Predictive Density Estimation for Two Ordered Normal Means Under α-Divergence Loss

Yuan-Tsung Chang (Mejiro University)
Nobuo Shinozaki (Keio University)
William, E. Strawderman (Rutgers University)

Grant-in-Aid for Scientific Research (A) 20H00576 "Innovative Developments of Theories and Methodologies for Large Complex Data" Symposium, Dec. 7-9/2023 at Tsukuba

Abstract

When the underlying loss metric is α-divergence, $D(\alpha)$, loss introduced by siszàr (1967), we consider stochastic and Pitman closeness domination in predictive density estimation problems when there are restrictions given on two means, in Section 3 and 5, respectively. The underlying distributions considered are normal location-scale models, including the distribution of the observables, the distribution of the variable whose density is to be predicted, and the estimated predictive density which will be taken to be of the plug-in type. The scales may be known or unknown. We

1. first introduce a general expression which derived by Chang and Strawderman (2014) for the α-divergence loss in this set-up and show that it is a concave monotone function of quadratic loss, and also of the variances (predicand, and plug-in).
2. Next, we demonstrate $D(\alpha)$ stochastic domination (Pitman closeness) of certain plug-in predictive densities over others for the entire class of metrics simultaneously when "usual" stochastic domination (Pitman closeness) holds in the related problem of estimating the mean with respect to quadratic loss.
3. We also establish $D(\alpha)$ Pitman closeness results for certain generalized Bayesian (best invariant) predictive density estimators. Examples of $D(\alpha)$ stochastic (Pitman closeness) domination presented relate to the problem of estimating the predictive density of the variable with the restrictions on two normal means.

keywords: Predictive density, α-divergence, stochastic dominance, ordered normal means, Pitman closeness criterion

1 Introduction

We consider stochastic and Pitman closeness domination in predictive density estimation problems when the underlying loss metric is α-divergence $\{D(\alpha)\}$, a loss introduced by Csiszàr (1967).The underlying distributions considered are normal, including the distribution of the observables, the distribution of the variable whose density is to be predicted, and the estimated predictive density which will be taken to be of the plug-in type. We demonstrate $\{D(\alpha)\}$ stochastic and Pitman closeness domination of certain plug-in predictive densities over others for the entire class of metrics simultaneously when related stochastic and Pitman's closeness domination holds in the problem of estimating the mean. We also consider $\{D(\alpha)\}$ Pitman domination of certain generalized Bayesian (best invariant) procedures.

Examples of Pitman closeness domination presented relate to the problem of estimating the predictive density of the variable with the larger mean. More precisely, let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ be two independent random normal variables, where $\mu_1 \leq \mu_2$. Under the above restriction we wish to predict a normal population with mean equal to the larger mean, μ_2, and variance equal to σ^2, $\tilde{Y} \sim N(\mu_2, \sigma^2)$. We consider different versions of this problem, depending on whether the $\sigma_i^2, i = 1, 2$ are known or are unknown but satisfy the additional order restriction, $\sigma_1^2 \leq \sigma_2^2$. The case of two ordered normal means with known covariance matrix is also considered.

Kullback- Leibler (KL) loss, given by

$$D_{KL}\{\hat{p}(\tilde{y}|y), p(\tilde{y}|\psi)\} = \int p(\tilde{y}|\psi) \log \frac{p(\tilde{y}|\psi)}{\hat{p}(\tilde{y}|y)} d\tilde{y}, \tag{1}$$

where $p(\tilde{y}|\psi)$ is the true density to be estimated and $\hat{p}(\tilde{y}|y)$ is the estimated predictive density based on observing $Y = y$, where $Y \sim P(y|\psi)$, is the most studied among losses for the predictive density estimation problem.

The associated KL risk is defined as

$$R_{KL} = \int D_{KL}\{\hat{p}(\tilde{y}|y), p(\tilde{y}|\psi)\} p(y|\psi) dy,$$

where $p(y|\psi)$ is the density of y.

As pointed out in Maruyama and Strawderman (2010) KL loss is essentially contained in the class of $\alpha-$divergence losses (D_α introduced by Csiszàr (1967)) given by

$$D_\alpha\{\hat{p}(\tilde{y}|y), p(\tilde{y}|\psi)\} = \int f_\alpha\Big(\frac{\hat{p}(\tilde{y}|y)}{p(\tilde{y}|\psi)}\Big) p(\tilde{y}|\psi) d\tilde{y}, \tag{2}$$

where, for $-1 \leq \alpha \leq 1$

$$f_\alpha(z) = \begin{cases} \frac{4}{1-\alpha^2}(1 - z^{(1+\alpha)/2}), & |\alpha| < 1 \\ z \log z, & \alpha = 1 \\ -\log z, & \alpha = -1. \end{cases} \tag{3}$$

Here KL loss corresponds to $\alpha = -1$. The case $\alpha = 1$ is sometimes referred to as reverse KL loss.

Chang and Strawderman (2014) have derived the general form of D_α loss for the case of normal models and have shown that it is a concave monotone function of quadratic loss and is also a function of the variances (observed, predicand, and plug-in). This is reviewed in Section 2.

An alternative criterion to evaluate the goodness of estimators was introduced by Pitman (1937) as follows:

Let T_1 and T_2 be two estimators of θ. Then T_1 is closer to θ than T_2 (or T_1 is prefered to T_2) if Pitman nearness (PN) of T_1 compared to T_2

$$PN_\theta(T_1, T_2) = P\{|T_1 - \theta| < |T_2 - \theta|\} > 1/2.$$

For the case, when the estimators are equal with positive probability, Nayak (1990) modified Pitman's criterion as follows :

T_1 is said to be closer to θ than T_2 if

$$P\{|T_1 - \theta| < |T_2 - \theta|\} > \frac{1}{2}P\{T_1 \neq T_2\}.$$

Motived by Nayak (1990), Gupta and Singh (1992) defined the modified Pitman nearness (MPN) of T_1 compared to T_2. Setting

$$MPN_\theta(T_1, T_2) = P\{|T_1 - \theta| < |T_2 - \theta| | T_1 \neq T_2\} = \frac{P\{|T_1 - \theta| < |T_2 - \theta|, T_1 \neq T_2\}}{P\{T_1 \neq T_2\}},$$

T_1 is closer to θ than T_2 if $MPN_\theta(T_1, T_2) > 1/2$. Many works related to Pitman's criterion were published in the special issue of Communications in Statistics - Theory and Methods A20 (11) in 1992 and were unified in the monograph by Keating, Mason and Sen (1993).

In Section 4 modified Pitman closeness domination results for the estimation problems of ordered means with ordered variances are reviewed, which have been previously established by Chang and Shinozaki (2015) for a broader class of estimators. As noted above, we apply the result of the the D_α loss metric for plug-in predictive density estimates in normal models to obtain stochastic and Pitman closeness domination for plug-in predictive density estimates in Section 3 and 5 respectively. Section 6 considers $\{D(\alpha)\}$ Pitman closeness domination of the best invariant (generalized Bayes) predictive density estimator.

