Afrika Statistika
ISSN 2316-0861
ISSN 2316-090X

Afrika Statistika
Vol. 14 (2), 2019, pages 1951 - 1976.
DOI: http://dx.doi.org/10.16929/as/2019.1951.144

Minimaxity and Admissibility of Predictive Density Estimators Under S-Hellinger Distances

Younes Ommane [(1,*)] **and Idir Ouassou**[(1,2)]

[1] Ecole Nationale des Sciences Appliquées, Marrakech, Université Cadi Ayyad (Morocco)
[2] Université Polytechnique Mohammed VI, 43140 Ben Guerir, (Morrocco)

Received on November 1, 2018; Accepted on May 25, 2019, Published Online on July 28, 2019

Copyright © 2019, Afrika Statistika and The Statistics and Probability African Society (SPAS). All rights reserved

Abstract. In this paper, we consider the study of the efficiency of predictive density estimators of multivariate observables measured by the frequentist risk corresponding to S-Hellinger distances as a set of loss functions (for every $\alpha \in [0, 1]$). The main themes, revolve around the inefficiency of minimum risk equivariant (MRE) predictors in high enough dimensions and about the inefficiency of plug-in estimators. We improve the plug-in for a dual point estimation loss with or without expanding the scale. A link between the S-Hellinger distances risk of plug-in type estimators and the risk under reflected normal loss for point estimation is established, bringing into play all the established literature on Stein type dominators. Further, we suggest dominant estimators with or without the presence of restrictions on the unknown mean parameter. Ultimately we prove under the new measure of goodness-of-fit dominance results under a restricted parameter space (multivariate and univariate).

Key words: S-Hellinger Distances, Minimaxity, Admissibility, Stein estimation, concave loss, Predictive density estimation.
AMS 2010 Mathematics Subject Classification : 62H12, 62C20, 62F10.

*Corresponding author: Younes Ommane, younes.ommane@ced.uca.ac.ma
Idir OUASSOU : i.ouassou@uca.ma

Résumé (French abstract) Dans cet article, nous considérons l'étude de l'efficacité des estimateurs de densité prédictives d'observables multivariés mesurés par le risque fréquentiste correspondant aux distances S-Hellinger en tant qu'ensemble de fonctions de perte (pour chaque $\alpha \in [0,1]$) . Les thèmes principaux tournent autour de l'inefficacité des prédicteurs ERM (équivariant à risque minimal) dans des dimensions suffisamment élevées, et de l'inefficacité des estimateurs plug-in. Nous améliorons l'estimateur plug-in pour le problème ponctuel dual avec ou sans extension du paramétre d'échelle. Un lien entre le risque de distance S-Hellinger des estimateurs de type plug-in et le risque sous perte normale refléchie pour l'estimation ponctuelle est établi, en mettant en jeu toute la littérature établie sur les dominateurs de type Stein. De plus, nous suggérons des estimateurs dominants avec ou sans la présence de restrictions sur le paramètre moyen inconnu. En fin de compte, nous prouvons que la nouvelle mesure de divergence permet d'obtenir des résultats de dominance dans un espace paramétrique restreint (multivarié et univarié).

1. Introduction

Let X and Y be two normal d-variate random variables, independently distributed, and let p and q be the respective pdf's (probability density functions), such that $X|\theta \sim \mathcal{N}_d(\theta, \sigma_x^2 I_d)$ and $Y|\theta \sim \mathcal{N}_d(\theta, \sigma_y^2 I_d)$. The pivotal problem is predicting the unknown mean vector θ, by observing X, where σ_x^2, σ_y^2, p and q are known. We will assess the goodness of prediction fit of a given predictive estimate $\hat{q}(y|x)$ from the target density $q(y|\theta)$, via the family of S-Hellinger Distances (introduced by Ghosh et al (2017)), defined as follows

$$\begin{aligned} D_{S_\alpha}(q,\hat{q}) &= \frac{2}{1+\alpha}\int_{\mathbb{R}^d}\left(\hat{q}^{\frac{\alpha+1}{2}}(y|x) - q^{\frac{\alpha+1}{2}}(y|\theta)\right)^2 dy \qquad (1)\\ &= \frac{2}{1+\alpha}||\hat{q}^{\frac{\alpha+1}{2}}(y|x) - q^{\frac{\alpha+1}{2}}(y|\theta)||_2^2 \end{aligned}$$

with one tuning parameter α taken in $[0,1]$, where $||.||_2$ is the usual L_2-norm. This is a generalized family of L_2 type distances, which generates the twice-squared Hellinger distance at $\alpha = 0$, and exactly the L_2-norm for $\alpha = 1$. In fact (1) connects the ordinary Hellinger distance to the L_2-norm smoothly through the parameter α. However, (1) is not exactly a distance, but rather after a slight makeup, i.e. this entity $\left(\frac{1+\alpha}{2}D_{S_\alpha}(q,\hat{q})\right)^{\frac{1}{2}}$ is now a genuine distance. Furthermore, we may use multivariate location and covariance estimation using the S-Hellinger distances, since it presents some easiness, which is mainly due to the fact that it corresponds to a distance metric.

An extension of the inefficiency of MRE predictors in high enough dimensions to our case is established, as well as the efficiency of plug-in estimators by either improving on the plug-in for a dual point estimation loss or expanding the scale. For the plug-in estimation problem, it is the reflected normal loss (introduced by

Spiring (2011)), that turns out to be the dual loss function to S-Hellinger distances for predictive estimation, which is also a bounded loss, this latter property seduces many statistical decision makers, and brings into play all the established results on Stein effect (Stein (1956)). Afterwards, we introduce dominating estimators with or without the presence of restrictions on the unknown mean parameter. Ultimately we extended under the new measure of goodness-of-fit dominance (S-Hellinger distances) Hartigan type results, under a restricted parameter space (multivariate and univariate).

The organization of this paper is outlined as follows. In Section 1, we introduce some preliminary identities and results, namely, an essential identity in general (matrix variate variance) and then a particular case (degenerate variance), then we deduce the expression of S-Hellinger Distances between two gaussian distributions. Afterwards, we establish the expression of the generalized Bayes estimator for a prior $\pi(\theta)$ under S-Hellinger distances, as a consequence we derive the MRE predictor for the flat prior ($\pi(\theta) = 1$), and we show its minimaxity, where we retrieve the established results for Kullback-Leibler loss (e.g. Kubokawa et al (2015)) and Liang et al (2004)), and for L_2-norm ($\alpha = 1$), Kubokawa et al (2015). We deduce the inadmissibility for $d \geq 3$ of the MRE estimator Stein (1956), finally we give an example in the univariate case. In Section 3, we swiftly move to study plug-in type estimators, we firstly evaluate the duality and efficiency of density estimators $\mathcal{N}_d(\hat{\theta}(x), c^2\sigma_y^2 I_d)$, where $c^2 > 1$, the scale-expanding factor. We established a sufficient condition of domination when the estimators of θ are $\hat{\theta}(x) = X$ and $\hat{\theta}(x) = aX$, with $0 < a \leq 1$. To emphasize these findings, we provide a bunch of classic dominating estimators when $d \geq 3$, we provide numerous numerical evaluations particularly for the positive part of the James-Stein estimator. For sake of avoiding congestion, we restricted our study on three main members of S-Hellinger family, namely for $\alpha \in \{0, 0.5, 1\}$, standing respectively for: twice-squared Hellinger distance, mid-range of the set $[0, 1]$ and L_2-norm, (Example (6) Baranchick type estimators), where we recover domination under the same sufficient dominance condition on the scale-expanding factor, the estimator of θ is either X or aX, as well as for other cases, emphasizing the existence of other areas of domination, with other conditions on the scale-expanding factor c^2 ($c^2 > 1$), and the ratio of variances $r = \sigma_x^2/\sigma_y^2$. Ultimately, we show that the Hartigan type results Hartigan (2004), shown previously in Kubokawa et al (2015), hold under S-hellinger distances.

2. Bayes, best equivariant and minimax estimation in the normal case

Let X and Y be two normal d-variate conditionally independent random variables given $\theta \in \mathbb{R}^d$, and let p and q be respectively the pdfs of X given θ and Y given θ, with σ_x^2 and σ_y^2 being their respective variances, such that

$$X|\theta \sim p(x|\theta) \stackrel{d}{=} \mathcal{N}_d(\theta, \sigma_x^2 I_d) \text{ and } Y|\theta \sim q(y|\theta) \stackrel{d}{=} \mathcal{N}_d(\theta, \sigma_y^2 I_d), \quad (2)$$

where "$\stackrel{d}{=}$" stands for equality in distribution.

In the next lemma we give a general expression of S-Hellinger distances as a loss function of two normal densities, for two normal distributions, namely, $q_1(y|\theta_1)$ and $q_2(y|\theta_2)$, where σ_1^2 and σ_2^2 are their respective variances.

Lemma 1. *For* $0 \leq \alpha \leq 1$, *and under the model (2) such that* $p(x|\theta) = q(y|\theta_1)$ *for* $\theta_1 \in \mathbb{R}_d$, $q(y|\theta) = q(y|\theta_2)$ *for* $\theta_2 \in \mathbb{R}_d$, $\sigma_x^2 = \sigma_1^2$ *and* $\sigma_y^2 = \sigma_2^2$, *we obtain*

$$
\begin{aligned}
D_{S_\alpha}(q_1, q_2) &= \frac{2}{1+\alpha} \int_{\mathbb{R}^d} \left(q_1^{\frac{1+\alpha}{2}}(y|\theta_1) - q_2^{\frac{1+\alpha}{2}}(y|\theta_2) \right)^2 dy \\
&= \frac{2(2\pi)^{-\frac{d\alpha}{2}}}{(1+\alpha)^{\frac{d}{2}+1}} \left((\sigma_1^2)^{-\frac{d\alpha}{2}} + (\sigma_2^2)^{-\frac{d\alpha}{2}} - 2 \left(\frac{\sigma_1^2 + \sigma_2^2}{2(\sigma_1^2\sigma_2^2)^{\frac{1-\alpha}{2}}} \right)^{-\frac{d}{2}} \right. \\
&\quad \times \left. \exp\left(-\frac{||\theta_1 - \theta_2||^2}{2\frac{2}{1+\alpha}(\sigma_1^2 + \sigma_2^2)} \right) \right).
\end{aligned} \tag{3}
$$

Proof. It is enough to directly apply Lemma 3, for $\alpha_1 = \alpha_2 = (1+\alpha)/2$. □

The next proposition gives a general expression of the bayesian predictive density estimator (BPDE) of Y given X, under S-Hellinger distances, for any prior $\pi(\theta)$ (it could be improper).

