COMBINING COORDINATES IN SIMULTANEOUS ESTIMATION OF NORMAL MEANS*

by

James O. Berger
Purdue University

and

Dipak K. Dey
University of Kentucky

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #80-32

November 1980

*Research supported by the National Science Foundation, Grant MCS 78-02300.

ABSTRACT

The problem of combining coordinates in Stein-type estimators, when simultaneously estimating normal means, is considered. The question of deciding whether to use all coordinates in one combined shrinkage estimator or to separate into groups and use separate shrinkage estimators on each group is considered. A Bayesian view point is (of necessity) taken, and it is shown that the "combined" estimator is, somewhat surprisingly, often superior.

AMS 1970 subject classification: Primary 62 F 15; Secondary 62 F 10, 62 F 23.

Key Words and Phrases: Shrinkage estimator, Robust Generalized Bayes estimator, Minimax estimator, Bayes risk, Normal prior, Flat prior.

1. INTRODUCTION AND SUMMARY

Let $X = (X_1,\ldots,X_k)^t$ have a k-variate normal distribution with mean vector $\theta = (\theta_1,\ldots,\theta_k)^t$ and known positive definite covariance matrix Σ. It is desired to estimate θ, using an estimator $\delta(X) = (\delta_1(X),\ldots,\delta_k(X))^t$, under a quadratic loss

$$L(\theta,\delta) = (\theta-\delta)^t Q(\theta-\delta),$$

Q being a known positive definite matrix. As usual we will evaluate an estimator δ in terms of its risk function (i.e., expected loss)

$$R(\theta,\delta) = E_\theta L(\theta,\delta(X)).$$

Stein (1955) showed that (for $Q = \Sigma = I_k$) the usual estimator $\delta^\circ(X) = X$ is inadmissible for $k \geq 3$. Estimators improving upon δ° for the above general case have been found by a variety of authors. (See Berger (1980a and 1980b) for references.) In actually selecting an alternative to δ°, however, it was shown in Berger (1980a) that prior information concerning θ must be taken into account (to choose the region of the parameter space in which significant improvement over δ° is to be obtained). In Berger (1980a) a robust generalized Bayes estimator was developed using, as inputs, a "prior mean" μ and a "prior covariance matrix" A. (Alternatively, μ and A can be thought of as specifying an ellipse in the parameter space in which significant improvement over δ° is desired.) The estimator is given by

$$\delta^{RB}(X) = \left(I_k - \frac{r((X-\mu)^t(\Sigma+A)^{-1}(X-\mu))}{(X-\mu)^t(\Sigma+A)^{-1}(X-\mu)}\,\Sigma\,(\Sigma+A)^{-1}\right)(X-\mu)+\mu, \tag{1.1}$$

where r is a certain increasing function which can be reasonably approximated by

$$r(z) = \min\{k-2, z\}.$$

This estimator was shown to be a very attractive alternative to δ°, but unfortunately is not necessarily uniformly better than δ° (and hence not minimax). For those desiring uniform dominance of δ°, a minimax Bayes estimator δ^{MB} (to be described later) was developed in Berger (1980b).

It should be noted that the above estimators, and hence the ensuing discussion, are concerned with the situation in which μ and A must be determined subjectively. In many "empirical Bayes" types of problems where the prior information about θ includes beliefs in certain relationships, such as exchangability, among the θ_i, the observation X can be used to estimate μ and A or certain facets of them. The estimators δ^{RB} and δ^{MB} are probably inadequate in such situations.

It was stated in Berger (1980b) that it might prove better, in certain situations, to divide the coordinates into groups and use separate estimators of the form δ^{RB} on each group, rather than combine all coordinates together as done in (1.1). This might prove superior, for example, if the coordinates fall naturally into two groups (say, right and left handed baseball players). In this paper, we investigate such a possibility for the estimators δ^{RB} and δ^{MB}.

To our knowledge, the only previous work on "combining of coordinates" is that of Efron and Morris (1973b). They considered use of the symmetric Stein estimator when the θ_i could be divided into two groups, suspected of having substantially different prior variances. Their conclusion was essestially that it is better to use separate Stein estimators for each group.

The setting is an empirical Bayes one, however, in that, one can hope to obtain fairly reasonable estimates of the prior variances of each group of θ_i. It is thus natural that the separate group Stein estimators (which are the appropriate empirical Bayes estimators) should out perform the overall Stein estimator. Since, in this paper, we are considering the situation in which no "empirical Bayes" type of knowledge is available, the results of Efron and Morris clearly do not apply.

The following notation will be used throughout the paper. Suppose that θ is divided into s groups of sizes $k_1,\ldots,k_s$ ($k_i \geq 3$, i=1,...,s), where if necessary the coordinates are relabeled so that the $\ell^{\underline{th}}$ group is given by (defining $k_o = 0$)

$$\theta_{(\ell)} = (\theta_{k_o+\ldots+k_{\ell-1}+1},\ldots\ldots,\theta_{k_o+\ldots+k_\ell})^t.$$

Similarly define $X_{(\ell)}$, and let $\delta^{RB}_{(\ell)}$ and $\delta^{MB}_{(\ell)}$ be the appropriate estimator for estimating $\theta_{(\ell)}$ based solely on $X_{(\ell)}$. The estimator δ^{RB} or δ^{MB} will be called the "combined estimator", while $\delta^{RBS} = (\delta^{RB}_{(1)},\ldots,\delta^{RB}_{(s)})^t$ or $\delta^{MBS} = (\delta^{MB}_{(1)},\ldots,\delta^{MB}_{(s)})^t$ will be called the "separate estimator".

Typically the combined estimator and the separate estimator will have risk functions which cross, so they can only be compared in terms of an average risk or Bayes risk. Since μ and A can be viewed as prior inputs, it is natural to incorporate μ and A into a prior distribution π, and then evaluate the combined and separate estimators in terms of Bayes risk

$$r(\pi,\delta) = E^{\pi}[R(\theta,\delta)].$$

The obvious problem is to choose a suitable π. Note first of all that δ^{RB} was developed as a generalized Bayes rule with respect to a prior π_o which incorporates μ and A. The prior π_o was prejudiced in favor of "combining", however, being of a form which imposed considerable dependence on the coordinates of θ. We will assume that Q, $\ddagger$ and A are block diagonal with $\ell^{\underline{th}}$ blocks $Q_{(\ell)}$, $\ddagger_{(\ell)}$ and $A_{(\ell)}$ (all $k_\ell \times k_\ell$ matrices) respectively, so that the groups of coordinates are actually roughly "independent". In comparing the combined and the separate estimators, therefore, we should choose π to reflect this independence. It is crucial to realize that π is, in a sense, an artifact, the only "true" prior information being μ and A. Thus we will be concerned with evaluation for a wide range of functional forms for π.