Here is a brief review of some of the relevant literature for the problem of estimating the mean.

Let

$$\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i, \quad s_i^2 = \sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2/(n_i - 1)$$

be the unbiased estimators of μ_i and σ_i^2, respectively, based on samples of size n_i from two normal populations, $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ respectively. When the means are equal (the common mean problem) $\mu_1 = \mu_2 = \mu$ and the variances are known, the UMVE of μ is

$$\hat{\mu} = \frac{n_1\sigma_2^2}{n_1\sigma_2^2 + n_2\sigma_1^2}\bar{X}_1 + \frac{n_2\sigma_1^2}{n_1\sigma_2^2 + n_2\sigma_1^2}\bar{X}_2.$$

When the variances are unknown, the unbiased estimator

$$\hat{\mu}^{GD} = \frac{n_1 s_2^2}{n_1 s_2^2 + n_2 s_1^2}\bar{X}_1 + \frac{n_2 s_1^2}{n_1 s_2^2 + n_2 s_1^2}\bar{X}_2$$

was proposed by Graybill and Deal (1959) and they gave a necessary and sufficient condition on n_1 and n_2 for $\hat{\mu}^{GD}$ to have a smaller variance than both $\bar{X}_1$ and $\bar{X}_2$.

When estimating the ordered means $\mu_1 \leq \mu_2$, Oono and Shinozaki (2005) proposed truncated estimators of $\mu_i, i = 1, 2$,

$$\hat{\mu}_1^{OS} = \min\{\bar{X}_1, \hat{\mu}^{GD}\}, \quad \hat{\mu}_2^{OS} = \max\{\bar{X}_2, \hat{\mu}^{GD}\}, \tag{4}$$

and showed that $\hat{\mu}_i^{OS}$ dominates the $\bar{X}_i$ in terms of MSE if and only if MSE of $\hat{\mu}^{GD}$ is not larger than that of $\bar{X}_i$ to estimate μ_i when $\mu_1 = \mu_2$.

When there are order restrictions given on both means and variances, $\mu_1 \leq \mu_2, \sigma_1^2 \leq \sigma_2^2$, Chang, Oono and Shinozaki (2012) have proposed

$$\hat{\mu}_1^{CS} = \begin{cases} \hat{\mu}_1^{OS}, & if\ s_1^2 \leq s_2^2 \\ \min\Big\{\bar{X}_1, \frac{n_1}{n_1+n_2}\bar{X}_1 + \frac{n_2}{n_1+n_2}\bar{X}_2\Big\}, & if\ s_1^2 > s_2^2 \end{cases} \tag{5}$$

and

$$\hat{\mu}_2^{CS} = \begin{cases} \hat{\mu}_2^{OS}, & if\ s_1^2 \le s_2^2 \\ \max\Big\{\bar{X}_2, \frac{n_1}{n_1+n_2}\bar{X}_1 + \frac{n_2}{n_1+n_2}\bar{X}_2\Big\}, & if\ s_1^2 > s_2^2. \end{cases} \tag{6}$$

They show that $\hat{\mu}_2^{CS}$ stochastically dominates $\hat{\mu}_2^{OS}$, but $\hat{\mu}_1^{CS}$ cannot dominate $\hat{\mu}_1^{OS}$ even in term of MSE when $\mu_2 - \mu_1$ is sufficiently large. We will show that that $\hat{\mu}_2^{CS}$ is Pitman closer to μ_2 than $\hat{\mu}_2^{OS}$ in Section 3.

When considering the estimation of ordered means of a normal distribution with a known covariance matrix, it has been recognized that the restricted MLEs do not always behave properly for general order restrictions and covariance matrices. See, for example, Lee (1981), Shinozaki and Chang (1999), Fernández et al. (2000) and Cohen and Sackrowitz (2004). Let $\boldsymbol{X}_i = (X_{1i}, X_{2i})', i = 1, \ldots, n$, be independent observations from $N(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu} = (\mu_1, \mu_2)'$, and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{7}$$

is a known covariance matrix. We assume that $|\rho| \neq 1$ and consider the estimation problem of $\mu_i, i = 1, 2$ when there is an order restriction, $\mu_1 \le \mu_2$. Using $\bar{X}_i = \sum_{i=1}^n X_i/n, i = 1, 2$, the restricted maximum likelihood estimators (MLE) of μ_1 and μ_2 are given as

$$\hat{\mu}_1^{MLE} = \bar{X}_1 - \beta(\bar{X}_1 - \bar{X}_2)^+ \ \text{ and } \ \hat{\mu}_2^{MLE} = \bar{X}_2 + \alpha(\bar{X}_1 - \bar{X}_2)^+, \tag{8}$$

where $\alpha = \omega_1/(\omega_1 + \omega_2)$ and $\beta = \omega_2/(\omega_1 + \omega_2)$ with $\omega_1 = \sigma_2^2 - \rho\sigma_1\sigma_2$, $\omega_2 = \sigma_1^2 - \rho\sigma_1\sigma_2$. We note that $\omega_1 + \omega_2 = \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2 > 0$ although ω_1 or ω_2 may be negative. Hwang and Peddada (1994) have proposed alternative estimators which are motivated by the case when a covariance matrix is diagonal. In our two-dimensional case the proposed estimators of μ_1 and μ_2 are given as

$$\hat{\mu}_1^{HP} = \min\left(\bar{X}_1, \alpha\bar{X}_1 + \beta\bar{X}_2\right) \quad \text{and} \quad \hat{\mu}_2^{HP} = \max\left(\bar{X}_2, \alpha\bar{X}_1 + \beta\bar{X}_2\right). \tag{9}$$

Clearly $\hat{\mu}_2^{HP} \ge \hat{\mu}_1^{HP}$ and Hwang and Peddada (1994) have shown that $\hat{\mu}_i^{HP}$ stochastically dominates the unrestricted MLE $\bar{X}_i, i = 1, 2$. Chang, Fukuda and Shinozaki (2017) have shown that $\hat{\mu}_i^{MLE}$ not only stochastically dominates $\hat{\mu}_i^{HP}$ but also dominates $\hat{\mu}_i^{HP}, i = 1, 2$ in the sense of Pitman closeness.

Broader reviews of statistical inference under order restrictions are given in Barlow et al. (1972), Robertson, Wright and Dykstra (1988) and the two monographs, Silvapulle and Sen (2004) and van Eeden (2006).

2 The Form of $D(\alpha)$ Loss for Normal Distributions

In this section we review the form of the $\{D(\alpha)\}$ loss when the density to be predicted and the predictive density estimate are both normal.

The aim here is to show that when $\hat{p}(\tilde{y}|y)$ and $p(\tilde{y}|\psi)$ are both normal distributions, then $D(\alpha)$ loss can be expressed as a concave monotone function of squared error loss.

