On the Choice of Coordinates in Simultaneous Estimation of Normal Means

by

Dipak K. Dey
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #80-23

August 1980

rch was supported by the NSF Grant MCS 78-02300.

CHAPTER I

INTRODUCTION

Section 1.1. Definitions and Notation

In this section, we briefly discuss the definitions and notation that are used throughout this paper. Let $X = (X_1,\ldots,X_k)$ represent an observable vector valued random variable with values x in a sample space $\mathcal{X}$. Assume X has a k-variate normal distribution with mean vector θ and covariance matrix Σ. Assume θ is unknown and let Θ represent the parameter space of all possible values of θ.

It is desired to estimate θ using an estimate d under the quadratic loss

$$L(\theta,d) = (d - \theta)^t \mathcal{D} (d - \theta), \tag{1.1.1}$$

where $\mathcal{D}$ is a known positive definite matrix. The risk of a nonrandomized estimator δ for a particular value of θ is defined to be

$$R(\theta,\delta) = E_\theta^X[L(\theta,\delta(x)] = \int_{\mathcal{X}} L(\theta,\delta(x))f(x|\theta)dx, \tag{1.1.2}$$

where $f(x|\theta)$ is the probability density function of X given θ with respect to Lebesgue measure. Here E stands for expectation, with subscripts denoting parameter values at which the expectation is to be taken and superscripts denoting random variables over which the expectation is to be taken.

An estimator δ_1 is defined to be as good as δ_2 if

$$R(\theta,\delta_1) \leq R(\theta,\delta_2) \tag{1.1.3}$$

for all $\theta \in \Theta$. The estimator δ_1 is said to be better than δ_2 (or dominates δ_2) if, in addition to (1.1.3),

$$R(\theta,\delta_1) < R(\theta,\delta_2) \tag{1.1.4}$$

for some $\theta \in \Theta$. The estimator δ is admissible if there exists no better estimator, and is inadmissible otherwise.

We will be concerned with prior information about the parameter θ, that is, information available before X is observed. A convenient way of describing information about θ is by means of a probability distribution called a prior distribution on Θ. We denote the corresponding prior density with respect to Lebesgue measure (if it exists) by $\pi(\theta)$. Given an estimator $\delta(x)$, the Bayes risk of δ is then defined to be

$$\gamma(\pi,\delta) = E^{\pi}[R(\theta,\delta)] = \int_{\Theta} R(\theta,\delta)\pi(\theta)d\theta \quad . \tag{1.1.5}$$

If the prior density π depends only on a parameter A, we will denote the Bayes risk of δ as $\gamma(A,\delta)$ instead of $\gamma(\pi,\delta)$.

An estimator δ_1 is said to be better than δ_2 with respect to a prior π if

$$\gamma(\pi,\delta_1) < \gamma(\pi,\delta_2) \; . \tag{1.1.6}$$

In this paper, we will often consider sum of squares error loss instead of quadratic loss as in (1.1.1). This can be done without loss of generality by transforming the problem by $\mathfrak{D}^{\frac{1}{2}}$. Defining $Y = \mathfrak{D}^{\frac{1}{2}}X$, $\eta = \mathfrak{D}^{\frac{1}{2}}\theta$ and $d^* = \mathfrak{D}^{\frac{1}{2}}d$ for the estimation of θ in the k-variate normal problem, we have $Y \sim N_k(\eta,\Sigma^*)$ where $\Sigma^* = \mathfrak{D}^{\frac{1}{2}}\Sigma\,\mathfrak{D}^{\frac{1}{2}}$. The loss (1.1.1) is then reduced to

$$L(\theta,d) = (d-\theta)^t\mathfrak{D}^{\frac{1}{2}}\,\mathfrak{D}^{\frac{1}{2}}(d-\theta) = (d^*-\eta)^t(d^*-\eta) \; ,$$

which is sum of squares error loss.

Section 1.2. History of the Problem

Let X be a k-variate ($k \geq 3$) vector normally distributed with mean θ and known positive definite covariance matrix Σ, that is $X \sim N_k(\theta, \Sigma)$. Assume the loss in estimating θ is (1.1.1). The classical estimator $\delta^0(X) = X$ is the MLE and MVUE. For the squared error loss function, that is $\mathcal{Q} = I$ in (1.1.1), and for $\Sigma = I$, James and Stein (1961) showed that δ^0 is inadmissible when $k \geq 3$ and that the estimator

$$\hat{\delta}(X) = (I - \frac{k-2}{X^tX})X \qquad (1.2.1)$$

has uniformly smaller risk than δ^0. Since the work of James and Stein, many generalizations have been done.

The James-Stein estimator shows the greatest improvement over the MVUE near the origin; likewise most alternative estimators show substantial improvement in only a particular region of the parameter space. Therefore if a user wants to find an estimator which is significantly better than the usual one for his problem, he should specify a region in which he would like the substantial improvement to occur. In other words, utilization of prior information concerning θ seems necessary for the development of good alternative estimators. If the user does not have any prior knowledge concerning the parameters to be estimated, then he may as well use the usual estimator, since any improved estimator will be unlikely to show much improvement at the true value of the parameter.

The James-Stein estimator allows incorporation of information about the prior mean but its main drawback is that it does not incorporate the higher order moments of the prior in order to adjust its region of significant improvement. From this point of view, Berger (1980) developed a generalized Bayes estimator for this situation which incorporates prior information in the form of a mean vector μ and a covariance matrix A. For $n > 0$, Berger considered the generalized prior density

$$g_n(\theta) = \int_0^1 |B(\lambda)|^{-\frac{1}{2}} \exp\{-\theta^t B(\lambda)^{-1}\theta/2\}\lambda^{(n-1-k/2)} d\lambda \quad (1.2.2)$$

where $B(\lambda) = \lambda^{-1}c-\Sigma$, $0 < \lambda < 1$ and $c = \rho(\Sigma+A)$ for some constant ρ. The above prior has extremely flat tails, so that the resulting generalized Bayes estimator is very robust. The estimator which Berger found can be written as

$$\delta^*(X) = \mu + \left(I - \frac{\gamma(||X-\mu||^2/\rho)\Sigma(\Sigma+A)^{-1}}{||X-\mu||^2}\right)(X-\mu) , \quad (1.2.3)$$

where $||X-\mu||^2 = (X-\mu)^t(\Sigma+A)^{-1}(X-\mu)$.

The estimator $\delta^*(X)$ is robust with respect to misspecification of prior information, is admissible and is sometimes stable in the ridge sense. Also when k, the dimension of the problem, is large and the prior information is correct, $\delta^*(X)$ will perform like the Bayes estimator using the conjugate normal prior.

Section 1.3. Work in this Paper

In this paper we will consider two types of robust generalized Bayes estimators. We will consider the robust generalized Bayes estimator given as

$$\delta^n(X) = \left(I - \frac{\gamma_n(||X||^2/\rho)\Sigma(\Sigma+A)^{-1}}{||X||^2}\right) X , \quad (1.3.1)$$

where $||X||^2 = X^t(\Sigma+A)^{-1}X$, $n = (k-2)/2$, $\rho = (k-2)/k$ and

$$\gamma_n(v) = \frac{v \int_0^1 \lambda^n \exp\{-\lambda v/2\}d\lambda}{\int_0^1 \lambda^{n-1} \exp\{-\lambda v/2\}d\lambda} . \quad (1.3.2)$$

Explicit formulas for γ_n exist and are given in Berger (1980). Also the motivation for the choice of n and ρ is given in Berger (1980). Often we will consider a simpler version of the above estimator to work with, namely

$$\delta(X) = (I - \frac{(k-2)\Sigma(\Sigma+A)^{-1}}{||X||^2}) X . \qquad (1.3.3)$$

The Bayes risk of (1.3.3) will be obtained easily and explicitly but the estimator definitely has some disadvantages compared to (1.3.1). The main drawback is for small X_i's, $||X||^2$ becomes small, so that the shrinking factor blows up which makes the estimator bad.

Subsection 1.3.1. Introduction

The major problem which does arise in the case of the robust generalized Bayes estimators is that the prior may be bad since it forces the coordinate to act together. The performance of the robust generalized Bayes estimator is best when all coordinates are similar or can be transformed so that they are similar. As for example, if there are two groups of similar coordinates, that is the variation within the groups is small and between the groups is large in some sense, it will probably be better to estimate each group separately. In terms of a prior, this can be interpreted as saying the θ_i's should be separated into independent similar groups.

We will assume that the k-parameters can be divided into s-many groups of sizes $k_1, k_2, \ldots, k_s$, where $k_\ell \geq 3$, for all $\ell = 1,2,\ldots,s$ and $\sum_{\ell=1}^{s} k_\ell = k$. Sometimes the groups are formed by a natural decomposition, like left handed or right handed players in baseball (Efron and Morris 1973). Also the groupings may be done on the basis of prior information.

We will use a robust generalized Bayes estimator in each group and also in the combined problem.

The estimator which will be used to estimate the whole parameter set will be called the "Combined Estimator" and the estimator which will be used to estimate the groups individually will be called the "Separate Estimator". The "combined" and the "separate" estimators generally have risk functions which cross, that is, the performance of one estimator is better than the other in one region and worse in another region. Therefore, we will use criterion involving Bayes risks to decide whether or not to separate.

The prior used to develop the robust generalized Bayes estimator is not felt to be absolutely correct, so we want to investigate the Bayes risks of the robust rules under an assortment of plausible priors. To obtain the Bayes risks of the combined and the separate estimators, we will assume that the prior distribution of θ has mean μ and covariance matrix Σ. We will consider first the conjugate normal prior so that the Bayes risks can be computed easily and then several flat tailed priors.

To compare the combined and the separate estimator in terms of the Bayes risk, we will consider the sum of squares error loss function. The original loss was transferred to sum of squares error loss in Section 1.1, so that the loss for the combined estimator is the sum of the losses of the separate estimators.

Subsection 1.3.2. Previous Results on Separation

In the following, some previous results concerning the problem of separation will be discussed. First, the results of Efron and Morris (1973b), concerning the problem of separation of James-Stein estimators will be discussed. Next, we will consider the results of Stein (1974) in which the modification of individual coordinate is sharply limited.

A. Efron and Morris' Results. Efron and Morris assumed that $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ could be naturally divided into two groups as $\theta = (\theta_{(1)}, \theta_{(2)})$. If $\delta = (\delta_{(1)}, \delta_{(2)})$ was an estimate of θ, then the loss function for estimating $\theta_{(i)}$ was assumed to be

$$L_{(i)}(\theta_{(i)}, \delta_{(i)}) = \frac{1}{k_i} |\delta_{(i)} - \theta_{(i)}|^2, \quad i = 1,2,$$

and that of estimating θ was

$$L(\theta, \delta) = \frac{1}{k} |\delta - \theta|^2 ,$$

where $k = k_1 + k_2$.

The authors suggested the scale invariant estimators

$$\delta_{(j)} = [1 - \frac{k-2}{S} \rho_j (\frac{S_j}{S})] X_{(j)}, \quad j = 1,2, \qquad (1.3.4)$$

where

$$S_j = |X_{(j)}|^2, \quad j = 1,2,$$

$$S = S_1 + S_2$$

and ρ_j, the "relevance function" determines the relevance of the two groups of the observation $X_{(1)}$ and $X_{(2)}$ to the estimation of $\theta_{(i)}$. The choices $\rho_j(X) \equiv 1$ and $\rho_j(X) = \frac{k_j - 2}{k-2} \cdot \frac{1}{X}$ will lead to the combined and the separate James-Stein rule respectively. The authors defined the concept of the "relative savings loss" and obtained the results of separation in terms of the relative savings loss.

The main drawbacks of Efron and Morris' results were, they did not use an additive loss function and the estimator they used was not a good one. Therefore, unlike the James-Stein estimator, we will choose the robust generalized Bayes estimator. Our approach will also differ in terms of the estimators. Instead of relative savings loss, we will consider the Bayes risks of the combined and the separate estimator.

B. Stein's Results. Stein (1974) considered a modification of the James-Stein estimate which limits the amount by which any coordinate of the estimate can differ from the corresponding coordinate of X. The estimate he considered was based on order statistics. He was trying to protect against flat tailed priors. He computed the relative efficiency of the estimate with respect to the James-Stein estimate in the case where the θ_i's were independently normally distributed with variance τ^2, just to make sure not too much was lost. From the numerical efficiencies, he concluded that the loss due to truncation (for a normal prior) was small enough.

We will explicitely consider the case when the prior has a flat tailed distribution and find the optimum truncation point for several such priors.

Subsection 1.3.3. Summary of Results Obtained in this Paper

Given a set of coordinates, it is known partially what shrinkage type estimators to use. It is not known however what coordinates to use. In the case of estimating the mean vector of a multivariate normal distribution, the question of choice of coordinates is considered. In this, the question often arises whether to use all coordinates in one combined shrinkage estimator or separate into groups and use shrinkage estimators on each group.

We consider the robust generalized Bayes estimator in the combined and the separate problems under various priors. We first consider the case of accurately specified priors and show that, somewhat surprisingly, the combined estimator is better than the separate estimator. In Chapter 2, it is assumed that the prior mean θ has a k-variate normal distribution with mean 0 and positive definite covariance matrix A. It is assumed that Σ and A are block diagonal matrices with block sizes $k_\ell \geq 3$, $\ell = 1,2,\ldots,S$; so that the problem is decomposed into S-many groups with $\ell^{\underline{th}}$ group size k_ℓ. If an individual group size is less than 3, then the maximum likelihood estimator is used instead of a shrinkage type estimator for that group. It is proved then that the combined estimator given in (1.3.3) is better than the separate estimator based on S-groups, in terms of Bayes risk under the normal prior. The separate estimator for the $\ell^{\underline{th}}$ group is defined as

$$\delta_{(\ell)}(X) = I_{k_\ell} - \frac{(k_\ell-2)\Sigma_{\ell\ell}(\Sigma_{\ell\ell}+A_{\ell\ell})^{-1}}{||X_{(\ell)}||^2}\ ) \ X_{(\ell)}, \qquad \ell = 1,2,\ldots,S, \tag{1.3.5}$$

where $\Sigma_{\ell\ell}$ and $A_{\ell\ell}$ are respectively the $\ell^{\underline{th}}$ partition matrix of Σ and A and $||X_{(\ell)}||^2 = X^t_{(\ell)}(\Sigma_{\ell\ell}+A_{\ell\ell})^{-1}X_{(\ell)}$.

The estimator of the form (1.3.1) is then considered; the corresponding estimator for the $\ell^{\underline{th}}$ group is then defined as

$$\delta^{n_\ell}_{(\ell)}(X) = (I_{k_\ell} - \frac{\gamma_{n_\ell}(||X_{(\ell)}||^2/\rho_\ell)\Sigma_{\ell\ell}(\Sigma_{\ell\ell}+A_{\ell\ell})^{-1}}{||X_{(\ell)}||^2})X_{(\ell)}, \qquad \ell=1,2,\ldots,S. \tag{1.3.6}$$

where $n_\ell = (k_\ell-2)/2$, $\rho_\ell = (k_\ell-2)/k_\ell$ and $\gamma_{n_\ell}(\cdot)$ is defined as $\gamma_n(\cdot)$ in (1.3.2) with n replaced by n_ℓ, $\ell = 1,2,\ldots,S$. By Monte Carlo

Simulations, it is indicated that the combined estimator is better than the separate estimator with respect to Bayes risk under normal prior for arbitrary groups.

In Section 2.2 of Chapter 2, a minimax estimator is considered which was developed by Berger (1979). It is assumed that Σ and A are diagonal with diagonal elements satisfying some conditions. It is then shown that under a normal prior, the combined minimax estimator is better than the separate minimax estimator. In Section 2.2.1, an example is considered for the two groups case which is not covered in the theorem, and it is indicated that the combined minimax estimator is better than the separate minimax estimator.

In Chapter 3, various flat tailed priors are considered. The θ_i are independently given flat tailed priors so that the coordinates are not forced to act together. Again it is indicated numerically that the combined estimator of the form (1.3.1) is better than the separate estimator (1.3.5) under the flat tailed prior. In Section 3.2 of Chapter 3, the asymptotic results for separation are discussed. The combined estimator of the form (1.3.2) and the separate estimator of the form (1.3.5) are considered. It is shown that for a normal prior, asymptotic separation is worse. In general, however, asymptotic separation is shown to be worse only when the fourth moment of the prior is small enough.

In Chapter 4, we consider the question of inclusion of extreme observations, when a flat tailed prior is suspected. The motivation for consideration of flat tailed priors is that most likely the θ_i will be occuring according to a flat tailed prior. This question was first studied by Stein (1974), who obtained partial answers. Stein considered a truncated estimator based on order statistics. We consider a broad class of flat tailed priors in Section 4.3 and get the optimum truncation points in each case for the shrinkage estimator.

Finally we consider in Chapter 5, the situation in which part of the prior information may be "misspecified", corresponding to a situation in which certain coordinates have much less prior information than others. It is assumed that $x|\theta \sim N_k(\theta,\Sigma)$, $\theta \sim N_k(0,A)$ and the problem has a decomposition into two groups of size k_1 and k_2 respectively. It is assumed that one of the groups, say the second group, is misspecified, that is, it has less certain prior information than the other. The amount of misspecification should be incorporated to develop the generalized Bayes estimator in this situation, as will be discussed in detail in Chapter 5. However for the sake of tractable calculation, we will assume that the amount of misspecification ρ, is reflected only in the unconditional covariance matrix of the second group. This amount of misspecification ρ can roughly be considered to be the ratio of the uncertainties in the guesses for the prior variances of the two groups. It is then shown that if $\rho \neq 1$, asymptotically (i.e., the $k_\ell \to \infty$, $\ell = 1,2$) separation is better. It is also shown asymptotically that if $k_1 = \varphi k_2$ and $\rho = 1 + \frac{\gamma}{\sqrt{k_1}}$ then under the misspecification model, $\lim_{k_1 \to \infty} \Delta \gtreqless 0$ according as $\gamma^2 \gtrless \frac{2(1+\varphi)}{\varphi}$, where Δ represents the difference of the Bayes risk of the combined to the separate estimator. It is clear that $\Delta > 0$ implies that separation is better. Thus asymptotically a region is obtained in terms of the amount of misspecification, where separation is better.

In Section 5.3 and 5.5, attention is restricted to the finite case and the region where the separation is better is obtained numerically.

Some tables are made for different cases. It is also indicated how large k_1 should be so that the asymptotic bounds coincide with the numerically obtained bounds.

CHAPTER II

SEPARATION UNDER NORMAL PRIORS

In this chapter we will consider the separation problem under normal priors. As before we have the problem of estimating the mean of a k-variate normal distribution under the squared error loss function, $L(\theta,\delta) = ||\theta-\delta||^2$. We will assume that θ has a k-variate normal distribution with mean 0 and positive definite covariance matrix A. Although a normal prior may not be very realistic, it allows explicit calculation of the Bayes risks of the estimators we are considering. Other more flat tailed priors will be considered later.

In Section 2.1, we will consider the estimator of the form (1.3.3), both in the combined and the separate problems for ease of calculation. We will show that the combined estimator is better than the separate estimator. Then we will consider the estimator of the form (1.3.2) and indicate similar results by Monte Carlo simulation.

In Section 2.2, similar results will be obtained for the minimax estimator as defined in (1.3.7).

Section 2.1. The Separation Problem for the Robust Generalized Bayes Estimator

Subsection 2.1.1. The Bayes Risk Calculation

We have $X|\theta \sim N_k(\theta,\Sigma)$ and $\theta \sim N_k(0,A)$. Assume Σ and A are positive definite with characteristic roots $d_1,d_2,\ldots,d_k$ and $a_1,a_2,\ldots,a_k$ respectively.

Without loss of generality we can assume the prior mean μ is zero (providing the prior mean has been correctly specified), so that as in Berger (1980) the risk of the estimator (1.2.3) reduces to

$$R(\theta,\delta^*)$$
$$= tr\Sigma + E_\theta[- \frac{2\gamma(||X||^2/\rho)}{||X||^2}\{tr\, \Sigma^2(\Sigma+A)^{-1} - \frac{2X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^2}\}$$
$$- \frac{4\gamma'(||X||^2/\rho)X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^2} + \frac{\gamma^2(||X||^2/\rho)X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^2}], \quad (2.1.1)$$

where $||X||^2 = X^t(\Sigma+A)^{-1}X$ and γ' is the derivative of γ. For the estimator (1.3.3), $\gamma(||X||^2) = k - 2$, so the risk of (1.3.3) is given as

$$R(\theta,\delta) = tr\Sigma + E_\theta[- \frac{2(k-2)}{||X||^2}\{tr\, \Sigma^2(\Sigma+A)^{-1} - \frac{2X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^2}\}$$
$$+ \frac{(k-2)^2X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^4}]. \quad (2.1.2)$$

Now it is clear that marginally $X \sim N_k(0,\Sigma+A)$, so that the Bayes risk of the estimator δ defined in (1.3.3) is given as

$$\gamma(A,\delta) = E^A[R(\theta,\delta)]$$
$$= tr\Sigma + E^X[- \frac{2(k-2)}{||X||^2}\{tr\, \Sigma^2(\Sigma+A)^{-1} - \frac{2X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^2}\}$$
$$+ \frac{(k-2)^2X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^4}], \quad (2.1.3)$$

where $E^X(\cdot)$ stands for the expectation under the marginal distribution of X.

To evaluate the Bayes risk in (2.1.3), we have to calculate the expectations of the terms in the bracket.

We know $X \sim N_k(0,\Sigma+A)$, so $X^t(\Sigma+A)^{-1}X$ has a chi-square distribution with k-degrees of freedom i.e., $||X||^2 \sim \chi_k^2$. Thus,

$$E^X\left[\frac{1}{||X||^2}\right] = \frac{1}{2\Gamma(\frac{k}{2})}\int_0^\infty \frac{1}{y}\, e^{-y/2}\left(\frac{y}{2}\right)^{\frac{k}{2}-1} dy$$

$$= \frac{\Gamma(k/2-1)}{2\Gamma(k/2)} = \frac{1}{k-2}. \tag{2.1.4}$$

Now let $\mathcal{O}$ be a k×k orthogonal matrix chosen so that

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

is diagonal, with diagonal elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$. Let a random variable Y be defined as $Y = \mathcal{O}(\Sigma+A)^{-\frac{1}{2}}X$. Thus we have

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

and

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

Therefore,

$$E^X\left[\frac{X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^4}\right] = E\left[\frac{\sum_{i=1}^k \lambda_i Y_i^2}{(\sum_{i=1}^k Y_i^2)^2}\right]$$

$$= \frac{1}{k(k-2)}\sum_{i=1}^k \lambda_i = \frac{1}{k(k-2)}\,\mathrm{tr}\,\Sigma^2(\Sigma+A)^{-1}, \tag{2.1.7}$$

using Lemma 1 of the appendix, since Y_i^2, $i = 1,2,\ldots,k$ are independent χ_1^2 and by the additive property of the chi-square distribution, $\sum_{i=1}^k Y_i^2 \sim \chi_k^2$.

