Electronic Journal of Statistics
Vol. 12 (2018) 4209–4238
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1493

On predictive density estimation with additional information

Éric Marchand[1,*] **and Abdolnasser Sadeghkhani**[2,**]

[1] *Département de mathématiques, Université de Sherbrooke, Sherbrooke Qc, Canada, J1K 2R1, e-mail:* [*] eric.marchand@usherbrooke.ca

[2] *Brock University, Department of Mathematics & Statistics, St. Catharines (Ontario), L2S 3A1 Canada, e-mail:* [**] asadeghkhani@brocku.ca

Abstract: Based on independently distributed $X_1 \sim N_p(\theta_1, \sigma_1^2 I_p)$ and $X_2 \sim N_p(\theta_2, \sigma_2^2 I_p)$, we consider the efficiency of various predictive density estimators for $Y_1 \sim N_p(\theta_1, \sigma_Y^2 I_p)$, with the additional information $\theta_1 - \theta_2 \in A$ and known $\sigma_1^2, \sigma_2^2, \sigma_Y^2$. We provide improvements on benchmark predictive densities such as those obtained by *plug-in*, by maximum likelihood, or as minimum risk equivariant. Dominance results are obtained for $\alpha-$divergence losses and include Bayesian improvements for Kullback-Leibler (KL) loss in the univariate case ($p = 1$). An ensemble of techniques are exploited, including variance expansion, point estimation duality, and concave inequalities. Representations for Bayesian predictive densities, and in particular for $\hat{q}_{\pi_{U,A}}$ associated with a uniform prior for $\theta = (\theta_1, \theta_2)$ truncated to $\{\theta \in \mathbb{R}^{2p} : \theta_1 - \theta_2 \in A\}$, are established and are used for the Bayesian dominance findings. Finally and interestingly, these Bayesian predictive densities also relate to skew-normal distributions, as well as new forms of such distributions.

Keywords and phrases: Additional information, α-divergence loss, bayes estimators, dominance, duality, frequentist risk, Kullback-Leibler loss, multivariate normal, plug-in, predictive density, restricted parameter, skew-normal, variance expansion.

Received January 2018.

1. Introduction

1.1. Problem and model

Consider independently distributed

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_{2p}\left(\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 I_p & 0 \\ 0 & \sigma_2^2 I_p \end{pmatrix}\right), Y_1 \sim \mathrm{N}_p(\theta_1, \sigma_Y^2 I_p), \quad (1)$$

where $X_1, X_2, \theta_1, \theta_2$ are $p-$dimensional, and with the additional information (or constraint) $\theta_1 - \theta_2 \in A \subset \mathbb{R}^p$, A, σ_1^2, σ_2^2, σ_Y^2 all known, the variances not necessarily equal. We investigate how to gain from the additional information in providing a predictive density $\hat{q}(\cdot; X)$ as an estimate of the density $q_{\theta_1}(\cdot)$ of Y_1. Such a density is of interest as a surrogate for q_{θ_1}, as well as for generating either future or missing values of Y_1. The additional information $\theta_1 - \theta_2 \in A$ renders X_2 useful in estimating the density of Y_1 despite the independence and the otherwise unrelated parameters. Moreover, one can anticipate that the potential usefulness of X_2 is mitigated by its precision relative to X_1. For instance, if the ratio $\frac{\sigma_2^2}{\sigma_1^2} \approx 0$, then efficient inference about θ_1, or efficient predictive density estimation for Y_1, should be governed by the high probability that "$\theta_1 - X_2 \in A$". At the other extreme if $\frac{\sigma_2^2}{\sigma_1^2}$ is much larger than 1, little benefit of making of X_2 should be expected. And, for intermediate values of $\frac{\sigma_2^2}{\sigma_1^2}$, such as 1, inference about θ_1 or for Y_1 should be calibrated somewhat by the observed value of X_2. Much of the above discussion is relevant to situations where data about X_2 is relatively plentiful or cheap to acquire.

The reduced X data of the above model is pertinent to summaries X_1 and X_2 that arise through a sufficiency reduction, a large sample approximation, or limit theorems. Specific forms of A include:

- **(i)** order constraints $\theta_{1,i} - \theta_{2,i} \geq 0$ for $i = 1, \ldots, p$; the $\theta_{1,i}$ and $\theta_{2,i}$'s representing the components of θ_1 and θ_2;
- **(ii)** rectangular constraints $|\theta_{1,i} - \theta_{2,i}| \leq m_i$ for $i = 1, \ldots, p$;
- **(iii)** spherical constraints $\|\theta_1 - \theta_2\| \leq m$;
- **(iv)** order and bounded constraints $m_1 \geq \theta_{1,i} \geq \theta_{2,i} \geq m_2$ for $i = 1, \ldots, p$.

There is a very large literature on statistical inference in the presence of such constraints, mostly for **(i)** (e.g., Hwang and Peddada [22]; Dunson and Neelon [15]; Park, Kalbfleisch and Taylor [35]) among many others). Other sources on estimation in restricted parameter spaces can be found in the review paper of Marchand and Strawderman [33], as well as the monograph by van Eeden [45]. There exist various findings for estimation problems with additional information, dating back to Blumenthal and Cohen [8] and Cohen and Sackrowitz [12], with further contributions by van Eeden and Zidek [43, 44], Marchand et al. [30], Marchand and Strawderman [33].

Remark 1.1. *Our set-up applies to various other situations that can be transformed or reduced to model (1) with $\theta_1 - \theta_2 \in A$. Here are some examples.*

- **(I)** *Consider model (1) with the linear constrained $c_1\theta_1 - c_2\theta_2 + d \in A$, c_1, c_2 being constants not equal to 0, and $d \in \mathbb{R}^p$. Transforming $X_1' = c_1 X_1, X_2' = c_2 X_2 - d$, and $Y_1' = c_1 Y_1$ leads to model (1) based on the triplet (X_1', X_2', Y_1'), expectation parameters $\theta_1' = c_1\theta_1, \theta_2' = c_2\theta_2 - d$, covariance matrices $c_i^2\sigma_i^2 I_p, i = 1, 2$ and $c_1^2\sigma_Y^2 I_p$, and with the additional information $\theta_1' - \theta_2' \in A$. With the class of losses being intrinsic (see Remark 1.2), and the study of predictive density estimation for Y_1' equivalent to that for Y_1, our basic model and the findings below in this paper will indeed apply for linear constrained $c_1\theta_1 - c_2\theta_2 + d \in A$.*
- **(II)** *Consider a bivariate normal model for X with means θ_1, θ_2, variances σ_1^2, σ_2^2, correlation coefficient $\rho > 0$, and the additional information $\theta_1 - \theta_2 \in A$. The transformation $X_1' = X_1$, $X_2' = \frac{1}{\sqrt{1+\rho^2}}(X_2 - \frac{\rho\sigma_2}{\sigma_1} X_1)$ leads to independent coordinates with means $\theta_1' = \theta_1, \theta_2' = \frac{1}{\sqrt{1+\rho^2}}(\theta_2 - \frac{\rho\sigma_2}{\sigma_1}\theta_1)$, and variances σ_1^2, σ_2^2. We thus obtain model (1) for (X_1', X_2') with the additional information $\theta_1 - \theta_2 \in A$ transformed to $c_1\theta_1' - c_2\theta_2' + d \in A$, as*

in part **(I)** *above, with* $c_1 = 1 + \frac{\rho\sigma_2}{\sigma_1}$, $c_2 = \sqrt{1+\rho^2}$, *and* $d = 0$.

1.2. Predictive density estimation

Several loss functions are at our disposal to measure the efficiency of estimate $\hat{q}(\cdot;x)$, and these include the class of $\alpha-$divergence loss functions (e.g., Csiszàr [14]) given by

$$L_\alpha(\theta,\hat{q}) \;=\; \int_{\mathbb{R}^p} h_\alpha\left(\frac{\hat{q}(y;x)}{q_{\theta_1}(y)}\right)\; q_{\theta_1}(y)\,dy\,, \tag{2}$$

with

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

Notable examples in this class include Kullback-Leibler (h_{-1}), reverse Kullback-Leibler (h_1), and Hellinger ($h_0/4$). The cases $|\alpha| < 1$ merit study with many fewer results available, and stand apart in the sense that these losses are typically bounded, whereas Kullback-Leibler loss is typically unbounded (see Remark 4.2). For an above given loss, we measure the performance of a predictive density $\hat{q}(\cdot;X)$ by the frequentist risk

$$R_\alpha(\theta,\hat{q}) \;=\; \int_{\mathbb{R}^{2p}} L_\alpha\left(\theta,\hat{q}(\cdot;x)\right)\; p_\theta(x)\,dx\,, \tag{3}$$

p_θ representing the density of X.

Such a predictive density estimation framework was outlined for Kullback-Leibler loss in the pioneering work of Aitchison and Dunsmore [2], as well as Aitchison [1], and has found its way in many different fields of statistical science such as decision theory, information theory, econometrics, machine learning, image processing, and mathematical finance. There has been much recent Bayesian and decision theory analysis of predictive density estimators, in particular for multivariate normal or spherically symmetric settings, as witnessed by the work of Komaki [24], George, Liang and Xu [18], Brown, George and Xu [10], Kato [23], Fourdrinier et al. [17], Ghosh, Mergel and Datta [19], Maruyama and Strawderman [34], Kubokawa, Marchand and Strawderman [26, 27], among others.

Remark 1.2. *We point out that losses in (2) are intrinsic in the sense that predictive density estimates of the density of* $Y' = g(Y)$, *with invertible* $g : \mathbb{R}^p \to \mathbb{R}^p$ *and inverse Jacobian* J, *lead to an equivalent loss with the natural choice* $\hat{q}(g^{-1}(y');x)\,|J|$ *as*

$$\int_{\mathbb{R}^p} h_\alpha\left(\frac{\hat{q}(g^{-1}(y');x)\,|J|}{q_{\theta_1}(g^{-1}(y'))\,|J|}\right)\; q_{\theta_1}(g^{-1}(y'))\,|J|\,dy' \;=\; \int_{\mathbb{R}^p} h_\alpha\left(\frac{\hat{q}(y;x)}{q_{\theta_1}(y)}\right)\; q_{\theta_1}(y)\,dy\,,$$

which is indeed $L_\alpha(\theta,\hat{q})$ *independently of* g.

1.3. Description of main findings

In our framework, we study and compare the efficiency of various predictive density estimators such as: **(i)** *plug-in* densities $N_p(\hat{\theta}_1(X), \sigma_Y^2 I_p)$, which include the maximum likelihood predictive density estimator $\hat{q}_{mle}$ obtained by taking $\hat{\theta}_1(X)$ to be the restricted (i.e., under the constraint $\theta_1 - \theta_2 \in A$) maximum likelihood estimator (mle) of θ_1; **(ii)** minimum risk equivariant (MRE) predictive density $\hat{q}_{mre}$; **(iii)** variance expansions $N_p(\hat{\theta}_1(X), c\sigma_Y^2 I_p)$, with $c > 1$, of *plug-in* predictive densities; and **(iv)** Bayesian predictive densities with an emphasis on the uniform prior for θ truncated to the information set A. The predictive mle density $\hat{q}_{mle}$ is a natural benchmark exploiting the additional information, but not the chosen divergence loss. On the other hand, the predictive density $\hat{q}_{mre}$ does optimize in accordance to the loss function, but ignores the additional information. It remains nevertheless of interest as a benchmark and the degrees attainable by improvements inform us on the value of the additional information $\theta_1 - \theta_2 \in A$. Our findings focus, except for Section 2, on the frequentist risk performance (3) and related dominance results, for Kullback-Leibler divergence loss and other L_α losses with $-1 \leq \alpha < 1$, as well as for various types of information sets A. [1]

Sections 2 and 3 relate to the Bayesian predictive density $\hat{q}_{\pi_{U,A}}$ with respect to the uniform prior restricted to A. Section 2 presents various representations for $\hat{q}_{\pi_{U,A}}$, with examples connecting not only to known skewed-normal distributions, but also to seemingly new families of skewed-normal type distributions. Section 3 contains Bayesian dominance results for Kullback-Leibler loss. For $p = 1$, making use of Section 2's representations, we show that the Bayes predictive density $\hat{q}_{\pi_{U,A}}$ improves on $\hat{q}_{mre}$ under Kullback-Leibler loss for both $\theta_1 \geq \theta_2$ or $|\theta_1 - \theta_2| \leq m$. For the former case, the dominance result is further proven in Theorem 3.3 to be robust with respect to various misspecifications of σ_1^2, σ_2^2, and σ_Y^2.

Subsection 4.1 provides Kullback-Leibler improvements on *plug-in* densities by variance expansion. Such variance (or scale) expansions refer to $N_p(\hat{\theta}_1, c\sigma_Y^2 I_p)$ densities with $c > 1$, in other words with a greater scale than the target $N_p(\theta_1, \sigma_Y^2 I_p)$ density. We make use of a technique due to Fourdrinier et al. [17], which is universal with respect to p and A and requiring a determination, or lower-bound, of the infimum mean squared error of the plug-in estimator. Such a determination is facilitated by a mean squared error decomposition (Lemma 4.2) expressed in terms of the risk of a one-population restricted parameter space estimation problem.

The dominance results of Subsection 4.2 apply to L_α losses and exploit point estimation duality. The targeted predictive densities to be improved upon include *plug-in* densities, $\hat{q}_{mre}$, and more generally predictive densities of the form $\hat{q}_{\hat{\theta}_1,c} \sim N_p(\hat{\theta}_1(X), c\sigma_Y^2 I_p)$. The focus here is on improving on *plug-in* estimates $\hat{\theta}_1(X)$ by exploiting a correspondence with the problem of estimating θ_1 under

[1] We refer to Sadeghkhani [38] for various results pertaining to reverse Kullback-Leibler divergence loss, which we do not further study here.

a dual loss. Kullback-Leibler loss leads to dual mean squared error performance. In turn, as in Marchand and Strawderman [33], the above risk decomposition relates this performance to a restricted parameter space problem. Results for such problems are thus borrowable to infer dominance results for the original predictive density estimation problem. For other $\alpha-$divergence losses, the strategy is similar, with the added difficulty that the dual loss relates to a reflected normal loss. But, this is handled through a concave inequality technique (e.g., Kubokawa, Marchand and Strawderman [26]) relating risk comparisons to mean squared error comparisons. Several examples complement the presentation of Section 4. Finally, numerical illustrations and an application are presented and commented upon in Section 5.

