Estimating the Common Mean of Two Multivariate Normal Distributions

by

Wei-Liem Loh
Purdue University

Technical Report #88-48

Department of Statistics
Purdue University

October 1988

AMS 1980 subject classifications: Primary 62F10, secondary 62C99.

Key words and phrases: Common mean, equivariant estimation, unbiased estimate of risk, Wishart distribution.

ESTIMATING THE COMMON MEAN OF TWO MULTIVARIATE NORMAL DISTRIBUTIONS

By Wei-Liem Loh

Purdue University

Let X_1, X_2 be two $p \times 1$ multivariate normal random vectors and S_1, S_2 be two $p \times p$ Wishart matrices where $X_1 \sim N_p(\xi, \Sigma_1)$, $X_2 \sim N_p(\xi, \Sigma_2)$, $S_1 \sim W_p(\Sigma_1, n)$ and $S_2 \sim W_p(\Sigma_2, n)$. We further assume that X_1, X_2, S_1, S_2 are stochastically independent. We wish to estimate the common mean ξ with respect to the loss function $L = (\hat{\xi} - \xi)'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \xi)$. By extending the methods of Stein (1975) and Haff (1982), alternative unbiased estimators to the usual generalized least squares estimator are obtained. However the risks of these estimators are not available in closed form. A Monte Carlo swindle is used instead to evaluate their risk performances. The results indicate that these alternative estimators perform very favorably against the usual estimator.

1 Introduction

In this paper we consider the problem of estimating the common mean of two multivariate normal distrbutions with unknown covariance matrices. The precise formulation of this problem is as follows:

Let X_1, X_2 be two $p \times 1$ multivariate normal random vectors and S_1, S_2 be two $p \times p$ Wishart matrices where $X_1 \sim N_p(\xi, \Sigma_1)$, $X_2 \sim N_p(\xi, \Sigma_2)$, $S_1 \sim W_p(\Sigma_1, n)$, $S_2 \sim W_p(\Sigma_2, n)$ with X_1, X_2, S_1, S_2 mutually independent. We wish to estimate ξ under the quadratic loss function:

$$L(\hat{\xi}; \xi, \Sigma_1, \Sigma_2) = (\hat{\xi} - \xi)'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \xi).$$

The above loss function is a natural symmetric extension of the following invariant loss function:

$$L_0(\hat{\xi}; \xi, \Sigma) = (\hat{\xi} - \xi)'\Sigma^{-1}(\hat{\xi} - \xi),$$

which was first considered by James and Stein (1961) in estimating the mean ξ of a multivariate normal distribution with unknown covariance matrix Σ.

The above problem is a canonical formulation of the following more common situation: Let $Y_1, \ldots, Y_{n+1}$ and $Z_1, \ldots, Z_{n+1}$ be two samples from two multivariate normal distribuions $N_p(\theta, \Sigma_1)$ and $N_p(\theta, \Sigma_2)$ respectively with θ, Σ_1 and Σ_2 unknown. We wish to estimate θ, or equivalently $\sqrt{n+1}\theta$. The sufficient statistics are

$$X_1 = \textstyle\sum_i Y_i/\sqrt{n+1}, \qquad S_1 = \displaystyle\sum_i (Y_i - X_1)^2,$$

$$X_2 = \textstyle\sum_i Z_i/\sqrt{n+1}, \qquad S_2 = \displaystyle\sum_i (Z_i - X_2)^2.$$

Then clearly we have for $i = 1, 2$,

$$X_i \sim N_p(\xi, \Sigma_i), \qquad S_i \sim W_p(\Sigma_i, n),$$

where $\xi = \sqrt{n+1}\theta$. Also, we observe that X_1, X_2, S_1 and S_2 are stochastically independent. This reduces to the above formulation.

When $p = 1$, there is a lot of research done on this problem. In particular, Graybill and Deal (1959) have shown that the unbiased estimator for ξ given by

$$\hat{\xi} = (S_2 X_1 + S_1 X_2)/(S_1 + S_2)$$

has smaller variance than either of X_1 or X_2 if $n > 10$. Other related literature include Brown and Cohen (1974), Cohen and Sackrowitz (1974), Khatri and Shah (1974).

When $p > 1$, Chiou and Cohen (1985) discuss this problem by evaluating unbiased estimators of ξ using their covariance matrices as a criterion. Also, Shinozaki (1978) considered the estimation of the common mean of k multivariate normal distributions where the covariance matrices are known up to an arbitrary constant.

We shall use the following notation throughout. If a matrix A has entries a_{ij}, we shall indicate it by (a_{ij}). Given a $r \times s$ matrix A, its $s \times r$ transpose is denoted by A'. $|A|$, A^{-1} denote the determinant, inverse of the square matrix A respectively. The trace of A is indicated by tr$A$ and I denotes the identity matrix. If the $p \times p$ matrix A is diagonal and has entries a_{ij}, we shall write it as $A = \text{diag}(a_{11}, \ldots, a_{pp})$. Finally, the expected value of a random vector X is denoted by EX.

2 Equivariant Estimators

The problem that we are considering is invariant under the group of affine transformations:

$$\xi \to A\xi + \alpha, \qquad X_i \to AX_i + \alpha,$$
$$\Sigma_i \to A\Sigma_i A', \qquad S_i \to AS_i A',$$

where $\alpha \in R^p$, $A \in GL(p,R)$ and $i = 1,2$. For simplicity of notation, if $x = (x_1, \ldots, x_p)'$ we define $| x |^i$ to be $(| x_1 |^i, \ldots, | x_p |^i)'$.

Theorem 1 *Let $X_i \sim N_p(\xi, \Sigma_i)$, $S_i \sim W_p(\Sigma_i, n)$, $i = 1,2$ with X_1, X_2, S_1, S_2 independent. Then under the group of affine transformations, $\hat{\xi}$ is an equivariant estimator for ξ if and only if $\hat{\xi}$ can be expressed as*

$$\hat{\xi}(X_1, X_2, S_1, S_2) = B^{-1}\Phi B X_1 + B^{-1}(I - \Phi)BX_2,$$

where $\Phi = \Phi(| B(X_1 - X_2) |^2, F)$ is a diagonal matrix, $B(S_1 + S_2)B' = I$, $BS_2B' = F = diag(f_1, \ldots, f_p)$ with $f_1 \geq \cdots \geq f_p$.

PROOF. Suppose that $\hat{\xi}$ is an equivariant estimator for ξ. First we observe that there exists a diagonal matrix Φ such that

$$\hat{\xi}(| B(X_1 - X_2) |, 0, I - F, F) = \Phi(| B(X_1 - X_2) |^2, F) \; | B(X_1 - X_2) | .$$

Then by equivariance, for $D = \text{diag}(\pm 1)$ satisfying $B(X_1 - X_2) = D \; | B(X_1 - X_2) |$, we have

$$\begin{aligned} \hat{\xi}(B(X_1 - X_2), 0, I - F, F) &= \hat{\xi}(D \; | B(X_1 - X_2) |, 0, I - F, F) \\ &= D\hat{\xi}(| B(X_1 - X_2) |, 0, I - F, F) \\ &= D\Phi(| B(X_1 - X_2) |^2, F) \; | B(X_1 - X_2) | \\ &= \Phi(| B(X_1 - X_2) |^2, F)B(X_1 - X_2). \qquad (1) \end{aligned}$$

