THE USE OF PRIOR INFORMATION
IN LINEAR REGRESSION

by

PUI MAN VICTOR YU, B.S., M.S.

A DISSERTATION

IN

MATHEMATICS

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

Approved

A
801
T3
1976
NO. 55
cop. 2

ACKNOWLEDGMENTS

I would like to express my gratitude to Professor Truman O. Lewis for his direction of this dissertation and to the other members of my committee, Professors Thomas L. Boullion, Benjamin S. Duran, Louis R. Hunt and Terry A. Watkins, for their helpful criticism.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
LIST OF TABLES	iv
I. INTRODUCTION	1
II. MATRIX DIFFERENTIATION	3
III. ESTIMATION WITH PRIOR INFORMATION	8
IV. PROBABILITY CONTENT OF USING PRIOR INFORMATION .	15
V. INTERVAL ON PROBABILITY CONTENT	41
VI. OPTIMAL DESIGN PROCEDURES	57
VII. SIMULATION RESULTS AND SUMMARY	77
LIST OF REFERENCES	82

LIST OF TABLES

Table	Page
6.1 Simulation results for five $\lambda_{\min}$'s	78

CHAPTER I

INTRODUCTION

1.1 Introduction

In this dissertation we consider the linear model

$$Y = X\beta + e$$

where β is a random variable. The statistical problem is to obtain an estimate of the value of β in a given experiment. When prior information is available, for example, an estimate of the value of β in previous experiments has already been computed and we want to utilize this information in addition to current experiments to obtain an improved estimate of the present value of β .

In chapter II, some matrix differentiation techniques are developed in order to derive the results in the following chapters. In chapter III, a method will be derived to combine two sets of data to obtain an estimate of the value of β in one of the experiments. This estimator is biased. Its mean square error is compared with the covariance matrix of the standard estimator.

In chapter IV, the probability that the use of prior information is beneficial is obtained. In chapter V, both upper and lower bounds for this probability are found. Monotonic properties of the bounds are also discussed.

In chapter VI, optimal design procedures are presented. In chapter VII, simulation results and a summary are given.

.....

CHAPTER II

MATRIX DIFFERENTIATION

2.1 Introduction

In this chapter we will present some matrix differentiation techniques. These results are in essence given in Dwyer [7], and Dwyer and MacPhail [6]. These results are presented for completeness and can be verified easily, hence proofs are omitted.

2.2 Matrix Differentiation

Definition Let f be a scalar function of the $m \times n$ matrix X and assume that each partial derivative $\partial f / \partial x_{ij}$ exists. Then the derivative of f with respect to the matrix X is denoted by $\partial f / \partial X$ and defined by

$$\frac{\partial f}{\partial X} = \left[\frac{\partial f}{\partial x_{ij}} \right] = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix} .$$

Notation Let X be a $m \times n$ matrix. The ij -th element of X is denoted by x_{ij} or $\langle X \rangle_{ij}$.

Let A and X be $n \times n$ square matrices with ij -th element a_{ij} and x_{ij} , respectively. We have the following

theorems.

Theorem 2.1

$$\frac{\partial}{\partial X} \langle XA \rangle_{ij} = \begin{bmatrix} & & \phi & & \\ & & & & \\ a_{1j} & a_{2j} & \dots & a_{nj} & \\ & & & & \\ & & \phi & & \end{bmatrix} \leftarrow i\text{-th row.}$$

Theorem 2.2 If X is a symmetric matrix, then

$$\frac{\partial}{\partial X} \langle XA \rangle_{ij} = \begin{bmatrix} & & & a_{1j} & & \\ & & & a_{2j} & \phi & \\ & \phi & & \vdots & & \\ & & & \vdots & & \\ a_{1j} & a_{2j} & \dots & a_{ij} & \dots & a_{nj} \\ & & & \vdots & & \\ & \phi & & a_{nj} & \phi & \end{bmatrix} \leftarrow i\text{-th row.}$$

$\uparrow$
 i-th column

Notation With f a scalar function of a matrix Y, which is itself some function of a matrix X, we write, subject to the existence of all derivatives

$$\frac{\partial f}{\partial x_{ij}} = \sum_k \sum_l \frac{\partial f}{\partial y_{kl}} \cdot \frac{\partial y_{kl}}{\partial x_{ij}}.$$

In matrix notation, this is written as

$$\frac{\partial f}{\partial X} = \sum_k \sum_l \left\langle \frac{\partial f}{\partial Y} \right\rangle_{kl} \frac{\partial}{\partial X} \langle Y \rangle_{kl}.$$

Some special cases found in [7] are:

Theorem 2.3

$$\frac{\partial}{\partial X} |Y| = \sum_i \sum_j |Y| \langle Y^{-t} \rangle_{ij} \frac{\partial}{\partial X} \langle Y \rangle_{ij}$$

where Y^{-t} denotes the transpose of the inverse of Y .