Now we get the Bayes risk of (1.3.3) by substituting (2.1.4) and (2.1.7) in (2.1.3) as

$$\gamma(A,\delta) = \text{tr}\ \Sigma - \left(\frac{k-2}{k}\right) \text{tr}\ \Sigma^2(\Sigma+A)^{-1}. \tag{2.1.8}$$

Similarly, the Bayes risk of (1.3.5) is given as

$$\gamma(A_\ell,\delta_{(\ell)}) = \text{tr}\ \Sigma_{\ell\ell} - \left(\frac{k_\ell-2}{k_\ell}\right)\text{tr}\ \Sigma_{\ell\ell}^2(\Sigma_{\ell\ell}+A_{\ell\ell})^{-1}, \quad \ell = 1,2,\ldots,s. \tag{2.1.9}$$

We will now use the above calculations to prove the following.

Subsection 2.1.2. The Separation Theorem Under Normal Prior

Theorem 2.1.1. Assume that $X|\theta \sim N_k(\theta,\Sigma)$, $\theta \sim N_k(0,A)$ with loss $L(\theta,\delta) = ||\theta - \delta||^2$. Suppose we have s groups of size $k_1,k_2,\ldots,k_s; k_\ell \geq 3$, $\ell = 1,2,\ldots,s$ and $\sum_{\ell=1}^{s} k_\ell = k$. Suppose $X^t = (X_{(1)}^t,\ldots,X_{(s)}^t)$, where $X_{(\ell)}$ is $k_\ell \times 1$, $\ell = 1,2,\ldots,s$, and the corresponding partition of Σ and A has the form

$$\Sigma = \begin{pmatrix} \Sigma_{11} & & & 0 \\ & \Sigma_{22} & & \\ & & \ddots & \\ 0 & & & \Sigma_{ss} \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} A_{11} & & & 0 \\ & A_{22} & & \\ & & \ddots & \\ 0 & & & A_{ss} \end{pmatrix}, \tag{2.1.10}$$

where $\Sigma_{\ell\ell}$ and $A_{\ell\ell}$ are $k_\ell X k_\ell$, $\ell = 1,2,\ldots,s$. Then the combined estimator (1.3.3) is better than the separate estimator (1.3.5) in terms of the Bayes risk.

Proof. Define

$$\Delta = \gamma(A,\delta) - \sum_{\ell=1}^{s}\gamma(A_\ell,\delta_{(\ell)}), \tag{2.1.11}$$

which represents the difference of the Bayes risks of the coimbined and the separate estimator. Using (2.1.8) and (2.1.9), (2.1.11) can be written as

$$\Delta = \text{tr}\,\Sigma - (\frac{k-2}{k})\,\text{tr}\,\Sigma^2(\Sigma+A)^{-1} - \sum_{\ell=1}^{s} \text{tr}\,\Sigma_{\ell\ell} + \sum_{\ell=1}^{s} (\frac{k_\ell-2}{k_\ell})\,\text{tr}\,\Sigma_{\ell\ell}^2(\Sigma_{\ell\ell}+A_{\ell\ell})^{-1}$$

$$< \frac{k-2}{k}\{\sum_{\ell=1}^{s} \text{tr}\,\Sigma_{\ell\ell}^2(\Sigma_{\ell\ell}+A_{\ell\ell})^{-1} - \text{tr}\,\Sigma^2(\Sigma+A)^{-1}\}, \quad (\text{since, } \frac{k_\ell-2}{k_\ell} < \frac{k-2}{k})$$

$$= 0.$$

This completes the proof. ||

Subsection 2.1.3. Monte Carlo Simulation Results

In this section a Monte Carlo simulation study for the separation problem will be discussed. Simulating normal random variables, the Bayes risk of the combined estimator (1.3.1) and the separate estimator (1.3.5) will be calculated. It will be indicated as in earlier results of this section, that the combined estimator is better than the separate estimator in terms of the Bayes risk under a normal prior.

Berger (1980) showed that if A is the true covariance matrix of the prior, then C is chosen as $C = \rho(\Sigma+A)$ where $\rho = (k-2)/k$. Therefore, in our simulation study, we have considered $\rho = (k-2)/k$ for the estimator (1.3.1) and $\rho = (k_\ell-2)/k_\ell$ for (1.3.5). We have considered the case where both Σ and A are diagonal with diagonal elements $(d_1,d_2,\ldots,d_k)$ and $(a_1,a_2,\ldots,a_k)$ respectively. It is assumed that we have two groups of size k_1 and k_2 with $k_1 + k_2 = k$. The simulation procedure consists of generating M standard normal random variables. Then making a nonsingular transformation, normal random variables with desired variances are obtained. The d_i's and a_i's are chosen in such a way that the unconditional variances of the X_i's are not much different within the groups but significantly different between the groups. This would intuitively be the case in which one would be most likely to want to estimate the groups separately.

Let us define the difference of the Bayes risk of the combined estimator (1.3.1) to the separate estimator (1.3.5) as

$$\Delta_n = \gamma(A,\delta^n) - \sum_{\ell=1}^{s} \gamma(A_\ell,\delta_\ell^{n_\ell}) , \qquad (2.1.12)$$

where $\gamma(A,\delta^n) = E^A[R(\theta,\delta^n)]$ and $\gamma(A_\ell,\delta_\ell^{n_\ell}) = E^{A_\ell}[R(\theta_{(\ell)},\delta^{n_\ell})]$, $\ell = 1,2$; and $R(\theta,\delta^n)$ are obtained from (2.1.1) by substituting $\gamma_n(||X||^2)$ for $\gamma(||X||^2)$.

In Table 1, values of Δ_n have been calculated. We have set k = 10 and M = 1000. We have given $c_i = d_i + a_i$, i = 1,2,...,k as input data. Subgroup sizes of 4 and 6 have been considered. The standard error of the values of Δ_n is found to be about .05. Table 1 shows that all values of Δ_n are negative, which indicates that the combined estimator is better than the separate estimator.

Table 1

Difference of Bayes risks for combined and separate estimators.
(under normal prior)
M = 1,000, K = 10

K_1	(C_i), i = 1,2,...,10	Δ
4	(2, 2, 2, 2, 2, 2, 2, 2, 2, 2)	- .78
4	(11, 11, 11, 11, 11, 11, 11, 11, 11, 11)	- .13
4	(2, 2, 2, 2, 11, 11, 11, 11, 11, 11)	- .53
4	(1.01, 1.01, 1.01, 1.01, 11, 11, 11, 11, 11, 11)	-1.06
6	(2, 2, 2, 2, 2, 2, 2, 2, 2, 2)	- .79
6	(2, 2, 2, 2, 11, 11, 11, 11, 11, 11)	- .34
6	(2, 2, 2, 2, 2, 2, 3, 3, 3, 3)	- .63

Section 2.2. The Separation Problem for the Minimax Estimator

In this section the separation problem for the minimax estimator will be considered. The prior will be assumed to have a normal distribution with mean 0 and covariance matrix A. As in Berger (1979), the minimax estimator is defined as $\delta'(X)$ whose $i^{\underline{th}}$ component is given by

$$\delta_i'(X) = \sum_{j=i}^{k} \alpha_i^j \delta_i^j(X) , \tag{2.2.1}$$

where $0 \leq \alpha_i^j \leq 1$, $\alpha_i^j = 0$ for $j < i$ and $\sum_{j=i}^{k} \alpha_i^j = 1$. The estimator $\delta^j(X)$ is chosen here as

$$\delta^j(X) = (I_j - \frac{(j-2)\Sigma^j(\Sigma^j+A^j)^{-1}}{||X^j||^2})X , \tag{2.2.2}$$

where $||X^j||^2 = X^{jt}(\Sigma^j+A^j)^{-1}X^j$, and Σ^j and A^j are the (jxj) upper left corner matrices of Σ and A respectively. The separate minimax estimator will be defined as in (2.2.1) with dimension replaced by k_ℓ, $\ell = 1,2,\ldots,s$.

As before, we have $X|\theta \sim N_k(\theta,\Sigma)$ and $\theta \sim N_k(0,A)$ so that $X \sim N_k(0,\Sigma+A)$. An orthogonal transformation $\mathcal{O}$ is made on $\Sigma(\Sigma+A)^{-1}\Sigma$ such that

$$\Lambda = \mathcal{O}^t\Sigma(\Sigma+A)^{-1}\Sigma\mathcal{O} \tag{2.2.3}$$

is diagonal, with diagonal elements

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0. \tag{2.2.4}$$

Let us now define as in Berger (1979)

$$\rho_i^* = \begin{cases} \lambda_i/\lambda_1 & \text{if} \quad i \leq k \\ 0 & \text{if} \quad i > k \end{cases} , \tag{2.2.5}$$

$$\tau_j = \min\{\frac{(j-2)}{2(j-\sum_{i=1}^{j}\rho_i^*)}, 1\} , \quad j \geq 3, \quad \tau_j = 1 \text{ for } j = 1,2, \tag{2.2.6}$$

$$\gamma_i = \begin{cases} \dfrac{[1-\tau_i(1-\rho^*_{i+1})]}{[1-\tau_i(1-\rho^*_i)]} , & \text{if} \quad i \leq k \\ 1 , & \text{if} \quad i > k \end{cases} , \tag{2.2.7}$$

and

$$\alpha_i^j = \begin{cases} 0 , & \text{if} \quad j \leq i \\ \dfrac{(\rho^*_j-\rho^*_{j+1})[1-\tau_j(1-\rho^*_i)]}{\rho^*_i[1-\tau_j(1-\rho^*_j)]} \prod_{\ell=j+1}^{k+1} \gamma_\ell , & \text{if} \quad j > i \end{cases} \tag{2.2.8}$$

We will now evaluate the Bayes risk of δ' as defined in (2.2.1) under the normal prior.

Subsection 2.2.1. The Bayes Risk Calculation for the Minimax Estimator

In this subsection we will compute the Bayes risk of the minimax estimator δ' given in (2.2.1).

Lemma 2.2.1. Let Σ = diag $(d_1,d_2,...,d_k)$, A = diag $(a_1,a_2,...,a_k)$ and $\gamma(A,\delta')$ be the Bayes risk of δ' as defined in (1.3.7). Then

$$\gamma(A,\delta') = \operatorname{tr} \Sigma - \sum_{i=3}^{k} \lambda_i \sum_{j=i}^{k} \alpha_i^j \left(\frac{j-2}{j}\right)\{2 \sum_{\ell=i}^{i} \alpha_i^\ell - \alpha_i^j\}$$

$$- \sum_{i=1}^{2} \lambda_i \sum_{j=3}^{k} \alpha_i^j \left(\frac{j-2}{j}\right)\{2 \sum_{\ell=i}^{j} \alpha_i^\ell - \alpha_i^j\} . \tag{2.2.9}$$

Proof. Clearly from (2.2.3), $\lambda_i = \dfrac{d_i^2}{a_i+d_i}$, $i = 1,2,...,k$. Then we have

$$\delta_i^j(X) = \left(1 - \frac{\lambda_i}{\alpha_i} \frac{j-2}{||X^j||^2}\right)X_i, \quad i = 1,2,...,k, \tag{2.2.10}$$

and hence from (1.3.7),

$$\delta_i'(X) = \sum_{j=i}^{k} \alpha_i^j \left(1 - \frac{\lambda_i}{d_i} \frac{(j-2)^+}{||X^j||^2}\right)X_i = \left[1 - \frac{\lambda_i}{d_i} \sum_{j=i}^{k} \frac{\alpha_i^j (j-2)^+}{||X^j||^2}\right]X_i ,$$

(2.2.11)

since $\sum_{j=i}^{k} \alpha_i^j = 1$, where $X^+ = \begin{cases} X & \text{if} \quad X > 0 \\ 0 & \text{if} \quad X \leq 0 \end{cases}$.

Now the Bayes risk of δ' is

$$\gamma(A,\delta') = E||\delta'-\theta||^2 = E^X[\sum_{i=1}^{k}\{(X_i-\theta_i) - (\frac{\lambda_i}{d_i} \sum_{j=i}^{k} \frac{\alpha_i^j(j-2)^+}{||X^j||^2})X_i\}^2]. \tag{2.2.12}$$

Let us define $g_i(X)$, for all $i = 1,2,\ldots,k$ as

$$g_i(X) = -\frac{\lambda_i}{d_i} \sum_{j=i}^{k} \frac{\alpha_i^j(j-2)^+}{||X^j||^2} X_i . \tag{2.2.13}$$

Now

$$\frac{\partial}{\partial X_i} g_i(X) = -\frac{\lambda_i}{d_i} \sum_{i=1}^{k} \frac{\alpha_i^j(j-2)^+}{||X^j||^2} + \frac{2\lambda_i}{d_i} \frac{X_i^2}{a_i+d_i} \sum_{j=i}^{k} \frac{\alpha_i^j(j-2)^+}{||X^j||^2} . \tag{2.2.14}$$

Thus by Stein's basic identity as in Hudson (1974),

$$\gamma(A,\delta') = \operatorname{tr}\Sigma + E^X\{\sum_{i=1}^{k} \frac{\lambda_i^2 X_i^2}{d_i^2} (\sum_{j=i}^{k} \frac{\alpha_i^j(j-2)^+}{||X^j||^2})^2\}$$

$$+ E^X[4\sum_{i=1}^{k} \frac{\lambda_i}{a_i+d_i} \sum_{j=i}^{k} \alpha_i^j(j-2)^+ \frac{X_j^2}{||X^j||^4} - 2\sum_{i=1}^{k} \lambda_i \sum_{j=i}^{k} \frac{\alpha_i^j(j-2)^+}{||X^j||^2}]. \tag{2.2.15}$$

Now using Lemma 1 of the appendix,

$$E^X(\frac{X_i^2}{||X^j||^4}) = \frac{a_i+d_i}{j(j-2)} , \quad \text{for} \quad j > 2 \tag{2.2.16}$$

and

$$E^X(\frac{X_i^2}{||X_1^j||^2||X_2^j||^2}) = \frac{a_i+d_i}{j_1(j_2-2)} , \quad \text{for} \quad j_2 > 2. \tag{2.2.17}$$

Therefore,

$$\gamma(A,\delta') = \text{tr}\,\Sigma + \sum_{i=1}^{k} \frac{\lambda_i^2}{d_i^2} \Big[\sum_{j=\max(i,3)}^{k} \alpha_i^{j^2} (j-2)^2 \frac{a_i+d_i}{j(j-2)}$$

$$+ 2 \sum_{i \le j_1 < j_2}^{k} \alpha_i^{j_1} \alpha_i^{j_2} (j_1-2)(j_2-2) \frac{a_i+d_i}{j_1(j_2-2)} \Big]$$

$$+ 4 \sum_{i=1}^{k} \frac{\lambda_i}{a_i+d_i} \sum_{j=\max(i,3)}^{k} \alpha_i^j (j-2) \frac{a_i+d_i}{j(j-2)}$$

$$- 2 \sum_{i=1}^{k} \lambda_i \sum_{j=\max(i,3)}^{k} \alpha_i^j (j-2) \frac{1}{(j-2)} \; .$$

Thus

$$\gamma(A,\delta') = \text{tr}\,\Sigma + \sum_{i=1}^{k} \frac{\lambda_i^2}{d_i^2} (a_i+d_i) \{ \sum_{j=i}^{k} \alpha_i^{j^2} \frac{(j-2)^+}{j} + 2 \sum_{j=i}^{k} \alpha_i^j \frac{(j-2)^+}{j} \sum_{\ell=j+1}^{k} \alpha_i^{\ell} \}$$

$$+ 4 \sum_{i=1}^{k} \lambda_i \sum_{j=\max(i,3)}^{k} \frac{\alpha_i^j}{j} - 2 \sum_{i=1}^{k} \lambda_i \sum_{j=n_i}^{k} \alpha_i^j$$

$$= \text{tr}\,\Sigma + \sum_{i=1}^{k} \lambda_i \sum_{j=i}^{k} \alpha_i^j \frac{(j-2)^+}{j} \{ \alpha_i^j + 2 \sum_{\ell=j+1}^{k} \alpha_i^{\ell} \} - 2 \sum_{i=1}^{k} \lambda_i \sum_{j=n_i}^{k} \alpha_i^j \frac{(j-2)}{j}$$

$$= \text{tr}\,\Sigma + \sum_{i=1}^{2} \lambda_i \sum_{j=3}^{k} \alpha_i^j \left(\frac{j-2}{j}\right) \{ \alpha_i^j + 2 \sum_{\ell=j+1}^{k} \alpha_i^{\ell} \} + \sum_{i=3}^{k} \lambda_i \sum_{j=i}^{k} \alpha_i^j \{ \alpha_i^j + 2 \sum_{\ell=j+1}^{k} \alpha_i^{\ell} \} \frac{(j-2)}{j}$$

$$- 2 \sum_{i=1}^{k} \lambda_i \sum_{j=\max(i,3)}^{k} \alpha_i^j \left(\frac{j-2}{j}\right)$$

$$= \text{tr}\,\Sigma + \sum_{i=1}^{2} \lambda_i \sum_{j=3}^{k} \alpha_i^j \left(\frac{j-2}{j}\right) \{ \alpha_i^j + 2 \sum_{\ell=j+1}^{k} \alpha_i^{\ell} - 2 \}$$

$$+ \sum_{i=3}^{k} \lambda_i \sum_{j=i}^{k} \alpha_i^j \left(\frac{j-2}{j}\right) \{ \alpha_i^j + 2 \sum_{\ell=j+1}^{k} \alpha_i^{\ell} - 2 \}$$

$$= \text{tr}\,\Sigma - \sum_{i=3}^{k} \lambda_i \sum_{j=i}^{k} \alpha_i^j \left(\frac{j-2}{j}\right)\{2 \sum_{\ell=i}^{j} \alpha_i^\ell - \alpha_i^j\}$$

$$- \sum_{i=j}^{2} \lambda_i \sum_{j=3}^{k} \alpha_i^j \left(\frac{j-2}{j}\right)\{2 \sum_{\ell=i}^{j} \alpha_i^\ell - \alpha_i^j\}\ , \quad (\text{since} \sum_{j=i}^{k} \alpha_i^j = 1).$$

This completes the proof of the lemma. ||

Subsection 2.2.2. The Separation Theorem for the Minimax Estimator

In the following theorem we will show that the combined minimax estimator is better than the separate minimax estimator under a normal prior.

Theorem 2.2.1. Assume that $X|\theta \sim N_k(\theta,\Sigma)$, $\theta \sim N_k(0,A)$ and the loss is $L(\theta,\delta) = ||\theta-\delta||^2$. Suppose Λ is defined as in (2.2.3) with diagonal elements λ_i satisfying (2.2.4). Assume we have s groups of size $k_1,k_2,\ldots,k_s$, $k_\ell \geq 3$, $\ell = 1,2,\ldots,s$, and $\sum_{\ell=1}^{s} k_\ell = k$ and that the groups are formed in such a way that within each group the λ_i's are the same, that is

$$\Lambda = \text{diag}\{\underbrace{\lambda_1,\ldots,\lambda_1}_{k_1}\ ;\ \underbrace{\lambda_2,\ldots,\lambda_2}_{k_2}\ ;\ \ldots\ ;\ \underbrace{\lambda_s,\ldots,\lambda_s}_{k_s}\}. \quad (2.2.18)$$

Then the combined minimax estimator is better than the separate minimax estimator in terms of Bayes risk with respect to the normal prior.

Proof. Let us now define, for convenience of notation, the sequence of partial sums

$$T_\ell = k_1 + k_2 + \ldots + k_\ell, \quad \ell = 1,2,\ldots,s, \quad T_0 \equiv 0. \quad (2.2.19)$$

Since the λ_i's are assumed to be equal within each group, it follows from (2.2.5) to (2.2.8) that

$$\alpha_i^j = 0 \quad \text{for} \quad j \neq T_1,T_2,\ldots,T_s$$

$$\alpha_i^{T_\ell} > 0 \quad \text{for} \quad \ell = 1,2,\ldots,s\ , \quad (2.2.20)$$

and

$$\sum_{\ell=1}^{s} \alpha_i^{T_\ell} = 1, \quad \text{for all } i.$$

Now to get an upper bound on $\gamma(A,\delta')$ given in (2.2.9), notice that (2.2.9) can also be written as

$$\gamma(A,\delta') = \text{tr}\ \Sigma + \sum_{i=1}^{k} \lambda_i \sum_{j=\max(i,3)}^{k} \alpha_i^{j^2} \left(\frac{j-2}{j}\right) - 2\sum_{i=1}^{k} \lambda_i \sum_{j=\max(i,3)}^{k} \left(\frac{j-2}{j}\right) \alpha_i^j \sum_{\ell=i}^{j} \alpha_i^{\ell}. \tag{2.2.21}$$

Now, using (2.2.18) to (2.2.20) we have

$$\sum_{i=1}^{k} \lambda_i \sum_{j=\max(i,3)}^{k} \alpha_i^{j^2} \left(\frac{j-2}{j}\right) = \sum_{\ell=1}^{s} \lambda_\ell \sum_{i=T_{\ell-1}+1}^{T_\ell} \sum_{j=\ell}^{s} \alpha_i^{T_j^2} \left(\frac{T_j-2}{T_j} \right). \tag{2.2.22}$$

Similarly using (2.2.18) through (2.2.20) we have

$$\sum_{i=1}^{k} \lambda_i \sum_{j=\max(i,3)}^{k} \left(\frac{j-2}{j}\right) \alpha_i^j \sum_{\ell=i}^{j} \alpha_i^{\ell} = \sum_{\ell=1}^{s} \lambda_\ell \sum_{i=T_{\ell-1}+1}^{T_\ell} \sum_{j=\ell}^{s} \left(\frac{T_j-2}{T_j}\right) \alpha_i^{T_j} \left(\sum_{h=\ell}^{j} \alpha_i^{T_h} \right). \tag{2.2.23}$$

Using (2.2.22) and (2.2.23) in (2.2.21) we have

$$\begin{aligned}
\gamma(A,\delta') &= \text{tr}\ \Sigma - \sum_{\ell=1}^{s} \lambda_\ell \sum_{i=T_{\ell-1}+1}^{T_\ell} \sum_{j=\ell}^{s} \left(\frac{T_j-2}{T_j}\right) \{2\alpha_i^{T_j} \sum_{h=\ell}^{j} \alpha_i^{T_h} - \alpha_i^{T_j^2}\} \\
&\leq \text{tr}\ \Sigma - \sum_{\ell=1}^{s} \lambda_\ell \left(\frac{T_\ell-2}{T_\ell}\right) \sum_{i=T_{\ell-1}+1}^{T_\ell} \sum_{j=\ell}^{s} \{\alpha_i^{T_j^2} + 2\alpha_i^{T_j} \sum_{h=\ell}^{j-1} \alpha_i^{T_h}\} \\
&= \text{tr}\ \Sigma - \sum_{\ell=1}^{s} \lambda_\ell \left(\frac{T_\ell-2}{T_\ell}\right) \sum_{i=T_{\ell-1}+1}^{T_\ell} \left(\sum_{j=\ell}^{s} \alpha_i^{T_j}\right)^2 = \text{tr}\ \Sigma - \sum_{\ell=1}^{s} \lambda_\ell \left(\frac{T_\ell-2}{T_\ell}\right) k ,
\end{aligned}$$

$$\left(\text{since, for } i = T_{\ell-1}+1,\ldots,T_\ell,\ \sum_{j=\ell}^{s} \alpha_i^{T_j} = 1\right).$$

Therefore an upper bound for $\gamma(A,\delta')$ is

$$\gamma(A,\delta') \leq \text{tr}\ \Sigma - \sum_{\ell=1}^{s} \lambda_\ell \left(\frac{T_\ell - 2}{T_\ell}\right) k_\ell \ . \tag{2.2.24}$$

Now using (2.2.5) to (2.2.8) we have for the $\ell^{\underline{th}}$ group

$$\alpha_{\ell i}^{j} = \begin{cases} 1 & \text{if} \quad j = T_\ell \\ 0 & \text{otherwise} \ , \end{cases} \tag{2.2.25}$$

$\ell = 1,2,\ldots,s$. Hence using (2.2.25) and Lemma 2.2.1, the Bayes risk of δ_ℓ', $\ell = 1,2,\ldots,s$, reduces to

$$\gamma(A\ ,\delta_\ell') = \text{tr}\ \Sigma_{\ell\ell} - \frac{\lambda_\ell (T_\ell - 2)}{T_\ell} k_\ell, \quad \ell = 1,2,\ldots,s. \tag{2.2.26}$$

Now the difference of the Bayes risk of the combined and the separate minimax estimator is defined as

$$\Delta = \gamma(A,\delta') - \sum_{\ell=1}^{s} \gamma(A_\ell,\delta_\ell'). \tag{2.2.27}$$

Now from (2.2.26) we have,

$$\sum_{\ell=1}^{s} \gamma(A_\ell,\delta_\ell') = \text{tr}\ \Sigma - \sum_{\ell=1}^{s} \lambda_\ell \left(\frac{T_\ell - 2}{T_\ell}\right) k_\ell \ . \tag{2.2.28}$$

Therefore using (2.2.24), (2.2.27) and (2.2.28), the proof is complete.