Finally we observe that

$$\begin{aligned} \hat{\xi}(X_1, X_2, S_1, S_2) &= \hat{\xi}((X_1 - X_2), 0, S_1, S_2) + X_2 \\ &= B^{-1}\hat{\xi}(B(X_1 - X_2), 0, I - F, F) + X_2 \\ &= B^{-1}\Phi(| B(X_1 - X_2) |^2, F)B(X_1 - X_2) + X_2 \\ &= B^{-1}\Phi B X_1 + B^{-1}(I - \Phi)BX_2. \end{aligned}$$

Here the second last equality follows from (1). This proves the necessity part. For the sufficiency part of the result, the proof is straightforward and is omitted. □

3 Calculus on Eigenstructure

Let $X_1 \sim N_p(\xi, \Sigma_1)$, $S_1 \sim W_p(\Sigma_1, n)$, $X_2 \sim N_p(\xi, \Sigma_2)$ and $S_2 \sim W_p(\Sigma_2, n)$. For $i = 1,\ 2$ and $1 \leq j,\ k \leq p$, we write:

$$\nabla^{(i)} = (\nabla_1^{(i)}, \ldots, \nabla_p^{(i)})', \qquad \tilde{\nabla}^{(i)} = (\tilde{\nabla}_{jk}^{(i)})_{p\times p},$$

where

$$x_j^{(i)} = (X_i)_j, \qquad \nabla_j^{(i)} = \partial/\partial x_j^{(i)},$$
$$s_{jk}^{(i)} = (S_i)_{jk}, \qquad \tilde{\nabla}_{jk}^{(i)} = (1/2)(1+\delta_{jk})\partial/\partial s_{jk}^{(i)},$$

and δ_{jk} denotes the Kronecker delta. We observe that there exists $B \in GL(p, R)$ such that $BS_1B' = I - F$ and $BS_2B' = F$ where $F = \mathrm{diag}(f_1, \ldots, f_p)$ and $f_1 \geq \cdots \geq f_p$. We shall now compute the partial derivatives of B, B^{-1} and F with respect to S_1 and S_2.

Theorem 2 *Let $X_i \sim N_p(\xi, \Sigma_i)$ and $S_i \sim W_p(\Sigma_i, n)$, $i = 1,\ 2$. Then with F, $B = (b_{il})$ and $B^{-1} = (b^{il})$ as defined above, we have*

$$\begin{aligned}
\tilde{\nabla}_{jk}^{(1)} f_i &= -f_i b_{ij} b_{ik}, \\
\tilde{\nabla}_{jk}^{(2)} f_i &= (1-f_i) b_{ij} b_{ik}, \\
\tilde{\nabla}_{jk}^{(1)} b_{il} &= -\frac{1}{2} b_{il} b_{ij} b_{ik} - \frac{1}{2} \sum_{k'\neq i} b_{k'l}(b_{ij}b_{k'k} + b_{ik}b_{k'j}) \frac{f_i}{f_i - f_{k'}}, \\
\tilde{\nabla}_{jk}^{(2)} b_{il} &= -\frac{1}{2} b_{il} b_{ij} b_{ik} + \frac{1}{2} \sum_{k'\neq i} b_{k'l}(b_{ij}b_{k'k} + b_{ik}b_{k'j}) \frac{1-f_i}{f_i - f_{k'}}, \\
\tilde{\nabla}_{jk}^{(1)} b^{il} &= \frac{1}{2} b^{il} b_{lj} b_{lk} - \frac{1}{2} \sum_{i'\neq l} b^{ii'}(b_{i'j}b_{lk} + b_{i'k}b_{lj}) \frac{f_{i'}}{f_l - f_{i'}}, \\
\tilde{\nabla}_{jk}^{(2)} b^{il} &= \frac{1}{2} b^{il} b_{lj} b_{lk} + \frac{1}{2} \sum_{i'\neq l} b^{ii'}(b_{i'j}b_{lk} + b_{i'k}b_{lj}) \frac{1-f_{i'}}{f_l - f_{i'}}.
\end{aligned}$$

PROOF. The proofs of the first, second, fifth and sixth equalities can be found in Loh (1988). It leaves us now only to consider the third and fourth equalities. To prove the third equality, we observe that

$$\begin{aligned}
0 &= \tilde{\nabla}_{jk}^{(1)}(BB^{-1}) \\
&= B(\tilde{\nabla}_{jk}^{(1)} B^{-1}) + (\tilde{\nabla}_{jk}^{(1)} B) B^{-1}.
\end{aligned}$$

This implies that
$$\tilde{\nabla}^{(1)}_{jk} B = -B(\tilde{\nabla}^{(1)}_{jk} B^{-1})B.$$
Considering the il'th component, we get
$$\begin{aligned}
&\tilde{\nabla}^{(1)}_{jk} b_{il} \\
= \quad & -\sum_{j',k'} b_{ij'}(\tilde{\nabla}^{(1)}_{jk} b^{j'k'}) b_{k'l} \\
= \quad & -\sum_{j',k'} b_{ij'} b_{k'l}[\frac{1}{2} b^{j'k'} b_{k'j} b_{k'k} + \frac{1}{2}\sum_{i'\neq k'} b^{j'i'}(b_{i'j}b_{k'k} + b_{i'k}b_{k'j})\frac{f_{i'}}{f_{i'} - f_{k'}}] \\
= \quad & -\frac{1}{2} b_{il} b_{ij} b_{ik} - \frac{1}{2}\sum_{k'\neq i} b_{k'l}(b_{ij}b_{k'k} + b_{ik}b_{k'j})\frac{f_i}{f_i - f_{k'}}.
\end{aligned}$$
The proof of the fourth equality is similar and hence is omitted. □

4 Two Identities

In this section we shall state the Normal and Wishart identities. Their proofs are given in Loh (1988) and hence are omitted. These identities are crucial in the derivation of the unbiased estimate of risk of an equivariant estimator in the problem that we are considering.

First we need some additional definitions. A function $g : R^{p\times n} \to R$ is almost differentiable if, for every direction, the restrictions to almost all lines in that direction are absolutely continuous. If g on $R^{p\times n}$ is vector-valued instead of being real-valued, then g is almost differentiable if each of its coordinate functions are.

Theorem 3 (Normal Identity) *Let $X = (X_1, \ldots, X_p)' \sim N_p(\xi, \Sigma)$ and $g : R^p \to R^p$ be an almost differentiable function such that $E[\sum_{i,j} \mid \partial g_i(X) / \partial X_j \mid]$ is finite. Then*
$$E[\Sigma^{-1}(X - \xi)g'(X)] = E[\nabla g'(X)],$$
where $\nabla = (\partial/\partial X_1, \ldots, \partial/\partial X_p)'$.

The Normal identity was first proved by Stein (1973).

Let S_p denote the set of $p \times p$ positive definite matrices. Also we write for $1 \le i, j \le p$,
$$\tilde{\nabla} = (\tilde{\nabla}_{ij})_{p\times p}, \text{ where } \tilde{\nabla}_{ij} = (1/2)(1 + \delta_{ij})\partial/\partial s_{ij},$$
where δ_{ij} denotes the Kronecker delta.

Theorem 4 (Wishart Identity) *Let $X = (X_1, \ldots, X_n)$ be a $p \times n$ random matrix, with the X_k independently normally distributed p-dimensional random vectors with mean 0 and unknown covariance matrix Σ. We suppose $n \geq p$. Let $g : S_p \to R^{p\times p}$ be such that $x \mapsto g(xx') : R^{p\times n} \to R^{p\times p}$ is almost differentiable. Then, with $S = (s_{ij}) = XX'$, we have*

$$Etr\Sigma^{-1}g(S) = Etr[(n-p-1)S^{-1}g(S) + 2\tilde{\nabla}g(S)],$$

provided the expectations of the two terms on the r.h.s. exist.