Theorem 2.4

$$\frac{\partial}{\partial X} \ln |Y| = \sum_i \sum_j \langle Y^{-t} \rangle_{ij} \frac{\partial}{\partial X} \langle Y \rangle_{ij}.$$

These results enable us to show the following theorems.

Theorem 2.5 If A is a symmetric square matrix, then

$$\frac{\partial}{\partial X} \ln |I + XA| = (I + AX')^{-1}A$$

where X' denotes the transpose of X .

Theorem 2.6 If A and X are symmetric matrices, then

$$\frac{\partial}{\partial X} \ln |I + XA| = 2(I + AX)^{-1}A - \text{diag } (I + AX)^{-1}A$$

where $\text{diag } (I + AX)^{-1}A$ denotes the diagonal matrix with diagonal elements $\langle (I+AX)^{-1}A \rangle_{ii}$.

Theorem 2.7 If A is a symmetric matrix, then

$$\frac{\partial}{\partial X} \langle XAX \rangle_{ij} = \begin{bmatrix} \phi \\ \sum_m a_{lm} x_{mj} \dots \sum_m a_{nm} x_{mj} \\ \phi \end{bmatrix} \leftarrow i\text{-th row} \\ + \begin{bmatrix} \sum_k x_{ik} a_{kl} \\ \phi \dots \phi \\ \sum_k x_{ik} a_{kn} \end{bmatrix} \cdot \\ \uparrow \\ j\text{-th column}$$

Theorem 2.8 If A and X are symmetric matrices, then

$$\frac{\partial}{\partial X} \langle XAX \rangle_{ij} = \begin{bmatrix} \phi & \sum_m a_{lm} x_{mj} & \phi \\ \sum_m a_{lm} x_{mj} & \dots & \sum_m a_{nm} x_{mj} \\ \phi & \sum_m a_{nm} x_{mj} & \phi \end{bmatrix} \leftarrow i\text{-th row} \\ \uparrow \\ i\text{-th column} \\ + \begin{bmatrix} \sum_k x_{ik} a_{kl} & \phi & \sum_k x_{ik} a_{kn} \\ \phi & \sum_k x_{ik} a_{kl} & \phi \\ \phi & \sum_k x_{ik} a_{kn} & \phi \end{bmatrix} \leftarrow j\text{-th row} \cdot \\ \uparrow \\ j\text{-th column}$$

Theorem 2.9 If A , C and X are symmetric matrices, then

$$\begin{aligned} & \frac{\partial}{\partial X} \ln |C + XAX| \\ &= 2[(C+XAX)^{-1}XA + AX(C+XAX)^{-1} - \text{diag } (C+XAX)^{-1}XA]. \end{aligned}$$

Theorem 2.10 If A , C and X are symmetric matrices, then

$$\begin{aligned} & \frac{\partial}{\partial X} \ln |C + X + XAX| \\ &= 2[(C+X+XAX)^{-1}XA + AX(C+X+XAX)^{-1} - \text{diag } (C+X+XAX)^{-1}XA] \\ & \quad + 2(C+X+XAX)^{-1} - \text{diag } (C+X+XAX)^{-1}. \end{aligned}$$

CHAPTER III

ESTIMATION WITH PRIOR INFORMATION

3.1 The Model

Consider the linear models

$$Y_i = X_i \beta_i + e_i, \quad i = 1, 2 \quad (3.1)$$

where Y_i is an $n_i \times 1$ random vector of observations, X_i is an $n_i \times r$ matrix of known fixed quantities and β_i is an $r \times 1$ random vector with mean $\bar{\beta}$ and covariance matrix V , and e_i is an $n_i \times 1$ random vector of unknown errors with zero mean and covariance matrix Σ_i . Also assume that β_i and e_i are independent. The matrix X_i is called the design matrix and is full column rank unless otherwise indicated.

The statistical problem is to obtain an estimate of the value of β_2 .

Consider the situation where we have some prior information, say, a set of observations from previous experiments. Our goal is to incorporate Y_1 and Y_2 to obtain an estimator of β_2 which has smaller mean square error than the standard estimator of β_2 based on Y_2 only.

It is well-known [10] from the Gauss-Markoff theorem that the best (minimum variance) linear unbiased estimate of β_i denoted by $\hat{\beta}_i$ is given by

$$\hat{\beta}_i = (X_i' \Sigma_i^{-1} X_i)^{-1} X_i' \Sigma_i^{-1} Y_i. \quad (3.2)$$

The covariance matrix V_i of this estimator is

$$V_i = \text{cov}(\hat{\beta}_i) = (X_i' \Sigma_i^{-1} X_i)^{-1}. \quad (3.3)$$

Our problem is to combine data Y_1 and data Y_2 to obtain a better estimate of β_2 . The combined model is assumed to be

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta_2 + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}. \quad (3.4)$$

Assuming that e_1 and e_2 are independent, we obtain the covariance matrix of $\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$ as

$$\text{cov} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}. \quad (3.5)$$