Subsection 2.2.3. Numerical Example

In this section we will compute the Bayes risk of the combined and the separate minimax estimator for the case of two groups in certain situations not covered by the theorem of the preceding subsection. The indication is again that the combined estimator is better.

In Table 2, we have considered $k = 10$, with two groups of different sizes. We have put the λ_i's as inputs and compute the α_i^j's and hence the Bayes risk (2.2.9). Similarly the Bayes risk of the separate estimator is computed. In all cases it is observed that the difference Δ of the Bayes risks of the combined and the separate estimator are negative which indicates that the combined estimator is better.

Table 2

Differences of Bayes risks for combined and separate minimax estimators. (under normal prior) k = 10

k_1	(λ_i), i = 1,2,...,10	Δ
3	(10, 10, 10, 10, 10, 1, 1, 1, 1)	-22.7857
4		-23.0741
6		- 5.5000
7		- 3.3571
3	(10, 10, 10, 1, 1, 1, .1, .1, .1, .001)	- 5.7437
4		- 1.3413
5		- 1.5669
6		- 0.2044
7		- 0.2360
5	(10, 10, 10, 10, 4, 4, 4, 1, 1, 1)	-11.1048
6	(4, 4, 4, 4, 4, 2, 2, 2, 2, 2)	- 5.3333
6	(4, 4, 4, 4, 4, 1, 1, 1, 1, 1)	- 4.0000
6	(2, 2, 2, 2, 2, 1, 1, 1, 1, 1)	- 2.6667

CHAPTER III

SEPARATION UNDER A FLAT PRIOR

In this chapter we will consider the separation problem for the robust generalized Bayes estimator under a flat prior. In Section 3.1, we will assume that $X_1, X_2, \ldots, X_k$ are independent normal with $E(X_i) = \theta_i$ and $v(X_i) = d_i$ (known), for all $i=i,2,\ldots,k$. Unlike Chapter 2, a class of flat priors will be assumed. This class of flat priors is no longer dependent on $|\theta|$ and has flat tails to reflect the possibility of outliers.

As in the last chapter, we will find the Bayes risk of the combined estimator (1.3.3) and the separate estimator (1.3.5) using the flat prior. A Romberg extrapolation method will be used to evaluate the integrals involved in the Bayes risks evaluation.

In Section 3.2, we will consider the asymptotic results for separation. We will consider the combined estimator of the form (1.3.3) and the separate estimator of the form (1.3.5). Then under the assumption of diagonal covariance matrix and finite eighth moment of the prior, we will evaluate the difference of the Bayes risk of the estimator (1.3.3) and the estimator (1.3.5) for large k. We will show that under a normal prior, the difference is negative. We will then show for general priors that, under the assumption of a finite eighth moment, the difference is positive if the fourth moment of the prior is too large. This is discouraging, since the fourth moment of the prior is rarely knowable.

Section 3.1. Numerical Results for a Flat-Tailed Prior

Assume that, $\theta_i \mid \lambda_i \overset{\text{indep.}}{\sim} N_1(0, b(\lambda_i))$, $i = 1,2,\ldots,k$ where the λ_i are independent and identically distributed with density $f(\lambda) = n\lambda^{n-1}$, for $n > 0$ and $0 < \lambda < 1$, and $b(\lambda_i) = \frac{c_i}{\lambda_i} - d_i$, $i = 1,2,\ldots,k$. The c_i's $i = 1,2,\ldots,k$ are such that $c_i - d_i \geq 0$ for all $i = 1,2,\ldots,k$. The generalized prior density for θ_i is

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

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 represents a class of flat prior densities.

Subsection 3.1.1. The Bayes Risk Calculation

Let us assume $\Sigma = \text{diag}(d_1,\ldots,d_k)$, $A = \text{diag}(a_1,\ldots,a_k)$ and $c = \text{diag}(c_1,\ldots,c_k)$. Clearly $X_i \mid \lambda_i \overset{\text{indep.}}{\sim} N_1(0, c_i/\lambda_i)$ for all $i = 1,2,\ldots,k$. Thus taking $c = \Sigma + A$ and using (2.1.3) we have that the Bayes risk of (1.3.3) is

$$\gamma(A,\delta) = \text{tr}\,\Sigma - 2(k-2)\sum_{i=1}^{k}\frac{d_i^2}{c_i}\, E^{\lambda}E^{X|\lambda}\Big[\frac{1}{\sum_{i=1}^{k}\frac{1}{c_i}X_i^2}\Big]$$

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

where λ is the vector $\lambda = (\lambda_1,\lambda_2,\ldots,\lambda_k)$ and where $E^{X|\lambda}(\cdot)$ stands for the expectation under the conditional distribution of X given λ.

Notice that the unconditional variance of X_i is

$$v(X_i) = \int_0^1 \frac{c_i}{\lambda_i} n\lambda_i^{n-1} d\lambda_i = \frac{nc_i}{n-1}, \quad i = 1,2,\ldots,k. \qquad (3.1.3)$$

Thus in order that $v(X_i) < \infty$, one should have $n \geq 2$.

It is clear that, to calculate the Bayes risk (3.1.2) we have to calculate the expectations

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

and

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

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

Now notice that since the λ_i's are iid, the unconditional distributions of X_i^2/c_i are independent of i. Therefore

$$kE^X\left[\frac{X_i^2/c_i}{\left(\sum_{i=1}^{k} X_i^2/c_i\right)^2}\right] = \sum_{i=1}^{k} E^X\left[\frac{X_i^2/c_i}{\left(\sum_{i=1}^{k} X_i^2/c_i\right)^2}\right] = E^X\left[\frac{1}{\sum_{i=1}^{k} X_i^2/c_i}\right]. \qquad (3.1.6)$$

Using (3.1.4), (3.1.5) and (3.1.6) in (3.1.2) we have

$$\gamma(A,\delta) = \text{tr}\,\Sigma - \frac{(k-2)^2}{k} v.S, \qquad (3.1.7)$$

where $v = \sum_{i=1}^{k} d_i^2/c_i$ and $S = E^X\left[\frac{1}{\sum_{i=1}^{k} X_i^2/c_i}\right]$.

We will compute S by a numerical integration technique.

It is clear that by choosing different values of n, several flat priors can be generated. Let us take $n = 2$, for simplicity, and compute the Bayes risk (3.1.2).

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

$$\mathfrak{L}(t|\lambda_i) = E^{Z_i|\lambda_i}[e^{-tZ_i/c_i}] = (1+\frac{2t}{\lambda_i})^{-\frac{1}{2}}. \qquad (3.1.8)$$

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

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

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

$$\mathfrak{L}(t) = 16t^2\int_0^{\theta^*} \sinh^4\theta d\theta = 16t^2\{\frac{1}{4}\sinh^3\theta^*\cosh\theta^* - \frac{3}{4}\int_0^{\theta^*}\sinh^2\theta d\theta\}$$

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

$$= \sqrt{1+2t} - 3t\sqrt{1+2t} + 6t^2\sinh^{-1}\frac{1}{\sqrt{2t}}. \qquad (3.1.9)$$

Now using the independence of the Z_i/c_i for $i = 1,2,\ldots,k$, 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

$$E[e^{-t\sum_{i=1}^k X_i^2/c_i}] = \{\sqrt{1+2t}(1-3t) + 6t^2\sinh^{-1}\frac{1}{\sqrt{2t}}\}^k. \qquad (3.1.10)$$

Therefore, by Lemma 2 of the appendix, we have

$$S = E\left[\frac{1}{\sum_{i=1}^{k} X_i^2/c_i}\right] = \int_0^\infty \{\sqrt{1+2t}\,(1-3t) + 6t^2 \sinh^{-1}\frac{1}{\sqrt{2t}}\}^k dt. \tag{3.1.11}$$

Notice that the expression (3.1.11) involves an improper integral. Thus to evaluate this by numerical integration using a computer, we need to approximate the integrand for large and small values.

Subsection 3.1.2. Approximation of the Integral

In this section we will approximate the integrand given in (3.1.11). Let us define

$$F(t) = \sqrt{1+2t}\,(1-3t) + 6t^2 \sinh^{-1}\frac{1}{\sqrt{2t}}\,. \tag{3.1.12}$$

We will now study the behavior of $F(t)$ near 0 and ∞.

Lemma 3.1.1. Let $F(t)$ be defined as in (3.1.12). Then

$$\lim_{t\to\infty} 2\sqrt{1+2t}\, F(t) = 1. \tag{3.1.13}$$

Proof. We have $\sinh^{-1}\frac{1}{\sqrt{2t}} = \log_e\left(\frac{1}{\sqrt{2t}} + \sqrt{1+\frac{1}{2t}}\right) = \log_e \frac{\sqrt{1+X}}{\sqrt{X}}\left(1+\frac{1}{\sqrt{1+X}}\right)$,

where $X = 2t$. Therefore,

$$\sinh^{-1}\frac{1}{\sqrt{2t}} = \tfrac{1}{2}\log_e\left(1+\frac{1}{X}\right) + \log_e\left(1+\frac{1}{\sqrt{1+X}}\right)$$

$$= \tfrac{1}{2}\left\{\frac{1}{X} - \frac{1}{2X^2} + o(X^{-3})\right\}$$

$$+ \frac{1}{\sqrt{1+X}} - \frac{1}{2(1+X)} + \frac{1}{3(1+X)^{3/2}} - \frac{1}{4(1+X)^2} + o(X-3)\}$$

$$= \tfrac{1}{2}\left\{\frac{1}{2t} - \frac{1}{8t^2} + o(t^{-3})\right\}$$

$$+ \left\{\frac{1}{\sqrt{1+2t}} - \frac{1}{2(1+2t)} + \frac{1}{3(1+2t)^{3/2}} - \frac{1}{4(1+2t)^2} + o(t^{-3})\right\}.$$

Therefore,

$$6t^2\sqrt{1+2t}\ \sinh^{-1}\frac{1}{\sqrt{2t}} = 3t^2\{\frac{1}{2t} - \frac{1}{8t^2} + o(t^{-3})\}\sqrt{2t}\ (1+\frac{1}{2t})^{\frac{1}{2}}$$

$$+ 6t^2\{1 - \frac{1}{2\sqrt{1+2t}} + \frac{1}{3(1+2t)} - \frac{1}{4(1+2t)^{3/2}} + o(t^{-5/2})\}$$

$$= a(t) + b(t), \text{ say.}$$

Now,

$$a(t) = \{\frac{3t}{2} - \frac{3}{8} + o(t^{-1})\}\{\sqrt{2t} + \frac{1}{2\sqrt{t}} - o(t^{-3/2})\} = \frac{3}{\sqrt{2}}\ t^{3/2} + o(t^{-\frac{1}{2}}),$$

and

$$b(t) = 6t^2 - \frac{3t^2}{\sqrt{2t}}\ (1+\frac{1}{2t})^{-\frac{1}{2}} + t(1+\frac{1}{2t})^{-1} - \frac{3t^2}{2(2t)^{3/2}}\ (1+\frac{1}{2t})^{-3/2} + o(t^{-\frac{1}{2}})$$

$$= 6t^2 - \frac{3}{\sqrt{2}}\ t^{3/2} + t - \tfrac{1}{2} + o(t^{-\frac{1}{2}}).$$

Thus it is clear that

$$6t^2\sqrt{1+2t}\ \sinh^{-1}\frac{1}{\sqrt{2t}} = a(t) + b(t) = 6t^2 + t - \tfrac{1}{2} + o(t^{-\frac{1}{2}}). \tag{3.1.14}$$

Therefore,

$$\lim_{t\to\infty} 2\sqrt{1+2t}\ F(t) = \lim_{t\to\infty} 2\{6t^2 + t - \tfrac{1}{2} + o(t^{-\frac{1}{2}}) + 1 - 6t^2 - t\} = 1,$$

which completes the proof of the lemma. ||

Lemma 3.1.2. Let F(t) be defined as in (3.1.12). Then

$$\lim_{t\to 0} F(t) = 1. \tag{3.1.15}$$

Proof. Clearly, $\lim_{t\to 0}\sqrt{1+2t}(1-3t) = 1$. To find $\lim_{t\to 0} t^2 \sinh^{-1}\frac{1}{\sqrt{2t}}$ let us define $X = \frac{1}{\sqrt{2t}}$. Then

$$\lim_{t\to 0} t^2 \sinh^{-1} \frac{1}{\sqrt{2t}} = \lim_{X\to\infty} \frac{1}{4X^4}\{\log(X+\sqrt{1+X^2})\}$$

$$= \lim_{X\to\infty} \frac{1}{16X^3} [\frac{1}{X+\sqrt{1+X^2}} \{1 + \frac{X}{\sqrt{1+X^2}}\}],$$

using L'Hospital rule. Therefore,

$$\lim_{t\to 0} t^2 \sinh^{-1} \frac{1}{\sqrt{2t}} = \lim_{X\to\infty} \frac{1}{16X^3} \{\frac{1}{\sqrt{1+X^2}}\} = 0. \qquad (3.1.16)$$

Hence,

$$\lim_{t\to 0} F(t) = \lim_{t\to 0} \sqrt{1+2t}\,(1-3t) + \lim_{t\to 0} 6t^2 \sinh^{-1} \frac{1}{\sqrt{2t}} = 1,$$

which completes the proof of the lemma. ||

Now notice from Lemma 3.1.2 that $\lim_{t\to 0}\{F(t)\}^k = 1$. Thus using Lemma 3.1.1 and Lemma 3.1.2 it is clear that near zero, F(t) behaves like 1 and near infinity, F(t) behaves like the function, $\frac{1}{2\sqrt{1+2t}}$. Therefore, there exists a large number M such that for all $t > M$, $F(t) \simeq \frac{1}{2\sqrt{1+2t}}$, where '$\simeq$' stands for approximate equality. Hence

$$\int_M^\infty \{F(t)\}^k dt \simeq \int_M^\infty \{\frac{1}{2\sqrt{1+2t}}\}^k = \frac{1}{2} \int_{2M+1}^\infty \{\frac{1}{2\sqrt{X}}\}^k \quad \text{(where } X = 1+2t)$$

$$= \frac{1}{2^k} \frac{1}{(k-2)(1+2M)^{\frac{k}{2}-1}} .$$

Subsection 3.1.3. Numerical Results

Let us define s_ℓ and v_ℓ for the $\ell^{\underline{th}}$ group as

$$S_\ell = E^{X(\ell)}[\frac{1}{\sum_{i=T_{\ell-1}+1}^{T_\ell} X_i^2/c_i}] , \quad \ell = 1,2,\ldots,s, \qquad (3.1.17)$$

and

$$V_\ell = \sum_{i=T_{\ell-1}+1}^{T_\ell} d_i^2/c_i, \quad \ell = 1,2,\ldots,s, \tag{3.1.18}$$

where T_0 is defined to be 0 and T_ℓ is defined as in (2.2.19). Now using (3.1.17), (3.1.18) and (3.1.2), the Bayes risk of $\delta_{(\ell)}$ as defined in (1.3.5) reduces to

$$\gamma(A_\ell,\delta_{(\ell)}) = \operatorname{tr} \Sigma_{\ell\ell} - \frac{(k_\ell-2)^2}{k_\ell} V_\ell S_\ell, \quad \ell = 1,2,\ldots,s. \tag{3.1.19}$$

Therefore the difference of the Bayes risks Δ reduces to

$$\Delta = \sum_{\ell=1}^{s} \frac{(k_\ell-2)^2}{k_\ell} V_\ell S_\ell - \frac{(k-2)^2}{k} VS. \tag{3.1.20}$$

Using the numerical integration technique given in the previous section, the S and S_ℓ's can be calculated and hence (3.2.20) can be calculated. Note that S and S_ℓ depend only on the k and k_ℓ respectively.

In the numerical study the dimension k is taken as 10 and group sizes 4, 5, 6 and 7 are considered. Different values of d_i and c_i, $i = 1,2,\ldots,10$ are taken as inputs. In Table 3, values of the differences of the Bayes risks of the combined and the separate estimator are computed for different inputs using the numerical integration technique. Table 3 shows that the differences are negative, which indicates that the combined estimator is better than the separate estimator with respect to the Bayes risk, even under a flat prior.

Note. For equal group sizes, that is $k_1 = k_2$, we have $S_1 = S_2$. Then (3.1.20) reduces to

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

Table 3

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

It is indicated numerically 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.

Section 3.2. The Asymptotic Results for Separation

In this section we will consider the robust generalized Bayes estimator of the form (1.3.3). Then under the assumption that Σ is a diagonal matrix and the prior has finite eighth moment, we will evaluate the difference of the Bayes risk of the combined estimator (1.3.3) and the separate estimator (1.3.5) for large k.

Subsection 3.2.1. Evaluation of the Difference of Bayes Risks for Large k

Assume that $X|\theta \sim N_k(\theta,\Sigma)$ where

$$\Sigma = \text{diag}(\underbrace{\sigma_1^2,\ldots,\sigma_1^2}_{k_1},\ldots,\underbrace{\sigma_s^2,\ldots,\sigma_s^2}_{k_s})$$

and the prior on θ is such that $\theta_1,\theta_2,\ldots,\theta_k$ are independent with $E(\theta_i) = 0$, $E(\theta_i^2) = \rho_\ell^2$ and $E(\theta_i^2 - \rho_\ell^2)^2 = V_\ell$ when $T_{\ell-1} + 1 \leq i \leq T_\ell$; $\ell=1,2,\ldots,s$ and the T_ℓ's are defined as in (2.2.19). Suppose also that $E(\theta_i^8) < \infty$, for all $i=1,2,\ldots,k$ and $\rho_\ell^2/\sigma_\ell^2 \leq T$, where T is a constant.

Lemma 3.2.1. $E^X[\dfrac{X_i^2}{\sigma_\ell^2+\rho_\ell^2} - 1]^2 = 2 + \dfrac{V_\ell - 2\rho_\ell^4}{(\sigma_\ell^2+\rho_\ell^2)^2}$ where E^X stands for expectation under the marginal distribution of X.

Proof. We have $X_i|\theta_i \sim N(\theta_i,\sigma_\ell^2)$, $T_{\ell-1} + 1 \leq i \leq T_\ell$. Therefore

$$E^X(X_i^2) = E^{\theta_i}E^{X_i|\theta_i}(X_i^2) = E^{\theta_i}(\sigma_\ell^2+\theta_i^2) = \sigma_\ell^2 + \rho_\ell^2 . \qquad (3.2.1)$$

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

$$E^{X_i|\theta_i}(X_i^4) = E^{X_i|\theta_i}[(X_i-\theta_i)^4 + 4(X_i-\theta_i)^2\theta_i^2 + \theta_i^4]$$

$$= 3\sigma_\ell^2 + 4\theta_i^2\sigma_\ell^2 + \theta_i^4 \ .$$

Thus,

$$E^X(X_i^4) = E^{\theta_i}[3\sigma_\ell^2 + 4\theta_i^2\sigma_\ell^2 + \theta_i^4]$$

$$= 3\sigma_\ell^4 + 6\sigma_\ell^2\rho_\ell^2 + V_\ell + \rho_\ell^4 \ ,$$

$$(\text{since, } E(\theta_i^4) = V_\ell + \rho_\ell^4). \qquad (3.2.2)$$

Now using (3.2.1) and (3.2.2) we have

$$E^X\left[\frac{X_i^2}{\sigma_\ell^2+\rho_\ell^2} - 1\right]^2 = E^X\left[\frac{X_i^4}{(\sigma_\ell^2+\rho_\ell^2)^2} + 1 - \frac{2X_i^2}{\sigma_\ell^2+\rho_\ell^2}\right]$$

$$= \frac{3\sigma_\ell^4+6\rho_\ell^2\rho_\ell^2+\rho_\ell^4+V_\ell}{(\sigma_\ell^2+\rho_\ell^2)^2} - 1$$

$$= \frac{2\sigma_\ell^4+4\sigma_\ell^2\rho_\ell^2+V_\ell}{(\sigma_\ell^2+\rho_\ell^2)^2}$$

$$= \frac{2(\sigma_\ell^2+\rho_\ell^2)^2+V_\ell-2\rho_\ell^4}{(\sigma_\ell^2+\rho_\ell^2)^2}$$

$$= 2 + \frac{V_\ell-2\rho_\ell^4}{(\sigma_\ell^2+\rho_\ell^2)^2} \ .$$

This completes the proof of the lemma. ||

Lemma 3.2.2. $E^X[\dfrac{k_\ell}{||X_{(\ell)}||^2}] = 1 + \dfrac{2}{k_\ell} + \dfrac{V_\ell - 2\rho_\ell^4}{k_\ell(\sigma_\ell^2 + \rho_\ell^2)^2} + o(\dfrac{1}{k_\ell})$.

Proof. We have,

$$||X_{(\ell)}||^2 = \sum_i \frac{X_i^2}{\sigma_\ell^2+\rho_\ell^2} = \sum_i Y_i^2 \quad \text{(say)}.$$

Therefore,

$$E^X[\frac{k_\ell}{||X_{(\ell)}||^2}] = E^Y[\frac{k_\ell}{\sum_i Y_i^2}]$$

$$= 1 - E^Y[\frac{1}{k_\ell} \sum_i (Y_i^2-1)] + E[\frac{1}{k_\ell} \sum_i (Y_i^2-1)]^2 + \varepsilon.$$

where ε is the error.

