Estimation, prediction and the Stein phenomenon under divergence loss

Malay Ghosh*, Victor Mergel, Gauri Sankar Datta

Department of Statistics, University of Florida, Gainesville, FL 32611-8545, United States
Department of Statistics, University of Georgia, Athens, GA 30602-1952, United States

Received 15 March 2007
Available online 8 February 2008

Abstract

We consider two problems: (1) estimate a normal mean under a general divergence loss introduced in [S. Amari, Differential geometry of curved exponential families — curvatures and information loss, Ann. Statist. 10 (1982) 357–387] and [N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, J. Roy. Statist. Soc. Ser. B. 46 (1984) 440–464] and (2) find a predictive density of a new observation drawn independently of observations sampled from a normal distribution with the same mean but possibly with a different variance under the same loss. The general divergence loss includes as special cases both the Kullback–Leibler and Bhattacharyya–Hellinger losses. The sample mean, which is a Bayes estimator of the population mean under this loss and the improper uniform prior, is shown to be minimax in any arbitrary dimension. A counterpart of this result for predictive density is also proved in any arbitrary dimension. The admissibility of these rules holds in one dimension, and we conjecture that the result is true in two dimensions as well. However, the general Baranchick [A.J. Baranchick, a family of minimax estimators of the mean of a multivariate normal distribution, Ann. Math. Statist. 41 (1970) 642–645] class of estimators, which includes the James–Stein estimator and the Strawderman [W.E. Strawderman, Proper Bayes minimax estimators of the multivariate normal mean, Ann. Math. Statist. 42 (1971) 385–388] class of estimators, dominates the sample mean under this loss and the improper uniform prior, is shown to be minimax in any arbitrary dimension. A counterpart of this result for predictive density is also proved in any arbitrary dimension. The admissibility of these rules holds in one dimension, and we conjecture that the result is true in two dimensions as well. However, the general Baranchick [A.J. Baranchick, a family of minimax estimators of the mean of a multivariate normal distribution, Ann. Math. Statist. 41 (1970) 642–645] class of estimators, which includes the James–Stein estimator and the Strawderman [W.E. Strawderman, Proper Bayes minimax estimators of the multivariate normal mean, Ann. Math. Statist. 42 (1971) 385–388] class of estimators, dominates the sample mean in three or higher dimensions for the estimation problem. An analogous class of predictive densities is defined and any member of this class is shown to dominate the predictive density corresponding to a uniform prior in three or higher dimensions. For the prediction problem, in the special case of Kullback–Leibler loss, our results complement to a certain extent some of the recent important work of Komaki [F. Komaki, A shrinkage predictive distribution for multivariate normal observations, Biometrika 88 (2001) 859–864] and George, Liang and Xu [E.I. George, F. Liang, X. Xu, Improved minimax predictive densities under Kullbak–Leibler loss, Ann. Statist. 34 (2006) 78–92].

* Corresponding address: Department of Statistics, University of Florida, 103 Griffin-Floyd Hall, P.O. Box 118545, Gainesville, FL 32611-8545, United States.
E-mail address: ghoshm@stat.ufl.edu (M. Ghosh).
While our proposed approach produces a general class of predictive densities (not necessarily Bayes, but not excluding Bayes predictors) dominating the predictive density under a uniform prior. We show also that various modifications of the James–Stein estimator continue to dominate the sample mean, and by the duality of estimation and predictive density results which we will show, similar results continue to hold for the prediction problem as well.

AMS 2000 subject classifications: 62C15; 62C20; 62C12

Keywords: Admissibility; Baranchick class; Bhattacharyya–Hellinger loss; Empirical Bayes; Kullback–Leibler loss; Minimaxity

1. Introduction

For estimating the normal mean, the classical estimator, namely the sample mean, meets many important frequentist desiderata. It is the UMVUE, the MLE, and the best equivariant estimator under translations of the sample space. Also, Blyth [7] proved minimaxity of this estimator in one dimension for a general class of losses, including but not limited to the squared error loss. He also proved admissibility of the estimator in one dimension while Stein [26] proved admissibility in two dimensions.

In his seminal 1956 paper, Stein, came up with the surprising result that the sample mean is an inadmissible estimator of normal population mean in dimensions three or more under squared error loss. Later James and Stein [21] provided an explicit estimator dominating the sample mean under the same loss. Subsequently, a very general class of minimax estimators dominating the sample mean was provided by Baranchick [3,4]. Strawderman [25] and Faith [16] provided a general class of proper Bayes minimax estimators dominating the sample mean. Brandwein and Strawderman [8] showed that the Baranchick class of estimators dominated the sample mean for spherically symmetric distributions under some stronger conditions than those required in the normal case.

A trivial extension of the above results is that the sample mean is also a minimax predictor of an observation drawn independently from the same normal distribution under squared error loss in any arbitrary dimension. Its admissibility continues to hold in one and two dimensions. However, in three and higher dimensions, the sample mean is dominated by the James–Stein predictor and its variants as mentioned in the previous paragraph.

The original results of Stein have been generalized in a variety of ways. Stein [27] proved admissibility of the Pitman estimator in one or two dimensions under squared error loss, while its inadmissibility in three or higher dimensions was proven in James and Stein [21]. Brown [9] proved the inadmissibility of the Pitman estimator under a wide class of losses for three or higher dimensions. Later, Brown [10] provided a necessary and sufficient condition for admissibility of the sample mean under losses with bounded risk.

While squared error loss has dominated most of the research related to estimation of the multivariate normal mean, there are other losses which are also of interest. Indeed, as pointed out by Robert [24], often it is natural to use losses which compare directly the densities \( f(\cdot|\theta) \) and \( f(\cdot|a) \), where \( \theta \) is a true parameter. Robert refers to such losses as “intrinsic losses”.

The two most well-used divergence measures between two distributions are the Kullback–Leibler (KL) or the entropy distance and the Bhattacharyya–Hellinger (BH) distance [19,5]. The KL distance has received more prominence in statistics literature than the Hellinger
distance. However if \( X_i \sim N(\theta_i, \sigma^2) \) and independent for \( i = 1, \ldots, p \), writing \( X = (X_1, \ldots, X_p)^T \) and denoting its pdf by \( f(x|\theta) \), the KL distance between \( f(x|\theta) \) and \( f(x|a) \) is given by

\[
E_\theta \left[ \log \frac{f(X|\theta)}{f(X|a)} \right] = \frac{1}{2\sigma^2} \| \theta - a \|^2,
\]

where \( \| \cdot \| \) denotes the Euclidean norm. Hence, point estimation based on the KL loss is tantamount to squared error loss for known \( \sigma^2 \). The other loss, namely the BH loss is given by

\[
L_{BH}(\theta, a) = \frac{1}{2} \int \left[ f^{1/2}(x|\theta) - f^{1/2}(x|a) \right]^2 \, dx = 1 - \exp \left\{ -\frac{1}{8\sigma^2} \| \theta - a \|^2 \right\}, \tag{1.1}
\]

where the second equality follows from Lemma 2.2 to be proved later in Section 2. This loss is different from the squared error loss and is not convex.

Both KL and BH losses are special cases of a more general divergence loss considered by many authors. Among others, we refer to Amari [2] and Cressie and Read [13]. This loss is given by

\[
L_\beta(\theta, a) = \frac{1 - \int f^{1-\beta}(x|\theta) f^\beta(x|a) \, dx}{\beta(1 - \beta)} = \frac{1 - \exp \left\{ -\frac{\beta(1-\beta)}{2\sigma^2} \| a - \theta \|^2 \right\}}{\beta(1 - \beta)}, \tag{1.2}
\]

which is again a consequence of Lemma 2.2. This above loss is to be interpreted as its limit when \( \beta \to 0 \) or \( \beta \to 1 \). In general, one is primarily interested in the case when \( \beta \to 0 \), i.e. \( L_{KL}(\theta, a) = E_\theta \log \frac{f_\theta(X)}{f_a(X)} \), namely the KL loss. For \( \beta = 1/2 \), the divergence loss is 4 times the BH loss given in (1.1). Throughout this article, we will perform the calculations with \( \beta \in (0, 1) \), and pass on to the endpoints only in the limit when needed.

Recently Komaki [22] and George, Liang and Xu [18] have considered improved minimax predictive densities under KL loss. They have developed various shrinkage versions of predictive densities which dominate under this loss the Bayes predictive density under the uniform prior for a future observation conditionally independent of the sampled observations. George et al. [18] have also explored various interesting duality results between multivariate estimation and prediction in the normal problem.

The objective of this paper is to consider simultaneously estimation of the normal mean and predictive density of a future observation conditionally independent of the sampled observations under general divergence loss. In this way, we are able to obtain some unifying results covering both estimation and prediction for a wide class of losses including important KL and BH losses. Prediction results for KL loss complement those of Komaki [22] and George et al. [18]. While the present results produce a general class of predictors (not necessarily Bayes) dominating normal predictive density under the uniform prior, Komaki [22] and George et al. [18] considered a class of Bayes predictive densities, and found sufficient conditions under which a similar dominance holds.

The organization of the remaining sections is as follows. We introduce the problems in Section 2, and prove some general results useful for the rest of the paper. In particular, we prove two lemmas in this section. Lemma 2.1 provides a general expression (not necessarily restricted to the normal case) of the predictive density under a general divergence loss, while Lemma 2.2 is used repeatedly for finding closed-form expressions of losses associated with various estimators and predictors.
of the population mean in any arbitrary dimension, and its counterpart, namely, the Bayes (under the uniform prior) predictive density of a future observation conditionally independent of the sampled ones, is a minimax rule in any arbitrary dimension. The admissibility results for both estimation and predictive density in one dimension are proved in Section 4. The theorems related to inadmissibility of the sample mean for estimating the population mean in dimensions three or more are stated in Section 5. The inadmissibility of its predictive density counterpart is also given in this section. In particular it is shown that the Baranchick class of estimators dominates the sample mean under essentially the same conditions as required under squared error loss. Predictive densities defined in an analogous fashion also dominate the Bayes predictive density under a uniform prior. Proofs of these results are presented in the Appendix.