Theorem 2.1. (Chang and Strawderman (2014), Theorem 2.1.) If the true density function of Y is $N(\mu, \sigma^2)$ and the estimated predictive density of Y, is $N(\hat{\mu}, \hat{\sigma}^2)$ then
a) for $-1 < \alpha < 1$,

$$D_\alpha(N(\tilde{y}|\hat{\mu}, \hat{\sigma}^2), N(\tilde{y}|\mu, \sigma^2)) = \frac{4}{1-\alpha^2}\Big(1 - d(\sigma^2, \hat{\sigma}^2)e^{-A(\sigma^2, \hat{\sigma}^2)\frac{(\hat{\mu}-\mu)^2}{2}}\Big), \tag{10}$$

where

$$d(\sigma^2, \hat{\sigma}^2) = \frac{\sigma^{(\alpha-1)/2}\tau}{\hat{\sigma}^{(\alpha+1)/2}}, \quad A(\sigma^2, \hat{\sigma}^2) = \Big(\frac{1-\alpha}{2\sigma^2}\Big)\Big(1 - \frac{(1-\alpha)\tau^2}{2\sigma^2}\Big) > 0, \qquad \frac{1}{\tau^2} = \Big(\frac{1+\alpha}{2\hat{\sigma}^2} + \frac{1-\alpha}{2\sigma^2}\Big).$$

Further, $d(\sigma^2, \hat{\sigma}^2) < 1$ and $A(\sigma^2, \hat{\sigma}^2) > 0$.

b) (Reverse KL)

$$D_{+1}(N(\tilde{y}|\hat{\mu}, \hat{\sigma}^2), N(\tilde{y}|\mu, \sigma^2)) = \frac{1}{2}\Big[\Big(\frac{\hat{\sigma}^2}{\sigma^2} - \log\frac{\hat{\sigma}^2}{\sigma^2} - 1\Big) + \frac{(\hat{\mu}-\mu)^2}{\sigma^2}\Big]. \tag{11}$$

c) (KL)

$$D_{-1}(N(\tilde{y}|\hat{\mu}, \hat{\sigma}^2), N(\tilde{y}|\mu, \sigma^2)) = \frac{1}{2}\Big[\Big(\frac{\sigma^2}{\hat{\sigma}^2} - \log\frac{\sigma^2}{\hat{\sigma}^2} - 1\Big) + \frac{(\hat{\mu}-\mu)^2}{\hat{\sigma}^2}\Big]. \tag{12}$$

Note : The first part of the RHS of (11) and (12) is a form of Stein's loss for estimating variances and the second part is the squared error loss $(\hat{\mu}-\mu)^2$ divided by either the true or estimated variance. Also note that in each case, the $\{D(\alpha)\}$ loss is a concave monotone function of squared error loss $|\hat{\mu}-\mu|^2$ and is also a function of the variances.

3 Stochastic Domination under the $D(\alpha)$ Loss Metric

In this section we will establish stochastic domination results under the $D(\alpha)$ loss metric for certain predictive density estimation problems involving two normal populations when the means are ordered. We handle the known and unknown variance cases in separate subsections. First we give a formal definition.

Definition 3.1. Given two predictive density estimates $\hat{f}_1(\tilde{y}|x)$ and $\hat{f}_2(\tilde{y}|x)$ of a density $f(\tilde{y}|\psi)$ based on data x from a distributions $X \sim g(X|\psi)$, $\psi \in \Omega$. $\hat{f}_2(\tilde{y}|x)$ stochastically dominate $\hat{f}_1(\tilde{y}|x)$ with respect to the $D(\alpha)$ metric, denoted , $\hat{f}_2 >_{SD(\alpha)} \hat{f}_1$, if $\forall\psi \in \Omega$ and $d \geq 0$,

$$P_\psi[D_\alpha(\hat{f}_2(\tilde{y}|x), f(\tilde{y}|\psi)) \leq d] \geq P_\psi[D_\alpha(\hat{f}_1(\tilde{y}|x), f(\tilde{y}|\psi)) \leq d],$$

with strict inequality for some d and ψ.

3.1 Results for the Known Variance Case

The data is

$$X_{ij} \sim N(\mu_i, \sigma_i^2), i = 1, 2, j = 1, \ldots, n_i.$$

with independent sufficient statistics

$$\bar{X}_i \sim N(\mu_i, \sigma_i^2/n_i), i = 1, 2.$$

where $\mu_1 \leq \mu_2$. The density we wish to predict is a future observation from some population with mean equal to the larger mean μ_2, i.e.

$$\tilde{Y} \sim N(\mu_2, \sigma^2),$$

where σ^2 is known. The following result due to Chang, Oono and Shinozaki (2012) shows stochastic dominance with respect to Euclidean metric of $\hat{\mu}_2^{MLE}$ over $\bar{X}_2$ in estimation of

μ_2. It is proved for completeness in Theorem 3.1 of of Chang and Strawderman (2014)(Appendix A.2).

Theorem 3.1. Suppose that

$$\hat{\mu}_i^{MLE} = \begin{cases} \bar{X}_i, & if\ \bar{X}_1 \le \bar{X}_2 \\ \frac{n_1\sigma_2^2}{n_1\sigma_2^2+n_2\sigma_1^2}\bar{X}_1 + \frac{n_2\sigma_1^2}{n_1\sigma_2^2+n_2\sigma_1^2}\bar{X}_2, & if\ \bar{X}_1 > \bar{X}_2, \end{cases} \quad i=1,2.$$

Then

$$P_{\mu_1,\mu_2}[|\hat{\mu}_i^{MLE} - \mu_i| \le d] \ge P_{\mu_1,\mu_2}[|\bar{X}_i - \mu_i| \le d],$$

for all $\mu_1 \le \mu_2$ and $d \ge 0$, i.e. $\hat{\mu}_i^{MLE}$ stochastically dominates $\bar{X}_i$ with respective to Euclidean metric.

We now consider comparison of plug-in estimators of the density of $\tilde{Y} \sim N(\mu_2, \sigma^2)$ of the form

$$\hat{f}_1(\tilde{y}|X_1, X_2) \sim N(\bar{X}_2, \nu^2)$$

and

$$\hat{f}_2(\tilde{y}|X_1, X_2) \sim N(\hat{\mu}_2^{MLE}, \nu^2),$$

where ν^2 is fixed (and not necessarily equal to σ^2). See Fourdrinier el al. (2011) for a discussion of why a choice of ν^2 different from (typically larger than) σ^2 is reasonable.

Our main result in this subsection is the following.

Theorem 3.2. $\hat{f}_2(\hat{y}|X_1, X_2) \sim N(\hat{\mu}_2^{MLE}, \nu^2)$ stochastically dominates $\hat{f}_1(\hat{y}|X_1, X_2) \sim N(\bar{X}_2, \nu^2)$ with respect to the $D(\alpha)$ metric in estimating the predictive density $f(\tilde{y}|\mu_2) \sim N(\mu_2, \sigma^2)$ for every $\alpha(-1 \le \alpha \le 1)$, σ^2 and ν^2.

