi})$. Then, $b = b_0 k^{(\xi_1 - 1)/2}$ and $a = \{(a_2 - a_1)/2\}\xi_2 + (a_1 + a_2)/2$ so that $\{[h_k^{-1}(\xi)]^{-1}(\sigma, \mu); (\sigma, \mu) \in P_k\} = \{(\sigma/b, (\mu - a)/b); (\sigma, \mu) \in P_k\}$ and σ/b, $(\mu - a)/b$ satisfy the inequalities

$$k^{-(1+\xi_1)/2} < \frac{\sigma}{b} < k^{(1-\xi_1)/2},$$

$$-\frac{a_2 - a_1}{2b_0}(1 + \xi_2)k^{-(\xi_1 - 1)/2} < \frac{\mu - a}{b} < \frac{a_2 - a_1}{2b_0}(1 - \xi_2)k^{-(\xi_1 - 1)/2},$$

both of which are satisfied by $k^{-\varepsilon/2} < \sigma/b < k^{\varepsilon/2}$ and

$$-\frac{a_2 - a_1}{2b_0}\varepsilon k^{-\varepsilon/2} < \frac{\mu - a}{b} < \frac{a_2 - a_1}{2b_0}\varepsilon k^{-\varepsilon/2}.$$

Hence, condition **(A6-3)** is satisfied and the minimaxity of $\hat{q}^{BI}$ is established. ■

Note that minimaxity still holds under the restriction $D_0 = \{(\sigma, \mu)|a_1 < \mu < a_2,\ 0 < \sigma\}$. However, we could not show minimaxity for the restriction $D_1 = \{(\sigma, \mu)|a_1 < \mu < a_2,\ b_0 < \sigma\}$, since we cannot take a sequence so that the lower and upper bounds of $(\mu - a)/b$ can be expanded to the whole real line in the proof of Theorem 4.4. We conjecture that the best equivariant estimator is not minimax under the restriction D_1. From Kubokawa (2005), we also guess that the best equivariant estimator is not minimax for the restriction $\{(\sigma, \mu)|a_1 < \mu/\sigma < a_2, \sigma > 0\}$.

4.3 Multidimensional case

As an extension to a multidimensional model, we consider density functions of the forms $p(\boldsymbol{\sigma}^{-1}(\boldsymbol{t}_x - \boldsymbol{\mu}), \boldsymbol{\sigma}^{-1}\boldsymbol{s}_x|\boldsymbol{u}_x)p_x(\boldsymbol{u}_x)$ and $q(\boldsymbol{\sigma}^{-1}(\boldsymbol{t}_y - \boldsymbol{\mu}), \boldsymbol{\sigma}^{-1}\boldsymbol{s}_y|\boldsymbol{u}_y)q_y(\boldsymbol{u}_y)$ where $\boldsymbol{u}_x$ and $\boldsymbol{u}_y$ are location-scale invariant statistics,

$$\boldsymbol{\sigma}^{-1}(\boldsymbol{t}_x - \boldsymbol{\mu}) = \Big(\frac{t_{x,1} - \mu_1}{\sigma_1}, \dots, \frac{t_{x,p} - \mu_p}{\sigma_p}\Big) \quad \text{and} \quad \boldsymbol{\sigma}^{-1}\boldsymbol{s}_x = \Big(\frac{s_{x,1}}{\sigma_1}, \dots, \frac{s_{x,p}}{\sigma_p}\Big),$$

and $\boldsymbol{\sigma}^{-1}(\boldsymbol{t}_y - \boldsymbol{\mu})$ and $\boldsymbol{\sigma}^{-1}\boldsymbol{s}_y$ are defined similarly.

[1] Ordered restriction of locations. We first treat the constraint given by

$$M_1 = \{(\boldsymbol{\sigma}, \boldsymbol{\mu})|\boldsymbol{B}\boldsymbol{\mu} \le \boldsymbol{\alpha}, \sigma_1 = \cdots = \sigma_p = \sigma\},$$

where $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_q)'$ is a $q \times p$ known matrix for $q \le p$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)'$ is a known vector, and the inequality $\boldsymbol{B\mu} \le \boldsymbol{\alpha}$ means that $\boldsymbol{b}_i'\boldsymbol{\mu} \le \alpha_i$ for $i = 1, \ldots, q$. This restriction means that the location parameters are restricted to the polyhedral convex cone and includes the positive orthant restriction $\mu_i \ge 0$, $i = 1, \ldots, p$, the simple order restriction $\mu_1 \le \mu_2 \le \cdots \le \mu_p$, and the tree order restriction $\mu_1 \le \mu_i$, $i = 2, \ldots, k$.