In section 2 it will be assumed that π is $N(\mu,A)$. This allows explicit calculation of the Bayes risks of the combined and separate estimators, and easy comparison of the Bayes risks.

In section 3, various flat tailed π will be considered. Monte Carlo simulation and asymptotics ($k \to \infty$) will be used to compare the Bayes risks of the combined and separate estimators. Section 4 discusses the conclusions.

In the remainder of the paper it will be assumed, without loss of generality, that $\mu = 0$. This can be accomplished by a simple translation of the problem. (Only translation invariant π will be considered.)

2. SEPARATION UNDER NORMAL PRIORS

2.1 The Robust Generalized Bayes Estimator

For normal prior distributions, linear transformations of the problem do not affect Bayes risk. It is easy to check that a linear transformation can be made which preserves the block diagonal structure of $\ddagger$ and A and for which the transformed loss is sum of squares error loss (i.e., $Q = I_k$). It will, therefore, be assumed in this section that $Q = I_k$ and π is $N(0,A)$.

The calculation of $r(\pi,\delta^{RB})$ is made difficult by the presence of the function r in (1.1). We will, therefore, replace r by $(k-2)$, and consider the estimator

$$\delta^{RB*}(X) = (I_k - \frac{(k-2)}{X^t(\Sigma+A)^{-1}X}\ \Sigma\,(\Sigma+A)^{-1})X. \qquad (2.1.1)$$

(Recall we set $\mu=0$, without loss of generality.) δ^{RB} and δ^{RB*} are very similar. (Indeed they have virtually identical risks for moderately large k.) Numerical studies in Dey (1980) (which for the sake of brevity will not be reported) indicate that the separation results for the two estimators are identical.

As in Berger (1980a), a calculation using integration by parts gives that

$$R(\theta,\delta^{RB*}) = \mathrm{tr}\,\Sigma + E_\theta[-\frac{2(k-2)}{||X||^2}\{\mathrm{tr}\,\Sigma^2\,(\Sigma+A)^{-1} - \frac{2X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^2}\}$$

$$+\frac{(k-2)^2X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^4}] \qquad (2.1.2)$$

where $||X||^2 = X^t(\Sigma+A)^{-1}X$ and tr stands for the trace of a matrix.

The following lemma then gives the desired Bayes risk.

Lemma 2.1.1. If $Q=I_k$ and π is $N(0,A)$, then

$$r(\pi,\delta^{RB*}) = \mathrm{tr}\,\Sigma - \frac{(k-2)}{k}\,\mathrm{tr}\,\Sigma^2(\Sigma+A)^{-1}. \qquad (2.1.3)$$

<u>Proof</u>. It is clear that marginally X is $N(0,\Sigma+A)$, and hence $X^t(\Sigma+A)^{-1}X$ has a chi-square distribution with k-degrees of freedom. Thus it can be easily shown that

$$E^X\left[\frac{1}{\|X\|^2}\right] = \frac{1}{k-2} . \qquad (2.1.4)$$

Now let O be a $k \times k$ orthogonal matrix chosen so that

$$\Lambda = O^t(\Sigma+A)^{-1/2}\ \Sigma^2(\Sigma+A)^{-1/2}O$$

is diagonal, with diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_k$. Let a random variable Y be defined as $Y = O(\Sigma+A)^{-1/2}X$. Then we have

$$\|X\|^2 = X^t(\Sigma+A)^{-1}X = Y^t Y \qquad (2.1.5)$$

and

$$X^t(\Sigma+A)^{-1}\ \Sigma^2(\Sigma+A)^{-1}X = Y^t \Lambda Y = \sum_{i=1}^{k} \lambda_i Y_i^2 . \qquad (2.1.6)$$

Therefore,

$$E^X\left[\frac{X^t(\Sigma+A)^{-1}\ \Sigma^2(\Sigma+A)^{-1}X}{\|X\|^4}\right] = \frac{1}{k(k-2)}\ \mathrm{tr}\ \Sigma^2(\Sigma+A)^{-1}, \qquad (2.1.7)$$

using Lemma 1 of the appendix, since Y_i^2, $i=1,\ldots,k$ are independent chi-square with 1 degree of freedom. Now (2.1.2) follows immediately using (2.1.4) and (2.1.7), which completes the proof.||

For completeness, we note that the separate estimator $\delta^{RB*S} = (\delta^{RB*}_{(1)}, \ldots, \delta^{RB*}_{(s)})^t$ is defined by

$$\delta^{RB*}_{(\ell)} = \left(I_{k_\ell} - \frac{(k_\ell-2)}{X^t_{(\ell)}(\Sigma_{(\ell)}+A_{(\ell)})^{-1}X_{(\ell)}}\ \Sigma_{(\ell)}(\Sigma_{(\ell)}+A_{(\ell)})^{-1}\right)X_{(\ell)}. \qquad (2.1.8)$$

Applying Lemma 2.1.1 to each group separately and summing shows that δ^{RB*S} has Bayes risk

$$r(\pi,\delta^{RB*S}) = \sum_{\ell=1}^{s} \{\mathrm{tr}\ \ddagger_{(\ell)} - \frac{(k_\ell-2)}{k_\ell}\mathrm{tr}\ \ddagger^2_{(\ell)}(\ddagger_{(\ell)}+A_{(\ell)})^{-1}\}$$

$$= \mathrm{tr}\ \ddagger - \sum_{\ell=1}^{s} (1-\frac{2}{k_\ell})\mathrm{tr}\ \ddagger^2_{(\ell)}(\ddagger_{(\ell)}+A_{(\ell)})^{-1} . \qquad (2.1.9)$$

The major result is given in the following theorem, namely that the combined estimator is always better than the separate estimator in this situation.

Theorem 2.1.1. Suppose $Q = I_k$ and π is $N(0,A)$. Then δ^{RB*} is better than δ^{RB*S} in terms of Bayes risk.

Proof. Comparing (2.1.3) and (2.1.9), noting that $k_\ell \leq k$, the conclusion follows. ||

2.2 The Minimax Bayes Estimator

In this section, it will be assumed that π is $N(0,A)$; that Q, $\ddagger$, and A are diagonal with diagonal elements q_i, d_i and a_i, respectively, and that the coordinates are indexed so that $q_1^* \geq q_2^* \geq \cdots \geq q_k^*$, where $q_i^* = q_i d_i^2/(d_i+a_i)$. In Berger (1980b) the following estimator was shown to be minimax (i.e., uniformly better than δ°), and yet allowed incorporation of μ and A: $\delta^{MB*} = (\delta_1^{MB*},\ldots,\delta_k^{MB*})$, where

$$\delta_i^{MB*}(X) = q_i^{*-1} \sum_{j=1}^{k} (q_j^* - q_{j+1}^*)[(1 - \frac{(j-2)^+}{||X^j - \mu^j||^2}\frac{d_i}{(d_i+a_i)})(X_i-\mu_i)+\mu_i], \qquad (2.2.1)$$

where $||X^j - \mu^j||^2 = \sum_{\ell=1}^{j} (X_\ell - \mu_\ell)^2/(d_\ell+a_\ell)$, $q_{k+1}^* \equiv 0$, and $(j-2)^+$ denotes the positive part of (j-2).