Proof: In each case, $D_\alpha(\hat{f}_2(\tilde{y}|X_1, X_2), f(\tilde{y}|\mu_2))$ is a non-negative monotone increasing of function of $(\hat{\mu} - \mu)^2$ by Theorem 2.1. Hence stochastic D_α dominance following immediately form Theorem 3.1.

The analogous result for estimating a predictive density for a population with mean μ_1 follows immediately. Hence in the case of known variances, stochastic domination holds both for quadratic and D_α metrics for populations corresponding to either the larger or smaller mean. This will not be so when the variances are ordered and unknown as will be seen in the next sub-section.

3.2 The Unknown variance case

In this subsection we consider the following setup. The data is

$$X_{ij} \sim N(\mu_i, \sigma_i^2), i = 1, 2, j = 1, \ldots, n_i.$$

with independent sufficient statistics

$$\bar{X}_i \sim N(\mu_i, \sigma_i^2/n_i), \quad s_i^2 \sim \sigma_i^2\chi^2_{n_i-1}, i = 1, 2.$$

It is assumed that $\mu_1 \le \mu_2$ and also that $\sigma_1^2 \le \sigma_2^2$. We wish to estimate the density of a future observation from a normal population with mean μ_2 and variance $\sigma^2 = a\sigma_2^2$, where a is known, i.e. we wish to estimate the density

$$f(\tilde{y}) \sim N(\mu_2, a\sigma_2^2).$$

The assumption $\sigma_1^2 \leq \sigma_2^2$ is made, because to the best of our knowledge, no stochastic domination results are known unless an order restriction is placed on σ_1^2 and σ_2^2. We note however that Oono and Shinozaki(2006) and Chang, Oono and Shinozaki (2012) have MSE domination results for both cases.

The known stochastic domination result, due to Chang , Oono and Shinozaki(2012) compares the two estimators $\hat{\mu}_2^{OS}(4)$ and $\hat{\mu}_2^{CS}$

We will need the following lemma which follows from the proof of the main result of Chang, Oono and Shinozaki (2012). In that paper the statement of the result claims stochastic domination holds, but the proof actually demonstrates the stronger conditional result, which we will require due to the form of the $D(\alpha)$ loss proved in Theorem 2.1.

Lemma 3.3. $\hat{\mu}_2^{CS}$ stochastically dominates $\hat{\mu}_2^{OS}$ unconditionally and conditionally, i.e.

$$P[|\hat{\mu}_2^{CS} - \mu_2| \leq d|\bar{X}_1 - \bar{X}_2 = e, s_1^2, s_2^2] \geq P[|\hat{\mu}_2^{OS} - \mu_2| \leq d|\bar{X}_1 - \bar{X}_2 = e, s_1^2, s_2^2]$$

for all $\mu_1 \leq \mu_2$, $\sigma_1^2 \leq \sigma_2^2$ and $d \geq 0$, strict inequality holds for $e > 0, d > 0$, equality holds otherwise.

Note To the best of our knowledge, no stochastic domination result of the sort that $\hat{\mu}^{new}$ stochastically dominates $\bar{X}_2$ is known, in the unknown variance case, when $\mu_1 \leq \mu_2$ whether the variances are restricted or not.

The next result is the main result of this subsection.

Theorem 3.4. Suppose

$$\bar{X}_i \sim N(\mu_i, \sigma_i^2/n_i), \quad s_i^2 \sim \sigma_i^2 \chi^2_{n_i-1}, i = 1, 2.$$

are independent, where $\mu_1 \leq \mu_2$ and $\sigma_1^2 \leq \sigma_2^2$, and it is desired to estimate the density of a future independent variable $Y \sim N(\mu_2, a\sigma_2^2)$ where $a > 0$ is known. Then the plug-in predictive density estimate $\hat{f}_2(y) \sim N(\hat{\mu}_2^{CS}, \hat{\sigma}^2)$ stochastically dominates $\hat{f}_1(y) \sim N(\hat{\mu}_2^{OS}, \hat{\sigma}^2)$ under the $D(\alpha)$ metric for all $-1 \leq \alpha \leq 1$, for every $\hat{\sigma}^2$ which is a function of s_1^2, s_2^2 and $\bar{X}_1 - \bar{X}_2$.

Proof The proof follows immediately from Lemma 3.3 and Theorem 2.1 since by Theorem 2.1 the $D(\alpha)$ loss is a monotone non-decreasing function of $(\hat{\mu}_2 - \mu_2)^2$ for each fixed σ^2 and $\hat{\sigma}^2$.

4 Results for estimation of two ordered normal means with unknown but ordered variances under modified Pitman closeness criterion

In this section we consider the problem of estimating the ordered means of two normal distributions with unknown but ordered variances under modified Pitman closeness criterion. We show that in estimating the mean with larger variance, the proposed estimator, $\hat{\mu}_2^{CS}$, given in (6), is closer to true mean than the usual one, $\hat{\mu}_2^{OS}$, given in (4), which ignores the order restriction on variances. However, while in estimating the mean with smaller variance, the usual estimator, $\hat{\mu}_1^{OS}$, is not improved upon by $\hat{\mu}_1^{CS}$. We also discuss simultaneous estimation of two ordered means when the unknown variances are ordered.

First, we show that $\hat{\mu}_2^{CS}$ is Pitman closer to μ_2 than $\hat{\mu}_2^{OS}$ under modified Pitman closeness criterion. Actually, if we set

$$\gamma = \frac{n_1 s_2^2}{n_1 s_2^2 + n_2 s_1^2} \tag{13}$$

in the Theorem 2 of Chang and Shinozaki (2015) then we have following.

Theorem 4.1. The estimator $\hat{\mu}_2^{CS}$ is closer to μ_2 than $\hat{\mu}_2^{OS}$, i.e., for all $\mu_1 \leq \mu_2$ and $\sigma_1^2 \leq \sigma_2^2$,

$$MPN_{\mu_2}(\hat{\mu}_2^{CS}, \hat{\mu}_2^{OS}) \geq 1/2,$$

with strict inequality for some $\mu_1 \leq \mu_2$ and $\sigma_1^2 \leq \sigma_2^2$.

Next we consider estimating the mean μ_1, the mean with smaller variance; we show that $\hat{\mu}_1^{CS}$ can not be closer to μ_1 than $\hat{\mu}_1^{OS}$ when $\mu_2 - \mu_1$ is sufficiently large. Similarly, if we set γ as (13) in the Theorem 3 of Chang and Shinozaki (2015) then we have following.

Theorem 4.2. When $\mu_2 - \mu_1$ is sufficiently large, the estimator $\hat{\mu}_1^{CS}$ can not be closer to μ_1 than $\hat{\mu}_1^{OS}$, i.e.,

$$MPN_{\mu_1}(\hat{\mu}_1^{CS}, \hat{\mu}_1^{OS}) < 1/2.$$

Chang and Shinozaki (2015) have obtained a broader class of results including the above.