The Wishart identity was proved by Stein (1975) and Haff (1977) independently.

5 Unbiased Estimate of Risk

Here we shall compute the unbiased estimate of risk of a subclass of equivariant estimators for ξ of the form:

$$\hat{\xi} = B^{-1}\Phi BX_1 + B^{-1}(I-\Phi)BX_2,$$

where $\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_p)$ depends only on F. It is clear that estimators of this kind are unbiased. First we need a couple of rather technical lemmas.

Lemma 1 *With the notation of Theorem 1, for $\Phi = \Phi(F)$ we have*

$$\begin{aligned} \nabla^{(1)'}[B^{-1}(I-\Phi)B(X_2-X_1)] &= -p + \sum_i \phi_i, \\ \nabla^{(2)'}[B^{-1}\Phi B(X_1-X_2)] &= -\sum_i \phi_i. \end{aligned}$$

PROOF. We observe that

$$\begin{aligned} & \nabla^{(1)'}[B^{-1}(I-\Phi)B(X_2-X_1)] \\ = & \sum_i \sum_j \frac{\partial}{\partial x_i^{(1)}}[B^{-1}(I-\Phi)B]_{ij}(X_2-X_1)_j \\ = & -\sum_i [B^{-1}(I-\Phi)B]_{ii} \\ = & -p + \sum_i \phi_i. \end{aligned}$$

The second part of the lemma can be proved similarly. □

Lemma 2 *With the notation of Theorem 1, for $\Phi = \Phi(F)$ we have*

1. $$tr\tilde{\nabla}^{(1)}\{[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]'\}$$
$$= \sum_i \{[B(X_1-X_2)]_i^2(1-\phi_i)^2 \sum_{j\neq i} \frac{f_j}{f_j-f_i}$$
$$+2[B(X_1-X_2)]_i^2(1-\phi_i)f_i\frac{\partial \phi_i}{\partial f_i}$$
$$-\sum_{j\neq i}[B(X_1-X_2)]_j^2(1-\phi_i)(1-\phi_j)\frac{f_i}{f_i-f_j}\},$$

2. $$tr\tilde{\nabla}^{(2)}\{[B^{-1}\Phi B(X_1-X_2)][B^{-1}\Phi B(X_1-X_2)]'\}$$
$$= \sum_i\{[B(X_1-X_2)]_i^2\phi_i^2\sum_{j\neq i}\frac{1-f_j}{f_i-f_j} + 2[B(X_1-X_2)]_i^2\phi_i(1-f_i)\frac{\partial\phi_i}{\partial f_i}$$
$$-\sum_{j\neq i}[B(X_1-X_2)]_j^2\phi_i\phi_j\frac{1-f_i}{f_j-f_i}\}.$$

PROOF. First we observe that

$$\{\tilde{\nabla}^{(1)}[B^{-1}(I-\Phi)B(X_2-X_1)]\}_i$$
$$= \sum_{j,k,l}\tilde{\nabla}_{ij}^{(1)}[b^{jl}(1-\phi_l)b_{lk}](X_2-X_1)_k$$
$$= \sum_{j,k,l}[(\tilde{\nabla}_{ij}^{(1)}b^{jl})(1-\phi_l)b_{lk} - b^{jl}(\tilde{\nabla}_{ij}^{(1)}\phi_l)b_{lk}$$
$$+b^{jl}(1-\phi_l)(\tilde{\nabla}_{ij}^{(1)}b_{lk})](X_2-X_1)_k$$
$$= \sum_{k,l}[\frac{1}{2}b_{li}b_{lk}(1-\phi_l)\sum_{i'\neq l}\frac{f_{i'}}{f_{i'}-f_l} + b^{jl}b_{lk}\sum_m(b_{mi}b_{mj}f_m\frac{\partial\phi_l}{\partial f_m})$$
$$-\frac{1}{2}\sum_{k'\neq l}b_{k'k}b_{k'i}(1-\phi_l)\frac{f_l}{f_l-f_{k'}}](X_2-X_1)_k$$
$$= \sum_l\{\frac{1}{2}b_{li}[B(X_2-X_1)]_l(1-\phi_l)\sum_{i'\neq l}\frac{f_{i'}}{f_{i'}-f_l} + b_{li}[B(X_2-X_1)]_lf_l\frac{\partial\phi_l}{\partial f_l}$$
$$-\frac{1}{2}\sum_{k'\neq l}b_{k'i}[B(X_2-X_1)]_{k'}(1-\phi_l)\frac{f_l}{f_l-f_{k'}}\}. \quad (2)$$

Here the second last equality follows from Theorem 2. Hence we have

$$\mathrm{tr}\tilde{\nabla}^{(1)}\{[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]'\}$$

$$
\begin{aligned}
&= 2\mathrm{tr}[\tilde{\nabla}^{(1)}B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]' \\
&= 2\sum_i[\tilde{\nabla}^{(1)}B^{-1}(I-\Phi)B(X_2-X_1)]_i[B^{-1}(I-\Phi)B(X_2-X_1)]_i \\
&= \sum_i\{[B(X_1-X_2)]_i^2(1-\phi_i)^2\sum_{j\neq i}\frac{f_j}{f_j-f_i} \\
&\quad +2[B(X_1-X_2)]_i^2(1-\phi_i)f_i\frac{\partial\phi_i}{\partial f_i} \\
&\quad -\sum_{j\neq i}[B(X_1-X_2)]_j^2(1-\phi_i)(1-\phi_j)\frac{f_i}{f_i-f_j}\}.
\end{aligned}
$$

The last equality follows from (2). The second part of this lemma can be proved similarly. □

Now we shall prove the main result of this section.

Theorem 5 *Let $\hat{\xi}$ be an estimator for ξ where*

$$\hat{\xi}(X_1,X_2,S_1,S_2) = B^{-1}\Phi BX_1 + B^{-1}(I-\Phi)BX_2,$$

$\Phi = \Phi(F) = diag(\phi_1,\ldots,\phi_p)$, $B(S_1+S_2)B' = I$, $BS_2B' = F = diag(f_1,\ldots, f_p)$ with $f_1 \geq \cdots \geq f_p$. Suppose Φ satisfies the Normal and Wishart identities in the following sense:

1. $E(X_1-\xi)'\Sigma_1^{-1}B^{-1}(I-\Phi)B(X_2-X_1)$
   $= E\nabla^{(1)'}[B^{-1}(I-\Phi)B(X_2-X_1)]$,
2. $E(X_2-\xi)'\Sigma_2^{-1}B^{-1}\Phi B(X_1-X_2) = E\nabla^{(2)'}[B^{-1}\Phi B(X_1-X_2)]$,
3. $E\ tr\Sigma_1^{-1}[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]'$
   $= E\ tr(n-p-1)S_1^{-1}[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]'$
   $+2\tilde{\nabla}^{(1)}\{[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]'\}$,
4. $E\ tr\Sigma_2^{-1}[B^{-1}\Phi B(X_1-X_2)][B^{-1}\Phi B(X_1-X_2)]'$
   $= E\ tr(n-p-1)S_2^{-1}[B^{-1}\Phi B(X_1-X_2)][B^{-1}\Phi B(X_1-X_2)]'$
   $+2\tilde{\nabla}^{(2)}\{[B^{-1}\Phi B(X_1-X_2)][B^{-1}\Phi B(X_1-X_2)]'\}$.