2. Bayesian predictive density estimators and skewed normal type distributions

2.1. Bayesian predictive density estimators

We provide here a general representation of the Bayes predictive density estimator of the density of Y_1 in model (1) associated with a uniform prior on the additional information set A. Multivariate normal priors truncated to A are plausible choices that are also conjugate, lead to similar results, but will not be further considered here. Throughout this manuscript, we denote ϕ as the $N_p(0, I_p)$ p.d.f., and Φ as the $N(0,1)$ c.d.f.

Lemma 2.1. *Consider model (1) and the Bayes predictive density $\hat{q}_{\pi_{U,A}}$ with respect to the (uniform) prior $\pi_{U,A}(\theta) = \mathbb{I}_A(\theta_1 - \theta_2)$ for α-divergence loss L_α in (2). Then, for $-1 \leq \alpha < 1$, we have*

$$\hat{q}_{\pi_{U,A}}(y_1; x) \propto \hat{q}_{mre}(y_1; x_1)\, I^{\frac{2}{1-\alpha}}(y_1; x)\,, \tag{4}$$

with $\hat{q}_{mre}(y_1; x_1)$ the minimum risk predictive density based on x_1 given by a $N_p(x_1, (\sigma_1^2 \frac{(1-\alpha)}{2} + \sigma_Y^2) I_p)$ density, and $I(y_1; x) = \mathbb{P}(T \in A)$, with $T \sim N_p\left(\mu_T, \sigma_T^2 I_p\right)$, $\mu_T = \beta(y_1 - x_1) + (x_1 - x_2)$, $\sigma_T^2 = \frac{2\sigma_1^2\sigma_Y^2}{(1-\alpha)\sigma_1^2 + 2\sigma_Y^2} + \sigma_2^2$, and $\beta = \frac{(1-\alpha)\sigma_1^2}{(1-\alpha)\sigma_1^2 + 2\sigma_Y^2}$.

Proof. See Appendix. □

The general form of the Bayes predictive density estimator $\hat{q}_{\pi_{U,A}}$ is thus a weighted version of $\hat{q}_{\mathrm{mre}}$, with the weight a multivariate normal probability raised to the $2/(1-\alpha)^{th}$ power which is a function of y_1 and which depends on x, α, A. Observe that the representation applies in the trivial case $A = \mathbb{R}^p$, yielding $I \equiv 1$ and $\hat{q}_{\mathrm{mre}}$ as the Bayes estimator. As expanded on in Subsection 2.2, the densities $\hat{q}_{\pi_{U,A}}$ for Kullback-Leibler loss relate to skew-normal distributions, and more generally to skewed distributions arising from selection (see for instance Arnold and Beaver [4]; Arellano-Valle, Branco and Genton [3]; among others). Moreover, it is known (e.g. Liseo and Loperfido [29]) that posterior distributions present here also relate to such skew-normal type distributions.

Lemma 2.1 does not address the evaluation of the normalization constant for the Bayes predictive density $\hat{q}_{\pi_{U,A}}$, but we now proceed with this for the particular cases of Kullback-Leibler and Hellinger losses, and more generally for cases where $\frac{2}{1-\alpha}$ is a positive integer, i.e., $\alpha = 1 - \frac{2}{n}$ where $n = 1, 2, \ldots$. In what follows, we denote 1_m as the m dimensional column vector with components equal to 1, and $\otimes$ as the usual Kronecker product.

Lemma 2.2. *For model (1), $\alpha-$divergence loss with $n = \frac{2}{1-\alpha} \in \{1, 2, \ldots\}$, the Bayes predictive density $\hat{q}_{\pi_{U,A}}(y_1; x)\,, y_1 \in \mathbb{R}^p$, with respect to the (uniform) prior $\pi_{U,A}(\theta) = \mathbb{I}_A(\theta_1 - \theta_2)$, is given by*

$$\hat{q}_{\pi_{U,A}}(y_1; x) \, = \hat{q}_{mre}(y_1; x_1) \, \frac{\{\mathbb{P}(T \in A)\}^n}{\mathbb{P}(\cap_{i=1}^n \{Z_i \in A\})} \,, \tag{5}$$

with $\hat{q}_{\mathrm{mre}}(y_1; x_1)$ a $N_p(x_1, (\sigma_1^2/n + \sigma_Y^2)\mathrm{I}_p)$ density, $T \sim N_p(\mu_T, \sigma_T^2 I_p)$ with $\mu_T = \beta(y_1 - x_1) + (x_1 - x_2)$, $\sigma_T^2 = \sigma_2^2 + n\sigma_Y^2\beta$, $\beta = \frac{\sigma_1^2}{\sigma_1^2 + n\sigma_Y^2}$, and $Z = (Z_1, \ldots, Z_n)' \sim N_{np}(\mu_Z, \Sigma_Z)$ with $\mu_Z = 1_n \otimes (x_1 - x_2)$ and $\Sigma_Z = (\sigma_T^2 + \sigma_Y^2\beta^2)I_{np} + (\frac{\beta^2\sigma_1^2}{n} 1_n 1_n' \otimes \mathrm{I}_p)\,.$

Remark 2.1. *The Kullback-Leibler case corresponds to $n = 1$ and the above form of the Bayes predictive density simplifies to*

$$\hat{q}_{\pi_{U,A}}(y_1; x) \, = \hat{q}_{mre}(y_1; x_1) \, \frac{\mathbb{P}(T \in A)}{\mathbb{P}(Z_1 \in A)} \,, \tag{6}$$

with $\hat{q}_{mre}(y_1; x_1)$ a $N_p(x_1, (\sigma_1^2 + \sigma_Y^2)I_p)$ density, $T \sim N_p(\mu_T, \sigma_T^2 I_p)$ with $\mu_T = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_Y^2}(y_1 - x_1) + (x_1 - x_2)$ and $\sigma_T^2 = \frac{\sigma_1^2\sigma_Y^2}{\sigma_1^2 + \sigma_Y^2} + \sigma_2^2$, and $Z_1 \sim N_p(x_1 - x_2, (\sigma_1^2 + \sigma_2^2)I_p)$. In the univariate case (i.e., $p = 1$), T is univariate normally distributed, and the expectation and covariance matrix of Z simplify to $1_n(x_1 - x_2)$ and $(\sigma_T^2 + \sigma_Y^2\beta^2)I_n + \beta^2\frac{\sigma_1^2}{n} 1_n 1_n'$ respectively. Finally, we point out that the diagonal elements of Σ_Z simplify to $\sigma_1^2 + \sigma_2^2$, a result which will arise below several times.

Proof of Lemma 2.2. It suffices to evaluate the normalization constant (say C) for the predictive density in (4). We have

$$\begin{aligned} C &= \int_{\mathbb{R}^p} \hat{q}_{\mathrm{mre}}(y_1; x_1) \, \{\mathbb{P}(T \in A)\}^n \, dy_1 \\ &= \int_{\mathbb{R}^p} \hat{q}_{\mathrm{mre}}(y_1; x_1) \, \mathbb{P}\left(\cap_{i=1}^n \{T_i \in A\}\right) \, dy_1 \,, \end{aligned}$$

with $T_1, \ldots, T_n$ independent copies of T. With the change of variables $u_0 = \frac{y_1 - x_1}{\sqrt{\sigma_1^2/n + \sigma_Y^2}}$ and letting $U_0, U_1, \ldots, U_n$ i.i.d. $N_p(0, I_p)$, we obtain

$$\begin{aligned} C &= \int_{\mathbb{R}^p} \phi(u_0) \, \mathbb{P}\left(\cap_{i=1}^n \{\sigma_T U_i + \beta u_0 \sqrt{\sigma_1^2/n + \sigma_Y^2} + x_1 - x_2\} \in A\right) \, du_0 \\ &= \mathbb{P}\left(\cap_{i=1}^n \{\sigma_T U_i + \beta U_0 \sqrt{\sigma_1^2/n + \sigma_Y^2} + x_1 - x_2\} \in A\right) \,, \\ &= \mathbb{P}\left(\cap_{i=1}^n \{Z_i \in A\}\right) . \end{aligned}$$

The result follows by verifying that the expectation and covariance matrix of $Z = (Z_1, \ldots, Z_n)'$ are as stated. $\square$

We conclude this section on a posterior distribution decomposition and with an accompanying representation of the posterior expectation $\mathbb{E}(\theta_1|x)$ in terms of a truncated multivariate normal expectation. The latter coincides with the expectation under the Bayes Kullback-Leibler predictive density $\hat{q}_{\pi_{U,A}}$.

Lemma 2.3. *Consider $X|\theta$ as in model (1) and the uniform prior $\pi_{U,A}(\theta) = \mathbb{I}_A(\theta_1 - \theta_2)$. Set $r = \frac{\sigma_2^2}{\sigma_1^2}$, $\omega_1 = \theta_1 - \theta_2$, and $\omega_2 = r\theta_1 + \theta_2$. Then, conditional on $X = x$, ω_1 and ω_2 are independently distributed with*

$$\omega_1 \sim N_p(\mu_{\omega_1}, \tau^2_{\omega_1}) \text{ truncated to } A, \qquad \omega_2 \sim N_p(\mu_{\omega_2}, \tau^2_{\omega_2}),$$

$\mu_{\omega_1} = x_1 - x_2$, $\mu_{\omega_2} = rx_1 + x_2$, $\tau^2_{\omega_1} = \sigma_1^2 + \sigma_2^2$, and $\tau^2_{\omega_2} = 2\sigma_2^2$. Furthermore, we have $\mathbb{E}(\theta_1|x) = \frac{1}{1+r}\left(\mathbb{E}(\omega_1|x) + \mu_{\omega_2}\right)$.

Proof. With the posterior density $\pi(\theta|x) \propto \phi(\frac{\theta_1 - x_1}{\sigma_1})\, \phi(\frac{\theta_2 - x_2}{\sigma_2})\, \mathbb{I}_A(\theta_1 - \theta_2)$, the result follows by transforming to (ω_1, ω_2). $\square$

2.2. Examples of Bayesian predictive density estimators

With the presentation of the Bayes predictive estimator $\hat{q}_{\pi_{U,A}}$ in Lemmas 2.1 and 2.2, which is quite general with respect to the dimension p, the additional information set A, and the $\alpha-$divergence loss, it is pertinent and instructive to continue with some illustrations. Moreover, various skewed-normal or skewed-normal type, including new extensions, arise as predictive density estimators. Such distributions have indeed generated much interest for the last thirty years or so, and continue to do so, as witnessed by the large literature devoted to their study. The most familiar choices of $\alpha-$divergence loss are Kullback-Leibler and Hellinger (i.e., $n = \frac{2}{1-\alpha} = 1, 2$ below) but the form of the Bayes predictive density estimator $\hat{q}_{\pi_{U,A}}$ is nevertheless expanded upon below in the context of Lemma 2.2, in view of the connections with an extended family of skewed-normal type distributions (e.g., Definition 2.1), which is also of independent interest. Subsections 2.2.1, 2.2.2, and 2.2.3. deal with Kullback-Leibler and $\alpha-$divergence losses for situations: (i) $p = 1, A = \mathbb{R}_+$; (ii) $p = 1, A = [-m, m]$; and (iii) $p \geq 1$ and A a ball of radius m centered at the origin.

2.2.1. Univariate case with $\theta_1 \geq \theta_2$

From (5), we obtain for $p = 1, A = \mathbb{R}_+$: $\mathbb{P}(T \in A) = \Phi(\frac{\mu_T}{\sigma_T})$ and

$$\hat{q}_{\pi_{U,A}}(y_1; x) \propto \frac{1}{\sqrt{\sigma_1^2/n + \sigma_Y^2}}\, \phi\Big(\frac{y_1 - x_1}{\sqrt{\sigma_1^2/n + \sigma_Y^2}}\Big)\, \Phi^n\Big(\frac{\beta(y_1 - x_1) + (x_1 - x_2)}{\sigma_T}\Big), \tag{7}$$

with β and σ_T^2 given in Lemma 2.2. These densities are of the following form.

Definition 2.1. *A generalized Balakrishnan type skewed-normal distribution, with shape parameters $n \in \mathbb{N}_+, \alpha_0, \alpha_1 \in \mathbb{R}$, location and scale parameters $\xi \in \mathbb{R}$ and $\tau \in \mathbb{R}_+$, denoted $\mathrm{SN}(n, \alpha_0, \alpha_1, \xi, \tau)$, has density on $\mathbb{R}$ given by*

$$\frac{1}{K_n(\alpha_0, \alpha_1)} \frac{1}{\tau} \phi(\frac{t-\xi}{\tau}) \Phi^n(\alpha_0 + \alpha_1 \frac{t-\xi}{\tau}), \tag{8}$$

with

$$K_n(\alpha_0, \alpha_1) = \Phi_n \left(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}}, \cdots, \frac{\alpha_0}{\sqrt{1+\alpha_1^2}}; \rho = \frac{\alpha_1^2}{1+\alpha_1^2} \right), \tag{9}$$

$\Phi_n(\cdot;\rho)$ representing the cdf of a $\mathrm{N}_n(0, \Lambda)$ distribution with covariance matrix $\Lambda = (1-\rho) I_n + \rho \, 1_n 1_n'$.

Remark 2.2. *(The case $n = 1$)*
$\mathrm{SN}(1, \alpha_0, \alpha_1, \xi, \tau)$ densities are given by (8) with $n = 1$ and $K_1(\alpha_0, \alpha_1) = \Phi(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}})$. Properties of $\mathrm{SN}(1, \alpha_0, \alpha_1, \xi, \tau)$ distributions were described by Arnold et al. [5], as well as Arnold and Beaver [4], with the particular case $\alpha_0 = 0$ reducing to the original skew normal density, modulo a location-scale transformation, as presented in Azzalini's seminal 1985 paper. Namely, the expectation of $T \sim \mathrm{SN}(1, \alpha_0, \alpha_1, \xi, \tau)$ is given by

$$\mathbb{E}(T) \; = \; \xi + \tau \frac{\alpha_1}{\sqrt{1+\alpha_1^2}} \, R(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}}), \tag{10}$$

with $R =: \frac{\phi}{\Phi}$ known as the inverse Mill's ratio.