Although the estimator $\hat{\mu}_1^{CS}$ is not closer to μ_1 than $\hat{\mu}_1^{OS}$ when $\mu_2 - \mu_1$ is sufficiently large, in the simultaneous estimation problem with $\mu_1 \leq \mu_2$ when the variances are also ordered, the next theorem shows that if $n_1 \geq n_2$ then $\hat{\boldsymbol{\mu}}^{CS} = (\hat{\mu}_1^{CS}, \hat{\mu}_2^{CS})'$ improves upon $\hat{\boldsymbol{\mu}}^{OS} = (\hat{\mu}_1^{OS}, \hat{\mu}_2^{OS})'$ under Pitman closeness based on the sum of normalized squared errors

$$\sum_{i=1}^{2} (\hat{\mu}_i - \mu_i)^2/\sigma_i^2.$$

Theorem 4.3. If $n_1 \geq n_2$ then $\hat{\boldsymbol{\mu}}^{CS} = (\hat{\mu}_1^{CS}, \hat{\mu}_2^{CS})$ is closer to (μ_1, μ_2) than $\hat{\boldsymbol{\mu}}^{OS} = (\hat{\mu}_1^{OS}, \hat{\mu}_2^{OS})$ as

$$\begin{aligned} MPN_{\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}^{CS}, \hat{\boldsymbol{\mu}}^{OS}) &= \frac{P\{\sum_{i=1}^{2}(\hat{\mu}_i^{CS} - \mu_i)^2/\sigma_i^2 \leq \sum_{i=1}^{2}(\hat{\mu}_i^{OS} - \mu_i)^2/\sigma_i^2, \hat{\boldsymbol{\mu}}^{CS} \neq \hat{\boldsymbol{\mu}}^{OS}\}}{P\{\hat{\boldsymbol{\mu}}^{CS} \neq \hat{\boldsymbol{\mu}}^{OS}\}} \\ &> 1/2. \end{aligned} \tag{14}$$

5 Pitman closeness in predicting density function under the $D(\alpha)$ Loss Metric

In this section we will establish Pitman closeness results under the $\{D(\alpha)\}$ loss metric for certain predictive density estimation problems involving two normal populations when the means are ordered. We handle the known and unknown variance cases in separate subsections.

First we give a formal definition.

Definition 5.1. Given two predictive density estimates $\hat{f}_1(\tilde{y}|x)$ and $\hat{f}_2(\tilde{y}|x)$ of a density $f(\tilde{y}|\psi)$ based on data x from a distributions $X \sim g(X|\psi)$, $\psi \in \Omega$, $\hat{f}_2(\tilde{y}|x)$ is closer to

$f(\tilde{y}|\psi)$ than $\hat{f}_1(\tilde{y}|x)$ with respect to the $D(\alpha)$ metric under the modified Pitman closeness criterion, if $\forall\psi \in \Omega$,

$$P_\psi\{D_\alpha(\hat{f}_2(\tilde{y}|x), f(\tilde{y}|\psi)) < D_\alpha(\hat{f}_1(\tilde{y}|x), f(\tilde{y}|\psi))|\hat{f}_2(\tilde{y}|x) \neq \hat{f}_1(\tilde{y}|x)\} \geq 1/2,$$

with strict inequality for some $\psi \in \Omega$.

We first consider the case when variances are known.

5.1 Case when variances are known

The data are

$$X_{ij} \sim N(\mu_i, \sigma_i^2),\ i = 1, 2,\ j = 1, \ldots, n_i \tag{15}$$

with independent sufficient statistics

$$\bar{X}_i \sim N(\mu_i, \sigma_i^2/n_i),\ i = 1, 2, \tag{16}$$

where $\mu_1 \leq \mu_2$.

First we show that the MLEs are closer to μ_i than the sample means under modified Pitman closeness criterion.

Theorem 5.1. The MLE of μ_i is

$$\hat{\mu}_i^{MLE} = \begin{cases} \bar{X}_i, & if\ \bar{X}_1 \leq \bar{X}_2 \\ \frac{n_1\sigma_2^2}{n_1\sigma_2^2+n_2\sigma_1^2}\bar{X}_1 + \frac{n_2\sigma_1^2}{n_1\sigma_2^2+n_2\sigma_1^2}\bar{X}_2, & if\ \bar{X}_1 > \bar{X}_2. \end{cases}$$

Then

$\hat{\mu}_i^{MLE}$ is Pitman closer to μ_i than $\bar{X}_i$, i.e.

$$P_{\mu_1,\mu_2}\{|\hat{\mu}_i^{MLE} - \mu_i| < |\bar{X}_i - \mu_i|\, |\hat{\mu}_i^{MLE} \neq \bar{X}_i\} \geq 1/2$$

for all $\mu_1 \leq \mu_2$, with strict inequality for some $\mu_1 \leq \mu_2$.

We wish to predict the density of a future observation from some population with mean equal to the larger mean μ_2, i.e.

$$\tilde{Y} \sim N(\mu_2, \sigma^2),$$

where σ^2 is known. We now consider comparison of plug-in estimators of the density of $\tilde{Y} \sim N(\mu_2, \sigma^2)$ of the form

$$\hat{f}_1(\tilde{y}|X_1, X_2) \sim N(\bar{X}_2, \nu^2)$$

and

$$\hat{f}_2(\tilde{y}|X_1, X_2) \sim N(\hat{\mu}_2^{MLE}, \nu^2),$$

where ν^2 is fixed (and not necessarily equal to σ^2). See Fourdrinier el al. (2011)for a discussion of why a choice of ν^2 different from (typically larger than) σ^2 is reasonable. Our main result in this subsection is the following.

Theorem 5.2. In estimating the predictive density $f(\tilde{y}|\mu_i) \sim N(\mu_i, \sigma^2)$, the density $\hat{f}_2(\tilde{y}|X_1, X_2) \sim N(\hat{\mu}_i^{MLE}, \nu^2)$ is Pitman closer to the predictive density $f(\tilde{y}|\mu_i) \sim N(\mu_i, \sigma^2)$ than $\hat{f}_1(\tilde{y}|X_1, X_2) \sim N(\bar{X}_i, \nu^2)$ with respect to the $\{D(\alpha)\}$ metric for every $\alpha(-1 \leq \alpha \leq 1)$

, σ^2 and ν^2.

Hence in the case of known variances, Pitman closeness domination holds both for mean estimation and prediction under D_α metrics for populations corresponding to either the larger or smaller mean.