Then the risk of $\hat{\xi}$ is given by

$$
\begin{aligned}
R(\hat{\xi};\xi,\Sigma_1,\Sigma_2) &= E\{\sum_i[B(X_1-X_2)]_i^2[\frac{n-p-1}{f_i}\phi_i^2+4(1-f_i)\phi_i\frac{\partial\phi_i}{\partial f_i} \\
&\quad +2\sum_{j\neq i}\phi_i(\phi_i-\phi_j)\frac{1-f_j}{f_i-f_j}+\frac{n-p-1}{1-f_i}(1-\phi_i)^2
\end{aligned}
$$

$$+4f_i(1-\phi_i)\frac{\partial \phi_i}{\partial f_i} + 2\sum_{j\neq i}(1-\phi_i)(\phi_i - \phi_j)\frac{f_j}{f_i - f_j}]\}.$$

PROOF. We observe that

$$\begin{aligned}
& R(\hat{\xi};\xi,\Sigma_1,\Sigma_2) \\
= \; & E(\hat{\xi}-\xi)'(\Sigma_1^{-1}+\Sigma_2^{-1})(\hat{\xi}-\xi) \\
= \; & E\mathrm{tr}\{2p + 2(X_1-\xi)'\Sigma_1^{-1}B^{-1}(I-\Phi)B(X_2-X_1) \\
& +\Sigma_1^{-1}[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]' \\
& +2(X_2-\xi)'\Sigma_2^{-1}B^{-1}\Phi B(X_1-X_2) \\
& +\Sigma_2^{-1}[B^{-1}\Phi B(X_1-X_2)][B^{-1}\Phi B(X_1-X_2)]'\}.
\end{aligned}$$

Since Φ satisfies the conditions of the Normal and Wishart identities, we have

$$\begin{aligned}
& R(\hat{\xi};\xi,\Sigma_1,\Sigma_2) \\
= \; & E\mathrm{tr}\{2p + 2\nabla^{(1)'}[B^{-1}(I-\Phi)B(X_2-X_1)] \\
& +(n-p-1)S_1^{-1}[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]' \\
& +2\tilde{\nabla}^{(1)}\{[B^{-1}(I-\Phi)B(X_2-X_1)][B^{-1}(I-\Phi)B(X_2-X_1)]'\} \\
& +2\nabla^{(2)'}[B^{-1}\Phi B(X_1-X_2)] \\
& +(n-p-1)S_2^{-1}[B^{-1}\Phi B(X_1-X_2)][B^{-1}\Phi B(X_1-X_2)]' \\
& +2\tilde{\nabla}^{(2)}\{[B^{-1}\Phi B(X_1-X_2)][B^{-1}\Phi B(X_1-X_2)]'\}\}.
\end{aligned}$$

Now it follows from Lemmas 1 and 2 that

$$\begin{aligned}
R(\hat{\xi};\xi,\Sigma_1,\Sigma_2) = \; & E\{\sum_i [B(X_1-X_2)]_i^2[\frac{n-p-1}{f_i}\phi_i^2 + 4(1-f_i)\phi_i\frac{\partial \phi_i}{\partial f_i} \\
& +2\sum_{j\neq i}\phi_i(\phi_i-\phi_j)\frac{1-f_j}{f_i-f_j} + \frac{n-p-1}{1-f_i}(1-\phi_i)^2 \\
& +4f_i(1-\phi_i)\frac{\partial \phi_i}{\partial f_i} + 2\sum_{j\neq i}(1-\phi_i)(\phi_i-\phi_j)\frac{f_j}{f_i-f_j}]\}.
\end{aligned}$$

This completes the proof. □

6 Generalized Least Squares Estimator

First suppose that the two covariance matrices Σ_1, Σ_2 are known. Then with respect to the loss function

$$L(\hat{\xi};\xi,\Sigma_1,\Sigma_2) = (\hat{\xi}-\xi)'(\Sigma_1^{-1}+\Sigma_2^{-1})(\hat{\xi}-\xi),$$

the best linear unbiased estimator $\hat{\xi}^{BE}$ for ξ is given by

$$\hat{\xi}^{BE} = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}(\Sigma_1^{-1}X_1 + \Sigma_2^{-1}X_2).$$

The following proposition is needed in the sequel.

Proposition 1 *With respect to the loss function L, the risk of $\hat{\xi}^{BE}$ is p.*

PROOF. The proof is straightforward and is omitted. □

However for the problem that we are concerned with in this paper, the covariance matrices Σ_1, Σ_2 are unknown. The usual practice would be to replace Σ_1, Σ_2 in $\hat{\xi}^{BE}$ by their maximum likelihood estimators S_1/n, S_2/n. This results in the usual estimator $\hat{\xi}^{LS}$ for ξ which is given by

$$\hat{\xi}^{LS} = (S_1^{-1} + S_2^{-1})^{-1}(S_1^{-1}X_1 + S_2^{-1}X_2).$$

Furthermore, we note that $\hat{\xi}^{LS}$ is the generalized least squares estimator in the following sense:

$$\hat{\xi}^{LS} = \min_{\hat{\xi}}{}^{-1} \sum_i (X_i - \hat{\xi})' S_i^{-1}(X_i - \hat{\xi}).$$

It is clear that this estimator is rather ad hoc and it is certainly plausible that by a more systematic treatment of the problem, better estimators can be obtained.

7 Alternative Estimators

It is well-known that the eigenvalues of $S_2(S_1 + S_2)^{-1}$ are more spread out than the eigenvalues of its expectation. By correcting for this eigenvalue distortion, we construct alternative estimators for ξ which compare favorably with the usual estimator. In particular, these estimators give substantial savings in risk when the eigenvalues of $\Sigma_2(\Sigma_1 + \Sigma_2)^{-1}$ are close together.

7.1 Stein-type Estimator

In this subsection, a method of Stein (1975), (1977a) is applied to obtain an alternative estimator $\hat{\xi}^{ST}$ for ξ. Let $\hat{\xi}$ be an equivariant estimator for ξ where

$$\hat{\xi}(X_1, X_2, S_1, S_2) = B^{-1}\Phi B X_1 + B^{-1}(I - \Phi)BX_2,$$

$\Phi = \Phi(F) = \mathrm{diag}(\phi_1, \ldots, \phi_p)$, $B(S_1 + S_2)B' = I$, $BS_2B' = F = \mathrm{diag}(f_1, \ldots, f_p)$ with $f_1 \geq \cdots \geq f_p$.

Lemma 3 *With the above notation, for $1 \le i \le p$ we have*

$$E^{S_1,S_2}[B(X_1 - X_2)]_i^2 = E^{S_1,S_2}[B(\Sigma_1 + \Sigma_2)B']_{ii},$$

where E^{S_1,S_2} denotes conditional expectation given S_1, S_2.