Observing that the Y_i^2 are independent with $E^Y(Y_i^2) = 1$ and using Lemma 3.2.1, we have

$$E^X[\frac{k_\ell}{||X_{(\ell)}||^2}] = 1 + \frac{1}{k_\ell}\{2 + \frac{V_\ell-2\rho_\ell^4}{(\sigma_\ell^2+\rho_\ell^2)^2}\} + \varepsilon .$$

Now by subtraction, we have

$$\varepsilon = E^Y\{\frac{k_\ell}{\sum_i Y_i^2} - 1 + \frac{\sum_i Y_i^2-k_\ell}{k_\ell} - [\frac{1}{k_\ell} \sum_i(Y_i^2-1)]^2\}$$

$$= E^Y\{\frac{[\sum_i(Y_i^2-1)]^2}{k_\ell \sum_i Y_i^2} - \frac{1}{k_\ell^2}[\sum_i(Y_i^2-1)]^2\}$$

$$= E^Y\{\frac{[\sum_i(Y_i^2-1)]^2}{k_\ell}(\frac{1}{\sum_i Y_i^2} - \frac{1}{k_\ell})\}$$

$$= -E^Y\{\frac{[\sum_i(Y_i^2-1)]^2\{\sum_i(Y_i^2-1)\}}{k_\ell^2 \sum_i Y_i^2}\}$$

$$= -E^Y\{\frac{[\sum_i(Y_i^2-1)^2][\sum_i(Y_i^2-1)]}{k_\ell^2 \sum_i Y_i^2} + \frac{[\sum_{i\neq j}(Y_i^2-1)(Y_j^2-1)][\sum_i(Y_i^2-1)]}{k_\ell^2 \sum_i Y_i^2}\} .$$

In order to show that $\varepsilon = o(\frac{1}{k_\ell})$, it is sufficient to show the following:

$$\text{(i)} \quad E^Y\left|\frac{(Y_i^2-1)^3}{\sum_i Y_i^2}\right| = o(1) \tag{3.2.3}$$

$$\text{(ii)} \quad E^Y[\frac{k_\ell(Y_i^2-1)^2(Y_j^2-1)}{\sum_i Y_i^2}] = o(1) \tag{3.2.4}$$

and

$$\text{(iii)} \quad E^Y[\frac{k_\ell(Y_i^2-1)^2(Y_j^2-1)(Y_k^2-1)}{\sum_i Y_i^2}] = o(1) . \tag{3.2.5}$$

To show the above, first notice that the conditional distribution of Y_i given $\theta_i(T_{\ell-1} + 1 \leq i \leq T_\ell)$ is

$$Y_i|\theta_i \sim N(\frac{\theta_i}{\sqrt{\sigma_\ell^2+\rho_\ell^2}}, \frac{\sigma_\ell^2}{\sigma_\ell^2+\rho_\ell^2}) .$$

If $g(\theta_i)$ is the prior density of θ_i then the marginal density of Y_i is

$$f(y_i) = \int_{-\infty}^{\infty} f(y_i|\theta_i)g(\theta_i)d\theta_i$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\frac{\sigma_\ell^2}{\sigma_\ell^2+\rho_\ell^2}}} e^{-\frac{\sigma_\ell^2+\rho_\ell^2}{2\sigma_\ell^2}\left(y_i - \frac{\theta_i}{\sqrt{\sigma_\ell^2+\rho_\ell^2}}\right)^2} g(\theta_i)d\theta_i$$

$$\leq \sqrt{\frac{1+T}{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{\sigma_\ell^2+\rho_\ell^2}{2\sigma_\ell^2}\left(y_i - \frac{\theta_i}{\sqrt{\sigma_\ell^2+\rho_\ell^2}}\right)^2} g(\theta_i)d\theta_i$$

$$\left(\text{since } \frac{\rho_\ell^2}{\sigma_\ell^2} < T\right)$$

$$\leq \sqrt{\frac{1+T}{2\pi}} \ . \tag{3.2.6}$$

Now to show (3.2.3) we have

$$E^Y \left| \frac{(Y_i^2-1)^3}{\sum_i Y_i^2} \right| = \int_{-\infty}^{\infty} \int_{\sum_{j\neq i} Y_j^2 > a^2} \frac{|y_i^2-1|^3}{\sum_i y_i^2} \prod_i f(y_i)dy_i$$

$$+ \int_{-\infty}^{\infty} \int_{\sum_{j\neq i} y_j^2 \leq a^2} \frac{|y_i^2-1|^3}{\sum_i y_i^2} \prod_i f(y_i)dy_i$$

$$= I_1 + I_2 \quad \text{(say)},$$

where a will be appropriately chosen. Now to show that $I_1 \to 0$, we have

$$\frac{Y_i^2-1}{\sum_i Y_i^2} = \frac{Y_i^2-1}{k_\ell} \frac{k_\ell}{\sum_i Y_i^2} \ .$$

Since, Y_i is fixed, $\frac{Y_i^2-1}{k_\ell} \to 0$ a.s. as $k_\ell \to \infty$, and by the Kolmogorov strong law of large numbers, $\frac{1}{k_\ell}\sum_i Y_i^2 \to 1$ a.s. Hence $\frac{k_\ell}{\sum_i Y_i^2} \to 1$ a.s. as $k_\ell \to \infty$. Also note that on the set $\sum_{j\neq i} Y_j^2 > a^2$,

$$\left| \frac{(Y_i^2-1)^3}{\sum_i Y_i^2} \right| = \left| \frac{Y_i^2-1}{\sum Y_i^2} \right| (Y_i^2-1)^2 \leq (Y_i^2-1)^2(1+\frac{1}{a^2})$$

and by Lemma 3.2.1

$$E^Y(Y_i^2-1)^2 < \infty .$$

Therefore, using the Dominated Convergence theorem $I_1 \to 0$. Now to show $I_2 \to 0$, we have

$$I_2 \leq \int_{-\infty}^{\infty} |y_i^2-1|^3 f(y_i)dy_i \int_{\sum_{j\neq i} y_j^2 \leq a^2} \frac{1}{\sum_{j\neq i} y_j^2} \prod_{j\neq i} f(y_j)dy_j$$

(by independence of Y_i's)

clearly, $\int_{-\infty}^{\infty} |y_i^2-1|^3 f(y_i)dy_i < \infty$ (by assumption).

Now using (3.2.6)

$$\int_{\sum_{j\neq i} y_i^2 \leq a^2} \frac{1}{\sum_{j\neq 2} y_j^2} \prod_{j\neq i} f(y_j)dy_j \leq (\frac{1+T}{2\pi})^{\frac{k_\ell-1}{2}} \int_{\sum_{j\neq i} y_j^2 \leq a^2} \frac{1}{\sum_{j\neq i} y_j^2} \prod_{j\neq i} dy_j$$

$$= (\frac{1+T}{2\pi})^{\frac{k_\ell-1}{2}} \frac{(a^2)^{\frac{k_\ell-3}{2}} \pi^{\frac{k_\ell-1}{2}}}{\Gamma(\frac{k_\ell+1}{2})}$$

(by Lemma 4 of the appendix).

Now choosing $a^2 = \frac{k_\ell - 3}{2M}$ where $\frac{(1+T)e}{2} < M < \infty$ and using the Stirling's approximation for $\Gamma(\frac{k_\ell+1}{2})$, it can be shown easily that $I_2 \to 0$. Thus (3.2.3) is proved.

To show (3.2.4), we will use the identity

$$\frac{1}{\sum_i y_i^2} = \frac{1}{\sum_{i\neq j} y_i^2} - \frac{y_j^2}{(\sum_{i\neq j} y_i^2)(\sum_i y_i^2)} . \tag{3.2.7}$$

Using (3.2.7) we have,

$$\begin{aligned} E^Y\Big[\frac{k_\ell (Y_i^2-1)^2(Y_j^2-1)}{\sum_i Y_i^2}\Big] &= \int_{\mathbb{R}^k} \frac{k_\ell (y_i^2-1)^2(y_j^2-1)}{\sum_{i\neq j} y_i^2} \prod_i f(y_i)dy_i \\ &\quad - \int_{\mathbb{R}^k} \frac{k_\ell (y_i^2-1)^2(y_j^2-1)y_j^2}{(\sum_{i\neq j} y_i^2)(\sum_i y_i^2)} \prod_i f(y_i)dy_i . \end{aligned} \tag{3.2.8}$$

Integrating over y_j, it is clear that the first integral in (3.2.8) is zero.

Consider the second integral in (3.2.8). First notice that,

$$\frac{1}{(\sum_{\ell\neq j} y_\ell^2)(\sum_\ell y_\ell^2)^2} \leq \frac{1}{(\sum_{\ell\neq j,i} y_\ell^2)^2} . \tag{3.2.9}$$

Now we can write,

$$\int_{\mathbb{R}^k} \frac{k_\ell (y_i^2-1)^2(y_j^2-1)y_j^2}{(\sum_{i\neq j} y_i^2)(\sum_i y_i^2)} \prod_i f(y_i)dy_i = \int_{-\infty}^{\infty} \int_{\sum_{\ell\neq j} y_\ell^2 > a^2} \frac{k_\ell (y_i^2-1)^2(y_j^2-1)y_j^2}{(\sum_{i\neq j} y_i^2)(\sum_i y_i^2)} \prod_i f(y_i)dy_i$$

$$+ \int_{-\infty}^{\infty} \int_{\sum_{\ell \neq j} y_\ell^2 \leq a^2} \frac{k_\ell (y_i^2-1)^2 (y_j^2-1) y_j^2}{(\sum_{i\neq j} y_i^2)(\sum_i y_i^2)} \prod_i f(y_i) dy_i$$

$$= I_3 + I_4 \quad \text{(say)},$$

where a will be appropriately chosen. Now to show $I_3 \to 0$, we have for fixed Y_i and Y_j,

$$\frac{k_\ell (Y_i^2-1)^2 (Y_j^2-1) Y_j^2}{(\sum_{i\neq j} Y_i^2)(\ _i\ Y_i^2)} \to 0 \quad \text{a.s.} \quad \text{as } k_\ell \to \infty,$$

$$\frac{(Y_i^2-1)^2}{\sum_i Y_i^2} \leq |Y_i^2-1|(1 + \frac{1}{a^2}) ,$$

and

$$\frac{k_\ell}{\sum_{i\neq j} Y_i^2} < \frac{k_\ell}{a^2} = \frac{2M}{1 - \frac{3}{k_\ell}} \quad (\text{choosing } a^2 = \frac{k_\ell - 3}{2M}).$$

Therefore by the Dominated convergence theorem, $I_3 \to 0$.

Now using (3.2.9) we have,

$$I_4 \leq \int_{-\infty}^{\infty} \int_{\sum_{\ell \neq j} y_\ell^2 \leq a^2} \frac{k_\ell (y_i^2-1)^2 (y_j^2-1) y_j^2}{(\sum_{\ell \neq i,j} y_\ell^2)^2} \prod_i f(y_i) dy_i$$

$$\leq \int_{-\infty}^{\infty} (y_i^2-1)^2 f(y_i) dy_i \int_{-\infty}^{\infty} y_j^2 (y_j^2-1) dy_j \int_{\sum_{\ell \neq i,j} y_\ell^2 \leq a^2} \frac{k_\ell}{(\sum_{\ell \neq i,j} y_\ell^2)^2} \prod_{\ell \neq i,j} f(y_\ell) dy_\ell$$

$$= E(Y_i^2-1)^2 E(Y_j^2-1) Y_j^2 \int_{\sum_{\ell \neq i,j} y_\ell^2 \leq a^2} \frac{k_\ell}{(\sum_{\ell \neq i,j} y_\ell^2)^2} \prod_{\ell \neq i,j} f(y_\ell) dy_\ell$$

Now using Lemma 4 of the appendix, we have

$$\int\limits_{\sum_{\ell\neq i,j} y_\ell^2 \leq a^2} \frac{k_\ell}{(\sum_{\ell\neq i,j} y_\ell^2)^2} \prod_{\ell\neq i,j} f(y_\ell)dy_\ell \leq \left(\frac{1+T}{2\pi}\right)^{\frac{k_\ell-2}{2}} \frac{k_\ell a^{k_\ell-6} (\sqrt{\pi})^{\frac{k_\ell-2}{2}}}{\Gamma(\frac{k_\ell}{2})} . \tag{3.2.10}$$

Again choosing $a^2 = \frac{k_\ell - 3}{2M}$ where $\frac{(1+T)}{2} < M < \infty$ and using the Stirling's approximation for $\Gamma(\frac{k_\ell}{2})$ for large k_ℓ, it is easy to show that (3.2.10) goes to zero. Thus $I_4 \to 0$ which completes the proof of (3.2.4).

Now (3.2.5) can be established as was (3.2.4). This completes the proof of the Lemma 3.2.2. ||

Lemma 3.2.3. $E^X\left[\frac{k}{||X||^2}\right] = 1 + \frac{2}{k} + \frac{1}{k^2}\sum_{\ell=1}^{s} \frac{(V_\ell - 2\rho_\ell^4)k}{(\sigma_\ell^2+\rho_\ell^2)^2} + o(\frac{1}{k})$.

Proof. Similar to the proof of Lemma 3.2.2. ||

Theorem 3.2.1. Suppose $X|\theta \sim N_k(\theta,\Sigma)$ where Σ and θ satisfy the condition given in the beginning of subsection 3.2.1. Then Δ, as defined in (2.1.11), is given as

$$\Delta = \sum_{\ell=1}^{s} \frac{\sigma_\ell^4}{\sigma_\ell^2+\rho_\ell^2} \{2(\tau_\ell - 1) + \frac{V_\ell - 2\rho_\ell^4}{(\sigma_\ell^2+\rho_\ell^2)^2} - \tau_\ell \sum_{j=1}^{s} \frac{(V_j - 2\rho_j^4)\tau_j}{(\sigma_j^2+\rho_j^2)^2}\} + o(1), \tag{3.2.11}$$

where $k_\ell = \tau_\ell k$, $\ell = 1,2,\ldots,s$.

Proof. From (2.1.3) we have

$$\gamma(A,\delta) = \text{tr}\,\Sigma + E^X[- \frac{2(k-2)}{||X||^2}\{\text{tr}\,\Sigma^2(\Sigma+A)^{-1} - \frac{2X^t(\Sigma+A)^{-1}\Sigma^2(\Sigma+A)^{-1}X}{||X||^2}\}$$

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

where $A = \text{diag}(\underbrace{\rho_1^2,\ldots,\rho_1^2}_{k_1},\ldots,\underbrace{\rho_s^2,\ldots,\rho_s^2}_{k_s})$.

Similarly from (2.1.9) we have

$$\gamma(A_{\rho_\ell},\delta_{(\rho_\ell)}) = k_\ell\sigma_\ell^2 - \frac{(k_\ell-2)^2\sigma_\ell^4}{\sigma_\ell^2+\rho_\ell^2} E[\frac{1}{||X_{(\rho_\ell)}||^2}], \quad \ell=1,2,\ldots,s.$$

Thus using Lemma 3.2.2, we have

$$\gamma(A_\ell,\delta_{(\ell)}) = k_\ell\sigma_\ell^2 - \frac{(k_\ell-2)^2}{k_\ell} \frac{\sigma_\ell^4}{\sigma_\ell^2+\rho_\ell^2}\{1 + \frac{2}{k_\ell} + \frac{V_\ell-2\rho_\ell^4}{k_\ell(\sigma_\ell^2+\rho_\ell^2)^2} + O(\frac{1}{k_\ell})\} ,$$

$$\ell=1,2,\ldots,s , \tag{3.2.12}$$

Similarly by Lemma 3.2.3, we have

$$\gamma(A,\delta) = \sum_{\ell=1}^{s} k_\ell\sigma_\ell^2 - \frac{(k-2)^2}{k^2}\{\sum_{\ell=1}^{s}\frac{\sigma_\ell^4}{\sigma_\ell^2+\rho_\ell^2}k_\ell\}\{1+\frac{2}{k}+\frac{1}{k^2}\sum_{\ell=1}^{s}\frac{k_\ell(V_\ell-2\rho_\ell^4)}{(\sigma_\ell^2+\rho_\ell^2)^2}+o(\frac{1}{k})\}. \tag{3.2.13}$$

Now from (3.2.12) and (3.2.13), the difference of Bayes risk (2.1.11) reduces to

$$\Delta = \sum_{\ell=1}^{s}\frac{(k_\ell-2)^2}{k_\ell^2}\frac{\sigma_\ell^4}{\sigma_\ell^2+\rho_\ell^2}k_\ell\{1 + \frac{2}{k_\ell} + \frac{V_\ell-2\rho_\ell^4}{k_\ell(\sigma_\ell^2+\rho_\ell^2)^2} + o(\frac{1}{k_\ell})\}$$

$$- \frac{(k-2)^2}{k^2}\{\sum_{\ell=1}^{s}\frac{\sigma_\ell^4}{\sigma_\ell^2+\rho_\ell^2}k_\ell\}\{1 + \frac{2}{k} + \frac{1}{k^2}\sum_{\ell=1}^{s}\frac{k_\ell(V_\ell-2\rho_\ell^4)}{(\sigma_\ell^2+\rho_\ell^2)^2} + o(\frac{1}{k})\}$$

$$= -4\sum_{\ell=1}^{s}\frac{\sigma_\ell^4}{\sigma_\ell^2+\rho_\ell^2} + 4\sum_{\ell=1}^{s}\frac{\sigma_\ell^4\tau_\ell}{\sigma_\ell^2+\rho_\ell^2} + \sum_{\ell=1}^{s}\frac{\sigma_\ell^4}{\sigma_\ell^2+\rho_\ell^2}\{2 + \frac{V_\ell-2\rho_\ell^4}{(\sigma_\ell^2+\rho_\ell^2)^2}\} - \sum_{\ell=1}^{s}\frac{2\sigma_\ell^4\tau_\ell}{\sigma_\ell^2+\rho_\ell^2}$$

$$- \{ \sum_{\ell=1}^{s} \frac{\sigma_\ell^4 \tau_\ell}{(\sigma_\ell^2+\rho_\ell^2)} \} \{ \sum_{\ell=1}^{s} \frac{(V_\ell - 2\rho_\ell^4)\tau_\ell}{(\sigma_\ell^2+\rho_\ell^2)^2} \} + o(1)$$

$$= \sum_{\ell=1}^{s} \frac{\sigma_\ell^4}{\sigma_\ell^2+\rho_\ell^2} \{2(\tau_\ell - 1) + \frac{V_\ell - 2\rho_\ell^4}{(\sigma_\ell^2+\rho_\ell^2)^2} - \tau_\ell \sum_{j=1}^{s} \frac{(V_j - 2\rho_j^4)\tau_j}{(\sigma_j^2+\rho_j^2)^2} \} + o(1) \tag{3.2.14}$$

which completes the proof of the theorem. ||

Comment 1. If $\sigma_\ell = \sigma$, $\rho_\ell = \rho$ and $V_\ell = V$ $\forall \ell = 1,2,\ldots,s$, then as k goes to infinity (3.2.14) reduces to

$$\Delta = \frac{\sigma^4}{\sigma^2+\rho^2} \sum_{\ell=1}^{s} (1-\tau_j) \{ \frac{V-2\rho^4}{(\sigma^2+\rho^2)^2} - 2\} + o(1)$$

$$= \frac{\sigma^4}{\sigma^2+\rho^2} \{ \frac{V-2\rho^4}{(\sigma^2+\rho^2)^2} - 2\} + o(1). \tag{3.2.15}$$

Subsection 3.2.2. Some Examples

Example 3.2.1. Assume the θ_i's are independent and identically distributed with a $N_1(0,\tau^2)$ distribution. Then marginally $X_i \overset{\text{indep.}}{\sim} N_1(0,\sigma^2+\tau^2)$ $i=1,2,\ldots,k$. In this case ρ^2 and V, as defined in Theorem 3.2.1, become

$$\rho^2 = E(\theta_i^2) = \tau^2$$

and

$$V = E^X(\theta_i^2 - \tau^2)^2 = E(\theta_i^4) + \tau^4 - 2\tau^2 E(\theta_i^2) = 2\tau^4 .$$

Therefore by (3.2.18), we have for large k

$$\Delta = - \frac{2\sigma^4(s-1)}{\sigma^2+\rho^2} < 0,$$

which shows that asymptotic separation is worse.

Example 3.2.2. Suppose the prior distribution of θ is defined as $\pi(0) = p$ and $\rho(\pm y) = \frac{1}{2}(1-p)$ where $0 < p < 1$. Then ρ_ℓ^2 and V_ℓ^2, as defined in Theorem 3.2.1, reduce to

$$\rho_\ell^2 = y^2(1-p)$$

and

$$V_\ell = y^4 p(1-p).$$

Therefore,

$$\frac{V_\ell - 2\rho_\ell^4}{(\sigma_\ell^2+\rho_\ell^2)^2} \to \frac{p}{1-p} - 2 \quad \text{as} \quad y \to \pm\infty .$$

Thus from (3.2.15)

$$\Delta \to \sum_{\ell=1}^{s} \frac{\sigma_\ell^4}{\sigma_\ell^2+\rho_\ell^2}\{(1-\tau_\ell)(\frac{p}{1-p} -4)\} > 0$$

if $p > \frac{4}{5}$. Thus for $p > \frac{4}{5}$, asymptotic separation is always better.

Example 3.2.3. Assume the prior distribution has a Student's t distribution with α degrees of freedom, i.e.,

$$\pi(\theta) = \frac{1}{aB(\frac{1}{2},\frac{\alpha}{2})}\left(1 + \frac{\theta^2}{a^2}\right)^{-\frac{\alpha+1}{2}}, \qquad -\infty \le \theta \le \infty .$$

Then clearly, $E(\theta) = 0$, $E(\theta^2) = \frac{a^2}{\alpha-2}$ and $E(\theta^4) = \frac{3a^4}{(\alpha-2)(\alpha-4)}$ where $\alpha > 4$. Thus ρ^2 and V, as defined in Theorem 3.2.1, reduce to

$$\rho^2 = E(\theta^2) = \frac{a^2}{\alpha-2}$$

and

$$V = E(\theta^4)-\{E(\theta^2)\}^2 = V-2\rho^4-2\sigma^4-4\sigma^2\rho^2$$

$$= \frac{1}{a^2}\{\frac{4a^6(4-\frac{\alpha+1}{2})}{(\alpha-2)^2(\alpha-4)} - \frac{2\sigma^4}{a^2} - \frac{4\sigma^2}{\alpha-2}\} .$$

Since σ is fixed, for $a \to \infty$ we have from (3.2.18) that $\Delta \geq 0$ provided $4 < \alpha < 7$. Hence for a Student's t prior with degrees of freedom between 4 and 7, the asymptotic separation is better.

<u>Comment 2</u>. If $\mu_2 = E(X_i^2)$ and $\mu_4 = E(X_i^4)$ then clearly we have $\mu_2 = \sigma_\ell^2$ and $\mu_4 = 3\mu_2^2 + (V_\ell-2\rho_\ell^4)$. Therefore under the assumption of $\sigma_\ell = \sigma$, $\rho_\ell = \rho$ and $V_\ell = V$, (3.2.15) reduces, in terms of the marginal moments of the X_i's, to

$$\Delta = \frac{\sigma^4(s-1)}{\mu_2} \frac{\mu_4-3\mu_2^2}{\mu_2^2} -2\} + o(1)$$

$$= \frac{\alpha^4(s-1)}{\mu_2} \frac{\mu_4-5\mu_2^2}{\mu_2^2} + o(1) . \qquad (3.2.16)$$

CHAPTER IV

INCLUSION OF EXTREME OBSERVATIONS

Section 4.1. Introduction

In this chapter we will consider the question of inclusion of extreme observations when a flat tailed prior is suspected. We will modify the combined and the separate estimators developed in the previous chapters using Stein's truncation method (Stein 1974), which essentially limits the amount by which any coordinate of the separate or combined estimate can differ from the corresponding coordinate of X.

