Afrika Statistika

Afrika Statistika
Vol. 14 (3), 2019, pages 2099 - 2125.
DOI: http://dx.doi.org/10.16929/as/2019.2099.150

ISSN 2316-090X

On the Stein Effect Under Density Power Divergence Loss

Younes OMMANE [1,*] **and Idir OUASSOU**[1,2]

[1] Cadi Ayyad University, National School of Applied Sciences of Marrakesh, Kingdom of Morocco
[2] Université Polytéchnique Mohamed VI, Ben guerir, Kingdom of Morocco

Received on October 5, 2019. Accepted on October 28, 2019

Abstract. In this paper we consider a decision-theoretical study of predictive density estimators of multivariate observables measured by the frequentist risk corresponding to Density Power Divergences (DPD) as a set of loss functions (for every α in $[0, 1]$). The main themes, revolve about the inefficiency of MRE (Minimum Risk Equivariant) predictors in high enough dimensions and about the efficiency of some families of plug-in estimators to be determined in the paper. The admissibility of both the sample mean, which is the Bayes estimate of the population mean under the DPD loss and under the improper flat prior, and its analog in the predictive problem is established for one or two dimensions, as well as their respective minimaxity in any dimension. (to be continued in page 2100).

Key words: minimaxity; admissibility; inadmissibility; Stein phenomenon; density power divergence; predictive density estimation; dominance; Baranchick estimators.
AMS 2010 Mathematics Subject Classification Objects : Primary 62F10; 62C10; 62C10; secondary; 62C20; 62F15.

---

*Corresponding author: Y. OMMANE. younes.ommane@ced.uca.ac.ma
I. OUASSOU: i.ouassou@uca.ma

Full Abstract. In this paper we consider a decision-theoretical study of predictive density estimators of multivariate observables measured by the frequentist risk corresponding to Density Power Divergences (DPD) as a set of loss functions (for every α in $[0, 1]$). The main themes, revolve about the inefficiency of MRE (Minimum Risk Equivariant) predictors in high enough dimensions and about the efficiency of some families of plug-in estimators to be determined in the paper. The admissibility of both the sample mean, which is the Bayes estimate of the population mean under the DPD loss and under the improper flat prior, and its analog in the predictive problem is established for one or two dimensions, as well as their respective minimaxity in any dimension. A class of shrinkage estimators is given either when the variance (degenerate case) is known or when variance-covariance matrix is unknown (including the degenerate case), under both the point estimation problem and the predictive problem. Finally, we extend these findings, to shrinkage toward a regression surface applied on modified version of the James-Stein estimator i.e. Lindley's estimator.Finally, we extend these findings to the case of unknown variance covariance matrix.

Résumé. (Abstract in French) Dans cet article, nous considérons une étude de décision théorique d'estimateurs de densité prédictifs d'observables multivariés mesurés par le risque fréquentiste correspondant aux divergences de puissance de densité (DPD) en tant qu'ensemble de fonctions de perte (pour chaque α en $[0, 1]$). Les thèmes principaux tournent autour de l'inefficacité des estimateurs prédictifs ERM (équivariant à Risque Minimal) dans des dimensions suffisamment élevées et de l'efficacité de certaines familles d'estimateurs plug-in à déterminer dans le document. L'admissibilité de la moyenne de l'échantillon, qui correspond à l'estimation de Bayes de la moyenne de la population dans le cadre de la perte de DPD et la loi a-priori non-informative impropre, et son analogue dans le problème prédictif est établie pour une ou deux dimensions, ainsi que leur minimaxité respective toute dimension. Une classe d'estimateurs dominants est donnée soit lorsque la variance (cas dégénéré) est connue, soit lorsque la matrice de variance-covariance est inconnue (y compris le cas dégénéré), à la fois sous le problème de l'estimation ponctuelle et sous celui de la prédiction. Enfin, nous étendons ces résultats à l'amélioration vers une surface de régression appliquée à la version modifiée de l'estimateur de James-Stein, c'est-à-dire à l'estimateur de Lindley. Enfin, nous étendons ces résultats.

1. Introduction

Let $\{Q_\theta\}$ be a parametric family of models indexed by the unknown d-variate parameter $\theta \in \Theta$, possessing densities $\{q_\theta\}$ with respect to Lebesgue measure.

We will limit our study to the family of normal densities as a model, namely, let $q, p \in Q_\theta$ two known densities with respect to Lebesgue measure:

$$q(Y|\theta) = N_d(\theta, v_x I_d) \text{ and } p(X|\theta) = N_d(\theta, v_y I_d), \tag{1}$$

where θ is the common location parameter, v_x and v_y being respective known variances of the independently normally distributed random vectors X and Y.

The pivotal problem is to predict $q(y|\theta)$ by a predictive density $\hat{q}(y)$ based on the observation of X. Our main concern lies in the evaluation of the performance of such predictors, first via the asymmetric family of Density Power Divergences (DPD) as our loss function, introduced in Basu (1998). It's a density-based divergence indexed by a single parameter $\alpha > 0$.

We now define the DPD loss $D_P(q, \hat{q})$ between two densities, classically between the target (true) unknown density $q(Y|\theta)$, and its predictive density estimator $\hat{q}(y)$, which states as

$$D_P(q,\hat{q}) = \int \left(\hat{q}^{\alpha+1}(y) - \frac{\alpha+1}{\alpha}\hat{q}^{\alpha}(y)q(y) + \frac{1}{\alpha}q^{\alpha+1}(y) \right) dy, \; (\alpha > 0). \tag{2}$$

The choice of the asymmetric DPD loss among many other measures of the discrepancy between two candidate densities, is firstly due to the fact that, to the best of our knowledge no author employed the DPD family as a loss function in a decision-theoretic approach, secondly, since the DPD family offers many appealing features among the aforementioned ones. Indeed, the author in Basu (1998) shed the light on the main intrinsic qualities of the asymmetric DPD family of loss functions such as equivariance, robustness and asymptotic efficiency. As for the equivariance property, the DPD loss is equivariant under linear transformations. Nonetheless, equivariance does not hold for other transformations in general, except for the peculiar case of ($\alpha = 0$), which corresponds to the Kullback-Leibler loss Kullback and Leibler (1951) (the full efficient member of the DPD family loss). Besides, concerning efficiency, the author in Basu (1998) emphasized that the DPD loss is mostly efficient for small values of the parameter α, i.e. for $0 \leq \alpha \leq 1$. This latter governs the compromise between robustness and asymptotic efficiency of the DPD loss, relating smoothly the Kullback-Leibler loss ($\alpha \to 0$), being the full efficient member, to the L_2 distance ($\alpha = 1$) (or integrated squared error loss, which a symmetric member of the asymmetric DPD family). However, efficiency becomes remarkably a fiasco when α exceeds 1, especially when ($\alpha \geq 2$). Whence the motivation of restricting our study to $0 < \alpha < 1$, excluding the boundary cases, for ($\alpha \to 0$) see Aitchison (1975), Brown *et al* (2008), George *et al* (2006), George *et al* (2012), or the works of Komaki, or Maruyama and Strawderman, among many others. For ($\alpha = 1$) see Kubokawa *et al* (2015), or under S-Hellinger distances Ommane and Ouassou (2019), for ($\alpha = 1$) as well, however, the S-Hellinger distances are obviously a symmetric family (it's the L_2 distance loss of two densities to the power of $(\alpha+1)/2$ for each), in which the authors attempted to extend most of the results established under the integrated squared error loss to the S-Hellinger distance family, for ($0 \leq \alpha \leq 1$).

Remark 1. To better justify our choice of DPD as a discrepancy measure, we zoom out to take a bigger shot, and locate the DPD loss as a member of a much bigger family, namely, the S-Divergence family is a two-parameterized class of divergences developed in Ghosh *et al* (2017), defined by

$$S_{(\alpha,\lambda)}(\hat{q},q) = A^{-1}(\alpha,\lambda)\int q^{\alpha+1}dy - \frac{\alpha+1}{A(\alpha,\lambda)B(\alpha,\lambda)}\int q^{B(\alpha,\lambda)}\hat{q}^{A(\alpha,\lambda)}dy + B^{-1}(\alpha,\lambda)\int \hat{q}^{\alpha+1}dy, \quad (3)$$

where $\alpha \in [0,1]$, $\lambda \in \mathbb{R}$, $A(\alpha,\lambda) = 1+\lambda(1-\alpha)$ and $B(\alpha,\lambda) = \alpha - \lambda(1-\alpha)$. We note that

$$D_P(\hat{q},q) = S_{(\alpha,0)}(\hat{q},q).$$

This family encompasses many famous and interesting subclasses of divergences, such as, the Density Power Divergence (DPD) for $\lambda = 0$, that generalizes the Power Divergence family (PD), introduced in Cressie and Read (1984) (recovered for $\alpha = 0$). Besides, the S-Kullback-Leibler Divergence (SKL) when $A \to 0$, Ghosh *et al* (2017), and the S-Likelihood Divergence (SLD) when $B \to 0$, both for $\alpha \in [0,1]$. These two subclasses are dual to each other (adjoint in the sense of Jimenez and Shao in their paper Jimenez and Shao (2001)), where we can obtain one from another simply by permuting $\hat{q}$ and q, i.e. $SKD_S(\hat{q},q) = SLD_\alpha(q,\hat{q})$, in fact SKL and SLD are symmetrically opposite to each other within the PD family (recovered for $\alpha = 0$ in SKL or SLD), with the Hellinger distance (HD) (actually, the twice-squared Hellinger distance to be rigorous) being the central member of the S-Divergence family (3).

2. Preliminary results

We begin by giving respectively, the general expression of the DPD loss between two arbitrary normal densities $q_1(y|\theta_1)$ and $q_2(y|\theta_2)$, with v_1 and v_2 their respective variances. This expression will be of a ubiquitous use throughout this section.

Lemma 1. *Let Y and X satisfy the model (1), and let (θ_1,θ_2) be in $\mathbb{R}^d \times \mathbb{R}^d$ and v_1, $v_2 > 0$, then we have:*

$$\begin{aligned} D_P(q_1,\hat{q}_2) = \quad & ((2\pi)^\alpha(\alpha+1))^{-\frac{d}{2}}\Bigg[(v_1)^{-\frac{d\alpha}{2}} + \frac{1}{\alpha}(v_2)^{-\frac{d\alpha}{2}} \\ & - \left(\frac{\alpha+1}{\alpha}\right)^{1+\frac{d}{2}}\left(v_1^\alpha(\frac{v_2}{v_1}+\frac{1}{\alpha})\right)^{-\frac{d}{2}}\exp\left(-\frac{||\theta_1-\theta_2||^2}{2(\frac{v_1}{\alpha}+v_2)}\right)\Bigg]. \end{aligned} \quad (4)$$

Proof. Identities (4) is readily verified, by exploiting the commonly used equality, using some properties of the squared error loss, for all θ_1, θ_2, y in $\mathbb{R}^d$:

$$\frac{||y-\theta_1||^2}{v_1} + \frac{||y-\theta_2||^2}{v_2} = \frac{||w-y||^2}{v_w} + \frac{||\theta_1-\theta_2||^2}{v_1+v_2},$$

where $w = B^*\theta_2 + (1-B^*)\theta_1$, $B^* = v_1/(v_1+v_2)$ and $v_w = v_1v_2/(v_1+v_2)$.

We can then derive the expression resp. of the DPD loss for the plug-in problem, namely, for $\theta_1 = \theta$, $\theta_2 = \hat{\theta}$ and $v_1 = v_2 = v_x$, where $\hat{\theta}(x)$ is the arbitrary efficient estimator of θ plugged in the predictive density estimator $\hat{q}(y|\theta = \hat{\theta}(x))$, based on X, then we obtain:

$$D_P(q,\hat{q}) = b_0\left[1 - \exp\left(-\frac{||\hat{\theta}(x)-\theta||^2}{2\frac{\alpha+1}{\alpha}v_x}\right)\right], \quad (5)$$

with $b_0 = ((2\pi v_x)^\alpha(\alpha+1))^{-\frac{d}{2}}\frac{\alpha+1}{\alpha}$.