Section 6 gives a variation of the above results and shows in particular that Lindley’s [23] modification of the James–Stein estimator also dominates the sample mean, and shows extensions of these results to regression problems. Corresponding results for predictive density are also mentioned. Section 7 contains some remarks regarding possible extensions of our results.

In view of Brown’s [9] result, inadmissibility of the sample mean under the general divergence loss is not that surprising, at least for estimation. However, the fact that the Baranchick class of estimators continues to dominate the sample mean under general divergence loss seems quite interesting. Second, the method of proof showing the dominance of the proposed class of estimators and predictors is quite non-standard. For squared error loss, proof of dominance either requires application of Stein’s identity or explicit evaluation of the risk based on the properties of non-central chisquares. For the prediction problem under KL loss, the basic approach of George et al. involves Stein’s identity and a use of the heat equation. In contrast, our method of proof requires evaluation of the risk for the Baranchick class of estimators and the analogous class of predictors; the final dominance proof follows from an integral inequality which seems to be quite new. The second interesting feature of our paper is that both estimation and predictive density results are unified through two key results, namely Theorems 5.1 and 5.2. The general technique for proving these theorems seems to be quite novel in its own right.

Eaton [14] considered admissibility of formal Bayes rules for estimation of bounded measurable functions of both the parameters and the data under quadratic loss. He considered also a wide class of prediction problems when the loss is a quadratic measure of the distance between two distributions. The general divergence loss as considered in this paper also measures the distance between two distributions, but it is quite different from the quadratic loss. More importantly, our main focus is to demonstrate the loss robustness of the Baranchick class of estimators and the corresponding predictors, an emphasis quite different from that of Eaton.

2. Some preliminary results

Let $X$ and $Y$ be conditionally independent given $\theta$ with corresponding pdf’s $p(x|\theta)$ and $p(y|\theta)$. We begin with a general expression for the predictive density of $Y$ based on $X$ under the divergence loss and a prior pdf $\pi(\theta)$, possibly improper. Under the KL loss and the prior pdf $\pi(\theta)$, the predictive density of $Y$ is given by

$$
\pi_{KL}(y|x) = \int p(y|\theta)\pi(\theta|x) \, d\theta,
$$
where $\pi(\theta|x)$ is the posterior of $\theta$ based on $X = x$ [1]. The predictive density is proper if and only if the posterior pdf is proper. We now provide a similar result based on the general divergence loss which includes the previous result of Aitchison as a special case when $\beta \to 0$. This result is also available in [12] with a somewhat different proof.

**Lemma 2.1.** Under divergence loss and the prior $\pi$, the Bayes predictive density of $Y$ is given by

$$
\pi_D(y|x) = k^{\frac{1}{1-\beta}}(y,x) \int k^{\frac{1}{1-\beta}}(y,x)dy,
$$

where $k(y,x) = \int p^{1-\beta}(y|\theta) \pi(\theta|x) d\theta$.

**Proof of Lemma 2.1.** Under divergence loss, the posterior risk of predicting $p(y|\theta)$, by a pdf $p(y|x)$, is $\beta^{-1}(1-\beta)^{-1}$ times

$$
1 - \int \left[ \int p^{1-\beta}(y|\theta)p^\beta(y|x)dy \right] \pi(\theta|x)d\theta
= 1 - \int p^\beta(y|x) \left\{ \int p^{1-\beta}(y|\theta) \pi(\theta|x) d\theta \right\} dy
= 1 - \int k(y,x) p^\beta(y|x)dy.
$$

An application of Holder’s inequality now shows that the integral in (2.2) is maximized at $p(y|x) \propto k^{\frac{1}{1-\beta}}(y,x)$. Again by the same inequality, the denominator of (2.1) is finite provided the posterior pdf is proper. This leads to the result noting that $\pi_D(y|x)$ has to be a pdf. □

The next lemma, to be used repeatedly in the sequel, provides an expression for the integral of the product of two normal densities each raised to a certain power. A proof of this result is given in the Appendix.

**Lemma 2.2.** Let $N_p(x|\mu, \Sigma)$ denote the pdf of a $p$-variate normal random variable with mean vector $\mu$ and positive definite variance–covariance matrix $\Sigma$. Then for $\alpha_1 > 0, \alpha_2 > 0$,

$$
\int [N_p(x|\mu_1, \Sigma_1)]^{\alpha_1} [N_p(x|\mu_2, \Sigma_2)]^{\alpha_2} dx
= (2\pi)^{\frac{p}{2}(1-\alpha_1-\alpha_2)} |\Sigma_1|^{\frac{1}{2}(1-\alpha_1)} |\Sigma_2|^{\frac{1}{2}(1-\alpha_2)} |\alpha_1 \Sigma_2 + \alpha_2 \Sigma_1|^{-\frac{1}{2}}
\times \exp \left[ -\frac{\alpha_1 \alpha_2}{2} (\mu_1 - \mu_2)^T (\alpha_1 \Sigma_2 + \alpha_2 \Sigma_1)^{-1} (\mu_1 - \mu_2) \right].
$$

The above results are now used to obtain the Bayes estimator of $\theta$ and the Bayesian predictive density of a future $Y \sim N(\theta, \sigma_\theta^2 I_p)$ under the general divergence loss and the $N(\mu, A I_p)$ prior for $\theta$. We continue to assume that conditional on $\theta, X \sim N(\theta, \sigma_\theta^2 I_p)$, where $\sigma_\theta^2 > 0$ is known. Then we have the following result.

**Lemma 2.3.** Under the loss given in (1.2), the Bayes estimator of $\theta$ is $(1 - B)X + B\mu$, and the Bayes predictive density of $Y$ given $X$ is $N(((1 - B)X + B\mu, (\sigma_\theta^2 (1 - B)(1 - \beta) + \sigma_\theta^2)I_p)$. 

Proof of Lemma 2.3. The Bayes estimator of $\theta$ is obtained by minimizing $1 - \int \exp[-\frac{\beta(1-\beta)}{2\sigma_x^2}\|\theta - a\|^2]N(\theta|(1 - B)X + B\mu, \sigma_x^2(1 - B)I_p)\,d\theta$ with respect to $a$, where $B = \sigma_x^2(\sigma_x^2 + A)^{-1}$. By Lemma 2.2,

$$\int \exp \left[-\frac{\beta(1-\beta)}{2\sigma_x^2}\|\theta - a\|^2\right]N(\theta|(1 - B)X + B\mu, \sigma_x^2(1 - B)I_p)\,d\theta$$

$$= \left(\frac{2\pi\sigma_x^2}{\beta(1-\beta)}\right)^{p/2} \int N(\theta|a, \sigma_x^2\beta^{-1}(1 - B)^{-1}I_p)N(\theta|(1 - B)X + B\mu, \sigma_x^2(1 - B)I_p)\,d\theta$$

$$+ B\mu, \sigma_x^2(1 - B)I_p)\,d\theta \propto \exp \left[-\frac{\|a - (1 - B)X + B\mu\|^2}{2\sigma_x^2(\beta^{-1}(1 - B)^{-1} + 1 - B)}\right]$$

(2.4)

which is maximized with respect to $a$ at $(1 - B)X + B\mu$. Hence, the Bayes estimator of $\theta$ under the $N(\mu, AI_p)$ prior and the general divergence loss is $(1 - B)X + B\mu$, the posterior mean. Also, by Lemma 2.2, the Bayes predictive density under the divergence loss is given by

$$\pi_D(y|X) \propto \left[\int N^{1-\beta}(\theta|y, \sigma_y^2I_p)N(\theta|(1 - B)X + B\mu, \sigma_x^2(1 - B)I_p)\,d\theta\right]^{1/\beta}$$

$$\propto N\left(y|(1 - B)X + B\mu, \left(\sigma_x^2(1 - B)(1 - \beta) + \sigma_y^2\right)I_p\right).$$

This proves the lemma. \qed

In the limiting ($B \to 0$) case, i.e. under the uniform prior $\pi(\theta) = 1$, the Bayes estimator of $\theta$ is $X$, and the Bayes predictive density of $Y$ is $N(y|X, (\sigma_x^2(1 - \beta) + \sigma_y^2)I_p)$.

We denote by $\delta_0$ the plug-in predictive density $N(X, \sigma_x^2I_p)$ of $N(y, \sigma_y^2I_p)$, and by $\delta_*$ the corresponding Bayes predictive density (under the uniform prior). The corresponding risk expressions are given in the following theorem.