Combining the arguments as in the proof of theorem 2.1 in Tsukuma and Kubokawa (2008) and the proof of Theorem 4.3, we can show the minimaxity of the best equivariant estimator.

Theorem 4.5 *Assume that $(\boldsymbol{\sigma}, \boldsymbol{\mu})$ is restricted to the polyhedral convex cone M_1 with unrestricted unknown scale σ. Then, the best equivariant estimator is minimax in the estimation of the predictive density under the restricted parameter space.*

[2] **Ordered restriction of scales.** We next consider the constraint given by

$$M_2 = \{(\boldsymbol{\sigma}, \boldsymbol{\mu}) | \boldsymbol{\mu} \in \mathbb{R}^p, \boldsymbol{B\eta} \le \boldsymbol{\alpha}\},$$

where $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)'$ for $\eta_i = \log \sigma_i$, and $\boldsymbol{B}$ and $\boldsymbol{\alpha}$ are the same as defined in M_1. This restriction means that $\boldsymbol{\eta}$ is restricted on the polyhedral convex cone and includes the positive orthant restriction $\sigma_i \ge 1$, $i = 1, \ldots, p$, the simple order restriction $\sigma_1 \le \sigma_2 \le \cdots \le \sigma_p$ and the tree order restriction $\sigma_1 \le \sigma_i$, $i = 2, \ldots, k$.

Since $\boldsymbol{\mu}$ is not restricted and the problem is invariant under location transformations, we can consider location equivariant estimators, which depend on $\boldsymbol{t}_x$ and $\boldsymbol{t}_y$ through $\boldsymbol{t}_y - \boldsymbol{t}_x$. Thus, the risk function of the location equivariant estimator does not depend on $\boldsymbol{\mu}$. Then, the problem can be reduced to estimation in the scale family with the restriction $\boldsymbol{B\eta} \le \boldsymbol{\alpha}$. Hence from the arguments as in the proof of Tsukuma and Kubokawa (2008), it follows that the best equivariant estimator is minimax.

Theorem 4.6 *Assume that $(\boldsymbol{\sigma}, \boldsymbol{\mu})$ is restricted into the polyhedral convex cone M_2 with unrestricted location parameters $\boldsymbol{\mu}$. Then, the best equivariant estimator is minimax in the estimation of the predictive density under the restricted parameter space.*

5 Concluding remarks

We have demonstrated that, for many restricted parameter space problems, the best equivariant predictive density $\hat{q}^{BI}$ under Kullback-Leibler loss remains minimax, with constant risk matching the minimax risk. We point out that versions of Theorem 2.1, 3.1, 3.3, and 4.1 also follow from the results of Marchand and Strawderman (2012).
For lower (or upper) bounded location or scale parameter problems, we have introduced a novel adaptation of Kubokawa's IERD technique to show that the generalized Bayes procedure $\hat{q}^U$ with respect to the truncation of the right Haar invariant measure onto the restricted parameter space dominates $\hat{q}^{BI}$ and is thus minimax. These findings are analogous to various point estimation results previously established. It seems plausible, but more research is required, that similar minimax results and $\hat{q}^{BI}$-$\hat{q}^U$ comparisons hold

for other choices of loss, such as for α-divergence losses (e.g., Csiszár, 1967; Corcuera and Guummole, 1999). Finally, further analysis of the efficiency of Bayes estimators for other restricted parameter spaces, such as for univariate compact interval restrictions (see the end of Example 3.1), represent challenging and interesting problems for further research.

Acknowledgments

The authors are grateful to the Associate Editor and referees for their valuable comments and helpful suggestions. Tatsuya Kubokawa's research is supported in part by Grant-in-Aid for Scientific Research Nos. 19200020 and 21540114 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).

References

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

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

[3] Casella, G. and Strawderman, W.E. (1981). Estimating a bounded mean. *Annals of Statistics*, **9**, 870-879.