Remark 2. Notice that we can rewrite the expressions (5) in function of the Reflected Normal Loss (RN loss) with one parameter γ, such that $\gamma = \frac{\alpha+1}{\alpha} v_x$, introduced in Spiring (1993), then

$$D_P(q,\hat{q}) = b_0 RN_\gamma(\theta,\hat{\theta}), \tag{6}$$

where $RN_\gamma(\theta,\hat{\theta}) = 1 - \exp\left(-\frac{||\hat{\theta}(x)-\theta||^2}{2\gamma}\right)$. RL is a concave loss in θ, the curve of its pdf is a exactly an up-side-down transformation of a normal pdf, hence its name "Reflected". We can give an illustration in the univariate case, i.e.

$$L(y) = K\left(1 - \exp\left(-\frac{(y-T)^2}{2\gamma^2}\right)\right),$$

where y represents the quality measurement, T the target value, y a shape parameter, and K the maximum-loss parameter. The target, shape (not a scale parameter, as $L(y)$ is not a density function), and maximum-loss parameters allow customization of the loss function to meet practitioners' requirements. The univariate form of the loss function (6) is illustrated in Figure 1.1 below.

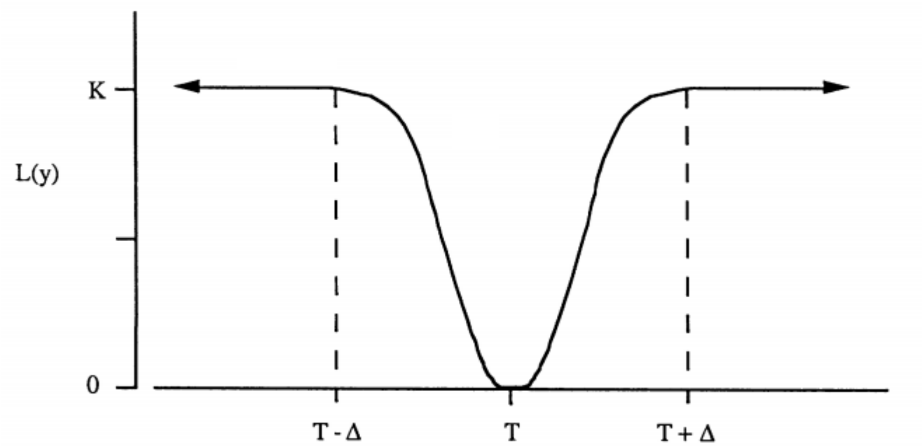

Fig. 1. The reflected normal loss function

where Δ is the distance from the target to the point where the maximum loss K first occurs.

The equality (6) yields a back-and-forth transfer of dominance between the RN loss and the DPD loss. It's worthy to recall the asymptotic relation between the squared error loss and the RN loss, that states:

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

which is equivalent to dominance under squared error loss by transitivity, thanks to this classical inequality $1 - e^{\omega} < \omega$, for any $\omega \neq 0$.

In the sequel, we will denote by $D_P(\theta, \hat{\theta})$ for the plug-in estimation problem under DPD loss, where $\hat{\theta}(x)$ is an efficient estimator of θ.

2.1. The Bayes estimator and Bayes predictive density estimator

The next proposition suggests the expression of the general Bayes predictive density estimator (BPDE) associated to an arbitrary prior (it could be improper) under the DPD loss, denoted by $\hat{q}_{\pi}(y|x)$. Obviously, the BPDE is proper if and only if the corresponding posterior pdf is proper. Where the BPDE turned out to be similar to the BPDE's under the divergence loss (also called alpha-divergence) Corcuera and Giummolè (1999), Mergel *et al* (2008); or under S-Hellinger distances Ommane and Ouassou (2019), in the sense that they all can be written in function of a normalization of the function $k_{\beta_2}^{\beta_1}(y, x)$, defined as:

$$k_{\beta_2}^{\beta_1}(y, x) = \left(\int q^{\beta_2}(y|\theta)\pi(\theta|x)d\theta \right)^{\beta_1},$$

where $\pi(\theta|x)$ is the posterior density.

Proposition 1. *Under DPD loss and any arbitrary prior π, the BPDE of Y is given by*

$$\hat{q}_{\pi}(y|x) \propto k_{\alpha}^{1/\alpha}(y, x). \tag{7}$$

Proof. Under DPD loss, the posterior risk associated to a predictor $\hat{q}(y)$ of the target density $q(y|\theta)$ is given by

$$\begin{aligned} \int D_P(q, \hat{q})\pi(\theta|x)d\theta &= \int \left(\frac{1}{\alpha} k_{\alpha+1}(y, x) - \frac{\alpha+1}{\alpha} \hat{q}(y) k_{\alpha}(y, x) + \hat{q}^{\alpha+1}(y) \right) dy \\ &= f(\hat{q}(y)) \end{aligned}$$

with f the corresponding function in $\hat{q}(y)$ to the posterior risk. It can be seen easily that the posterior risk attains its minimum over $\hat{q}(y)$ at $\left(\frac{1}{\alpha} k_{\alpha}^{1/\alpha}(y, x) \right)^{\frac{1}{\alpha}}$, normalizing this latter concludes the proof.

By virtue of (1), the next corollary derives the expression of the GBEE (General Best Equivariant Estimator), or the GMRE (General Minimum Risk Equivariant) predictive density estimator of Y , denoted $\hat{q}_{bee}(y|x)$. The latter corresponds to the noninformative prior and the minimaxity of this latter will be shown in the next section.

Corollary 1. *Under DPD loss and the noninformative prior $\pi(\theta) = 1$, the GBEE is given by:*

$$\hat{q}_{gbee}(y|x) \propto k_{*}^{1/\alpha}(y, x), \tag{8}$$

where $k_{*}(y, x) = \displaystyle\int q^{\alpha}(y|\theta)p(x|\theta)d\theta$.

Proof. This is an immediate consequence of (1), for $\pi(\theta) = 1$, which implies that $\pi(\theta|x) = p(x|\theta)$.

In the sequel we denote by ϕ the pdf of the standard gaussian distribution $N_d(0, I_d)$. We assume the model (1), and we consider a normal prior

$$\pi_G(\theta) = (v_\theta)^{-\frac{d}{2}} \phi\left(\frac{\theta - \mu}{\sqrt{v_\theta}}\right),$$

where μ and v_θ are known parameters, then the next lemma establishes the expression of both the Bayesian estimator (BE) and the Bayesian Predictive Density estimator (BPDE) .

Lemma 2. *Under the model (1) and resp. under the DPD loss, for the prior π_G, we have:*

1. *The BE of θ is $\hat{\theta}_{\pi_G}(X) = (1 - A)X + A\mu$,*
2. *The BPDE of Y given X is given by*

$$\hat{q}_{\pi_G}(y|x) = (\alpha v_\theta A + v_y)^{-\frac{d}{2}} \phi\left(\frac{y - \hat{\theta}_{\pi_G}(X)}{\sqrt{\alpha v_\theta A + v_y}}\right), \tag{9}$$

where $A = v_x/(v_x + v_\theta)$.

Proof. For (1): Since the posterior density states as

$$\pi_G(\theta|x) = (v_\theta A)^{-\frac{d}{2}} \phi\left(\frac{\theta - (1 - A)X + A\mu}{\sqrt{v_\theta A}}\right),$$

then the posterior risk of a given efficient point estimator $\hat{\theta}(x)$ of θ is given by

$$b\left[1 - \exp\left(-\frac{||\hat{\theta}(x) - (1 - A)X + A\mu||^2}{2\left(\frac{\alpha+1}{\alpha} v_y + A\right)}\right)\right],$$

where $b = ((2\pi v_y)^\alpha(\alpha + 1))^{-\frac{d}{2}}$. The BE is obtained by minimizing the posterior risk. It's clear that for $\hat{\theta}_{\pi_G}(X) = (1 - A)X + A\mu$, the posterior risk is minimal.

For (2): It's an immediate consequence of (1), placed under the model (1), for $\pi(\theta) = \pi_G(\theta)$, which concludes the proof.

3. Minimaxity results

Minimaxity is a frequentist (θ is deterministic) optimality principle, we owe this principle to game theory. Briefly, it's a criterion that suggests the best decision (estimator) at the worst scenario, originally conceived for zero-sum-games (i.e. a gain for a player is a loss for another). From a decision-theoretic perspective, there are many ways to prove minimaxity. For instance, when the frequentist risk associated to a decision (estimator) is constant (which is our case), a sufficient condition for minimaxity is to show that this estimator is a limit of Bayes estimators corresponding to a sequence of priors (called the least favorable priors).

We start off by expliciting the frequentist risk of the plug-in estimator of Y $\hat{q}_*(y|\theta = x)$ (i.e. considering the MLE X as a point estimator of θ), which is the UMVUE, the MRE, the BEE under translations of the sample space, as well as the corresponding BPDE under the noninformative prior, denoted by $\hat{q}_{bee}(y|x)$. The first can be obtained by limiting over $A \to 0$, $\hat{\theta}_{\pi_G}(X) = (1-A)X + A\mu \to X$, and since

$$\hat{q}_*(y|\theta = x) = v_x^{-\frac{d}{2}} \phi\left(\frac{y-x}{\sqrt{v_x}}\right), \tag{10}$$

while the predictive BEE corresponding to noninformative prior can be deduced from (9) for $\pi(\theta) = 1$,

$$\hat{q}_{bee}(y|x) = ((\alpha r + 1)v_y)^{-\frac{d}{2}} \phi\left(\frac{y-x}{\sqrt{(\alpha r+1)v_y}}\right), \tag{11}$$

with $r = v_x/v_y$.

By Lemma 1 under DPD loss, the frequentist risk of $\hat{q}_*(y|\theta = X)$ (i.e. of X) is given by

$$R_0 = R_\alpha(\theta, X) = b_0 \left[1 - \left(\frac{\alpha r}{\alpha+1} + 1\right)^{-d/2}\right] \tag{12}$$

Theorem 1. *X is a minimax estimator of θ in all dimensions, under DPD loss, equivalently, under scaled reflected normal loss $b_0 RN_{\gamma_1}$ (6).*

Proof. See Appendix (7).

We give the constant frequentist risk under DPD loss associated to $\hat{q}_{bee}(y|x)$:

$$R_U = R_\alpha(q, \hat{q}_{bee}) = b\left[1 + \frac{(\alpha r+1)^{-\frac{\alpha d}{2}}}{\alpha} - \left(\frac{\alpha+1}{\alpha}\right)(\alpha r+1)^{-\frac{d}{2}}\right]. \tag{13}$$

This latter will be employed in the next theorem.

Remark 3. We recall that the BEE $\hat{q}_{bee}(y|x)$ should naturally improve upon the plug-in estimator $\hat{q}_*(y|\theta = x)$, which can be seen by comparing their respective constant frequentist risks given in (12) and (13). It can easily be seen that $R_\alpha(\theta, \hat{q}_*) > R_\alpha(\theta, \hat{q}_{bee})$.

Likewise, the minimaxity of $\hat{q}_{bee}(y|x)$ will be shown in the next theorem.

Theorem 2. *$\hat{q}_{bee}(y|x)$ is a minimax predictive density estimator of $q(y|\theta)$ in all dimensions under DPD loss.*

Proof. See Appendix (7).

Moving on now we will argue the minimaxity from a broader perspective (broader than the model (1)), by considering the general BPDE (GBEE) $\hat{q}_{gbee}(y|x)$ corresponding to the noninformative prior given by (8), in corollary 1, the proof lies on a Girshick and Savage technique, see Girshick and Savage (1951).

Theorem 3. *The general BEE $\hat{q}_{gbee}(y|x)$ given by (8) is a minimax predictive density estimator of $q(y|\theta)$ for all dimensions.*

Proof. See Appendix (7).