The new estimate of β_2 in the combined model given in equation (3.4) can be found by the Gauss-Markoff theorem. Denoting this new estimator by $\hat{\hat{\beta}}_2$, we have

$$\begin{aligned} \hat{\hat{\beta}}_2 &= \left\{ \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}' \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \right\}^{-1} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}' \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix}^{-1} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \\ &= (X_1' \Sigma_1^{-1} X_1 + X_2' \Sigma_2^{-1} X_2)^{-1} (X_1' \Sigma_1^{-1} Y_1 + X_2' \Sigma_2^{-1} Y_2). \end{aligned} \quad (3.6)$$

By equations (3.2) and (3.3), equation (3.6) can be written as

$$\hat{\hat{\beta}}_2 = (V_1^{-1} + V_2^{-1})^{-1} (V_1^{-1} \hat{\beta}_1 + V_2^{-1} \hat{\beta}_2). \quad (3.7)$$

Note that this estimator is biased.

3.2 Comparison between the Estimators

We now have two estimators for β_2 , namely $\hat{\beta}_2$ and $\hat{\hat{\beta}}_2$. $\hat{\beta}_2$ is the estimator using Y_2 only while $\hat{\hat{\beta}}_2$ is the estimator using both Y_1 and Y_2 . The question now arises: 'Which estimator is better?'.

In order to answer this question, we will set up a criterion. There are two cases that need to be considered.

Case I. Prior data is taken.

In this case Y_1 have been observed. Since $\hat{\hat{\beta}}_2$ is biased for a given experiment, we obtain the mean square error as:

$$\begin{aligned} & \text{MSE}(\hat{\hat{\beta}}_2 | \beta_2, Y_1) \\ &= E[(\hat{\hat{\beta}}_2 - \beta_2)(\hat{\hat{\beta}}_2 - \beta_2)' | \beta_2, Y_1] \\ &= E\{[(V_1^{-1} + V_2^{-1})^{-1}(V_1^{-1}\hat{\beta}_1 + V_2^{-1}\hat{\beta}_2) - \beta_2] \\ & \quad \cdot [(V_1^{-1} + V_2^{-1})^{-1}(V_1^{-1}\hat{\beta}_1 + V_2^{-1}\hat{\beta}_2) - \beta_2]' | \beta_2, Y_1\} \end{aligned}$$

$$\begin{aligned}
&= (V_1^{-1} + V_2^{-1})^{-1} E\{ [V_1^{-1}(\hat{\beta}_1 - \beta_2) + V_2^{-1}(\hat{\beta}_2 - \beta_2)] \\
&\quad \cdot [(\hat{\beta}_1 - \beta_2)' V_1^{-1} + (\hat{\beta}_2 - \beta_2)' V_2^{-1}] | \beta_2, Y_1 \} (V_1^{-1} + V_2^{-1})^{-1} \\
&= (V_1^{-1} + V_2^{-1})^{-1} [V_1^{-1}(\hat{\beta}_1 - \beta_2)(\hat{\beta}_1 - \beta_2)' V_1^{-1} + V_2^{-1}(\hat{\beta}_2 - \beta_2)(\hat{\beta}_2 - \beta_2)' V_2^{-1}] (V_1^{-1} + V_2^{-1})^{-1} \\
&= (I + V_1 V_2^{-1})^{-1} [(\hat{\beta}_1 - \beta_2)(\hat{\beta}_1 - \beta_2)' + V_1 V_2^{-1} V_1] (I + V_2^{-1} V_1)^{-1},
\end{aligned}
\tag{3.8}$$

where we have used the facts that

$$E[(\hat{\beta}_2 - \beta_2) | \beta_2] = 0,$$

and $E[(\hat{\beta}_2 - \beta_2)(\hat{\beta}_2 - \beta_2)' | \beta_2] = V_2.$

We have also assumed that the value of β varies from experiment to experiment with mean $\bar{\beta}$ and covariance matrix V . In this case the average of the mean square error over the ensemble of experiments produces

$$\begin{aligned}
&\text{MSE}(\hat{\beta}_2 | Y_1) \\
&= E[\text{MSE}(\hat{\beta}_2 | \beta_2, Y_1) | Y_1] \\
&= E\{ (I + V_1 V_2^{-1})^{-1} [(\hat{\beta}_1 - \beta_2)(\hat{\beta}_1 - \beta_2)' + V_1 V_2^{-1} V_1] (I + V_2^{-1} V_1)^{-1} | Y_1 \} \\
&= (I + V_1 V_2^{-1})^{-1} \{ E[(\hat{\beta}_1 - \beta_2)(\hat{\beta}_1 - \beta_2)' | Y_1] + V_1 V_2^{-1} V_1 \} (I + V_2^{-1} V_1)^{-1} \\
&= (I + V_1 V_2^{-1})^{-1} \{ E[(\hat{\beta}_1 - \bar{\beta}) + (\bar{\beta} - \beta_2)] [(\hat{\beta}_1 - \bar{\beta}) + (\bar{\beta} - \beta_2)]' | Y_1 \\
&\quad + V_1 V_2^{-1} V_1 \} (I + V_2^{-1} V_1)^{-1}
\end{aligned}$$