**Theorem 2.4.** Under the divergence loss (1.2),

$$R(\theta, \delta_0) = \frac{1}{\beta(1-\beta)} \left[1 - (\sigma_y^2)^{\frac{p_0}{2}}(\sigma_x^2)^{\frac{p(1-\beta)}{2}}\{1 - \beta^2\sigma_x^2 + \beta\sigma_x^2\}^{-p/2}\right];$$

$$R(\theta, \delta_*) = \frac{1}{\beta(1-\beta)} \left[1 - (\sigma_y^2)^{\frac{p_0}{2}}\{1 - \beta\sigma_x^2 + \sigma_y^2\}^{-p/2}\right].$$

**Proof of Theorem 2.4.** From Lemma 2.2 with $\alpha_1 = 1 - \beta$ and $\alpha_2 = \beta$, the divergence loss for the plug-in predictive density $N(\theta, \sigma_x^2I_p)$, which we denote by $\delta_0$, is

$$L(\theta, \delta_0) = \frac{1}{\beta(1-\beta)} \left[1 - \int N^{1-\beta}(y|\theta, \sigma_y^2I_p)N^{\beta}(y|X, \sigma_x^2I_p)\,dy\right]$$

$$= \frac{1}{\beta(1-\beta)} \left[1 - (\sigma_y^2)^{\frac{p_0}{2}}(\sigma_x^2)^{\frac{p(1-\beta)}{2}}\{1 - \beta\sigma_x^2 + \beta\sigma_x^2\}^{-p/2}\right.\left.\times \exp \left\{-\frac{\beta(1-\beta)\|X - \theta\|^2}{2((1 - \beta)\sigma_x^2 + \beta\sigma_x^2)}\right\}\right].$$

(2.5)
Noting that $\|X - \theta\|^2 \sim \sigma_X^2 \chi_p^2$, the corresponding risk is given by

$$R(\theta, \delta_0) = \frac{1}{\beta(1-\beta)} \left[ 1 - (\sigma_\gamma^2)^{\frac{p \beta}{2}} (\sigma_X^2)^{\frac{p(1-\beta)}{2}} \left\{ (1-\beta)\sigma_X^2 + \beta\sigma_\gamma^2 \right\}^{-p/2} \right]$$

$$\times \left\{ 1 + \frac{\beta(1-\beta)\sigma_X^2}{(1-\beta)\sigma_X^2 + \beta\sigma_\gamma^2} \right\}^{-\frac{p}{2}}$$

$$= \frac{1}{\beta(1-\beta)} \left[ 1 - (\sigma_\gamma^2)^{\frac{p \beta}{2}} (\sigma_X^2)^{\frac{p(1-\beta)}{2}} \left\{ (1-\beta)\sigma_X^2 + \beta\sigma_\gamma^2 \right\}^{-p/2} \right]. \quad (2.6)$$

On the other hand, by Lemma 2.2 again, the divergence loss for the Bayes predictive density (under uniform prior) of $N(y|\theta, \sigma_\gamma^2 I_p)$ which we denote by $\delta_\delta$ is

$$L(\theta, \delta_\delta) = \frac{1}{\beta(1-\beta)} \left[ 1 - \int N^{1-\beta}(y|\theta, \sigma_\gamma^2 I_p) N^\beta(y|X, ((1-\beta)\sigma_X^2 + \sigma_\gamma^2) I_p) \, dy \right]$$

$$= \frac{1}{\beta(1-\beta)} \left[ 1 - (\sigma_\gamma^2)^{\frac{p \beta}{2}} (1-\beta)^2 (\sigma_\gamma^2)^{\frac{p(1-\beta)}{2}} \left\{ (1-\beta)\sigma_X^2 + \sigma_\gamma^2 \right\}^{-p/2} \right]$$

$$\times \exp \left\{ -\frac{\beta(1-\beta)\|X - \theta\|^2}{2((1-\beta)^2\sigma_X^2 + \sigma_\gamma^2)} \right\}. \quad (2.7)$$

The corresponding risk

$$R(\theta, \delta_\delta) = \frac{1}{\beta(1-\beta)} \left[ 1 - (\sigma_\gamma^2)^{\frac{p \beta}{2}} (1-\beta)^2 (\sigma_\gamma^2)^{\frac{p(1-\beta)}{2}} \left\{ (1-\beta)\sigma_X^2 + \sigma_\gamma^2 \right\}^{-\frac{p}{2}} \right]$$

$$\times \left\{ 1 + \frac{\beta(1-\beta)\sigma_X^2}{(1-\beta)^2\sigma_X^2 + \sigma_\gamma^2} \right\}^{-\frac{p}{2}}$$

$$= \frac{1}{\beta(1-\beta)} \left[ 1 - (\sigma_\gamma^2)^{\frac{p \beta}{2}} (1-\beta)^2 (\sigma_\gamma^2)^{\frac{p(1-\beta)}{2}} \left\{ (1-\beta)\sigma_X^2 + \sigma_\gamma^2 \right\}^{-\frac{p}{2}} \right]. \quad (2.8)$$

To show that $R(\theta, \delta_0) > R(\theta, \delta_\delta)$ for all $\theta$, $\sigma_X^2 > 0$ and $\sigma_\gamma^2 > 0$, it suffices to show that

$$(\sigma_\gamma^2)^{\frac{p \beta}{2}} (\sigma_X^2)^{\frac{p(1-\beta)}{2}} \left\{ (1-\beta)\sigma_X^2 + \beta\sigma_\gamma^2 \right\}^{-p/2} < (\sigma_\gamma^2)^{\frac{p \beta}{2}} (\sigma_X^2)^{\frac{p(1-\beta)}{2}} \left\{ (1-\beta)\sigma_X^2 + \sigma_\gamma^2 \right\}^{-\frac{p}{2}}, \quad (2.9)$$

or equivalently that

$$1 + \beta(\sigma_X^2/\sigma_\gamma^2 - \beta) > (1 + \sigma_X^2/\sigma_\gamma^2 - \beta)^\beta, \quad (2.10)$$

for all $0 < \beta < 1$, $\sigma_X^2 > 0$ and $\sigma_\gamma^2 > 0$. But the last inequality is a consequence of the elementary inequality $(1 + z)^u < 1 + uz$ for all real $z$ and $0 < u < 1$. \quad \square

In the next section, we prove the minimaxity of $X$ as an estimator of $\theta$ and the minimaxity of $N(y|X, ((1-\beta)\sigma_X^2 + \sigma_\gamma^2) I_p)$ as the predictive density of $Y$ in any arbitrary dimension.
3. Minimaxity results

Suppose \( X \sim N(\theta, \sigma^2_n\mathbf{I}_p) \), where \( \theta \in \mathbb{R}^p \). By Lemma 2.2 under the general divergence loss given in (1.2), the risk of \( X \) is given by

\[
R(\theta, X) = \frac{1}{\beta(1-\beta)}[1 - \{1 + \beta(1-\beta)\}^{-p/2}]
\]

for all \( \theta \). We now prove the minimaxity of \( X \) as an estimator of \( \theta \).

**Theorem 3.1.** \( X \) is a minimax estimator of the \( \theta \) in any arbitrary dimension under the divergence loss given in (1.2).

**Proof of Theorem 3.1.** Consider the sequence of proper priors \( N(0, \sigma^2_n\mathbf{I}_p) \) for \( \theta \), where \( \sigma^2_n \rightarrow \infty \) as \( n \rightarrow \infty \). We denote this sequence of priors by \( \pi_n \). The Bayes estimator of \( \theta \), namely the posterior mean, under the prior \( \pi_n \) is

\[
\delta^{\pi_n}(X) = (1 - B_n)X,
\]

with \( B_n = \sigma^2_n(\sigma^2 + \sigma^2_n)^{-1} \).

The Bayes risk of \( \delta^{\pi_n} \) under the prior \( \pi_n \) is given by:

\[
r(\pi_n, \delta^{\pi_n}) = \frac{1}{\beta(1-\beta)} \left[ 1 - E \left[ \exp \left\{ -\frac{\beta(1-\beta)}{2\sigma^2_n} \| \delta^{\pi_n}(X) - \theta \|^2 \right\} \right] \right],
\]

where expectation is taken over the joint distribution of \( X \) and \( \theta \), with \( \theta \) having the prior \( \pi_n \).

Since under the prior \( \pi_n \),

\[
\theta | X = x \sim N \left( \delta^{\pi_n}(x), \sigma^2_n(1 - B_n)\mathbf{I}_p \right),
\]

it follows that

\[
\| \theta - \delta^{\pi_n}(X) \|^2 \mid X = x \sim \sigma^2_n(1 - B_n)\chi^2_p,
\]

which does not depend on \( x \). Accordingly, from (3.3),

\[
r(\pi_n, \delta^{\pi_n}) = \frac{1}{\beta(1-\beta)} \left[ 1 - \{1 + \beta(1-\beta)(1 - B_n)\}^{-p/2} \right].
\]

Since \( B_n \rightarrow 0 \) as \( n \rightarrow \infty \), it follows from (3.4) that \( r(\pi_n, \delta^{\pi_n}) \rightarrow \frac{1}{\beta(1-\beta)} \left[ 1 - \{1 + \beta(1-\beta)\}^{-p/2} \right] \) as \( n \rightarrow \infty \).

Noting (3.1), an appeal to a result of Hodges and Lehmann [20] now shows that \( X \) is a minimax estimator of \( \theta \) for all \( p \). \( \square \)

Next we prove the minimaxity of the predictive density \( \delta_*(X) = N(y | X, ((1 - \beta)\sigma^2 + \sigma^2_n)\mathbf{I}_p) \) of \( Y \) having pdf \( N(y | \theta, \sigma^2_n\mathbf{I}_p) \).

**Theorem 3.2.** \( \delta_*(X) \) is a minimax predictive density of \( N(y | \theta, \sigma^2_n\mathbf{I}_p) \) in any arbitrary dimension under the general divergence loss given in (1.2).