So far this chapter has focused on the minimaxity of the two BEE estimators X and $\hat{\theta}_U(y|x)$ under the model (1). And even in general as shown in Theorem 3. So as these latters attain the full optimality, at least for $d \leq 2$. It is now necessary to argue their admissibility under the model (1).

4. Admissibility of the benchmark estimators X and $\hat{q}_U(y|x)$ in one dimension

The proof of the admissibility of the benchmark estimator X (resp. $\hat{q}_{bee}(y|x)$) in one dimension is based on Blyth's method, see Blyth (1951), Stein in his paper Stein (1955), as well as page 386 of Berger's book Berger (1985). It consists of supposing that X (resp. $\hat{q}_{bee}(y|x)$) is inadmissible, as a consequence, there exists a dominant estimator $\hat{\theta}_0$ smaller in risk for all θ, with a strict inequality for a certain θ_0. Then by considering a sequence of normal priors, concluding by a contradiction of the bayesian character of the corresponding estimator $\hat{\theta}_{\pi_n}$.

Theorem 4. *X is an admissible estimator of θ for $d = 1$, under DPD loss.*

Proof. See Appendix (7).

Similarly, we show in the next theorem the admissibility of $\hat{q}_{bee}(y|x)$ in one dimension.

Theorem 5. *$\hat{q}_{bee}(y|x)$ is an admissible estimator of $q(y|\theta)$ for $d = 1$, under DPD loss given in (2).*

Proof. See Appendix (7).

Remark 4. For $d = 2$, what about admissibility of the benchmark estimators X and $\hat{q}_{bee}(y|x)$?

The two-dimensional case requires a different technique for proving the admissibility of X or $\hat{q}_{bee}(y|x)$. In fact, if one tries out the Blyth's method, the ratios given in Appendix (53) and (54) they tend to a constant as $n \to \infty$. Yet, we are intuitively sure that X and $\hat{q}_{bee}(y|x)$ are admissible estimators of respectively θ and $q(y|\theta)$ when $d = 2$. Indeed, we took the plunge to demonstrate it via a more original but fastidious method suggested in Brown and Fox (1974), re-used elegantly ten years later in Gatsonis (1984) in his quest of deriving posterior distributions for location parameter under the L_2 distance (a special case of the DPD loss $\alpha = 1$). Unfortunately, we postponed this proof to an upcoming article.

5. The Stein effect and the inadmissibility of the benchmark estimators, when the dimension is greater than 3

In this section, we argue the inadmissibility of the benchmark estimators X and $\hat{\theta}(X)$ when the dimension goes beyond two dimensions, respectively under the plug-in and the predictive problem. It will be studied from two different angles, first, by considering the Baranchick family of estimators (which is a huge family containing many famous and useful estimators, as it will be illustrated below).

5.1. *The Stein effect for* $\hat{\theta}(X) = g(\frac{||X||^2}{v_x})X$ *class of dominators (e.g. Baranchick type dominators)*

Let $\hat{\theta}_g(X) = g(Z)\,X$, where g an arbitrary function and $Z = \frac{||X||^2}{v_x}$. In this framework, we rewrite the DPD loss respectively corresponding to the plug-in problem and the predictive problem, using (5), where $b_0 = \frac{\alpha+1}{\alpha}\left((2\pi v_x)^\alpha(\alpha+1)\right)^{-\frac{d}{2}}$, $\gamma_1 = \frac{\alpha+1}{\alpha r}$, $b = \left((2\pi v_x)^\alpha(\alpha+1)\right)^{-\frac{d}{2}}$ and $\gamma_2 = \alpha + \frac{1}{r} + \frac{1}{\alpha r}$. For the plug-in problem:

$$D_\alpha(\theta, \hat{\theta}_g) = b_0\left[1 - e^{-\frac{||\hat{\theta}_g(x)-\theta||^2}{2v_x\gamma_1}}\right],$$

whence its corresponding risk R_0^*:

$$R_0^* = R_\alpha(\theta, \hat{\theta}_g) = b_0\left[1 - E_\theta\left(e^{-\frac{||\hat{\theta}_g(x)-\theta||^2}{2v_x\gamma_1}}\right)\right]. \tag{14}$$

For the predictive problem, denoting $\hat{q}_g(y|x) = ((\alpha r+1)v_y)^{-\frac{d}{2}}\phi\left(\frac{y-\hat{\theta}_g(x)}{\sqrt{(\alpha r+1)v_y}}\right)$:

$$D_P(q, \hat{q}_g) = b\left[1 + \frac{1}{\alpha}(\alpha r+1)^{-\frac{d}{2}} - \left(\frac{\alpha+1}{\alpha}\right)^{-\frac{d}{2}}\left(\alpha r + 1 + \frac{1}{\alpha}\right)^{-\frac{d}{2}} e^{-\frac{||\hat{\theta}_g(x)-\theta||^2}{2v_x\gamma_2}}\right],$$

whence its corresponding risk R_U^*:

$$R_U^* = R_\alpha(q, \hat{q}_g) = b\left[1 + \frac{1}{\alpha}(\alpha r+1)^{-\frac{d}{2}} - \left(\frac{\alpha+1}{\alpha}\right)^{-\frac{d}{2}}\left(\alpha r + 1 + \frac{1}{\alpha}\right)^{-\frac{d}{2}} E_\theta\left(e^{-\frac{||\hat{\theta}_g(x)-\theta||^2}{2v_x\gamma_2}}\right)\right]. \tag{15}$$

The first result of this subsection gives an explicit form of the expectation

$$\mathrm{I}_\gamma = E_\theta\left[\exp\left(-\frac{||g(Z)T-\eta||^2}{2\gamma}\right)\right], \tag{16}$$

with $\gamma \in \{\gamma_1, \gamma_2\}$, simplified into an integral on $(0, \infty)$. For $T = \frac{X}{\sigma_x}$, where $\sigma_x^2 = v_x$ and $Z = ||T||^2 = \frac{||X||^2}{v_x} \sim \chi_d^2(\delta^2)$, where $\delta^2 = ||\eta||^2 = \frac{||\theta||^2}{v_x}$, with $\eta = \frac{\theta}{\sigma_x}$, considering the conditional expectation decomposition we have:

$$E_\theta\left[\exp\left(-\frac{||g(Z)T-\eta||^2}{2\gamma}\right)\right] = E^Z\left\{E_\theta\left[\exp\left(-\frac{||g(Z)T-\eta||^2}{2\gamma}\right)\right]|Z\right\}. \tag{17}$$

The pdf of variable Z is described as:

$$\begin{aligned} f(z) &= \frac{1}{2}\left(\frac{z}{\delta^2}\right)^{\frac{d-2}{4}} \mathcal{I}_{\frac{d-2}{4}}(\delta\sqrt{z})e^{-\frac{\delta+z}{2}} \\ &= 2^{-\frac{d}{2}}\frac{z^{\frac{d-2}{2}}}{\Gamma(\frac{d}{2})}{}_0F_1(;\frac{d}{2};\frac{z\delta^2}{4})e^{-\frac{\delta+z}{2}}, \end{aligned}$$

with $\mathcal{I}_a(b)$ the modified Bessel function of the first type of order a given by

$$\mathcal{I}_a(b) = \frac{(\frac{b}{2})^a}{\Gamma(a+1)}{}_0F_1(;b+1;\frac{b^2}{4}),$$

where ${}_0F_1$ the hypergeometric function defined as:

$${}_0F_1(-;\nu;\mu) = \sum_{k=0}^{\infty}\frac{\mu^k}{k!(\nu)_k},\ \nu>0, \mu\in\mathbb{R},$$

where

$$(\nu)_k = \begin{cases} \nu(\nu+1)...(\nu+k-1) & /k>0 \\ 1 & /k=0. \end{cases}$$

The following lemma is an auxiliary result to simplify (16) into an integral on $\mathbb{R}_+$.

Lemma 3. *Whenever,* $X\sim N_d(\theta, v_x I_d)$, $\delta=\frac{||\theta||}{\sigma_x}$, $W=\frac{\theta' X}{||\theta||||X||}$. *Then, we have*

$$E(Z\big|||X||=r) \propto e^{\delta r t}(1-t^2)^{\frac{d-3}{2}}\mathbb{I}_{[-1,1]}(t).$$

Furthermore, for $s>0$,

$$E(e^{st}\big|||X||=r) = \frac{{}_0F_1(;\frac{d}{2};\frac{(s+\delta r)^2}{4})}{{}_0F_1(;\frac{d}{2};\frac{(\delta r)^2}{4})} = \left(\frac{\delta r}{\delta r+s}\right)^{\frac{d}{2}-1}\frac{\mathcal{I}_{\frac{d}{2}-1}(\delta r+s)}{\mathcal{I}_{\frac{d}{2}-1}(\delta r)}. \qquad (18)$$

Proof. It suffices to set $\sigma_x\delta=(||\theta||,0,\ldots,0)$ and $W=\frac{X_1}{||X||}$, then we have $X_1\sim N(\delta,1)$ and $||X||^2-X_1^2\sim\chi^2_{d-1}$, together with the independence between X_1 and $||X||^2-X_1^2$, we can derive the joint distribution of $(X_1,||X||)$, as well as the marginal density of X_1 given $||X||=r$ from the joint distribution of $(X_1,||X||^2-X_1^2)$. Finally an appeal of the following result that appeared in (Abramovitz ans Stegun (1966), p. 376), that stated that: For $c\in\mathbb{R}$, $a>0$ and $B(.,.)$ the Beta function, we have:

$$\begin{aligned} \int_{-1}^{1} e^{ct}(1-t^2)^{a-1}dt &= 2\int_0^1 \cosh(ct)(1-t^2)^{a-1}dt \\ &= \frac{\Gamma(\frac{1}{2})\Gamma(a)}{(\frac{c}{2})^{\frac{a-1}{2}}}\mathcal{I}_{a-\frac{1}{2}}(c) = B(\frac{1}{2},a){}_0F_1(;a+\frac{1}{2};\frac{c^2}{4}), \end{aligned}$$

which brings about the expression (18), which ends of the proof.

The next proposition proposes a primarily more convenient form of (16).

Proposition 2. *For $\hat{\theta}(X) = g(\frac{||X||^2}{v_x})X$ and I_γ defined as in (16), we have:*

$$\begin{aligned}
\mathrm{I}_\gamma &= e^{-\frac{\delta^2}{2\gamma}} E^Z \left(e^{-\frac{g^2(z)z}{2\gamma}} \frac{{}_0F_1(;\frac{d}{2};\frac{z\delta^2}{4}(1+\frac{g(z)}{\gamma})^2)}{{}_0F_1(;\frac{d}{2};\frac{z\delta^2}{4})} \right) \\
&= e^{-\frac{\delta^2}{2\gamma}} E^Z \left(e^{-\frac{g^2(z)z}{2\gamma}} \left(\frac{\gamma}{\gamma+g(z)} \right)^{\frac{d}{2}-1} \frac{\mathcal{I}_{\frac{d}{2}-1}(\delta\sqrt{z}(1+\frac{g(z)}{\gamma}))}{\mathcal{I}_{\frac{d}{2}-1}(\delta\sqrt{z})} \right) \\
&= e^{-\frac{\delta^2(1+\gamma)}{2\gamma}} \frac{2^{-\frac{d}{2}}}{\Gamma(\frac{d}{2})} \int_0^\infty z^{\frac{d}{2}-1} e^{-\frac{z}{2}(1+\frac{g^2(z)}{\gamma})} {}_0F_1(;\frac{d}{2};\frac{z\delta^2}{4}) dz. \qquad (19)
\end{aligned}$$

Proof. The expression (19) is readily verified, it suffices to consider the conditional decomposition in (17), then by expanding the squared norm $||g(Z)T - \eta||^2$, gathered with (3) ends the proof.

We can get a more convenient formula of I_γ by considering the case of Baranchick estimators, denoted by

$$\hat{\theta}_B(X) = \left(1 - \frac{\tau(Z)}{Z}\right) X,$$

with τ an arbitrary function, in order to establish suitable dominance conditions under both the plug-in and the predictive problem. The next proposition suggests a much more convenient expression than the previous, deduced from (19).