Remark 2.3. *For $\alpha_0 = 0, n = 2, 3, \ldots$, the densities were proposed by Balakrishnan as a discussant of Arnold and Beaver [4], and further analyzed by Gupta and Gupta [20]. We are not aware of an explicit treatment of such distributions in the general case, but standard techniques may be used to derive the following properties. For instance, as handled more generally above in the proof of Lemma 2.2, the normalization constant K_n may be expressed in terms of a multivariate normal c.d.f. by observing that*

$$\begin{aligned} K_n(\alpha_0, \alpha_1) &= \int_{\mathbb{R}} \phi(z) \Phi^n(\alpha_0 + \alpha_1 z) \, dz \\ &= \mathbb{P}(\cap_{i=1}^n \{U_i \leq \alpha_0 + \alpha_1 U_0\}) \\ &= \mathbb{P}(\cap_{i=1}^n \{W_i \leq \frac{\alpha_0}{\sqrt{1+\alpha_1^2}}\}), \end{aligned} \tag{11}$$

with $(U_0, \ldots, U_n) \sim N_{n+1}(0, I_{n+1})$, $W_i \overset{d}{=} \frac{U_i - \alpha_1 U_0}{\sqrt{1+\alpha_1^2}}$, for $i = 1, \ldots, n$, and $(W_1, \ldots, W_n) \sim \mathrm{N}_n(0, \Lambda)$.

In terms of expectation, we have, for $T \sim SN(n, \alpha_0, \alpha_1, \xi, \tau)$, $\mathbb{E}(T) = \xi + \tau \mathbb{E}(W)$ where $W \sim SN(n, \alpha_0, \alpha_1, 0, 1)$ and

$$\mathbb{E}(W) \; = \; \frac{n\alpha_1}{\sqrt{1+\alpha_1^2}} \, \phi(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}}) \, \frac{K_{n-1}(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}}, \frac{\alpha_1}{\sqrt{1+\alpha_1^2}})}{K_n(\alpha_0, \alpha_1)}. \tag{12}$$

This can be obtained via Stein's identity $\mathbb{E}\,Ug(U) = \mathbb{E}g'(U)$ for differentiable g and $U \sim N(0,1)$. Indeed, we have

$$\int_{\mathbb{R}} u\phi(u)\,\Phi^n(\alpha_0+\alpha_1 u)\,du \quad = \quad n\alpha_1 \int_{\mathbb{R}} \phi(u)\phi(\alpha_0+\alpha_1 u)\,\Phi^{n-1}(\alpha_0+\alpha_1 u)\,du\,,$$

and the result follows by using the identity $\phi(u)\,\phi(\alpha_0+\alpha_1 u) = \phi(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}})\,\phi(v)$, with $v = \sqrt{1+\alpha_1^2}\,u + \frac{\alpha_0\alpha_1}{\sqrt{1+\alpha_1^2}}$, the change of variables $u \to v$, and the definition of K_{n-1}.

The connection between the densities of Definition 2.1 and the predictive densities in (7) is thus explicitly stated as follows, with the Kullback-Leibler and Hellinger cases corresponding to $n = 1, 2$ respectively.

Corollary 2.1. *For $p = 1, A = \mathbb{R}_+$, $\pi_{U,A}(\theta) = \mathbb{I}_A(\theta_1 - \theta_2)$, the Bayes predictive density estimator $\hat{q}_{\pi_{U,A}}$ under $\alpha-$divergence loss, with $n = \frac{2}{1-\alpha} \in \mathbb{N}_+$ positive integer, is given by a $\mathrm{SN}(n, \alpha_0 = \frac{x_1-x_2}{\sigma_T}, \alpha_1 = \frac{\beta\tau}{\sigma_T}, \xi = x_1, \tau = \sqrt{\frac{\sigma_1^2}{n} + \sigma_Y^2})$ density, with $\sigma_T^2 = \sigma_2^2 + n\beta\sigma_Y^2$ and $\beta = \frac{\sigma_1^2}{\sigma_1^2 + n\sigma_Y^2}$.*

Remark 2.4. **(a)** *For the equal variances case with $\sigma_1^2 = \sigma_2^2 = \sigma_Y^2 = \sigma^2$, the above predictive density estimator is a $\mathrm{SN}(n, \alpha_0 = \sqrt{\frac{n+1}{(2n+1)\sigma}}(x_1 - x_2), \alpha_1 = \sqrt{\frac{1}{n(2n+1)}}, \xi = x_1, \tau = \sqrt{\frac{n+1}{n}}\sigma)$ density.*

(b) *Under the conditions of Corollary 2.1 with $A = \mathbb{R}_-$ instead, an analogous calculation yields a $\mathrm{SN}(n, \alpha_0 = \frac{x_2-x_1}{\sigma_T}, \alpha_1 = -\frac{\beta\tau}{\sigma_T}, \xi = x_1, \tau = \sqrt{\frac{\sigma_1^2}{n} + \sigma_Y^2})$ density as the Bayes density $\hat{q}_{\pi_{U,A}}$.*

2.2.2. Univariate case with $|\theta_1 - \theta_2| \le m$

From (5), we obtain for $p = 1, A = [-m, m]$: $\mathbb{P}(T \in A) = \Phi(\frac{\mu_T+m}{\sigma_T}) - \Phi(\frac{\mu_T-m}{\sigma_T})$, and we may write

$$\hat{q}_{\pi_{U,A}}(y_1; x) = \frac{1}{\tau}\phi(\frac{t-\xi}{\tau})\;\frac{\{\Phi(\alpha_0+\alpha_1\,\frac{t-\xi}{\tau}) - \Phi(\alpha_2+\alpha_1\,\frac{t-\xi}{\tau})\}^n}{J_n(\alpha_0,\alpha_1,\alpha_2)}\,, \tag{13}$$

with $\xi = x_1, \tau = \sqrt{\sigma_1^2/n + \sigma_Y^2}, \alpha_0 = \frac{x_1-x_2+m}{\sigma_T}$, $\alpha_1 = \frac{\beta\tau}{\sigma_T}$ $\alpha_2 = \frac{x_1-x_2-m}{\sigma_T}$, β, μ_T, and σ_T^2 given in Lemma 2.2, and $J_n(\alpha_0,\alpha_1,\alpha_2)$ (independent of ξ,τ) a special case of the normalization constant given in (5). For fixed n, the densities in (13) form a five-parameter family of densities with location and scale parameters $\xi \in \mathbb{R}$ and $\tau \in \mathbb{R}_+$, and shape parameters $\alpha_0, \alpha_1, \alpha_2 \in \mathbb{R}$ such that $\alpha_0 > \alpha_2$. The Kullback-Leibler predictive densities ($n = 1$) match densities introduced by Arnold et al. [5] with the normalization constant in (13) simplifying to:

$$\begin{aligned} J_1(\alpha_0,\alpha_1,\alpha_2) &= \Phi(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}}) - \Phi(\frac{\alpha_2}{\sqrt{1+\alpha_1^2}}) \\ &= \Phi(\frac{m-(x_1-x_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}) - \Phi(\frac{-m-(x_1-x_2)}{\sqrt{\sigma_1^2+\sigma_2^2}})\,. \end{aligned} \tag{14}$$

The corresponding expectation which is readily obtained as in (10) equals

$$\begin{aligned} \mathbb{E}(T) &= \xi + \tau \frac{\alpha_1}{\sqrt{1+\alpha_1^2}} \frac{\phi(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}}) - \phi(\frac{\alpha_2}{\sqrt{1+\alpha_1^2}})}{\Phi(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}}) - \Phi(\frac{\alpha_2}{\sqrt{1+\alpha_1^2}})} \\ &= x_1 + \frac{\sigma_1^2}{\sqrt{\sigma_1^2+\sigma_2^2}} \frac{\phi(\frac{x_1-x_2+m}{\sqrt{\sigma_1^2+\sigma_2^2}}) - \phi(\frac{x_1-x_2-m}{\sqrt{\sigma_1^2+\sigma_2^2}})}{\Phi(\frac{x_1-x_2+m}{\sqrt{\sigma_1^2+\sigma_2^2}}) - \Phi(\frac{x_1-x_2-m}{\sqrt{\sigma_1^2+\sigma_2^2}})}\,, \end{aligned} \tag{15}$$

by using the above values of $\xi, \tau, \alpha_0, \alpha_1, \alpha_2$.

Hellinger loss yields the Bayes predictive density in (13) with $n = 2$, and a calculation as in Remark 2.3 leads to the evaluation

$$J_2(\alpha_0,\alpha_1,\alpha_2) \;=\; \Phi_2(\alpha_0',\alpha_0';\alpha_1') + \Phi_2(\alpha_2',\alpha_2';\alpha_1') - 2\Phi_2(\alpha_0',\alpha_2';\alpha_1')$$

with $\alpha_i' = \frac{\alpha_i}{\sqrt{1+\alpha_1^2}}$ for $i = 0, 1, 2$.

2.2.3. *Multivariate case with* $||\theta_1 - \theta_2|| \leq m$

For $p \geq 1$, the ball $A = \{t \in \mathbb{R}^p : ||t|| \leq m\}$, μ_T and σ_T^2 as given in Lemma 13, the Bayes predictive density in (5) under $\alpha-$divergence loss with $\frac{2}{1-\alpha} = n \in \mathbb{N}_+$ is expressible as

$$\hat{q}_{\pi_{U,A}} \;\propto\; \hat{q}_{\text{mre}}(y_1;x_1)\,\{\mathbb{P}(||T||^2 \leq m^2)\}^n$$

with $T \sim \sigma_T^2 \chi_p^2(\|\mu_T\|^2/\sigma_T^2)$, i.e., the weight attached to $\hat{q}_{\text{mre}}$ is proportional to the n^{th} power of the c.d.f. of a non-central chi-square distribution. For Kullback-Leibler loss, we obtain from (5)

$$\begin{aligned} \hat{q}_{\pi_{U,A}}(y_1;x) &= \hat{q}_{\text{mre}}(y_1;x_1)\,\frac{\mathbb{P}(||T||^2 \leq m^2)}{\mathbb{P}(||Z_1||^2 \leq m^2)} \\ &= \hat{q}_{\text{mre}}(y_1;x_1)\,\frac{\mathbb{F}_{p,\lambda_1(x,y_1)}(m^2/\sigma_T^2)}{\mathbb{F}_{p,\lambda_2(x)}(m^2/(\sigma_1^2+\sigma_2^2))}\,, \end{aligned} \tag{16}$$

where $F_{p,\lambda}$ represents the c.d.f. of a $\chi_p^2(\lambda)$ distribution, $\lambda_1(x,y_1) = \frac{\|\mu_T\|^2}{\sigma_T^2} = \frac{\|\beta(y_1-x_1)+(x_1-x_2)\|^2}{\sigma_T^2}$; with $\beta = \frac{\sigma_1^2}{\sigma_1^2+\sigma_Y^2}$, $\sigma_T^2 = \sigma_2^2 + \beta\sigma_Y^2$, and $\lambda_2(x) = \frac{\|x_1-x_2\|^2}{\sigma_1^2+\sigma_2^2}$. Observe that the non-centrality parameters λ_1 and λ_2 are random, and themselves non-central chi-square distributed as $\lambda_1(X,Y_1) \sim \chi_p^2(\frac{||\theta_1-\theta_2||^2}{\sigma_T^2})$ and $\lambda_2(X) \sim \chi_p^2(\frac{||\theta_1-\theta_2||^2}{\sigma_1^2+\sigma_2^2})$. Of course, the above predictive density (16) matches the Kullback-Leibler predictive density given in (13) for $n = 1$, and represents an otherwise interesting multivariate extension.

3. Bayesian dominance results

We focus here on Bayesian improvements for Kullback-Leibler divergence loss of the benchmark minimum risk equivariant predictive density. We establish that the uniform Bayes predictive density estimator $\hat{q}_{\pi_{U,A}}$ dominates $\hat{q}_{mre}$ for the univariate cases where $\theta_1 - \theta_2$ is either restricted to a compact interval, lower-bounded, or upper-bounded. We also investigate situations where the variances of model (1) are misspecified, but where the dominance persists. There is no loss in generality in taking $A = [-m, m]$ in the former case, and $A = [0, \infty)$ for the latter two cases. We begin with the lower bounded case.

Theorem 3.1. *Consider model (1) with $p = 1$ and $A = [0, \infty)$. For estimating the density of Y_1 under Kullback-Leibler loss, the Bayes predictive density $\hat{q}_{\pi_{U,A}}$ dominates the minimum risk equivariant predictive density estimator $\hat{q}_{mre}$. The Kullback-Leibler risks are equal iff $\theta_1 = \theta_2$.*

Proof. Making use of Corollary 2.1's representation of $\hat{q}_{\pi_{U,A}}$, the difference in risks is given by

$$\begin{aligned}
\Delta(\theta) &= R_{KL}(\theta, \hat{q}\mathrm{mre}) - R_{KL}(\theta, \hat{q}_{\pi_{U,A}}) \\
&= \mathbb{E}^{X,Y_1} \log\left(\frac{\hat{q}_{\pi_{U,A}}(Y_1; X)}{\hat{q}\mathrm{mre}(Y_1; X)}\right) \\
&= \mathbb{E}^{X,Y_1} \log\left(\Phi(\alpha_0 + \alpha_1 \frac{Y_1 - X_1}{\tau})\right) - \mathbb{E}^{X,Y_1} \log\left(\Phi(\frac{\alpha_0}{\sqrt{1+\alpha^2}})\right) \qquad (17)
\end{aligned}$$

with $\alpha_0 = \frac{X_1 - X_2}{\sigma_T}$, $\alpha_1 = \frac{\beta\tau}{\sigma_T}$, $\tau = \sqrt{\sigma_1^2 + \sigma_Y^2}$, $\beta = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_Y^2}$, and $\sigma_T^2 = \sigma_2^2 + \beta\sigma_Y^2$. Now, observe that

$$\alpha_0 + \alpha_1 \frac{Y_1 - X_1}{\tau} = \frac{X_1 - X_2 + \beta(Y_1 - X_1)}{\sigma_T} \sim N(\frac{\theta_1 - \theta_2}{\sigma_T}, 1)\,, \qquad (18)$$

and

$$\frac{\alpha_0}{\sqrt{1+\alpha_1^2}} = \frac{X_1 - X_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sim N(\frac{\theta_1 - \theta_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}, 1)\,. \qquad (19)$$

We thus can write

$$\Delta(\theta) \;=\; \mathbb{E}\, G(Z)\,,$$

$$\text{with } G(Z) = \log\Phi(Z + \frac{\theta_1 - \theta_2}{\sigma_T}) - \log\Phi(Z + \frac{\theta_1 - \theta_2}{\sqrt{\sigma_1^2 + \sigma_2^2}})\,, Z \sim N(0,1)\,.$$

With $\theta_1 - \theta_2 \geq 0$ and $\sigma_T^2 < \sigma_1^2 + \sigma_2^2$, we infer that $\mathbb{P}_\theta(G(Z) \geq 0) = 1$ and $\Delta(\theta) \geq 0$ for all θ such that $|\theta_1 - \theta_2| \leq m$, with equality iff $\theta_1 - \theta_2 = 0$. ☐

We now obtain an analogue dominance result in the univariate case for the additional information $\theta_1 - \theta_2 \in [-m, m]$.