**Proof of Theorem 3.2.** We have shown already that the predictive density \( \delta_*(X) \) of \( N(y | \theta, \sigma^2_n\mathbf{I}_p) \) has constant risk \( \frac{1}{\beta(1-\beta)} \left[ 1 - (\sigma^2_n )^{p/2} ((1 - \beta)\sigma^2 + \sigma^2_n )^{p/2} \right] \) under the divergence loss given in (1.2). Under the same sequence \( \pi_n \) of priors considered earlier in this section, by
Lemma 2.2, the Bayes predictive density of $N(y|\theta, \sigma^2_{\gamma}I_p)$ is given by $N(y|(1 - B_n)X, \{(1 - \beta)(1 - B_n)\sigma^2_{\chi} + \sigma^2_{\gamma}\}I_p)$. By Lemma 2.2 once again, one gets the identity

$$\int N^{1-\beta}(y|\theta, \sigma^2_{\gamma}I_p)N^\beta(y|(1 - B_n)X, \{(1 - \beta)(1 - B_n)\sigma^2_{\chi} + \sigma^2_{\gamma}\}I_p)\,dy = (\sigma^2_{\gamma})^{\frac{p\beta}{2}} ((1 - \beta)(1 - B_n)\sigma^2_{\chi} + \sigma^2_{\gamma})^{\frac{p(1-\beta)}{2}} \frac{(1 - \beta)^2(1 - B_n)\sigma^2_{\gamma} + \sigma^2_{\gamma} - p/2}{2((1 - \beta)^2(1 - B_n)\sigma^2_{\chi} + \sigma^2_{\chi})} \times \exp \left[ -\frac{\beta(1 - \beta)\|\theta - (1 - B_n)X\|^2}{2((1 - \beta)^2(1 - B_n)\sigma^2_{\chi} + \sigma^2_{\chi})} \right].$$

(3.5)

Noting once again that $\|\theta - (1 - B_n)X\|^2|X = x \sim \sigma^2_{\chi}(1 - B_n)\chi^2_p$, the posterior risk of $\delta_\ast(X)$ simplifies to

$$\frac{1}{\beta(1 - \beta)} \left[ 1 - (\sigma^2_{\gamma})^{\frac{p\beta}{2}} ((1 - \beta)(1 - B_n)\sigma^2_{\chi} + \sigma^2_{\gamma})^{\frac{p(1-\beta)}{2}} \right] \times ((1 - \beta)^2(1 - B_n)\sigma^2_{\chi} + \sigma^2_{\gamma})^{-\frac{p}{2}} \left[ 1 + \frac{\beta(1 - \beta)\sigma^2_{\chi}(1 - B_n)}{(1 - \beta)^2(1 - B_n)\sigma^2_{\chi} + \sigma^2_{\chi}} \right]^{-\frac{p}{2}}$$

$$= \frac{1}{\beta(1 - \beta)} \left[ 1 - (\sigma^2_{\gamma})^{\frac{p\beta}{2}} ((1 - \beta)(1 - B_n)\sigma^2_{\chi} + \sigma^2_{\gamma})^{\frac{p(1-\beta)}{2}} \right].$$

(3.6)

Since the expression does not depend on $x$, this is also the same as the Bayes risk of $\delta_\ast(X)$. The Bayes risk converges to

$$\frac{1}{\beta(1 - \beta)} \left[ 1 - (\sigma^2_{\gamma})^{\frac{p\beta}{2}} ((1 - \beta)\sigma^2_{\chi} + \sigma^2_{\gamma})^{\frac{p(1-\beta)}{2}} \right].$$


4. Admissibility for $p = 1$

We use Blyth’s [7] original technique for proving admissibility. First consider the estimation problem. Suppose that $X$ is not an admissible estimator of $\theta$. Then there exists an estimator $\delta_0(X)$ of $\theta$ such that $R(\theta, \delta_0) \leq R(\theta, X)$ for all $\theta$ with strict inequality for some $\theta = \theta_0$. Let $\eta = R(\theta_0, X) - R(\theta_0, \delta_0(X)) > 0$. Due to continuity of the risk function, there exists an interval $[\theta_0 - \epsilon, \theta_0 + \epsilon]$ with $\epsilon > 0$ such that $R(\theta, X) - R(\theta, \delta_0(X)) \geq \frac{1}{2}\eta$ for all $\theta \in [\theta_0 - \epsilon, \theta_0 + \epsilon]$. Now with the same prior $\pi_n(\theta) = N(\theta|0, \sigma^2_n)$,

$$r(\pi_n, X) - r(\pi_n, \delta_0(X)) = \int_{\mathbb{R}} [R(\theta, X) - R(\theta, \delta_0(X))]\pi_n(d\theta)$$

$$\geq \int_{\theta_0 - \epsilon}^{\theta_0 + \epsilon} [R(\theta, X) - R(\theta, \delta_0(X))]\pi_n(d\theta)$$

$$\geq \frac{1}{2}\eta \int_{\theta_0 - \epsilon}^{\theta_0 + \epsilon} (2\pi \sigma^2_n)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma^2_n} \theta^2 \right\} \,d\theta$$

$$\geq \frac{1}{2}\eta (2\pi \sigma^2_n)^{-\frac{1}{2}} \left( \frac{\epsilon}{2} \right).$$

(4.1)
Again,

\[
r(\pi_n, X) - r(\pi_n, \delta^{\pi_n}(X)) = \frac{[(1 + \beta(1 - \beta)(1 - B_n))^{-1/2} - (1 + \beta(1 - \beta))^{-1/2}]}{\beta(1 - \beta)} = O_\varepsilon(B_n)
\]

for large \(n\), where \(O_\varepsilon\) denotes the exact order. Since \(B_n = \sigma^2(\sigma^2 + \sigma_n^2)^{-1}\) and \(\sigma_n^2 \to \infty\) as \(n \to \infty\), denoting \(C(>0)\) as a generic constant, it follows from (4.1) and (4.2) that for large \(n\), say \(n \geq n_0\),

\[
\frac{r(\pi_n, X) - r(\pi_n, \delta_0(X))}{r(\pi_n, X) - r(\pi_n, \delta^{\pi_n}(X))} \geq \frac{1}{4} (2\pi)^{-1/2} \sigma_n^{-1} CB_n \to \infty
\]

as \(n \to \infty\).

Hence, for large \(n\), \(r(\pi_n, \delta^{\pi_n}(X)) > r(\pi_n, \delta_0(X))\) which contradicts the Bayesness of \(\delta^{\pi_n}(X)\) with respect to \(\pi_n\). This proves the admissibility of \(X\) for \(p = 1\).

For the prediction problem, suppose there exists a density \(p(y|v(X))\) which dominates \(N(y|X, ((1 - \beta)\sigma^2_x + \sigma^2_y)^{-1})\). Since

\[
[(1 - \beta)(1 - B_n)\sigma^2_x + \sigma^2_y)^{1/2} - (1 - \beta)\sigma^2_x + \sigma^2_y)^{-1/2}] = O_\varepsilon(B_n)
\]

for large \(n\) under the same prior \(\pi_n\), using a similar argument,

\[
\frac{r(\pi_n, N(y|X, ((1 - \beta)\sigma^2_x + \sigma^2_y))) - r(\pi_n, p(y|v(X)))}{r(\pi_n, N(y|X, ((1 - \beta)(1 - B_n)\sigma^2_x + \sigma^2_y))) - r(\pi_n, N(y|X, ((1 - \beta)(1 - B_n)\sigma^2_x + \sigma^2_y)))} = O(\sigma_n^{-1}B_n^{-1}) \to \infty,
\]

as \(n \to \infty\). An argument similar to the previous result now completes the proof. \(\square\)

**Remark 1.** The above technique of proving admissibility does not work for \(p = 2\). This is because for \(p = 2\), the ratios in the left hand side of (4.3) and (4.4) are greater than or equal to some constant times \(\sigma_n^{-2}B_n^{-1}\) for large \(n\) which tends to a constant as \(n \to \infty\). We conjecture the admissibility of \(X\) or \(N(y|X, ((1 - \beta)\sigma^2_x + \sigma^2_y)I_p)\) for \(p = 2\) under the general divergence loss for the respective problems of estimation and prediction. It is possible that the technique of Brown and Fox [11] (see also [17]) can be applied in this context. But that yet remains to be explored.

### 5. Inadmissibility results for \(p \geq 3\)

Let \(S = \|X\|^2/\sigma^2_x\). The Baranchick class of estimators for \(\theta\) is given by \(\delta^\tau(X) = (1 - \tau\frac{S}{\sigma^2_x})X\), where one needs some restrictions on \(\tau\). The special choice \(\tau(S) = p - 2\) (with \(p \geq 3\)) leads to the James–Stein estimator.

It is important to note that the class of estimators \(\delta^\tau(X)\) can be motivated from an empirical Bayes (EB) point of view. To see this, we first note that with the \(N(0, AI_p)\) \((A > 0)\) prior for \(\theta\), the Bayes estimator of \(\theta\) under the divergence loss is \((1 - B)X\), where \(B = \sigma^2_x(A + \sigma^2_y)^{-1}\).

An EB estimator of \(\theta\) estimates \(B\) from the marginal distribution of \(X\). Marginally, \(X \sim N(0, \sigma^2_xB^{-1}I_p)\) so that \(S\) is minimal sufficient for \(B\). Thus, a general EB estimator of \(\theta\) can
be written in the form $\delta^T(X)$. In particular, the UMVUE of $B$ is $(p - 2)/S$ which leads to the James–Stein estimator [15].