A complication arises in this situation when attempting to divide θ into groups for separate estimation. The complication is that it is important to retain the ordering of the q_i^* in each group. (The q_i^* reflect the "importance" of the coordinates to improvement in simultaneous estimation. See Berger (1980b) for further discussion.) We will therefore only consider groups formed from the given ordering; i.e., the first group will be $\theta_{(1)} = (\theta_1,\ldots,\theta_{k_1})^t$, the second group will be $\theta_{(2)} = (\theta_{k_1+1},\ldots,\theta_{k_1+k_2})^t$, etc., where the θ_i are as above (with corresponding q_i^* that are decreasing). Such grouping in terms of decreasing q_i^* is natural, in any case, since "similar" coordinates should have similar q_i^*. The following theorem shows that the combined estimator δ^{MB*} has smaller Bayes risk than the separate estimator $\delta^{MB*S} = (\delta^{MB*}_{(1)},\ldots,\delta^{MB*}_{(s)})^t$, where $\delta^{MB*}_{(\ell)}$ is given componentwise as

$$\delta^{MB*S}_{(\ell)i} = q_i^{*-1} \sum_{j=T_{\ell-1}+1}^{T_\ell} (q_j^* - q_{j+1}^*)\left[\left(1 - \frac{\left(j-2-T_{\ell-1}\right)^+}{\|X^j - \mu^j\|_\ell^2} \frac{d_i}{d_i+a_i}\right)(X_i - \mu_i)+\mu_i\right],$$

where $T_\ell = k_1+\ldots+k_\ell (\ell=1,\ldots,s)$, $T_o \equiv 0$, $\|X^j - \mu^j\|_\ell^2 = \sum_{i=T_{\ell-1}+1}^{j} (X_i-\mu_i)^2/(d_i+a_i)$,

and (by an abuse of notation) $q^*_{T_\ell+1}$ in the above expression is understood to be zero for each ℓ.

Theorem 2.2.1. In the above situation, where π is $N(0,A)$, δ^{MB*} has smaller Bayes risk than δ^{MB*S}.

<u>Proof</u>. For simplicity, we will drop the "*" from q_i^*. Equation (A1) of Berger (1980b) shows that

$$r(\pi,\delta^{MB*}) = \text{tr } Q\ddagger - \sum_{i=1}^{k}\sum_{j=i}^{k} \frac{(j-2)^+}{j} \frac{(q_j - q_{j+1})}{q_i} [2q_i - (q_j + q_{j+1})]$$

$$\geq \text{tr } Q\ddagger - \sum_{i=1}^{k} \sum_{j=i}^{k} \frac{(q_j - q_{j+1})}{q_i} [2q_i - (q_j + q_{j+1})]$$

$$= \text{tr } Q\ddagger - \sum_{i=1}^{k} \sum_{j=i}^{k} 2(q_j - q_{j+1}) + \sum_{i=1}^{k} \sum_{j=i}^{k} \frac{(q_j^2 - q_{j+1}^2)}{q_i}$$

$$= \text{tr } Q\ddagger - \sum_{i=1}^{k} 2q_i + \sum_{i=1}^{k} \frac{q_i^2}{q_i}$$

$$= \text{tr } Q\ddagger - \sum_{i=1}^{k} q_i. \tag{2.2.2}$$

This inequality clearly also holds for the $\delta_{(\ell)}^{MB*S}$, implying that

$$r(\pi,\delta^{MB*S}) = \sum_{\ell=1}^{s} r(\pi,\delta_{(\ell)}^{MB*S})$$

$$\geq r(\pi,\delta_{(1)}^{MB*S}) + \sum_{\ell=2}^{s} \{\text{tr } Q_{(\ell)}\Sigma_{(\ell)} - \sum_{i=T_{\ell-1}+1}^{T_\ell} q_i\}. \tag{2.2.3}$$

Lemma 2 of Berger (1980b) states that

$$r(\pi,\delta_{(1)}^{MB*}) = \text{tr } Q_{(1)}\Sigma_{(1)} - \sum_{i=3}^{T_1} q_i - 2\sum_{i=3}^{T_1} \frac{q_i}{i} [1 - \frac{q_i}{(i-1)} \sum_{j=1}^{(i-1)} \frac{1}{q_j}].$$

Using this in (2.2.3) together with the fact that

$$1 - \frac{q_i}{(i-1)} \sum_{j=1}^{(i-1)} \frac{1}{q_j} \geq 0, \quad i=1,\ldots,k$$

(since the q_i^* are nonincreasing), we obtain

$$r(\pi,\delta^{MB*S}) \geq \text{tr } Q\ddagger - \sum_{i=3}^{k} q_i - 2\sum_{i=3}^{T_1} \frac{q_i}{i} [1 - \frac{q_i}{(i-1)} \sum_{j=1}^{(i-1)} \frac{1}{q_j}]$$

$$\geq \text{tr}\, Q\ddagger - \sum_{i=3}^{k} q_i - 2\sum_{i=3}^{k} \frac{q_i}{i}\left[1 - \frac{q_i}{(i-1)} \sum_{j=1}^{(i-1)} \frac{1}{q_j}\right]$$

$$= r(\pi,\delta^{MB*}) \quad \text{(Lemma 2 of Berger (1980b))}.$$

This establishes the theorem. ||

3. SEPARATION UNDER A FLAT PRIOR

The results of the previous section are somewhat surprising. Even for a normal prior which reflects independence of the various groups of coordinates, the combined estimators seem better than the separate estimators. To alleviate concerns that this result may be due to the sharp tails of the normal prior, we consider in this section priors which incorporate μ and A but have flat tails. We restrict consideration to δ^{RB*} and δ^{RB*S}, and again assume (w.l.o.g.) that $\mu = 0$.