Proposition 3.

$$\mathrm{I}_\gamma = \left(\frac{\gamma+1}{\gamma}\right)^{-\frac{d}{2}} \sum_{k=0}^{\infty} \frac{e^{-h_\gamma} h_\gamma^k}{k!} \beta_\gamma(k). \qquad (20)$$

where $h_\gamma = \delta^2 \left(\frac{\gamma+1}{2\gamma}\right)$ and

$$\beta_\gamma(k) = \int_0^\infty \frac{e^{(-\varphi(u) - \frac{\tau_0^2(u)}{4\gamma u})}}{\Gamma(k+d/2)} (\varphi(u))^{k+\frac{d}{2}-1} \left(1 - \frac{\tau_0(u)}{2\gamma u}\right)^{-(d-2)} du,$$

with $u = \frac{\gamma+1}{2\gamma} z$, $\tau_0(u) = \tau\left(\frac{2\gamma z}{\gamma+1}\right)$ and $\varphi(u) = u(1 - \frac{\tau_0(u)}{2\gamma u})^2$.

Proof. It suffices to replace the hypergeometric function by its discrete form, taking into account the substitutions:$\tau_0(u) = \tau\left(\frac{2\gamma z}{\gamma+1}\right)$, $h_\gamma = \delta^2\left(\frac{\gamma+1}{2\gamma}\right)$ and $u = \frac{\gamma+1}{2\gamma} z$. The expression (20) stems from Lebesgue's theorem, with an appeal to an established result in (Muirhead (2005), p. 23), which ends the proof.

Thanks to (20), we can nicely rewrite respectively the risk expressions R_0^* and R_U^* defined by (14) and (15), under the plug-in and the predictive problem respectively, after simplification they state as:

$$R_0^* = b_0 \left[1 - \left(\frac{\alpha r}{\alpha+1} + 1 \right)^{-\frac{d}{2}} \sum_{k=0}^{\infty} \frac{e^{-h_{\gamma_1}} h_{\gamma_1}^k}{k!} \beta_{\gamma_1}(k) \right], \qquad (21)$$

with $\gamma_1 = \frac{\alpha+1}{\alpha r}$ and

$$R_U^* = b\left[1 + \frac{1}{\alpha}(\alpha r+1)^{-\frac{d}{2}} - \left(\frac{\alpha+1}{\alpha}\right)(\alpha r+1)^{-\frac{d}{2}}\sum_{k=0}^{\infty}\frac{e^{-h_{\gamma_2}}h_{\gamma_2}^k}{k!}\beta_{\gamma_2}(k)\right], \tag{22}$$

with $\gamma_2 = \alpha + \frac{1}{r} + \frac{1}{\alpha r}$.

The next theorem argues a sufficient condition of dominance of the Baranchick type estimators over the two benchmark estimators X and $\hat{q}_{bee}(X)$, under the plug-in and the predictive problem respectively, obstruding some regularities on the arbitrary function τ.

Remark 5. It's worthy to highlight that the condition of dominance meets with the one established in Mergel *et al* (2008) under divergence loss, except that our proof of the expression (20) as shown above is conveniently shorter.

Theorem 6. *Under the model (1), for $d \geq 3$ and τ an arbitrary function. Provided:*

1. $0 < \tau(t) < 2(d-2)$ *for all* $t > 0$*;*
2. $\tau(t)$ *is a differentiable nondecreasing function on* t*.*

Then, $\hat{\theta}_B(X)$ improves upon X resp. under DPD loss, equivalently under scaled reflected normal loss $b_i RN_{\gamma_1}(\theta, \hat{\theta}_B(x))$ (6). Also, $\hat{q}_B(y|x) = (v_y)^{-\frac{d}{2}}\phi\left(\frac{y-\hat{\theta}_B(x)}{\sqrt{(\alpha r+1)v_y}}\right)$ improves upon $\hat{q}_{bee}(y|x) = (v_y)^{-\frac{d}{2}}\phi\left(\frac{y-x}{\sqrt{(\alpha r+1)v_y}}\right)$.

Proof. It can be seen according to (14) and (12) (resp. (15) and (13)), that a sufficient condition of dominance of $\hat{\theta}_B(X)$ over X (resp. $\hat{q}_B(y|x)$ over $\hat{q}_{bee}(y|x)$) under DPD loss, for both problems, is that $\left(\frac{\gamma+1}{\gamma}\right)^{\frac{d}{2}} I_\gamma \geq 1$. For this end, we set $u_0 = \sup\{u > 0 : \tau_0(u)/u \geq 2\gamma\}$, since $\frac{\tau_0(u)}{u}$ is continuous in u, moreover, $\lim_{u\to 0}\frac{\tau_0(u)}{u} = +\infty$ and $\lim_{u\to\infty}\frac{\tau_0(u)}{u} = 0$, peculiarly for $\frac{\tau_0(u)}{u} = 2\gamma$. Now we may adapt a lemma in Mergel *et al* (2008) to our case, stating as the following:

Lemma 4. *For $u \geq u_0$ and $\gamma > 0$, if $\tau_0(u)$ verifies (6) and $\varphi(u)$ is defined as in (20), then*

$$e^{-\frac{\tau_0^2(u)}{4\gamma u}}\left(1 - \frac{\tau_0(u)}{2\gamma u}\right)^{-(d-2)} \geq \varphi'(u). \tag{23}$$

Proof. For $u \geq u_0$, we have the classical inequality $(1-\omega)^{-\varepsilon} \geq e^{\varepsilon\omega}$, whenever, $\varepsilon > 0$ and $0 < \omega < 1$. Together with condition (1) in (6), makes the left hand side of the inequality (23) satisfy:

$$e^{-\frac{\tau_0^2(u)}{4\gamma u}}\left(1 - \frac{\tau_0(u)}{2\gamma u}\right)^{-(d-2)} \geq e^{-\left(\frac{\tau_0^2(u)}{4\gamma u} - \frac{d-2}{2\gamma}\frac{\tau_0(u)}{u}\right)}.$$

Additionally, the derivative of $\varphi(u)$ is described by

$$\varphi^{'}(u) = 1 - \left(\frac{\tau_0(u)}{\gamma u} + \frac{\tau_0(u)\tau_0^{'}(u)}{2\gamma^2 u} - \frac{\tau_0^{'}(u)}{\gamma} - \frac{\tau_0^2(u)}{2\gamma^2 u^2} \right).$$

Finally, if $\varphi^{'}(u) \leq 1$ is true, it concludes the proof. Indeed, since:

$$\varphi^{'}(u) \leq 1 \Leftrightarrow \frac{1}{\gamma} \left(1 - \frac{\tau_0(u)}{2\gamma u}\right) \left(\tau_0^{'}(u) - \frac{\tau_0(u)}{u}\right) \geq 0.$$

In fact, $u > u_0 \Rightarrow \left(1 - \frac{\tau_0(u)}{2\gamma u}\right) \geq 0$, besides, $\frac{\tau_0(u)}{u} \geq 2\gamma \Rightarrow \left(\tau_0^{'}(u) - \frac{\tau_0(u)}{u}\right) \geq 0$.

Hence, if we set $U = \varphi(u)$ (i.e. $dU = \varphi^{'}(u)du$) the theorem follows thanks to (23), such that:

$$\begin{aligned} \beta_\gamma(k) &\geq (\Gamma(k + d/2))^{-1} \int_{t_0}^{\infty} e^{-U} (U)^{k+\frac{d}{2}-1} dU \\ &\geq (\Gamma(k + d/2))^{-1} \Gamma(k + d/2) = 1. \end{aligned}$$

Therefore, by (20), we get

$$\left(\frac{\gamma + 1}{\gamma}\right)^{\frac{d}{2}} I_\gamma \geq \sum_{k=0}^{\infty} \frac{e^{-h_\gamma} h_\gamma^k}{k!} = 1.$$

Remark 6. We note that the class of Baranchick estimators is a very rich class, which was the central object of numerous previous works, that reached the zenith during the seventies of the last century. Many authors had established sufficient conditions of dominance on the $\tau(.)$, such as, Baranchick (1970), Strawderman (1971), Alam (1973), Stein (1973, 1981), Berger (1976a, 1976b), Efron and Morris (1976), Faith (1978), DasGupta and Strawderman (1997) and Fourdrinier, Strawderman and Wells (1998). Moreover, several minimax estimators were introduced. The most famous member of the Baranchick class is the James-Stein estimator (1961), as well as its positive part found later by Baranchick (1964), among many others, for instance Strawderman (1971), Alam (1973), Berger (1976a), Li and Kuo (1982), Kubokawa (1991), Guo and Pal (1992), Shao and Strawderman (1994), Maruyama (1998, 2004, 2007) and Kuriki and Takemura (2000). Nevertheless, among this dizzying multitude of competitors it remains hard to set an exhaustive ranking of all of them. For instance, to the best of our knowledge, the only rival that out-performs the positive part of the James-Stein estimator is the one of Shao and Strawderman (1994). Few works tackled the ranking problem, we cite Magnus (2002), which only considered the univariate case (i.e. no quest for minimax estimators). Recently Hansen (2015) provided a sharp efficiency bound for $d \geq 3$ based the sufficient condition shown by Efron and Morris (1976). He went further by constructing a novel shrinkage estimator, having a considerably smaller "MaxRegret" (the maximum of the regret function corresponding to an estimator $\hat{\theta}$, which is the difference between its risk and the efficiency bound). We illustrate in the examples below the different forms that the function $\tau(.)$ could take.

Example 1. **Various cases of the function $\tau(.)$, chronologically throughout history :**

- James-Stein estimator (1961):
$$\tau(z) = \tau_{JS}(z) = d - 2. \tag{24}$$
- Baranchick (1964, 1970):
$$\tau(z) = \tau_B(z) = \min(z, d-2). \tag{25}$$
- Li and Kuo (1982):
$$\tau(z) = \tau_{LK}(z) = d - 2 - c_1 z^{-\frac{\alpha_1}{2}}; \tag{26}$$
where $c_1 = \alpha_1 2^{\frac{\alpha_1}{2}} \frac{\Gamma(d/2 - (1+\alpha_1/2))}{\Gamma(d/2 - (1+\alpha_1))}$, with $0 < \alpha_1 < d/2 - 1$.
- Guo and Pal (1992):
$$\tau(z) = \tau_{GP}(z) = d - 2 - \sum_{j=1}^{n} c_j z^{-\frac{\alpha_j}{2}}; \tag{27}$$
with $0 < \alpha_1 < \ldots < \alpha_n < d/2 - 1$. The authors gave also explicit formulas for the constants c_j.
- Shao and Strawderman (1994). Under gamma distribution model (not the normal model):
$$\tau(z) = \tau_{SS}(z) = \min(z, d-2) + ag(|z|); \tag{28}$$
where
$$g(t) = \begin{cases} 2\omega - 1 + t & , 0 \le t \le \omega \\ t - 1 & , \omega < t \le 1 \\ 0 & , 1 < t < \infty. \end{cases}$$
The authors provided the values of a and ω ensuring the dominance of $\hat{\theta}_{SS}(X)$ over $\hat{\theta}_B(X)$.
- Kuriki and Takemura (2000):
$$\tau(z) = \tau_{KT}(z) = \begin{cases} d - 2 - \frac{r}{\sqrt{z} - r} & , z \ge \left(\frac{d-1}{d-2} r\right)^2 \\ 0 & , z < \left(\frac{d-1}{d-2} r\right)^2, \end{cases} \tag{29}$$
with $r \ge 0$.
- Strawderman (1971), Alam (1973), Kubokawa (1991) and Maruyama (1998, 2004):
$$\tau(z) = \tau_M(z) = z \frac{B(b+1, d/2 - a + 2) {}_1F_1(b+1; d/2 - a + b + 3; z/2) + \beta}{B(b+1, d/2 - a + 1) {}_1F_1(b+1; d/2 - a + b + 2; z/2) + \beta}, \tag{30}$$
where $B(.,.)$ is the beta function and ${}_1F_1(.;.;.)$ the confluent hypergeometric function, with $3 - d/2 \le a \le 1 + d/2$, $b \ge -\frac{a + d/2 - 3}{3d/2 + 1 - a}$ and
$$0 \le \beta \le B(d/2 - a + 1, b - a + d/2 + 2) \times \left(\sqrt{1 + \frac{D_M(a,b)(d/2 - a + 1)(b+1)}{(b - a + d/2 + 2)(b - a + d/2 + 3)}} - 1 \right),$$

such that

$$D_M(a,b) = \begin{cases} a + d/2 - 3 & , b \leq 0 \\ \frac{b(3d/2+1-a)+(a+d/2-3)}{b+1} & , b < 0. \end{cases}$$

We owe this latter shrinkage function to Maruyama (2004). The rest of the cases are special case (when $\beta = 0$), namely, for Maruyama (2004) a and b are arbitrary. For Kubokawa (1991) $a = 2$ and $b = 0$. For Strawderman (1971) $b = 0$ and a is arbitrary. Alam (1973) $b = \nu - 1$ and $a = \nu + 1$ where ν is arbitrary.