Theorem 3.2. *Consider model (1) with $p = 1$ and $A = [-m, m]$. For estimating the density of Y_1 under Kullback-Leibler loss, the Bayes predictive density $\hat{q}_{\pi_{U,A}}$ (strictly) dominates the minimum risk equivariant predictive density estimator $\hat{q}_{mre}$.*

Proof. Making use of (13) and (14) for the representation of $\hat{q}_{\pi_{U,A}}$, the difference in risks is given by

$$\begin{aligned}\Delta(\theta) &= R_{KL}(\theta, \hat{q}_{\mathrm{mre}}) - R_{KL}(\theta, \hat{q}_{\pi_{U,A}}) \\ &= \mathbb{E}^{X,Y_1} \log\left(\frac{\hat{q}_{\pi_{U,A}}(Y_1;X)}{\hat{q}_{\mathrm{mre}}(Y_1;X)}\right) \\ &= \mathbb{E}^{X,Y_1} \log\left(\Phi(\alpha_0 + \alpha_1\frac{Y_1 - X_1}{\tau}) - \Phi(\alpha_2 + \alpha_1\frac{Y_1 - X_1}{\tau})\right) \\ &- \mathbb{E}^{X,Y_1} \log\left(\Phi(\frac{\alpha_0}{\sqrt{1+\alpha_1^2}}) - \Phi(\frac{\alpha_2}{\sqrt{1+\alpha_1^2}})\right)\end{aligned}$$

with the α_i's given in Section 2.2. Now, observe that

$$\alpha_0 + \alpha_1\frac{Y_1 - X_1}{\tau} = \frac{m + X_1 - X_2 + \beta(Y_1 - X_1)}{\sigma_T} \sim N(\delta_0 = \frac{m + \theta_1 - \theta_2}{\sigma_T}, 1)\,, \tag{20}$$

and

$$\frac{\alpha_0}{\sqrt{1+\alpha_1^2}} = \frac{m + (X_1 - X_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sim N(\delta_0' = \frac{m + \theta_1 - \theta_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}, 1)\,. \tag{21}$$

Similarly, we have $\alpha_2 + \alpha_1 \frac{Y_1 - X_1}{\tau} \sim N(\delta_2 = \frac{-m+\theta_1-\theta_2}{\sigma_T}, 1)$ and $\frac{\alpha_2}{\sqrt{1+\alpha_1^2}} \sim N(\delta_2' = \frac{-m+\theta_1-\theta_2}{\sqrt{\sigma_1^2+\sigma_2^2}}, 1)$. We thus can write for $Z \sim N(0,1)$

$$\Delta(\theta) = \mathbb{E}H(Z)\,,$$

$$\text{with } H(Z) = \log\left(\Phi(Z+\delta_0) - \Phi(Z+\delta_2)\right) - \log\left(\Phi(Z+\delta_0') - \Phi(Z+\delta_2')\right)\,.$$

With $-m \leq \theta_1 - \theta_2 \leq m$ and $\sigma_T^2 < \sigma_1^2 + \sigma_2^2$, we infer that $\delta_0 \geq \delta_0'$ with equality iff $\theta_1 - \theta_2 = -m$ and $\delta_2 \leq \delta_2'$ with equality iff $\theta_1 - \theta_2 = m$, so that $\mathbb{P}_\theta(H(Z) > 0) = 1$ and $\Delta(\theta) > 0$ for all θ such that $|\theta_1 - \theta_2| \leq m$. □

We now investigate situations where the variances in model (1) are misspecified. To this end, we consider σ_1^2, σ_2^2 and σ_Y^2 as the nominal variances used to construct the predictive density estimates $\hat{q}_{\pi_{U,A}}$ and $\hat{q}_{\mathrm{mre}}$, while the true variances, used to assess frequentist Kullback-Leibler risk, are, unbeknownst to the investigator, given by $a_1^2\sigma_1^2$, $a_2^2\sigma_2^2$ and $a_Y^2\sigma_Y^2$ respectively. We exhibit, below in Theorem 3.3, many combinations of the nominal and true variances such that the Theorem 3.1's dominance result persists. Such conditions for the dominance to persist includes the case of equal a_1^2, a_2^2 and a_Y^2 (i.e., the three ratios true variance over nominal variance are the same), among others.

We require the following intermediate result.

Lemma 3.1. *Let $U \sim N(\mu_U, \sigma_U^2)$ and $V \sim N(\mu_V, \sigma_V^2)$ with $\mu_U \geq \mu_V$ and $\sigma_U^2 \leq \sigma_V^2$. Let H be a differentiable function such that both H and $-H'$ are increasing. Then, we have $\mathbb{E}H(U) \geq \mathbb{E}H(V)$.*

Proof. Suppose without loss of generality that $\mu_V = 0$, and set $s = \frac{\sigma_U}{\sigma_V}$. Since U and $\mu_U + sV$ share the same distribution and $\mu_U \geq 0$, we have:

$$\begin{aligned}
\mathbb{E}H(U) &= \mathbb{E}H(\mu_U + sV) \\
&\geq \mathbb{E}H(sV) \\
&= \int_{\mathbb{R}_+} (H(sv) + H(-sv)) \, \frac{1}{\sigma_V}\phi(\frac{v}{\sigma_V}) \, dv \, .
\end{aligned}$$

Differentiating with respect to s, we obtain

$$\frac{d}{ds} \, \mathbb{E}H(sV) \, = \int_{\mathbb{R}_+} v \, (H'(sv) - H'(-sv)) \, \frac{1}{\sigma_V}\phi(\frac{v}{\sigma_V}) \, dv \, \leq 0$$

since H' is decreasing. We thus conclude that

$$\mathbb{E}H(U) \geq \mathbb{E}H(sV) \geq \mathbb{E}H(V) \, ,$$

since $s \leq 1$ and H is increasing by assumption. □

Theorem 3.3. *Consider model (1) with $p = 1$ and $A = [0, \infty)$. Suppose that the variances are misspecified and that the true variances are given by $\mathbb{V}(X_1) = a_1^2\sigma_1^2, \mathbb{V}(X_2) = a_2^2\sigma_2^2, \mathbb{V}(Y_1) = a_Y^2\sigma_Y^2$. For estimating the density of Y_1 under Kullback-Leibler loss, the Bayes predictive density $\hat{q}_{\pi_{U,A}}$ dominates the minimum risk equivariant predictive density $\hat{q}_{mre}$ whenever $\sigma_U^2 \leq \sigma_V^2$ with*

$$\sigma_U^2 = \frac{a_2^2\sigma_2^2 + (1-\beta)^2 a_1^2\sigma_1^2 + \beta^2 a_Y^2\sigma_Y^2}{\sigma_2^2 + \beta\sigma_Y^2} \, , \, \sigma_V^2 = \frac{a_1^2\sigma_1^2 + a_2^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \, , \beta = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_Y^2} . \tag{22}$$

In particular, dominance occurs for cases : **(i)** $a_1^2 = a_2^2 = a_Y^2$, **(ii)** $a_Y^2 \leq a_1^2 = a_2^2$, **(iii)** $\sigma_1^2 = \sigma_2^2 = \sigma_Y^2$ *and* $\frac{a_2^2 + a_Y^2}{2} \leq a_1^2$.

Remark 3.1. *Conditions* **(i)**, **(ii)** *and* **(iii)** *are quite informative. One common factor for the dominance to persist, especially seen by* **(iii)**, *is for the variance of X_1 to be relatively large compared to the variances of X_2 and Y_1.*

Proof. Particular cases **(i)**, **(ii)**, **(iii)** follow easily from (22). To establish condition (22), we prove, as in Theorem 3.1, that $\Delta(\theta)$ given in (17) is greater or equal to zero. We apply Lemma 3.1, with $H \equiv \log \Phi$ increasing and concave as required, showing that $\mathbb{E} \log(\Phi(U)) \geq \mathbb{E} \log(\Phi(V))$ with $U = \alpha_0 + \alpha_1 \frac{Y_1 - X_1}{\tau} \sim N(\mu_U, \sigma_U^2)$ and $V = \frac{\alpha_0}{\sqrt{1+\alpha_1^2}} \sim N(\mu_V, \sigma_V^2)$. Since $\mu_U = \frac{\theta_1 - \theta_2}{\sigma_T} > \frac{\theta_1 - \theta_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \mu_V$, the inequality $\sigma_U^2 \leq \sigma_V^2$ will suffice to have dominance. Finally, the proof is complete by checking that σ_U^2 and σ_V^2 are as given in (22), when the true variances are given by $\mathbb{V}(X_1) = a_1^2\sigma_1^2, \mathbb{V}(X_2) = a_2^2\sigma_2^2, \mathbb{V}(Y_1) = a_Y^2\sigma_Y^2$. □

Remark 3.2. *In opposition to the above robustness analysis, the dominance property of $\hat{q}_{\pi_{U,A}}$ versus $\hat{q}_{mre}$ for the restriction $\theta_1 - \theta_2 \geq 0$ does not persist for parameter space values such that $\theta_1 - \theta_2 < 0$, i.e., the additional information difference is misspecified. In fact, it is easy to see following the proof of Theorem*

3.1 that $R_{KL}(\theta,\hat{q}_{mre}) - R_{KL}(\theta,\hat{q}_{\pi_{U,A}}) < 0$ for θ's such that $\theta_1-\theta_2<0$. A potential protection is to use the predictive density estimator $\hat{q}_{\pi_{U,A'}}$ with $A'=[\epsilon,\infty)$, $\epsilon<0$, and with dominance occurring for all θ such that $\theta_1-\theta_2\geq\epsilon$ (Remark 1.1 and Theorem 3.1).

4. Further dominance results

We exploit different channels to obtain predictive density estimation improvements on benchmark procedures such as the maximum likelihood predictive density $\hat{q}_{\text{mle}}$ and the minimum risk equivariant predictive density $\hat{q}_{\text{mre}}$. These predictive densities are members of the broader class of densities

$$q_{\hat{\theta}_1,c} \sim N_p(\hat{\theta}_1(X), c\sigma_Y^2 I_p)\,, \tag{23}$$

with, for instance, the choice $\hat{\theta}_1(X) = \hat{\theta}_{1,\text{mle}}(X), c=1$ yielding $\hat{q}_{\text{mle}}$, and $\hat{\theta}_1(X) = X, c = 1+\frac{(1-\alpha)\sigma_1^2}{2\sigma_Y^2}$ yielding $\hat{q}_{\text{mre}}$ for loss L_α. In opposition to the previous section, our analysis and findings apply to all $\alpha-$divergence losses with $-1\leq\alpha<1$.

Two main strategies are exploited to produce improvements: **(A)** scale expansion and **(B)** point estimation duality.

- **(A)** *Plug–in* predictive densities $q_{\hat{\theta}_1,1}$ were shown in Fourdrinier et al. [17], in models where X_2 is not observed and for Kullback-Leibler loss, to be universally deficient and improved upon uniformly in terms of risk by a subclass of scale expansion variants $q_{\hat{\theta}_1,c}$ with $c-1$ positive and bounded above by a constant depending on the infimum mean squared error of $\hat{\theta}_1$. An adaptation of their result leads to dominating predictive densities of $\hat{q}_{\text{mle}}$, as well as other *plug–in* predictive densities which exploit the additional information $\theta_1-\theta_2\in A$, in terms of Kullback-Leibler risk. This is expanded upon in Subsection 4.1.
- **(B)** By duality, we mean that the frequentist risk performance of a predictive density $q_{\hat{\theta}_1,c}$ is equivalent to the point estimation frequentist risk of $\hat{\theta}_1$ in estimating θ_1 under an associated dual loss (e.g., Robert [36]). For Kullback-Leibler risk, the dual loss is squared error (Lemma 4.3) and our problem connects to the problem of estimating θ_1 with $\theta_1-\theta_2\in A$ based on model (1). In turn, as expanded upon in Marchand and Strawderman [33], improvements for the latter problem can be generated by improvements for a related restricted parameter space problem. Findings for $\alpha-$divergence loss with $\alpha\in(-1,1)$ are also obtained by exploiting a dual relationship with reflected normal loss. Details and illustrations are provided in Subsection 4.2.

4.1. Improvements by variance expansion

For Kullback-Leibler divergence loss, improvements on *plug–in* predictive densities by variance expansion stem from the following result.

Lemma 4.1. *Consider model (1) with $\theta_1 - \theta_2 \in A$, a given estimator $\hat{\theta}_1$ of θ_1, and the problem of estimating the density of Y_1 under Kullback-Leibler loss by a predictive density estimator $q_{\hat{\theta}_1,c}$ as in (23). Let $\underline{R} = \inf_\theta\{\mathbb{E}_\theta[\|\hat{\theta}_1(X) - \theta_1\|^2]\}/(p\sigma_Y^2)$, where the infimum is taken over the parameter space, i.e. $\{\theta \in \mathbb{R}^{2p} : \theta_1 - \theta_2 \in A\}$, and suppose that $\underline{R} > 0$.*

- **(a)** *Then, $q_{\hat{\theta}_1,1}$ is inadmissible and dominated by $q_{\hat{\theta}_1,c}$ for $1 < c < c_0(1+\underline{R})$, with $c_0(s)$, for $s > 1$, the root $c \in (s, \infty)$ of $G_s(c) = (1 - 1/c)\, s - \log c$.*
- **(b)** *Furthermore, we have $s^2 < c_0(s) < e^s$ for all $s > 1$, and $\lim_{s\to\infty} \frac{c_0(s)}{e^s} = 1$.*

Proof. See Appendix. □

Remark 4.1. *Part* **(b)** *above is indicative of the large allowance in the degree of expansion that leads to improvement on the* plug–in *procedure. A minimal complete subclass of predictive densities $q_{\hat{\theta}_1,c}$ is given by the values $c \in [1 + \underline{R}, 1 + \overline{R}]$, with $\overline{R} = \sup_\theta\{\mathbb{E}_\theta[\|\hat{\theta}_1(X) - \theta_1\|^2]\}/(p\sigma_Y^2)$, where the supremum is taken over the restricted parameter space, i.e., $\theta_1 - \theta_2 \in A$ (see Fourdrinier [17], Remark 5.1).*

The above result is, along with Theorem 4.1 below, universal with respect to the choice of the *plug-in* estimator $\hat{\theta}_1$, the dimension p and the constraint set A. We will otherwise focus below on the *plug-in* maximum likelihood predictive density $\hat{q}_{\mathrm{mle}}$, and the next result will be useful. Its first part presents a decomposition of $\hat{\theta}_{1,mle}$, while the second and third parts relate to a squared error risk decomposition of estimators given by Marchand and Strawderman [33].