PROOF. We observe that

$$\begin{aligned} E^{S_1,S_2}[B(X_1 - X_2)]_i^2 &= E^{S_1,S_2} \sum_{j,k} b_{ij} b_{ik} (x_j^{(1)} - x_j^{(2)})(x_k^{(1)} - x_k^{(2)}) \\ &= E^{S_1,S_2} \sum_{j,k} b_{ij} b_{ik} (\Sigma_1 + \Sigma_2)_{jk} \\ &= E^{S_1,S_2}[B(\Sigma_1 + \Sigma_2)B']_{ii}. \end{aligned}$$

This completes the proof. □

For $i = 1, 2$, by replacing Σ_i in the above lemma with its maximum likelihood estimator S_i/n, we get the following approximation:

$$\begin{aligned} E^{S_1,S_2}[B(X_1 - X_2)]_i^2 &\approx E^{S_1,S_2}[B(S_1 + S_2)B']_{ii}/n \\ &= 1/n. \end{aligned} \tag{3}$$

Next it follows from Theorem 5 that the unbiased estimate of the risk of $\hat{\xi}$ is

$$\begin{aligned} \hat{R} = & \sum_i [B(X_1 - X_2)]_i^2 [\frac{n-p-1}{f_i} \phi_i^2 + 4(1 - f_i)\phi_i \frac{\partial \phi_i}{\partial f_i} \\ & + 2 \sum_{j \ne i} \phi_i (\phi_i - \phi_j) \frac{1 - f_j}{f_i - f_j} + \frac{n-p-1}{1 - f_i}(1 - \phi_i)^2 \\ & + 4 f_i (1 - \phi_i) \frac{\partial \phi_i}{\partial f_i} + 2 \sum_{j \ne i} (1 - \phi_i)(\phi_i - \phi_j) \frac{f_j}{f_i - f_j}]. \end{aligned}$$

From (3), we observe that $\hat{R}$ can be approximated by

$$\begin{aligned} \hat{R} \approx & \frac{1}{n} \sum_i [\frac{n-p-1}{f_i} \phi_i^2 + 4(1 - f_i)\phi_i \frac{\partial \phi_i}{\partial f_i} \\ & + 2 \sum_{j \ne i} \phi_i (\phi_i - \phi_j) \frac{1 - f_j}{f_i - f_j} + \frac{n-p-1}{1 - f_i}(1 - \phi_i)^2 \\ & + 4 f_i (1 - \phi_i) \frac{\partial \phi_i}{\partial f_i} + 2 \sum_{j \ne i} (1 - \phi_i)(\phi_i - \phi_j) \frac{f_j}{f_i - f_j}] \\ = & \tilde{R}, \text{ say.} \end{aligned} \tag{4}$$

By ignoring the derivative terms in $\tilde{R}$, we get

$$\begin{aligned}\check{R} &= \frac{1}{n}\sum_i[\frac{n-p-1}{f_i}\phi_i^2 + 2\sum_{j\neq i}\phi_i(\phi_i-\phi_j)\frac{1-f_j}{f_i-f_j} \\ &+\frac{n-p-1}{1-f_i}(1-\phi_i)^2 + 2\sum_{j\neq i}(1-\phi_i)(\phi_i-\phi_j)\frac{f_j}{f_i-f_j}].\end{aligned}$$

Now we minimize $\check{R}$ with respect to ϕ_i, $i = 1,\ldots,p$. This gives

$$\begin{aligned}\frac{\partial\phi_i}{\partial f_i} &= 0 \\ &= \frac{n-p-1}{f_i}\phi_i + 2\phi_i\sum_{j\neq i}\frac{1-f_j}{f_i-f_j} - \sum_{j\neq i}\phi_j \\ &-\frac{n-p-1}{1-f_i}(1-\phi_i) + 2(1-\phi_i)\sum_{j\neq i}\frac{f_j}{f_i-f_j} + \sum_{j\neq i}(1-\phi_j).\end{aligned}$$

For computational simplicity, we ignore the third and sixth terms in the last equation. We observe that these terms do not contribute significantly to the r.h.s. of the above equation. Solving for ϕ_i, we get

$$\phi_i = [\beta_i^{ST}/(1-f_i)]/\{[\alpha_i^{ST}/f_i] + [\beta_i^{ST}/(1-f_i)]\},$$

where

$$\begin{aligned}\alpha_i^{ST} &= n-p-1+2\sum_{j\neq i}\frac{f_i(1-f_j)}{f_i-f_j}, \\ \beta_i^{ST} &= n-p-1-2\sum_{j\neq i}\frac{(1-f_i)f_j}{f_i-f_j}.\end{aligned}$$

Unfortunately the natural ordering of the ϕ_i's may be altered. The natural ordering is given by $\phi_1 \geq \cdots \geq \phi_p \geq 0$. To correct for this, Stein's (1975) isotonic regression is applied to the α_i^{ST}/f_i's and the $\beta_i^{ST}/(1-f_i)$'s. This results in φ_i^{ST} and ψ_i^{ST}, $i = 1,\ldots,p$ respectively where $0 \leq \varphi_1^{ST} \leq \cdots \leq \varphi_p^{ST}$ and $0 \leq \psi_p^{ST} \leq \cdots \leq \psi_1^{ST}$. For a detailed description of Stein's isotonic regression, see for example Lin and Perlman (1985). Now we define

$$\hat{\xi}^{ST} = B^{-1}\Phi^{ST}BX_1 + B^{-1}(I-\Phi^{ST})BX_2,$$

with $\Phi^{ST} = \text{diag}(\phi_1^{ST},\ldots,\phi_p^{ST})$ and $\phi_i^{ST} = \psi_i^{ST}/(\varphi_i^{ST}+\psi_i^{ST})$, whenever $i = 1,\ldots p$. It is easy to see that in this case we have $\phi_1^{ST} \geq \cdots \geq \phi_p^{ST} \geq 0$.

7.2 Haff-type Estimator

Here we shall extend a method of Haff (1982) to derive an alternative estimator, denoted by $\hat{\xi}^{HF}$, for ξ. Again we consider estimators of the form

$$\hat{\xi} = B^{-1}\Phi BX_1 + B^{-1}(I-\Phi)BX_2,$$

where $\Phi(F) = \text{diag}(\phi_1,\ldots,\phi_p)$, $B(S_1+S_2)B' = I$, $BS_2B' = F = \text{diag}(f_1,\ldots,f_p)$ with $f_1 \geq \cdots \geq f_p$. Next we put a prior distribution on the parameter space $\{(\xi,\Sigma_1,\Sigma_2) : \xi \in R^p$ and Σ_1,Σ_2 being positive definite matrices.$\}$ and let $m(F)$ denote the marginal density of F. It follows from (4) that the unbiased estimate of the risk of $\hat{\xi}$ can be approximated by $\tilde{R}$. This implies that the approximate average risk of this estimator is

$$\int \hat{G}(f_1,\ldots,f_p;\phi_1,\ldots,\phi_p;\partial\phi_1/\partial f_1,\ldots,\partial\phi_p/\partial f_p)m(F)\mathrm{d}F,$$

where

$$\hat{G} = m\tilde{R}.$$

The solution of the Euler-Lagranges equations minimizes the above integral. These equations are given by

$$\hat{G}_{\phi_i} \;=\; \sum_j \frac{\partial}{\partial f_j}\hat{G}_{\frac{\partial\phi_i}{\partial f_j}}, \qquad \forall i = 1,\ldots p,$$

where $\hat{G}_{\phi_i} = \partial\hat{G}/\partial\phi_i$, etc. Evaluating the above equations gives for $i = 1,\ldots,p$,

$$\begin{aligned}
0 \;=\; & \frac{1}{f_i}[(n-p-1)\phi_i + 2\phi_i\sum_{j\neq i}\frac{f_i(1-f_j)}{f_i-f_j} + 2f_i\phi_i \\
& -f_i\sum_{j\neq i}\phi_j - 2f_i(1-f_i)\phi_i\frac{\partial\log m}{\partial f_i}] - \frac{1}{1-f_i}[(n-p-1)(1-\phi_i) \\
& -2(1-\phi_i)\sum_{j\neq i}\frac{(1-f_i)f_j}{f_i-f_j} + 2(1-f_i)(1-\phi_i) \\
& -(1-f_i)\sum_{j\neq i}(1-\phi_j) + 2f_i(1-f_i)(1-\phi_i)\frac{\partial\log m}{\partial f_i}].
\end{aligned}$$