Proof. The proof follows immediately from Theorem 2.1 since

$$\begin{aligned} & D_\alpha(N(\hat{\mu}_i^{MLE}, \nu^2), N(\mu_i, \sigma^2)) < D_\alpha(N(\bar{X}_i, \nu^2), N(\mu_i, \sigma^2)) \\ \Leftrightarrow \quad & (\hat{\mu}_i^{MLE} - \mu_i)^2 < (\bar{X}_i - \mu_i)^2 \\ \Leftrightarrow \quad & |\hat{\mu}_i^{MLE} - \mu_i| < |\bar{X}_i - \mu_i|. \end{aligned}$$

From Theorem 5.1 we have

$$P\{|\hat{\mu}_i^{MLE} - \mu_i| < |\bar{X}_i - \mu_i| \, |\hat{\mu}_i^{MLE} \neq \bar{X}_i\} \geq 1/2.$$

This completes the proof.

In the next section we extend the above results to two ordered means when a covariance matrix is known.

5.2 Case when the covariance matrix is known

In this section we consider the case when two normal means are ordered and covariance matrix defined in (7) is known. We show that plug-in predictive density with $\hat{\mu}_i^{MLE}$, (8), is Pitman closer to the true predictive density than plug-in predictive density with $\hat{\mu}_i^{HP}$, (9), under D_α loss. The following result from Chang, Fukuda and Shinozaki (2017) is the basis for our study.

Theorem 5.3. (Chang, Fukuda and Shinozaki (2017), Theorem 3.1) $\hat{\mu}_i^{MLE}$ is Pitman closer to μ_i than $\hat{\mu}_i^{HP}, i = 1, 2$.

Based on the above theorem we have the following main result.

Theorem 5.4. In estimating the predictive density $f_i(\tilde{y}|\mu_i) \sim N(\mu_i, \sigma^2)$, the density $\hat{f}_i^{MLE}(\tilde{y}|X_1, X_2) \sim N(\hat{\mu}_i^{MLE}, \nu^2)$ is Pitman closer to the true predictive density $f_i(\tilde{y}|\mu_i) \sim N(\mu_i, \sigma^2)$ than $\hat{f}_i^{HP}(\tilde{y}|X_1, X_2) \sim N(\hat{\mu}_i^{HP}, \nu^2)$ with respect to the $\{D(\alpha)\}$ metric for every $\alpha(-1 \leq \alpha \leq 1)$, σ^2 and ν^2.

Proof. This follows directly from Theorem 5.3, since from Theorem 2.1, $D_\alpha(N(\tilde{y}|\hat{\mu}, \hat{\sigma}^2)$, $N(\tilde{y}|\mu, \sigma^2))$ is monotone function of $|\hat{\mu} - \mu|$, where $\hat{\mu}$ is an estimator of μ.

Next we consider the cases of unknown variances.

5.3 Case when variances are unknown and unrestricted

In this subsection we consider the same setup as (15) and (16). It is assumed that $\mu_1 \leq \mu_2$ and that no restriction is given on unknown σ_i^2. It is shown that plug-in predictive density with $\hat{\mu}_i^{OS}$, (4), is Pitman closer to the true predictive density than plug-in predictive density with $\bar{X}_i$ under D_α loss.

From the result of Chang and Shinozaki (2015), we note that with respect to modified Pitman criterion, the most critical case for $\hat{\mu}_i^{OS}$ to be closer to μ_i than $\bar{X}_i$ is the case when $\mu_1 = \mu_2 = \mu$. Surprisingly, this result reduces the dominance problem in estimating two ordered means to that in estimating the common mean, that is $\hat{\mu}_i^{OS}$ improves upon $\bar{X}_i$ under modified Pitman closeness criterion if and only if $\hat{\mu}^{GD}$ is closer to μ than $\bar{X}_i$ under Pitman closeness criterion. Kubokawa (1989) has given a sufficient condition on $n_i (n_i \geq 5)$ so that $\hat{\mu}^{GD}$ is closer to μ than both $\bar{X}_1$ and $\bar{X}_2$.

As matter of fact, if we set $\gamma = n_1 S_2^2/(n_1 S_2^2 + n_2 S_1^2)$ as (3.1) in the Theorem 1 of Chang and Shinozaki (2015) then we have the following.

Theorem 5.5. $MPN_{\mu_i}(\hat{\mu}_i^{OS}, \bar{X}_i) \geq 1/2$ for all $\mu_1 \leq \mu_2$ and for all σ_1^2 and σ_2^2 if and only if for all σ_1^2 and σ_2^2, $PN_{\mu}(\hat{\mu}^{GD}, \bar{X}_i) \geq 1/2$ when $\mu_1 = \mu_2 = \mu$.

We wish to predict the density of a future observation from a normal population with mean μ_i and unknown variance σ^2, i.e. we wish to predict the density

$$f(\tilde{y}) \sim N(\mu_i, \sigma^2).$$

Let

$$\bar{X}_i \sim N(\mu_i, \sigma_i^2/n_i), \quad S_i^2 \sim \sigma_i^2 \chi^2_{n_i - 1}, \; i = 1, 2 \tag{17}$$

are independent, where $\mu_1 \leq \mu_2$ and it is desired to predict the density of a future independent variable $Y_i \sim N(\mu_i, a\sigma^2), i = 1, 2$, where $a > 0$ is known.

Based on the above Theorem 5.5 we have the following main result.

Theorem 5.6. The plug-in predictive density estimate $\hat{f}_i^{OS}(\tilde{y}) \sim N(\hat{\mu}_i^{OS}, \widehat{a\sigma^2})$ is Pitman closer to $f(\tilde{y}|\mu_i, a\sigma^2)$ than $\hat{f}_i^{\bar{X}_i}(\tilde{y}) \sim N(\bar{X}_i, \widehat{a\sigma^2})$ for all $\mu_1 \leq \mu_2$ and $\sigma_i^2, i = 1, 2$ under the $D(\alpha)$ metric for all $-1 \leq \alpha \leq 1$ and every estimator $\widehat{\sigma^2}$ if and only if $\hat{\mu}^{GD}$ is Pitman closer to μ than $\bar{X}_i$ for all σ_1^2 and σ_2^2 when $\mu_1 = \mu_2 = \mu$.

5.4 Case when variances are ordered

In this subsection we consider the same setup as (17). We give Pitman closeness domination results in predictive density estimation when unknown variances σ_1^2 and σ_2^2 satisfy the order restriction $\sigma_1^2 \leq \sigma_2^2$. We note, however, that Oono and Shinozaki (2005) and Chang, Oono and Shinozaki (2012) have MSE domination results in estimating means when an order restriction on σ_1^2 and σ_2^2 is present or absent.

First we estimate the density of a future observation from a normal population with mean μ_2 and variance $\sigma^2 = a\sigma_2^2$, where a is known, i.e. we estimate the density

$$f(\tilde{y}) \sim N(\mu_2, a\sigma_2^2).$$

Based on the Theorem 4.1 and Theorem 2.1 we have the following.

Theorem 5.7. The plug-in predictive density estimate $\hat{f}^{CS}(\tilde{y}) \sim N(\hat{\mu}_2^{CS}, \widehat{a\sigma_2^2})$ is Pitman closer to $f(\tilde{y}|\mu_2, a\sigma_2^2)$ than $\hat{f}^{OS}(\tilde{y}) \sim N(\hat{\mu}_2^{OS}, \widehat{a\sigma_2^2})$ under the $D(\alpha)$ metric for all $-1 \leq \alpha \leq 1$ and for any estimator $\widehat{\sigma_2^2}$.