Lemma 4.2. *Consider the problem of estimating θ_1 in model (1) with $\theta_1 - \theta_2 \in A$ and based on X. Set $r = \sigma_2^2/\sigma_1^2$, $\mu_1 = (\theta_1 - \theta_2)/(1+r)$, $\mu_2 = (r\theta_1 + \theta_2)/(1+r)$, $W_1 = (X_1 - X_2)/(1+r), W_2 = (rX_1 + X_2)/(1+r)$, and consider the subclass of estimators of θ_1*

$$\mathrm{C} = \{\delta_\psi :\ \delta_\psi(W_1, W_2) = W_2 + \psi(W_1)\}\ . \tag{24}$$

Then,

- **(a)** *The maximum likelihood estimator (mle) of θ_1 is a member of C with $\psi(W_1)$ the mle of μ_1 based on $W_1 \sim N_p(\mu_1, \sigma_1^2/(1+r) I_p)$ and $(1+r)\mu_1 \in A$;*
- **(b)** *The frequentist risk under squared error loss $\|\delta - \theta_1\|^2$ of an estimator $\delta_\psi \in \mathrm{C}$ is equal to*

$$R(\theta, \delta_\psi) \ =\ \mathbb{E}_{\mu_1}[\|\psi(W_1) - \mu_1\|^2] + \frac{p\sigma_2^2}{1+r}\,; (1+r)\mu_1 \in A; \tag{25}$$

- **(c)** *Under squared error loss, the estimator δ_{ψ_1} dominates δ_{ψ_2} iff $\psi_1(W_1)$ dominates $\psi_2(W_1)$ as an estimator of μ_1 under loss $\|\psi - \mu_1\|^2$ and the constraint $(1+r)\mu_1 \in A$.*

Proof. Part **(c)** follows immediately from part **(b)**. Part **(b)** follows since

$$\begin{aligned} R(\theta, \delta_\psi) &= \mathbb{E}_\theta\left[\|W_2 + \psi(W_1) - \theta_1\|^2\right] \\ &= \mathbb{E}_\theta\left[\|\psi(W_1) - \mu_1\|^2\right] + \mathbb{E}_\theta\left[\|W_2 - \mu_2\|^2\right]\ , \end{aligned}$$

yielding (25) given that W_1 and W_2 are independently distributed with $W_2 \sim N_p(\mu_2, (\sigma_2^2/(1+r))I_p)$. Similarly, for part **(a)**, we have $\hat{\theta}_{1,mle} = \hat{\mu}_{1,mle} + \hat{\mu}_{2,mle}$ with $\hat{\mu}_{2,mle}(W_1, W_2) = W_2$. Finally, since W_1 and W_2 are independent, the estimator $\hat{\mu}_{1,mle}(W_1, W_2)$ depends only on $W_1 \sim N_p(\mu_1, (\sigma_1^2/(1+r))I_p)$. □

Combining Lemmas 4.1 and 4.2, we obtain the following.

Theorem 4.1. *Lemma 4.1 applies to* plug-in *predictive densities $q_{\delta_\psi,1} \sim N_p(\delta_\psi, \sigma_Y^2 I_p)$ with $\delta_\psi \in C$, as defined in (24), and*

$$\underline{R} \,=\, \frac{1}{\sigma_Y^2}\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} \,+\, \frac{1}{p}\inf_{\mu_1}\mathbb{E}[\|\psi(W_1)-\mu_1\|^2]\right). \tag{26}$$

Namely, $q_{\delta_\psi,c} \sim N_p(\delta_\psi, c\sigma_Y^2 I_p)$ dominates $q_{\delta_\psi,1}$ for $1 < c < c_0(1+\underline{R})$. Moreover, we have $c_0(1+\underline{R}) \geq (1+\underline{R})^2 \geq (1+\frac{1}{\sigma_Y^2}\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2})^2$. Finally, the above applies to the maximum likelihood predictive density

$$\hat{q}_{mle} \sim N_p(\hat{\theta}_{1,mle}, \sigma_Y^2 I_p)\,, \;\; with\; \hat{\theta}_{1,mle}(X) = W_2 + \hat{\mu}_{1,mle}(W_1)\,, \tag{27}$$

and

$$\underline{R} \,=\, \frac{1}{\sigma_Y^2}\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} \,+\, \frac{1}{p}\inf_{\mu_1}\mathbb{E}[\|\hat{\mu}_{1,mle}(W_1)-\mu_1\|^2]\right), \tag{28}$$

where $\hat{\mu}_{1,mle}(W_1)$ the mle of μ_1 based on $W_1 \sim N_p(\mu_1, (\sigma_1^2/(1+r))I_p)$ and under the restriction $(1+r)\mu_1 \in A$.

We pursue with a pair of examples. With the above dominance result quite general, one further issue is the determination of $\underline{R}$ in (26). An analytical assessment is a challenge in general, but the univariate order restriction case leads to an explicit solution as detailed upon in Example 4.1. Otherwise a numerical evaluation of (26) is quite feasible and such a strategy is illustrated with Example 4.2.

Example 4.1. *(Univariate case with $\theta_1 \geq \theta_2$)*
Consider model (1) with $p = 1$ and $A = [0, \infty)$. The maximum likelihood predictive density $\hat{q}_{mle}$ is given by (27) with $\hat{\mu}_{1,mle}(W_1) = \max(0, W_1)$. The mean squared error of $\hat{\theta}_{1,mle}(X)$ may be derived from (25) and equals

$$R(\theta, \hat{\theta}_{1,mle}) \,=\, \mathbb{E}_{\mu_1}[\,|\hat{\mu}_{1,mle}(W_1) - \mu_1|^2] + \frac{\sigma_2^2}{1+r}\,, \mu_1 \geq 0.$$

A standard calculation for the mle of a non-negative normal mean based on $W_1 \sim N\left(\mu_1, \sigma_{W_1}^2 = \sigma_1^2/(1+r)\right)$ yields the expression

$$\begin{aligned}\mathbb{E}_{\mu_1}[\,|\hat{\mu}_{1,mle}(W_1)-\mu_1|^2] \;&=\; \mu_1^2\,\Phi(-\frac{\mu_1}{\sigma_{W_1}}) \,+\, \int_0^\infty (w_1-\mu_1)^2\,\phi(\frac{w_1-\mu_1}{\sigma_{W_1}})\,\frac{dw_1}{\sigma_{W_1}}\\ &=\; \sigma_{W_1}^2\left\{\frac{1}{2} \,+\, \rho^2\Phi(-\rho) \,+\, \int_0^\rho t^2\,\phi(t)\,dt\right\},\end{aligned}$$

with the change of variables $t = (w_1 - \mu_1)/\sigma_{W_1}$*, and setting* $\rho = \mu_1/\sigma_{W_1}$*. Furthermore, the above risk increases in* μ_1*, as* $\frac{d}{d\rho}\left\{\rho^2\Phi(-\rho) + \int_0^\rho t^2\,\phi(t)\,dt\right\} = 2\rho\Phi(-\rho) > 0$ *for* $\rho > 0$*, ranging from a minimum value of* $\sigma^2_{W_1}/2$ *to a supremum value of* $\sigma^2_{W_1}$*. Theorem 4.1 thus applies with*

$$\underline{R} \,=\, \frac{1}{\sigma_Y^2}\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} + \frac{\sigma_{W_1}^2}{2}\right) \,=\, \frac{\sigma_1^2}{\sigma_Y^2(\sigma_1^2+\sigma_2^2)}\,(\sigma_2^2 + \sigma_1^2/2)\,.$$

Similarly, Remark 4.1 applies with $\overline{R} = \sigma_1^2/\sigma_Y^2$*.*

As a specific illustration of Theorem 4.1 and Remark 4.1, consider the equal variances case with $\sigma_1^2 = \sigma_2^2 = \sigma_Y^2$ *for which the above yields* $\underline{R} = 3/4, \overline{R} = 1$ *and for which we can infer that:*

- **(a)** $q_{\hat{\theta}_{1,mle},c}$ *dominates* $\hat{q}_{mle}$ *under Kullback-Leibler loss for* $1 < c < c_0(7/4) \approx 3.48066$*;*
- **(b)** *A minimal complete subclass among the* $q_{\hat{\theta}_{1,mle},c}$*'s is given by the choices* $c \in [1+\underline{R}, 1+\overline{R}] = [7/4, 2]$.

Example 4.2. *Here is a further illustration of the dominance results in a multivariate setting for* Y_1*. Consider* X, Y_1 *as distributed as in (1) with* $p = 3, \sigma_1^2 = 1, \sigma_2^2 = 2, \sigma_Y^2 = 1$*. Further suppose that we can stipulate as additional information that* $\|\theta_1 - \theta_2\| \le m$ *for some known* m*. We apply Lemma 4.2 and Theorem 4.1 to obtain improvements to the maximum likelihood predictive density*

$$\hat{q}_{mle} \sim N_3(\hat{\theta}_{1,mle}(X), I_3)\,.$$

From Lemma 4.2, we have for $W_2 = \frac{X_1+2X_2}{3}$ *and* $W_1 = \frac{2(X_1-X_2)}{3}$*:*

$$\hat{\theta}_{1,mle}(X_1, X_2) \,=\, W_2 + \psi_{mle}(W_1)\,,$$

with $\psi_{mle}(W_1) = \min\{\frac{2m}{3}, \|W_1\|\}\frac{W_1}{\|W_1\|}$ *(e.g., Marchand and Perron [32]) the mle of* $\mu_1 = \frac{2(\theta_1-\theta_2)}{3}$ *based on* $W_1 \sim N_3(\mu_1, \frac{2}{3}I_3)$ *and the parametric restriction* $\|\mu_1\| \le \frac{2m}{3}$*.*

Theorem 4.1 tells us that variance expansion variants $q_{\hat{\theta}_{1,mle},c} \sim N_3(\hat{\theta}_{1,mle}(X), cI_3)$ *dominate* $\hat{q}_{mle}$ *for Kullback-Leibler loss and for* $\|\theta_1 - \theta_2\| \le m$ *as long as* $1 < c \le c_0(1+\underline{R})$*, with*

$$\underline{R} \,=\, \frac{1}{3}\left(1 \,+\, \inf_{\mu_1}\mathbb{E}[\|[\psi_{mle}(W_1) - \mu_1\|^2]\right)\,;$$

with the infimum taken over $\{\mu_1 \in \mathbb{R}^3 : \|\mu_1\| \le \frac{2m}{3}\}$*. Taking* $m = 2$*, a numerical evaluation tells us that* $\underline{R} \approx 0.672$ *and* $\overline{R} \approx 0.763$ *(Lemma 4.1)). The dominance of* $q_{\hat{\theta}_{1,mle},c}$ *over* $\hat{q}_{mle}$ *will thus occur for* $1 < c \le c_0(1+\underline{R}) \approx 3.108$*. But, among these choices of variance expansion, only* $c \in [1+\underline{R}, 1+\overline{R}] \approx [1.672, 1.763]$ *are admissible (Remark 4.1). Figure 1 reproduces the Kullback-Leibler risks of* $\hat{q}_{mre}, \hat{q}_{mle}$ *and* $q_{\hat{\theta}_{1,mle},c}$ *for* $c = 1.672$ *and* $c = 2$*, as a function of* $\|\mu_1\| =$

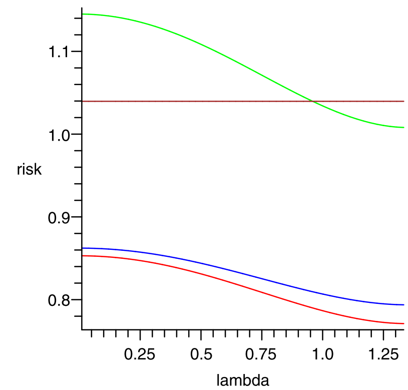

FIG 1. *Kullback-Leibler risks, as a function of* $\lambda = \|\mu_1\| \in [0, 4/3]$ *of* $\hat{q}_{mre}$ *(brown),* $q_{\hat{\theta}_{1,mle},c}$ *for* $c = 1$ *(green),* $c = 2$ *(blue),* $c = 1.672 \approx 1 + \underline{R}$ *(red), for* $p = 3$, $\sigma_1^2 = \sigma_2^2 = 0.5, \sigma_Y^2 = 1$

$\frac{2}{3}\|\theta_1 - \theta_2\| \in [0, 4/3]$. *The dominance of the two variance expansions densities* $q_{\hat{\theta}_{1,mle},c}$ *is clear and impressive, with gains of between* 31% *and* 34%*, or so, of* $q_{\hat{\theta}_{1,mle},1.672}$ *over* $\hat{q}_{mle}$. *The complete subclass result is also illustrated with the slight ordering between the two variance expansion densities. The density* $\hat{q}_{mre} \sim N_3(X_1, 2I_3)$ *ignores the additional information on the means (A), but does incorporate an expansion of variance (B). As seen by the comparison with* $\hat{q}_{mle}$*, which does the opposite, the latter quality (A) compensates to a significant extent for the former deficiency (B). Supplementary details on the computation of the Kullback-Leibler risks and the mean squared error of* ψ_{mle} *are provided in the Appendix (see in particular expressions (31) and (35)).*

4.2. Improvements through duality

We consider again here predictive density estimators $q_{\hat{\theta}_1,c}$, as in (23), but focus rather on the role of the plugged-in estimator $\hat{\theta}_1$. We seek improvements on benchmark choices such as $\hat{q}_{mre}$, and *plug–in* predictive densities with $c = 1$. We begin with the Kullback-Leibler case which relates to squared error loss.