Note that for the estimation problem,
\begin{equation}
L(\theta, \delta^T(X)) = \frac{1 - \exp\left[-\frac{\beta(1-\beta)}{2\sigma^2} \|\delta^T(X) - \theta\|^2\right]}{\beta(1 - \beta)},
\end{equation}
while for the prediction problem,
\begin{equation}
L(N(y|\theta, I_p), N(y|\delta^T(X), ((1 - \beta)(\sigma^2 + \sigma^2)I_p)) = 1 - (\sigma^2)^{\frac{p(1-\beta)}{\beta}}((1 - \beta)(\sigma^2 + \sigma^2))^{\frac{p(1-\beta)}{\beta}}(1 - \beta)^2(\sigma^2 + \sigma^2)^{-p/2}/\beta(1 - \beta) \times \exp\left[-\frac{\beta(1-\beta)\|\delta^T(X) - \theta\|^2}{2((1 - \beta)2\sigma^2 + \sigma^2)^2}\right].
\end{equation}

The first result of this section finds an expression for $E_\theta \left\{\exp\{-b\|\delta^T(X) - \theta\|^2\}\right\}$, $b > 0$. We will need $b = \beta(1 - \beta)$ and $b = \frac{\beta(1-\beta)\sigma^2}{2((1 - \beta)(\sigma^2 + \sigma^2)^2}$ to evaluate (5.1) and (5.2).

**Theorem 5.1.**
\begin{equation}
E_\theta \left\{\exp\{-\frac{b}{\sigma^2} \|\delta^T(X) - \theta\|^2\}\right\} = (2b + 1)^{-p/2} \sum_{r=0}^{\infty} \{\exp(-\phi)\phi^r/r!\} I_b(r),
\end{equation}
where $\phi = (b + \frac{1}{2})\|\theta\|^2/\sigma^2$ and
\begin{equation}
I_b(r) = \int_0^{\infty} \left\{1 - \frac{b}{t} \left(\frac{2t}{2b + 1}\right)\right\} \frac{2r^{r+\frac{p}{2}-1}}{2^r \Gamma(r + \frac{p}{2})} \times \exp\left[-t - \frac{b(b + \frac{1}{2})}{t} \frac{2t}{2b + 1} + 2b\tau \frac{2t}{2b + 1}\right] dt.\end{equation}

The proof of the result is technical, and is deferred to the Appendix. As a consequence of this theorem, putting $b = \beta(1 - \beta)$, it follows from (5.1) and (5.3) that
\begin{equation}
R(\theta, \delta^T(X)) = \frac{1 - (1 + \beta(1 - \beta))^{-p/2} \sum_{r=0}^{\infty} \{\exp(-\phi)\phi^r/r!\} I_{\beta(1-\beta)/2}(r)}{\beta(1 - \beta)},
\end{equation}
while putting $b = \frac{\beta(1-\beta)\sigma^2}{2((1 - \beta)(\sigma^2 + \sigma^2)^2}$, it follows from (5.2) and (5.3) that
\begin{equation}
R(N(y|\theta, \sigma^2 I_p), N(y|\delta^T(X), ((1 - \beta)(\sigma^2 + \sigma^2)I_p)) = \frac{1 - (\sigma^2)^{p\beta/2}(\sigma^2 + \sigma^2)^{-p\beta/2} \sum_{r=0}^{\infty} \{\exp(-\phi)\phi^r/r!\} I_{\beta(1-\beta)\sigma^2}{2((1 - \beta)(\sigma^2 + \sigma^2)^2}(r)}{\beta(1 - \beta)}.\end{equation}
Hence, proving \( I_b(r) > 1 \) for all \( b > 0 \) under certain conditions on \( \tau \) leads to

\[
R(\theta, \delta^r(X)) < R(\theta, X)
\]

and

\[
R(N(y|\theta, \sigma_0^2 I_p), N(y|\delta^r(X), ((1 - \beta)\sigma_x^2 + \sigma_y^2)I_p)) < R(N(y|\theta, \sigma_0^2 I_p), N(y|X, ((1 - \beta)\sigma_x^2 + \sigma_y^2)I_p))
\]

for all \( \theta \). In the limiting case when \( \beta \to 0 \), i.e. for the KL loss, one gets \( R_{KL}(\theta, \delta^r(X)) < p/2 = R_{KL}(\theta, X) \) for all \( \theta \), since as shown in Section 1, for estimation, the KL loss is half of the squared error loss. Similarly, for the prediction problem, as \( \beta \to 0 \),

\[
R_{KL}(N(y|\theta, \sigma_0^2 I_p), N(y|\delta^r(X), ((1 - \beta)\sigma_x^2 + \sigma_y^2)I_p)) < \frac{p}{2} \log \left( \frac{\sigma_x^2 + \sigma_y^2}{\sigma_y^2} \right) = R_{KL}(N(y|\theta, \sigma_0^2 I_p), N(y|X, ((1 - \beta)\sigma_x^2 + \sigma_y^2)I_p))
\]

for all \( \theta \).

The following theorem provides sufficient conditions on the function \( \tau(\cdot) \) which guarantee \( I_b(r) > 1 \) for all \( r = 0, 1, \ldots \).

**Theorem 5.2.** Let \( p \geq 3 \). Suppose

(i) \( 0 < \tau(t) < 2(p - 2) \) for all \( t > 0 \);

(ii) \( \tau(t) \) is a differentiable nondecreasing function of \( t \).

Then \( I_b(r) > 1 \) for all \( b > 0 \).

The proof of Theorem 5.2 is also deferred to Appendix.

**Remark 2.** Baranchick [3], under squared error loss, proved the dominance of \( \delta^r(X) \) over \( X \) under (i) and (ii). We may note that the special choice \( \tau(t) = p - 2 \) for all \( t \) leading to the James–Stein estimator, satisfies both conditions (i) and (ii) of the theorem.

**Remark 3.** We may note that the Baranchick class of estimators shrinks the sample mean \( X \) towards 0. Instead one can shrink \( X \) towards any arbitrary constant \( \mu \). In particular, if we consider the \( N(\mu, AI_p) \) prior for \( \theta \), where \( \mu \in \mathbb{R}^p \) is known, then the Bayes estimator of \( \theta \) is \( (1 - B)X + B\mu \), where \( B = \sigma_x^2(A + \sigma_x^2)^{-1} \). A general EB estimator of \( \theta \) is then given by

\[
\delta^{**}(X) = \left( 1 - \frac{\tau(S')}{S'} \right) X + \frac{\tau(S')}{S'} \mu,
\]

where \( S' = ||X - \mu||^2/\sigma_x^2 \), and Theorem 5.2 with obvious modifications will then provide the dominance of the EB estimator \( \delta^{**}(X) \) over \( X \) under the divergence loss. The corresponding prediction result is also true.

**Remark 4.** The special case with \( \tau(t) = c \) satisfies conditions of the theorem if \( 0 < c < 2(p - 2) \). This is the original James–Stein result.

**Remark 5.** Strawderman [25] considered the hierarchical prior

\[
\theta|A \sim N(0, AI_p),
\]
where $A$ has pdf

$$
\pi(A) = \delta(1 + A)^{-1-\delta} I_{[A>0]}
$$

with $\delta > 0$.

Under the above prior, assuming squared error loss, and recalling that $S = ||X||^2/\sigma^2_x$, the Bayes estimator of $\theta$ is given by

$$
\tau(S) = \left(1 - \frac{\tau(S)}{\delta}\right)X,
$$

where

$$
\tau(S) = p + 2\delta - \int_0^{\frac{\delta}{2}} \frac{2 \exp\left(-\frac{t}{2}\right)}{\lambda^{\frac{p}{2}+\delta-1} \exp\left(-\frac{\lambda}{2}t\right)} d\lambda.
$$

(5.5)

Under the general divergence loss, it is not clear whether this estimator is the hierarchical Bayes estimator of $\theta$, although its EB interpretation continues to hold. Besides, as it is well known, this particular $\tau$ satisfies conditions of Theorem 5.2 if $p > 4 + 2\delta$. Thus the Strawderman class of estimators dominates $X$ under the general divergence loss. The corresponding predictive density also dominates $N(y|X, ((1 - \beta)\sigma_x^2 + \sigma_y^2)I_p)$. In a preliminary version of the paper, the authors proved a weaker result using the increasing failure rate (IFR) property [6] of certain distributions. For the special KL loss, the present results complement those of Komaki [22] and George et al. [18]. The predictive density obtained by these authors under the Strawderman prior, (and Stein’s superharmonic prior as a special case) are quite different from the general class of EB predictive densities of this paper. One of the virtues of the latter is that the expressions are in closed form, and thus it is easy to generate samples from these densities.

6. Lindley’s estimator

Lindley [23] considered a modification of the James–Stein estimator. Rather then shrinking $X$ towards an arbitrary point, say $\mu$, he proposed shrinking $X$ towards $\bar{X}1_p$, where $\bar{X} = \frac{1}{p} \sum_{i=1}^p X_i$ and $1_p$ is a $p$-component column vector with each element equal to 1. Writing $R = \sum_{i=1}^p (X_i - \bar{X})^2/\sigma^2_x$, Lindley’s estimator is given by

$$
\delta(X) = X - \frac{p - 3}{R} (X - \bar{X}1_p), \quad p \geq 4.
$$

(6.1)

The above estimator has a simple EB interpretation. Suppose $X|\theta \sim N(\theta, \sigma^2_x I_p)$ and $\theta$ has the $N_p(\mu 1_p, AI_p)$ prior. Then the Bayes estimator of $\theta$ is given by $(1 - B)X + B\mu 1_p$ where $B = \sigma^2_x (A + \sigma^2_x)^{-1}$. Now if both $\mu$ and $A$ are unknown, since marginally $X \sim N(\mu 1_p, \sigma^2_x B^{-1} I_p)$, $(X, R)$ is complete sufficient for $\mu$ and $B$, and the UMVUE of $\mu$ and $B^{-1}$ are given by $\bar{X}$ and $(p - 3)/R$, $p \geq 4$.