$$= (I+V_1V_2^{-1})^{-1}[(\hat{\beta}_1-\bar{\beta})(\hat{\beta}_1-\bar{\beta})' + V + V_1V_2^{-1}V_1](I+V_2^{-1}V_1)^{-1}, \quad (3.9)$$

where we have used the facts that

$$E(\beta_2 - \bar{\beta}) = 0,$$

and
$$E[(\beta_2 - \bar{\beta})(\beta_2 - \bar{\beta})'] = V.$$

The difference between V_2 and $MSE(\hat{\beta}_2|Y_1)$ is

$$\begin{aligned} & V_2 - MSE(\hat{\beta}_2|Y_1) \\ &= (I+V_1V_2^{-1})^{-1}[2V_1+V_2-V-(\hat{\beta}_1-\bar{\beta})(\hat{\beta}_1-\bar{\beta})'](I+V_2^{-1}V_1)^{-1}. \end{aligned} \quad (3.10)$$

The criterion is: if $V_2 - MSE(\hat{\beta}_2|Y_1)$ is positive definite, the use of prior information is beneficial. If $V_2 - MSE(\hat{\beta}_2|Y_1)$ is negative definite, the use of prior information is detrimental.

Case II. Prior data has not been taken.

An interesting situation presents itself when trial runs are allowed before the final experiment. In this case, prior data Y_1 have not been observed. The experimenter can design his prior experiments to observe Y_1 so that the use of prior information is always beneficial.

In this case, the mean square error is

$$MSE(\hat{\beta}_2) = E[MSE(\hat{\beta}_2|Y_1)]. \quad (3.11)$$

From equation (3.9), we have

$$\begin{aligned}
& E\{[MSE(\hat{\beta}_2|Y_1)]|\beta_1\} \\
&= (I+V_1V_2^{-1})^{-1}\{E[(\hat{\beta}_1-\bar{\beta})(\hat{\beta}_1-\bar{\beta})'|\beta_1]+V+V_1V_2^{-1}V_1\}(I+V_2^{-1}V_1)^{-1} \\
&= (I+V_1V_2^{-1})^{-1}\{E[(\hat{\beta}_1-\beta_1)+(\beta_1-\bar{\beta})][(\hat{\beta}_1-\beta_1)+(\beta_1-\bar{\beta})']|\beta_1 \\
&\quad +V+V_1V_2^{-1}V_1\}(I+V_2^{-1}V_1)^{-1} \\
&= (I+V_1V_2^{-1})^{-1}[V_1+(\beta_1-\bar{\beta})(\beta_1-\bar{\beta})'+V+V_1V_2^{-1}V_1](I+V_2^{-1}V_1)^{-1}
\end{aligned} \tag{3.12}$$

where we have used the facts that

$$E[(\hat{\beta}_1-\beta_1)|\beta_1] = 0,$$

and $E[(\hat{\beta}_1-\beta_1)(\hat{\beta}_1-\beta_1)'|\beta_1] = V_1.$

Taking expectation with respect to β_1 in equation (3.12), equation (3.11) becomes

$$\begin{aligned}
& MSE(\hat{\beta}_2) \\
&= (I+V_1V_2^{-1})^{-1}(V_1+2V+V_1V_2^{-1}V_1)(I+V_2^{-1}V_1)^{-1},
\end{aligned} \tag{3.13}$$

since $E[(\beta_1-\bar{\beta})(\beta_1-\bar{\beta})'] = V.$

Hence, the difference between V_2 and $MSE(\hat{\beta}_2)$ in this case is

$$\begin{aligned}
& V_2 - \text{MSE}(\hat{\beta}_2) \\
&= (I + V_1 V_2^{-1})^{-1} (V_1 + V_2 - 2V) (I + V_2^{-1} V_1)^{-1}. \quad (3.14)
\end{aligned}$$

As before, the criterion is: if $V_2 - \text{MSE}(\hat{\beta}_2)$ is positive definite, the use of prior information is beneficial. If $V_2 - \text{MSE}(\hat{\beta}_2)$ is negative definite, the use of prior information is detrimental.

CHAPTER IV

PROBABILITY CONTENT OF USING PRIOR INFORMATION

4.1 Introduction

In chapter III we set up the criteria for determining whether or not the use of prior information is beneficial. Case I in which prior data has already been taken will be considered in this chapter and in chapter V. Case II in which prior data has not been taken will be considered in chapter VI.

In the case that prior data is available, the criterion is given by equation (3.10),

$$\begin{aligned} V_2 - \text{MSE}(\hat{\beta}_2 | Y_1) \\ = (I + V_1 V_2^{-1})^{-1} [2V_1 + V_2 - V - (\hat{\beta}_1 - \bar{\beta})(\hat{\beta}_1 - \bar{\beta})'] (I + V_2^{-1} V_1)^{-1}. \end{aligned} \quad (4.1)$$

If $V_2 - \text{MSE}(\hat{\beta}_2 | Y_1)$ is positive definite, the use of prior information is beneficial. If $V_2 - \text{MSE}(\hat{\beta}_2 | Y_1)$ is negative definite, the use of prior information is detrimental.