Lemma 4.3. *For model (1), the frequentist risk of the predictive density estimator* $q_{\hat{\theta}_1,c}$ *of the density of* Y_1*, under Kullback-Leibler divergence loss, is dual to the frequentist risk of* $\hat{\theta}_1(X)$ *for estimating* θ_1 *under squared error loss* $\|\hat{\theta}_1 - \theta_1\|^2$. *Namely,* $q_{\hat{\theta}_{1,A},c}$ *dominates* $q_{\hat{\theta}_{1,B},c}$ *under loss* L_α *iff* $\hat{\theta}_{1,A}(X)$ *dominates* $\hat{\theta}_{1,B}(X)$ *under squared error loss.*

Proof. See for instance Fourdrinier [17]. □

As exploited by Ghosh, Mergel and Datta [19], and also more recently by Marchand, Perron and Yadegari [31], for other $\alpha-$divergence losses, it is reflected normal loss which is dual for *plug–in* predictive density estimators, as well as scale expansions in (23).

Lemma 4.4. *(Duality between $\alpha-$divergence and reflected normal losses) For model (1), the frequentist risk of the predictive density estimator $q_{\hat{\theta}_1,c}$ of the density of Y_1 under $\alpha-$divergence loss (2), with $|\alpha| < 1$, is dual to the frequentist risk of $\hat{\theta}_1(X)$ for estimating θ_1 under reflected normal loss*

$$L_{\gamma_0}(\theta_1, \hat{\theta}_1) = 1 - e^{-\|\hat{\theta}_1 - \theta_1\|^2/2\gamma_0}\,, \tag{29}$$

with $\gamma_0 = 2\,(\frac{c}{1+\alpha} + \frac{1}{1-\alpha})\,\sigma_Y^2$. Namely, $q_{\hat{\theta}_{1,A},c}$ dominates $q_{\hat{\theta}_{1,B},c}$ under loss L_α iff $\hat{\theta}_{1,A}(X)$ dominates $\hat{\theta}_{1,B}(X)$ under loss L_{γ_0} as above.

Proof. See for instance Ghosh, Mergel and Datta [19]. □

Remark 4.2. *Observe that $\lim_{\gamma_0\to\infty} 2\gamma_0\, L_{\gamma_0}(\theta_1, \hat{\theta}_1) = \|\hat{\theta}_1 - \theta_1\|^2$, so that the point estimation performance of θ_1 under reflected normal loss L_{γ_0} should be expected to match that of squared error loss when $\gamma_0 \to \infty$. In view of Lemma 4.3 and Lemma 4.4, this in turn suggests that the $\alpha-$divergence performance of $\hat{q}_{\hat{\theta}_1,c}$ will match that of Kullback-Leibler when taking $\alpha \to -1$. Finally, we point out that the boundedness nature of the loss in (29) stands out in contrast to KL loss and its dual unbounded squared-error loss.*

Now, pairing Lemma 4.3 and Lemma 4.2 leads immediately to the following general dominance result for Kullback-Leibler loss.

Theorem 4.2. *Consider model (1) with $\theta_1 - \theta_2 \in A$ and the problem of estimating the density of Y_1 under Kullback-Leibler loss. Set $r = \sigma_2^2/\sigma_1^2$, $W_1 = (X_1 - X_2)/(1+r), W_2 = (rX_1 + X_2)/(1+r), \mu_1 = (\theta_1 - \theta_2)/(1+r)$, and further consider the subclass of predictive densities $q_{\delta_\psi,c}$, as in (23) for fixed c, with δ_ψ an estimator of θ_1 of the form $\delta_\psi(W_1, W_2) = W_2 + \psi(W_1)$. Then, $q_{\delta_{\psi_A},c}$ dominates $q_{\delta_{\psi_B},c}$ if and only if ψ_A dominates ψ_B as an estimator of μ_1 under loss $\|\psi - \mu_1\|^2$, for $W_1 \sim N_p(\mu_1, \frac{\sigma_1^2}{1+r} I_p)$ and the parametric restriction $(1+r)\mu_1 \in A$.*

Proof. The result follows from Lemma 4.3 and Lemma 4.2. □

The above result connects three problems, namely:

- **(I)** the efficiency of $q_{\delta_\psi,c}$ under KL loss as a predictive density for Y_1 with the additional information $\theta_1 - \theta_2 \in A$;
- **(II)** the efficiency of $\delta_\psi(X)$ as an estimator of θ_1 under squared error loss $\|\delta_\psi - \theta_1\|^2$ with the additional information $\theta_1 - \theta_2 \in A$;
- **(III)** the efficiency of $\psi(W_1)$ for $W_1 \sim N_p(\mu_1, \sigma_1^2/(1+r)I_p)$ as an estimator of μ_1 under squared error loss $\|\psi - \mu_1\|^2$ with the parametric restriction $(1+r)\mu_1 \in A$.

Previous authors (Blumenthal and Cohen [8]; Cohen and Sackrowitz [12]; van Eeden and Zidek [43, 44], for $p = 1$; Marchand and Strawderman [33], for $p \geq 1$) have exploited the **(II)-(III)** connection (i.e., Lemma 4.2) to obtain findings for problem **(II)** based on restricted parameter space findings for **(III)**. The above Theorem further exploits connections **(I)-(II)** (i.e., Lemma 4.3) and permits one to derive findings for predictive density estimation problem **(I)** from restricted parameter space findings for **(III)**. An example, which will also be illustrative of $\alpha-$divergence results, is provided below at the end of this section.

For other $\alpha-$divergence losses, the above scheme is not immediately available for the dual reflected normal loss since Lemma 4.2 is intimately linked to squared error loss. However, the following slight extension of a result due to Kubokawa, Marchand and Strawderman [26], exploiting a concave loss technique dating back to Brandwein and Strawderman [9], permits us to connect reflected normal loss to squared error loss, and consequently the efficiency of predictive densities under α-divergence loss to point estimation in restricted parameter spaces as in **(III)** above. The proof of the next Lemma is omitted, but is quite analogous to that given by Kubokawa, Marchand and Strawderman [26].

Lemma 4.5. *Consider model (1) and the problem of estimating θ_1 based on X, with $\theta_1 - \theta_2 \in A$ and reflected normal loss as in (29) with $|\alpha| < 1$. Then $\hat{\theta}_1(X)$ dominates X_1 whenever $\hat{\theta}_1(Z)$ dominates Z_1 as an estimate of θ_1, under squared error loss $\|\hat{\theta}_1 - \theta_1\|^2$, with $\theta_1 - \theta_2 \in A$, for the model*

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathrm{N}_{2p}\left(\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \Sigma_Z = \left(\begin{smallmatrix} \sigma^2_{Z_1} I_p & 0 \\ 0 & \sigma_2^2 I_p \end{smallmatrix}\right)\right), \tag{30}$$

with $\sigma^2_{Z_1} = \frac{\gamma\sigma_1^2}{\gamma+\sigma_1^2}$.

Theorem 4.3. *Consider model (1) with $\theta_1 - \theta_2 \in A$ and the problem of estimating the density of Y_1 under $\alpha-$divergence loss with $|\alpha| < 1$. Set $r = \sigma_2^2/\sigma_1^2$, $W_1 = (X_1 - X_2)/(1+r), W_2 = (rX_1 + X_2)/(1+r), \mu_1 = (\theta_1 - \theta_2)/(1+r)$, and further consider the subclass of predictive densities $q_{\delta_\psi,c}$, as in (23) for fixed c, with δ_ψ an estimator of θ_1 of the form $\delta_\psi(W_1, W_2) = W_2 + \psi(W_1)$. Then, $q_{\delta_{\psi_A},c}$ dominates $q_{\delta_{\psi_B},c}$ as long as ψ_A dominates ψ_B as an estimator of μ_1 under loss $\|\psi - \mu_1\|^2$, for $W_1 \sim N_p(\mu_1, \frac{\sigma^2_{Z_1}}{1+r} I_p)$, the parametric restriction $(1+r)\mu_1 \in A$, and $\sigma^2_{Z_1} = \frac{\{(1+\alpha)+c(1-\alpha)\}\,\sigma_1^2}{\{(1+\alpha)+c(1-\alpha)\} + (1-\alpha^2)\,\sigma_1^2/(2\sigma_Y^2)}$.*

Proof. The result follows from Lemma 4.4 and its dual reflected normal loss L_{γ_0}, the use of Lemma 4.5 applied to $\sigma^2_{Z_1} = \frac{\gamma_0\sigma_1^2}{\gamma_0+\sigma_1^2}$, and an application of part **(c)** of Lemma 4.2 to Z as distributed in (30). □

As for Kullback-Leibler loss, the above $\alpha-$ divergence result connects several problems, analogous to **(I)**, **(II)**, and **(III)**, and additionally

(IB) the efficiency of $\delta_\psi(X)$ as an estimator of θ_1 under reflected normal loss L_{γ_0} with $\gamma_0 = 2\,(\frac{c}{1+\alpha} + \frac{1}{1-\alpha})\,\sigma_Y^2$ with the additional information $\theta_1 - \theta_2 \in A$.

Here is an illustration.

Example 4.3. *Here is an illustration of both Theorems 4.2 and 4.3. Consider model (1) with A a convex set with a non-empty interior, and $\alpha-$divergence* *loss ($-1 \le \alpha < 1$) for assessing a predictive density for Y_1. Further consider the minimum risk predictive density $\hat{q}_{mre}$ as a benchmark procedure, which is of the form $q_{\delta_{\psi_B}}$ as in Theorem 4.3 with $\delta_{\psi_B} \in C$, $\psi_B(W_1) = W_1$ and $c = c_{mre} = 1 + (1-\alpha)\sigma_1^2/(2\sigma_Y^2)$. Now consider the Bayes estimator $\psi_U(W_1)$ under squared error loss of μ_1 associated with a uniform prior on the restricted parameter space $(1+r)\mu_1 \in A$, for $W_1 \sim N_p((\mu_1, \frac{\sigma^2_{Z_1}}{1+r} I_p)$ as in Theorem 4.3. It follows from Hartigan's theorem (Hartigan [21]; Marchand and Strawderman [33]) that $\psi_A(W_1) \equiv \psi_U(W_1)$ dominates $\psi_B(W_1)$ under loss $\|\psi - \mu_1\|^2$ and for $(1+r)\mu_1 \in A$. It thus follows from Theorem 4.3 that the predictive density $N_p(\delta_{\psi_B}(X), (\frac{1-\alpha}{2}\sigma_1^2 + \sigma_Y^2) I_p)$ dominates $\hat{q}_{mre}$ under $\alpha-$divergence loss with $\delta_{\psi_B}(X) = \frac{rX_1+X_2}{1+r} + \psi_U(\frac{X_1-X_2}{1+r})$. The dominance result is unified with respect to $\alpha \in [-1,1)$, the dimension p, and the set A.*

We conclude this section with an adaptive two-step strategy, building on both variance expansion and improvements through duality, to optimise potential Kullback-Leibler improvements on $\hat{q}_{mle} \sim N_p(\hat{\theta}_{1,mle}, \sigma_Y^2 I_p)$, in cases where point estimation improvements on $\hat{\theta}_{1,mle}(X)$ under squared error loss are readily available.

- **(I)** Select an estimator δ^* which dominates $\hat{\theta}_{1,mle}$ under squared error loss. This may be achieved via part **(c)** of Lemma 4.2 resulting in a dominating estimator of the form $\delta^*(X) = W_2 + \psi^*(W_1) = (rX_1 + X_2)/(1+r) + \psi^*((X_1 - X_2)/(1+r))$ where $\psi^*(W_1)$ dominates $\hat{\mu}_{1,mle}(W_1)$ as an estimator of μ_1 under squared error loss and the restriction $(1+r)\mu_1 \in A$.
- **(II)** Now, with the *plug-in* predictive density estimator $q_{\delta^*,1}$ dominating $\hat{q}_{mle}$, further improve $q_{\delta^*,1}$ by a variance expanded $q_{\delta^*,c}$. Suitable choices of c are prescribed by Theorem 4.1 and given by $c_0(1+\underline{R})$, with $\underline{R}$ given in (26). The evaluation of $\underline{R}$ hinges on the infimum risk $\inf_{\mu_1} \mathbb{E}[\|\psi^*(W_1) - \mu_1\|^2]$, and such a quantity can be either estimated by simulation, derived in some cases analytically, or safely underestimated by 0.

Examples where the above can be applied include the cases: **(i)** $A = [0, \infty)$ with the use of Shao and Strawderman's [39] dominating estimators, and **(ii)** A the ball of radius m centered at the origin with the use of Marchand and Perron's [32] dominating estimators.

5. Further examples and illustrations

We present and comment numerical evaluations of Kullback-Leibler risks in the univariate case for both $\theta_1 \ge \theta_2$ (Figures 2, 3) and $|\theta_1 - \theta_2| \le m, m = 1, 2$. (Figures 4, 5). Each of the figures consists of plots of risk ratios, as functions of $\Delta = \theta_1 - \theta_2$ with the benchmark $\hat{q}_{mre}$ as the reference point. The variances are set equal to 1, except for Figure 3 which highlights the effect of varying σ_2^2.