Assume that $Y|\theta \sim N(\theta,\Sigma)$ and θ is such that $E(\theta) = 0$ and $V(\theta) = A$. Also assume that $\Sigma = \sigma^2 I$ and $A = \rho^2 I$. Then

$$\delta(Y) = \left(1 - \frac{(k-2)\sigma^2}{\sum_{i=1}^{k} Y_i^2}\right)Y$$

is a usual shrinkage estimator for θ. Suppose $X_i = \frac{Y_i}{\sqrt{\sigma^2+\rho^2}}$ and $\eta_i = \frac{\theta_i}{\sqrt{\sigma^2+\rho^2}}$ for all $i=1,2,\ldots,k$. Then

$$\delta(X) = \left(1 - \frac{(k-2)\sigma^2}{(\sigma^2+\rho^2)\sum_{i=1}^{k} X_i^2}\right)X \qquad (4.1.1)$$

is a usual shrinkage estimator for η. The truncated estimator for η will be defined coordinatewise by

$$\hat{\eta}_i^{(\ell)}(X) = X_i - \frac{(\ell-2)\sigma^2}{(\sigma^2+\rho^2)\sum_{j=1}^{k}(X_j^2 \wedge Z_{(\ell)}^2)}(\operatorname{sgn} X_i)(|X_i| \wedge Z_{(\ell)}), \tag{4.1.2}$$

where ℓ is a large fraction of k,

$$Z_i = |X_i|, \quad i=1,2,\ldots,k \tag{4.1.3}$$

and

$$Z_{(1)} < Z_{(2)} \cdots < Z_{(k)}$$

are the order statistics of $Z_1, Z_2, \ldots, Z_k$. (4.1.4)

Under the above set up, the Bayes risk of $\hat{\theta}^{(\ell)} = (\Sigma+A)^{\frac{1}{2}}\hat{\eta}^{(\ell)}$ under squared error loss can be shown, as in (2.1.3), to be

$$r(A,\hat{\theta}^{(\ell)}) = k - (\ell-2)^2 \frac{\sigma^4}{\sigma^2+\rho^2} E^X\left[\frac{1}{\sum_{j=1}^{k}(X_j^2 \wedge Z_{(\ell)}^2)}\right]. \tag{4.1.5}$$

Therefore, the improvement in the risk of $\hat{\theta}^{(\ell)}$ compared to the risk of X is

$$\Delta_k^{(\ell)} = \frac{\sigma^4}{\sigma^2+\rho^2} E^X\left[\frac{(\ell-2)^2}{\sum_{j=1}^{k}(X_j^2 \wedge Z_{(\ell)}^2)}\right], \tag{4.1.6}$$

where E^X stands for expectation under the marginal distribution of X. Also note that the improvement in the risk of the estimator (4.1.1) is given by

$$\Delta_k = \frac{\sigma^4}{\sigma^2+\rho^2} E^X\left[\frac{(k-2)^2}{\sum_{j=1}^{k} X_j^2}\right]. \tag{4.1.7}$$

It happens that frequently the θ_i will occur according to a flat prior. If the prior is flat some of the X_i's will be extremely large. Therefore from (4.1.7) the improvement, Δ_k, will be small. It is thus natural to truncate the large X_i's.

We will assume that the θ_i's are independent and have a common flat distribution, so that, marginally, the X_i's have independent and identical flat distributions. It would be aesthetically more pleasing to work with a given class of flat priors, but the marginals would then be messy and hard to work with. Luckily t-marginals are similar to marginals obtained for many standard flat priors. Therefore in Section 4.3 the marginal distribution of the X_i's will be chosen to be the t-distribution. We will then find the optimum truncation points (y) for different degrees of freedom of the t-distribution.

Section 4.2. Evaluation and Asymptotic Properties of $\Delta_k^{(\ell)}$

Suppose $X_1, X_2, \ldots, X_k$ are marginally independent and identically distributed random variables and $Z_1, Z_2, \ldots, Z_k$ are defined as in (4.1.3). Then $Z_1, Z_2, \ldots, Z_k$ are independent and identically distributed random variables with marginal distribution function, say $G(z)$. We will derive the asymptotic properties of $\Delta_k^{(\ell)}$ as defined in (4.1.6).

Lemma 4.2.1. Suppose $\ell = [yk]$, where $0 < y \leq 1$ and $[x]$ denotes the nearest integer to x. Also let $\alpha(y)$ denote the $y^{\underline{th}}$ fractile of G that is

$$G(\alpha(y)) = y. \tag{4.2.1}$$

Then as $k \to \infty$,

$$Z_{(\ell)} \to \alpha(y) \text{ almost surely.} \tag{4.2.2}$$

Proof. See Rao (1973). ||

<u>Theorem 4.2.1</u>. Suppose G(z) admits a probability density g(z) which is bounded then

$$\lim_{k\to\infty} \frac{1}{k} E^X\left[\frac{(\ell-2)^2}{\sum_{j=1}^{k} (X_j^2 \wedge Z_{(\ell)}^2)} \right] = \frac{y^2}{\mu_2} \tag{4.2.4}$$

where $\mu_2 = E(X_j^2 \wedge \alpha^2(y)) < \infty$ and $\ell = [yk]$.

<u>Proof</u>. Define,

$$U_{j,k} = X_j^2 \wedge Z_{(\ell)}^2 \quad \text{and} \quad V_j = X_j^2 \wedge \alpha^2(y) .$$

We will first show that

$$\frac{k}{\sum_{j=1}^{k} U_{j,k}} - \frac{k}{\sum_{j=1}^{k} V_j} = o(1) . \tag{4.2.5}$$

Using the inequality

$$|X_j^2 \wedge Z_{(\ell)}^2 - X_j^2 \wedge \alpha^2(y)| \le |Z_{(\ell)}^2 - \alpha^2(y)| ,$$

we have,

$$\left|\frac{k}{\sum_{j=1}^{k} U_{j,k}} - \frac{k}{\sum_{j=1}^{k} V_j}\right| = \left|\frac{\frac{1}{k}\sum_{j=1}^{k}(V_j - U_{j,k})}{(\frac{1}{k}\sum_{j=1}^{k} U_{j,k})(\frac{1}{k}\sum_{j=1}^{k} V_j)}\right| \le \frac{\frac{1}{k}\sum_{j=1}^{k}|Z_{(\ell)}^2 - \alpha^2(y)|}{(\frac{1}{k}\sum_{j=1}^{k} U_{j,k})(\frac{1}{k}\sum_{j=1}^{k} V_j)} . \tag{4.2.6}$$

Using (4.2.2), we have

$$\frac{1}{k}\sum_{j=1}^{k}|Z_{(\ell)}^2 - \alpha^2(y)| \to 0 \quad \text{a.s.} \quad \text{as } k \to \infty , \tag{4.2.7}$$

and by the strong law of large numbers,

$$\frac{1}{k}\sum_{j=1}^{k} V_j \to \frac{1}{\mu_2} \quad \text{a.s.} \quad \text{as } k \to \infty . \tag{4.2.8}$$

Combining (4.2.7) and (4.2.8) we have

$$\frac{1}{k}\sum_{j=1}^{k} X_j^2 \wedge Z_{(\ell)}^2 \to \frac{1}{\mu_2} \quad \text{a.s.} \quad \text{as } k \to \infty . \tag{4.2.9}$$

Thus from (4.2.7), (4.2.8) and (4.2.9) we have (4.2.5).

From the assumption of the theorem we have $g(z) \le B$, where B is a constant. Thus $G(z) \le Bz$. Now, we can choose $\lambda > 0$, such that $\lambda B < 1$ and $\lambda < \alpha^2(y)$. Then using the fact $\sum_{j=1}^{k} X_j^2 \wedge Z_{(\ell)}^2 \ge (k-\ell) Z_{(\ell)}^2$ we have, on the set $\{Z_{(\ell)}^2 > \lambda\}$,

$$\frac{k}{\sum_{j=1}^{k} U_{j,k}} \le \frac{k}{k-\ell} \frac{1}{Z_{(\ell)}^2} \le \frac{k}{(k-\ell)\lambda} \le \frac{c}{\lambda}$$

where c is a constant, and

$$\frac{k}{\sum_{j=1}^{k} V_j} \le \frac{k}{\sum_{j=1}^{k} V_j \wedge Z_{(\ell)}^2} \le \frac{k}{(k-\ell)} \frac{1}{Z_{(\ell)}^2 \wedge \alpha^2(y)} \le \frac{k}{(k-\ell)} \frac{1}{\lambda \wedge \alpha^2(y)} \le \frac{c}{\lambda \wedge \alpha^2(y)}$$

Using (4.2.7), (4.2.8) and the Dominated convergence theorem, the two integrals converge on $\{Z_{(\ell)}^2 > \lambda\}$ as $k \to \infty$. Now on the set $\{Z_{(\ell)}^2 < \lambda\}$, we have

$$\int_{Z_{(\ell)}^2 < \lambda} \frac{k}{\sum_{j=1}^{k} U_{j,k}} \prod_{j=1}^{k} g(z_j) dz_j \le \int_{Z_{(\ell)}^2 < \lambda} \frac{c}{Z_{(\ell)}^2} \prod_{j=1}^{k} g(z_j) dz_j$$

and

$$\int_{Z_{(\ell)}^2 < \lambda} \frac{k}{\sum_{j=1}^{k} V_j} \prod_{j=1}^{k} g(z_j) dz_j \le \int_{Z_{(\ell)}^2 < \lambda} \frac{c}{Z_{(\ell)}^2} \prod_{j=1}^{k} g(z_j) dz_j \quad (\text{since } \lambda < \alpha^2(y)).$$

Now,

$$\int_{Z_{(\ell)}^2 < \lambda} \frac{1}{Z_{(\ell)}^2} \prod_{j=1}^{k} g(z_j) dz_j = \frac{k!}{(\ell-1)!(k-\ell)!} \int_0^{\lambda} z^{-2} [G(z)]^{\ell-1} [1-G(z)]^{k-\ell} g(z) dz$$

$$\le \frac{k!}{(\ell-1)!(k-\ell)!} \frac{1}{\lambda^2} \frac{(\lambda B)^{\ell}}{\ell-2} \; (\text{since } G(z) \le B). \quad (4.2.10)$$

Using the Stirling's approximation, (4.2.10) goes to zero, since $\lambda B < 1$. Therefore the expectation of the left hand side of (4.2.5) goes to zero as $k \to \infty$. Therefore,

$$\lim_{k\to\infty} \frac{1}{k} E^X\left[\frac{(\ell-2)^2}{\sum_{j=1}^{k} X_j^2 \wedge Z_{(\ell)}^2}\right] = \lim_{k\to\infty} \left(\frac{\ell-2}{k}\right)^2 E\left[\frac{k}{\sum_{j=1}^{k} X_j^2 \wedge Z_{(\ell)}^2}\right] = \frac{y^2}{\mu^2}, \quad (4.2.11)$$

which completes the proof of the theorem. ||

Now notice that

$$\mu_2 = E(X_j^2 \wedge \alpha^2(y)) = 2[\int_0^{\alpha(y)} x^2 f(x)dx + \alpha^2(y) \int_{\alpha(y)}^{\infty} f(x)dx] . \tag{4.2.12}$$

where f(x) is the common density of X_i's. Using the definition of $\alpha(y)$, (5.2.5) can be written as

$$\mu_2 = 2 \int_0^{\alpha(y)} x^2 f(x)dx + \alpha^2(y)(1-y). \tag{4.2.13}$$

Define,

$$r(y) = \frac{y^2}{2 \int_0^{\alpha(y)} x^2 f(x)dx + \alpha^2(y)(1-y)} . \tag{4.2.14}$$

Thus r(y) is the asymptotic improvement of the truncated James-Stein estimator. The following lemma will give the behavior of r(y) near 0 and 1.

Lemma 4.2.2. If r(y) is defined as in (4.2.14) then

(i) $\lim_{y \to o} r(y) = 4[f(o)]^2$

and

(ii)
$$\lim_{y \to 1} r(y) = \begin{cases} \frac{1}{V(X)} & \text{if} \quad V(X) < \infty \\ 0 & \text{if} \quad V(X) = \infty. \end{cases} \tag{4.2.15}$$

Proof. We have

$$G(z) = P\{-z < X_i < -z\} = 2 \int_0^z f(x)dx. \tag{4.2.16}$$

Define,

$$G^{-1}(u) = \inf\{z: G(z) \geq u\} . \tag{4.2.17}$$

Then clearly the $y^{\underline{th}}$ fractile of G is given by

$$\alpha(y) = G^{-1}(y) \ . \tag{4.2.18}$$

Hence,

$$\lim_{y\to o} r(y) = \lim_{\alpha(y)\to o} \frac{[G(\alpha(y))]^2}{2\int_0^{\alpha(y)} x^2 f(x)dx + \alpha^2(y)[1-G(\alpha(y))]}$$

$$= \lim_{\alpha(y)\to o} \frac{4G(\alpha(y))f(\alpha(y))}{2\alpha(y)[1-G(\alpha(y))]} \text{ (by L'Hospital's rule),}$$

$$= \lim_{\alpha(y)\to o} \frac{8[f(\alpha(y))]^2 + 4G(\alpha(y))f'(\alpha(y))}{2[1-G(\alpha(y))] - 4\alpha(y)f(\alpha(y))}$$

(by L'Hospital's rule),

$$= 4[f(0)]^2 \ ,$$

which completes the proof of part (i) of the lemma.

To prove part (ii), notice that as $y \to 1$, $\alpha(y) \to \infty$. Now clearly,

$$V(X) = 2\int_0^{\infty} x^2 f(x)dx > 2\int_0^{\alpha(y)} x^2 f(x)dx + [\alpha(y)]^2[1-G(\alpha(y))]$$

$$= 2\int_0^{\alpha(y)} x^2 f(x)dx + [\alpha(y)]^2(1-y).$$

Therefore if $V(X) < \infty$, then

$$\lim_{\alpha(y)\to\infty} [\alpha(y)]^2(1-y) = 0 \ . \tag{4.2.19}$$

Using (4.2.19) in (4.2.14), part (ii) of the lemma follows. ||

Now using Lemma 4.2.2, r(y) is between 0 and 1 and seems to look like a bell shaped curve with a maximum tending to occur somehwere between 0 and 1.

Section 4.3. t-marginal

We will now consider the case where $X_1, X_2, \ldots, X_k$ are independent and identically distributed, with X_i having (marginally) a t-distribution with α-degrees of freedom; that is, X_i has density function

$$f(x) = c_{\alpha,\sigma}\left(1 + \frac{x^2}{\alpha\sigma^2}\right)^{-(\alpha+1)/2}, \quad \alpha \geq 1, \tag{4.3.1}$$

where

$$c_{\alpha,\sigma} = \frac{\Gamma((\alpha+1)/2)}{\sqrt{\alpha\pi}\,\Gamma(\alpha/2)\sigma}.$$

Observe that

$$E(x_i) = 0 \quad \text{and } v(x_i) = \frac{\alpha}{\alpha-2}\sigma^2, \quad (\alpha \geq 2),\ i = 1,2,\ldots,k. \tag{4.3.2}$$

Now the distribution function $G(z)$ of Z as defined in (4.1.2) is given as

$$G(z) = P\{-z < x_i < z\} = 2c_{\alpha,\sigma}\int_0^z \left(1 + \frac{x^2}{\alpha\sigma^2}\right)^{-(\alpha+1)/2} dx$$

$$= 2c_{\alpha,\sigma}\sqrt{\alpha}\,\sigma \int_0^{\tan^{-1} z/\sqrt{\alpha}\,\sigma} \cos^{\alpha-1}\theta\, d\theta. \quad (\text{substituting } x = \sqrt{\alpha}\,\sigma \tan\theta) \tag{4.3.3}$$

Also to calculate $\gamma(y)$ given in (4.2.11) we have

$$\int_0^{\alpha(y)} x^2 f(x)dx = c_{\alpha,\sigma}\int_0^{\alpha(y)} x^2\left(1 + \frac{x^2}{\sigma^2\alpha}\right)^{-(\alpha+1)/2} dx$$

$$= c_{\alpha,\sigma}\sigma^3\alpha^{3/2}\int_0^{\tan^{-1}\alpha(y)/\sigma\sqrt{\alpha}} \tan^2\theta \cos^{\alpha-1}\theta\, d\theta. \tag{4.3.4}$$

Subsection 4.3.1. Cauchy Marginal Problem

Suppose that marginally the X_i's have independent Cauchy distributions, i.e., t-distributions with 1 degree of freedom. We then have,

$$f(x) = \frac{1}{\pi\sigma(1+x^2/\sigma^2)}\ , \text{ mean and variance undefined.} \tag{4.3.5}$$

Substituting $\alpha = 1$ in (4.3.3) and (4.3.4) we have

$$G(z) = \frac{2}{\pi}\tan^{-1} z/\sigma\ , \tag{4.3.6}$$

and

$$\int_0^{\alpha(y)} x^2 f(x)dx = \frac{1}{\pi\sigma}\int_0^{\alpha(y)} x^2\left(1+\frac{x^2}{\sigma^2}\right)^{-1} dx. \tag{4.3.7}$$

Using (4.2.1) , we have

$$y = G(\alpha(y)) = \frac{2}{\pi}\tan^{-1}\alpha(y)/\sigma\ .$$

Therefore,

$$z(y) = \sigma \tan \frac{\pi y}{2}\ . \tag{4.3.8}$$

Now to evaluate $\gamma(y)$ as defined in (4.2.14)we have

$$2\int_0^{\alpha(y)} x^2 f(x)dx = \frac{2}{\pi\sigma}\int_0^{\alpha(y)/\sigma}\frac{\sigma^2x^2}{1+x^2}\sigma dx = \frac{2}{\pi}\sigma^2\int_0^{\alpha(y)/\sigma}\frac{1+x^2-1}{1+x^2}dx$$

$$= \frac{2}{\pi}\sigma^2\left\{\frac{\alpha(y)}{\sigma} - \tan^{-1}\frac{\alpha(y)}{\sigma}\right\} = \sigma^2\left\{\frac{2}{\pi}\tan\frac{\pi y}{2} - y\right\}. \tag{4.3.9}$$

Substituting (4.3.8) and (4.3.9) in (4.2.14) it follows that

$$\gamma(y) = \frac{y^2}{\sigma^2\left\{\frac{2}{\pi}\tan\frac{\pi y}{2} - y + (1-y)\tan^2\frac{\pi y}{2}\right\}}\ . \tag{4.3.10}$$

Using Lemma 4.2.2, we get

$$\lim_{y\to 0}\gamma(y) = \frac{4}{\sigma^2\pi^2} \quad\text{and}\quad \lim_{y\to 1}\gamma(y) = 0. \tag{4.3.11}$$

Now to get the optimum truncation point in this case we have to find the maximum of $\gamma(y)$ and the maximizing point will be the optimum truncation point.

It is found numerically that the funcion $\gamma(y)$ is maximized at $y \simeq 0.24$. In Table 4, values of $\sigma^2\gamma(y)$ are calculated for different values of y.

Subsection 4.3.2. t_2-marginal Problem

Suppose that marginally the X_i's have t-distributions with 2-degrees of freedom, i.e., the density is

$$f(x) = \frac{1}{2\sqrt{2}\,\sigma}\left(1 + \frac{x^2}{2\sigma^2}\right)^{-3/2} . \tag{4.3.12}$$

Then

$$E(X) = 0 \quad \text{and} \quad v(X) = \infty. \tag{4.3.13}$$

Substituting $\alpha = 2$ in (4.3.3), we get

$$G(z) = 2\sqrt{2}\,\frac{\sigma}{2\sqrt{2}\,\sigma}\int_0^{\tan^{-1} z/\sqrt{2}\sigma} \cos\theta\, d\theta = \sin\tan^{-1}\frac{z}{\sigma\sqrt{2}} = \frac{z}{\sqrt{z^2+2\sigma^2}} \tag{4.3.14}$$

and from (4.3.4)

$$\int_0^{\alpha(y)} x^2 f(x)dx = \sigma^2 \int_0^{\tan^{-1}\alpha(y)/\sigma\sqrt{2}} \tan^2\theta\cos\theta\, d\theta$$

$$= \sigma^2 \int_0^{\tan^{-1}\alpha(y)/\sigma\sqrt{2}} (\sec\theta - \cos\theta)d\theta, \tag{4.3.15}$$

where $\alpha(y) = G^{-1}(y)$. Therefore

$$y = G(\alpha(y)) = \frac{\alpha(y)}{\sqrt{\alpha^2(y)+2\sigma^2}}, \tag{4.3.16}$$

and hence

$$\alpha(y) = \frac{y\sigma\sqrt{2}}{\sqrt{1-y^2}} . \tag{4.3.17}$$

It follows from (4.3.15) that

$$\int_0^{\alpha(y)} x^2 f(x)\, dx = \sigma^2[\ell n\,(\frac{\sqrt{\alpha^2(y)+2\sigma^2}}{\sigma\sqrt{2}} + \frac{\alpha(y)}{\sigma\sqrt{2}} - \frac{\alpha(y)}{\sqrt{\alpha^2(y)+2\sigma^2}}],$$

$$\text{(since, } \alpha^2(y)+2\sigma^2 = \frac{2\sigma^2}{1-y^2})$$

$$= \sigma^2[\ell n\,(\frac{1}{\sqrt{1-y^2}} + \frac{y}{\sqrt{1-y^2}}) - y] . \tag{4.3.18}$$

Substituting (4.3.17) and (4.3.18) in (4.2.14), it follows that

$$\gamma(y) = \frac{y^2}{2\sigma^2[\ell n\,(\frac{1}{\sqrt{1-y^2}} + \frac{y}{\sqrt{1-y^2}})-y] + (1-y)\sigma^2\,\frac{2y^2}{1-y^2}}$$

$$= \frac{y^2}{2\sigma^2[\ell n\,(\frac{1+y}{\sqrt{1-y^2}}) - \frac{y}{1+y}]} . \tag{4.3.19}$$

Using Lemma 4.2.2 , we have

$$\lim_{y\to 0} \gamma(y) = \frac{1}{2\sigma^2} \tag{4.3.20}$$

and

$$\lim_{y\to 1} \gamma(y) = 0 .$$

It was found numerically that $\gamma(y)$ is maximized at $y \simeq 0.44$. Thus $y \simeq 0.44$ is the optimum truncation point. In Table 4, the values of $\gamma(y)$ are shown for different values of y.