Since V_1 and V_2 are positive definite, we observe from equation (4.1) that $V_2 - \text{MSE}(\hat{\beta}_2 | Y_1)$ is positive definite if and only if $2V_1 + V_2 - V - (\hat{\beta}_1 - \bar{\beta})(\hat{\beta}_1 - \bar{\beta})'$ is positive

definite.

Consider the following three cases:

- (1) $2V_1 + V_2 - V$ is negative definite. In this case, clearly $V_2 - \text{MSE}(\hat{\beta}_2 | Y_1)$ is negative definite. The use of prior information is detrimental.
- (2) $2V_1 + V_2 - V$ is positive definite. We will consider this case in section 4.2 of this chapter.
- (3) $2V_1 + V_2 - V$ is neither positive definite nor negative definite. We will discuss this case in section 4.5 of this chapter.

4.2 The case when $2V_1 + V_2 - V$ is Positive Definite

Denote

$$A^{-1} = 2V_1 + V_2 - V$$

and

$$Z = \hat{\beta}_1 - \bar{\beta} \quad (4.2)$$

where A^{-1} is an $r \times r$ positive definite matrix and Z is an $r \times 1$ random vector. Equation (4.1) can be written as

$$V_2 - \text{MSE}(\hat{\beta}_2 | Y_1) = (I + V_1 V_2^{-1})^{-1} (A^{-1} - Z Z') (I + V_2^{-1} V_1)^{-1}. \quad (4.3)$$

We will derive the probability that $V_2 - \text{MSE}(\hat{\beta}_2 | Y_1)$ is positive definite, that is the probability that the use of prior information is beneficial. The following lemma is due to Watkins [24].

Lemma 4.1 If B is an $r \times r$ positive definite symmetric matrix and Z is an $r \times 1$ vector, then $B - ZZ'$ is positive definite if and only if $Z'B^{-1}Z < 1$.

By lemma 4.1 and equation (4.3), we conclude with the following lemma.

Lemma 4.2 If A^{-1} given in equation (4.3) is positive definite, (obviously A^{-1} is symmetric), then $V_2 - \text{MSE}(\hat{\beta}_2|Y_1)$ is positive definite if and only if $Z'AZ < 1$. Hence,

$$\begin{aligned} & \Pr [V_2 - \text{MSE}(\hat{\beta}_2|Y_1) \text{ is positive definite}] \\ &= \Pr [Z'AZ < 1]. \end{aligned}$$

We will develop the distribution of $Z'AZ$ so that $\Pr [Z'AZ < 1]$ can be evaluated. The following lemma is a direct consequence of the r -variate normal distribution.

Lemma 4.3 If B is an $r \times r$ positive definite matrix and Z is an $r \times 1$ vector, then

$$\frac{1}{(2\pi)^{r/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-Z'BZ/2} dz_1 \dots dz_r = |B|^{-1/2}$$

where $|B|$ denotes the determinant of B .

Lemma 4.4 If a vector Z is distributed $N(0, I)$ and A is a positive definite matrix, then the characteristic function of $y = Z'AZ$ is given by

$$\psi_y(t) = |I - 2itA|^{-1/2}.$$

Lemma 4.5

$$\binom{-\frac{1}{2}}{n} = (-1)^n \binom{2n}{n} 2^{-2n}.$$

Proof

$$\begin{aligned} \binom{-\frac{1}{2}}{n} &= \frac{(-\frac{1}{2})(-\frac{1}{2}-1)\dots(-\frac{1}{2}-n+1)}{n!} \\ &= \frac{(-1)^n}{n!} \left(\frac{1}{2}\right)\left(\frac{1}{2}+1\right)\dots\left(\frac{1}{2}+n-1\right) \\ &= \frac{(-1)^n}{n!} \frac{1}{2^n} 1.3.5\dots(2n-1) \\ &= \frac{(-1)^n}{n!} \frac{1}{2^n} \frac{1.2.3.4.5\dots(2n-1)(2n)}{2.4.6.8\dots(2n)} \\ &= \frac{(-1)^n}{n!} \frac{1}{2^n} \frac{(2n)!}{2^n(1.2.3\dots n)} \\ &= \frac{(-1)^n}{n!} \frac{1}{2^n} \frac{(2n)!}{n!} \frac{1}{2^n} \\ &= (-1)^n 2^{-2n} \binom{2n}{n}. \end{aligned}$$

The following theorem is a modification of Ruben's result [21]. However, we give a different proof here.