3.1 Numerical Results for a Certain Flat-tailed Prior

We will assume that Σ and A are diagonal with diagonal elements d_i and a_i, respectively. Also assume that given λ_i, the θ_i's are independent $N(0,b(\lambda_i))$, i=1,...,k, where the λ_i's are independently distributed with density $f(\lambda) = n\lambda^{n-1}$, for $n > 0$ and $0 < \lambda < 1$, and $b(\lambda_i) = \frac{c_i}{\lambda_i} - d_i$, i=1,...,k where c_i is defined here as $c_i = d_i + a_i$. Thus the generalized prior density for θ_i is

$$g_n(\theta_i) = \frac{n}{\sqrt{2\pi}} \int_0^1 (b(\lambda_i))^{-1/2} \exp\{-\frac{\theta_i^2}{2b(\lambda_i)}\} \lambda_i^{n-1}\, d\lambda_i. \qquad (3.1.1)$$

It can be shown asymptotically (for large θ_i) that $g_n(\theta_i)$ behaves like $C_1(\theta_i)^{-2n}$, for some constant C_1. Thus g_n is a prior density with a tail considerably flatter than that of a normal density. (This particular density is chosen for its comparative ease in calculation.) Clearly, given λ_i, the X_i's are independent $N(0, c_i/\lambda_i)$ for all $i=1,\ldots,k$. Thus from (2.1.2) the Bayes risk of δ^{RB*} with respect to the above prior is given as

$$r(g_n,\delta^{RB*}) = \mathrm{tr}\,\ddagger - 2(k-2)\sum_{i=1}^{k}\frac{d_i^2}{c_i}E^{\lambda}E^{X|\lambda}\left[\frac{1}{\sum_{i=1}^{k} X_i^2/c_i}\right]$$

$$+ (k^2-4)\,E^{\lambda}E^{X|\lambda}\left[\frac{\sum_{i=1}^{k} d_i^2X_i^2/c_i^2}{(\sum_{i=1}^{k} X_i^2/c_i)^2}\right] \qquad (3.1.2)$$

where $\lambda = (\lambda_1,\ldots,\lambda_k)$ and $E^{X|\lambda}(\cdot)$ stands for the expectation under the conditional distribution of X given λ.

By choosing different values of n, several flat priors can be generated. We take $n=2$ for simplicity. Note that

$$E^{\lambda}E^{X|\lambda}\left[\frac{1}{\sum_{i=1}^{k} X_i^2/c_i}\right] = E^{X}\left[\frac{1}{\sum_{i=1}^{k} X_i^2/c_i}\right] \qquad (3.1.3)$$

and

$$E^{\lambda}E^{X|\lambda}\left[\frac{\sum_{i=1}^{k} d_i^2X_i^2/c_i^2}{(\sum_{i=1}^{k} X_i^2/c_i)^2}\right] = \sum_{i=1}^{k}\frac{d_i^2}{c_i}E^{X}\left[\frac{X_i^2/c_i}{(\sum_{i=1}^{k} X_i^2/c_i)^2}\right]. \qquad (3.1.4)$$

Since the λ_i's are independent and identically distributed, the unconditional distributions of X_i^2/c_i are independent of i. Therefore

$$k\,E^X\left[\frac{X_i^2/c_i}{\left(\sum_{i=1}^{k} X_i^2/c_i\right)^2}\right] = E^X\left[\frac{1}{\sum_{i=1}^{k} X_i^2/c_i}\right] . \tag{3.1.5}$$

Using (3.1.3), (3.1.4) and (3.1.5) in (3.1.2) we have

$$r(g_1,\delta^{RB*}) = \text{tr}\ \ddagger - \frac{(k-2)^2}{k} V\ S , \tag{3.1.6}$$

where $V = \sum_{i=1}^{k} d_i^2/c_i$ and $S = E^X\left[\frac{1}{\sum_{i=1}^{k} X_i^2/c_i}\right]$. S can be computed by a numerical integration technique using Lemma 2 in the appendix. Table 1 gives the values of S for various k. (Using the conditional representation in (3.1.3) it is easy to see that S does not depend on the c_i.)

In calculating the Bayes risk of the separate estimator, define S_ℓ and V_ℓ for the ℓth group as

$$S_\ell = E^{X(\ell)}\left[\frac{1}{\sum_{i=T_{\ell-1}+1}^{T_\ell} X_i^2/c_i}\right], \tag{3.1.7}$$

and

$$V_\ell = \sum_{i=T_{\ell-1}+1}^{T_\ell} d_i^2/c_i, \quad \ell=1,2,\ldots,s, \tag{3.1.8}$$

where T_ℓ is defined as in section 2. Note that S_ℓ can be found from Table 1. Now using (3.1.7), (3.1.8) and (3.1.2), the Bayes risk of $\delta_{(\ell)}^{RB*}$ is given by

$$r(g_1,\delta_{(\ell)}^{RB*}) = \text{tr}\ \ddagger_{(\ell)} - \frac{(k_\ell-2)^2}{k_\ell} V_\ell S_\ell, \quad \ell=1,\ldots,s. \tag{3.1.9}$$

Table I

Values of S for different values of k.

k	S	k	S
3	.38339	13	.05117
4	.23961	14	.04670
5	.17750	15	.04295
6	.13959	20	.03055
7	.11382	25	.02366
8	.09540	30	.01928
9	.08173	35	.01646
10	.07129	40	.01406
11	.06310	45	.01236
12	.05654	50	.01020

Table 2

Difference of Bayes risks of combined and separate estimators. $k = 10$

k_1	(d_i), $i = 1,2,\ldots,10$	(c_i), $i = 1,2,\ldots,10$	Δ
5	(1,1,1,1,1,1,1,1,1,1)	(10,10,10,10,10,10,10,10,10,10)	- .1256
5	(.0001,.01,.1,.1,1,2,2,2,2,4)	(10,10,10,10,10,12,12,12,100,100)	- .1766
5	(1,1,1,1,1,2,2,2,2,2)	(10,10,10,10,10,10,10,10,10,10)	- .3139
5	(1,1,1,1,1,2,2,2,2,2)	(1,1,1,1,1,1000,1000,1000,1000,1000)	- .5558
5	(.1,.1,.1,.1,.1,1,1,1,1,1)	(1,1,1,1,1,1,1,1,1,1)	- .5592
5	(1,1,1,1,1,2,2,2,2,2)	(1,1,1,1,1,200,200,200,200,200)	- .5647
5	(1,1,1,1,1,10,10,10,10,10)	(1,1,1,1,1,1000,1000,1000,1000,1000)	- .6090
5	(1,1,1,1,1,1,1,1,1,1)	(2,2,2,2,2,2,2,2,2,2)	- .6278
5	(1,1,1,1,1,2,2,2,2,2)	(1,1,1,1,1,10,10,10,10,10)	- .7751
5	(1,1,1,1,1,10,10,10,10,10)	(1,1,1,1,1,200,200,200,200,200)	- .8305
5	(1,1,1,1,1,10,10,10,10,10)	(1,1,1,1,1,100,100,100,100,100)	- 1.1073
5	(1,1,1,1,1,1,1,1,1,1)	(1,1,1,1,1,1,1,1,1,1)	- 1.2556
5	(1,1,1,1,1,10,10,10,10,10)	(1,1,1,1,1,10,10,10,10,10)	- 6.0902
5	(1,1,1,1,1,10,10,10,10,10)	(10,10,10,10,10,10,10,10,10,10)	- 6.3406
4	(.0001,.0001,.0001,.01,.01,10,10,10,10,10)	(1,1,1,2,2,10,10,100,100,100)	- 11.6132
5	(.0001,.0001,.0001,.01,.01,10,10,10,10,10)	(1,1,1,2,2,10,10,100,100,100)	- 25.4680
6	(.0001,.0001,.0001,.01,.01,10,10,10,10,10)	(1,1,1,2,2,10,10,100,100,100)	- 62.2560
7	(.0001,.0001,.0001,.01,.01,10,10,10,10,10)	(1,1,1,2,2,10,10,100,100,100)	- 158.4928