Example 2. ***Superharmonic priors.*** By virtue of Stein (1981), we have that when the prior π is superharmonic, then the corresponding Bayes estimator $\hat{\theta}_\pi(X)$ dominates X under DPD loss, equivalently, $\hat{\theta}_\pi(X)$ dominates X under scaled reflected normal loss. Again $\hat{\theta}_\pi(W)$ dominates W under quadratic loss, where $W \sim N_d(\theta, v_w I_d)$, with $v_w = \frac{\gamma_1 v_x}{\gamma_1 + v_x}$, thanks to (6). A broader case is available as well, by considering this time, the square root of the marginal density of X (resp. W) under π, to be superharmonic, then the minimaxity follows, Fourdrinier *et al* (1998) gave several illustrations of such estimators.

Example 3. ***Hierarchical priors.*** Strawderman (1971) introduced a hierarchical prior such that: $\theta|\lambda \sim N_d(0, \lambda I_d)$, where $\lambda \sim \pi(\lambda) = a(1+\lambda)^{-(1+a)} I_{a>0}$, with $a > 0$. The corresponding estimator is dominant when $d > 4 + 2a$.

5.2. Application to generalized Lindley's estimator

In a discussion of Stein's paper '*Confidence sets for the mean of a multivariate normal distribution*', Lindley introduced a new rival based on the James-Stein estimator, in his paper Lindley (1962), which shrinks toward the sample mean $\bar{X}\mathbf{1}_d$, instead of shrinking X toward an arbitrary μ, where $\mathbf{1}_d$ is a d-variate column vector with ones everywhere. Assuming $Z_L = \frac{||X - \bar{X}\mathbf{1}_d||^2}{v_x}$, Lindley's estimator states as

$$\hat{\theta}_L(X) = X - \frac{d-3}{Z_L}(X - \bar{X}\mathbf{1}_d), d \geq 4. \tag{31}$$

We can naturally upgrade this latter estimator to a broader case, by rewriting (31) as a general Baranchick class of estimators, such that:

$$\hat{\theta}_{LB}(X) = X - \frac{\tau(Z_L)}{Z_L}(X - \bar{X}\mathbf{1}_d), d \geq 4. \tag{32}$$

Henceforth, we can easily recover a minimaxity result similar to the one established in the previous section, by a similar technique, thanks to Proposition 2 and Proposition 3, which we formulate in the next theorem:

Theorem 7. *Under the model (1), for $d \geq 4$. Provided:*

1. $0 < \tau(t) < 2(d-3)$ *for all* $t > 0$;
2. $\tau(t)$ *is a differentiable nondecreasing function on* t.

Then, $\hat{\theta}_{LB}(X)$ improves upon X under DPD loss, equivalently under scaled reflected normal loss $b_0 RN_{\gamma_1}(\theta, \hat{\theta}_{LB}(x))$ (6). Also, $\hat{q}_{LB}(y|x) = (v_y)^{-\frac{d}{2}}\phi\left(\frac{y-\hat{\theta}_{LB}(x)}{\sqrt{(\alpha r+1)v_y}}\right)$ improves upon

$\hat{q}_{bee}(y|x) = (v_y)^{-\frac{d}{2}}\phi\left(\frac{y-x}{\sqrt{(\alpha r+1)v_y}}\right)$ *under DPD loss (4).*

Proof. Let $\kappa = \frac{||\bar{\theta}\mathbf{1}_d - \theta||^2}{v_x}$ and $T = \frac{X}{\sigma_x}$. We note that Proposition 2 is not applicable here, since $\hat{\theta}_{LB}(X)$ does not shrink toward X, but instead shrinks toward $\bar{X}$, i.e. $\hat{\theta}_{LB}(X)$ is not of the form $g(\kappa)X$, but rather $g(\kappa)X + g(\kappa)\bar{X}$. Therefore, we will abide by a somewhat similar outline, exploiting again the invariance of our setting toward orthogonal transformations (geometrically speaking: invariance toward rotations). Let $\bar{\theta}$ be the sample mean of θ_i's for $i = 1, \ldots, d$, and $\bar{\eta} = \bar{\theta}/\sigma_x$ and $\rho^2 = v_x^{-1}\sum_{i=1}^{d}(\theta_i - \bar{\theta})^2$, thus, the quadratic form in $I_{\gtrsim}$ can rewritten as:

$$\begin{aligned} v_x^{-1}||\hat{\theta}_{LB}(X) - \theta||^2 &= \left|\left|(\bar{T}-\bar{\eta})\mathbf{1}_d - (\eta - \bar{\eta}\mathbf{1}_d) - \left(1-\frac{\tau(\kappa)}{\kappa}\right)(T - \bar{T}\mathbf{1}_d)\right|\right|^2 \\ &= d(\bar{T}-\bar{\eta})^2 + \rho^2 + 2(\eta-\bar{\eta}\mathbf{1}_d)^t\left(1-\frac{\tau(\kappa)}{\kappa}\right)(T-\bar{T}\mathbf{1}_d) \\ &+ \left(1-\frac{\tau(\kappa)}{\kappa}\right)^2 T. \end{aligned} \tag{33}$$

Now, performing a $\mathbf{C}$-orthogonal transformation ($\mathbf{G} = \mathbf{C}Z_L = (G_1, \ldots, G_d)^t$), with $\mathbf{C}$ is an orthogonal matrix, where its first two rows are respectively $\sqrt{d}^{-1}\mathbf{1}_d$ and $((\eta_1-\bar{\eta})/\rho, \ldots, (\eta_d - \bar{\eta})/\rho)$. As a result, by (34) and denoting $\tilde{\kappa} = G_2^2 + \sum_{i=3}^{d} G_i^2$, where $\mathbf{G}$ and $\sum_{i=3}^{d} G_i^2$ are mutually independent, with $G_1 \sim N(\sqrt{d}\bar{\eta}, 1)$, and $(G_2, \ldots, G_d)$ are i.i.d standard gaussian variables, we have

$$v_x^{-1}||\hat{\theta}_{LB}(X) - \theta||^2 = \left(1-\frac{\tau(\tilde{\kappa})}{\tilde{\kappa}}\right)\left[\left(1-\frac{\tau(\tilde{\kappa})}{\tilde{\kappa}}\right)\tilde{\kappa} - 2\rho G_2\right] + (G_2 - \sqrt{d}\bar{\eta})^2 + \rho^2. \tag{34}$$

Gathering the independence of G_1 with $(G_2, \ldots, G_d)$, with (34) and repeating the same technique of Proposition 3, then $\bar{\mathrm{I}}_\gamma$ ends up to be expressed by

$$\begin{aligned} \bar{\mathrm{I}}_\gamma &= \left(\frac{\gamma+1}{\gamma}\right)^{-\frac{d}{2}} E_\theta\left\{\left(1-\frac{\tau(\tilde{\kappa})}{\tilde{\kappa}}\right)\left[\left(1-\frac{\tau(\tilde{\kappa})}{\tilde{\kappa}}\right)\tilde{\kappa} - 2\rho G_2\right] + \rho^2\right\} \\ &= \left(\frac{\gamma+1}{\gamma}\right)^{-\frac{d}{2}} \sum_{k=0}^{\infty} \frac{e^{-\bar{h}_\gamma}\bar{h}_\gamma^k}{k!}\beta_\gamma(k), \end{aligned} \tag{35}$$

where $\bar{h}_\gamma = \left(\frac{\gamma+1}{2\gamma}\right)\rho^2$, while $\beta_\gamma(k)$, $\tau_0(u)$ and $\varphi(u)$ remain unchanged. In conclusion, an appeal to Lemma 4 given assumptions (1) and (2), finishes the proof.

We can go further with the generalized Lindley's estimator, by generalizing Theorem 7 to shrinkage toward a regression surface. For this end, we consider a hierarchical model, such that

$$X|\theta \sim N_d(\theta, v_x I_d) \textit{ and } \theta \sim N_d(M\beta, N I_d),$$

where M is given $d \times n$ matrix of rank $n < d$ and β a n-variate column vector as a regression coefficient. Let $P = M(M^tM)^{-1}M^t$, the projection of X on the regression surface described by n columns of M which is a $d \times d$ idempotent symmetric matrix or rank n, then $PX = M\hat{\beta}_{ls}$, where $\hat{\beta}_{ls} = (M^tM)^{-1}M^t$ the classic least squares estimator of β. Whence we may construct a general class of estimators described as

$$\hat{\theta}_R(X) = X - \frac{\tau(\kappa_*)}{\kappa_*}(X - PX), \tag{36}$$

where $\tilde{\kappa}_*^2 = \frac{||X-PX||^2}{v_x}$. Thus, an extension of Theorem 7 is formulated in the next theorem:

Theorem 8. *Under the model (1), for $d \geq n+3$. Provided:*

1. $0 < \tau(t) < 2(d-n-2)$ *for all $t > 0$;*
2. $\tau(t)$ *is a differentiable nondecreasing function on t.*

Then, $\hat{\theta}_R(X)$ improves upon X under DPD loss (4), equivalently under scaled reflected normal loss $b_0RN_{\gamma_1}(\theta, \hat{\theta}_R(x))$ (6). Also, $\hat{q}_R(y|x) = (v_y)^{-\frac{d}{2}}\phi\left(\frac{y - \hat{\theta}_R(x)}{\sqrt{(\alpha r+1)v_y}}\right)$ improves upon $\hat{q}_{bee}(y|x) = (v_y)^{-\frac{d}{2}}\phi\left(\frac{y-x}{\sqrt{(\alpha r+1)v_y}}\right)$ under DPD loss.