Figure 2 illustrates the effectiveness of variance expansion (Theorem 4.1), as well as the dominance finding of Theorem 3.1. More precisely, the Figure relates to Example 4.1 where $\hat{q}_{mle}$ is improved by the variance expansion version $\hat{q}_{mle,2}$,

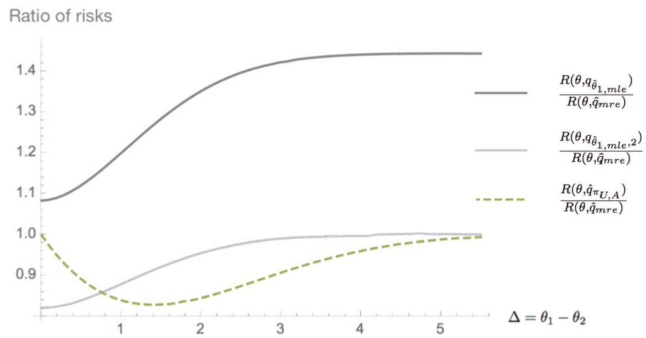

FIG 2. *Kullback-Leibler risk ratios for* $p = 1$, $A = [0, \infty)$, *and* $\sigma_1^2 = \sigma_2^2 = \sigma_Y^2 = 1$

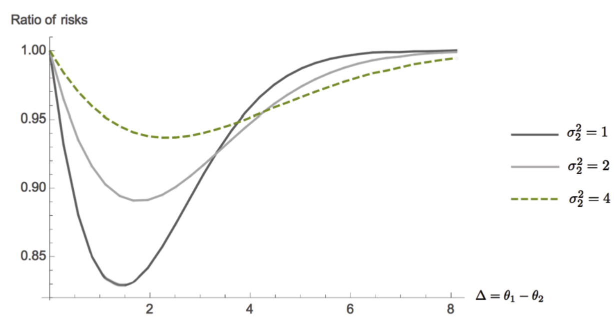

FIG 3. *Kullback-Leibler risk ratios for* $p = 1$, $A = [0, \infty)$, $\sigma_1^2 = \sigma_Y^2 = 1$ *and* $\sigma_2^2 = 1, 2, 4$

which belongs both to the subclass of dominating densities $\hat{q}_{mle,c}$ as well as to the complete subclass of such predictive densities. The gains are impressive ranging from a minimum of about 8% at $\Delta = 0$ to a supremum value of about 44% for $\Delta \to \infty$. Moreover, the predictive density $\hat{q}_{mle,2}$ also dominates $\hat{q}_{mre}$ by duality, but the gains are more modest. Interestingly, the penalty of failing to expand is more severe than the penalty for using an inefficient *plug-in* estimator of the mean. In accordance with Theorem 3.1, the Bayes predictive density $\hat{q}_{\pi_{U,A}}$ improves uniformly on $\hat{q}_{mre}$ except at $\Delta = 0$ where the risks are equal. As well, $\hat{q}_{\pi_{U,A}}$ compares well to $\hat{q}_{mle,2}$, except for small Δ, with $R(\theta, \hat{q}_{mle,2}) \leq R(\theta, \hat{q}_{\pi_{U,A}})$ if and only if $\Delta \leq \Delta_0$ with $\Delta_0 \approx 0.76$.

Figure 3 compares the efficiency of the predictive densities $\hat{q}_{\pi_{U,A}}$ and $\hat{q}_{mre}$ for varying σ_2^2. Smaller values of σ_2^2 represent more precise estimation of θ_2 and translates to a tendency for the gains offered by $\hat{q}_{\pi_{U,A}}$ to be greater for smaller σ_2^2; but the situation is slightly reversed for larger Δ.

Figures 4 and 5 compare the same estimators as in Figure 2, but they are adapted to the restriction to compact interval. Several of the features of Figure 2 are reproduced with the noticeable inefficiency of $\hat{q}_{mle}$ compared to both $\hat{q}_{mle,2}$ and $\hat{q}_{\pi_{U,A}}$. For the larger parameter space (i.e. $m = 2$), even $\hat{q}_{mre}$ outperforms $\hat{q}_{mle}$ as illustrated by Figure 5, but the situation is reversed for $m = 1$ where the

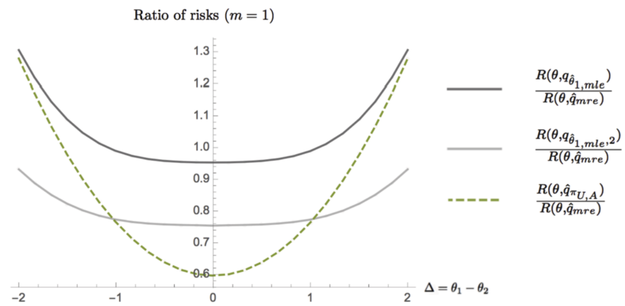

FIG 4. *Kullback-Leibler risk ratios for* $p = 1$, $A = [-1, 1]$, *and* $\sigma_1^2 = \sigma_2^2 = \sigma_Y^2 = 1$

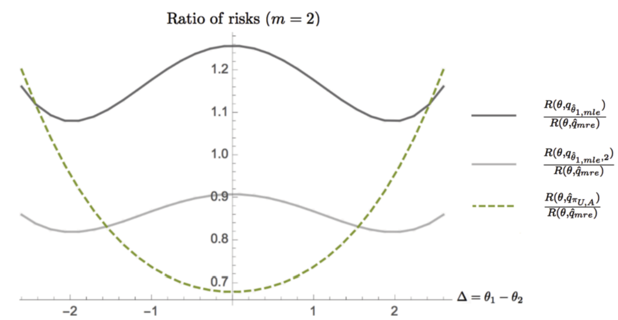

FIG 5. *Kullback-Leibler risk ratios for* $p = 1$, $A = [-2, 2]$, *and* $\sigma_1^2 = \sigma_2^2 = \sigma_Y^2 = 1$

efficiency of better point maximum likelihood estimates plays a more important role. The Bayes performs well, dominating $\hat{q}_{mre}$ in accordance with Theorem 3.2, especially for small of moderate Δ, and even improving on $\hat{q}_{mle,2}$ for $m = 1$. Finally, we have extended the plots outside the parameter space which is useful for assessing performance for slightly incorrect specifications of the additional information.

We conclude with an application.

Example 5.1. *We consider data presented in Silvapulle and Sen [40] and originating in Zelazo, Zelazo and Kolb [46]. The data consists of ages (in months) of infants for walking alone and divided into active-exercise and passive-exercise groups.*

Active exercise	*9 9.5 9.75 10 13 9.5*	*Mean = 10.125* (X_1)
Passive exercise	*11 10 10 11.75 10.5 15*	*Mean = 11.375* (X_2)

Let $Y_1 \sim N(\theta_1, \sigma_Y^2)$ be the age for walking alone for a new infant and for the active exercise group. Consider choosing a predictive density for Y_1 and Kullback-Leibler loss. Assume that the observed means are actually occurrences of $X_1 \sim N(\theta_1, \sigma_1^2)$ and $X_2 \sim N(\theta_2, \sigma_2^2)$ and that one is willing to assume that $\theta_1 \leq \theta_2$ (i.e., the active exercise regime can only lower the average age for walking). We are thus in set-up (1) with $A = (-\infty, 0]$, and such an application is both a continuation of, and relates to, Corollary 2.1, Theorem 3.1, Example 4.1, and the risk evaluation of this Section.

Suppose that we are willing to assume that the variances at the level of the individual measurements are equal (i.e., $6\sigma_1^2 = 6\sigma_2^2 = \sigma_Y^2$), and that past data suggests that $\sigma_Y^2 \approx 2$. Consider the restricted maximum likelihood estimator of θ_1 given by

$$\hat{\theta}_{1,mle}(X) \,=\, \frac{X_1 + X_2}{2} \,+\, \max\{0, \frac{X_1 - X_2}{2}\}$$

(e.g., Lemma 4.2). Under squared error loss, with the additional information $\theta_1 \leq \theta_2$, $\hat{\theta}_{1,mle}(X)$ dominates X_1 as an estimator of θ_1 (e.g., Lee, 1981) and, consequently, the predictive density $\hat{q}_{\hat{\theta}_{1,mle},c} \sim N(\hat{\theta}_{1,mle}, 2c)$ dominates the density $N(X_1, 2c)$ for all $c > 0$ under Kullback-Leibler loss (Lemma 4.3). This includes the case $c = 2$ where the latter of these densities is $\hat{q}_{mre}$.

With the given data, and supported by theoretical findings of this paper summarized by **(I)** and **(II)** below, two plausible predictive densities for Y_1 are $\hat{q}_{\hat{\theta}_{1,mle},2} \sim N(10.125, 4)$ and $\hat{q}_{\pi_{U,A}}$ given by a $\mathrm{SN}(1, \alpha_0 = \sqrt{525/208}, \alpha_1 = -\sqrt{13}/13, 10.125, \sqrt{7/3})$ density (Remark 2.4, **(b)**). The expectation for the Bayes density is slightly less (10.071 vs. 10.125). There is more variability for the Bayes density than that of Y_1 (standard deviation of 1.5155 versus $\sqrt{2}$), but less than for $\hat{q}_{\hat{\theta}_{1,mle},2}$ which has standard deviation equal to 2. Finally, summarizing some of the theoretical support for these choices, we have that:

- **(I)** $\hat{q}_{\pi_{U,A}}$ dominates $\hat{q}_{mre}$ for Kullback-Leibler loss (Theorem 3.1);
- **(II)** $\hat{q}_{\hat{\theta}_{1,mle},2}$ dominates $\hat{q}_{mre}$ for Kullback-Leibler loss, and belongs to the complete class of variance expansions consisting of densities $\hat{q}_{\hat{\theta}_{1,mle},c}$ with $c \in [7/4, 2]$ (Example 4.1).

6. Concluding remarks

For multivariate normal observables $X_1 \sim N_p(\theta_1, \sigma_1^2 I_p)$, $X_2 \sim N_p(\theta_2, \sigma_2^2 I_p)$, we have provided findings concerning the efficiency of predictive density estimators $Y_1 \sim N_p(\theta_1, \sigma_1^2 I_p)$ with the added parametric information $\theta_1 - \theta_2 \in A$. Several findings provide improvements on benchmark predictive densities, such those obtained as *plug-in's* or as minimum risk equivariant. Namely, for *plug-in* densities, we have illustrated the benefits of variance expansion for Kullback-Leibler loss. The efficiency of the *plug-in* estimator matters, and other methods than maximum likelihood, such as a weighted likelihood approach are worth investigating, but the Kullback-Leibler performance will still remain defective. The results obtained range over a class of $\alpha-$divergence losses, different settings for A, and include Bayesian improvements for Kullback-Leibler divergence loss. The various techniques used lead to novel connections between different problems, which are described following both Theorem 4.2 and Theorem 4.3.

Although the Bayesian dominance results for Kullback-Leibler loss for $p = 1$ extend to the rectangular case with $\theta_{1,i} - \theta_{2,i} \in A_i$ for $i = 1, \ldots, p$ and the $A_i's$

either lower bounded, upper bounded, or bounded to intervals $[-m_i, m_i]$ (since the Kullback-Leibler divergence for the joint density of Y factors and becomes the sum of the marginal Kullback-Leibler divergences, and that the posterior distributions of the $\theta_{1,i}$'s are independent), a general Bayesian dominance result of $\hat{q}_{\pi_{U,A}}$ over $\hat{q}_{mre}$, is lacking and would be of interest. As well, comparisons of predictive densities for the case of homogeneous, but unknown variance (i.e., $\sigma_1^2 = \sigma_2^2 = \sigma_Y^2$), are equally of interest. Finally, the analyses carried out here should be useful as benchmarks in situations where the constraint set A has an anticipated form, but yet is unknown. In such situations, a reasonable approach would be to consider priors that incorporate uncertainty on A, such as setting $A = \{\theta \in \mathbb{R}^{2p} | \, \|\theta_1 - \theta_2\| \leq m\}$, $A = [m, \infty)$, with prior uncertainty specified for m.

Appendix

Proof of Lemma 2.1

As shown by Corcuera and Giummolè [13], the Bayes predictive density estimator of the density of Y_1 in (1) under loss L_α, $\alpha \neq 1$, is given by

$$\hat{q}_{\pi_{U,A}}(y_1; x) \propto \left\{ \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \phi^{(1-\alpha)/2}(\frac{y_1 - \theta_1}{\sigma_Y}) \, \pi(\theta_1, \theta_2 | x) \, d\theta_1 \, d\theta_2 \right\}^{2/(1-\alpha)}.$$

With prior measure $\pi_{U,A}(\theta) = \mathbb{I}_A(\theta_1 - \theta_2)$, we obtain

$$\hat{q}_{\pi_{U,A}}(y_1; x)$$
$$\propto \left\{ \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} \phi(\frac{y_1 - \theta_1}{\sqrt{\frac{2}{1-\alpha}\sigma_Y^2}}) \, \phi(\frac{\theta_1 - x_1}{\sigma_1}) \, \phi(\frac{\theta_2 - x_2}{\sigma_2}) \, \mathbb{I}_A(\theta_1 - \theta_2) \, d\theta_1 \, d\theta_2 \right\}^{2/(1-\alpha)},$$

given that $\phi^m(z) \propto \phi(m^{1/2}z)$ for $m > 0$. By the decomposition

$$\frac{\|\theta_1 - y_1\|^2}{a} + \frac{\|\theta_1 - x_1\|^2}{b} = \frac{\|y_1 - x_1\|^2}{a + b} + \frac{\|\theta_1 - w\|^2}{\sigma_w^2},$$

with $a = \frac{2\sigma_Y^2}{1-\alpha}$, $b = \sigma_1^2$, and $w = \frac{by_1 + ax_1}{a+b} = \beta y_1 + (1-\beta)x_1$, $\sigma_w^2 = \frac{ab}{a+b} = \frac{2\sigma_1^2\sigma_Y^2}{2\sigma_Y^2 + (1-\alpha)\sigma_1^2}$, we obtain

$$\begin{aligned} \hat{q}_{\pi_{U,A}}(y_1; x) \quad &\propto \quad \phi^{2/(1-\alpha)}(\frac{y_1 - x_1}{\sqrt{\frac{2\sigma_Y^2}{1-\alpha} + \sigma_1^2}}) \\ &\qquad \times \left\{ \int_{\mathbb{R}^{2p}} \phi(\frac{\theta_1 - w}{\sigma_w}) \, \phi(\frac{\theta_2 - x_2}{\sigma_2}) \, \mathbb{I}_A(\theta_1 - \theta_2) \, d\theta_1 \, d\theta_2 \right\}^{2/(1-\alpha)} \\ &\propto \quad \hat{q}\mathrm{mre}(y_1; x_1) \, \{\mathbb{P}(Z_1 - Z_2 \in A)\}^{2/(1-\alpha)}, \end{aligned}$$

with Z_1, Z_2 independently distributed as $Z_1 \sim N_p(w, \sigma_w^2)$, $Z_2 \sim N_p(x_2, \sigma_2^2)$. The result follows by setting $T =^d Z_1 - Z_2$. □

Proof of Lemma 4.1

See Fourdrinier et al. ([17], Theorem 5.1) for part **(a)**. For the first part of **(b)**, it suffices to show that **(i)** $G_s(s^2) > 0$ and **(ii)** $G_s(e^s) < 0$, given that $G_s(\cdot)$ is, for fixed s, a decreasing function on (s, ∞). We have indeed $G_s(e^s) = -se^{-s} < 0$, while $G_s(s^2)|_{s=1} = 0$ and $\frac{\partial}{\partial s} G_s(s^2) = (1-1/s)^2 > 0$, which implies **(i)**. Finally, set $k_0(s) = \log c_0(s), s > 1$, and observe that the definition of c_0 implies that $u(k_0(s)) = \frac{k_0(s)}{1-e^{-k_0(s)}} = s$. Since $u(k)$ increases in $k \in (1, \infty)$, it must be the case that $k_0(s)$ increases in $s \in (1, \infty)$ with $\lim_{s\to\infty} k_0(s) \geq \lim_{s\to\infty} \log s^2 = \infty$. The result thus follows since $\lim_{s\to\infty} k_0(s)/s = \lim_{s\to\infty}(1 - e^{-k_0(s)}) = 1$. $\square$

The risk evaluation of Example 4.2

We expand here on the Kullback-Leibler risk computation of densities $\hat{q}_{mle,c} \sim N_3(\hat{\theta}_{1,mle}(X), cI_3)$ in the setting of Example 4.2 (i.e., $\sigma_1^2 = \sigma_Y^2 = 1, \sigma_2^2 = 1/2, \|\theta_1 - \theta_2\| \leq m$). Namely, we present a risk decomposition in terms of one-dimensional integral which facilitates the numerical evaluations required for Figure 1.