Then we can conclude that the difference of the Bayes risks of the combined and separate estimators is given by

$$\Delta = r(g_1, \delta^{RB*}) - r(g_1, \delta^{RB*S})$$

$$= r(g_1, \delta^{RB*}) - \sum_{\ell=1}^{s} r(g_1, \delta^{RB*}_{(\ell)}) = \sum_{\ell=1}^{s} \frac{(k_\ell - 2)^2}{k_\ell} V_\ell S_\ell - \frac{(k-2)^2}{k} V\, S. \quad (3.1.10)$$

Using Table 1, Δ can be computed for different c_i, k, and k_ℓ. A representative sample of such calculations is given in Table 2 for s=2. All calculated values of Δ are negative, which indicates that the combined estimator is better than the separate estimator with respect to the Bayes risk, under the given flat prior.

Remark 3.1. If s=2 and $k_1=k_2$, we have $S_1=S_2$. Then (3.1.10) reduces to

$$\Delta = -V \left\{\frac{(k-2)^2}{k} S - \frac{(k_1-2)^2}{k_1} S_1\right\}.$$

Numerical calculations indicate that the expression within the brackets is always positive and it is clearly constant when k_1 and k are fixed. Thus for fixed k_1 and k, Δ is proportional to V.

3.2. Asymptotic (as $k \to \infty$) Results for Separation

Under the assumption that Σ is a diagonal matrix and the prior has uniformly bounded tenth moments, we will approximate the Bayes risk of the combined estimator δ^{RB*} and the separate estimator δ^{RB*S} for large k. Assume that

$$\Sigma = \text{diag}(\underbrace{d_1, \ldots, d_1}_{k_1}, \ldots, \underbrace{d_s, \ldots, d_s}_{k_s})$$

and that the prior π on θ is such that $\theta_1,\ldots,\theta_k$ are independent with $E(\theta_i)=0$, $E(\theta_i^2)=a_\ell$ and $E(\theta_i^2-a_\ell)^2=\nu_\ell$ when $T_{\ell-1}+1 \le i \le T_\ell$; $\ell=1,\ldots,s$ (the T_ℓ being defined as before). Suppose also that there exists a $T<\infty$ such that, for all i, $E(\theta_i^{10})<T$ and $a_i/d_i<T$. Finally, assume that $\tau_\ell = \lim_{k\to\infty}(k_\ell/k)$ exists for all ℓ, and that $0<\tau_\ell<1$.

Theorem 3.2.1. If $\ddagger$ and π are as above, then Δ, defined as the difference of the Bayes risks of δ^{RB*} and δ^{RB*S}, is given by

$$\Delta = \sum_{\ell=1}^{s} \frac{d_\ell^2}{d_\ell+a_\ell}\{2(\tau_\ell-1) + \frac{\nu_\ell-2a_\ell^2}{(d_\ell+a_\ell)^2} - \tau_\ell \sum_{j=1}^{s} \frac{(\nu_j-2a_j^2)\tau_j}{(d_j+a_j)^2}\} + o(1) \qquad (3.2.1)$$

where o(1) converges to zero as $k\to\infty$.

Proof. Given in the appendix.

Remark 3.2. For normal priors, an easy calculation shows that $\nu_\ell = 2a_\ell^2$. Hence $\Delta<0$ asymptotically, agreeing with the more explicit results of section 2.

Remark 3.3. If $\nu_\ell=\nu$, $a_\ell=a$, and $d_\ell=d$ ($\ell=1,\ldots,s$), then

$$\Delta = \frac{d^2}{d+a}(s-1)\{\frac{\nu-2a^2}{(d+a)^2} - 2\}.$$

Thus separation is asymptotically better, even in this symmetric situation, if

$$\frac{\nu-2a^2}{(d+a)^2} > 2. \qquad (3.2.2)$$

This inequality can be satisfied for very flat tailed priors. For example, it can be shown to hold for the truncated t priors

$$\pi(\theta_i) = c(1 + \frac{\theta_i^2}{\beta})^{-(\alpha+1)/2} I_{\{|\theta_i| \leq M\}},$$

providing β and M are large enough and $4 < \alpha < 7$. (The truncation at M is to ensure that the moment assumptions of Theorem 3.2.1 are satisfied.)

This is somewhat discouraging, in that it shows that separation can be better if the fourth moment of the prior is large enough. Trying a variety of possible forms for π, however, will convince the reader that it is quite rare for (3.2.2) to be satisfied (providing the appropriate numbers of moments exist). Since it is rare to have accurate enough prior knowledge to be able to specify a fourth moment, use of the combined estimator is again indicated therefore.

Remark 3.4. It is natural to question the assumption that the prior even has finite moments. Priors with very flat tails are not at all unreasonable. But if the prior does not have (say) finite variances, then it is clear that the estimators considered here are all inadequate, since $(k-2)/(X-\mu)^t(\ddagger+A)^{-1}(X-\mu)$ becomes infinitely small (with probability one) as $k \to \infty$. To deal with this problem, Stein (1974) proposed (for the symmetric case) truncating excessively large values of the X_i. Determination of the optimal truncation point is an interesting problem discussed in Dey (1980). Indications are however, that for properly truncated versions, the combined estimator is still better than the separate estimator.

4. CONCLUSIONS AND EXTENSIONS

The results obtained in this paper came as a considerable surprise. It was initially believed that separation would quite often turn out to be superior. Instead it was indicated that, in the vast majority of cases, the combined estimator is better than the separate estimator. An important qualification for this result must be made, however. The qualification is that it may be desirable, in some situations, to evaluate the Bayes risks of the various estimators for varying values of μ and A. For example, suppose there are two groups, with specified prior information $\mu_{(1)}$ and $A_{(1)}$ for the first group and $\mu_{(2)}$ and $A_{(2)}$ for the second. Suppose $\mu_{(1)}$ and $A_{(1)}$ are thought to have been accurately determined (subjectively), but that $\mu_{(2)}$ and $A_{(2)}$ were hard to specify and could be wrong (in terms of true prior beliefs). It may pay to use separate estimators for each group, provided the possible variation in $\mu_{(2)}$ and $A_{(2)}$ is large enough. Another way of saying this is that if robustness with respect to possible misspecification of the prior beliefs is taken into account, the separate estimator may well be better than the combined estimator. The investigation of this will be pursued elsewhere.