Subsection 4.3.3. t_3-marginal Problem

Suppose that marginally the X_i's have t-distributions with 3 degrees of freedom , i.e.,

$$f(x) = \frac{2}{\pi\sigma\sqrt{3}} \left(1 + \frac{x^2}{3\sigma^2}\right)^{-2} , \quad \sigma > 0. \tag{4.3.21}$$

Then,

$$E(X) = 0 \quad \text{and} \quad v(X) = 3\sigma^2. \tag{4.3.22}$$

Substituting $\alpha = 3$, in (4.3.3) we have

$$G(z) = \frac{2\sqrt{3}\sigma}{\pi\sigma\sqrt{3}} \int_0^{\tan^{-1} z/\sigma\sqrt{3}} \cos^2\theta \, d\theta = \frac{2}{\pi} \left[\tan^{-1} \frac{z}{\sigma\sqrt{3}} + \frac{\sigma z\sqrt{3}}{z^2+3\sigma^2}\right] \tag{4.3.23}$$

and from (4.3.4)

$$\int_0^{\alpha(y)} x^2 f(x)dx = \frac{6}{\pi} \int_0^{\tan^{-1}\alpha(y)/\sigma\sqrt{3}} \sin^2\theta \, d\theta , \tag{4.3.24}$$

where $\alpha(y) = G^{-1}(y)$. Therefore,

$$y = G(\alpha(y)) = \frac{2}{\pi} \left[\tan^{-1} \frac{\alpha(y)}{\sigma\sqrt{3}} + \frac{\sigma\alpha(y)\sqrt{3}}{\alpha^2(y)+3\sigma^2}\right]$$

$$= \frac{2}{\pi} \left[\tan^{-1}\left(\frac{\alpha(y)}{\sigma}\right) \frac{1}{\sqrt{3}} + \frac{\frac{\alpha(y)}{\sigma}\sqrt{3}}{\left(\frac{\alpha(y)}{\sigma}\right)^2+3}\right]$$

$$\equiv H\left(\frac{\alpha(y)}{\sigma}\right) \qquad \text{(definition).} \tag{4.3.25}$$

Then,

$$\alpha(y) = \sigma H^{-1}(y) . \tag{4.3.26}$$

It follows from (4.3.24) that

$$\int_0^{\alpha(y)} x^2 f(x)dx = \frac{3\sigma^2}{\pi} \left[\tan^{-1} \frac{\alpha(y)}{\sigma\sqrt{3}} - \frac{\alpha(y)\sigma\sqrt{3}}{\alpha^2(y)+3\sigma^2}\right] . \tag{4.3.27}$$

Clearly from (4.3.26), $\frac{\alpha(y)}{\sigma}$ is independent of σ. Now define,

$$b(y) = \frac{\alpha(y)}{\sigma\sqrt{3}} . \qquad (4.3.28)$$

Then from (4.3.27),

$$\int_0^{\alpha(y)} x^2 f(x)dx = \frac{3\sigma^2}{\pi} [\tan^{-1}b(y) - \frac{b(y)}{1+b^2(y)}], \qquad (4.3.29)$$

where b(y) will be obtained from

$$y = \frac{2}{\pi} [\tan^{-1}b(y) + \frac{b(y)}{1+b^2(y)}] . \qquad (4.3.30)$$

Now using (4.3.28) and (4.3.29) in (4.2.14) we have

$$\gamma(y) = \frac{y^2}{\frac{6\sigma^2}{\pi} [\tan^{-1}b(y) - \frac{b(y)}{1+b^2(y)}] + 3\sigma^2 b^2(y)[1-y]} . \qquad (4.3.31)$$

Using Lemma 4.2.2. , we have

$$\sigma^2 \lim_{y\to 0} \gamma(y) = 4(\frac{2}{\sqrt{3}\,\pi})^2 = \frac{16}{3\pi^2} = 0.54, \qquad (4.3.32)$$

and

$$\sigma^2 \lim_{y\to 1} \gamma(y) = \frac{1}{V(X)} = \frac{1}{3} = 0.33.$$

It was found numerically that, $\gamma(y)$ is maximized at $y \simeq 0.55$. Thus $y \simeq 0.55$ is the optimum truncation point. Table 4 shows the values of $\gamma(y)$ for different values of y.

Subsection 4.3.4. t_4-marginal Problem

Suppose that marginally the X_i's have t-distributions with 4 degrees of freedom, i.e.,

$$f(x) = \frac{2}{\pi\sigma\sqrt{3}} (1 + \frac{x^2}{3\sigma^2})^{-5/2} , \quad \sigma > 0. \qquad (4.3.33)$$

Then,

$$E(X) = 0 \quad \text{and} \quad v(X) = 2\sigma^2. \tag{4.3.34}$$

Substituting $\alpha = 4$ in (4.3.3) we have

$$G(z) = \frac{3\sigma}{2}\int_0^{\tan^{-1}z/2\sigma} \cos^3\theta\, d\theta = \frac{3}{2}\int_0^{\tan^{-1}z/2\sigma} (\tfrac{1}{4}\cos 3\theta + \tfrac{3}{4}\cos\theta)d\theta$$

$$= [\tfrac{3}{2}\sin\theta - \tfrac{1}{2}\sin^3\theta]_0^{\tan^{-1}z/2\sigma} = \frac{z(z^2+6\sigma^2)}{(z^2+4\sigma^2)^{3/2}}, \tag{4.3.35}$$

and from (4.3.4)

$$\int_0^{\alpha(y)} x^2 f(x)dx = 3\sigma^2 \int_0^{\tan^{-1}\alpha(y)/2\sigma} (\cos\theta - \cos^3\theta)d\theta, \tag{4.3.36}$$

where $\alpha(y) = G^{-1}(y)$. Therefore,

$$y = G(\alpha(y)) = \frac{\alpha(y)\{\alpha^2(y)+6\sigma^2\}}{\{\alpha^2(y)+4\sigma^2\}^{3/2}},$$

$$= \frac{\frac{\alpha(y)}{\sigma}\{(\frac{\alpha(y)}{\sigma})^2+6\}}{\{(\frac{\alpha(y)}{\sigma})^2+4\}^{3/2}}$$

$$\equiv H(\frac{\alpha(y)}{\sigma}) \quad \text{(definition)}. \tag{4.3.37}$$

Thus,

$$\alpha(y) = \sigma H^{-1}(y). \tag{4.3.38}$$

It follows from (4.3.36) that

$$\int_0^{\alpha(y)} x^2 f(x)dx = \sigma^2[3\sin\tan^{-1}\frac{\alpha(y)}{2\sigma} - 3\int_0^{\tan^{-1}\alpha(y)/2\sigma} \cos^3\theta\, d\theta]$$

$$= \sigma^2 \sin^3 \tan^{-1}\frac{\alpha(y)}{2\sigma} = \sigma^2\{\frac{\alpha(y)}{\sqrt{\alpha^2(y)+4\sigma^2}}\}^3. \tag{4.3.39}$$

Clearly from (4.3.38), $\frac{\alpha(y)}{\sigma}$ is independent of σ.

Now define,

$$b(y) = \frac{\alpha(y)}{\sigma} . \tag{4.3.40}$$

Then from (4.3.39)

$$\int_0^{\alpha(y)} x^2 f(x)dx = \sigma^2 \{\frac{b(y)}{\sqrt{b^2(y)+4}}\}^3 , \tag{4.3.41}$$

where b(y) will be obtained from

$$y = \frac{b(y)\{b^2(y)+6\}}{\{b^2(y)+4\}^{3/2}} . \tag{4.3.42}$$

Now using (4.3.40) and (4.3.41) in (4.2.14), we have

$$\gamma(y) = \frac{y^2}{2\sigma^2\{\frac{b(y)}{\sqrt{b^2(y)+4}}\}^3 + \sigma^2 b^2(y)[1-y]} . \tag{4.3.43}$$

Using Lemma 4.2.2. , we have

$$\lim_{y\to 0} \gamma(y) = 4[f(0)]^2 = \frac{9}{16} = 0.56 \tag{4.3.44}$$

and

$$\lim_{y\to 1} \gamma(y) = \frac{1}{V(X)} = 0.5.$$

It was found numerically that $\gamma(y)$ is maximized at $y \simeq 0.65$. Thus $y \simeq 0.65$ is the optimum truncation point. In Table 4, values of $\gamma(y)$ are calculated for different values of y.

Section 4.4. Conclusion

In Table 4, the values of $\gamma(y)$ are plotted against the values of y when the marginal distributions are Cauchy, Student's t with degrees of freedom 2, 3, 4 and normal. From Table 4, it appears that y = 0.6 is a reasonable compromise value for truncation.

Comment 1. For the normal case, Stein (1974) computed the asymptotic relative efficiency of the estimated improvement in the risk of the estimate $\hat{\theta}^{(\ell)}$ with respect to the estimated improvementin the risk of the James-Stein estimate (nonzero since the prior variance is finite). We are considering the quantity $\gamma(y)$, since in our case the variance may not exist always. The two measures are clearly proportional when the variance does exist.

Table 4

Values of $r(y)$ for Cauchy, t_2, t_3, t_4 and normal marginals

y	Cauchy	t_2	t_3	t_4	Normal
0	.40	.50	.54	.56	.63
.1	.43	.53	.57	.59	.70
.2	.44	.55	.61	.63	.74
.3	.43	.57	.63	.65	.76
.4	.42	.58	.64	.68	.79
.5	.39	.57	.65	.69	.83
.6	.35	.56	.66	.70	.87
.7	.29	.53	.65	.71	.91
.8	.21	.49	.62	.69	.94
.9	.11	.40	.56	.67	.97
1	0	0	.33	.50	1

CHAPTER V

SEPARATION UNDER MISSPECIFICATION OF NORMAL PRIORS

Section 5.1. Introduction

In this chapter we will consider the separation problem under misspecification of normal priors. We will consider a misspecification model in which under certain conditions separation is better. Suppose a priori it is known that some of the θ_i's are quite different than the rest, in that there is more uncertainty in their prior variances. In that case θ can be partitioned as $\theta = (\theta_{(1)}, \theta_{(2)})$, where $\theta_{(i)}$ is $k_i \times 1$, $i = 1,2$. Suppose the random vector X is partitioned as $X = (X_{(1)}, X_{(2)})$ where $X_{(i)}$ is $k_i \times 1$, $i = 1,2$. Suppose $X_{(1)}$ and $X_{(2)}$ are independent, so that the covariance matrix reduces to

$$\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} .$$

Suppose the prior θ is felt to have a k-variate distribution with mean zero and known positive definite covariance matrix A of the form

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} .$$

Suppose for the two groups, we know the prior information within each group is correct in terms of the mean, and is proportionally correct in terms of the variances, i.e., the true covariance matrix for the

$i^{\underline{th}}$ group might be $\rho_i A_i$, $i = 1,2$. The ρ_i could differ, corresponding to a condition in which there is a differing amount of confidence in the prior information for each group.

An appealing way to proceed would be to proceed as in Berger (1980), but assume that the covariance matrix of θ given ρ_1 and ρ_2 is

$$B(\rho_1,\rho_2) = \begin{pmatrix} \rho_1(\Sigma_1+A_1) & 0 \\ 0 & \rho_2(\Sigma_2+A_2) \end{pmatrix} - \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix},$$

where ρ_1 and ρ_2 have some (possibly joint) prior distribution. By introducing the ρ_i, one achieves, robustness, admitting there is uncertainty about A. The introduction of two different ρ's allows one to deal with a situation in which one is more uncertain about A_1 than about A_2. The generalized prior density of θ, in this case would be

$$g(\theta) = \int_0^1 \int_0^\infty |B(\rho_1,\rho_2)|^{-\frac{1}{2}} \exp\{-\theta^t B(\rho_1,\rho_2)^{-1}\theta/2\} d\mu(\rho_1) d\nu(\rho_2),$$

where μ and ν are probability measures on ρ_1 and ρ_2 respectively. Unfortunately, in this case the calculations become too hard to work with. Also one could not specify measures μ and ν easily.

We thus will restrict ourselves to consideration of the usual combined and separate estimators, and evaluate their Bayes risks under a simpler misspecification model. In particular, we will assume for simplicity that the prior is normal, and furthermore pretend that the misspecification is in the marginal of X. If A were the prior covariance matrix, then (marginally)

$$X \sim N(0, \begin{pmatrix} \Sigma_1+A_1 & 0 \\ 0 & \Sigma_2+A_2 \end{pmatrix}) .$$

Usually A_i is much larger than Σ_i, so misspecifying A_i by a factor ρ_i will approximately correspond to misspecifying $\Sigma_i + A_i$ by a factor ρ_i. Thus our assumption in the calculation of Bayes risks will be that

$$X \sim N\left(0, \begin{pmatrix} \rho_1(\Sigma_1+A_1) & 0 \\ 0 & \rho_2(\Sigma_2+A_2) \end{pmatrix}\right) .$$

Furthermore, it turns out that the desirability of separation depends only on the ratio ρ_2/ρ_1. Hence it suffices to consider

$$X \sim N\left(0, \begin{pmatrix} (\Sigma_1+A_1) & 0 \\ 0 & \rho(\Sigma_2+A_2) \end{pmatrix}\right) . \qquad (5.1.1)$$

We will consider (5.1.1) as our misspecification model. It is also assumed that the two groups are homogeneous, in the sense that

$$\Lambda = \Sigma^2(\Sigma+A)^{-1} = \text{diag}(\underbrace{\lambda_1,\ldots,\lambda_1}_{k_1} ; \underbrace{\lambda_2,\ldots,\lambda_2}_{k_2}). \qquad (5.1.2)$$

In section 5.2, the Bayes risks of the combined and the separate estimators are obtained under the misspecification model. In section 5.3, theoretical results are obtained for finite group sizes, which determine a region where the separation is better. In section 5.4, theoretical results on separation are obtained asymptotically as the group sizes go to infinity. It is proved first that for $\rho \neq 1$, separation is asymptotically better. Then it is shown that if $\varphi = \lim_{k\to\infty} \frac{k_1}{k_2}$ exists, the boundary for separation is of the form

$$\rho = 1 + \frac{\sqrt{2(1+\varphi)/\varphi}}{\sqrt{k_2}} ,$$

in the sense that if

$$\lim_{k\to\infty} \sqrt{k_2}\,(\rho-1) = \gamma \ ,$$

separation is better (or worse) as $\gamma^2 \geq \frac{2(1+\varphi)}{\varphi}$ (or $\gamma^2 \leq \frac{2(1+\varphi)}{\varphi}$). In section 5.5, the exact boundary is obtained numerically for finite group sizes. It is observed that as the group sizes increase, the asymptotic theoretical bound obtained in section 5.4 is approached.

Comment. The quantity ρ introduced in (5.1.1) can roughly be measured as follows. Suppose

$$\theta \sim N_k(0, \begin{pmatrix} \tau_1 A_1 & 0 \\ 0 & \tau_2 A_2 \end{pmatrix}),$$

then if I_1 and I_2 are confidence intervals for τ_1 and τ_2 respectively then ρ can roughly be taken as the ratio of the lengths of I_1 and I_2.

Section 5.2. The Bayes Risk Calculation

In this section we will evaluate the Bayes risks of the separate and the combined estimator under the misspecification model.

Let us partition X as $X = (X^t_{(1)}, X^t_{(2)})$, where $X_{(i)}$ is x_ix1. Then from (5.1.1) we have,

$$\begin{aligned} X^t_{(1)}(\Sigma_1+A_1)^{-1}X_{(1)} &\sim \chi^2_{k_1} \ , \\ X^t_{(2)}(\Sigma_2+A_2)^{-1}X_{(2)} &\sim \rho\chi^2_{k_2} \ , \end{aligned} \qquad (5.2.1)$$

and the two quadratic forms are independent. Thus

$$||X||^2 = X^t(\Sigma+A)^{-1}X \sim \chi^2_{k_1} + \rho\chi^2_{k_2} \ . \qquad (5.2.2)$$

Using (5.1.1) and (5.2.2), the Bayes risk (2.1.3) reduces to

$$\gamma(A,\delta) = \text{tr}\ \Sigma - 2(k-2)(k_1\lambda_1+k_2\lambda_2)E[\frac{1}{\chi^2_{k_1}+\rho\chi^2_2}]$$

$$+ (k^2-4)E[\frac{\lambda_1\chi^2_{k_1}+\rho\lambda_2\chi^2_{k_2}}{(\chi^2_{k_1}+\rho\chi^2_{k_2})}] , \qquad (5.2.3)$$

where E stands for the expectation over the chi square random variables.

Now using (5.1.1) and (5.1.2), the Bayes risk of the separate estimator reduces to

$$\gamma(A_1,\delta_1) = \text{tr}\ \Sigma_1 - (k_1-2)\lambda_1$$

and

$$\gamma(A_2,\delta_2) = \text{tr}\ \Sigma_2 - (k_2-2)\lambda_2/\rho\ . \qquad (5.2.4)$$

Thus the difference of the Bayes risks of the combined and the separate estimator reduces to

$$\Delta = \{(k_1-2)\lambda_1+(k_2-2)\ \frac{\lambda_2}{\rho}\} - 2(k-2)E[\frac{k_1\lambda_1+k_2\lambda_2}{\chi^2_{k_1}+\rho\chi^2_{k_2}}]$$

$$+ (k^2-4)E[\frac{\lambda_1\chi^2_{k_1}+\rho\lambda_2\chi^2_{k_2}}{(\chi^2_{k_1}+\rho\chi^2_{k_2})^2}] \ . \qquad (5.2.5)$$

Section 5.3. Separation for Finite Group Sizes

In this section we will assume that the group sizes k_1 and k_2 are finite and λ_1 and λ_2 as defined in (5.1.2) are equal. We will get regions in terms of the group sizes, where in one region separation is better and in the other region separation is worse. Unfortunately the regions which will be obtained are not complementary, so the exact boundary is not determined.

Theorem 5.3.1. If $\lambda_1 = \lambda_2$ then

$$\Delta > 0 \quad \text{if} \quad \rho \notin \left(\frac{(k_1-2)(k_2-2)}{k_2(2k_1+k_2-4)} , \frac{k_1(k_1+2k_2-4)}{(k_1-2)(k_2-2)} \right)$$

and

$$\Delta < 0 \quad \text{if} \quad \rho \in \left(\frac{k_2-2}{k_2} , \frac{k_1}{k_1-2} \right). \tag{5.3.1}$$

Proof. Using $\lambda_1 = \lambda_2$ in (5.2.5) we have

$$\frac{\Delta}{\lambda_2} = (k_1-2) + \frac{(k_2-2)}{\rho} - (k-2)^2 E\left[\frac{1}{\chi^2_{k_1} + \rho \chi^2_{k_2}}\right]. \tag{5.3.2}$$

Now using the fact that $\chi^2_{k_1} + \rho \chi^2_{k_2} > \chi^2_{k_1}$, we have

$$E\left[\frac{1}{\chi^2_{k_1} + \rho \chi^2_{k_2}}\right] < E\left[\frac{1}{\chi^2_{k_1}}\right] = \frac{1}{k_1-2} .$$

Therefore

$$\frac{\Delta}{\lambda_2} > (k_1-2) + \frac{k_2-2}{\rho} - \frac{(k-2)^2}{k_1-2} .$$

Hence

$$\rho < \frac{(k_2-2)(k_1-2)}{k_2(2x_1+k_2-4)} \quad \text{implies} \quad \Delta > 0. \tag{5.3.3}$$

Using $\chi^2_{k_1} + \rho \chi^2_{k_2} > \rho \chi^2_{k_2}$, we have

$$E\left[\frac{1}{\chi^2_{k_1} + \rho \chi^2_{k_2}}\right] < \frac{1}{\rho (k_2-2)} .$$

Therefore

$$\frac{\Delta}{\lambda_2} > (k_1-2) + \frac{(k_2-2)}{\rho} - \frac{(k-2)^2}{\rho(k_2-2)} .$$

Hence,

$$\rho > \frac{(2k_2+k_1-4)k_1}{(k_1-2)(k_2-2)} \quad \text{implies} \quad \Delta > 0. \tag{5.3.4}$$

Now clearly $\rho > 1$ implies $\chi^2_{k_1} + \rho\chi^2_{k_2} < \rho\chi^2_{k_1} + \rho\chi^2_{k_2}$

and

$$E[\frac{1}{\chi^2_{k_1}+\rho\chi^2_{k_2}}] > E[\frac{1}{\rho\chi^2_k}] = \frac{1}{\rho(k-2)} .$$

Therefore,

$$\frac{\Delta}{\lambda_2} < (k_1-2) + \frac{(k_2-2)}{\rho} - \frac{k-2}{\rho} .$$

Hence,

$$\rho < \frac{k_1}{k_1-2} \quad \text{implies} \quad \Delta < 0. \tag{5.3.5}$$

For $\rho < 1$, $\chi^2_{k_1} + \rho\chi^2_{k_2} < \chi^2_{k_1} + \chi^2_{k_2}$ which implies that

$$E[\frac{1}{\chi^2_{k_1}+\rho\,\chi^2_{k_2}}] > \frac{1}{k-2} .$$

Therefore,

$$\frac{\Delta}{\lambda_2} < (k_1-2) + \frac{(k_2-2)}{\rho} - (k-2) .$$

Hence,

$$\rho > \frac{k_2-2}{k_2} \quad \text{implies} \quad \Delta > 0 . \tag{5.3.6}$$

Combining (5.3.3) through (5.3.6), the proof is complete. ||

Section 5.4. Asymptotic Results for Separation

In this section, we will compute the Bayes risk (5.2.3) asymptotically when k_1 and k_2 go to infinity. It is clear that $E(\frac{\chi^2_{k_i}}{k_i}) = 1$ and $v(\frac{\chi^2_{k_i}}{k_i}) = \frac{2}{k_i} \to 0$ as $k_i \to \infty$, $i = 1,2$. Thus

$$\frac{\chi^2_{k_i}}{k_i} \to 1 \quad \text{in probability,} \quad i = 1,2. \tag{5.4.1}$$

Using (5.4.1), it is clear that

$$\frac{k_1+\rho k_2}{\chi^2_{k_1}+\rho\, \chi^2_{k_2}} \to 1 \quad \text{in probability.} \tag{5.4.2}$$

We will now show that the above convergence is also in mean.

Lemma 5.4.1.

$$E\left|\frac{k_1+\rho\, k_2}{\chi^2_{k_1}+\rho\chi^2_{k_2}}\right| - 1 \quad = O(k^{-\frac{1}{2}}) \ . \tag{5.4.3}$$

Proof. Using the Cauchy Schwarz inequality, it is clear that

$$E\left|\frac{k_1+\rho\, k_2 - (\chi^2_{k_1}+\rho\chi^2_{k_2})}{\chi^2_{k_1}+\chi^2_{k_2}}\right| \le [E(\chi^2_{k_1}+\rho\chi^2_{k_2}-k_1-\rho k_2)]^{\frac{1}{2}} \quad [E(\frac{1}{\chi^2_{k_1}+\rho\chi^2_{k_2}})^2]^{\frac{1}{2}}.$$

Now,

$$[E(\chi^2_{k_1}+\rho\chi^2_{k_2}-k_1-\rho k_2)^2]^{\frac{1}{2}} = [v(\chi^2_{k_1}+\rho\chi^2_{k_2})]^{\frac{1}{2}} = (2k_1+2\rho^2_{k_2})^{\frac{1}{2}} \le c_1 k^{\frac{1}{2}} ,$$

where c_1 is a constant, and

$$[E(\frac{1}{\chi^2_{k_1}+\rho\chi^2_{k_2}})^2]^{\frac{1}{2}} \le c_2[E(\frac{1}{\chi^2_k})^2]^{\frac{1}{2}} = c_2[\frac{\Gamma(\frac{k}{2}-2)}{\Gamma(\frac{k}{2})}]^{\frac{1}{2}} < \frac{c_2}{k} ,$$

where c_2 is a constant. Thus

$$E\left|\frac{k_1+\rho k_2}{\chi^2_{k_1}+\rho\chi^2_{k_2}}\right| - 1 \quad < c_1 c_2 k^{-\frac{1}{2}} ,$$

which completes the proof of the lemma. ||

Now again using (5.4.1) and 5.4.2) it is clear that

$$\frac{\lambda_1\chi^2_{k_1}+\rho\lambda_2\chi^2_{k_2}}{\lambda_1 k_1+\rho\lambda_2 k_2}\left(\frac{k_1+\rho k_2}{\chi^2_{k_1}+\rho\chi^2_{k_2}}\right)^2 \to 1 \quad \text{in probability.} \tag{5.4.4}$$

The following lemma will show that the above convergence is also in mean.