Proposition 1. *Under S-Hellinger distances and for a prior* π *(it can be an improper prior), the bayesian predictive density estimator of* $q(y|\theta)$ *is given by*

$$
\hat{q}_\pi(y|x) = \frac{k_{\frac{\alpha+1}{2}}^{\frac{2}{\alpha+1}}(y,x)}{\int_{\mathbb{R}^d} k_{\frac{\alpha+1}{2}}^{\frac{2}{\alpha+1}}(y,x)dy}, \tag{4}
$$

with

$$
k_{(*)}(y,x) = \int_{\mathbb{R}^d} q^{(*)}(y|\theta)\pi(\theta|x)d\theta,
$$

whenever (4) exists.

Proof. Let $\hat{q}(y|x)$ be a BPDE of $q(y|\theta)$, and let $\rho(\hat{q})$ be its corresponding posterior risk, which is stated according to Fubini's theorem, denoting $\omega = \hat{q}^{\frac{1+\alpha}{2}}(y|x)$, as the following

$$
\begin{aligned}
\rho(\hat{q}) &= \int_{\mathbb{R}^d} D_{S_\alpha}(q, \hat{q})\pi(\theta|x)d\theta \\
&= \frac{2}{1+\alpha} \int_{\mathbb{R}^d} (k_{1+\alpha}(y,x) + \hat{q}^{\alpha+1}(y|x) - 2k_{\frac{\alpha+1}{2}}(y,x)\hat{q}^{\frac{\alpha+1}{2}}(y|x))dy \\
&= \frac{2}{1+\alpha} \int_{\mathbb{R}^d} (\omega^2 - 2\omega k_{\frac{\alpha+1}{2}}(y,x) + k_{1+\alpha}(y,x))dy \\
&= h(\omega),
\end{aligned}
$$

by Holder's inequality we establish that the function h reaches its unnormalized minimum at $k^{\frac{2}{1+\alpha}}_{\frac{1+\alpha}{2}}(y,x)$. Therefore, the BPDE denoted by $\hat{q}_\pi(y|x)$ satisfies $\hat{q}_\pi(y|x) \propto k^{\frac{2}{1+\alpha}}_{\frac{1+\alpha}{2}}(y,x)$, thus, by normalizing this latter we retrieve Equation (4), which concludes the proof. □

Remark 1. We highlight that the predictive density is proper if and only if the posterior density is proper. The BPDE's expression in (4) includes the established expression of the BPDE under L_2-norm in Kubokawa et al (2015) (for $\alpha = 1$ under S-Hellinger distances), which is similar to Aitchison's expression under KL in Aitchison (1975).

Example 1 (**Normal prior**)**.** We suppose that the distribution of θ is $\mathcal{N}_d(0, t^2 I_d)$, then we obtain that the posterior density is $\mathcal{N}_d\left(\frac{t^2}{t^2+\sigma_x^2}, \frac{t^2\sigma_x^2}{t^2+\sigma_x^2} I_d\right)$ and

$$k_{\frac{1+\alpha}{2}}(y,x) \propto \mathcal{N}_d\left(\frac{t^2}{t^2+\sigma_x^2}, \left(\frac{t^2\sigma_x^2}{t^2+\sigma_x^2} + \frac{2\sigma_y^2}{1+\alpha}\right) I_d\right),$$

and hence,

$$k^{\frac{2}{1+\alpha}}_{\frac{1+\alpha}{2}}(y,x) \propto \mathcal{N}_d\left(\frac{t^2}{t^2+\sigma_x^2}, \left(\frac{1+\alpha}{2}\frac{t^2\sigma_x^2}{t^2+\sigma_x^2} + \sigma_y^2\right) I_d\right).$$

Therefore,

$$\hat{q}_\pi(y|x,t) \sim \mathcal{N}_d\left(\frac{t^2}{t^2+\sigma_x^2}, \left(\frac{1+\alpha}{2}\frac{t^2\sigma_x^2}{t^2+\sigma_x^2} + \sigma_y^2\right) I_d\right).$$

Example 2 (***Non informative prior***)*.* In this example, we consider the flat prior $\pi(\theta) = 1$, under the model (2), then we compute its corresponding BPDE, which coincides with the minimum risk equivariant (MRE) estimator, denoted by $\hat{q}_{mre}(y|x)$, becomes

$$\hat{q}_{mre}(y|x) = \hat{q}_{\pi=1}(y|x) = \left(\sigma_y^2\left(\frac{1+\alpha}{2}r+1\right)\right)^{-\frac{d}{2}} \phi\left(\frac{y-x}{\sigma_y\sqrt{\frac{1+\alpha}{2}r+1}}\right), \quad (5)$$

with $r = \sigma_x^2/\sigma_y^2$. We can also provide the expressions of $D_{S_\alpha}(q, \hat{q}_{mre})$ and its corresponding frequentist risk $R_{S_\alpha}(q, \hat{q}_{mre})$, under S-Hellinger distances:

$$D_{S_\alpha}(q, \hat{q}_{mre}) = \frac{2(2\pi\sigma_y^2)^{-\frac{d}{2}}}{(1+\alpha)^{\frac{d}{2}+1}} \times \left[1 + \left(\frac{1+\alpha}{2}r + 1\right)^{-\frac{d\alpha}{2}} - 2\left(\frac{\frac{1+\alpha}{2}r + 2}{2\left(\frac{1+\alpha}{2}r+1\right)^{\frac{1-\alpha}{2}}}\right)^{-\frac{d}{2}} e^{-\frac{||x-\theta||^2}{2\sigma_y^2(r+\frac{4}{1+\alpha})}}\right], \tag{6}$$

$$R_{S_\alpha}(q, \hat{q}_{mre}) = \frac{2(2\pi\sigma_y^2)^{-\frac{d}{2}}}{(1+\alpha)^{\frac{d}{2}+1}} \times \left[1 + \left(\frac{1+\alpha}{2}r + 1\right)^{-\frac{d\alpha}{2}} - 2\left(\frac{1+\alpha}{2}r+1\right)^{-\frac{d}{4}(1+\alpha)}\right]. \tag{7}$$

Remark 2. (a) For $\alpha = 1$ we find the same expression of MRE estimator under L_2-norm established in Kubokawa et al (2015), which coincides with Aitchison's BPDE, under Kullback-Leibler loss, indeed:

$$\hat{q}_{mre}(y|x) = (\sigma_y^2(r+1))^{-\frac{d}{2}}\phi\left(\frac{y-x}{\sigma_y\sqrt{r+1}}\right).$$

We notice that for $\alpha = 1$ we retrieve as well the risk expression of the MRE estimator under L_2-norm established in Kubokawa et al (2015):

$$R_{S_\alpha}(q, \hat{q}_{mre}) = (4\pi\sigma_y^2)^{-\frac{d}{2}}(1-(r+1)^{-\frac{d}{2}}).$$

(b) For $\alpha = 0$ under S-Hellinger distances, we find the expression of the minimum risk equivariant estimator under twice-squared Hellinger distance:

$$\hat{q}_{mre}(y|x) = (\sigma_y^2(r/2+1)t)^{-\frac{d}{2}}\phi\left(\frac{y-x}{\sigma_y\sqrt{\frac{r}{2}+1}}\right). \tag{8}$$

Therefore, the risk expression of the MRE estimator under twice-squared Hellinger distance, where the MRE estimator corresponding to the squared Hellinger distance:

$$R_{S_\alpha}(q, \hat{q}_{mre}) = 2(1-(1+r/2)^{-\frac{d}{4}}).$$

*Example 3 (**Univariate Laplace distribution**).* In this example we consider the Laplace distribution, so that X and Y are independently Laplace distributed, and the prior distribution of θ is also Laplace distributed (centered for sake of simplicity), i.e.:

$$\begin{aligned} p(x|\theta) &= \frac{1}{2\gamma_x} e^{-\frac{|x-\theta|}{\gamma_x}}, \\ q(y|\theta) &= \frac{1}{2\gamma_y} e^{-\frac{|y-\theta|}{\gamma_y}}, \\ \pi(\theta) &= \frac{1}{2\gamma_\theta} e^{-\frac{|\theta|}{\gamma_\theta}}, \end{aligned}$$

with γ_x, γ_x and γ_θ (a hyperparameter) are respectively the scale parameters of X, Y and θ. The posterior density turns out to be also Laplace distributed, i.e.

$$Laplace\left(\frac{\gamma_\theta}{\gamma_x+\gamma_\theta}x, \frac{\gamma_x\gamma_\theta}{\gamma_x+\gamma_\theta}\right),$$

Hence the unnormalized BPDE $k^{\frac{2}{1+\alpha}}_{\frac{1+\alpha}{2}}(y,x)$ is proportional to

$$Laplace\left(\frac{\gamma_\theta}{\gamma_x+\gamma_\theta}x, \frac{1+\alpha}{2}\left(\frac{\gamma_x\gamma_\theta}{\gamma_x+\gamma_\theta}+\frac{2\gamma_y}{1+\alpha}\right)\right),$$

Thus, the corresponding BPDE states as follows:

$$\hat{q}_\pi(y|x,\gamma_\theta) \sim Laplace\left(\frac{\gamma_\theta}{\gamma_x+\gamma_\theta}x, \frac{1+\alpha}{2}\left(\frac{\gamma_x\gamma_\theta}{\gamma_x+\gamma_\theta}+\frac{2\gamma_y}{1+\alpha}\right)\right). \tag{9}$$

The next proposition provides another formula for the BPDE, presenting a convoluting form between $q^{\frac{1+\alpha}{2}}$ and a given arbitrary function g, when the posterior density coincides with g at the point $(\theta-\hat{\theta}(x))$, and gives the expression of the MRE (Minimum Risk equivariant) estimator (i.e. for $\pi(\theta)=1$) and establishes its minimaxity.