Theorem 4.1 If a vector Z is distributed $N(0, I)$ and A is a symmetric positive definite matrix, then the p.d.f. of $y = Z'AZ$ is given by

$$f_y(y) = |A|^{-1/2} \sum_{k=0}^{\infty} \frac{c_k}{\lambda_{\min}^k} \frac{1}{2^{k+r/2} \Gamma(k+r/2)} e^{-y/(2\lambda_{\min})} y^{k+r/2-1},$$

$y > 0$

(4.4)

where the λ_i 's are eigenvalues of A , $i = 1, 2, \dots, r$;

$\lambda_{\min} = \min \{\lambda_1, \lambda_2, \dots, \lambda_r\}$, and

$$c_k = 2^{-2k} \sum_{k_1 + \dots + k_r = k} \binom{2k_1}{k_1} \dots \binom{2k_r}{k_r} \left(1 - \frac{\lambda_{\min}}{\lambda_1}\right)^{k_1} \dots \left(1 - \frac{\lambda_{\min}}{\lambda_r}\right)^{k_r}.$$

(4.5)

Proof Since A is symmetric and positive definite, there exists an orthogonal matrix P such that

$$P'AP = \begin{bmatrix} \lambda_1 & & & \phi \\ & \lambda_2 & & \\ & & \ddots & \\ \phi & & & \lambda_r \end{bmatrix}$$

(4.6)

where $\lambda_1, \lambda_2, \dots, \lambda_r$ are eigenvalues of A and all positive.

Since $Z \sim N(0, I)$, by lemma 4.4, the characteristic function of $y = Z'AZ$ is

$$\psi_Y(t) = |I - 2itA|^{-1/2}. \quad (4.7)$$

Using equation (4.6), we write equation (4.7) as

$$\begin{aligned} \psi_Y(t) &= |P'(I - 2itA)P|^{-1/2} \\ &= \left| \begin{array}{cccc} 1-2it\lambda_1 & & & \\ & \cdot & & \\ \phi & & \cdot & \phi \\ & & \cdot & \\ & & & 1-2it\lambda_r \end{array} \right|^{-1/2} \\ &= \prod_{j=1}^r (1 - 2it\lambda_j)^{-1/2}. \end{aligned} \quad (4.8)$$

The p.d.f. of y will be

$$\begin{aligned} f_Y(y) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \psi_Y(t) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \prod_{j=1}^r (1 - 2it\lambda_j)^{-1/2} dt. \end{aligned} \quad (4.9)$$

Observe that

$$\begin{aligned} &1 - 2it\lambda_j \\ &= 1 - 2it\lambda_{\min} \frac{\lambda_j}{\lambda_{\min}} \\ &= \frac{\lambda_j}{\lambda_{\min}} (1 - 2it\lambda_{\min}) \left[1 - \left(1 - \frac{\lambda_{\min}}{\lambda_j} \right) (1 - 2it\lambda_{\min})^{-1} \right]. \end{aligned}$$

Taking square roots on both sides, we obtain

$$\begin{aligned}
& (1 - 2it\lambda_j)^{-1/2} \\
&= \left(\frac{\lambda_j}{\lambda_{\min}} \right)^{-1/2} (1-2it\lambda_{\min})^{-1/2} \left[1 - \left(1 - \frac{\lambda_{\min}}{\lambda_j} \right) (1-2it\lambda_{\min})^{-1} \right]^{-1/2}.
\end{aligned} \tag{4.10}$$

Since $\frac{\lambda_{\min}}{\lambda_j} \leq 1$ for $j = 1, 2, \dots, r$, we can apply the binomial theorem and write equation (4.10) as

$$\begin{aligned}
& (1 - 2it\lambda_j)^{-1/2} \\
&= \left(\frac{\lambda_j}{\lambda_{\min}} \right)^{-1/2} (1-2it\lambda_{\min})^{-1/2} \sum_{k=0}^{\infty} (-1)^k \binom{-\frac{1}{2}}{k} \left(1 - \frac{\lambda_{\min}}{\lambda_j} \right)^k (1-2it\lambda_{\min})^{-k}.
\end{aligned}$$

Taking products on both sides,

$$\begin{aligned}
& \prod_{j=1}^r (1 - 2it\lambda_j)^{-1/2} \\
&= \prod_{j=1}^r \left(\frac{\lambda_j}{\lambda_{\min}} \right)^{-1/2} (1-2it\lambda_{\min})^{-r/2} \sum_{k=0}^{\infty} (-1)^k \binom{-\frac{1}{2}}{k} \left(1 - \frac{\lambda_{\min}}{\lambda_j} \right)^k \\
& \quad \cdot (1-2it\lambda_{\min})^{-k} \\
&= \prod_{j=1}^r \left(\frac{\lambda_j}{\lambda_{\min}} \right)^{-1/2} (1-2it\lambda_{\min})^{-r/2} \sum_{k_1, \dots, k_r=0}^{\infty} \prod_{j=1}^r (-1)^{k_j} \binom{-\frac{1}{2}}{k_j} \\
& \quad \cdot \left(1 - \frac{\lambda_{\min}}{\lambda_j} \right)^{k_j} (1-2it\lambda_{\min})^{k_j}.
\end{aligned}$$