Lemma 5.4.2.

$$E\left[\frac{\lambda_1\chi^2_{k_1}+\rho\lambda_2\chi^2_{k_2}}{\lambda_1 k_1+\rho\lambda_2 k_2}\left(\frac{k_1+\rho k_2}{\chi^2_{k_1}+\rho\chi^2_{k_2}}\right)^2\right] = 1 + 0(k^{-\frac{1}{2}}). \tag{5.4.5}$$

Proof. Let us define for simplicity

$$A = \frac{\lambda_1\chi^2_{k_1}+\rho\lambda_2\chi^2_{k_2}}{\lambda_1 k_1+\rho\lambda_2 k_2} .$$

Hence

$$E\left[\frac{\lambda_1\chi^2_{k_1}+\rho\lambda_2\chi^2_{k_2}}{\lambda_1 k_1+\rho\lambda_2 k_2}\left(\frac{k_1+\rho k_2}{\chi^2_{k_1}+\rho\chi^2_{k_2}}\right)^2\right] = E\left[A + A\left\{\left(\frac{k_1+\rho k_2}{\chi^2_{k_1}+\rho\chi^2_{k_2}}\right)^2 - 1\right\}\right]$$

$$= 1 + E\left[A\,\frac{\{k_1+\rho k_2-\chi^2_{k_1}-\rho\chi^2_{k_2}\}\{k_1+\rho k_2+\chi^2_{k_1}+\rho\chi^2_{k_2}\}}{(\chi^2_{k_1}+\rho\chi^2_{k_2})^2}\right].$$

Let us define, $B = k_1+\rho k_2-\chi^2_{k_1}-\rho\chi^2_{k_2}$, $c = k_1+\rho k_2+\chi^2_{k_1}+\rho\chi^2_{k_2}$ and

$D = (\chi^2_{k_1}+\rho\chi^2_{k_2})^2$. In order to prove the lemma, it is sufficient to show that $E(\frac{ABC}{D}) = O(k^{-\frac{1}{2}})$.

Now applying the Cauchy Scharz inequality we have,

$$|E(\frac{BAC}{D})| \le [E(B^2)E(\frac{AC}{D})^2]^{\frac{1}{2}} = \sqrt{v(\chi^2_{k_1}+\rho\chi^2_{k_2})}[E(\frac{AC}{D})^2]^{\frac{1}{2}} . \quad (5.4.6)$$

Now,

$$A \le \frac{\max\{\lambda_1,\lambda_2\}}{\min\{\lambda_1,\lambda_2\}} \frac{(\chi^2_{k_1}+\rho\chi^2_{k_2})}{k_1+\rho k_2} .$$

Hence,

$$\frac{AC}{D} \le \frac{\max\{\lambda_1,\lambda_2\}}{\min\{\lambda_1,\lambda_2\}} [\frac{1}{\chi^2_{k_1}+\rho\chi^2_{k_2}} + \frac{1}{k_1+\rho k_2}]. \quad (5.4.7)$$

Also,

$$\frac{1}{\chi^2_{k_1}+\rho\chi^2_{k_2}} \le \frac{1}{\min(1,\rho)(\chi^2_{k_1}+\chi^2_{k_2})} .$$

Therefore, from (5.4.7)

$$\frac{AC}{D} \le \frac{c_1}{\chi^2_{k_1}+\chi^2_{k_2}} + \frac{c_2}{k_1+\rho k_2},$$

where c_1 and c_2 are constants.

Now we have,

$$E(\frac{AC}{D})^2 \le c_1^2 E(\frac{1}{\chi^2_k})^2 + \frac{c_2^2}{(k_1+\rho k_2)^2} + \frac{2c_1c_2}{k_1+\rho k_2}\frac{1}{k-2}$$

$$= \frac{c_1^2}{(k-2)(k-4)} + \frac{c_2^2}{(k_1+\rho k_2)^2} + \frac{2c_1c_2}{(k_1+\rho k_2)(k-2)} < \frac{c_3}{k^2} \quad (5.4.8)$$

where c_3 is a constant. Combining (5.4.6) and (5.4.8) we have

$$E(\frac{ABC}{D}) = 0(k^{-\frac{1}{2}}),$$

which completes the proof of the lemma. ||

We will now establish the following theorem concerning separation. The theorem will establish for $\rho \neq 1$, that separation is asymptotically better.

<u>Theorem 5.4.1</u>. Define $\beta = \lambda_2/\lambda_1$ and $k_i = [n\alpha_i]$, $i = 1,2$ where α_1 and α_2 are constants such that $0 < \alpha_i < 1$, $i = 1,2$, $\alpha_1 + \alpha_2 = 1$ and $[n\alpha_i]$ is the nearest integer to $n\alpha_i$. Then for any fixed $\rho \neq 1$,

$$\lim_{n\to\infty} \Delta > 0. \tag{5.4.9}$$

<u>Proof</u>. Define $\varphi = \alpha_1/\alpha_2$. Using Lemma 5.4.1, we have

$$E\left[\frac{(k_1\lambda_1+k_2\lambda_2)}{\chi^2_{k_1}+\rho\chi^2_{k_2}}\right] = \frac{k_1\lambda_1+k_2\lambda_2}{k_1+\rho k_2} E\left[\frac{k_1+\rho k_2}{\chi^2_{k_1}+\rho\chi^2_{k_2}}\right]$$

$$= \frac{\lambda_1\alpha_1+\lambda_2\alpha_2}{\alpha_1+\rho\alpha_2} + 0(n^{-\frac{1}{2}}),$$

$$= \frac{(\varphi\beta+1)\lambda_2}{\varphi+\rho} + 0(n^{-\frac{1}{2}}). \tag{5.4.10}$$

Similarly using Lemma 5.4.2 we have

$$E\left[\frac{\lambda_1\chi^2_{k_1}+\rho\lambda_2\chi^2_{k_2}}{(\chi^2_{k_1}+\rho\chi^2_{k_2})^2}\right] = \frac{\lambda_1 k_1+\rho\lambda_2 k_2}{(k_1+\rho k_2)^2} E\left[\frac{(k_1+\rho k_2)^2}{(\lambda_1 k_1+\rho\lambda_2 k_2)} \frac{\lambda_1\chi^2_{k_1}+\rho\lambda_2\chi^2_{k_2}}{(\chi^2_{k_1}+\chi^2_{k_2})^2}\right]$$

$$= \frac{\lambda_1\alpha_1+\rho\lambda_2\alpha_2}{n(\alpha_1+\rho\alpha_2)} + 0(n^{-3/2}).$$

Therefore,

$$(k+2)E\left[\frac{\lambda_1\chi^2_{k_1}+\rho\lambda_2\chi^2_{k_2}}{(\chi^2_{k_1}+\rho\chi^2_{k_2})^2}\right] = \{n(\alpha_1+\alpha_2)+2\}\left\{\frac{(\beta\varphi+\rho)\lambda_2}{n(\varphi+\rho)^2\alpha_2} + O(n^{-3/2})\right\}$$

$$= \frac{(1+\varphi)(\beta\varphi+\rho)\lambda_2}{(\varphi+\rho)^2} + \frac{2(\beta\varphi+\rho)\lambda_2}{n(\varphi+\rho)^2\alpha_2} + O(n^{-\frac{1}{2}}). \tag{5.4.11}$$

Therefore (5.2.5) reduces to

$$\Delta = \{(n\alpha_1-2)\beta + \frac{n\alpha_2-2}{\rho}\}\lambda_2$$

$$+ \lambda_2\{n\alpha_2(1+\varphi)-2\}\{\frac{(1+\varphi)(\beta\varphi+\rho)}{(\varphi+\rho)^2} + \frac{2(\beta\varphi+\rho)}{n(\varphi+\rho)^2\alpha_2} + O(n^{-\frac{1}{2}}) - \frac{2(\beta\varphi+1)}{\varphi+\rho} - O(n^{-\frac{1}{2}})\}.$$

Hence,

$$\frac{\Delta}{\lambda_2} = (n\varphi\alpha_2-2)\beta + \frac{n\alpha_2-2}{\rho} - \frac{2(1+\varphi)(\beta\varphi+\rho)}{(\varphi+\rho)^2} - \frac{4(\beta\varphi+\rho)}{n(\varphi+\rho)^2\alpha_2} + \frac{4(\beta\varphi+1)}{\varphi+\rho}$$

$$+ \frac{n\alpha_2(1+\varphi)^2(\beta\varphi+\rho)}{(\varphi+\rho)^2} + \frac{2(1+\varphi)(\beta\varphi+\rho)}{(\varphi+\rho)^2} - \frac{2n\alpha_2(1+\varphi)(\beta\varphi+1)}{\varphi+\rho} + O(n^{\frac{1}{2}}).$$

Hence,

$$\frac{\Delta}{n\lambda_2} = \alpha_2\{\varphi\beta + \frac{1}{\rho} + \frac{(1+\varphi)^2(\beta\varphi+\rho)}{(\varphi+\rho)^2} - \frac{2(1+\varphi)(\beta\varphi+1)}{\varphi+\rho}\} + O(n^{-1})$$

$$= \alpha_2\,\frac{\varphi(\varphi+\beta\rho)(1-\rho)^2}{\rho(\varphi+\rho)^2}\} + O(n^{-1}).$$

Now letting n go to infinity, the proof is complete. ||

Therefore in the misspecification model, separation is asymptotically better for $\rho \neq 1$. Now naturally one should search for a boundary where the separation is better. It turns out that the boundary is of the form $\rho = 1 + \gamma/\sqrt{k_2}$. In the following theorem, the values of γ will determine an asymptotic region within which the combined estimator is better and outside of which the separate estimator is better.

Theorem 5.4.2. Let $\beta = \lambda_1/\lambda_2$ and $\rho = 1 + \gamma/\sqrt{k_2}$ and assume $\varphi = \lim_{k\to\infty} \frac{k_1}{k_2}$ exists. Then under the misspecification model

$$\lim_{k\to\infty} \Delta \gtrless 0 \quad \text{according as } \gamma^2 \gtrless \frac{2(1+\varphi)}{\varphi} . \qquad (5.4.12)$$

Proof. Let us define,

$$S = E[\frac{1}{\chi^2_{k_1}+\rho\chi^2_{k_2}}], \quad T_1 = E[\frac{\chi^2_{k_1}}{(\chi^2_{k_1}+\rho\chi^2_{k_2})^2}] \quad \text{and} \quad T_2 = E[\frac{\rho\chi^2_{k_2}}{(\chi^2_{k_1}+\rho\chi^2_{k_2})^2}]. \qquad (5.4.13)$$

Thus from (5.2.5),

$$\Delta = \{(k_1-2)\lambda_1+(k_2-2)\frac{\lambda_2}{\rho}\} + (k^2-4)(\lambda_1T_1+\lambda_2T_2) - 2(k-2)(k_1\lambda_1+k_2\lambda_2)S. \qquad (5.4.14)$$

Now we have

$$S = E[\frac{1}{(\chi^2_{k_1}+\chi^2_{k_2}+\frac{\gamma}{\sqrt{k_2}}\chi^2_{k_2})}] \qquad (\text{since } \rho = 1 + \frac{\gamma}{\sqrt{k_2}})$$

$$= E\{\frac{1}{\chi^2_{k_1}+\chi^2_{k_2}} - \frac{\gamma}{\sqrt{k_2}}\frac{\chi^2_{k_2}}{(\chi^2_{k_1}+\chi^2_{k_2})^2} + \frac{\gamma^2}{k_2}\frac{\chi^4_{k_2}}{(\chi^2_{k_1}+\chi^2_{k_2})^3}\} + o(k^{-5/2}).$$

Using the fact that $\chi^2_{k_1} + \chi^2_{k_2} \sim \chi^2_k$ and is independent of

$$\frac{\chi^2_{k_2}}{\chi^2_{k_1}+\chi^2_{k_2}} \sim Be(\frac{k_2}{2}, \frac{k_1}{2}),$$

we have

$$S = \frac{1}{k-2}\{1 - \frac{\gamma}{\sqrt{k_2}} \frac{1}{1+\varphi} + \frac{\gamma^2}{k_2} \frac{1}{1+\varphi} \frac{k_2+2}{(1+\varphi)k_2+2}\} + o(k^{-5/2}).$$

(5.4.15)

Now using the fact that $1 - \rho = - \frac{\gamma}{\sqrt{k_2}}$ we have

$$T_1 = E[\frac{\chi^2_{k_1}}{\{\rho(\chi^2_{k_1}+\chi^2_{k_2})+(1-\rho)\chi^2_{k_1}\}^2}]$$

$$= E[\frac{\chi^2_{k_1}}{\{\rho(\chi^2_{k_1}+\chi^2_{k_2}) - \frac{\gamma}{\sqrt{k_2}} \chi^2_{k_1}\}^2}]$$

$$= \frac{1}{\rho^2} E[\frac{\chi^2_{k_1}}{(\chi^2_{k_1}+\chi^2_{k_2})^2} + \frac{2\gamma}{\rho\sqrt{k_2}}\{\frac{1}{\chi^2_{k_1}+\chi^2_{k_2}} (\frac{\chi^2_{k_1}}{\chi^2_{k_1}+\chi^2_{k_2}})^2\}$$

$$+ \frac{3\gamma^2}{\rho^2 k_2}\{\frac{1}{\chi^2_{k_1}+\chi^2_{k_2}} (\frac{\chi^2_{k_1}}{\chi^2_{k_1}+\chi^2_{k_2}})^3\}] + o(k^{-5/2}) .$$

Using the fact that

$$E(\frac{\chi^2_{k_1}}{\chi^2_{k_1}+\chi^2_{k_2}})^s = \frac{B(\frac{k_1}{2}+s, \frac{k_2}{2})}{B(\frac{k_1}{2}, \frac{k_2}{2})} ,$$

we have

$$T_1 = \frac{1}{\rho^2(k-2)} \frac{\varphi}{1+\varphi} [1 + \frac{2\gamma}{\rho\sqrt{k_2}} \frac{\varphi k_2+2}{k+2} + \frac{3\gamma^2}{\rho^2 k_2} \frac{(\varphi k_2+2)(\varphi k_2+4)}{(k+2)(k+4)}] + o(k^{-5/2}).$$

(5.4.16)

Similarly,

$$T_2 = E[\frac{\rho\chi^2_{k_2}}{\{(\chi^2_{k_1}+\chi^2_{k_2})+\frac{\gamma}{\sqrt{k_2}}\chi^2_{k_2}\}^2}]$$

$$= \rho E[\frac{\chi^2_{k_2}}{(\chi^2_{k_1}+\chi^2_{k_2})^2} - \frac{2\gamma}{\sqrt{k_2}}\{\frac{1}{\chi^2_{k_1}+\chi^2_{k_2}}(\frac{\chi^2_{k_2}}{\chi^2_{k_1}+\chi^2_{k_2}})^2\} + \frac{3\gamma^2}{k_2}\{\frac{1}{\chi^2_{k_1}+\chi^2_{k_2}}(\frac{\chi^2_{k_2}}{\chi^2_{k_1}+\chi^2_{k_2}})^3\}]$$

$$+ o(k^{-5/2})$$

$$= \rho[\frac{k_2}{k(k-2)} - \frac{2\gamma}{\sqrt{k_2}}\frac{1}{k-2}\frac{k_2(k_2+2)}{k(k+2)} + \frac{3\gamma^2}{k_2}\frac{1}{k-2}\frac{k_2(k_2+2)(k_2+4)}{k(k+2)(k+4)}]$$

$$+ o(k^{-5/2})$$

$$= \frac{\rho}{k-2}\frac{1}{1+\varphi}[1 - \frac{2\gamma}{\sqrt{k_2}}\frac{k_2+2}{k+2} + \frac{3\gamma^2}{k_2}\frac{(k_2+2)(k_2+4)}{(k+2)(k+4)}]$$

$$+ o(k^{-5/2}). \tag{5.4.17}$$

Now we will consider (5.4.14) term by term. First note that

$$(k_1-2)\lambda_1+(k_2-2)\frac{\lambda_2}{\rho} = \beta\varphi k_2 - 2\beta+(k_2-2)\{1-\frac{\gamma}{\sqrt{k_2}}+\frac{\gamma^2}{k_2} - o(k_2^{-3/2})\}$$

$$= -2(1+\beta)+k_2(1+\beta\varphi)-\frac{\gamma}{\sqrt{k_2}}(k_2-2)+\frac{\gamma^2}{k_2}(k_2-2)+o(k^{-\frac{1}{2}}).$$

$$\tag{5.4.18}$$

Next note that

$$(k^2-4)\,\frac{k}{k_1}\,T_1 = (k+2) - \frac{2\gamma(k+2)}{\sqrt{k_2}} + \frac{3\gamma^2}{k_2}(k+2) - o(k^{-\frac{1}{2}})$$

$$+\frac{2\gamma}{\sqrt{k_2}}(\varphi k_2+2) - \frac{6\gamma^2}{k_2}(\varphi k_2+2) + o(k^{-\frac{1}{2}})$$

$$+\frac{3\gamma^2}{k_2}(\varphi k_2+2)(\varphi k_2+4)\{\frac{1}{k} - o(k^{-2})\}\{1-o(k^{-\frac{1}{2}})\} + o(k^{-\frac{1}{2}})$$

$$= (k+2) + \frac{2\gamma}{\sqrt{k_2}}(\varphi k_2+2-k-2) + \frac{3\gamma^2}{k_2}(k+2-2\varphi k_2-4)$$

$$+\frac{3\gamma^2}{k_2}(\varphi k_2+2)(\varphi k_2+4)\{\frac{1}{k} - o(k^{-2})\} + o(k^{-\frac{1}{2}})$$

$$= (k-2) - 2\gamma\sqrt{k_2} + 3\gamma^2(1-\varphi+\frac{\varphi^2}{1+\varphi}) + o(k^{-\frac{1}{2}}). \qquad (5.4.19)$$

Similarly,

$$(k^2-4)\,\frac{k}{k_2}\,T_2 = (k+2)(1+\frac{\gamma}{\sqrt{k_2}}) - \frac{2\gamma}{\sqrt{k_2}}(k_2+2)(1+\frac{\gamma}{\sqrt{k_2}})$$

$$+\frac{3\gamma^2}{k_2}(1+\frac{\gamma}{\sqrt{k_2}})(k_2+2)(k_2+4)\{\frac{1}{k} - o(k^{-2})\} + o(k^{-\frac{1}{2}})$$

$$= (k+2) + \frac{\gamma}{\sqrt{k_2}}(k+2) - \frac{2\gamma}{\sqrt{k_2}}(k_2+2) - \frac{2\gamma^2}{k_2}(k_2+2)$$

$$+\frac{3\gamma^2}{k_2}(1+\frac{\gamma}{\sqrt{k_2}})(\frac{k_2^2+6k_2+8}{k}) + o(k^{-\frac{1}{2}})$$

$$= (k+2) + \frac{\gamma}{\sqrt{k_2}}(\varphi k_2-k_2-2) + \frac{\gamma^2}{k_2}(-2k_2-4+\frac{3k_2}{1+\varphi}) + o(k^{-\frac{1}{2}}).$$

$$(5.4.20)$$

Finally,

$-2(k-2)(\beta k_1+k_2)S$

$$= -2(\beta\varphi+1)k_2\{1-\frac{\gamma}{\sqrt{k_2}}\frac{1}{1+\varphi}+\frac{\gamma^2}{k_2}\frac{1}{1+\varphi}\frac{k_2+2}{k+2}\}+o(k^{-\frac{1}{2}})$$

$$= -2(\beta\varphi+1)k_2+\frac{\gamma}{1+\varphi}2\sqrt{k_2}(\beta\varphi+1)-\frac{\gamma^2}{1+\varphi}2(\beta\varphi+1)(k_2+2)\{\frac{1}{k}-o(k^{-2})\}$$

$$+ o(k^{-\frac{1}{2}})$$

$$= -2(\beta\varphi+1)+\frac{2\gamma}{1+\varphi}\sqrt{k_2}(\beta\varphi+1)-\frac{\gamma^2}{k_2}\frac{2(\beta\varphi+1)}{(1+\varphi)^2}(k_2+2)$$

$$+ o(k^{-\frac{1}{2}}). \qquad (5.4.21)$$

Therefore using (5.4.18) to (5.4.21) and (5.4.14) we have,

$$\frac{\Delta}{\lambda_2} = -2(1+\beta)-k_2(1+\beta\varphi)+(k+2)\frac{1+\beta\varphi}{1+\varphi}$$

$$+\frac{\gamma}{\sqrt{k_2}}\{\frac{1+\beta\varphi}{1+\varphi}2k_2-k_2+2-\frac{2k_2\beta\varphi}{1+\varphi}+\frac{\varphi k_2-k_2-2}{1+\varphi}\}$$

$$+\frac{\gamma^2}{k_2}\{(k_2-2)-\frac{1+\beta\varphi}{(1+\varphi)^2}2(k_2+2)$$

$$+\frac{3\beta\varphi}{1+\varphi}(k_2-\varphi k_2+\frac{\varphi^2k_2}{1+\varphi})-\frac{2(k_2+2)}{1+\varphi}+\frac{3k_2}{(1+\varphi)^2}\}+o(k^{-\frac{1}{2}})$$

$$= -2(1+\beta)+\frac{2(1+\beta\varphi)}{1+\varphi}+\frac{\gamma}{\sqrt{k_2}}\frac{2\varphi}{1+\varphi}$$

$$+\frac{\gamma^2}{k_2}\{(k_2-2)+\frac{3k_2\beta\varphi-3k_2\beta\varphi^2-2k_2-4}{1+\varphi}$$

$$+\frac{3\beta\varphi^3k_2-2k_2-2\beta\varphi k_2-4-4\beta\varphi+3k_2}{(1+\varphi)^2}\}+o(k^{-\frac{1}{2}})$$

$$= -\frac{2(\beta+\varphi)}{1+\varphi} + \gamma^2\varphi\frac{\beta+\varphi}{(1+\varphi)^2} + o(k^{-\frac{1}{2}}) = \frac{\beta+\varphi}{1+\varphi}\{\frac{\gamma^2\varphi}{1+\varphi} - 2\} + o(k^{-\frac{1}{2}}).$$

(5.4.22)

Now letting k go to infinity completes the proof. ||

Note 1. The conclusion of Theorem 5.4.2 does not depend on β. It depends only on how far ρ is from one, and on the ratio of the group sizes.

Section 5.5. Numerical Results for Separation

In this section we will compute the Bayes risk given in (5.1.6) numerically. We will find the cut off points for ρ which will determine a region in which separation is better. First we will consider the case $\lambda_1 = \lambda_2$ and then will consider the case $\lambda_1 = \beta\lambda_2$ for $\beta \neq 1$.