Proposition 2. *We have the following facts.*

(1) Under the model (2), if the posterior density $\pi(\theta|x)$ is of the form $g(\theta-\hat{\theta}(x))$, where g an arbitrary function, and $\hat{\theta}(x)$ any point estimator of θ, then:
For any prior distribution $\pi(\theta)$, the BPDE under S-Hellinger distances, states as the following

$$\hat{q}_\pi(y|x) = \frac{(q^{\frac{1+\alpha}{2}} * g)^{\frac{2}{1+\alpha}}(y-\hat{\theta}(x))}{\int_{\mathbb{R}^d}(q^{\frac{1+\alpha}{2}} * g)^{\frac{2}{1+\alpha}}(y-\hat{\theta}(x))dy} \tag{10}$$

(2) The MRE estimator associated to (10) under S-Hellinger distances is given by

$$\hat{q}_{mre}(y|x) = \frac{(q^{\frac{1+\alpha}{2}} * p)^{\frac{2}{1+\alpha}}(y|x)}{\int_{\mathbb{R}^d}(q^{\frac{1+\alpha}{2}} * p)^{\frac{2}{1+\alpha}}(y|x)dy}. \tag{11}$$

Furthermore, $\hat{q}_{mre}(y|x)$ is minimax.

proof. See the appendix. □

*Example 4 (**Univariate Laplace distribution**).* According to Example (3), we deduce the expression of the MRE estimator under S-Hellinger distances, by noticing that it suffices to substitute $\frac{\gamma_x \gamma_\theta}{\gamma_x+\gamma_\theta}$ by γ_x, and $\frac{\gamma_x}{\gamma_x+\gamma_\theta}x$ by x, as pointed out by (9), since $\pi(\theta|x)$ reduces to $p(x|\theta)$ when $\pi(\theta)=1$, therefore,

$$\hat{q}_{mre}(y|x) \sim Laplace\left(x, \left(\frac{1+\alpha}{2}r+1\right)\gamma_y\right),$$

with $r=\gamma_x/\gamma_y$.

3. Plug-in type estimators in the normal case

3.1. Duality and efficiency of density estimators $\mathcal{N}_d(\hat{\theta}(x), c^2\sigma_y^2 I_d)$

We consider in this subsection the normal model, in which we aim to assess the performance of density estimators $\hat{q}_{c^2,\hat{\theta}}(y|\hat{\theta} \sim \mathcal{N}_d(\hat{\theta}(x), c^2\sigma_y^2 I_d)$ which combines both a plug-in component with $\hat{\theta}(x)$ being an estimate of θ, and a modification of variance component for $c^2 \neq 1$ Kubokawa et al (2015). We give a sufficient condition on the scale-expanding factor c of the efficiency of such estimators related to the efficiency of the point estimator $\hat{\theta}(x)$ in estimating θ, as well as the degree of variance expansion governed by the choice of $c^2>1$. With respect to the duality with the point estimation problem, it remains the reflected normal loss, under our loss function, S-Hellinger distances, denoted as

$$D_\gamma(\theta,\hat{\theta}) = 1-\exp\left(-\frac{||\hat{\theta}-\theta||^2}{2\gamma}\right), \tag{12}$$

with $\gamma>0$, which brings back the established results in Kubokawa et al (2015), whereas it's the quadratic loss $(||.-.||^2)$ that intervenes as a dual loss for Kullback-Leibler (George et al (2006); Brown et al (2008)), it's worthy to mention that

$$\lim_{\gamma\to\infty} 2\gamma D_\gamma(\theta,\hat{\theta}) = ||\theta-\hat{\theta}||^2.$$

The expression of S-Hellinger distances are given by

$$\begin{aligned} D_{S_\alpha}(q, \hat{q}(y|\hat{\theta}(x))) &= \frac{4(2\pi\sigma_y^2)^{-\frac{d\alpha}{2}}}{(1+\alpha)^{\frac{d}{2}+1}}\left(1-\exp\left(-\frac{||\hat{\theta}(x)-\theta||^2}{2\frac{4\sigma_y^2}{1+\alpha}}\right)\right) \\ &= \frac{4(2\pi\sigma_y^2)^{-\frac{d\alpha}{2}}}{(1+\alpha)^{\frac{d}{2}+1}} D_{\gamma_s}(\theta,\hat{\theta}), \end{aligned} \tag{13}$$

where $\gamma_s = \frac{4\sigma_y^2}{1+\alpha}$.

In the next lemma we compute the S-Hellinger distances of our new candidate $\hat{q}_{c^2,\hat{\theta}}$, using an auxiliary lemma:

Lemma 2. *Let $\hat{q}_{c^2,\hat{\theta}}(y|\hat{\theta} \sim \mathcal{N}_d(\hat{\theta}(x), c^2\sigma_y^2 I_d)$ be a scale-expanded estimator of $q(y|\theta)$, if $\hat{q}_{c^2,\hat{\theta}}(y|\hat{\theta} \sim \mathcal{N}_d(\hat{\theta}(x), c^2\sigma_y^2 I_d)$, then its corresponding S-Hellinger distances are given by*

$$D_{S_\alpha}(q, \hat{q}_{c^2,\hat{\theta}}) = \frac{2(2\pi\sigma_y^2)^{-\frac{d\alpha}{2}}}{(1+\alpha)^{\frac{d}{2}+1}} \Bigg(1 + (c^2)^{-\frac{d\alpha}{2}} - 2\left(\frac{1+c^2}{2(c^2)^{\frac{1-\alpha}{2}}}\right)^{-\frac{d}{2}} \times \exp\left(-\frac{||\hat{\theta}(x)-\theta||^2}{2\frac{2\sigma_y^2}{1+\alpha}(1+c^2)}\right)\Bigg). \tag{14}$$

Proof This is a direct application of Lemma 3, by making these substitutions $q_1 = q$, $q_2 = \hat{q}_{c^2,\hat{\theta}}$, $\theta_1 = \theta$, $\theta_2 = \hat{\theta}(x)$, $\sigma_1^2 = \sigma_y^2$ and $\sigma_2^2 = c^2\sigma_y^2$.$\square$

Theorem 1. *Under the model (2), we have:*

1. *For a fixed c^2. We have $\hat{q}_{c^2,\hat{\theta}_1} \sim \mathcal{N}_d(\hat{\theta}_1, c^2\sigma_y^2 I_d)$ improves on $\hat{q}_{c^2,\hat{\theta}_2} \sim \mathcal{N}_d(\hat{\theta}_2, c^2\sigma_y^2 I_d)$ under S-Hellinger distances iff $\hat{\theta}_2$ improves on $\hat{\theta}_1$ under reflected normal loss $a + bD_{\gamma_s}(\hat{\theta}, \theta)$, with $\gamma_s = 4\sigma_y^2(1+c^2)/(1+\alpha)$, where*

$$a = \frac{2(2\pi\sigma_y^2)^{-\frac{d\alpha}{2}}}{(1+\alpha)^{\frac{d}{2}+1}} \left(1 + (c^2)^{-\frac{d\alpha}{2}} - 2\left(\frac{1+c^2}{2(c^2)^{\frac{1-\alpha}{2}}}\right)^{-\frac{d}{2}}\right),$$

   and

$$b = \frac{4}{1+\alpha}\left(\frac{(1+\alpha)(2\pi\sigma_y^2)^{\alpha}(1+c^2)}{2(c^2)^{\frac{1-\alpha}{2}}}\right)^{-\frac{d}{2}}.$$

2. *For $\hat{\theta}(x) = x$, and let r be the ratio of variances, i.e. $r = \sigma_x^2/\sigma_y^2$. The risk $R_{S_\alpha}(q, \hat{q}_{c^2,x})$ is constant, and states as*

$$\frac{2(2\pi\sigma_y^2)^{-\frac{d\alpha}{2}}}{(1+\alpha)^{\frac{d}{2}+1}} \left(1 + (c^2)^{-\frac{d\alpha}{2}} - 2\left(\frac{\frac{1+\alpha}{2}r + 1 + c^2}{2(c^2)^{\frac{1-\alpha}{2}}}\right)^{-\frac{d}{2}}\right). \tag{15}$$

3. *For all d, the constant (and minimax) risk of $\hat{q}_{mre}$, corresponding to the optimal choice of c^2, $c_*^2 = \frac{1+\alpha}{2}r + 1$ is equal to*

$$\frac{2(2\pi\sigma_y^2)^{-\frac{d\alpha}{2}}}{(1+\alpha)^{\frac{d}{2}+1}} \left(1 + (c_*^2)^{-\frac{d\alpha}{2}} - 2(c_*^2)^{-\frac{d}{4}(\alpha+1)}\right). \tag{16}$$

4. *Furthermore, for $0 \leq \alpha \leq 1$, all estimators $\hat{q}_{c^2,\theta}$ dominate the plug-in estimator $\hat{q}_{1,\hat{\theta}}$ for all d, whenever $1 < c^2 < (\frac{1-\alpha}{1+\alpha})c_*^2$.*

Proof. Let us proceed by parts.
Part (1) follows directly from Lemma (2).

For part (2), using the fact that $\hat{\theta}(x) = x$ and the fact that $\hat{q}_{c^2,x}(y|x) \stackrel{d}{=} \mathcal{N}_d(x,(c^2\sigma_y^2)I_d)$ we obtain (14), the use of (21) leads to (15).

Part (3) The value of c^2 that minimizes the risk in (15) is attained at $\frac{1+\alpha}{2}r+1$, indeed, if we put $u = c^2 > 1$ and let φ_s be a function of u such that:

$$\varphi_s(u) = \left(1 + (u)^{-\frac{d\alpha}{2}} - 2\left(\frac{u+c_*}{2(u)^{\frac{1-\alpha}{2}}}\right)^{-\frac{d}{2}}\right),$$

the risk in (15) denoting that $A_* = \frac{2(2\pi\sigma_y^2)^{-\frac{d\alpha}{2}}}{(1+\alpha)^{\frac{d}{2}+1}}$, becomes

$$R_{S_\alpha}(\theta, \hat{q}_{u,x}) = A_*\varphi_s(u).$$

It is easy then to check that $\varphi'(c_*^2) = 0$. We obtain (16) by inserting c_*^2 in (15).