5. APPENDIX

Lemma 1. If $\{X_i\}$, i=1,2,...,k, is a sequence of independent and identically distributed chi-square random variables with 1 degree of freedom, then

$$E\left[\frac{\sum_{i=1}^{k} \ell_i X_i}{\left(\sum_{i=1}^{k} X_i\right)^2} \right] = \frac{1}{k(k-2)} \sum_{i=1}^{k} \ell_i , \qquad \text{(A1)}$$

where the ℓ_i are scalars.

Proof. We have

$$E\left[\frac{\ell_i X_i}{(\sum_{i=1}^{k} X_i)^2}\right] = \ell_i \, E\left[\frac{\frac{1}{k}\sum_{i=1}^{k} X_i}{(\sum_{i=1}^{k} X_i)^2}\right]$$

$$= \frac{\ell_i}{k} E\left[\frac{1}{\sum_{i=1}^{k} X_i}\right] = \frac{\ell_i}{k(k-2)} .$$

Now summing over i, the proof is complete. ||

Lemma 2.

$$S = E^X\left[\frac{1}{\sum_{i=1}^{k} X_i^2/c_i}\right] = \int_0^\infty \{(1+2t)^{1/2}(1-3t) + 6t^2 \sin h^{-1} \frac{1}{(2t)^{-1/2}}\}^k dt. \quad \text{(A2)}$$

Proof. Suppose $Z_i = X_i^2$, $i=1,\ldots,k$. Then $Z_i|\lambda_i$ is $\frac{c_i}{\lambda_i}\chi_1^2$ where χ_1^2 represents a chi-square random variable with 1 degree of freedom. Thus given λ_i, Z_i/c_i is $\frac{1}{\lambda_i}\chi_1^2$. Hence given λ_i, the Laplace transform of Z_i/c_i is

$$\phi(t|\lambda_i) = E^{Z_i|\lambda_i}[e^{-tZ_i/c_i}] = (1 + \frac{2t}{\lambda_i})^{-1/2}$$

It follows that the unconditional Laplace transform of Z_i/c_i is

$$\phi(t) = \int_0^1 2(1 + \frac{2t}{\lambda_i})^{-1/2} \lambda_i d\lambda_i = 4t \int_0^{\theta^*} \tanh\theta \sinh^2\theta \; 4tc \sinh\theta \cosh\theta \, d\theta,$$

the last step following by the change of variables $\lambda = 2t \sinh^2\theta$ and defining $\theta^* = \sinh^{-1}(2t)^{-1/2}$. Integrating by parts gives

$$\phi(t) = 16t^2 \int_0^{\theta^*} \sinh^4 \theta \, d\theta = 16t^2\{\tfrac{1}{4} \sinh^3 \theta^* \cosh \theta^* - \tfrac{3}{4} \int_0^{\theta^*} \sinh^2 \theta \, d\theta\}$$

$$= 16t^2\{\tfrac{1}{4}(2t)^{-2}(1+2t)^{1/2} - \tfrac{3}{4}[\tfrac{1}{4} \sinh 2\theta - \tfrac{\theta}{2}]_0^{\theta^*}\}$$

$$= (1+2t)^{1/2} - 3t(1+2t)^{1/2} + 6t^2 \sinh^{-1} (2t)^{-1/2}.$$

Now using the independence of the Z_i/c_i for $i=1,\ldots,k$, the Laplace transform of $\sum_{i=1}^{k} Z_i/c_i$ (or $\sum_{i=1}^{k} X_i^2/c_i$) is given as

$$E[e^{-t \sum_{i=1}^{k} X_i^2/c_i}] = (1+t)^{1/2} (1-3t) + 6t^2 \sinh^{-1} (2t)^{-1/2} .$$

Finally by Fubini's theorem,

$$E^X[\frac{1}{\sum_{i=1}^{k} X_i^2/c_i}] = \int_0^\infty E[e^{-t \sum_{i=1}^{k} X_i^2/c_i}]dt$$

$$= \int_0^\infty\{(1+2t)^{1/2} (1-3t) +6t^2 \sinh^{-1} (2t)^{-1/2}\}^k dt,$$

which completes the proof of the lemma. ||

The following lemmas are needed in the proof of Theorem 3.2.1. All relevant notation and conditions are given in Subsection 3.2.

Lemma 3. For $T_{\ell-1}+1 \leq i \leq T_\ell$,

$$E^X[\frac{X_i^2}{d_\ell+a_\ell} - 1] = 2 + \frac{\nu_\ell-2a_\ell^2}{(d_\ell+a_\ell)^2} \qquad \text{(A3)}$$

where E^X stands for expectation under the marginal distribution of X.

Proof. We have $X_i|\theta_i$ is $N(\theta_i, d_\ell)$, $T_{\ell-1} \leq i \leq T_\ell$. Therefore

$$E^X(X_i^2) = E^{\theta_i} E^{X_i|\theta_i}(X_i^2) = E^{\theta_i}(d + \theta_i^2) = d_\ell + a_\ell .$$

We know that $E^{X_i|\theta_i}(X_i - \theta_i)^4 = 3d_\ell^2$. Therefore,

$$E^{X_i|\theta_i}(X_i^4) = 3d_\ell^2 + 6\theta_i^2 d_\ell + \theta_i^4 .$$

Thus,

$$E^X(X_i^4) = 3d_\ell^2 + 6d_\ell a_\ell + \nu_\ell + a_\ell^2 \quad (\text{since } E(\theta_i^4) = \nu_\ell + a_\ell^2).$$

Therefore

$$E^X\left[\frac{X_i^2}{d_\ell + a_\ell} - 1\right]^2 = E^X\left[\frac{X_i^4}{(d_\ell + a_\ell)^2} + 1 - \frac{2X_i^2}{d_\ell + a_\ell}\right] = 2 + \frac{\nu_\ell - 2a_\ell^2}{(d_\ell + a_\ell)^2} .$$

This completes the proof of the lemma. ||

Lemma 4. Under the assumptions given at the beginning of Section 3.2,

$$E^X\left[\frac{k_\ell}{\|X_{(\ell)}\|^2}\right] = 1 + \frac{2}{k_\ell} + \frac{\nu_\ell - 2a_\ell^2}{k_\ell(d_\ell + a_\ell)^2} + o\left(\frac{1}{k_\ell}\right). \tag{A4}$$

Proof. Define $Y_i^2 = X_i^2/(d_\ell + a_\ell)$ and

$$\varepsilon = E^{Y(\ell)}\left\{\frac{\left[\sum_{i=T_{\ell-1}+1}^{T_\ell} (Y_i^2 - 1)\right]^3}{k_\ell \sum_{i=T_{\ell-1}+1}^{T_\ell} Y_i^2}\right\} . \tag{A5}$$