Subsection 5.5.1. Numerical Results for $\lambda_1 = \lambda_2$

For $\lambda_1 = \lambda_2$, the difference of the Bayes risks of the combined and the separate estimator is given by (5.2.5). We know that $\chi^2_{k_1}$ is independent of $\chi^2_{k_2}$ in the expression (5.2.2). Thus the joint Laplace transform of $\chi^2_{k_1}$ and $\rho\chi^2_{k_2}$ is

$$E[e^{-t(\chi^2_{k_1}+\rho\chi^2_{k_2})}] = (1+2t)^{-\frac{k_1}{2}}(1+2t\rho)^{-\frac{k_2}{2}}.$$

Hence by Lemma 1 of the appendix, we have

$$E[\frac{1}{\chi^2_{k_1}+\rho\chi^2_{k_2}}] = \int_0^\infty (1+2t)^{-\frac{k_1}{2}}(1+2t\rho)^{-\frac{k_2}{2}} = \tfrac{1}{2}\int_0^\infty (1+t)^{-\frac{k_1}{2}}(1+t\rho)^{-\frac{k_2}{2}}dt.$$

(5.5.1)

Clearly from (5.3.2), for given values of k_1 and k_2, Δ is a function of ρ.

Empirical studies indicate that $\Delta(\rho)$ is inverted bell shaped and has two roots. We will determine the roots which will define an interval outside of which the separation is better.

In Table 5, the root between 0 and 1 of the equation $\Delta(\rho) = 0$ is given for various group sizes. The other root is obtained from symmetry considerations as indicated in the follwoing comment.

Comment 1. If ρ_0 is a solution to $\Delta(\rho) = 0$ when the first group size is k_1 and the second group size is k_2 then $1/\rho_0$ is a solution to $\Delta(\rho) = 0$ when the group sizes are interchanged. In particular when $k_1 = k_2$, if ρ_0 is a solution to $\Delta(\rho) = 0$, then $1/\rho_0$ is the other solution.

In Table 5, we have considered group sizes from 3 to 20. The roots of $\Delta(\rho)$ are also obtained for higher group sizes in Table 6. It is clear from Theorem 5.4.2 that the asymptotic lower bound when $k_1 = k_2$ is $\rho_0^* = 1 - 2/\sqrt{k_1}$. In Table 6, it is indeed observed that as the group size k_1 increases, the theoretical lower bound is approached.

Table 5

The cut off point for the separation problem under misspecification of Normal prior.
[The entries are the ρ values.]

k_2 vs. k_1 (under $\beta = 1$)

k_2 \ k_1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.14	.22	.28	.32	.36	.37	.40	.43	.43	.44	.45	.45	.46	.46	.47	.47	.47	.48
4	.15	.23	.29	.33	.37	.39	.41	.44	.45	.46	.46	.47	.48	.48	.49	.49	.50	.50
5	.15	.24	.30	.34	.38	.40	.42	.45	.46	.47	.48	.48	.49	.50	.51	.51	.52	.52
6	.15	.25	.31	.35	.39	.41	.43	.46	.47	.48	.49	.50	.51	.52	.52	.53	.53	.54
7	.16	.25	.32	.36	.40	.42	.44	.46	.48	.49	.50	.51	.52	.53	.53	.54	.55	.55
8	.16	.26	.32	.37	.40	.43	.45	.47	.49	.50	.51	.52	.53	.54	.54	.55	.56	.56
9	.16	.26	.33	.37	.41	.44	.46	.48	.49	.51	.52	.53	.54	.55	.55	.56	.56	.57
10	.16	.26	.33	.38	.41	.44	.46	.48	.50	.51	.52	.54	.54	.55	.56	.57	.57	.58
11	.17	.27	.33	.38	.42	.45	.47	.49	.51	.52	.53	.54	.55	.56	.57	.57	.58	.58
12	.17	.27	.33	.38	.42	.45	.47	.49	.51	.52	.53	.55	.56	.56	.57	.58	.59	.59
13	.17	.27	.34	.39	.42	.45	.48	.50	.52	.53	.54	.55	.56	.57	.58	.58	.59	.59
14	.17	.27	.34	.39	.43	.45	.48	.50	.52	.53	.54	.55	.56	.57	.58	.59	.60	.60
15	.17	.27	.34	.39	.43	.46	.48	.50	.52	.53	.55	.56	.57	.58	.58	.59	.60	.60
16	.17	.27	.34	.39	.43	.46	.49	.50	.52	.54	.55	.56	.57	.58	.59	.59	.60	.61
17	.17	.27	.34	.39	.43	.46	.49	.51	.53	.54	.55	.56	.57	.58	.59	.60	.61	.61
18	.17	.28	.34	.39	.43	.46	.49	.51	.53	.54	.55	.57	.58	.58	.59	.60	.61	.61
19	.17	.28	.35	.40	.43	.46	.49	.51	.53	.54	.56	.57	.58	.59	.60	.60	.61	.62
20	.17	.28	.35	.40	.44	.47	.49	.51	.53	.55	.56	.57	.58	.59	.60	.61	.61	.62

k = # of parameters

k_1 = 1st subgroup size

k_2 = 2nd subgroup size

$k = k_1 + k_2$

Table 6

Comparison between theoretical and numerical lower bounds

$k_1 = k_2$ and $\lambda_1 = \lambda_2$

k_1	20	25	36	49	64	100
ρ_0	.62	.65	.71	.74	.77	.81
ρ_0^*	.55	.60	.66	.71	.75	.80

Subsection 5.5.2. Numerical Results for $\lambda_1 \neq \lambda_2$

In this section we will consider $\lambda_1 = \beta\lambda_2$ for $\beta \neq 1$. To find the cut off points for the separation we will calculate T_1 and T_2 numerically in the expression (5.4.14). Clearly the joint Laplace transform of $\chi^2_{k_1}$ and $\chi^2_{k_2}$ is

$$\varphi(t_1,t_2) = (1+2t_1)^{-\frac{k_1}{2}} (1+2t_2\rho)^{-\frac{k_2}{2}} . \qquad (5.5.2)$$

The partial derivatives of $\varphi(t_1,t_2)$ are

$$\frac{\partial\varphi(t_1,t_2)}{\partial t_1} = -k_1(1+2t_1)^{-\frac{k_1+2}{2}} (1+2t_2\rho)^{-\frac{k_2}{2}}$$

and

$$\frac{\partial\varphi(t_1,t_2)}{\partial t_2} = -k_2\rho(1+2t_1)^{-\frac{k_1}{2}} (1+2t_2\rho)^{-\frac{k_2+2}{2}} . \qquad (5.5.3)$$

Now using Lemma 3 of the appendix we have,

$$T_1 = \int_0^\infty t_2 k_1 (1+2t_2)^{-\frac{k_1+2}{2}} (1+2t_2\rho)^{-\frac{k_2}{2}} dt_2$$

$$= \frac{k_1}{4} \int_0^\infty t(1+t)^{-\frac{k_1+2}{2}} (1+t\rho)^{-\frac{k_2}{2}} dt, \qquad (5.5.4)$$

and

$$T_2 = \rho \int_0^\infty t_1 k_2 (1+2t_1)^{-\frac{k_1}{2}} (1+2t_1\rho)^{-\frac{k_2+2}{2}} dt_1$$

$$= \frac{k_2\rho}{4} \int_0^\infty t(1+t)^{-\frac{k_1}{2}} (1+t\rho)^{-\frac{k_2+2}{2}} dt. \qquad (5.5.5)$$

It is easy to observe that in (5.5.4) and (5.5.5) we need to approximate the integrand near zero. Clearly for t near zero

$$t(1+t)^{-\frac{k_1+2}{2}} (1+t\rho)^{-\frac{k_2}{2}} \sim t(1+t)^{-\frac{k+2}{2}} \qquad (5.5.6)$$

where '~' means approximately equal. Therefore using (5.5.6) we have

$$\frac{4T_1}{k_1} \sim \int_0^\varepsilon t(1+t)^{-\frac{k+2}{2}} dt + \int_\varepsilon^M t(1+t)^{-\frac{k_1+2}{2}} (1+t\rho)^{-\frac{k_2}{2}} dt$$

where ε is small and M is large. Thus

$$T_1 \sim \frac{k_1}{4} [\frac{4}{k(k-2)} + 2(1+\varepsilon)^{-\frac{k}{2}} (\frac{1}{k} - \frac{1+\varepsilon}{k-2}) + \int_\varepsilon^M t(1+t)^{-\frac{k_1+2}{2}} (1+t\rho)^{-\frac{k_2}{2}} dt]. \qquad (5.5.7)$$

Similarly we have,

$$T_2 \sim \frac{k_2\rho}{4} [\frac{4}{k(k-2)} + 2(1+\varepsilon)^{-\frac{k}{2}} (\frac{1}{k} - \frac{1+\varepsilon}{k-2}) + \int_\varepsilon^M t(1+t)^{-\frac{k_1}{2}} (1+t\rho)^{-\frac{k_2+1}{2}} dt]. \qquad (5.5.8)$$

Now substituting (5.5.1), (5.5.7) and (5.5.8) in (5.4.14), we notice that Δ is a function of k_1, k_2, β and ρ. In Table 7, we have considered $k_1 = k_2$ and, for different values of β, the roots of $\Delta(\rho) = 0$ are obtained. Corresponding to each β and k_1, only one root is given in Table 7. The other root is obtained as indicated in the following comment.

Comment 2. If ρ_0 is a solution to $\Delta(\rho) = 0$ when the first group size is k_1 and the second group size is k_2 and the ratio λ_1/λ_2 is β then $1/\rho_0$ is a solution to $\Delta(\rho) = 0$ when the group sizes are interchanged and the ratio λ_1/λ_2 becomes $1/\beta$ (for $\beta < \infty$).

Comment 3. For $\beta = \infty$, (5.4.14) reduces to

$$\frac{\Delta}{\lambda_1} = k_1 - 2 + (k^2-4)T_1 - 2(k-2)k_1 S.$$

In Tables 7, 8 and 9, the cut off points are obtained for different values of β and k_1. In Table 8, k_2 is taken as $k_1 + 2$ and in Table 9, k_2 is taken as $k_1 + 4$.

Comment 4. The Tables 7, 8 and 9 show that for $\beta = 0$ the combined estimator is better than the separate estimator when ρ belongs to the region obtained from the tables. This will not be true if the estimator (1.3.1) is considered instead of (1.3.3). Indeed if the estimator (1.3.1) is considered then for $\beta = 0$ the separate estimator is always better than the combined estimator. This difficulty with (1.3.3) will occur mainly for extreme β.

Table 7
cut off points
β vs. k_1 $(k = 2k_1)$

β \ k_1	3	4	5	6	7	8	9	10	15	20
∞	.10	.17	.21	.23	.25	.26	.27	.28	.31	.33
20	.11	.18	.22	.23	.25	.27	.28	.29	.32	.34
15	.11	.18	.22	.23	.25	.27	.28	.29	.32	.35
10	.11	.18	.22	.23	.26	.27	.28	.29	.33	.35
7	.11	.18	.22	.24	.26	.28	.29	.30	.34	.36
6	.11	.18	.23	.24	.26	.28	.29	.30	.34	.36
5	.11	.18	.23	.25	.27	.29	.30	.31	.35	.37
4	.11	.19	.23	.25	.27	.29	.30	.31	.36	.38
3	.11	.19	.24	.26	.29	.30	.32	.33	.37	.40
2	.12	.20	.25	.29	.31	.33	.34	.36	.41	.44
1	.14	.23	.30	.36	.40	.43	.46	.48	.57	.62
0.5	.19	.30	.43	.55	.64	.58	.72	.75	.83	.87
0.3333	.23	.37	.55	.67	.74	.78	.80	.83	.88	.91
0.25	.26	.43	.61	.73	.78	.81	.84	.85	.90	.92
0.2	.29	.48	.65	.76	.80	.83	.85	.87	.91	.93
0.1666	.31	.51	.68	.77	.82	.84	.86	.88	.92	.93
0.1428	.33	.53	.69	.79	.83	.85	.87	.88	.92	.94
0.1	.37	.56	.79	.80	.85	.86	.88	.89	.93	.94
0.0666	.40	.60	.74	.82	.85	.87	.89	.90	.93	.94
0.05	.42	.62	.75	.82	.86	.88	.89	.90	.93	.95
0	.47	.66	.85	.86	.88	.89	.90	.91	.94	.95

Table 8
cut off points
β vs. k_1 ($k_2 = k_1+2$, i.e. $k = 2(k_1+1)$)

β \ k_1	3	4	5	6	7	8	9	10	15	20
∞	.12	.19	.23	.26	.27	.29	.30	.30	.32	.34
20	.12	.19	.24	.26	.28	.29	.30	.31	.32	.34
15	.12	.19	.24	.26	.28	.29	.30	.31	.32	.36
10	.12	.20	.24	.26	.28	.30	.31	.32	.33	.37
7	.12	.20	.24	.27	.29	.30	.31	.32	.34	.38
6	.13	.20	.25	.27	.29	.31	.32	.32	.35	.39
5	.13	.20	.25	.27	.29	.31	.32	.33	.36	.40
4	.13	.20	.25	.28	.30	.31	.33	.33	.37	.41
3	.13	.21	.26	.29	.31	.32	.34	.35	.39	.41
2	.13	.22	.27	.30	.33	.35	.36	.38	.43	.45
1	.15	.25	.32	.37	.41	.44	.47	.49	.57	.62
0.5	.18	.32	.44	.55	.61	.67	.70	.73	.82	.86
0.3333	.22	.41	.56	.67	.73	.77	.80	.82	.88	.91
0.25	.27	.49	.63	.73	.78	.81	.84	.85	.90	.92
0.2	.31	.54	.68	.76	.80	.83	.85	.87	.91	.93
0.1666	.34	.58	.70	.78	.82	.84	.86	.87	.91	.93
0.1428	.37	.60	.72	.79	.83	.85	.87	.88	.92	.94
0.1	.43	.65	.75	.81	.84	.86	.88	.89	.93	.94
0.0666	.49	.68	.77	.83	.85	.87	.89	.90	.93	.95
0.05	.52	.70	.79	.84	.86	.88	.89	.90	.93	.95
0	.59	.74	.81	.85	.88	.89	.90	.91	.94	.95

Table 9
cut off points
β vs. k_1 ($k_2 = k_1+4$, i.e. $k = 2(k_1+2)$)

β \ k_1	3	4	5	6	7	8	9	10	15	20
∞	.13	.20	.25	.27	.29	.31	.32	.32	.35	.37
20	.13	.21	.25	.28	.30	.31	.32	.33	.36	.37
15	.13	.21	.25	.28	.30	.31	.33	.33	.36	.38
10	.13	.21	.25	.28	.30	.31	.33	.34	.37	.39
7	.13	.21	.26	.28	.30	.32	.33	.34	.37	.39
6	.14	.21	.26	.29	.31	.32	.34	.35	.38	.40
5	.14	.21	.26	.29	.31	.33	.34	.35	.38	.41
4	.14	.22	.26	.29	.32	.33	.35	.36	.39	.41
3	.14	.22	.27	.30	.32	.34	.36	.37	.41	.43
2	.14	.23	.28	.32	.34	.36	.38	.39	.44	.47
1	.16	.26	.32	.38	.42	.45	.48	.50	.58	.63
0.5	.19	.33	.44	.54	.60	.65	.69	.72	.81	.86
0.3333	.23	.42	.56	.66	.72	.76	.79	.81	.87	.90
0.25	.28	.50	.64	.72	.77	.80	.83	.84	.89	.92
0.2	.34	.56	.71	.76	.80	.83	.85	.86	.91	.93
0.1666	.39	.60	.74	.78	.81	.84	.86	.87	.91	.93
0.1428	.43	.63	.76	.79	.82	.85	.87	.88	.92	.94
0.1	.51	.68	.79	.81	.84	.86	.88	.89	.92	.94
0.0666	.57	.71	.81	.83	.86	.87	.89	.90	.93	.95
0.05	.60	.73	.82	.84	.86	.88	.89	.90	.93	.95
0	.68	.77	.85	.86	.88	.89	.90	.91	.94	.95

BIBLIOGRAPHY

[1] Berger, J. (1975). Miminax 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. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. Ann. Statist. 8, No. 4.

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

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

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

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

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

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

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

[13] Rao, C. R. (1973). Linear Statistical Inference and Its Application. Second edition, John Wiley and Sons.

[14] Robbins, H. (1955). An emperical Bayes approach to statistics. Proc. Third Berkeley Symposium Math. Stat. Prob. 1, 157-163, University of California Press.

[15] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist., 35, 1-20.

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

[17] Stein, C. (1973). Estimation of the mean of a multivariate distribution. Proc. Prague Symp. Asymptotic Statist. 345-381.

[18] Stein, C. (1974). Estimation of the parameters of a multivariate normal distribution. I. Estimation of the means. Technical Report No. 63, Stanford University.

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

[20] Strawderman, W. E. (1973). Proper Bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances. Ann. Statist. 1, 1189-1194.

[21] Strawderman, W. E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. J. Multivariate Anal. 4, 255-264.

APPENDIX

Lemma 1. If $\{X_i\}$ $i = 1,2,\ldots,S$ is a sequence of independent random variables with $X_i \sim \chi^2_{k_i}$, $i = 1,2,\ldots,S$ and $a_1, a_2, \ldots, a_S$ are scalars, then

$$E\left[\frac{\sum_{i=1}^{S} a_i X_i}{\left(\sum_{i=1}^{S} X_i\right)^2} \right] = \frac{1}{k(k-2)} \sum_{i=1}^{S} a_i k_i \tag{A1}$$

where $k = \sum_{i=1}^{S} k_i$.

Proof. By the additive property of the chi-square distribution, we know that $\sum_{i=1}^{S} X_i \sim \chi^2_k$ and it can be easily shown that

$$\frac{X_i}{\sum_{i=1}^{S} X_i} \sim B_e\left(\frac{k_i}{2}, \frac{k-k_i}{2} \right) .$$

Therefore

$$E\left(\frac{X_i}{\sum_{i=1}^{S} X_i} \right) = \frac{k_i/2}{k/2} = \frac{k_i}{k}, \quad i = 1,2,\ldots,S .$$

Now by Basu's theorem $\frac{X_i}{\sum_{i=1}^{S} X_i}$ is independent of $\sum_{i=1}^{S} X_i$, $i = 1,2,\ldots,S$.

Therefore

$$E\left[\frac{X_i}{(\sum_{i=1}^{S} X_i)^2}\right] = E\left[\frac{1}{\sum_{i=1}^{S} X_i} \quad \frac{X_i}{\sum_{i=1}^{S} X_i}\right] = \frac{k_i}{(k-2)k}, \quad i = 1,2,\ldots,S.$$

Thus we have

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

which completes the proof. ||

Lemma 2. Suppose $Z > 0$ is a random variable such that its Laplace transform $E(e^{-tZ})$ exists. Then

$$E(\tfrac{1}{Z}) = \int_0^\infty E(e^{-tZ})dt. \tag{A2}$$

Proof. Let $F(Z)$ be the distribution function of Z. Then

$$\int_0^\infty E(e^{-tZ})dt = \int_0^\infty \int_0^\infty e^{-tZ} dF(Z)dt$$

$$= \int_0^\infty \left(\int_0^\infty e^{-tZ}dt\right)dF(Z) , \quad \text{(by Fubini's theorem)}$$

$$= E\left[- \frac{e^{-tZ}}{Z} \Big|_0^\infty \right] = E(\tfrac{1}{Z})$$

which completes the proof. ||

Lemma 3. If $Z_1, Z_2, \ldots, Z_k$ is a sequence of independent random variables such that

$$\varphi(t_1,t) = E\left[e^{-t_1 Z_i - t \sum_{j \neq i} Z_j}\right]$$

is finite and differentiable, then

$$E\left[\frac{Z_i}{(\sum_{j=1}^{k} Z_j)^2} \right] = - \int_0^\infty t \left. \frac{\partial \varphi (t_1,t)}{\partial t_1} \right|_{t_1=t} dt. \qquad (A3)$$

Proof. We have,

$$\frac{\partial \varphi}{\partial t_1} = - E[Z_i e^{-t_1 Z_i - t \sum_{j \neq i} Z_j}] ,$$

and hence

$$\left. \frac{\partial \varphi}{\partial t_1} \right|_{t_1=t} = - E[Z_i e^{-t \sum_{j=1}^{k} Z_j}] .$$

Therefore,

$$- \int_0^\infty t \left. \frac{\partial \varphi (t_1,t)}{\partial t_1} \right|_{t_1-t} = \int_0^\infty t\, E[Z_i e^{-t \sum_{j=1}^{k} Z_j}]\, dt$$

$$= E[\int_0^\infty t Z_i e^{-t \sum_{j=1}^{k} Z_j} dt]$$

$$= E[Z_i \{[-t \frac{e^{-t \sum_{j=1}^{k} Z_j}}{\sum_{j=1}^{k} Z_j}]_0^\infty + \int_0^\infty \frac{e^{-t \sum_{j=1}^{k} Z_j}}{\sum_{j=1}^{k} Z_j} dt\}]$$

$$= E[\frac{Z_i}{(\sum_{j=1}^{k} Z_j)^2}],$$

which completes the proof. ||

Lemma 4.
$$\int_{\sum_{i=1}^{n} X_i^2 \le a^2} \frac{1}{\sum_{i=1}^{n} X_i^2} \prod_{i=1}^{n} dX_i = \frac{a^{n-2}(\sqrt{\pi})^n}{\Gamma(\frac{n}{2}+1)} \qquad \text{(A4)}$$

and

$$\int_{\sum_{i=1}^{n}} \frac{1}{(\sum_{i=1}^{n} X_i^2)^2} \prod_{i=1}^{n} dX_i = \frac{a^{n-4}(\sqrt{\pi})^n}{\Gamma(\frac{n}{2}+1)} \qquad \text{(A5)}$$

Proof. Consider the polar transformation in the spherical neighborhood of the origin as:

$$x_1 = r \cos \theta_1 \cos \theta_2 \ldots \cos \theta_{n-2} \cos \theta_{n-1}, \qquad 0 \le \theta_{n-1} \le 2\pi$$

$$x_2 = r \cos \theta_1 \cos \theta_2 \ldots \cos \theta_{n-2} \sin \theta_{n-1}, \qquad -\frac{\pi}{2} \le \theta_{n-2} \le \frac{\pi}{2}$$

$\cdot$

$\cdot$

$$x_n = r \sin \theta_1, \qquad -\frac{\pi}{2} \le \theta_1 \le \frac{\pi}{2}.$$

Then it can be shown that the Jacobian is

$$J = r^{n-1} \prod_{i=1}^{n-2} (\cos \theta_i)^{n-i-1}$$

and $\sum_{i=1}^{n} x_i^2 = r^2$. Then (A4) reduces to

$$\int_0^a r^{n-3} dr \int_{-\pi/2}^{\pi/2} \cos^{n-2}\theta_1 d\theta_1 \ldots \int_{-\pi/2}^{\pi/2} \cos \theta_{n-2} d\theta_{n-2} \int_0^{2\pi} d\theta_{n-1} = \frac{a^{n-2}(\sqrt{\pi})^n}{\Gamma(\frac{n}{2}+1)}.$$

(A5) follows similarly. ||