Following [3] a more general class of EB estimators is given by

$$
\delta^*_\tau(X) = X - \frac{\tau(R)}{R} (X - \bar{X}1_p), \quad p \geq 4.
$$

(6.2)

**Theorem 6.1.** Assume

(i) $0 < \tau(t) < 2(p - 3)$ for all $t > 0$, $p \geq 4$;

(ii) $\tau(t)$ is a nondecreasing differentiable function of $t$. 

Then the estimator $\delta^*_X(X)$ dominates $X$ under the divergence loss given in (1.2). Similarly, $N(y|\delta^*_X(X), ((1 - \beta)\sigma^2_x + \sigma^2_y)I_p)$ dominates $N(y|X, ((1 - \beta)\sigma^2_x + \sigma^2_y)I_p)$ as the predictor of $N(y|\theta, \sigma^2_y I_p)$.

The proof of Theorem 6.1 is included in the Appendix.

The above result can immediately be extended to shrinkage towards an arbitrary regression surface. Suppose now that $X|\theta \sim N(\theta, \sigma^2_x I_p)$ and $\theta \sim N_p(K\beta, A I_p)$ where $K$ is a known $p \times r$ matrix of rank $r(< p)$ and $\beta$ is $r \times 1$ regression coefficient. Writing $P = K(K^T K)^{-1}K^T$, the projection of $X$ on the regression surface is given by $PX = \hat{K}\beta$, where $\hat{K} = (K^T K)^{-1}K^T X$ is the least squares estimator of $\beta$. Now we consider the general class of estimators given by

$$X - \frac{\tau(R^*)}{R^*}(X - PX),$$

where $R^* = \|X - PX\|^2/\sigma^2_x$. The above estimator also has an EB interpretation noting that marginally $(\hat{\beta}, R^*)$ is complete sufficient for $(\beta, A)$.

The following theorem now extends Theorem 6.1.

**Theorem 6.2.** Let $p \geq r + 3$ and

(i) $0 < \tau(t) < 2(p - r - 2)$ for all $t > 0$;
(ii) $\tau(t)$ is a nondecreasing differentiable function of $t$.

Then the estimator $X - \frac{\tau(R^*)}{R^*}(X - PX)$ dominates $X$ under the divergence loss. A similar dominance result holds for prediction of $N(y|\theta, \sigma^2_y I_p)$.

## 7. Summary and conclusion

The paper considers estimation of the normal mean and prediction of the $N(y|\theta, \sigma^2_y I_p)$ density under a general divergence loss. It is shown that the sample mean is a minimax estimator of a population mean in any arbitrary dimension, and is admissible in one dimension. The same results hold for the $N(y|X, ((1 - \beta)\sigma^2_x + \sigma^2_y)I_p)$ predictor of $N(y|\theta, \sigma^2_y I_p)$. However, the sample mean is inadmissible in three or higher dimensions. A general class of minimax estimators dominating the sample mean is provided. The dual set of predictors dominating $N(y|X, ((1 - \beta)\sigma^2_x + \sigma^2_y)I_p)$ is also given. The divergence loss considered in this paper includes both KL and BH losses.

One important open question that we have already posed is whether the admissibility results for estimation and prediction hold in dimension 2. We wish to address this problem. Moreover, we wish to extend estimation and prediction results when $X \sim N(\theta, \sigma^2_x I_p)$ pdf when $\theta$ and $\sigma^2_x$ are both unknown, and subsequently to the more general $N(\theta, \Sigma)$ pdf when both $\theta$ and $\Sigma$ are unknown.

**Acknowledgments**

Thanks are due to two reviewers for their comments.

## Appendix

**Proof of Lemma 2.2.** Writing $H = \alpha_1\Sigma_1^{-1} + \alpha_2\Sigma_2^{-1}$ and $g = H^{-1}(\alpha_1\Sigma_1^{-1}\mu_1 + \alpha_2\Sigma_2^{-1}\mu_2)$, it follows after some simplification that
\[
\int [N_p(x|\mu_1, \Sigma_1)]^{\alpha_1} [N_p(x|\mu_2, \Sigma_2)]^{\alpha_2} \, dx \\
= (2\pi)^{-\frac{p}{2}}(\alpha_1+\alpha_2) |\Sigma_1|^{-\frac{1}{2}} |\Sigma_2|^{-\frac{1}{2}} \alpha_2 \int \exp \left\{ -\frac{1}{2} (x - g)^T H (x - g) \right\} \\
- \frac{1}{2} \left\{ \alpha_1 (\mu_1^T \Sigma_1^{-1} \mu_1) + \alpha_2 (\mu_2^T \Sigma_2^{-1} \mu_2) - g^T H g \right\} \, dx \\
= (2\pi)^{-\frac{p}{2}}(1-\alpha_1-\alpha_2) |\Sigma_1|^{-\frac{1}{2}} |\Sigma_2|^{-\frac{1}{2}} \alpha_2 |H|^{-\frac{1}{2}} \\
\times \exp \left\{ -\frac{1}{2} \left\{ \alpha_1 (\mu_1^T \Sigma_1^{-1} \mu_1) + \alpha_2 (\mu_2^T \Sigma_2^{-1} \mu_2) - g^T H g \right\} \right\} . \tag{A.1} \]

It can be checked that
\[
\alpha_1 (\mu_1^T \Sigma_1^{-1} \mu_1) + \alpha_2 (\mu_2^T \Sigma_2^{-1} \mu_2) - g^T H g \\
= \alpha_1 \alpha_2 (\mu_1 - \mu_2)^T (\alpha_1 \Sigma_2 + \alpha_2 \Sigma_1)^{-1} (\mu_1 - \mu_2) , \tag{A.2} \]
and
\[
|H|^{-1/2} = |\Sigma_1|^{1/2} |\Sigma_2|^{1/2} |\alpha_1 \Sigma_2 + \alpha_2 \Sigma_1|^{-1/2} . \tag{A.3} \]

Then by (A.2) and (A.3),
\[
\text{right hand side of (A.1) } = (2\pi)^{p(1-\alpha_1-\alpha_2)/2} |\Sigma_1|^{(1-\alpha_1)/2} |\Sigma_2|^{(1-\alpha_2)/2} |\alpha_1 \Sigma_2 + \alpha_2 \Sigma_1|^{-1/2} \\
\times \exp \left[ -\frac{\alpha_1 \alpha_2}{2} (\mu_1 - \mu_2)^T (\alpha_1 \Sigma_2 + \alpha_2 \Sigma_1)^{-1} (\mu_1 - \mu_2) \right] .
\]

This proves the lemma. \(\Box\)

**Proof of Theorem 5.1.** Recall that \(S = \|X\|^2/\sigma_x^2\). For \(\|\theta\| = 0\), proof is straightforward. So we consider \(\|\theta\| > 0\). Let \(Z = X/\sigma_x\) and \(\eta = \theta/\sigma_x\). First we reexpress \(\frac{1}{\sigma_x^2} \left\| \left(1 - \frac{\tau(S)}{S} \right) X - \theta \right\|^2\) as
\[
\left\| \left(1 - \frac{\tau(\|Z\|^2)}{\|Z\|^2} \right) Z - \eta \right\|^2 \\
= S + \frac{\tau^2(S)}{S} - 2\tau(S) + \|\eta\|^2 - 2\|\eta\| \left(1 - \frac{\tau(S)}{S} \right) . \tag{A.4} \]

We begin with the orthogonal transformation \(Y = CZ\) where \(C\) is an orthogonal matrix with its first row given by \(\{\theta_1/\|\theta\|, \ldots, \theta_p/\|\theta\|\}\). Writing \(Y = (Y_1, \ldots, Y_p)^T\), the right hand side of (A.4) can be written as
\[
S + \frac{\tau^2(S)}{S} - 2\tau(S) + \|\eta\|^2 - 2\|\eta\| Y_1 \left(1 - \frac{\tau(S)}{S} \right) , \tag{A.5} \]
where \(S = \|Y\|^2\). Also we note that \(Y_1, \ldots, Y_p\) are mutually independent with \(Y_1 \sim N(\|\eta\|, 1)\), and \(Y_2, \ldots, Y_p\) are iid \(N(0, 1)\). Now writing \(Z = \sum_{i=2}^p Y_i^2 \sim \chi_{p-1}^2\) we have
\[
E \left[ \exp \left\{ -\frac{b}{\sigma_x^2} \|\delta^T(X) - \theta\|^2 \right\} \right] \\
= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp \left\{ -b \left\{ y_1^2 + z + \frac{\tau^2(y_1^2 + z)}{(y_1^2 + z)} - 2\tau(y_1^2 + z) + \|\eta\|^2 - 2\|\eta\| y_1 \right\} \right\} \\
\times \left(1 - \frac{\tau(y_1^2 + z)}{y_1^2 + z} \right) \, \right\} \right]\]
\[
\times (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2} (y_1 - \|\eta\|)^2 \right\} \exp \left\{ -\frac{1}{2} z^2 \right\} \frac{z^{\frac{1}{2}(p-1)-1}}{2^{p-1} \Gamma \left( \frac{p-1}{2} \right)} dy_1 dz
\]
\[
= \int_0^{+\infty} \int_{-\infty}^{+\infty} (2\pi)^{-1/2} \exp \left[ -\left( b + \frac{1}{2} \right) (y_1 - \|\eta\|)^2 - \frac{b\tau^2(y_1^2 + z)}{(y_1^2 + z)} \right] dy_1 dz
\]
\[
+ 2b\tau(y_1^2 + z) - 2b\|\eta\|y_1 \frac{\tau(y_1^2 + z)}{y_1^2 + z} \exp \left\{ -\left( b + \frac{1}{2} \right) z \right\} \frac{z^{\frac{p-3}{2}}}{2^{p-1} \Gamma \left( \frac{p-1}{2} \right)} dy_1 dz.
\]  