Next we consider estimating the predictive density with smaller variance σ_1^2, $N(\mu_1, a\sigma_1^2)$. Based on the Theorem 4.2 and Theorem 2.1 we have the following.

Theorem 5.8. The plug-in predictive density estimate $\hat{f}^{CS}(\tilde{y}) \sim N(\hat{\mu}_1^{CS}, a\widehat{\sigma}_1^2)$ can not be Pitman closer to $f(\tilde{y}|\mu_1, a\sigma_1^2)$ than $\hat{f}^{OS}(\tilde{y}) \sim N(\hat{\mu}_1^{OS}, a\widehat{\sigma_1^2})$ when $\mu_2 - \mu_1$ is sufficiently large, under the $\{D(\alpha)\}$ metric for all $-1 \leq \alpha \leq 1$ and for any estimator $\widehat{\sigma_1^2}$.

Proof. The proof follows immediately from Theorem 4.2.

Finally, we consider estimation of the predictive density function $p(\boldsymbol{y}|\boldsymbol{\mu}, \Sigma) = N\Big(\boldsymbol{\mu}, \Sigma\Big)$ and show that $N\Big(\hat{\boldsymbol{\mu}}^{CS}, \hat{\Sigma}\Big)$ dominates $N\Big(\hat{\boldsymbol{\mu}}^{OS}, \hat{\Sigma}\Big)$ in terms of Pitman closeness under reverse Kullback-Leibler loss D_{+1}, where

$$\boldsymbol{\mu} = (\mu_1, \mu_2)', \;\; \Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \text{ and } \hat{\Sigma} = \begin{pmatrix} \hat{\sigma}_1^2 & 0 \\ 0 & \hat{\sigma}_2^2 \end{pmatrix}.$$

We need the following Lemma.

Lemma 5.1. The reverse Kullback-Leibler loss D_{+1} when we predict $p(\tilde{\boldsymbol{y}}|\boldsymbol{\mu}, \Sigma) = N\Big(\boldsymbol{\mu}, \Sigma\Big)$ by $p(\tilde{\boldsymbol{y}}|\hat{\boldsymbol{\mu}}, \hat{\Sigma}) = N\Big(\hat{\boldsymbol{\mu}}, \hat{\Sigma}\Big)$ is given as

$$D_{+1}\Big(N\Big(\hat{\boldsymbol{\mu}}, \hat{\Sigma}\Big), N\Big(\boldsymbol{\mu}, \Sigma\Big)\Big) = 1/2\Bigg[\sum_{i=1}^{2}\Big(\frac{\hat{\sigma}_i^2}{\sigma_i^2} - \log\frac{\hat{\sigma}_i^2}{\sigma_i^2} - 1 + \frac{(\mu_i - \hat{\mu}_i)^2}{\sigma_i^2}\Big)\Bigg].$$

Proof. Straightforward calculation.

Theorem 5.9. If $n_1 \geq n_2$ then $\hat{f}^{CS}(\tilde{\mathrm{y}}) \sim N\Big(\hat{\boldsymbol{\mu}}^{CS}, \hat{\Sigma}\Big)$ dominates $\hat{f}^{OS}(\tilde{\mathrm{y}}) \sim N\Big(\hat{\boldsymbol{\mu}}^{OS}, \hat{\Sigma}\Big)$ in terms of Pitman closeness under reverse Kullback-Leibler metric.

Proof. From Lemma 5.1 and Theorem 4.3 the result follows.

6 Extension to generalized Bayesian predictive densities

In this section we discuss improving the generalized Bayesian predictive densities suggested by Corcuera and Giummole (1999) under $D(\alpha)$ loss.

Based on the data

$$X_{ij} \sim N(\mu_i, \sigma_i^2), i = 1, 2, \; j = 1, \cdots, n_i,$$

we predict the density $\tilde{Y} \sim N(\mu_i, \sigma_i^2), i = 1, 2.$ We denote its density function by $p(\tilde{y}; \mu_i, \sigma_i)$, where μ_i and σ_i^2 are unknown.

When $-1 \leq \alpha < 1$, Corcuera and Giummole (1999) have established that the best invariant predictive density of $p(\tilde{y}; \mu_i, \sigma_i)$ based solely on $x_{i1}, \cdots x_{in_i}$ is

$$\hat{p}_\alpha(\tilde{y}; \bar{x}_i, \tilde{\sigma}_i) \propto \left[1 + \frac{1-\alpha}{2n_i + 1 - \alpha}\left(\frac{y - \bar{x}_i}{\tilde{\sigma}_i}\right)^2\right]^{-(2n_i - 1 - \alpha)/2(1-\alpha)}, \tag{18}$$

where $\bar{x}_i$ is the sample mean and $\tilde{\sigma}_i^2 = ((n_i - 1)/n_i)s_i^2$ is the sample variance. Corcuera and Giummole (1999) have also shown that $\hat{p}_\alpha(\tilde{y}; \bar{x}_i, \tilde{\sigma}_i)$ is the generalized Bayesian predictive density for the prior density $f(\mu_i, \sigma_i) \propto 1/\sigma_i, 0 < \sigma_i < \infty$. It is to be noted that

$\hat{p}_\alpha(\tilde{y};\bar{x}_i,\tilde{\sigma}_i)$ is not a normal distribution, although the plug-in density $N(\bar{x}_i,s_i^2)$ is the generalized Bayes rule when $\alpha=1$.

We consider the following two cases separately where order restrictions on μ_i and/or σ_i^2 are present,

i) Case when $\mu_1 \leq \mu_2$.

ii) Case when $\mu_1 \leq \mu_2$ and $\sigma_1^2 \leq \sigma_2^2$.

We consider to improve $\hat{p}_\alpha(\tilde{y};\bar{x}_i,\tilde{\sigma}_i)$ or $\hat{p}_\alpha(\tilde{y};\hat{\mu}_i^{OS},\tilde{\sigma}_i)$ by replacing $\bar{x}_i$ with $\hat{\mu}_i^{OS}$ or $\hat{\mu}_i^{OS}$ with $\hat{\mu}_i^{CS}$, respectively.

The next lemma is usefully for improving the generalized Bayesian predictive densities (18). We give its proof for completeness.

Lemma 6.1. Let $f(\cdot)$ be the probability density function of $X \sim N(0,\tau^2)$. Assume that $g(t) \geq 0$ is symmetric about the origin and is a strictly decreasing function of $|t|$ such that $\int_{-\infty}^{\infty} g(x)f(x)dx < \infty$. Then

$$\int_{-\infty}^{\infty} g(y-x)f(y-\mu)dy$$

is a strictly decreasing function of $|x-\mu|$.