As in the Stein-type estimator, we ignore the last term in each of the square brackets. This leaves us with the equation

$$0 \;=\; \frac{1}{f_i}[(n-p-1)\phi_i + 2\phi_i\sum_{j\neq i}\frac{f_i(1-f_j)}{f_i-f_j} + 2f_i\phi_i]$$

$$-\frac{1}{1-f_i}[(n-p-1)(1-\phi_i)-2(1-\phi_i)\sum_{j\neq i}\frac{(1-f_i)f_j}{f_i-f_j}$$
$$+2(1-f_i)(1-\phi_i)].$$

Solving for ϕ_i, we get

$$\phi_i=[\beta_i^{HF}/(1-f_i)]/\{[\alpha_i^{HF}/f_i]+[\beta_i^{HF}/(1-f_i)]\},$$

where

$$\begin{aligned}\alpha_i^{HF} &= n-p-1+2f_i+2\sum_{j\neq i}\frac{f_i(1-f_j)}{f_i-f_j},\\ \beta_i^{HF} &= n-p-1+2(1-f_i)-2\sum_{j\neq i}\frac{(1-f_i)f_j}{f_i-f_j}.\end{aligned}$$

Again, as in the Stein-type estimator, the natural ordering of the ϕ_i's may be changed. To correct for this, Stein's isotonic regression is applied to the α_i/f_i's and the $\beta_i/(1-f_i)$'s. This results in φ_i^{HF} and ψ_i^{HF}, $i=1,\ldots,p$ respectively where $0\leq\varphi_1^{HF}\leq\cdots\leq\varphi_p^{HF}$ and $0\leq\psi_p^{HF}\leq\cdots\leq\psi_1^{HF}$. Now we define

$$\hat{\xi}^{HF}=B^{-1}\Phi^{HF}BX_1+B^{-1}(I-\Phi^{HF})BX_2,$$

with $\Phi^{HF}=\mathrm{diag}(\phi_1^{HF},\ldots,\phi_p^{HF})$ and $\phi_i^{HF}=\psi_i^{HF}/(\varphi_i^{HF}+\psi_i^{HF})$, whenever $i=1,\ldots,p$. Finally it is easy to see that $\phi_1^{HF}\geq\cdots\geq\phi_p^{HF}\geq 0$.

7.3 Asymptotics

As a rough check to see that the approximations used in the previous two subsections are not too inaccurate, we observe the following special case. Let p be fixed and let n tend to infinity. Then for a fixed set of parameters Σ_1, Σ_2 we have

$$\begin{aligned}\alpha_i^{ST}&\sim n, & \beta_i^{ST}&\sim n,\\ \alpha_i^{HF}&\sim n, & \beta_i^{HF}&\sim n.\end{aligned}$$

Hence $\phi_i^{ST}\sim f_i$ and $\phi_i^{HF}\sim f_i$ for $1\leq i\leq p$. This implies that $\hat{\xi}^{ST}\sim\hat{\xi}^{LS}$ and $\hat{\xi}^{HF}\sim\hat{\xi}^{LS}$. This is a reassuring result since $\hat{\xi}^{LS}$ is asymptotically efficient under these conditions.

8 Monte Carlo Study

Due to the rather complicated nature of their construction, at this time we have not been able to give an analytical treatment of the risk performance of the estimators $\hat{\xi}^{ST}$ and $\hat{\xi}^{HF}$. We shall instead observe the risk behaviour of these estimators via a Monte Carlo study.

For the simulations, the following variance-reduction technique is used: First let

$$\hat{\xi} = B^{-1}\Phi BX_1 + B^{-1}(I-\Phi)BX_2,$$

where $\Phi(F) = \mathrm{diag}(\phi_1, \ldots, \phi_p)$, $B(S_1+S_2)B' = I$, $BS_2S' = F = \mathrm{diag}(f_1, \ldots, f_p)$ with $f_1 \geq \cdots \geq f_p$.

Proposition 2 *With the above notation,*

$$R(\hat{\xi}; \xi, \Sigma_1, \Sigma_2) = p + E(\hat{\xi} - \hat{\xi}^{BE})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \hat{\xi}^{BE}).$$

PROOF. We observe that

$$\begin{aligned} R(\hat{\xi}; \xi, \Sigma_1, \Sigma_2) &= E[(\hat{\xi} - \hat{\xi}^{BE})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \hat{\xi}^{BE}) \\ &\quad +(\hat{\xi}^{BE} - \xi)'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi}^{BE} - \xi) \\ &\quad +2(\hat{\xi} - \hat{\xi}^{BE})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi}^{BE} - \xi)] \\ &= p + E(\hat{\xi} - \hat{\xi}^{BE})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \hat{\xi}^{BE}). \end{aligned}$$

Since $E(\hat{\xi} - \hat{\xi}^{BE})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi}^{BE} - \xi) = 0$, the last equality follows from Proposition 1. □

Lemma 4 *With the above notation, let*

$$\begin{aligned} Y &= (\Sigma_1^{-1} + \Sigma_2^{-1})^{1/2}(\hat{\xi}^{BE} - \xi), \\ Z &= (\Sigma_1^{-1} + \Sigma_2^{-1})^{1/2}(\hat{\xi} - \hat{\xi}^{BE}). \end{aligned}$$

Then Y and Z are conditionally independent given S_1, S_2.

PROOF. It is straightforward to show that $E^{S_1,S_2}YZ' = 0$. Also we note that conditional on S_1 and S_2, both Y and Z follow a multivariate normal distribution. Hence we conclude that Y and Z are conditionally independent. □

Proposition 3 *With the above notation, let*

$$\begin{aligned} V &= (\hat{\xi} - \hat{\xi}^{BE})'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \hat{\xi}^{BE}), \\ W &= (\hat{\xi} - \xi)'(\Sigma_1^{-1} + \Sigma_2^{-1})(\hat{\xi} - \xi). \end{aligned}$$

Then $var[E(V \mid S_1, S_2)] \leq var(V) \leq var(W)$.

PROOF. From Lemma 4 and the proof of Proposition 2, we have

$$\begin{aligned} \mathrm{var}(W) &= E\mathrm{var}(W \mid S_1, S_2) + \mathrm{var}E(W \mid S_1, S_2) \\ &\geq E\mathrm{var}(V \mid S_1, S_2) + \mathrm{var}E(V \mid S_1, S_2) \\ &= \mathrm{var}(V) \\ &\geq \mathrm{var}E(V \mid S_1, S_2). \end{aligned}$$

This completes the proof. □

For the simulations, independent standard normal variates are generated by the IMSL subroutine DRNNOA and the eigenvalue decomposition uses the IMSL subroutine DEVCSF. Also we take $p = 5,\ n = 7,\ 15,\ 30$ and $p = 10,\ n = 12,\ 25,\ 50$. Since Propositions 2 and 3 show that $\mathrm{var}E(V \mid S_1, S_2) \leq \mathrm{var}(W)$ and $E(W) = p + E(V)$, a simulation is done to estimate $E(V)$; this will permit the estimation of $R(\hat{\xi}; \xi, \Sigma_1, \Sigma_2)$. We do this by computing the mean $\bar{V}$ of $E(V \mid S_1, S_2)$ based on 500 independent replications. Then we calculate the average loss of the estimator $\hat{\xi}$ by the formula $\bar{L} = p + \bar{V}$. Tables 1 to 6 give the average losses and their standard deviations of the estimators $\hat{\xi}^{LS}$, $\hat{\xi}^{ST}$ and $\hat{\xi}^{HF}$. The risk of $\hat{\xi}^{BE}$ is also shown in the tables for convenience. This serves as a lower bound on the risks of these estimators.

Compared with the naive Monte Carlo, this variance-reduction technique on the average reduces the estimated standard deviations here by a factor of roughly around 10.