For part (4), we distinguish two cases, namely, case 1: $0 \leq \alpha < 1$ and case 2: $\alpha = 1$:

case 1: The derivative function of φ_s satisfies

$$\frac{2\varphi_{S_\alpha}^{'}(u)}{dA_* u^{-\frac{d\alpha}{2}-1}} = 2^{1+\frac{d}{4}(1-\alpha)}(1+\alpha)u^{\frac{d\alpha}{4}+1}(u+c_*)^{1-\frac{d}{2}}(u - (\frac{1-\alpha}{1+\alpha})c_*^2) - \alpha,$$

it can be seen that for $u < (\frac{1-\alpha}{1+\alpha})c_*^2$, the risk in (15) becomes a decreasing function of u, whenever $0 \leq \alpha < 1$.

case 2: A stronger condition on the c^2 is available in Kubokawa et al (2015) (not only sufficient but also necessary), where $c_*^2 = 1 + r$, which concludes the proof. ■

3.2. Dominating plug-in type estimators of the form $\hat{\theta}(x) = ax$

In this section, we consider the performance of estimators of the type $\hat{q}_{c^2,\hat{\theta}} \sim \mathcal{N}_d(\hat{\theta}, c^2\sigma_y^2 I_d)$, and with more development for the affine linear case $\hat{\theta}(x) = ax$, with $0 < a \leqslant 1$. As in subsection 3.1, there exists an optimal choice of the expansion factor c^2, when $a = 1$ (i.e. for $c^2 = c_*^2 = \frac{1+\alpha}{2}r + 1$ under S-Hellinger distances), under the conditions of Theorem (1). Here, the objective is to assess whether such results hold for other choices of $\hat{\theta}(x)$, and more specifically: to determine a range of variance expansions of values c^2 that leads to improvement, and to determine whether there exists a universal dominance results for sufficiently large d (i.e. for all $c^2 > 1$). For any $\hat{\theta}(x) \in \mathbb{R}^d$, Lemma (2) implies that

$$R_{S_\alpha}(q,\hat{q}_{c^2,\hat{\theta}}) = A_* \left(1 \;+\; (u)^{-\frac{d\alpha}{2}} - 2^{\frac{d}{2}+1}\left(\frac{1+u}{(u)^{\frac{1-\alpha}{2}}}\right)^{-\frac{d}{2}} \times\; \mathbb{E}^X \exp\left(-\frac{||\hat{\theta}(x)-\theta||^2}{2\frac{2\sigma_y^2}{1+\alpha}(1+u)}\right)\right). \tag{17}$$

More precisely, we aim to study the case where $\hat{\theta}(X) = aX$, consequently (17) becomes by virtue of Lemma (3):

$$R_{S_\alpha}(q,\hat{q}_{c^2,\hat{\theta}}) = \varphi_a(u) = A_* \left(1 \quad + \quad (u)^{-\frac{d\alpha}{2}} - 2\left(\frac{u+u_0}{2(u)^{\frac{1-\alpha}{2}}}\right)^{-\frac{d}{2}} \exp\left(-\frac{T_a}{a^2 r + \frac{1+u}{a^2(1+\alpha)}}\right)\right), \tag{18}$$

where $u_0 = a^2(1+\alpha)r+1$ and $T_a = ||\theta||^2(a-1)^2/\sigma_y^2$. Thus, in the following proposition we established a sufficient condition of domination when $\hat{\theta}(X) = \; aX$:

Proposition 3. *All estimators of the form* $\hat{q}_{c^2,aX}$ *dominates* $\hat{q}_{1,aX}$, *whenever* $1 < c^2 < \frac{1-\alpha}{1+\alpha}u_0$.

proof. We derive the derivative of the risk function, denoted $\varphi_a(u)$ in (18), verifying this expression

$$\frac{2u}{dA_*}\varphi_a'(u) \;=\; \left(\frac{u+u_0}{2(u)^{\frac{1+\alpha}{2}}}\right)^{-\frac{d}{2}}\left(\left(\frac{1+\alpha}{u+u_0}\right)\left(u-\frac{1-\alpha}{1+\alpha}u_0\right) - \frac{4T_a u}{d(u+a^2(u_0-1)+1)}\right) \times\; \exp\left(-\frac{T_a}{2(a^2 r + \frac{u+1}{a^2(1+\alpha)})}\right) - \alpha. \tag{19}$$

As in (19), a sufficient condition of domination would be
$1 < u < \frac{1-\alpha}{1+\alpha}u_0$, which concludes the proof. □

3.3. Dominating estimators over the MRE estimator

According to the established parallel between plug-in estimation under S-hellinger distances in Theorem (1), with the point estimation under reflected normal loss $a + bD_{\gamma_s}$, this latter being concave in $||\hat{\theta}-\theta||$, however the fact that it's strictly bowled-shaped in $||\hat{\theta}-\theta||$ brings into play all the valid results for loss functions alike. It's worthy to mention that the previous works tackled a wider class of losses, namely those of the form $f(||\hat{\theta}-\theta||^2)$, where f is a concave nondecreasing function (Brandwein *et al* (1981), Brandwein et al (1991) and Brandwein et al (1993)), besides, Kubokawa et al (2015), Kubokawa and Saleh proved the dominance of the bayesian estimate $\hat{\theta}_U(x)$ corresponding to uniform prior on (a, b), or (a, ∞) over the MRE estimator X under reflected normal loss, in the univariate case, as mentioned in Kubokawa et al (2015), which remains valid for plug-in estimation under S-Hellinger distances, in the normal case, for example:

Theorem 2 (Univariate normal case). *For $d = 1$ and $\theta \in (a, b)$ (resp. $\theta \in (a, \infty)$), the plug-in estimation under S-Hellinger distances of $q(y|\theta)$, based on X, the plug-in type estimator $\frac{1}{c\sigma_y}\phi\left(\frac{y-\hat{\theta}_U(x)}{c\sigma_y}\right)$ improves over $\frac{1}{c\sigma_y}\phi\left(\frac{y|x}{c\sigma_y}\right)$ under reflected normal loss $D_{\gamma_s}(\theta, \hat{\theta})$, where $\hat{\theta}_U(x)$ is the Bayes point estimator of θ associated with a uniform prior on $[a, b]$ (on $[a, \infty)$), with $\gamma_s = 4\sigma_y^2(1 + c^2)/(1 + \alpha)$.*

proof. Since $\hat{\theta}_U(x)$ improves upon the MRE estimator as shown in Marchand et al (2005), the result follows from part (1) of Theorem (1). □

By virtue of an intuitive lemma in Kubokawa et al (2015), together with the established duality in Theorem (1) between plug-in density estimators under S-hellinger distances, and the corresponding point estimator under reflected normal loss, we are led to the following theorem, in the multivariate case, which is net extension of the same result under L_2-norm in Kubokawa et al (2015).

Theorem 3 (d-variate normal case). *Considering the estimation of $q(y|\theta)$ based on X: Under S-Hellinger distances, the MRE estimator $\hat{q}_{mre}(y|x) \sim \mathcal{N}_d(x, \sigma_y^2 c_*^2 I_d)$ is inadmissible when $d \geq 3$, and dominated by $\hat{q}(y - \hat{\theta}(x) \sim \mathcal{N}_d(\hat{\theta}(x), \sigma_y^2 c_*^2 I_d)$, as long as $\hat{\theta}(W)$ dominates W, where $W \sim \mathcal{N}_d(\theta, \sigma_s^2 I_d)$ under the quadratic loss $||\hat{\theta} - \theta||^2$, such that $\sigma_s^2 = \sigma_x^2 \frac{2(1+c_0^2)}{2(1+c_0^2)+(1+\alpha)r}$, with $r = \sigma_x^2/\sigma_y^2$ and $c_0^2 = \frac{1+\alpha}{2}r + 1$.*

proof. The result is readily verified, by virtue of Lemma 3.3 in Kubokawa et al (2015) and part (1) of Theorem (1), by making the these substitutions: $c^2 = c_*^2 = \frac{1+\alpha}{2}r + 1$ and $\gamma_s = 2\sigma_y^2(1 + c^2)/(1 + \alpha)$ for part (1). □

The results previously shown emphasize the inadmissibility of the MRE estimator as a benchmark estimator for ($d \geq 3$), which brings into play all the established results on the Stein estimation under quadratic loss, more precisely, we can provide several explicit dominating plug-in type density estimators of the form $\hat{q}(y - \hat{\theta}(x) \sim \mathcal{N}_d(\hat{\theta}(x), \sigma_y^2 c_*^2 I_d)$, where $c_*^2 = (\frac{1+\alpha}{2}r + 1)$ under S-Hellinger distances, as the following:

Example 5 (Bayes estimators under Superharmonic prior). Stein in 1981 showed that when the prior π is superharmonic, the Bayes estimator $\hat{\theta}_\pi(W)$ improves upon W, with $W \sim \mathcal{N}_d(\theta, \sigma_w^2 I_d)$, when $d \geq 3$, under quadratic loss, as a consequence, the corresponding plug-in density
$\hat{q}(y|\ \hat{\theta}(x))\ \sim \mathcal{N}_d\ (\hat{\theta}_\pi(x), \sigma_*^2 I_d)$ dominates $\hat{q}_{mre}(y|x)$ under S-Hellinger distances. Moreover, Stein's result also brings about the dominance of $\hat{\theta}_\pi(x)$ over X. Furthermore, we can widen more the family of dominating estimators, by taking the square root of the marginal density of W under π, denoted by $\sqrt{m_\pi(W)}$, to be superharmonic as well Fourdrinier et al (1998).

Example 6 (Baranchick type estimators). Baranchick in Baranchick (1966), suggested a better class of dominating estimators (it includes the James-Stein estimator) over the MLE estimator (i.e. $W \sim \mathcal{N}_d(\theta, \sigma_*^2 I_d)$), where $\sigma_*^2 = \sigma_s^2$ under S-Hellinger distances, namely, Baranchick type estimators, this class of estimators is of the form

$$\hat{\theta}_{\beta,r(.)}(W) = \left(1 - \frac{r(W'W)}{W'W}\beta\right) W, \tag{20}$$

where $r(.)$ is an increasing function, such that $\beta \in]0, 2(d-2)\sigma_*^2]$ and $0 < r(.) \leq 1$. It is easily seen that such estimators fully satisfy the conditions of dominance in theorem (3), thus, the corresponding plug-in type estimators, i.e. $\hat{q}_{brc} \sim \mathcal{N}_d(\hat{\theta}_{\beta,r(.)}(W), \sigma_y^2 c_*^2 I_d)$, dominate the MRE estimator $\hat{q}_{mre}$ when $d \geq 3$ under S-Hellinger distances .