Proof. By making the transformation $z=y-\mu$ we see that

$$\int_{-\infty}^{\infty} g(y-x)f(y-\mu)dy = \int_{-\infty}^{\infty} g(z-v)f(z)dz = h(v),$$

where $v=x-\mu$. Then $h(v)$ satisfies

i) $h(v)=h(-v)$. (Since f and g are symmetric about the origin.)

ii) $h(v)$ is a strictly decreasing function of $|v|$.

We prove *ii)* here. We need only to show that $h(v)-h(v+\Delta)>0$ for any $v \geq 0$ and for any $\Delta > 0$. We have

$$h(v)-h(v+\Delta) = \int_{-\infty}^{\infty} k(z;v,\Delta)f(z)dz,$$

where

$$k(z;v,\Delta) = g(z-v) - g(z-v-\Delta).$$

We notice that $k(z;v,\Delta)$ satisfies

1) $k(v+\Delta/2;v,\Delta)=0$.

2) When $z > v+\Delta/2$, $k(z;v,\Delta) < 0$.

3) When $z < v+\Delta/2$, $k(z;v,\Delta) > 0$.

4) $k(v+\Delta/2+(z-v-\Delta/2;v,\Delta) = -k(v+\Delta/2-(z-v-\Delta/2;v,\Delta)$.

Thus we see that $h(v) - h(v+\Delta) > 0$.

Note: Lemma 6.1 can be generalized to p dimensional case. See Lemma (A.6) of Fourdrinier, Strawderman and Wells (2018), which is an extension of Anderson's Theorem due to Chou and Strawderman (1990).

Let $\hat{\mu}_i$ denote an estimator of $\mu_i, i = 1, 2$ in general. Now we show that for any $1 \le \alpha < 1$, $D_\alpha(\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i))$ is a strictly increasing function of $|\hat{\mu}_i - \mu_i|$.

From Lemma 6.1, we see that for $|\alpha| < 1$,

$$D_\alpha(\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i)) \propto 1 - \int_\infty^\infty g(\tilde{y} - \hat{\mu}_i) f(\tilde{y} - \mu_i) d\tilde{y}$$

is a strictly increasing function of $|\hat{\mu}_i - \mu_i|$, where

$$g(y - x) = \left[1 + \frac{1-\alpha}{2n_i + 1 - \alpha}\left(\frac{y - x}{\hat{\sigma}_i}\right)^2\right]^{-(2n_i - 1 - \alpha)(1+\alpha)/4(1-\alpha)}$$

and

$$f(y - \mu) \propto \exp\left\{-\frac{(1-\alpha)(y-\mu)^2}{4\sigma^2}\right\}.$$

For $\alpha = -1$, in order to show that

$$D_{-1}(\hat{p}_{-1}(\tilde{y}; \hat{\mu}_i, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i)) = -E_{\tilde{y}}\left\{\log\left[\frac{\hat{p}_{-1}(\tilde{y}; \hat{\mu}_i, \hat{\sigma}_i)}{p(\tilde{y}; \mu_i, \sigma_i)}\right]\right\}$$

is a strictly increasing function of $|\hat{\mu}_i - \mu_i|$, we need only to notice that

$$\begin{aligned} &\int_{-\infty}^\infty \log\left[1 + \frac{1}{n_i + 1}\left(\frac{\tilde{y} - \hat{\mu}_i}{\hat{\sigma}_i}\right)^2\right] \exp\left\{-\frac{(\tilde{y} - \mu_i)^2}{2\sigma^2}\right\} d\tilde{y} \\ = &\int_{-\infty}^\infty \log\left[1 + \frac{1}{n_i + 1}\left(\frac{z - v}{\hat{\sigma}_i}\right)^2\right] \exp\left\{-\frac{z^2}{2\sigma^2}\right\} dz \end{aligned}$$

is a strictly increasing function of $v = |\hat{\mu}_i - \mu_i|$ from Lemma 6.1.

6.1 Case when $\mu_1 \le \mu_2$

In this case we have the following result.

Theorem 6.1. The predictive density estimate $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i^{OS}, \hat{\sigma}_i), i = 1, 2$ is closer to the predictive density $p(\tilde{y}; \mu_i, \sigma_i)$ than $\hat{p}_\alpha(\tilde{y}; \bar{x}_i, \hat{\sigma}_i)$, respectively, under the $\{D(\alpha)\}$ metric for all $-1 \le \alpha < 1$ and for every estimator $\hat{\sigma}_i$ if and only if $\hat{\mu}^{GD}$ is Pitman closer to μ than $\bar{X}_i$ for all σ_1^2 and σ_2^2 when $\mu_1 = \mu_2 = \mu$.

Proof. Let $\hat{\mu}_i$ denote an estimator of $\mu_i, i = 1, 2$ in general. Since $D_\alpha(\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i))$ is a strictly increasing function of $|\hat{\mu}_i - \mu_i|$, we see that

$$D_\alpha(\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i^{OS}, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i)) < D_\alpha(\hat{p}_\alpha(\tilde{y}; \bar{x}_i, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i))$$

if and only if

$$|\hat{\mu}_i^{OS} - \mu_i| < |\bar{x}_i - \mu_i|.$$

Thus from Theorem 5.5 we have the desired result.

6.2 Case when $\mu_1 \le \mu_2$ and $\sigma_1^2 \le \sigma_2^2$

Here we give domination results in predictive density estimation when unknown variances σ_1^2 and σ_2^2 satisfy the order restriction $\sigma_1^2 \le \sigma_2^2$. Then we have following results.

Theorem 6.2. The predictive density estimate $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_2^{CS}, \hat{\sigma}_2)$ is closer to the predictive density $p(\tilde{y}; \mu_2, \sigma_2)$ than $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_2^{OS}, \hat{\sigma}_2)$ under the $\{D(\alpha)\}$ metric for all $-1 \le \alpha < 1$ and for every estimator $\hat{\sigma}_2^2$.

Theorem 6.3. The predictive density estimate $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_1^{CS}, \hat{\sigma}_1)$ is not Pitman closer to $p(\tilde{y}; \mu_1, s_1)$ than $\hat{p}_\alpha(\tilde{y}; \hat{\mu}_1^{OS}, \hat{\sigma}_1)$ when $\mu_2 - \mu_1$ is sufficiently large, under the $\{D(\alpha)\}$ metric for all $-1 \le \alpha < 1$ and for any estimator $\hat{\sigma}_1^2$.

Proof of Theorems 6.2 and 6.3. Since for $-1 \le \alpha < 1$

$$D_\alpha(\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i^{CS}, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i)) < D_\alpha(\hat{p}_\alpha(\tilde{y}; \hat{\mu}_i^{OS}, \hat{\sigma}_i), p(\tilde{y}; \mu_i, \sigma_i))$$

if and only if

$$|\hat{\mu}_i^{CS} - \mu_i| < |\hat{\mu}_i^{OS} - \mu_i|,$$

from Theorems 4.1 and 4.2, Theorems 6.2 and 6.3 are established, respectively.