Collecting terms, where $k_1 + k_2 + \dots + k_r = k$, we obtain

$$\begin{aligned}
& \prod_{j=1}^r (1 - 2it\lambda_j)^{-1/2} \\
&= \prod_{j=1}^r \left(\frac{\lambda_j}{\lambda_{\min}} \right)^{-1/2} (1 - 2it\lambda_{\min})^{-r/2} \\
&\quad \cdot \sum_{k=0}^{\infty} \sum_{k_1+k_2+\dots+k_r=k} (-1)^k \binom{-\frac{1}{2}}{k_1} \dots \binom{-\frac{1}{2}}{k_r} \left(1 - \frac{\lambda_{\min}}{\lambda_1}\right)^{k_1} \\
&\quad \dots \left(1 - \frac{\lambda_{\min}}{\lambda_r}\right)^{k_r} (1 - 2it\lambda_{\min})^{-k} \\
&= \prod_{j=1}^r \left(\frac{\lambda_j}{\lambda_{\min}} \right)^{-1/2} \sum_{k=0}^{\infty} c_k (1 - 2it\lambda_{\min})^{-(k+r/2)}, \tag{4.11}
\end{aligned}$$

where

$$c_k = \sum_{k_1+\dots+k_r=k} (-1)^k \binom{-\frac{1}{2}}{k_1} \dots \binom{-\frac{1}{2}}{k_r} \left(1 - \frac{\lambda_{\min}}{\lambda_1}\right)^{k_1} \dots \left(1 - \frac{\lambda_{\min}}{\lambda_r}\right)^{k_r}.$$

By lemma 4.5, we can write

$$c_k = 2^{-2k} \sum_{k_1+\dots+k_r=k} \binom{2k_1}{k_1} \dots \binom{2k_r}{k_r} \left(1 - \frac{\lambda_{\min}}{\lambda_1}\right)^{k_1} \dots \left(1 - \frac{\lambda_{\min}}{\lambda_r}\right)^{k_r}. \tag{4.12}$$

Substituting equation (4.11) into equation (4.9),

we obtain

$$\begin{aligned}
& f_Y(y) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} \prod_{j=1}^r \left(\frac{\lambda_j}{\lambda_{\min}} \right)^{-1/2} \sum_{k=0}^{\infty} c_k (1 - 2it\lambda_{\min})^{-(k+r/2)} dt
\end{aligned}$$

$$= \prod_{j=1}^r \left(\frac{\lambda_j}{\lambda_{\min}} \right)^{-1/2} \sum_{k=0}^{\infty} c_k \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ity} (1-2it\lambda_{\min})^{-(k+r/2)} dt. \quad (4.13)$$

Let $t\lambda_{\min} = u$. Equation (4.13) can be written as

$$f_Y(y) = \prod_{j=1}^r \left(\frac{\lambda_j}{\lambda_{\min}} \right)^{-1/2} \sum_{k=0}^{\infty} c_k \frac{1}{\lambda_{\min}} \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iuy/\lambda_{\min}} (1-2iu)^{-(k+r/2)} du. \quad (4.14)$$

It is easy to see from the relationship between the characteristic function and the p.d.f. of a chi-square distribution with n degrees of freedom that

$$\begin{aligned} & \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} (1-2iu)^{-n/2} du \\ &= \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} e^{-x/2} x^{n/2-1}, & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases} \end{aligned} \quad (4.15)$$

Using equation (4.15), equation (4.14) becomes

$$\begin{aligned} f_Y(y) &= \prod_{j=1}^r \left(\frac{\lambda_j}{\lambda_{\min}} \right)^{-1/2} \sum_{k=0}^{\infty} \frac{c_k}{\lambda_{\min}} \frac{1}{2^{k+r/2} \Gamma(k+r/2)} e^{-y/(2\lambda_{\min})} \left(\frac{y}{\lambda_{\min}} \right)^{k+r/2-1} \\ &= \prod_{j=1}^r \lambda_j^{-1/2} \sum_{k=0}^{\infty} \frac{c_k}{\lambda_{\min}^k} \frac{1}{2^{k+r/2} \Gamma(k+r/2)} e^{-y/(2\lambda_{\min})} y^{k+r/2-1} \end{aligned}$$

$$= |A|^{-1/2} \sum_{k=0}^{\infty} \frac{c_k}{\lambda_{\min}^k} \frac{1}{2^{k+r/2} \Gamma(k+r/2)} e^{-y/(2\lambda_{\min})} y^{k+r/2-1},$$

where $|A| = |P'AP| = \prod_{j=1}^r \lambda_j$ by equation (4.6).

This completes the proof of theorem 4.1.

Corollary 4.1 In theorem 4.1, if $\lambda_i = 1$ for $i = 1, \dots, r$; then y is distributed as χ_r^2 .

Proof If $\lambda_i = 1$ for $i = 1, \dots, r$, we then have

$$c_0 = 1, c_k = 0 \text{ for } k > 1 \text{ and } |A| = 1.$$

Hence, the p.d.f. of y is

$$f_y(y) = \frac{1}{2^{r/2} \Gamma(r/2)} e^{-y/2} y^{r/2-1},$$

which is the p.d.f. of χ_r^2 .