In order to get a closer view, we singled out of the Baranchick class of estimators, a modified version of the most famous and historical member of this class, which is considered to be the precursor to the Baranchick class of estimator, namely the positive part of James-Stein estimator, i.e.

$$\hat{\theta}_{pjs}(W) = h(W)W,$$

where

$$h(W) = 0 \vee \left(1 - \frac{(d-2)\sigma_y^2 c_*^2}{W'W}\right) \ and\ d \geq 3,$$

to obtain numerical simulations based on Theorem (3), for the corresponding plug-in estimator $\hat{q}_{pjs} \sim \mathcal{N}_d(\hat{\theta}_{pjs}(W), \sigma_y^2 c_*^2 I_d)$ versus the MRE estimator $\mathcal{N}_d(X, \sigma_y^2 c_*^2 I_d)$, setting fixed values for the dimension d, the ratio of variances r, the scale-expanding factor c and the tuning parameter α, mainly for twice-squared Hellinger distance (for $\alpha = 0$), and L_2-norm (for $\alpha = 1$), tuning over $\lambda = ||\theta||$ and the ratio $r = \sigma_x^2/\sigma_y^2$, and for sake of simplification we take $\sigma_x^2 = 1$, then we assess the ratio $Ratio = R_{S_\alpha}(\theta, \hat{q}_{pjs})/R_{S_\alpha}(\theta, \hat{q}_{mre})$.

We present the 3D figures corresponding to three members of the S-Hellinger family, namely: $\alpha \in \{0, 0.5, 1\}$, assessing the latter ratio ($Ratio$) for both (λ, r) and (λ, c), where $d \in \{3, 5, 10\}$, $\lambda \in [0, 10]$ and $r, c \in [1, 10]$.

Comments on the figures:

Let us make the following comments.

(a) The fact that the barrier of dominance $\frac{1-\alpha}{1+\alpha}c_*^2$ (avoiding $\alpha = 1$, to go on details Kubokawa et al (2015)) of the expanding factor c, is an nondecreasing function

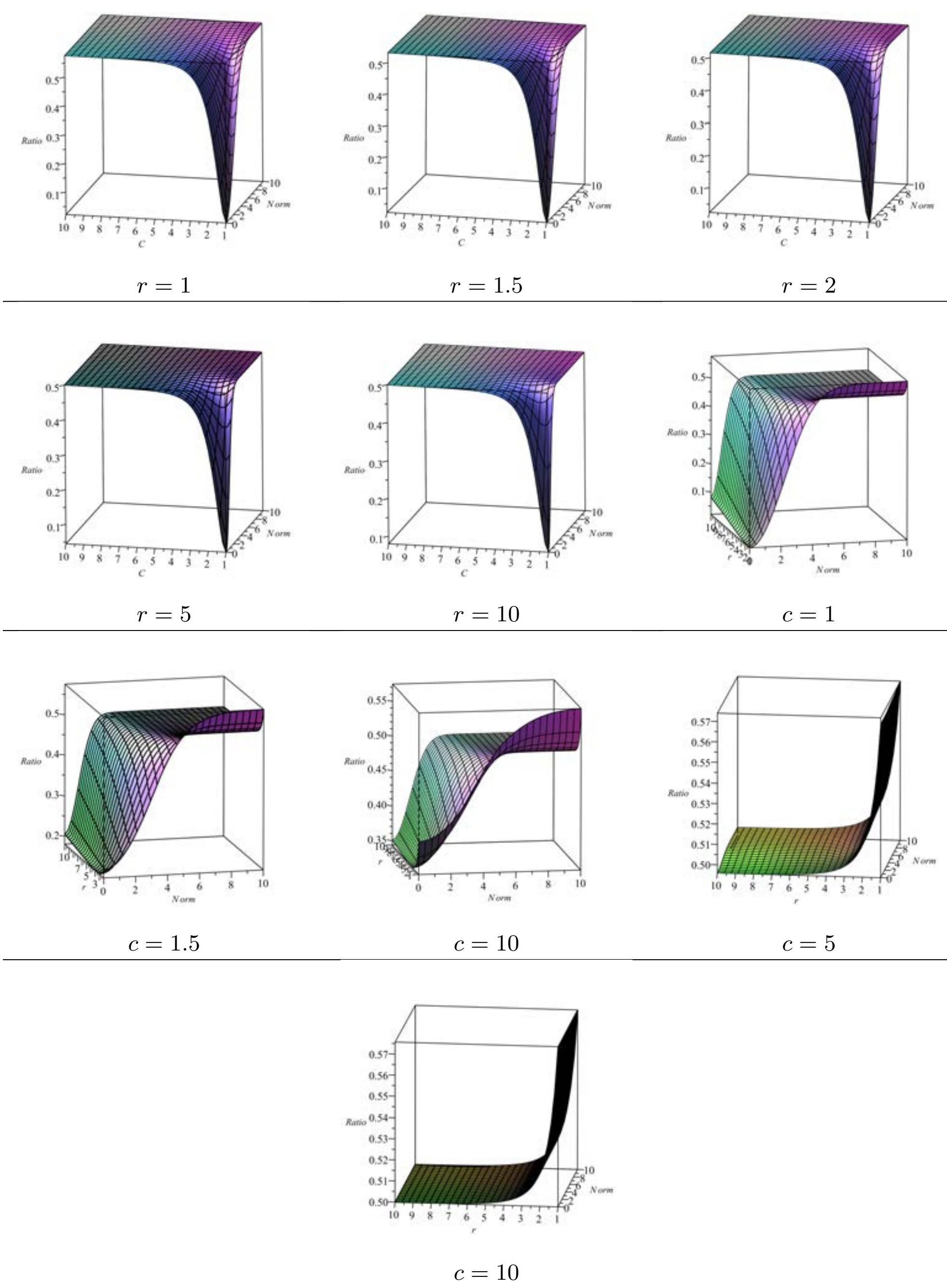
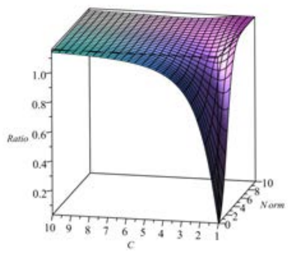
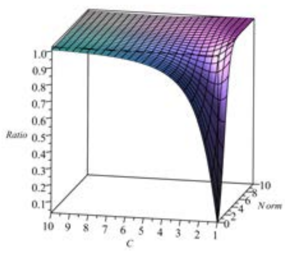


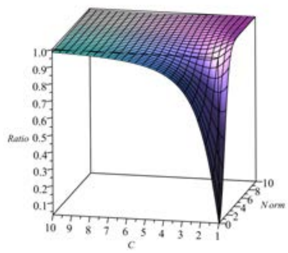
Table 1. Ratio: $d = 10$, $\lambda \in [0, 10]$, $r \in [1, 10]$ and $c \in [1, 10]$, for $\alpha = 0$ (twice-squared Hellinger distance)



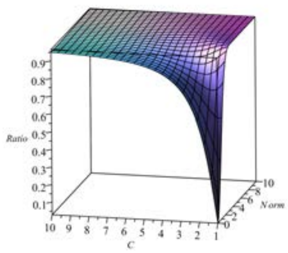
$r = 5$



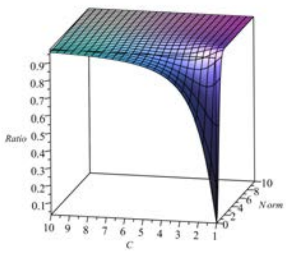
$r = 1.5$



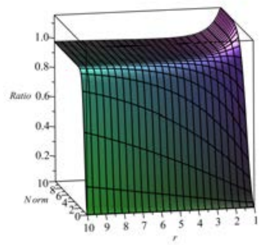
$r = 2$



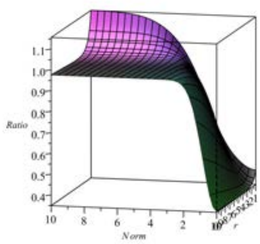
$r = 5$



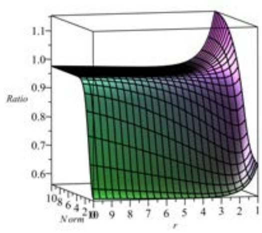
$r = 10$



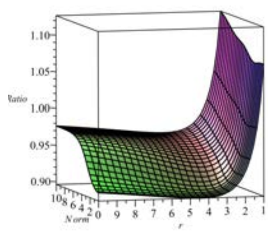
$c = 1$



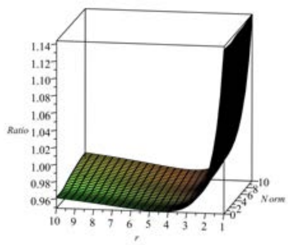
$c = 1.5$



$c = 2$

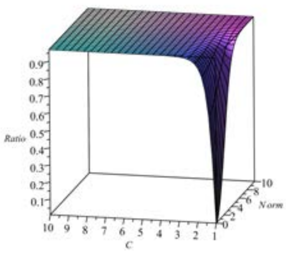


$c = 5$

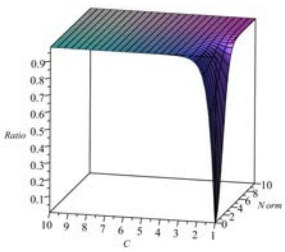


$c = 10$

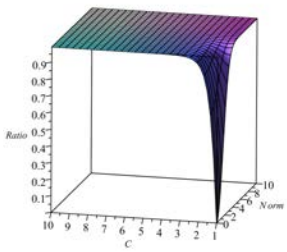
Table 2. Ratio: $d = 3$, $\lambda \in [0, 10]$, $r \in [1, 10]$ and $c \in [1, 10]$, for $\alpha = 0.5$ (the mid-range)