Proof. Let $\kappa_* = \frac{||\bar{\theta}\mathbf{1}_d - \theta||^2}{v_x}$ and $T = \frac{X}{\sigma_x}$. We notice that $\hat{\theta}_R(X) = g(\kappa_*)X$, with $g(\kappa_*) = \left(I_d - \frac{\tau(\kappa_*)}{\kappa_*}(I_d - P)\right)$, where g $\mathbb{R}_+ \to \mathbb{R}^d$ is now a function. Proposition 2 is now applicable componentwise, except that we need a different algebraic tool to prepare the quadratic norm than the one used in Lemma 3.Namely the spectral decomposition theorem, such that if we denote $\tilde{\rho} = \eta^t(I_d - P)\eta$, exploiting the fact that $P(I_d - P) = 0$, we have:

$$\begin{aligned} v_x^{-1}||\hat{\theta}_R(X) - \theta||^2 &= \left|\left|\left(1 - \frac{\tau(\tilde{\kappa}_*)}{\tilde{\kappa}_*}\right)(T - TP) + P(T-\eta) - (I_d - P)\eta\right|\right|^2 \\ &= \left(1 - \frac{\tau(\tilde{\kappa}_*)}{\tilde{\kappa}_*}\right)\left[1 - \frac{\tau(\tilde{\kappa}_*)}{\tilde{\kappa}_*}\tilde{\kappa}_* - 2\eta^t(I_d - P)(T - PT) + \tilde{\rho}\right. \\ &+ \left.(P(T-\eta))^t(P(T-\eta))\right] \end{aligned} \tag{37}$$

By virtue of the spectral decomposition theorem we get $P = \sum_{i=1}^n \varrho_i\varrho_i^t$, where ϱ_i's are orthonormal vectors. Through which we rotate the setting, via a orthogonal transformation on T, such that $U_i = \varrho_i^tT$ and $\eta_i = \varrho_i^t\eta$, for $i = 1, \ldots, n$, where $U_i \sim N(\eta_i, 1)$ are mutually independent, then we obtain $(P(T-\eta))^t(P(T-\eta)) = \sum_{i=1}^n (U_i - \eta_i)^2 \sim \chi_n^2$. Expressing $(I_d - P) = \sum_{i=n+1}^d \varrho_i\varrho_i^t$, where the ϱ_i's are orthonormal with $\varrho_{n+1} = (I_d - P)\eta/(\eta(I_d - P)\eta)$, together with the fact that $(U_1, \ldots, U_d)$ are mutually independent with $U_{n+1} = \varrho_{n+1}^tT \sim N(\varrho_{n+1}^t\eta, 1)$, whereas, $(U_{n+2}, \ldots, U_d)$ are i.i.d standard gaussian variables, thus (37) can be rewritten as

$$\left(1-\frac{\tau(\tilde{\kappa}_*)}{\tilde{\kappa}_*}\right)\left[1-\frac{\tau(\tilde{\kappa}_*)}{\tilde{\kappa}_*}\tilde{\kappa}_*-2\eta^t(I_d-P)(T-PT)+\tilde{\rho}\right].$$

Combining the above with Propositions 2 and 3, and adopting the same outline as in in the proof of Proposition 3 in computing I_γ, denoted now as $\tilde{\mathrm{I}}_\gamma$,

$$\tilde{\mathrm{I}}_\gamma=\left(\frac{\gamma+1}{\gamma}\right)^{-\frac{d}{2}}\sum_{k=0}^{\infty}\frac{e^{-\tilde{h}_\gamma}\tilde{h}_\gamma^k}{k!}\tilde{\beta}_\gamma(k), \tag{38}$$

where $\tilde{h}_\gamma=\left(\frac{\gamma+1}{2\gamma}\right)\tilde{\rho}$ and

$$\tilde{\beta}_\gamma(k)=\int_0^\infty\frac{e^{\left(-\varphi(u)-\frac{\tau_0^2(u)}{4\gamma u}\right)}}{\Gamma(k+(d-n)/2)}(\varphi(u))^{k+\frac{d-n}{2}-1}\left(1-\frac{\tau_0(u)}{2\gamma u}\right)^{-(d-2)}du,$$

while $\tau_0(u)$ and $\varphi(u)$ remain unchanged. In conclusion, an appeal to Lemma 4 given assumptions (1) and (2), finishes the proof.

6. Inadmissibility findings when the variance covariance matrix is unknown

6.1. The degenerate case i.e. $\Sigma=vI_d$

We remain under the model (1), where v_x is unknown. Hence, we introduce a natural estimator of v_x independently of X, usually denoted by S_1, such that: $S_1\sim\frac{v_x}{n+2}\chi^2_n$. Actually this situation models classically a one-way ANOVA with a balanced fixed effects, such that:

$$X_{ij}=\theta_i+\varepsilon_{ij},\ (i=1,\ldots,p;j=1,\ldots,m), \tag{39}$$

where the ε_{ij}'s are i.i.d, with $\varepsilon\sim N_d(0,v_0I_d)$. Then, the minimal sufficient estimator is typically $(\bar{X}_1,\ldots,\bar{X}_p,S)$, where $X_i=m^{-1}\sum_{j=1}^m X_{ij}$ for $i=1,\ldots,p$ and the estimator S states as:

$$S_1=(2+(n-1)p)^{-1}\sum_{i=1}^p\sum_{j=1}^m(X_{ij}-\bar{X}_i)^2.$$

Whence, we consider the following setting: $X=(X_1,\ldots,X_p)^t$, $\theta=(\theta_1,\ldots,\theta_p)^t$, $v=\frac{v_0}{m}$ and $n=(m-1)p$.

Given the above and assuming that $Z_s=\frac{||X||^2}{S}$, we may introduce this Baranchick type class, as an opponent class of shrinkage estimators

$$\hat{\theta}_S(X)=\left(1-\frac{\tau(Z_s)}{Z_s}\right)X, \tag{40}$$

where τ is an arbitrary function to be restricted in the sequel.

The following finding cites sufficient restrictions on τ to achieve dominance versus the benchmark estimator X. Thereby, demonstrating that Theorem (6) holds even when v_x is unknown, as it is the case under divergence loss in Mergel *et al* (2008), using similar arguments, indeed:

Theorem 9. *Under the model (1). For $d \geq 3$, assuming:*

1. $0 < \tau(t) < 2(d-2)$ *for all* $t > 0$;
2. $\tau(t)$ *is a differentiable nondecreasing function on* t.

Then, $\hat{\theta}_S(X)$ *improves upon* X *under DPD loss, equivalently under scaled reflected normal loss* $b_0 RN_{\gamma_1}(\theta, \hat{\theta}_S(x))$.

Proof. Assuming these substitutions: $T = \frac{X}{\sigma_x}$, $\eta = \frac{\theta}{\sigma_x}$ and $Z = \frac{S}{v_x}$ (note that Z is independently distributed of T); thanks to (5) the frequentist risk corresponding to $\hat{\theta}_S(X)$ is described as

$$R_\alpha(\theta, \hat{\theta}_S) = b_0 \left[1 - E_\theta \exp\left(- \frac{\left|\left| \left(1 - \frac{Z\tau(||T||^2/Z)}{||T||^2}\right) T - \eta \right|\right|^2}{2\gamma_1} \right) \right], \tag{41}$$

where $T \sim N_d(\eta, I_d)$, $Z \sim \sigma_x (n+1)^{-1} \chi^2_n$ and $\gamma_1 = \frac{\alpha+1}{\alpha}$. Considering a conditional decomposition over a variable $L = \frac{n+2}{2} Z$ in the expectation above (41) and exploiting the independence of T and L, we can rewrite it as follows

$$E\left[E_\theta \exp\left(- \frac{\left|\left| \left(1 - \frac{Z\tau(||T||^2/Z)}{||T||^2}\right) T - \eta \right|\right|^2}{2\gamma_1} \right) |L \right] = E \exp\left(- \frac{\left|\left| \left(1 - \frac{L\tau_0(||T||^2/L)}{||T||^2}\right) T - \eta \right|\right|^2}{2\gamma_1} \right), \tag{42}$$

where $\tau_0(s/l) = \frac{2}{n+2}\tau\left(\frac{(n+2)s}{2l}\right)$ satisfying all conditions in the proof Theorem (6). Now by virtue of Proposition 3, the expectation (42) can also be formulated as:

$$\hat{I}_{\gamma_1} = \left(\frac{\gamma+1}{\gamma}\right)^{-\frac{d}{2}} \sum_{k=0}^{\infty} \frac{e^{-\hat{h}_{\gamma_1}} \hat{h}^k_{\gamma_1}}{k!} \hat{\beta}_{\gamma_1}(k), \tag{43}$$

where $\hat{h}_{\gamma_1} = \left(\frac{\gamma+1}{\gamma}\right)^{-\frac{d}{2}} \frac{||\eta||^2}{2}$. Moreover, given that (10) and via Proposition 2 and Proposition 3, $\hat{\beta}_{\gamma_1}(k)$ is now given by,

$$\hat{\beta}_{\gamma_1}(k) = \int_0^\infty \left(1 - \frac{\tau_0(u)}{2\gamma_1 u}\right)^{2k} \frac{u^{k+\frac{d}{2}-1}}{B\left(k+\frac{d}{2}, \frac{n}{2}\right)} \left(u + 1 + \frac{(\gamma_1+1)}{4\gamma_1^2 u}\tau_0(u) - \frac{\tau_0(u)}{2\gamma_1 u}\right)^{-(k+\frac{d}{2}+\frac{n}{2})} du, \tag{44}$$

with $u = \frac{s}{l}$. It's clear that it suffices again that $\hat{\beta}_{\gamma_1}(k) > 1$, to achieve dominance. Let u_0 be defined as in the proof Theorem 6. Thus, for $u > u_0$ and $\tau_0(u)/u > 2\gamma_1$, by (44) we obtain

$$\begin{aligned}
\hat{\beta}_{\gamma_1}(k) &\geq \int_{u_0}^\infty \left(1 - \frac{\tau_0(u)}{2\gamma_1 u}\right)^{2k} \frac{u^{k+\frac{d}{2}-1}}{B\left(k+\frac{d}{2}, \frac{n}{2}\right)} \left(u + 1 + \frac{(\gamma_1+1)}{4\gamma_1^2 u}\tau_0(u) - \frac{\tau_0(u)}{2\gamma_1 u}\right)^{-(k+\frac{d}{2}+\frac{n}{2})} du \\
&= \int_{u_0}^\infty \frac{(\psi(u))^{k+\frac{d}{2}-1}(1+\psi(u))^{-(k+\frac{d}{2}-\frac{n}{2})}}{B\left(k+\frac{d}{2}, \frac{n}{2}\right)} \frac{\left(1 - \frac{\tau_0(u)}{2\gamma_1 u}\right)^{-(d-2)}}{\left(1 + \frac{\tau_0^2(u)}{4\gamma_1 u}\right)^{-(1+\frac{m}{2})}} du, \qquad (45)
\end{aligned}$$

where

$$\psi(u) = u\left(1 - \frac{\tau_0(u)}{2 \gtrsim_1 u}\right)^2 \Big/ \left(1 + \frac{\tau_0^2(u)}{4\gamma_1 u}\right).$$

Using the same classical inequality we have:

$$\left(1 - \frac{\tau_0(u)}{2\gamma_1 u}\right)^{-(d-2)} \geq \exp\left((d-2)\frac{\tau_0(u)}{2\gamma_1 u}\right)$$

and

$$\left(1 + \frac{\tau_0^2(u)}{4\gamma_1 u}\right)^{-(1+\frac{m}{2})} \geq \exp\left(-(1+\frac{n}{2})\frac{\tau_0^2(u)}{4\gamma_1 u}\right).$$

It implies that

$$\left(1 - \frac{\tau_0(u)}{2\gamma_1 u}\right)^{-(d-2)} \left(1 + \frac{\tau_0^2(u)}{4\gamma_1 u}\right)^{-(1+\frac{m}{2})} \geq \exp\left[-\frac{(n+2)\tau_0(u)}{8\gamma_1 u}\left(\tau_0(u) - 4\frac{d-2}{n+2}\right)\right] > 1, \quad (46)$$

because $0 \leq \tau_0(u) \leq 4\dfrac{d-2}{n+2}$. On the other hand

$$\psi^{'}(u) = \frac{4(2\gamma_1 u - \tau_0(u))}{(4\gamma_1 u + \tau_0^2(u))^2}\left(2\gamma_1 u + \tau_0(u) + \tau_0^2(u) - 2u\tau_0^{'}(u) - u\tau_0(u)\tau_0^{'}(u)\right).$$

As a result

$$\psi^{'}(u) \leq 1 \Leftrightarrow \tau_0^{'2}(u)(2 + \tau_0(u)) + 4u\tau_0^{'}(u)(2\gamma_1 - \tau_0(u)) \geq 0,$$

which is true, according to the assumptions: $u \geq u_0$ and $u \geq \dfrac{\tau_0(u)}{2\gamma_1}$; together with the assumption (2) of the theorem. By and large, setting $H = \psi(u)$, (i.e. $dH = \psi^{'}(u)du$), we obtain for every k in $\mathbb{N}$

$$\hat{\beta}_{\gamma_1}(k) \geq \int_0^{\infty} (1+H)^{k+\frac{d}{2}+\frac{n}{2}} \frac{H^{k+\frac{d}{2}+1}}{B\left(k+\frac{d}{2}, \frac{n}{2}\right)} dH = 1,$$

$$\hat{I}_{\gamma_1} \geq \left(\frac{\gamma_1+1}{\gamma_1}\right)^{-\frac{d}{2}} \sum_{k=0}^{\infty} \frac{e^{-h_{\gamma_1}} h_{\gamma_1}^k}{k!} = \left(\frac{\gamma_1+1}{\gamma_1}\right)^{-\frac{d}{2}},$$

which concludes the proof.