An easy calculation then gives that

$$E^X[\frac{k_\ell}{||X_{(\ell)}||^2}] = 1 - E^{Y(\ell)}[\frac{1}{k_\ell}\sum_{i=T_{\ell-1}+1}^{T_\ell}(Y_i^2-1)] + E^{Y(\ell)}[\frac{1}{k_\ell}\sum_{i=T_{\ell-1}+1}^{T_\ell}(Y_i^2-1)]^2 - \frac{\varepsilon}{k_\ell} .$$

Observing that the Y_i^2 are independent with $E^Y(Y_i^2) = 1$ and using (A3), it follows that

$$E^X[\frac{k_\ell}{||X_{(\ell)}||^2}] = 1 + \frac{1}{k_\ell}\{2 + \frac{\nu_\ell - 2a_\ell^2}{(d_\ell + a_\ell)^2}\} - \frac{\varepsilon}{k_\ell} .$$

Hence it is only necessary to prove that $\varepsilon = o(1)$.

Expanding the numerator in (A5) gives

$$\varepsilon = E^{Y(\ell)}\left\{\frac{\sum_i (Y_i^2-1)^3}{k_\ell \sum_i Y_i^2}\right\} + 3E^{Y(\ell)}\left\{\frac{\sum_{i\neq j}\sum (Y_i^2-1)^2(Y_j^2-1)}{k_\ell \sum_i Y_i^2}\right\}$$

$$+ 6E^{Y(\ell)}\left\{\frac{\sum_{i\neq j\neq m}\sum\sum (Y_i^2-1)(Y_j^2-1)(Y_m^2-1)}{k_\ell \sum_i Y_i^2}\right\}$$

$$= I_1 + I_2 + I_3 \quad \text{(say)}. \tag{A6}$$

(In the above expression and in the following, it is to be understood that summations or products are only over the integers $T_{\ell-1}+1,\ldots,T_\ell$).

To deal with I_1, define for some $a > 0$

$$\Omega = \{(Y_{T_{\ell-1}+1},\ldots,Y_{T_\ell})^t : \sum_{i\neq j} Y_i^2 \geq a^2 \text{ for } j = T_{\ell-1}+1,\ldots,T_\ell\},$$

and write (letting $f(Y_i)$ denote the marginal density of Y_i)

$$I_1 = \int_\Omega \frac{\sum_i (Y_i^2-1)^3}{k_\ell \sum_i Y_i^2} \prod_i [f(Y_i)dY_i] + \int_{\Omega^c} \frac{\sum_i (Y_i^2-1)^3}{k_\ell \sum_i Y_i^2} \prod_i [f(Y_i)dY_i]$$

$$= I_1^1 + I_1^2 \quad \text{(say)}. \tag{A7}$$

Now on Ω,

$$\frac{|\sum_i (Y_i^2-1)^3|}{k_\ell \sum_i Y_i^2} \leq \sum_i \frac{(Y_i^2-1)^2 |Y_i^2-1|}{k_\ell (a^2+Y_i^2)} \leq \sum_i \frac{(Y_i^2-1)^2}{k_\ell} (1 + \frac{1}{a^2}) \equiv g_{k_\ell}(Y_{(\ell)}) \quad \text{(say)}.$$

From the assumption of uniformly bounded $10^{\underline{th}}$ moments, it can be assumed without loss of generality that $E(Y_i^2-1)^2 \leq T$ for all i. An easy argument, using the strong law of large numbers and the extended Lebesgue dominated convergence theorem (see Rao (1973) for a statement of this theorem due to V. Johns and J. Pratt), then shows that $|I_1^1| = o(1)$.

To deal with I_1^2 from (A7), define

$$\Omega_j = \{Y_{(\ell)} : \sum_{i \neq j} Y_i^2 < a^2\},$$

observe that $\Omega^c = \bigcup_{j=T_{\ell-1}+1}^{T_\ell} \Omega_j$, and note from this and the independence of the Y_i that

$$|I_1^2| \leq \sum_i \int_{\Omega^c} \frac{|Y_i^2-1|^3}{k_\ell \sum_{j \neq i} Y_j^2} \prod_j [f(Y_j)dY_j]$$

$$\leq \sum_i \{(E|Y_i^2-1|^3) \sum_m \int_{\Omega_m} \frac{1}{k_\ell \sum_{j\neq i,m} Y_j^2} \prod_{j\neq i} [f(Y_j)dY_j]\}$$

$$\leq \sum_i \sum_m [E|Y_i^2-1|^3] \int_{\{\sum_{j\neq i,m} Y_j^2 < a^2\}} \frac{1}{k_\ell \sum_{j\neq i,m} Y_j^2} \prod_{j\neq i,m} [f(Y_j)dY_j].$$

Using the fact that $a_\ell/d_\ell \leq T$, it is clear that

$$f(Y_i) = E^{\pi(\theta_i)}[\{\frac{(d_\ell+a_\ell)}{2\pi d_\ell}\}^{1/2} \exp\{-\frac{(d_\ell+a_\ell)}{2d_\ell}(Y_i - \frac{\theta_i}{(d_\ell+a_\ell)^{1/2}})^2\}]$$

$$\leq \{\frac{(d_\ell+a_\ell)}{2\pi d_\ell}\}^{1/2} \leq (\frac{1+T}{2\pi})^{1/2} .$$

Together with the uniform bound T on the moments $E|Y_i^2-1|^3$, it follows that

$$|I_1^2| \leq \sum_i \sum_m T(\frac{1+T}{2\pi})^{(k_\ell-2)/2} \int_{\{\sum_{j\neq i,m} Y_j^2 < a^2\}} \frac{1}{k_\ell \sum_{j\neq i,m} Y_j^2} \prod_{j\neq i,m} dY_j$$

$$= k_\ell^2\, T(\frac{1+T}{2\pi})^{(k_\ell-2)/2} \frac{a^{k_\ell-4}\pi^{(k_\ell-2)/2}}{k_\ell\Gamma(k_\ell/2)}$$

$$= o(1).$$

To deal with I_2 from (A6), observe first that

$$\frac{1}{\sum_i Y_i^2} = \frac{1}{\sum_{i\neq j} Y_i^2} - \frac{Y_j^2}{(\sum_{i\neq j} Y_i^2)(\sum_i Y_i^2)} .$$

It follows that

$$I_2 = 3 \sum_{i \neq j} \sum [E^{Y(\ell)} \left\{ \frac{(Y_i^2-1)^2(Y_j^2-1)}{k_\ell \sum_{i\neq j} Y_i^2} \right\} - E^{Y(\ell)} \left\{ \frac{(Y_i^2-1)^2(Y_j^2-1)Y_j^2}{k_\ell (\sum_{i\neq j} Y_i^2)(\sum_i Y_i^2)} \right\}]$$

$$= -3 \sum_{i \neq j} \sum E^{Y(\ell)} \left\{ \frac{(Y_i^2-1)^2(Y_j^2-1)Y_j^2}{k_\ell (\sum_{i\neq j} Y_i^2)(\sum_i Y_i^2)} \right\} \qquad (\text{since } E(Y_j^2) = 1).$$

The argument that $|I_2| = o(1)$ proceeds as did that for I_1.