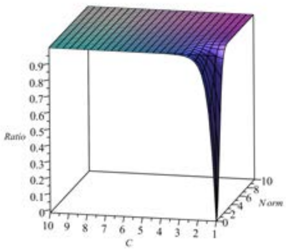
$r = 5$



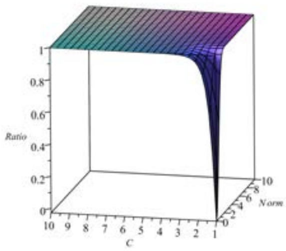
$r = 1.5$



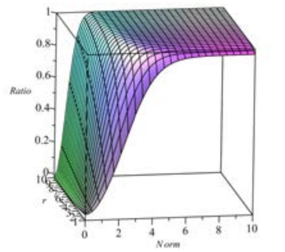
$r = 2$



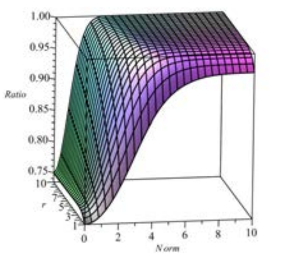
$r = 5$



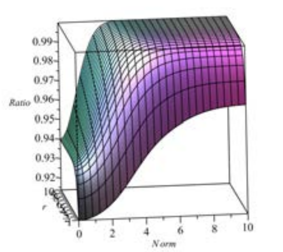
$r = 10$



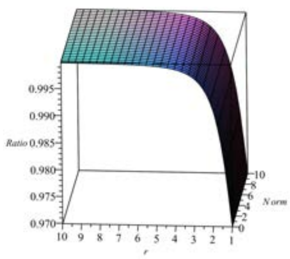
$c = 1$



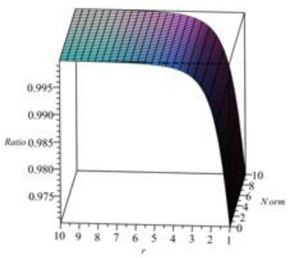
$c = 1.5$



$c = 2$



$c = 5$



$c = 10$

Table 3. Ratio: $d = 10$, $\lambda \in [0, 10]$, $r \in [1, 10]$ and $c \in [1, 10]$, for $\alpha = 0.5$ (the mid-range)

of the ratio r, explains the growth of the area of dominance of the PJS estimator over the benchmark MRE estimator, is in concordance with the growth of r, further, we emphasize that this phenomenon is omnipresent at all dimensions.

(b) The second general phenomenon is that our two rivals (PJS vs MRE) become twin estimators (superposition), once the norm λ and the expanding factor c take big enough values, the Ratio converges to 1 for low values of r (≤ 2), for greater values of r (≥ 2), they become proportional with a remarkably constant improvement of PJS estimate over the benchmark estimate (limit < 1), accordingly, the greater the dimension gets, the sooner the Ratio converges to 1.

(c) Thereupon, coming to the effect of the tuning parameter $\alpha \in [0, 1[$, whenever α tunes from 0 to 1, the area of dominance shrinks more and more (where PJS beats MRE), in both cases: either when ($r \in \{1, 5, 10\}$, $c \in [1, 10]$) or ($c \in \{1, 5, 10\}$, $r \in [1, 10]$). For sake of avoiding congestion, we singled out only 3 members of the S-Hellinger family, e.g. ($\alpha = 0$)corresponding to twice-squared Hellinger distance, ($\alpha = 0.5$)corresponding to the mid-range distance and ($\alpha = 1$) corresponding to L_2-norm.

Remark 3. Another class of dominating estimators which is wider than the Baranchick class, in the sense that the latter class shrinks the sample mean X towards 0, instead one can shrink X towards any arbitrary θ. The authors in Ghosh et al (2008), stated the explicit expression of the Bayes estimator $\hat{\theta}_\pi(x)$ of θ under S-Hellinger distances and a normal prior $\pi(\theta) \sim \mathcal{N}_d(\mu, AI_p)$ with ($A > 0$), such that $\hat{\theta}_\pi(x) = (1 - B)x + B\mu$, where $B = \sigma_x^2(A + \sigma_x^2)^{-1}$. Therefore a more general Bayes estimator of θ will be given by

$$\hat{\theta}_*(x) = \left(1 - \frac{r(S)}{S}\right) x + \frac{r(S)}{S} \mu,$$

where $S = ||x - \mu||^2/\sigma_x^2$, the authors established also the dominance of such estimators over X under S-Hellinger distances.

Theorem 4 (Hartigan type result). *Let $W \sim \mathcal{N}_d(\theta, \sigma_w^2 I_d)$ with $\sigma_w^2 = \sigma_s^2\sigma_x^2/\sigma_s^2 + \sigma_x^2$, and let $\hat{\theta}_\pi(W)$ be the Bayes estimators of θ associated to the prior π and the quadratic loss. For estimating the density of $Y \sim \mathcal{N}_d(\theta, \sigma_y^2 I_d)$ based on $X \sim \mathcal{N}_d(\theta, \sigma_x^2 I_d)$ under S-Hellinger distances, whenever θ is constrained to any convex subset C of $\mathbb{R}^d$ with non-empty interior.*

(a) The estimator $\hat{q}(.; X) \sim \mathcal{N}_d(\hat{\theta}_{\pi_U}(W), \sigma_y^2 c_*^2 I_d)$ dominates $\hat{q}_{mre}(.; X) \sim \mathcal{N}_d(X, \sigma_y^2 c_*^2 I_d)$, with π_U being the uniform prior on C.

(b) Univariate case ($C = [a, b]$), the dominance of $\hat{q}_{mre}(.; X)$ is attained at any $\hat{q}(.; X) \sim \mathcal{N}_d(\hat{\theta}_\pi(W), \sigma_y^2 c_*^2 I_d)$, as long as the prior density π is absolutely continuous and symmetric around the mid-range of C ,i.e. $\frac{a+b}{2}$.

Proof. This is a direct consequence of Kubokawa et al (2015), for $c_*^2 = \frac{1+\alpha}{2}r + 1$ under S-Hellinger distances and Part $(1.)$ of Theorem 1, combined with point estimation results of Hartigan (2004) for Part (a), and Kubokawa (2005), or Marchand et al (2011) for part (b).

4. Conclusion and perspectives

By and large, the results of this paper established essential findings for assessing the efficiency of predictive density estimators of multivariate observables for S-Hellinger distances as a set of loss functions (for every $\alpha \in [0, 1]$). Thus, by widening the scope of investigation from L_2-norm (integrated squared error loss, for $\alpha = 1$), to a broader perspective, such as S-Hellinger distances, a family of symmetric divergences for any α in $[0, 1]$.

The main topics, revolved around the inefficiency of MRE predictors in high enough dimensions and about the inefficiency of plug-in estimators by either improving on the plug-in for a dual point estimation loss or expanding the scale. Another key point, would be to upgrade these results to scale mixture of normals as a model distribution, we already made an attempt, but faced significant hardships. Last but not least, considering such models with unknown scale represents one of several challenging and interesting problems worthwhile pursuing.

Acknowledgement

We greatly show our gratitude to Professor Éric Marchand from Université de Sherbrooke for supporting financially and academically a scientific visit to his department for one of the authors, during the summer session of 2015, and also for providing insight and for sharing his pearls of wisdom in predictive density estimation with us during the course of this research, this research is mainly a fruit of this scientific encounter.

5. Appendix

5.1. Auxiliary lemma

Lemma 3 (Degenerate case). *For $\alpha_1, \alpha_2 > 0$, and if $q_1(y|\theta_1) \sim \mathcal{N}_d(\theta_1, \sigma_1^2 I_d)$ and $q_2(y|\theta_2) \sim \mathcal{N}_d(\theta_2, \sigma_2^2 I_d)$, we have*

$$\int_{\mathbb{R}^d} q_1^{\alpha_1}(y|\theta_1) q_2^{\alpha_2}(y|\theta_2) dy = \left((2\pi)^{(1-\alpha_1-\alpha_2)} \frac{(\sigma_1^2)^{1-\alpha_1} (\sigma_2^2)^{1-\alpha_2}}{\alpha_1 \sigma_2^2 + \alpha_2 \sigma_1^2} \right)^{\frac{d}{2}} \times \exp\left(-\frac{||\theta_1 - \theta_2||^2}{2\sigma_{1,2}^2} \right) \tag{21}$$

with $\sigma_{1,2}^2 = \frac{\sigma_1^2}{\alpha} + \frac{\sigma_2^2}{1-\alpha}$, where ϕ denotes the probability density function of a standard normal random variable.

Proof. It suffices to make these substitutions $\Sigma_1 = \sigma_1^2 I_d$ and $\Sigma_2 = \sigma_2^2 I_d$ in Lemma 2.2 in Ghosh et al (2008), and (21) follows immediately, we clarify the calculus in our degenerate case:

Given that $w_{1,2} = \frac{\alpha_2 \sigma_1^2 \theta_1 + \alpha_1 \sigma_2^2 \theta_2}{\alpha_2 \sigma_1^2 + \alpha_1 \sigma_2^2}$ and $\sigma_{1,2}^2 = \frac{\alpha_2 \sigma_1^2 \alpha_1 \sigma_2^2}{\alpha_2 \sigma_1^2 + \alpha_1 \sigma_2^2}$, we can check with ease the commonly used equality

$$\frac{||y - \theta_1||^2}{\frac{\sigma_1^2}{\alpha_1}} + \frac{||y - \theta_2||^2}{\frac{\sigma_2^2}{\alpha_2}} = \frac{||y - w_{1,2}||^2}{\sigma_{1,2}^2} + \frac{||\theta_1 - \theta_2||^2}{\frac{\sigma_1^2}{\alpha_1} + \frac{\sigma_2^2}{\alpha_2}}, \tag{22}$$

and by virtue of (22), we have the passage from second line to third line in the calculation right below:

$$
\begin{aligned}
\int_{\mathbb{R}^d} q_1^{\alpha_1}(y|\theta_1)q_2^{\alpha_2}(y|\theta_2)dy &= \int_{\mathbb{R}^d}(2\pi\sigma_1^2)^{-\frac{d}{2}\alpha_1}\exp\left(-\frac{||y-\theta_1||^2}{2\frac{\sigma_1^2}{\alpha_1}}\right)\\
&\times\ (2\pi\sigma_2^2)^{-\frac{d}{2}\alpha_2}\exp\left(-\frac{||y-\theta_2||^2}{2\frac{\sigma_2^2}{\alpha_2}}\right)dy\\
&= \int_{\mathbb{R}^d}\left((2\pi)^{-(\alpha_1+\alpha_2)}(\sigma_1^2)^{-\alpha_1}(\sigma_2^2)^{-\alpha_2}\right)^{\frac{d}{2}}\\
&\times\ \exp\left(-\frac{1}{2}\left(\frac{||y-\theta_1||^2}{\frac{\sigma_1^2}{\alpha_1}}+\frac{||y-\theta_2||^2}{\frac{\sigma_2^2}{\alpha_2}}\right)\right)dy\\
&= \int_{\mathbb{R}^d}\left((2\pi)^{-(\alpha_1+\alpha_2)}(\sigma_1^2)^{-\alpha_1}(\sigma_2^2)^{-\alpha_2}\right)^{\frac{d}{2}}\\
&\times\ \exp\left(-\frac{1}{2}\left(\frac{||y-w_{1,2}||^2}{\sigma_{1,2}^2}+\frac{||\theta_1-\theta_2||^2}{\frac{\sigma_1^2}{\alpha_1}+\frac{\sigma_2^2}{\alpha_2}}\right)\right)dy\\
&= \left((2\pi)^{1-(\alpha_1+\alpha_2)}\frac{(\sigma_1^2)^{1-\alpha_1}(\sigma_2^2)^{1-\alpha_2}}{\alpha_2\sigma_1^2+\alpha_1\sigma_2^2}\right)^{\frac{d}{2}}\exp\left(-\frac{||\theta_1-\theta_2||^2}{2(\frac{\sigma_1^2}{\alpha_1}+\frac{\sigma_2^2}{\alpha_2})}\right)\\
&\times\ \underbrace{\int_{\mathbb{R}^d}(\sigma_{1,2}^2)^{-\frac{d}{2}}\phi\left(\frac{y-w_{1,2}}{\sigma_{1,2}}\right)dy}_{=1}\\
&= \left((2\pi)^{1-(\alpha_1+\alpha_2)}\frac{(\sigma_1^2)^{1-\alpha_1}(\sigma_2^2)^{1-\alpha_2}}{\alpha_2\sigma_1^2+\alpha_1\sigma_2^2}\right)^{\frac{d}{2}}\exp\left(-\frac{||\theta_1-\theta_2||^2}{2(\frac{\sigma_1^2}{\alpha_1}+\frac{\sigma_2^2}{\alpha_2})}\right),
\end{aligned}
$$

which concludes the proof.

5.2. Proof of proposition (2)

For part (1), by making this substitution $u = \hat{\theta}(x) - \theta$, our result follows easily:

$$\begin{aligned}
\hat{q}_{(\pi)}(y|x) &= \frac{1}{m_\pi(x)} k^{\frac{2}{1+\alpha}}_{\frac{1+\alpha}{2}}(y,x) \\
&= \frac{1}{m_\pi(x)} \left(\int_{\mathbb{R}^d} q^{\frac{1+\alpha}{2}}(y|\theta) g(\hat{\theta}(x)-\theta) d\theta \right)^{\frac{2}{1-\alpha}} \\
&= \frac{1}{m_\pi(x)} \left(\int_{\mathbb{R}^d} q^{\frac{1+\alpha}{2}}(y-u-\hat{\theta}(x)) g(u) du \right)^{\frac{2}{1-\alpha}} \\
&= \frac{1}{m_\pi(x)} (q^{\frac{1+\alpha}{2}} * g)^{\frac{2}{1+\alpha}}(y-\hat{\theta}(x)).
\end{aligned}$$

where $m_\pi(x) = \int_{\mathbb{R}^d} k^{\frac{2}{1+\alpha}}_{\frac{1+\alpha}{2}}(y,x)dy$.

For part (2), when $\pi(\theta) = 1$, we have $\pi(\theta|x) = g(\theta - \hat{\theta}(x)) = p(x|\theta)$, moreover, the expression in (11) is readily verified by making these substitutions $\hat{\theta}(x) = x$ and $g = p$ in (4), thus we get (5).

We give the proof in both the normal and the general case, proceeding via a direct approach in the normal case (i.e. finding a least favourable sequence of priors to show that $\hat{q}_{mre}(y|x)$ is minimax under the S-Hellinger distances), we suggested a somewhat different proof from the one given in Ghosh et al (2008) in the normal case, this lying basically on a technique introduced by Girshick et al (1951), which is identical to the one used in Kubokawa et al (2015).

Normal case For $\pi(\theta) = 1$, we have $\pi(\theta|x) = p(x|\theta)$, then we consider the sequence of priors $\pi_m \sim \mathcal{N}_d(0, m^2 I_d)$, and its corresponding posterior density

$$\begin{aligned}
\pi_m(\theta|x) &= \left(\frac{m^2+\sigma_x^2}{m^2\sigma_x^2}\right)^{\frac{d}{2}} \phi\left(\frac{x-\theta}{\sigma_x}\right) \phi\left(\frac{\theta}{m}\right) / \phi\left(\frac{x}{\sqrt{m^2+\sigma_x^2}}\right), \\
&= \left(\frac{m^2\sigma_x^2}{m^2+\sigma_x^2}\right)^{-\frac{d}{2}} \phi\left(\frac{\frac{m^2}{m^2+\sigma_x^2}x - \theta}{\sqrt{\frac{m^2\sigma_x^2}{m^2+\sigma_x^2}}}\right),
\end{aligned}$$

we notice that, $\lim_{m\to\infty}\pi_m(\theta|x) = p(x|\theta)$, besides, since

$$\begin{aligned}
k_{\pi_m}(y,x) = (2\pi\sigma_y^2)^{-\frac{d}{4}(1+\alpha)}(4\pi\sigma_y^2)^{\frac{d}{2}} &\times \left(2\sigma_y^2 + 1 + \alpha\frac{m^2\sigma_x^2}{m^2+\sigma_x^2}\right)^{-\frac{d}{2}} \\
&\times \phi\left(\frac{y - \frac{m^2}{m^2+\sigma_x^2}x}{\sqrt{\sigma_y^2 + \frac{1+\alpha}{2}\frac{m^2\sigma_x^2}{m^2+\sigma_x^2}}}\right),
\end{aligned}$$

then

$$k_{\pi_m}^{\frac{2}{1+\alpha}}(y,x) \propto (\sigma_y^2 + \frac{1+\alpha}{2}\frac{m^2\sigma_x^2}{m^2+\sigma_x^2})^{-\frac{d}{2}}\phi\left(\frac{y - \frac{m^2}{m^2+\sigma_x^2}x}{\sqrt{\sigma_y^2 + \frac{1+\alpha}{2}\frac{m^2\sigma_x^2}{m^2+\sigma_x^2}}}\right),$$

and we have as a consequent

$$\hat{q}_{\pi_m}(y|x) \sim \mathcal{N}_d\left(\frac{m^2}{m^2+\sigma_x^2}x, \sigma_y^2 + \frac{1+\alpha}{2}\frac{m^2\sigma_x^2}{m^2+\sigma_x^2}\right),$$

according to example (2), we notice that $\lim_{m\to\infty}\hat{q}_{\pi_m}(y|x) = \hat{q}_{mre}(y|x)$; thus, the corresponding posterior risk will be

$$\begin{aligned}
\rho_m(q,\hat{q}_{\pi_m}) &= \int D_\alpha(q,\hat{q}_{\pi_m})\pi_m(\theta|x)d\theta \\
&= \frac{2(2\pi\sigma_y^2)^{-\frac{d}{2}}}{(1+\alpha)^{\frac{d}{2}+1}}\left(1 + (\frac{1+\alpha}{2}\frac{m^2 r}{m^2+\sigma_x^2}+1)^{-\frac{d\alpha}{2}}\right. \\
&- \left. 2(\frac{1+\alpha}{2}\frac{m^2 r}{m^2+\sigma_x^2}+1)^{-\frac{d}{4}(1+\alpha)}\right),
\end{aligned}$$

$\rho_m(q,\hat{q}_{\pi_m}) = r_{S_\alpha}(q,\hat{q}_{\pi_m}) \to R_{S_\alpha}(q,\hat{q}_{mre}) = r_{S_\alpha}(\hat{q}_{mre})$, since $R_{S_\alpha}(q,\hat{q}_{\pi_m})$ and $R_{S_\alpha}(q,\hat{q}_{mre})$ are constant, which proves the minimaxity of $\hat{q}_{mre}(y|x)$, and that the sequence π_m is least favourable.

General proof. We firstly state the expression of S-Hellinger distances corresponding to the MRE predictor given in (5), by virtue of (1):

$$D_{S_\alpha}(q,\hat{q}_{mre}) = \frac{2}{1+\alpha}\int_{\mathbb{R}^d}\left(q^{\frac{1+\alpha}{2}}(y|\theta) - \hat{q}_{mre}^{\frac{1+\alpha}{2}}(y|x)\right)^2 dy.$$

Then the corresponding frequentist risk given by

$$R_{S_\alpha}(q,\hat{q}_{mre}) = \frac{2}{1+\alpha}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\left(q^{\frac{1+\alpha}{2}}(y|\theta) - \hat{q}_{mre}^{\frac{1+\alpha}{2}}(y|x)\right)^2 dy\right)p(x|\theta)dx.$$

Consequently the corresponding bayesian risk states as, where $\pi(\theta) = 1$,

$$\begin{aligned}
r_{S_\alpha}(q,\hat{q}_{mre}) &= \frac{2}{1+\alpha}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\left(q^{\frac{1+\alpha}{2}}(y|\theta) - \hat{q}_{mre}^{\frac{1+\alpha}{2}}(y|x)\right)^2 dy\right) \\
&\times\ p(x|\theta)dx\pi(\theta)d\theta \\
&= \frac{2}{1+\alpha}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\left(q^{\frac{1+\alpha}{2}}(y|\theta) - \hat{q}_{mre}^{\frac{1+\alpha}{2}}(y|x)\right)^2 dy\right) \\
&\times\ p(x|\theta)dxd\theta. \qquad (23)
\end{aligned}$$