The next subsection is essentially an extension of the findings in the degenerate case $\Sigma = vI_d$, to matrix variate case, for $n \geq d \geq 3$. As will be shown in the main result of this subsection, by rotating the problem in order to yield a new coordinate system representing some variables having certain optimal properties, so as to adopt a somewhat similar technique to the one in the degenerate case.

6.2. The matrix variate case

Let $(X_i)_{0\leq i\leq n}$ be a sequence of i.i.d. d-variate random variables, such that for every i, $X_i \sim N_d(\theta, \Sigma)$, where Σ is an unknown positive definite matrix. In the remainder, we will call two classical estimators respectively of the mean θ and the variance-covariance matrix Σ, namely, $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i \sim N_d(\theta, n^{-1}\Sigma)$ as a benchmark estimator of θ (being: MLE, BEE, UMVUE) and

$$S_2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^t,$$

as a benchmark estimator of Σ (also being: MLE, BEE, UMVUE), where S is independently distributed of $\bar{X}$. The expression of frequentist risk associated to $\bar{X}$ will be equal to the corresponding risk in the degenerate case, since (12) doesn't depend on the variance-covariance matrix, also since $n(\bar{X}-\theta))^t\Sigma^{-1}(\bar{X}-\theta) \sim \chi^2_d$. Let us now consider the general class of shrinkage estimators of θ, given by:

$$\hat{\theta}_*(\bar{X}, S_2) = \left(1 - \frac{\tau(n\bar{X}^t S_2^{-1}\bar{X})}{n\bar{X}^t S_2^{-1}\bar{X}}\right)\bar{X}. \tag{47}$$

Now we present the main result of this subsection:

Theorem 10. *Under the model (1). For $n \geq d \geq 3$, assuming:*

1. $0 < \tau(t) < 2(d-2)(n-1)/(n-d+2)$*, for all* $t > 0$*;*
2. $\tau(t)$ *is a differentiable nondecreasing function on* t*.*

Then, $\hat{\theta}_*(\bar{X}, S_2)$ *improves upon* $\bar{X}$ *under DPD loss (4), equivalently under scaled reflected normal loss* $b_0 RN_{\gamma_1}(\theta, \hat{\theta}_{S_2}(x))$.

Proof. It suffices to rotate the problem, in order to exploit Theorem 9, using a suitable orthogonal similarity transformation of the sequence $(X_i)_{0\leq i\leq n}$ (with an incremental or a decremental algorithm, appropriate to the structure of $\bar{X}$), denoted $(\mathbf{G}_i)_{0\leq i\leq n}$, and defined as:

$$\left\{\begin{array}{ll} \mathbf{G}_1 & = \frac{\Sigma^{-\frac{d}{2}}}{\sqrt{2}}(X_2 + X_1) \\ \mathbf{G}_2 & = \frac{\Sigma^{-\frac{d}{2}}}{\sqrt{6}}(2X_3 + X_2 + X_1) \\ & \dots \\ \mathbf{G}_{n-1} & = \frac{\Sigma^{-\frac{d}{2}}}{\sqrt{n(n-1)}}((n-1)X_n + X_{n-1} + \ldots + 2X_3 + X_2 + X_1) \\ \mathbf{G}_n & = \frac{\Sigma^{-\frac{d}{2}}}{\sqrt{n}}\sum_{i=1}^{n} X_i = \sqrt{n}\Sigma^{-\frac{d}{2}}\bar{X} \end{array}\right\},$$

where $(\mathbf{G}_1, \ldots, \mathbf{G}_n)$ forms an orthogonal algebraic base as a consequence. Consequently, $(\mathbf{G}_i)_{0\leq i\leq n}$ are mutually independent random vectors (i.i.d), in addition, $\mathbf{G}_1 \sim N_d(0, I_d)$, ..., $\mathbf{G}_n \sim N_d(\sqrt{n}\Sigma^{-\frac{d}{2}}\theta, I_d)$ then (47) becomes:

$$\hat{\theta}_*(\bar{X}, S) = \left(1 - \frac{\tau((n-1)\mathbf{G}_n^t\left(\sum_{i=1}^{n-1}\mathbf{G}_i\mathbf{G}_i^t\right)^{-1}\mathbf{G}_n)}{(n-1)\mathbf{G}_n^t\left(\sum_{i=1}^{n-1}\mathbf{G}_i\mathbf{G}_i^t\right)^{-1}\mathbf{G}_n}\right)\bar{X}. \tag{48}$$

As shown in (Muirhead (2005), p. 27), we have

$$\mathbf{G}_n^t \left(\sum_{i=1}^{n-1} \mathbf{G}_i \mathbf{G}_i^t\right)^{-1} \mathbf{G}_n \overset{d}{=} \frac{||\mathbf{G}_n||^2}{W},$$

where, $W \sim \chi^2_{n-d}$, and is independently distributed of $\mathbf{G}_n$.

Now the frequentist risk associated to (48) under DPD loss (4), is described by

$$R_\alpha(\theta, \hat{\theta}_*(\bar{X}, S)) = b_0 \left[1 - E_\theta \exp\left(-\frac{\left|\left|\left(1 - \frac{\tilde{R}\tau_0(||\mathbf{G}_n||^2/\tilde{R})}{||\mathbf{G}_n||^2}\right)\mathbf{G}_n - \eta\right|\right|^2}{2\gamma_1}\right)\right], \tag{49}$$

where $\tilde{R} = W/2$ and $\tau_0(u)(u) = \frac{2}{n-1}\tau\left(\frac{2}{n-1}\right)$. In conclusion, gathering (49) with the assumptions (1) and (2), an appeal Theorem 9 terminates the proof.

7. Appendix

Proof of Theorem 1. Let $\pi_n(\theta)$ be a class of priors on θ, such that $\pi_n(\theta) = v_n^{-\frac{d}{2}}\phi\left(\frac{\theta}{\sqrt{v_n}}\right)$, where $v_n \to \infty$ as $n \to \infty$, then the corresponding BE of θ will be of the form $\hat{\theta}_{\pi_n}(X) = (1 - A_n)X$, with $A_n = v_x/(v_x + v_n)$. Consequently, the Bayes risk associated to $\hat{\theta}_{\pi_n}$ is given by

$$\begin{aligned} r_{\pi_n}(\theta, \hat{\theta}_{\pi_n}) &= \int R_\alpha(\theta, \hat{\theta}_{\pi_n})\pi_n(\theta)d\theta \\ &= \int\int D_P(\theta, \hat{\theta}_{\pi_n})p(x|\theta)dx\pi_n(\theta)d\theta. \end{aligned}$$

After a direct application of Lemma 1, we obtain

$$r_{\pi_n}(\theta, \hat{\theta}_{\pi_n}) = b_0\left[1 - \left(\frac{\alpha+1}{\alpha}\frac{v_n A_n^2}{v_y} + \frac{\alpha+1}{\alpha}r(1 - A_n)^2 + 1\right)^{-\frac{d}{2}}\right].$$

Since $v_n A_n^2 \to 0$ and $(1 - A_n)^2 \to 1$ as $n \to \infty$, thus according to (12), $r_{\pi_n}(\theta, \hat{\theta}_{\pi_n}) \to R_\alpha(\theta, \hat{\theta}_*)$. Thus, the $\pi_n(\theta)$'s is a least favourable sequence of priors associated to the minimaxity of X; which concludes the proof.

Proof of Theorem 2. We consider the same class of priors as in the proof of Theorem 1, then we give the corresponding BPDE denoted by $\hat{q}_{\pi_n}(y|x)$:

$$\hat{q}_{\pi_n}(y|x) = (v_y + \alpha v_n A_n)^{-\frac{d}{2}}\phi\left(\frac{y - \hat{\theta}_{\pi_n}(x)}{\sqrt{v_y + \alpha v_n A_n}}\right).$$

As a consequence, we give the Bayes risk associated to $\hat{q}_{\pi_n}(y|x)$:

$$\begin{aligned} r_{\pi_n}(q, \hat{q}_{\pi_n}) &= \int R_\alpha(q, \hat{q}_{\pi_n})\pi_n(\theta)d\theta \\ &= \int\int D_P(q, \hat{q}_{\pi_n})p(x|\theta)dx\pi_n(\theta)d\theta. \end{aligned}$$

A direct application of Lemma 1, gives

$$r_{\pi_n}(q,\hat{q}_{\pi_n}) = b\left[1+\frac{1}{\alpha}\left(\frac{\alpha v_n A_n}{v_y}+1\right)^{-\frac{d\alpha}{2}} - \left(\frac{\alpha+1}{\alpha}\right)^{1+\frac{d}{2}} \times \left(\frac{v_n A_n^2}{v_x}+(1-A_n^2)+\frac{\alpha v_n A_n}{v_y}+1\right)^{-\frac{d}{2}}\right]. \quad (50)$$

Since $v_n A_n \to v_x$, $v_n A_n^2 \to 0$ and $(1-A_n)^2 \to 0$, as $n \to \infty$, together with (13) implies that

$$r_{\pi_n}(q,\hat{q}_{\pi_n}) \underset{n\to\infty}{\to} b\left[1+\frac{1}{\alpha}(\alpha r+1)^{-\frac{\alpha d}{2}} - \left(\frac{\alpha+1}{\alpha}\right)(\alpha r+1)^{-\frac{d}{2}}\right].$$

Whence, $r_{\pi_n}(q,\hat{q}_{\pi_n}) \to R_\alpha(\theta,\hat{q}_{bee})$; which concludes the proof.