The argument for I_3 follows similarly after applying the identity

$$\frac{1}{\sum_i Y_i^2} = \frac{1}{\sum_{i\neq j} Y_i^2} - \frac{Y_j^2}{(\sum_{i\neq j} Y_i^2)^2} + \frac{Y_j^4}{(\sum_{i\neq j} Y_i^2)^2 (\sum_i Y_i^2)} .$$

It is for this term that uniformly bounded tenth moments are required. ||

Lemma 5.

$$E^X[\frac{k}{||X||^2}] = 1 + \frac{2}{k} + \frac{1}{k^2} \sum_{\ell=1}^{s} \frac{(\nu_\ell - 2a_\ell^2)}{(d_\ell + a_\ell)^2} + o(\tfrac{1}{k}). \qquad \text{(A8)}$$

<u>Proof</u>. Similar to the proof of Lemma 4. ||

<u>Proof of Theorem 3.2.1.</u>

Using (2.1.2) and Lemma 5 we have that the Bayes risk of δ^{RB*} is

$$r(\pi,\delta^{RB*}) = \sum_{\ell=1}^{s} k_\ell d_\ell - \frac{(k-2)^2}{k^2} \{ \sum_{\ell=1}^{s} \frac{d_\ell^2}{d_\ell + a_\ell} k_\ell \} \{1 + \frac{2}{k} + \frac{1}{k^2} \sum_{\ell=1}^{s} \frac{k_\ell(\nu_\ell - 2a_\ell^2)}{(d_\ell + a_\ell)^2} + o(\tfrac{1}{k}) \}.$$

(A9)

Similarly by Lemma 4, we have

$$r(\pi,\delta^{RB*}_{(\ell)}) = k_\ell d_\ell - \frac{(k_\ell-2)^2 d_\ell^2}{k_\ell(d_\ell+a_\ell)}\{1+\frac{2}{k_\ell}+\frac{\nu_\ell-2a_\ell^2}{k_\ell(d_\ell+a_\ell)^2}+o(\frac{1}{k_\ell})\}.$$

Thus the Bayes risk of the separate estimator is

$$r(\pi,\delta^{RB*S}) = \sum_{\ell=1}^{s} k_\ell d_\ell - \sum_{\ell=1}^{s}\frac{(k_\ell-2)^2}{k_\ell}\frac{d_\ell^2}{d_\ell+a_\ell}\{1+\frac{2}{k_\ell}+\frac{\nu_\ell-2a_\ell^2}{k_\ell(d_\ell+a_\ell)^2}+o(\frac{1}{k_\ell})\}. \quad \text{(A10)}$$

From (A9) and A10), it follows that the difference of Bayes risks is

$$\Delta = -4\sum_{\ell=1}^{s}\frac{d_\ell}{d_\ell+a_\ell} + 4\sum_{\ell=1}^{s}\frac{d_\ell^2\tau_\ell}{d_\ell+a_\ell} + \sum_{\ell=1}^{s}\frac{d_\ell^2}{d_\ell+a_\ell}\{2+\frac{\nu_\ell-2a_\ell^2}{(d_\ell+a_\ell)^2}\}$$

$$- \sum_{\ell=1}^{s}\frac{2d_\ell^2\tau_\ell}{d_\ell+a_\ell} - \{\sum_{\ell=1}^{s}\frac{d_\ell^2\tau_\ell}{d_\ell+a_\ell}\sum_{\ell=1}^{s}\frac{(\nu_\ell-2a_\ell^2)\tau_\ell}{(d_\ell+a_\ell)^2}\} + o(1)$$

$$= \sum_{\ell=1}^{s}\frac{d_\ell^2}{d_\ell+a_\ell}\{2(\tau_\ell-1) + \frac{\nu_\ell-2a_\ell^2}{(d_\ell+a_\ell)^2} - \tau_\ell\sum_{j=1}^{s}\frac{(\nu_j-2a_j^2)\tau_j}{(d_j+a_j)^2}\} + o(1),$$

which completes the proof of the theorem. ||

BIBLIOGRAPHY

[1] Berger, J. (1975). Minimax estimation of location vectors of a wide class of densities. Ann. Statist. 3, 1318-1328.

[2] Berger, J. (1976a). Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss. Ann. Statist. 4, 223-226.

[3] Berger, J. (1976b). Minimax estimation of a multivariate normal mean under arbitrary quadratic loss. J. Multivariate Anal. 6, 256-264.

[4] Berger, J. (1979). Multivariate estimation with nonsymmetric loss functions. In Optimizing Methods in Statistics, J.S. Rustagi (Ed.) Academic Press, New York.

[5] Berger, J. (1980a). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. Ann. Statist. 8, 716-761.

[6] Berger, J. (1980b). Selecting a minimax estimator of a multivariate normal mean. Mimeograph Series #80-24, Department of Statistics, Purdue University.

[7] Bhattacharya, P.K. (1966). Estimating the mean of a multivariate normal population with general quadratic loss function. Ann. Math. Statist. 37, 1819-1927.

[8] Bock, M.E. (1975). Minimax estimators of the mean of a multivariate normal distribution. Ann. Statist. 3, 209-218.

[9] Chung, K.L. (1968). A Course in Probability Theory. Harcourt, Brace & World, Inc.

[10] Dey, D.K. (1980). Ph.D. Thesis, Purdue University.

[11] Efron, B. and Morris, C. (1973a). Stein's estimation rule and its competitors - an empirical Bayes approach. J. Amer. Statist. Assoc. 68, 117-130.

[12] Efron, B. and Morris, C. (1973b). Combining possibly related estimation problems. J. Roy. Statist. Soc. B, 35, 379-421.

[13] Hudson, M. (1974). Empirical Bayes Estimation. Technical Report #58, Stanford University.

[14] James, W. and Stein, C. (1960). Estimation with quadratic loss. Proc. Fourth Berkeley Symposium Math. Stat. Prob. 1, 361-379. University of Calfiornia Press

[15] Rao, C.R. (1973). Linear Statistical Inference and Its Applications. Second Edition, John Wiley and Sons.

[16] Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Stat. Prob. 1, 197-206. University of California Press.

[17] Strawderman, W.E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. Ann. Math. Statist. 42, 385-388.