(A.6)

We first simplify
\[
\int_0^{+\infty} (2\pi)^{-1/2} \exp \left[ -\left( b + \frac{1}{2} \right) (y_1 - \|\eta\|)^2 - \frac{b\tau^2(y_1^2 + z)}{(y_1^2 + z)} \right] dy_1
\]
\[
+ 2b\tau(y_1^2 + z) - 2b\|\eta\|y_1 \frac{\tau(y_1^2 + z)}{y_1^2 + z} \exp \left\{ -\left( b + \frac{1}{2} \right) z \right\} dy_1
\]
\[
= \int_0^{+\infty} (2\pi)^{-1/2} \exp \left[ -\left( b + \frac{1}{2} \right) (y_1^2 + \|\eta\|^2) - \frac{b\tau^2(y_1^2 + z)}{(y_1^2 + z)} + 2b\tau(y_1^2 + z) \right]
\]
\[
\times \left[ \exp \left\{ 2 \left( b + \frac{1}{2} \right) \|\eta\|y_1 - 2b\|\eta\|y_1 \frac{\tau(y_1^2 + z)}{y_1^2 + z} \right\} \right.
\]
\[
+ \exp \left\{ -2 \left( b + \frac{1}{2} \right) \|\eta\|y_1 + 2b\|\eta\|y_1 \frac{\tau(y_1^2 + z)}{y_1^2 + z} \right\} \right] dy_1
\]
\[
= 2 \int_0^{+\infty} (2\pi)^{-1/2} \exp \left[ -\left( b + \frac{1}{2} \right) (y_1^2 + \|\eta\|^2) - \frac{b\tau^2(y_1^2 + z)}{(y_1^2 + z)} + 2b\tau(y_1^2 + z) \right]
\]
\[
\times \left[ \sum_{r=0}^{\infty} \frac{(2\|\eta\|y_1)^{2r}}{(2r)!} \left\{ \left( b + \frac{1}{2} \right) - \frac{b\tau(y_1^2 + z)}{y_1^2 + z} \right\}^{2r} \right] dy_1
\]
\[
= 2 \int_0^{+\infty} (2\pi)^{-1/2} \exp \left[ -\left( b + \frac{1}{2} \right) (w + \|\eta\|^2) - \frac{b\tau^2(w + z)}{(w + z)} + 2b\tau(w + z) \right]
\]
\[
\times \left[ \sum_{r=0}^{\infty} \frac{2\|\eta\|^{2r}w^{r-\frac{1}{2}}}{(2r)!} \left\{ \left( b + \frac{1}{2} \right) - \frac{b\tau(w + z)}{w + z} \right\}^{2r} \right] dw,
\]  

(A.7)

where \( w = y_1^2 \).

With the substitution \( v = w + z \) and \( u = w/(w + z) \), it follows from (A.6) and (A.7) that
\[
E \left[ \exp \left\{ -\frac{b}{\sigma^2} \|\delta^\tau(X) - \theta\|^2 \right\} \right]
\]
\[
= \int_0^{+\infty} \int_0^1 (2\pi)^{-1/2} \sum_{r=0}^{\infty} \exp \left( -\left( b + \frac{1}{2} \right) \|\eta\|^2 \right) \frac{(2b\|\eta\|)^{2r}}{(2r)!} \]
\begin{equation}
\times \left\{ \left( (2b)^{-1} + 1 \right) - \frac{\tau(v)}{v} \right\}^{2r} \exp \left[ - \left( b + \frac{1}{2} \right) v - \frac{b\tau^2(v)}{v} + 2b\tau(v) \right] v^{r+\frac{p}{2}-1} \\
\times \frac{u^{r+\frac{p}{2}-1}(1-u)^{\frac{p-1}{2}}}{2^{\frac{p}{2}} \Gamma \left( \frac{p-1}{2} \right)} \, du \, dv.
\end{equation}

(A.8)

By the Legendre duplication formula, namely,

\[(2r)! = \Gamma(2r + 1) = \Gamma \left( r + \frac{1}{2} \right) \Gamma(r + 1) 2^{2r} \pi^{-1/2},\]

(A.8) simplifies into

\begin{align*}
E \left[ \exp \left\{ - \frac{b}{\sigma^2_x} \| \delta^r(X) - \theta \|^2 \right\} \right] &= \int_0^{+\infty} \int_0^{1} (2\pi)^{-1/2} \sum_{r=0}^{\infty} \exp \left( - \left( b + \frac{1}{2} \right) \| \eta \|^2 \right) \\
\times \frac{(2b\| \eta \|)^{2r} \sqrt{\pi}}{r! \Gamma \left( r + \frac{1}{2} \right)} \left\{ \left( (2b)^{-1} + 1 \right) - \frac{\tau(v)}{v} \right\}^{2r} \\
\times v^{r+\frac{p}{2}-1} \exp \left[ - \left( b + \frac{1}{2} \right) v - \frac{b\tau^2(v)}{v} + 2b\tau(v) \right] \\
\times \frac{2^{\frac{p}{2}} \Gamma \left( r + \frac{p}{2} \right)}{2^{\frac{p}{2}} \Gamma \left( r + \frac{p}{2} \right) \Gamma \left( r + \frac{p}{2} \right)} \, du \, dv.
\end{align*}

(A.9)

Integrating with respect to $u$, (A.9) leads to

\begin{align*}
E \left[ \exp \left\{ - \frac{b}{\sigma^2_x} \| \delta^r(X) - \theta \|^2 \right\} \right] &= \sum_{r=0}^{\infty} \exp \left\{ - \phi \frac{\phi^r}{r!} \int_0^{+\infty} \left( b + \frac{1}{2} \right) \left\{ 1 - \frac{\tau(v)}{(2b)^{-1} + 1)} v \right\}^{2r} \\
\times v^{r+\frac{p}{2}-1} \exp \left[ - \left( b + \frac{1}{2} \right) v - \frac{b\tau^2(v)}{v} + 2b\tau(v) \right] \, dv
\end{align*}

(A.10)

where $\phi = \left( b + \frac{1}{2} \right) \| \eta \|^2$. Now putting $t = \left( b + \frac{1}{2} \right) v$, we get from (A.10)

\begin{equation}
E \left[ \exp \left\{ - \frac{b}{\sigma^2_x} \| \delta^r(X) - \theta \|^2 \right\} \right] = (2b + 1)^{-\frac{p}{2}} \sum_{r=0}^{\infty} \exp \left\{ - \phi \frac{\phi^r}{r!} \right\}
\end{equation}
\[ \times \int_0^{+\infty} \exp \left\{ -t - \frac{b (b + \frac{1}{2}) \tau^2 (\frac{2r}{2b+1})}{t} + 2b \tau \left( \frac{2t}{2b+1} \right) \right\} \]
\[ \times \left( 1 - \frac{b}{t} \tau \left( \frac{2t}{2b+1} \right) \right)^{2r} \frac{t^{r + \frac{c}{2}}}{\Gamma(r + \frac{c}{2})} \, dr. \]

The theorem follows. \( \square \)

**Proof of Theorem 5.2.** Define \( \tau_0(t) = \tau(\frac{2t}{2b+1}) \). Notice that \( \tau_0(t) \) will also satisfy conditions of Theorem 5.2. Now

\[ t + \frac{b (b + \frac{1}{2}) \tau^2 (\frac{2t}{2b+1})}{t} - 2b \tau \left( \frac{2t}{2b+1} \right) = t \left( 1 - \frac{b \tau_0(t)}{t} \right)^2 + \frac{b}{2t} \tau_0^2(t) \]

and

\[ I_b(r) = \int_0^{+\infty} \frac{\exp \left\{ -t \left( 1 - \frac{b \tau_0(t)}{t} \right)^2 \right\} \left( t \left( 1 - \frac{b \tau_0(t)}{t} \right)^2 \right)^{r + \frac{c}{2}}}{\Gamma(r + \frac{c}{2})} \times \exp \left\{ - \frac{b \tau_0^2(t)}{2t} \right\} \left( 1 - \frac{b \tau_0(t)}{t} \right)^{-(p-2)} \, dr. \]

Define \( t_0 = \sup\{t > 0 : \tau_0(t)/t \geq b^{-1} \} \). Since \( \tau_0(t)/t \) is continuous in \( t \) with \( \lim_{t \to 0} \tau_0(t)/t = +\infty \) and \( \lim_{t \to \infty} \tau_0(t)/t = 0 \), there exists such a \( t_0 \) which also satisfies \( \tau_0(t_0)/t_0 = b^{-1} \). We now need the following lemma.

**Lemma A.1.** For \( t \geq t_0, b > 0 \) and \( \tau_0(t) \) satisfying conditions of Theorem 5.2 the following inequality holds:

\[ \exp \left\{ - \frac{b \tau_0^2(t)}{2t} \right\} \left( 1 - \frac{b \tau_0(t)}{t} \right)^{-(p-2)} - q'(t) \geq 0, \]

where \( q(t) = t \left( 1 - \frac{b \tau_0(t)}{t} \right)^2. \)

**Proof of Lemma A.1.** Notice first that for \( t \geq t_0 \), by the inequality, \( (1 - z)^{-c} \geq \exp(cz) \) for \( c > 0 \) and \( 0 < z < 1 \), one gets

\[ \exp \left\{ - \frac{b \tau_0^2(t)}{2t} \right\} \left( 1 - \frac{b \tau_0(t)}{t} \right)^{-(p-2)} \geq \exp \left\{ - \frac{b \tau_0^2(t)}{2t} + \frac{b(p - 2) \tau_0(t)}{t} \right\}, \]

for \( 0 < \tau_0(t) < 2(p - 2) \).