We shall now summarize the results of this numerical study:

1. The risks of the estimators $\hat{\xi}^{ST}$ and $\hat{\xi}^{HF}$ compare very favorably with that of $\hat{\xi}^{LS}$. In particular when p and n are of comparable magnitude, significant savings in risk are achieved in most parts of the parameter space. This is most evident when the eigenvalues of $(\Sigma_1 + \Sigma_2)^{-1}\Sigma_2$ are close together. For example, in the case of $p = 10$, $n = 12$ and the eigenvalues of $\Sigma_2\Sigma_1^{-1}$ being all equal to 1, approximately 40% savings in risk is achieved with the use of $\hat{\xi}^{ST}$ or $\hat{\xi}^{HF}$ over that of $\hat{\xi}^{LS}$.

2. However, when the eigenvalues of $(\Sigma_1 + \Sigma_2)^{-1}\Sigma_2$ are far apart, there does not appear to be any significant difference in risk among the three estimators: $\hat{\xi}^{LS}$, $\hat{\xi}^{ST}$ and $\hat{\xi}^{HF}$.

3. For a fixed set of parameters $(\xi, \Sigma_1, \Sigma_2)$, the study also shows that the savings in risk of $\hat{\xi}^{ST}$ and $\hat{\xi}^{HF}$ over $\hat{\xi}^{LS}$ increases with p and decreases with n.

Also we note that in our simulation, for a fixed set of eigenvalues of $\Sigma_2\Sigma_1^{-1}$, the estimators are computed from the same set of 500 independently generated samples. This suggests that there is a high correlation among the average losses of these estimators. Since we are more interested in the relative risk ordering of these estimators, we conclude that the estimated standard deviation (as given in Tables 1 to 6) is probably a pessimistic indicator of the variability of the relative magnitude of the average losses.

9 Final Remarks

The main obstacle in using the unbiased estimate of risk to get good estimators is the fact that risk is a 'smooth version' of the unbiased estimate of risk and hence the unbiased estimate of risk does not reflect exactly the behavior of the risk. This is self-evident since we need to integrate the unbiased estimate of risk to get to the risk. Thus except for special cases, proving theoretical dominance over the usual estimator is generally not possible with this method, assuming of course that the usual estimator is inadmissible. However as this paper indicates, the unbiased estimate of risk does possess a good deal of useful information. If this is exploited carefully, one can obtain alternative estimators which beat the usual estimator over most parts of the parameter space.

10 Acknowledgments

This paper forms a part of my Ph.D. dissertation at Stanford University. I would like to express my deep gratitude to my advisor, Professor Charles Stein, whose constant support, encouragement and guidance made this dissertation possible. Special thanks go to Professors Ingram Olkin and Iain Johnstone for reading my dissertation and giving me several enlightening suggestions. Finally I would like to thank Professors Art Cohen and L. R. Haff for alerting me to relevant literature which I have overlooked and for sending me copies of their work.

TABLE 1

$p = 5 \qquad n = 7$

Average losses of estimators for the common mean
(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\xi}^{BE}$	$\hat{\xi}^{LS}$	$\hat{\xi}^{ST}$	$\hat{\xi}^{HF}$
(1,1,1,1,1)	5.00	8.13	5.46	5.38
		(0.05)	(0.02)	(0.02)
(10,0.1,0.1,0.1,0.1)	5.00	8.96	7.77	7.88
		(0.13)	(0.08)	(0.07)
(10,10,10,0.1,0.1)	5.00	8.56	8.24	8.33
		(0.09)	(0.08)	(0.07)
(10,1,1,1,0.1)	5.00	8.24	7.46	7.44
		(0.07)	(0.04)	(0.04)
(10,10,1,0.1,0.1)	5.00	8.37	8.18	8.23
		(0.07)	(0.06)	(0.06)
(20,5,1,0.5,0.05)	5.00	8.30	8.33	8.41
		(0.07)	(0.07)	(0.07)
(10^{10},5,1,0.5,10^{-10})	5.00	8.30	7.93	7.88
		(0.07)	(0.06)	(0.06)
(5,2,1,0.5,0.2)	5.00	8.23	6.70	6.63
		(0.06)	(0.03)	(0.02)
(10^{10},10^{-10},10^{-10},10^{-10},10^{-10})	5.00	7.54	7.54	7.54
		(0.11)	(0.11)	(0.11)

TABLE 2

$p = 5 \qquad n = 15$

Average losses of estimators for the common mean
(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\xi}^{BE}$	$\hat{\xi}^{LS}$	$\hat{\xi}^{ST}$	$\hat{\xi}^{HF}$
(1,1,1,1,1)	5.00	6.11	5.25	5.23
		(0.02)	(0.01)	(0.01)
(10,0.1,0.1,0.1,0.1)	5.00	6.03	5.77	5.81
		(0.02)	(0.02)	(0.02)
(10,10,10,0.1,0.1)	5.00	6.11	6.01	6.04
		(0.02)	(0.02)	(0.02)
(10,1,1,1,0.1)	5.00	6.13	5.94	5.93
		(0.02)	(0.02)	(0.02)
(10,10,1,0.1,0.1)	5.00	6.13	6.11	6.13
		(0.02)	(0.02)	(0.02)
(20,5,1,0.5,0.05)	5.00	6.13	6.12	6.12
		(0.02)	(0.02)	(0.02)
$(10^{10},5,1,0.5,10^{-10})$	5.00	6.12	6.07	6.06
		(0.02)	(0.02)	(0.02)
(5,2,1,0.5,0.2)	5.00	6.13	5.89	5.88
		(0.02)	(0.01)	(0.01)
$(10^{10},10^{-10},10^{-10},10^{-10},10^{-10})$	5.00	5.70	5.70	5.70
		(0.02)	(0.02)	(0.02)

TABLE 3

$p = 5 \qquad n = 30$

Average losses of estimators for the common mean
(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\xi}^{BE}$	$\hat{\xi}^{LS}$	$\hat{\xi}^{ST}$	$\hat{\xi}^{HF}$
(1,1,1,1,1)	5.00	5.53	5.14	5.13
		(0.01)	(0.01)	(0.01)
(10,0.1,0.1,0.1,0.1)	5.00	5.42	5.33	5.34
		(0.01)	(0.01)	(0.01)
(10,10,10,0.1,0.1)	5.00	5.50	5.45	5.46
		(0.01)	(0.01)	(0.01)
(10,1,1,1,0.1)	5.00	5.53	5.43	5.42
		(0.01)	(0.01)	(0.01)
(10,10,1,0.1,0.1)	5.00	5.51	5.50	5.51
		(0.01)	(0.01)	(0.01)
(20,5,1,0.5,0.05)	5.00	5.52	5.52	5.52
		(0.01)	(0.01)	(0.01)
(10^{10},5,1,0.5,10^{-10})	5.00	5.51	5.50	5.50
		(0.01)	(0.01)	(0.01)
(5,2,1,0.5,0.2)	5.00	5.53	5.49	5.48
		(0.01)	(0.01)	(0.01)
(10^{10},10^{-10},10^{-10},10^{-10},10^{-10})	5.00	5.30	5.30	5.30
		(0.01)	(0.01)	(0.01)