We consider now the following sequence of sets

$$S_k = \{\theta / |\theta_i| < k/2, i = 1, .., d, k \in \mathbb{N}^*\}.$$

We notice that $S_k \to \mathbb{R}^d$ when $k \to \infty$, and let π_k be a sequence of priors given by

$$\pi_k(\theta) = k^{-d}\delta_\theta(S_k) \tag{24}$$

with $\delta_\theta(.)$ being the dirac function. The corresponding Bayesian estimators are of the form

$$\hat{q}_{\pi_k}(y|x) = \frac{k_{\pi_k}^{\frac{2}{1+\alpha}}(y,x)}{\int_{\mathbb{R}^d} k_{\pi_k}^{\frac{2}{1+\alpha}}(y,x)dy} = \frac{k_{S_k}^{\frac{2}{1+\alpha}}(y,x)}{m_{S_k}(x)}, , \tag{25}$$

where

$$k_{\pi_k}(y,x) = \frac{k^{-d}\int_{S_k} q^{\frac{1+\alpha}{2}}(y|a)p(x|a)da}{k^{-d}m_{\pi_k}(x)} = \frac{k_{S_k}(y,x)}{m_{\pi_k}(x)},$$

with $m_{\pi_k}(x) = \int_{S_k} p(x-\theta)d\theta$, $k_{S_k}(y,x) = \int_{S_k} q^{\frac{1+\alpha}{2}}(y|\theta)p(x-\theta)d\theta$, and $m_{S_k}(x) = \int_{\mathbb{R}^d} k_{S_k}^{\frac{2}{1+\alpha}}(y,x)dy$, provided that $k_{S_k}(y,x)$ and $m_{\pi_k}(x)$ are finite on S_k.
According to (1) the S-Hellinger distances associated to $\hat{q}_{\pi_k}(y|x)$ is:

$$D_{S_\alpha}(q, \hat{q}_{\pi_k}) = \frac{2}{1+\alpha}\int_{\mathbb{R}^d}\left(q^{\frac{1+\alpha}{2}}(y|\theta) - \hat{q}_{\pi_k}^{\frac{1+\alpha}{2}}(y|x)\right)^2 dy,$$

besides, given that

$$\pi_k(\theta|x) = \frac{k^{-d}p(x|\theta)\delta_{S_k}(\theta)}{k^{-d}m_{\pi_k}(x)} = \frac{p(x|\theta)\delta_{S_k}(\theta)}{m_{\pi_k}(x)},$$

the corresponding posterior risk is

$$\begin{aligned}
\rho_{S_\alpha}(q, \hat{q}_{(\pi_k)}) &= \int_{\mathbb{R}^d} D_{S_\alpha}(q, \hat{q}_{\pi_k})\pi_k(\theta|x)d\theta \\
&= \int_{S_k}\left(\frac{2}{1+\alpha}\int_{\mathbb{R}^d}\left(q^{\frac{1+\alpha}{2}}(y|\theta) - \hat{q}_{\pi_k}^{\frac{1+\alpha}{2}}(y|x)\right)^2 dy\right)\frac{p(x|\theta)}{m_{\pi_k}(x)}d\theta.
\end{aligned}$$

Further, after the following substitutions: $z = x - \theta$, $h = a - \theta$, $\nu = \frac{\theta}{k}$, the corresponding Bayes risk will be

$$\begin{aligned}
r_{S_\alpha}(q,\hat{q}_{\pi_k}) &= \int_{\mathbb{R}^d}\rho_{S_\alpha}(\hat{q}_{\pi_k})k^{-d}m_{\pi_k}(x)dx \\
&= \frac{2}{1+\alpha}\int_{S_k}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}k^{-d}\left(q^{\frac{1+\alpha}{2}}(y|\theta)-\hat{q}_{\pi_k}^{\frac{1+\alpha}{2}}(y|x)\right)^2dy\right) \\
&\times \frac{p(x|\theta)}{m_{\pi_k}(x)}(1+k^{-d})m_{\pi_k}(x)dxd\theta, \\
&= \frac{2}{1+\alpha}\int_{S_\nu}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}(1+k^{-d})\left(q^{\frac{1+\alpha}{2}}(y|\theta)-\hat{q}_{\pi_{k_*}}^{\frac{1+\alpha}{2}}(y|x)\right)^2dy\right) \\
&\times p(z)dzkd\nu,
\end{aligned}$$

with

$$\hat{q}_{\pi_{k_*}}(y|x)=\frac{k_{S_h}^{\frac{1+\alpha}{2}}(y,x)}{m_{S_h}(x)},$$

where,

$$S_\nu=\{\nu/|\nu_i|<1/3\},$$

$$S_h=\{h/h+k\nu\in S_k\},$$

$$k_{S_h}(y,x)=\int_{S_h}q^{\frac{1+\alpha}{2}}(y|a)p(x|a)da$$

and

$$m_{S_h}(x)=\int_{\mathbb{R}^d}k_{S_h}^{\frac{1+\alpha}{2}}(y,x)dy.$$

Provided that $r_{S_\alpha}(\hat{q}_{\pi_k})\leq r_{S_\alpha}(\hat{q}_{mre})$, all we need is to prove that

$$\liminf_{k\to\infty}r_{S_\alpha}(q,\hat{q}_{\pi_k})\geq r_{S_\alpha}(\hat{q}_{mre}).$$

For $\epsilon>0$, we have the inequality, given that $S_*=\{\nu/|\nu_i|<(1-\epsilon)/2\}$:

$$\begin{aligned}
r_{S_\alpha}(q,\hat{q}_{\pi_k}) &= \int_{S_\nu}\int_{\mathbb{R}^d}\left(\frac{2}{1+\alpha}\int_{\mathbb{R}^d}k^{1-d}\left(q^{\frac{1+\alpha}{2}}(y|\theta)-\hat{q}_{\pi_{k_*}}^{\frac{1+\alpha}{2}}(y|x)\right)^2dy\right)p(z)dzd\nu \\
&\geq \int_{S_*}\int_{\mathbb{R}^d}\left(\frac{2}{1+\alpha}\int_{\mathbb{R}^d}k^{1-d}\left(q^{\frac{1+\alpha}{2}}(y|\theta)-\hat{q}_{\pi_{k_*}}^{\frac{1+\alpha}{2}}(y|x)\right)^2dy\right)p(z)dzd\nu.
\end{aligned}$$

It is easy to check for $|\nu_i|<(1-\epsilon)/3$, that $\{|h|<k\epsilon/2\}\subset S_h$, which insures that $\hat{q}_{\pi_k}(y|x)\to\hat{q}_{mre}(y|x)$ when $k\to\infty$ by construction. Now by virtue of Fatou's lemma and (23), we obtain

$$\begin{aligned}
\liminf_{k\to\infty} r_{S_\alpha}(\hat{q}_{\pi_k}) &\geq \liminf_{k\to\infty} \frac{2}{1+\alpha}\int_{S_*}\int_{\mathbb{R}^d} \\
&\times \left(\int_{\mathbb{R}^d} k^{1-d}\left(q^{\frac{1+\alpha}{2}}(y|\theta) - \hat{q}_{\pi_{k_*}}^{\frac{1+\alpha}{2}}(y|x)\right)^2 dy\right) p(z)dzd\nu \\
&\geq \frac{2}{1+\alpha}\int_{S_*}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} \liminf_{k\to\infty} k^{1-d}\right. \\
&\times \left.\left(q^{\frac{1+\alpha}{2}}(y|\theta) - \hat{q}_{\pi_{k_*}}^{\frac{1+\alpha}{2}}(y|x)\right)^2 dy\right) p(z)dzd\nu \\
&= (1-\epsilon)^d r_{S_\alpha}(q, \hat{q}_{mre}),
\end{aligned}$$

thus, $\liminf_{k\to\infty} r_{S_\alpha}(\hat{q}_{\pi_k}) \geq r_{S_\alpha}(q, \hat{q}_{mre})$, hence the minimaxity of (11), which concludes the proof.

References

Aitchison, J. (1975) Goodness of prediction fit.*Biometrika*, Number 3, Oxford University Press, 62, 547-554.

Baranchick, A.J. (1970) A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics.* 41, 642-645.

Brandwein, A.C and Strawderman, W.E. (1981) Minimax estimation of location parameters for spherically symmetric distributions with oncave loss. *Annals of Statistics*, 8, 279-284.

Brandwein, A.C and Strawderman,W.E. (1991) Generalization of James-Stein estimators under spherical symmetry. *Annals of tatistics.* 19, 639-650.

Brandwein, A.C and Ralescu S. and Strawderman, W. E. (1993) Shrinkage estimators of the location parameter for Certain Spherically Symmetric Distributions. *Ann. Inst. Statist. Math.* 45, 3, 551-56.

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

Corcuera, J.M. and Giummolè, F. (1999) A Generalized Bayes Rule for Prediction. *Scand. J. Stat.*, 2, 26, 65-279.

Csiszár, I. and Shields, P. (2004) Foundations and Trends in Communications and Information Theory. *Information Theory nd Statistics*. 4, 417-528.

Fourdrinier D., Strawdermann W.E. and Wells M. (1998) On the construction of Bayes minimax estimators. *Annals of tatistics*. 26, 660-671.

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

Girshick, M. A. and Savage, L. J. (1951) Bayes and minimax estimates for quadratic loss functions. *Proc. Second Berkeley Symp. Math. Stat. Prob.*, 1, 53-73.

Ghosh M., Mergel V. and Datta G.S. (2008) Estimation, prediction and the Stein phenomenon under divergence loss. *J. ultivar. Anal.*, 9, 99, 1941-1961.

Ghosh, A., Harris, I. R., Maji, A., Basu, A. and Pardo, L. (2013) A Generalized Divergence for Statistical Inference. textitBernoulli, 23(4A), 2746–2783.

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

Kubokawa T. (2005) Estimation of a mean of a normal distribution with a ounded coefficient of variation. *The indian Journal f Statistics.* 3, 67, 499-525.

Kubokawa T., Marchand É. and Strawderman W. E. (2015) On predictive density estimation for location families under integrated quared error loss. *J. Multivar. Anal.* Elsevier Inc.

Kubokawa T., Marchand É., Strawderman W.E. and Turcotte J-P. (2013) Minimaxity in predictive density estimation with parametric onstraints. *Journal of Multivariate Analysis.* 116, 382-397.

Liang F. and Barron A. (2004) Exact minimax strategies for predictive density estimation, data compression and model selection. *IEEE Trans. Inform. Theory.*

Marchand, É. and Amir P.N. (2011) Bayesian improvements of a MRE estimator of a bounded location parameter. *Electron. . Stat.* 5, 1495-1502.

Marchand, É. and Strawderman W.E. (2005) On improving on the minimum risk equivariant estimator of a locationparameter which s constrained to an interval or half-interval. *Ann. inst. Statist. Math.* 57, 129-143.

Spiring, F. A. (1993) The reflected normal loss function. *Scan. J. Stat.* 3, 21, 321-330.

Stein C. (1956) Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. *Proc. Third Berkeley Symp. on Math. Statist. and Prob Univ. of Calif. Press.* 1, 197-206.