(A) With $q_{\theta_1}(y) = (2\pi)^{-3/2}\, e^{-\frac{\|y-\theta\|^2}{2}}$ and

$$\hat{q}_{mle,c}(y;x) \;=\; (2\pi c)^{-3/2}\, e^{-\frac{\|y-\hat{\theta}_{1,mle}(x)\|^2}{2c}},$$

we have

$$\begin{aligned}
R_{KL}(\theta, \hat{q}_{mle,c}) &= \mathbb{E}^{X,Y_1} \log \frac{q_{\theta_1}(Y_1)}{\hat{q}_{mle,c}(Y_1;X)} \\
&= \frac{3}{2}\log c + \frac{1}{2c}\mathbb{E}^{X,Y_1}(\|Y_1 - \hat{\theta}_{1,mle}(X)\|^2) - \frac{1}{2}\mathbb{E}^{Y_1}\|Y_1 - \theta_1\|^2 \\
&= \frac{3}{2}\log c + \frac{1}{2c}\mathbb{E}^{X,Y_1}(\|Y_1 - \theta_1\|^2 + \|\hat{\theta}_{1,mle}(X) - \theta_1\|^2) \\
&\quad - \frac{1}{2}\mathbb{E}^{Y_1}\|Y_1 - \theta_1\|^2 \\
&= \frac{3}{2}(\log c + \frac{1}{c} - 1) \;+\; \frac{1}{2c}\mathbb{E}^{X}(\|\hat{\theta}_{1,mle}(X) - \theta_1\|^2)\,,
\end{aligned}$$

by making use of the independence of X and Y_1 and since $\mathbb{E}^{Y_1}\|Y_1 - \theta_1\|^2 = 3$. Now, make use of (25) to obtain from the above

$$R_{KL}(\theta, \hat{q}_{mle,c}) \;=\; \frac{3}{2}(\log c - 1) + \frac{2}{c} + \frac{1}{2c}\mathbb{E}(\|\psi_{mle}(W_1) - \mu_1\|^2)\,, \tag{31}$$

with $W_1 \sim N(\mu_1, \sigma_{W_1}^2 I_3)$, $\sigma_{W_1}^2 = \frac{2}{3}$, $\mu_1 = \frac{2}{3}(\theta_1 - \theta_2)$, and $\|\mu_1\| \leq 2m/3$. As in Fourdrinier et al. (2011), the above provides the risk of $\hat{q}_{mle,c}$ in terms of the expected squared-error loss of the restricted mle $\psi_{mle}(W_1)$ of μ_1.

(B) As in Marchand and Perron (2001), the risk $\mathbb{E}(\|\psi_{mle}(W_1) - \mu_1\|^2)$ can be evaluated by conditioning on $R = \|W_1\| \sim \sqrt{\sigma^2_{W_1} \chi^2_3(\frac{\|\mu_1\|^2}{\sigma^2_{W_1}})}$ which has density on $\mathbb{R}_+$ given by:

$$f_{W_1}(r) = \frac{r}{\sigma^2_{W_1}} \, (\frac{r}{\lambda})^{1/2} \, I_{1/2}(\frac{\lambda r}{\sigma^2_{W_1}}) \, e^{-\frac{\lambda^2+r^2}{2\sigma^2_{W_1}}} , \tag{32}$$

with $\lambda = \|\mu_1\|$ and $I_\nu(\cdot)$ is the modified Bessel function of order ν. To proceed, it is convenient to write

$$\psi_{mle}(W_1) \; = \; h_{mle}(\|W_1\|) \, \frac{W_1}{\|W_1\|} \, ,$$

with $h_{mle}(t) = \min\{\frac{2m}{3}, t\}$. We obtain

$$\begin{aligned} \mathbb{E}(\|\psi_{mle}(W_1) - \mu_1\|^2) \; &= \; \mathbb{E}^R \, \mathbb{E}[\|h_{mle}(\|W_1\|) \, \frac{W_1}{\|W_1\|} - \mu_1\|^2 | \, R \,] \\ &= \; \lambda^2 \, + \, \mathbb{E} \, h^2_{mle}(R) & (33) \\ &\quad - 2 \, \lambda \, \mathbb{E} \left[h_{mle}(R) \, \frac{I_{3/2}(\frac{\lambda R}{\sigma^2_{W_1}})}{I_{1/2}(\frac{\lambda R}{\sigma^2_{W_1}})} \right] , & (34) \end{aligned}$$

where we make use of the identity (e.g., Berry, 1990; Robert, 1990)

$$\mathbb{E}\left[\frac{\mu' S}{\|S\|} \, | \, \|S\| = r \right] \; = \; \|\mu_S\| \, \frac{I_{p/2}(\|\mu_S\| r / \sigma^2_S)}{I_{(p-2)/2}(\|\mu_S\| r / \sigma^2_S)} \, ,$$

for $S \sim N_p(\mu_S, \sigma^2_S I_p)$. Finally, combining the density of R in (32) with (33) yields the expression

$$\begin{aligned} &\mathbb{E}(\|\psi_{mle}(W_1) - \mu_1\|^2) \\ &\quad = \lambda^2 \, + \, \int_0^\infty f_{W_1}(r) \, h_{mle}(r) \left\{ h_{mle}(r) - 2\lambda \, \frac{I_{3/2}(\lambda r / \sigma^2_{W_1})}{I_{1/2}(\lambda r / \sigma^2_{W_1})} \right\} dr \, , \quad (35) \end{aligned}$$

which generates an explicit expression for the Kullback-Leibler risk in (31).[2]

Acknowledgements

Author Marchand gratefully acknowledges the research support from the Natural Sciences and Engineering Research Council of Canada. We are grateful to Bill Strawderman for useful discussions. Finally, we are thankful to two referees for constructive comments and queries that led to an improved version of the manuscript.

[2] We point out that the identities $I_{1/2}(z) = \sqrt{\frac{2}{\pi z}} \, \sinh(z)$ and $I_{3/2}(z) = \sqrt{\frac{2}{\pi z}} \{\cosh(z) - \frac{\sinh(z)}{z}\}$, $z \geq 0$, lead to further simplifications.

References

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

[2] Aitchison, J. & Dunsmore, I.R., (1975). *Statistical Prediction Analysis*. Cambridge University Press. MR0408097

[3] Arellano-Valle, R.B., Branco, M.D., & Genton, M.G. (2006). A unified view on skewed distributions arising from selections. *Canadian Journal of Statistics*, **34**, 581-601. MR2347047

[4] Arnold, B.C. & Beaver, R.J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test*, **11**, 7-54. MR1915776

[5] Arnold, B.C., Beaver, R.J., Groeneveld, R.A., & Meeker, W.Q. (1993). The nontruncated marginal of a truncated bivariate normal distribution. *Psychometrika*, **58**, 471-488. MR1237615

[6] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178. MR0808153

[7] Berry, C. (1990). Minimax estimation of a bounded normal mean vector. *Journal of Multivariate Analysis*, **35**, 130-139. MR1084946

[8] Blumenthal, S. & Cohen, A. (1968). Estimation of the larger translation parameter. *Annals of Mathematical Statistics*, **39**, 502-516. MR0223006

[9] Brandwein, A.C. & Strawderman, W.E. (1980). Minimax estimation of location parameters for spherically symmetric distributions with concave loss. *Annals of Statistics*, **8**, 279-284. MR0560729

[10] Brown, L.D., George, E.I., & Xu, X. (2008). Admissible predictive density estimation. *Annals of Statistics*, **36**, 1156-1170. MR2418653

[11] Brown, L.D. (1986). *Foundations of Exponential Families*. IMS Lecture Notes, Monograph Series **9**, Hayward, California. MR0882001

[12] Cohen, A. & Sackrowitz, H.B. (1970). Estimation of the last mean of a monotone sequence. *Annals of Mathematical Statistics*, **41**, 2021-2034. MR0270483

[13] Corcuera, J.M. & Giummolè, F. (1999). A generalized Bayes rule for prediction. Scandinavian Journal of Statistics, **26**, 265-279. MR1707658

[14] Csiszàr, I. (1967). Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar. **2**, 299-318. MR0219345

[15] Dunson, D.B. & Neelon, B. (2003). Bayesian inference on order-constrained parameters in generalized linear models. *Biometrics*, **59**, 286-295. MR1987395

[16] Fourdrinier, D. & Marchand, É. (2010). On Bayes estimators with uniform priors on spheres and their comparative performance with maximum likelihood estimators for estimating bounded multivariate normal means. *Journal of Multivariate Analysis*, **101**, 1390-1399. MR2609500

[17] Fourdrinier, D., Marchand, É., Righi, A. & Strawderman, W.E. (2011). On improved predictive density estimation with parametric constraints. *Electronic Journal of Statistics*, **5**, 172-191. MR2792550

[18] George, E.I., Liang, F. & Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Annals of Statistics*, **34**, 78-91. MR2275235

[19] Ghosh, M., Mergel, V. & Datta, G.S. (2008). Estimation, prediction and the Stein phenomenon under divergence loss. *Journal of Multivariate Analysis*, **99**, 1941-1961. MR2466545

[20] Gupta, R.C. & Gupta, R.D. (2004). Generalized skew normal model. *Test*, **13**, 501-524. MR2154011

[21] Hartigan, J. (2004). Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, **67**, 285-288. MR2060127

[22] Hwang, J.T.G. & Peddada, S.D. (1994). Confidence interval estimation subject to order restrictions. *Annals of Statistics*, **22**, 67-93. MR1272076

[23] 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. MR2529965

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

[25] Kubokawa, T. (2005). Estimation of bounded location and scale parameters. *Journal of the Japanese Statistical Society*, **35**, 221-249. MR2328426

[26] Kubokawa, T., Marchand, É., & Strawderman, W.E. (2015). On predictive density estimation for location families under integrated squared error loss. *Journal of Multivariate Analysis*, **142**, 57-74. MR3412739

[27] Kubokawa, T., Marchand, É. & Strawderman, W.E. (2017). On predictive density estimation for location families under integrated absolute value loss. *Bernoulli*, **23**, 3197-3212. MR3654804

[28] Lee, C.I.C. (1981). The quadratic loss of isotonic regression under normality. *Annals of Statistics*, **9**, 686-688. MR0615447

[29] Liseo, B. & Loperfido, N. (2003). A Bayesian interpretation of the multivariate skew-normal distribution. *Statistics & Probability Letters*, **61**, 395-401. MR1959075

[30] Marchand, É., Jafari Jozani, M. & Tripathi, Y.M. (2012). On the inadmissibility of various estimators of normal quantiles and on applications to two-sample problems with additional information, *Contemporary Developments in Bayesian analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman*, Institute of Mathematical Statistics Volume Series, **8**, 104-116. MR3202506

[31] Marchand, É., Perron, F., & Yadegari, I. (2017). On estimating a bounded normal mean with applications to predictive density estimation. *Electronic Journal of Statistics*, **11**, 2002-2025. MR3651022

[32] Marchand, É. & Perron, F. (2001). Improving on the MLE of a bounded normal mean. *Annals of Statistics*, **29**, 1078-1093. MR1869241

[33] Marchand, É. & Strawderman, W.E. (2004). Estimation in restricted parameter spaces: A review. *Festschrift for Herman Rubin*, IMS Lecture Notes-Monograph Series, **45**, 21-44. MR2126884

[34] Maruyama, Y. & Strawderman, W.E. (2012). Bayesian predictive densities for linear regression models under $\alpha-$divergence loss: Some results and open

problems. *Contemporary Developments in Bayesian analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman*, Institute of Mathematical Statistics Volume Series, **8**, 42-56. MR3202501

[35] Park, Y., Kalbfleisch, J.D. & Taylor, J. (2014). Confidence intervals under order restrictions. *Statistica Sinica*, **24**, 429-445. MR3183692

[36] Robert, C. (1996). Intrinsic loss functions. *Theory and Decision*, **40**, 192-214. MR1385186

[37] Robert, C. (1990). Modified Bessel functions and their applications in probability and statistics. *Statistics and Probability Letters*, **9**, 155-161. MR1045178

[38] Sadeghkhani, A. (2017). Estimation d'une densité prédictive avec information additionnelle. Ph.D. thesis. Université de Sherbrooke (`http://savoirs.usherbrooke.ca/handle/11143/11238`)

[39] Shao, P.Y.-S. & Strawderman, W. (1996). Improving on the mle of a positive normal mean. *Statistica Sinica*, **6**, 275-287. MR1379061

[40] Silvapulle, M.J. & Sen, P.K. (2005). *Constrained statistical inference. Inequality, order, and shape restrictions.* Wiley Series in Probability and Statistics. Wiley-Interscience. MR2099529

[41] Spiring, F.A. (1993). The reflected normal loss function. *Canadian Journal of Statistics*, **31**, 321-330. MR1241064

[42] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, **9**, 1135-1151. MR0630098

[43] van Eeden, C. & Zidek, J.V. (2001). Estimating one of two normal means when their difference is bounded. *Statistics & Probability Letters*, **51**, 277-284. MR1822735

[44] van Eeden, C. & Zidek, J.V. (2003). Combining sample information in estimating ordered normal means. *Sankhyā A*, **64**, 588-610. MR1985401

[45] van Eeden, C. (2006). *Restricted parameter space problems: Admissibility and minimaxity properties.* Lecture Notes in Statistics, **188**, Springer. MR2265239

[46] Zelazo, P.R., Zelazo, N.A. & Kolb, S. (1972). "Walking in the newborn". *Science*, **176**, 314-315.