[4] Corcuera, J. M. and Giummole, F. (1999). A generalized Bayes rule for prediction. *Scandinavian Journal of Statistics*, **26**, 265-279.

[5] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect obsevations. *Studia Sci. Math. Hungary*, **2**, 299-315.

[6] Farrell, R.H. (1964). Estimators of a location parameter in the absolutely continuous case. *Annals of Mathematical Statistics*, **35**, 949-998.

[7] 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.

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

[9] George, E. I., Liang, F. and Xu, X. (2012). From minimax shrinkage estimation to minimax shrinkage prediction. *Statistical Science*, **27**, 82-94.

[10] Girshick, M.A. and Savage, L.J. (1951). Bayes and minimax estimates for quadratic loss functions. In *Proc. Second Berkeley Symp. Math. Statist. Probab.*, **1**, 53-74. University of California Press, Berkeley.

[11] Hartigan, J. (2004). Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, **67**, 285-288.

[12] Hora, R.B., and Buehler, R.J. (1966). Fiducial theory and invariant estimation. *Annals of Mathematical Statistics*, **37**, 643-656.

[13] Hora, R.B., and Buehler, R.J. (1967). Fiducial theory and invariant prediction. *Annals of Mathematical Statistics*, **38**, 795-801.

[14] 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.

[15] Katz, M. (1961). Admissible and minimax estimates of parameters in truncated spaces. *Annals of Mathematical Statistics*, **32**, 136-142.

[16] Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Annals of Mathematical Statistics*, **28**, 573-601.

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

[18] Kubokawa, T. (1994A). A unified approach to improving equivariant estimators. *Annals of Statistics*, **22**, 290-299.

[19] Kubokawa, T. (1994B). Double shrinkage estimation of ratio of scale parameters. *Annals of the Institute of Statistical Mathematics*, **46**, 95-119.

[20] Kubokawa, T. (2004). Minimaxity in estimation of restricted parameters. *Journal of the Japanese Statistical Society*, **34**, 229-253.

[21] Kubokawa, T. (2005A). Estimation of a mean of a normal distribution with a bounded coefficient of variation. *Sankhyā*, **67**, 499-525.

[22] Kubokawa, T. (2005B). Estimation of bounded location and scale parameters. *Journal of the Japanese Statistical Society*, **35**, 221-249.

[23] Kubokawa, T. and Saleh, A.K.MD.E. (1998). Estimation of location and scale parameters under order restrictions. *Journal of Statistical Research*, **28**, 41-51.

[24] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation.* Springer-Verlag, 2nd edition.

[25] Liang, F. and Barron, A. (2004). Exact minimax strategies for predictive density estimation, data compression, and model selection. *IEEE Trans. Inform. Theory*, **50**, 2708-2726.

[26] Marchand, É., and Payandeh Najafabadi, A.T. (2011). Bayesian improvements of a MRE estimator of a bounded location parameter. *Electronic Journal of Statistics*, **5**, 1495-1502.

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

[28] Marchand, É. and Strawderman, W.E. (2004). Estimation in restricted parameter spaces: A review. *Festschrift for Herman Rubin*, IMS Lecture Notes-Monograph Series, **45**, 21-44.

[29] Marchand, É., and Strawderman, W. E. (2005A). On improving on the minimum risk equivariant estimator of a location parameter which is constrained to an interval or a half-interval. *Annals of the Institute of Statistical Mathematics*, **57**, 129-143.

[30] Marchand, É., and Strawderman, W. E. (2005B). On improving on the minimum risk equivariant estimator of a scale parameter under a lower bound constraint. *Journal of Statistical Planning and Inference*, **134**, 90-101.

[31] Marchand, É., and Strawderman, W.E. (2012). A unified minimax result for restricted parameter spaces. *Bernoulli*, **18**, 635-643.

[32] Tsukuma, H. and Kubokawa, T. (2008). Stein phenomenon in estimation of means restricted to a polyhedral convex cone. *Journal of Multivariate Analysis*, **99**, 141-164.

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

[34] Wijsman, R.A. (1985). A useful inequality on ratios of integrals, with application to maximum likelihood estimation. *Journal of the American Statistical Association*, **80**, 472-475.