TABLE 4

$p = 10 \qquad n = 12$

Average losses of estimators for the common mean
(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\xi}^{BE}$	$\hat{\xi}^{LS}$	$\hat{\xi}^{ST}$	$\hat{\xi}^{HF}$
(1,1,1,1,1,	10.00	17.86	10.69	10.62
1,1,1,1,1)		(0.08)	(0.03)	(0.03)
(10,0.1,0.1,0.1,0.1,	10.00	22.89	14.81	14.97
0.1,0.1,0.1,0.1,0.1)		(0.23)	(0.10)	(0.10)
(10,10,10,0.1,0.1,	10.00	19.37	16.32	16.36
0.1,0.1,0.1,0.1,0.1)		(0.15)	(0.10)	(0.09)
(10,10,10,10,10,	10.00	18.15	16.42	16.43
0.1,0.1,0.1,0.1,0.1)		(0.11)	(0.08)	(0.08)
(10,10,10,10,10,	10.00	19.90	15.49	15.58
10,10,10,0.1,0.1)		(0.16)	(0.08)	(0.08)
(10,9/2,8/3,7/4,6/5,	10.00	18.03	14.24	14.19
5/6,4/7,3/8,2/9,1/10)		(0.09)	(0.04)	(0.04)
(10,10,10,1,1,	10.00	18.12	15.69	15.67
1,1,0.1,0.1,0.1)		(0.10)	(0.06)	(0.06)
$(10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10},$	10.00	15.12	15.12	15.12
$10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10})$		(0.36)	(0.36)	(0.36)
$(10^{10}, 10^{10}, 10^{10}, 10^{10}, 10^{10},$	10.00	18.18	18.18	18.18
$10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10})$		(0.14)	(0.14)	(0.14)

TABLE 5

$p = 10 \qquad n = 25$

Average losses of estimators for the common mean
(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\xi}^{BE}$	$\hat{\xi}^{LS}$	$\hat{\xi}^{ST}$	$\hat{\xi}^{HF}$
(1,1,1,1,1,	10.00	12.68	10.34	10.32
1,1,1,1,1)		(0.02)	(0.01)	(0.01)
(10,0.1,0.1,0.1,0.1,	10.00	12.33	11.16	11.20
0.1,0.1,0.1,0.1,0.1)		(0.03)	(0.02)	(0.02)
(10,10,10,0.1,0.1,	10.00	12.56	12.03	12.04
0.1,0.1,0.1,0.1,0.1)		(0.03)	(0.02)	(0.02)
(10,10,10,10,10,	10.00	12.66	12.31	12.32
0.1,0.1,0.1,0.1,0.1)		(0.02)	(0.02)	(0.02)
(10,10,10,10,10,	10.00	12.47	11.68	11.71
10,10,10,0.1,0.1)		(0.03)	(0.02)	(0.02)
(10,9/2,8/3,7/4,6/5,	10.00	12.69	12.06	12.05
5/6,4/7,3/8,2/9,1/10)		(0.02)	(0.02)	(0.02)
(10,10,10,1,1,	10.00	12.69	12.29	12.28
1,1,0.1,0.1,0.1)		(0.02)	(0.02)	(0.02)
$(10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10},$	10.00	10.97	10.97	10.97
$10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10})$		(0.02)	(0.02)	(0.02)
$(10^{10},10^{10},10^{10},10^{10},10^{10},$	10.00	12.60	12.60	12.60
$10^{-10},10^{-10},10^{-10},10^{-10},10^{-10})$		(0.03)	(0.03)	(0.03)

TABLE 6

$p = 10 \qquad n = 50$

Average losses of estimators for the common mean

(Estimated standard errors are in parenthesis)

Eigenvalues of $\Sigma_2\Sigma_1^{-1}$	$\hat{\xi}^{BE}$	$\hat{\xi}^{LS}$	$\hat{\xi}^{ST}$	$\hat{\xi}^{HF}$
(1,1,1,1,1,	10.00	11.18	10.14	10.14
1,1,1,1,1)		(0.01)	(0.01)	(0.01)
(10,0.1,0.1,0.1,0.1,	10.00	10.81	10.47	10.48
0.1,0.1,0.1,0.1,0.1)		(0.01)	(0.01)	(0.01)
(10,10,10,0.1,0.1,	10.00	11.06	10.88	10.88
0.1,0.1,0.1,0.1,0.1)		(0.01)	(0.01)	(0.01)
(10,10,10,10,10,	10.00	11.15	11.01	11.01
0.1,0.1,0.1,0.1,0.1)		(0.01)	(0.01)	(0.01)
(10,10,10,10,10,	10.00	10.96	10.71	10.72
10,10,10,0.1,0.1)		(0.01)	(0.01)	(0.01)
(10,9/2,8/3,7/4,6/5,	10.00	11.17	11.04	11.04
5/6,4/7,3/8,2/9,1/10)		(0.01)	(0.01)	(0.01)
(10,10,10,1,1,	10.00	11.16	10.99	10.99
1,1,0.1,0.1,0.1)		(0.01)	(0.01)	(0.01)
$(10^{10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10},$	10.00	10.41	10.41	10.41
$10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10})$		(0.01)	(0.01)	(0.01)
$(10^{10}, 10^{10}, 10^{10}, 10^{10}, 10^{10},$	10.00	11.12	11.12	11.12
$10^{-10}, 10^{-10}, 10^{-10}, 10^{-10}, 10^{-10})$		(0.01)	(0.01)	(0.01)

References

[1] BROWN, L. D. and COHEN, A. (1974). Point and confidence estimation of a common mean and recovery of interblock information. *Ann. Statist.* **2** 963-976.

[2] CHIOU, W. and COHEN, A. (1985). On estimating a common multivariate normal mean vector. *Ann. Inst. Statist. Math.* **37** 499-506.

[3] COHEN, A. and SACKROWITZ, H. B. (1974). On estimating the common mean of two normal distributions. *Ann. Statist.* **2** 1274-1282.

[4] GRAYBILL, F. A. and DEAL, R. B. (1959). Combining unbiased estimators. *Biometrics* **15** 543-550.

[5] HAFF, L. R. (1977). Minimax estimators for a multinormal precision matrix. *J. Multivariate Anal.* **7** 374-385.

[6] HAFF, L. R. (1982). Solutions of the Euler-Lagrange equations for certain multivariate normal estimation problems. Unpublished manuscript.

[7] HAFF, L. R. (1988). The variational form of certain Bayes estimators. Unpublished manuscript.

[8] JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1**, 361-380. Univ. of California Press.

[9] KHATRI, C. G. and SHAH, K. R. (1974). Estimation of location parameters from two linear models under normality. *Comm. Statist.B* **3** 647-665.

[10] LIN, S. P. and PERLMAN, M. D. (1985). A Monte Carlo comparision of four estimators for a covariance matrix. *Multivariate Analy.* **6**, 411-429, ed. P. R. Krishnaiah, North Holland, Amsterdam.

[11] LOH, W. L. (1988). Estimating covariance matrices I. Technical Report No. 88-37, Department of Statistics, Purdue University.

[12] SHINOZAKI, N. (1978). A note on estimating the common mean of k normal distributions and the Stein problem. *Comm. Statist. A–Theory Methods* **7** 1421-1432.

[13] STEIN, C. (1973). Estimation of the mean of a multivariate normal distribution. *Proc. Prague Symp. Asymptotic Statist.* 345-381.

[14] STEIN, C. (1975). Rietz lecture. 39th annual meeting IMS. Atlanta, Georgia.

[15] STEIN, C. (1977a). Unpublished notes on estimating the covariance matrix.

[16] STEIN, C. (1977b). Lectures on the theory of estimation of many parameters. (In Russian.) *Studies in the Statistical Theory of Estimation, Part I* (Ibragimov, I. A. and Nikulin, M. S., eds.), *Proceedings of Scientific Seminars of the Steklov Institute, Leningrad Division* **74**, 4-65.

[17] STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135-1151.