Notice that

\[ q'(t) = 1 - 2b \tau_0'(t) + \frac{2b^2 \tau_0(t) \tau_0'(t)}{t} - \frac{b^2 \tau_0^2(t)}{t^2}. \]

Thus \( q'(t) \leq 1 \) for \( t \geq t_0 \) if and only if

\[ 2b \tau'_0(t) \left( 1 - \frac{b \tau_0(t)}{t} \right) + \frac{b^2 \tau_0^2(t)}{t^2} \geq 0. \]
The last inequality is true since \( \tau'_0(t) \geq 0 \) for all \( t > 0 \) and \( \tau_0(t)/t \leq b^{-1} \) for all \( t \geq t_0 \). The lemma follows.

In view of previous lemma, it follows from (A.12) that

\[
I_b(r) \geq \frac{1}{\Gamma\left(r + \frac{p}{2}\right)} \int_{t_0}^{+\infty} \exp\left(-q(t)\right)\left(q(t)\right)^{r + \frac{p}{2} - 1} q'(t) \, dt = 1.
\]

This proves Theorem 5.2.

Proof of Theorem 6.1. Let \( \hat{\theta} = p^{-1} \sum_{i=1}^{p} \theta_i, \hat{\eta} = \hat{\theta}/\sigma_x \) and \( \zeta^2 = \frac{1}{\sigma^2} \sum_{i=1}^{p} \left( \theta_i - \hat{\theta} \right)^2 \). As in the proof of Theorem 5.1, we assume \( \zeta > 0 \). We first rewrite

\[
\frac{1}{\sigma_x^2} \|\delta_x^*(X) - \theta\|^2 = \left\| Z - \eta - \frac{\tau(R)}{R} (Z - \bar{Z})_{1_p} \right\|^2
\]

\[
= \left( Z - \bar{Z}_{1_p} \right) - \left( \eta - \bar{\eta} \right)_{1_p} + \left( \bar{Z} - \bar{\eta} \right)_{1_p} - \frac{\tau(R)}{R} (Z - \bar{Z})_{1_p} \right\|^2
\]

\[
= \left[ 1 - \frac{\tau(R)}{R} \right]^2 R + p(\bar{Z} - \bar{\eta})^2 + \zeta^2 - 2(\eta - \bar{\eta})^T \left( 1 - \frac{\tau(R)}{R} \right) (Z - \bar{Z})_{1_p}. \quad (A.16)
\]

By the orthogonal transformation \( G = (G_1, \ldots, G_p)^T = CZ \), where \( C \) is an orthogonal matrix with first two rows given by \( (p^{-\frac{1}{2}}, \ldots, p^{-\frac{1}{2}}) \) and \( ((\eta_1 - \bar{\eta})/\zeta, \ldots, (\eta_p - \bar{\eta})/\zeta) \). We can rewrite

\[
\frac{1}{\sigma_x^2} \|\delta_x^*(X) - \theta\|^2 = \left[ 1 - \frac{\tau(G_2 + Q)}{G_2^2 + Q} \right]^2 (G_2^2 + Q) + (G_1 - \sqrt{p} \bar{\eta})^2
\]

\[
+ \zeta^2 - 2\zeta G_2 \left( 1 - \frac{\tau(G_2 + Q)}{G_2^2 + Q} \right), \quad (A.17)
\]

where \( Q = \sum_{i=3}^{p} G_i^2 \) and \( G_1, G_2, \ldots, G_p \) are mutually independent with \( G_1 \sim N(\sqrt{p} \bar{\eta}, 1) \), \( G_2 \sim N(\zeta, 1) \) and \( G_3, \ldots, G_p \) are iid \( N(0, 1) \). Hence due to the independence of \( G_1 \) with \( (G_2, \ldots, G_p) \), and the fact that \( (G_1 - \sqrt{p} \bar{\eta})^2 \sim \chi_1^2 \), from (A.17),

\[
E \left[ \exp \left\{ -\frac{b}{\sigma_x^2} \|\delta_x^*(X) - \theta\|^2 \right\} \right]
\]

\[
= (2b + 1)^{-\frac{p}{2}} E \left[ \exp \left\{ -b \left( 1 - \frac{\tau(G_2^2 + Q)}{G_2^2 + Q} \right)^2 (G_2^2 + Q) \right. \right.
\]

\[
+ \zeta^2 - 2\zeta G_2 \left( 1 - \frac{\tau(G_2 + Q)}{G_2^2 + Q} \right) \right\}
\]

\[
= (2b + 1)^{-\frac{p}{2}} \sum_{r=0}^{\infty} \exp\{-\phi_s\} \frac{\phi_s^r}{r!} \int_0^{+\infty} \exp \left\{ -t - b \left( 1 + \frac{1}{2} \right) \frac{\tau_0^2(t)}{t} + 2b \tau_0(t)4 \right\}
\]

\[
	imes \left( 1 - b \frac{\tau_0(t)}{t} \right)^{2r} t^{r + \frac{p}{2} - 1} \frac{dr}{\Gamma(r + \frac{p}{2})}, \quad (A.18)
\]
where $\varphi_\ast = (b + \frac{1}{2})\zeta^2$ and as before $\tau_0(t) = \tau(\frac{t}{1+b})$. The second equality in (A.18) follows after long simplification proceeding as in the proof of Theorem 5.1.

Hence, by (A.18), the dominance of $\delta_\ast^2(X)$ over $X$ follows if the right hand side of (A.18) $\geq (2b + 1)^{-\nu/2}$. This however is an immediate consequence of Theorem 5.2. □

**Proof of Theorem 6.2.** We start with

$$\frac{1}{\sigma^2} \left\| X - \frac{\tau(R^*)}{R^*} (X - PX) - \theta \right\|^2$$

$= \left\| (1 - \frac{\tau(R^*)}{R^*}) (Z - PZ) + P(Z - \eta) - (I_p - P) \eta \right\|^2$

$= \left(1 - \frac{\tau(R^*)}{R^*}\right)^2 R^* + \eta^T (I_p - P) \eta + [P(Z - \eta)]^T [P(Z - \eta)]$

$- 2 \left(1 - \frac{\tau(R^*)}{R^*} \right) \eta^T (I_p - P) (Z - PZ), \quad (A.19)$

since $P(I - P) = 0$.

Since $P$ is a symmetric idempotent matrix of rank $r$, by the Spectral Decomposition Theorem, $P = \sum_{i=1}^r \xi_i \xi_i^T$ where $\xi_1, \ldots, \xi_r$ are $(p \times 1)$ orthonormal vectors. Now, letting $H_i = \xi_i^T Z$ ($i = 1, \ldots, r$), and $\eta_i = \xi_i^T \eta$ ($i = 1, \ldots, r$), we have

$$[P(Z - \eta)]^T [P(Z - \eta)]$$

$$= \left[\sum_{i=1}^r \xi_i(H_i - \eta_i)^T \right]^T \left[\sum_{i=1}^r \xi_i(H_i - \eta_i) \right] = \sum_{i=1}^r (H_i - \eta_i)^2 \sim \chi_r^2. \quad (A.20)$$

Now write $\eta(I_p - P) \eta = \zeta^2$. The case $\zeta^2 = 0$ is straightforward. Assume $\zeta^2 > 0$. Note that $H_i \overset{ind}{\sim} N(\eta_i, 1)$ ($i = 1, \ldots, r$). Also, writing $I_p - P = \sum_{j=r+1}^p \xi_j \xi_j^T$, where the $\xi_j$ are orthonormal vectors with $\xi_{r+1} = (I_p - P) \eta/\zeta$, we can reexpress the right hand side of (A.19) as

$$\left[1 - \frac{\tau(R^*)}{R^*}\right]^2 R^* + \zeta^2 + \sum_{i=1}^r (H_i - \eta_i)^2 - 2 \left(1 - \frac{\tau(R^*)}{R^*}\right) \zeta H_{r+1},$$

where $H_{r+1} = \xi_{r+1}^T Z$. Also, $R^* = \sum_{j=r+1}^p H_j^2$, where $H_1, \ldots, H_r, H_{r+1}, \ldots, H_p$ are mutually independent with $H_{r+1} \sim N(\xi_{r+1}^T \eta, 1)$ and $H_{r+2}, \ldots, H_p$ are iid $N(0, 1)$. Now, repeating the proof of Theorem 5.1,

$$E \left[ \exp \left\{ \frac{-b}{\sigma^2} \left\| X - \frac{\tau(R^*)}{R^*} (X - PX) - \theta \right\|^2 \right\} \right]$$

$$= (2b + 1)^{-\frac{1}{2}} \sum_{j=0}^{\infty} \exp\{-\varphi_{**}\} \varphi_{**}^j \int_0^{+\infty} \exp \left\{ -t - b \left( b + \frac{1}{2} \right) \frac{\tau_0^2(t)}{t} + 2b \tau_0(t) \right\}$$

$$\times \left(1 - b \frac{\tau_0(t)}{t}\right)^{2j} \frac{t^{j + \frac{b-1}{2}}}{\Gamma(j + \frac{b-1}{2})} \, dt, \quad (A.21)$$

and $\varphi_{**} = (b + \frac{1}{2}) \eta^T (I_p - P) \eta$. This completes the proof. □
References