Proof of Theorem 3. According to corollary 1, $\hat{q}_{bee}(y|x) \propto k_*^{1/\alpha}(y,x)$, where $k_*(y,x) = \int q^\alpha(y|\theta)p(x|\theta)d\theta$. As indicated, we will employ a Girshick and Savage technique, which consists of taking a sequence of priors of the form $\pi_m(\theta) = m^{-d}\delta_{S_m}(\theta)$, where for $m \in \mathbb{N}^*$

$$S_m = \{\theta : |\theta_i| < m/2; i = 1,\dots,d\}$$

and δ is the Dirac function. Notice that $S_m \to \mathbb{R}^d$, as $m \to \infty$. Then we can give the corresponding BPDE

$$\hat{q}_{\pi_m}(y|x) \propto k_{S_m}^{1/\alpha}(y,x),$$

where $k_{S_m}(y,x) = \int_{S_m} q^\alpha(y|u)p(x|u)du$, since the corresponding posterior density is given by

$$\pi_m(\theta|x) = \frac{p(x|u)\delta_{S_m}(u)}{\int_{S_m} p(x|u)du}.$$

It follows that the corresponding Bayes risk verifies

$$r_{\pi_m}(q,\hat{q}_{\pi_m}) = m^{-d}\int_{S_m}\int D_P(q,\hat{q}_{\pi_m})p(x|u)dxdu \le R_\alpha(q,\hat{q}_{gbee}) = R_0,$$

which stems from the Bayesness of $\hat{q}_{\pi_m}(y|x)$ (i.e. translating the privilege of having a prior knowledge ($\pi_m(\theta)$) over having a noninformative prior). Thereby, it suffices to demonstrate that $\liminf_m r_{\pi_m}(q,\hat{q}_{\pi_m}) \ge R_0$; for that by making the following substitutions: ($z = x-\theta$), ($t = u-\theta$) and ($\nu_i = \theta_i/m$), and by denoting $\hat{q}^*_{\pi_m}(y|x) \propto \left(\int_{t+m\nu\in S_m} \hat{q}^\alpha(y|t)(x|t)dt\right)^{1/\alpha}$, we obtain

$$\begin{aligned} r_{\pi_m}(q,\hat{q}_{\pi_m}) &= \int_{\{\nu_i<1/2\}}\int D_P(q,\hat{q}^*_{\pi_m})p(x|\theta)dxd\theta \\ &\ge \int_{\{\nu_i<(1-\epsilon)/2\}}\int D_P(q,\hat{q}^*_{\pi_m})p(x|\theta)dxd\theta. \end{aligned}$$

On the other hand, for $|\nu_i| < (1-\epsilon)/2$, again we notice this inclusion $\{|t_i| < m\epsilon/2; i = 1, \ldots, d\} \subset \{t + m\nu \in S_m\}$. This ensures that $\hat{q}^*_{\pi_m}(y|x) \to \hat{q}_{bee}(y|x)$, as $m \to \infty$, along with an appeal to Fatou's lemma leads to

$$\begin{aligned}
\liminf_m r_{\pi_m}(q, \hat{q}_{\pi_m}) &\geq \liminf_m \int_{\{\nu_i<(1-\epsilon)/2\}} \int D_P(q, \hat{q}^*_{\pi_m}) p(x|\theta) dx d\theta \\
&\geq \int_{\{\nu_i<(1-\epsilon)/2\}} \int \liminf_m D_P(q, \hat{q}^*_{\pi_m}) p(x|\theta) dx d\theta \\
&= (1-\epsilon)^d R_\alpha(q, \hat{q}_{bee}) = (1-\epsilon)^d R_0.
\end{aligned}$$

By the arbitrariness of ϵ, one gets that $\liminf_m r_{\pi_m}(q, \hat{q}_{\pi_m}) \geq R_0$; thus establishing the minimaxity of $\hat{q}_{gbee}(y|x)$.

Proof of Theorem 4. Suppose that X is inadmissible, then there exist a dominant estimator $\hat{\theta}_0$, such that for all $\theta \in \mathbb{R}$: $R_\alpha(\theta, \hat{\theta}_0) \leq R_\alpha(\theta, X)$, with strict inequality for a certain $\bar{\theta}_0$. We denote Δ the risk difference of the risks respectively associated X and $\hat{\theta}_0$, $\Delta_0 = R_\alpha(\bar{\theta}_0, \hat{\theta}_0) - R_\alpha(\bar{\theta}_0, X) > 0$. By continuity of the risk function $\theta \mapsto R_\alpha(\theta, .)$, there exist a neighborhood $[\bar{\theta}_0 - \epsilon, \bar{\theta}_0 + \epsilon]$ of $\bar{\theta}_0$, such that $\Delta_*(\theta) = R_\alpha(\theta, \hat{\theta}_0) - R_\alpha(\theta, X) > \Delta/2$, for all $\theta \in [\bar{\theta}_0 - \epsilon, \bar{\theta}_0 + \epsilon]$. On the other hand, we consider a centered version of the same class of priors as in the last section, $\pi_n(\theta) \sim N_d(0, v_n)$, if we denote $\Delta_{\pi_n} = r_{\pi_n}(\theta, X) - r_{\pi_n}(\theta, \hat{\theta}_0)$ and ϕ the pdf of a standard gaussian $N(0,1)$, then

$$\begin{aligned}
\Delta^\circ_{\pi_n} &= \int_{\mathbb{R}} \Delta_*(\theta) \pi_n(\theta) d\theta \\
&\geq \int_{\bar{\theta}_0-\epsilon}^{\bar{\theta}_0+\epsilon} \Delta_*(\theta) \pi_n(\theta) d\theta \\
&\geq (\Delta_0/2) \int_{\bar{\theta}_0-\epsilon}^{\bar{\theta}_0+\epsilon} v_n^{-d/2} \phi\left(\frac{\theta}{\sqrt{v_n}}\right) d\theta \\
&\geq (\epsilon\Delta_0/4)(2\pi v_n)^{-1/2}. \qquad (51)
\end{aligned}$$

Otherwise, if we denote $\Delta^*_{\pi_n} = r_{\pi_n}(\theta, X) - r_{\pi_n}(\theta, \hat{\theta}_{\pi_n})$, $b_* = (2\pi v_y)^{-1/2} \alpha^{-1} \sqrt{\frac{\alpha+1}{v_y}}$, then

$$\Delta^*_{\pi_n} = b_* \left\{ \left[\frac{\alpha+1}{\alpha} \left(\frac{v_n A_n^2}{v_x} + (1-A_n)^2 \right) + 1 \right]^{-1/2} - \left[\frac{\alpha}{\alpha+1} + 1 \right]^{-1/2} \right\} \qquad (52)$$

Let K be a nonnegative generic constant, together with (51) and (52), for large enough n ($n \geq N$, N positive integer), we obtain that :

$$\frac{\Delta^\circ_{\pi_n}}{\Delta^*_{\pi_n}} \geq \frac{1}{4} K (2\pi v_n)^{-1/2} (\Delta^*_{\pi_n})^{-1} \to \infty, \text{ as } n \to \infty, \qquad (53)$$

Since $(v_n)^{-\frac{1}{2}}(\Delta^*_{\pi_n})^{-1} \to \infty$, when $v_n \to \infty$. Therefore, for $n \geq N$, $\Delta^\circ_{\pi_n} > \Delta^*_{\pi_n}$, i.e. $r_{\pi_n}(\theta, \hat{\theta}_{\pi_n}) > r_{\pi_n}(\theta, \hat{\theta}_0)$, which breaks the Bayesness of $\hat{\theta}_{\pi_n}$; which concludes the proof.

Proof of Theorem 5. As for the predictive density estimation, we suppose that there exist an improving predictive density estimator denoted by $\hat{q}_\star(y|x)$ over $\hat{q}_{bee}(y|x)$.

Likewise, if we set $\Delta_{\pi_n} = r_{\pi_n}(\theta, \hat{\theta}_U(y|x)) - r_{\pi_n}(\theta, \hat{q}_\star(y|x))$ and $\Delta_0 = R_\alpha(\bar{\theta}_0, \hat{\theta}_0) - R_\alpha(\bar{\theta}_0, X) > 0$, then the inequality (51) holds for the new Δ_{π_n}. On the other hand, if we denote $\Delta^*_{\pi_n} = r_{\pi_n}(\theta, \hat{q}_{bee}(y|x)) - r_{\pi_n}(\theta, \hat{q}_\star(y|x))$ and $b_* = ((2\pi v_y)^\alpha(\alpha+1))^{-1/2}$, according to (50) $\Delta^*_{\pi_n}$ states as

$$\begin{aligned}\Delta^*_{\pi_n} &= b_*\Bigg\{\frac{1}{\alpha}\left[\left(\frac{\alpha v_n A_n}{v_y}+1\right)^{-\frac{\alpha}{2}} - (\alpha r+1)^{-\alpha/2}\right] - \left(\frac{\alpha+1}{\alpha}\right)^{\frac{3}{2}}\left(\frac{\alpha v_n A_n}{v_y}+1\right)^{\frac{1-\alpha}{2}} \\ &\quad \times \left(\frac{\alpha v_n A_n}{v_y}+\frac{\alpha+1}{\alpha}+1\right)^{-\frac{1}{2}}\left(\frac{v_n A_n^2}{v_y}+r\left(\frac{A_n}{1-A_n}\right)^2+\frac{\alpha v_n A_n}{v_y}+1\right)^{-\frac{1}{2}} \\ &\quad - \frac{(\alpha r+1)^{-\alpha 1/2}}{\alpha} - \frac{\alpha+1}{\alpha}\left(\frac{\alpha r}{\alpha+1}+1\right)^{-1/2}\Bigg\}.\end{aligned}$$

Let $K^{'}$ be a positive generic constant, and we have

$$\frac{\Delta^\circ_{\pi_n}}{\Delta^*_{\pi_n}} \geq \frac{1}{4}K(2\pi v_n)^{-1/2}(\Delta^*_{\pi_n})^{-1} \to \infty, \; \textit{as } n \to \infty. \tag{54}$$

Since $(v_n)^{-\frac{1}{2}}(\Delta^*_{\pi_n})^{-1} \to \infty$, when $v_n \to \infty$. By a similar argument to the one in Theorem (4), the proof is concluded.

References

Abramovitz M. and Stegun I.A., 1966. Handbook of mathematical functions, *American Journal of Physics*.

Aitchison J., 1975. Goodness of prediction fit, *Biometrika*, 62 (3), 547–554.

Alam K., 1973. A family of admissible minimax estiamtors of the mean of a multivariate normal distribution, *The Annals of Statistics*, 1 (3), 517–525.

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

Basu A., 1998. Robust and efficient estimation by minimizing a density power divergence, *Biometrika*, 85, 549–559.

Berger J. O., 1985. Statistical Decision Theory and Bayesian Analysis, 2nd Edition, *New York: Springer-Verlag*.

Blyth. C. R., 1951. On minimax statistical decision procedures and their inadmissibility, *Annals of Mathematical Statistics*, 8, 279–284.

Brown L.D. and Fox M., 1974. Admissibility of procedures in two-dimensional location parameter problems, *The Annals of Statistics*, 2, 248–266.

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

Corcuera J. and Giummolè F., 1999. A generalized bayes rule for prediction, *Scand. J. Stat.*, 26 (2), 265–279.

Cressie N. and T. R. C. Read T.R.C (1984). Multinomial Goodnessof-Fit Tests. *Journal of the Royal Statistical Society*. Series B (Methodological), 46, 3, 440-464.

Gatsonis C., 1984. Deriving posterior distributionsfor a location parameter: A decision theoretic approach, *The Annals of Statistics*, 3, 958–970.

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

George E. I., Liang F. and Xu X., 2012. From minimax shrinkage estimation to minimax shrinkage prediction, *Statistical Science*, 27 (1), 82–94.

Ghosh A., Harris I. R., Maji A., Basu A., Pardo L., 2017. A generalized divergence for statistical inference, *Bernoulli*, 4A (23), 2746–2783.

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.

Hansen B. E.,2015. Shrinkage efficiency bounds, *Econometric Theory*, 31, 860–440.

Jimenez R. and Shao Y. 2001. On robustness and efficiency of minimum divergence estimators. *Test*, 10(2), 241-248.

Kubokawa T., Marchand É. and Strawderman W. E., On predictive density estimation for location families under integrated squared error loss, *J. Multivar. Anal.*

Kullback S. and Leibler R. A., 1951. On information and sufficiency, *The Annals of Mathematical Statistics*, 22 (1), 79–86.

Lindley D.V., 1962. Discussion of professor stein's paper'confidence sets for the mean of a multivariate distribution, *Journal of Royal Statistical Society B*, (24), 445, 265–296.

Mergel. V., Ghosh M. and Datta. G.S., 2008. Estimation, prediction and the stein phenomenon under divergence loss, *Journal of Multivariate Analysis*, 9 (99), 1941–1961.

Muirhead R., 2005. Aspects of multivariate statistical theory, *John Wiley and Sons, Inc, New York.*

Ommane Y. and Ouassou I. 2019. Minimaxity and Admissibility of Predictive Density Estimators Under S-Hellinger Distances, *Afrika Statistika*, 14, 2, 1951-1976.

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

Stein C., 1955. A necessary and sufficient condition for admissibility, *Annals of Mathematical Statistics*, 26, 518–526.

Stein C., 1981. Estimation of the mean of a multivariate normal distribution, *The Annals of Statistics*, 9, 1135–1151.

Strawderman W.E., 1971. Proper bayes minimax estimators of the multivariate normal mean, *The Annals of Statistics*, 42 (1), 385–388.