From equation (4.2), it is easy to see that $E(Z) = 0$ and $\text{cov}(Z) = V_1 + V$. We have the following theorem.

Theorem 4.2 If Z is distributed $N(0, V_1 + V)$ and A is a symmetric positive definite matrix, then the p.d.f. of $y = Z'AZ$ is given by

$$f_y(y) = |A(V_1 + V)|^{-1/2} \sum_{k=0}^{\infty} \frac{c_k}{\lambda_{\min}^k} \frac{1}{2^{k+r/2} \Gamma(k+r/2)} e^{-y/(2\lambda_{\min})} y^{k+r/2-1},$$

$y > 0.$

where the λ_i 's are eigenvalues of $A(V_1+V)$, $i = 1, 2, \dots, r$;
 $\lambda_{\min} = \min \{\lambda_1, \dots, \lambda_r\}$, and

$$c_k = 2^{-2k} \sum_{k_1 + \dots + k_r = k} \binom{2k_1}{k_1} \cdots \binom{2k_r}{k_r} \left(1 - \frac{\lambda_{\min}}{\lambda_1}\right)^{k_1} \cdots \left(1 - \frac{\lambda_{\min}}{\lambda_r}\right)^{k_r}.$$

Proof Let $W = (V_1+V)^{-1/2}Z$. Then $W \sim N(0, I)$ and

$$y = Z'AZ = W'(V_1+V)^{1/2}A(V_1+V)^{1/2}W.$$

By lemma 4.4, the characteristic function of y is

$$\psi_y(t) = |I - 2it(V_1+V)^{1/2}A(V_1+V)^{1/2}|^{-1/2}.$$

Using this characteristic function and replacing A by $(V_1+V)^{1/2}A(V_1+V)^{1/2}$ in the proof of theorem 4.1, we obtain the result of this theorem.

Notation Let P be the probability that the use of prior information is beneficial.

Recall from equation (4.2) that $A = (2V_1+V_2-V)^{-1}$ is symmetric and positive definite. We can apply theorem 4.1 to prove the following important result.

Theorem 4.3 The probability P that the use of prior information is beneficial is given by

$$P = \frac{\lambda_{\min}^{r/2}}{\sqrt{|A(V_1+V)|}} \sum_{k=0}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \quad (4.16)$$

where $\lambda_{\min}$ and c_k are given in theorem 4.2 and

$$A = (2V_1 + V_2 - V)^{-1}.$$

Proof By lemma 4.2,

$$\begin{aligned} P &= \Pr [V_2 - \text{MSE}(\hat{\beta}_2 | Y_1) \text{ is positive definite}] \\ &= \Pr (Z'AZ < 1). \end{aligned}$$

Applying theorem 4.2 and writing $y = Z'AZ$, we have

$$\begin{aligned} P &= \Pr (y < 1) \\ &= \int_0^1 f_Y(y) dy \\ &= |A(V_1 + V)|^{-1/2} \int_0^1 \sum_{k=0}^{\infty} \frac{c_k}{\lambda_{\min}^k} \frac{1}{2^{k+r/2} \Gamma(k+r/2)} e^{-y/(2\lambda_{\min})} \\ &\quad \cdot y^{k+r/2-1} dy \\ &= \frac{1}{\sqrt{|A(V_1 + V)|}} \sum_{k=0}^{\infty} \frac{c_k}{\lambda_{\min}^k} \frac{1}{2^{k+r/2} \Gamma(k+r/2)} \int_0^1 e^{-y/(2\lambda_{\min})} \\ &\quad \cdot y^{k+r/2-1} dy. \end{aligned}$$

Let $y = \lambda_{\min} z$, we obtain

$$\begin{aligned} P &= \frac{\lambda_{\min}^{r/2}}{\sqrt{|A(V_1 + V)|}} \sum_{k=0}^{\infty} c_k \frac{1}{2^{k+r/2} \Gamma(k+r/2)} \int_0^{1/\lambda_{\min}} e^{-z/2} z^{k+r/2-1} dz \\ &= \frac{\lambda_{\min}^{r/2}}{\sqrt{|A(V_1 + V)|}} \sum_{k=0}^{\infty} c_k \Pr (\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}), \end{aligned}$$

which is the required result.

Corollary 4.2 In theorem 4.3, if $\lambda_i = \lambda$ for $i = 1, \dots, r$, then $P = \Pr (\chi_r^2 \leq \frac{1}{\lambda})$.

Proof It follows directly from theorem 4.2.

Note that P depends on the eigenvalues of $A(V_1+V)$ only.

4.3 A Recursive Relationship for c_k

The constants c_k in theorem 4.2 are expressed as:

$$c_k = 2^{-2k} \sum_{k_1 + \dots + k_r = k} \binom{2k_1}{k_1} \dots \binom{2k_r}{k_r} \left(1 - \frac{\lambda_{\min}}{\lambda_1}\right)^{k_1} \dots \left(1 - \frac{\lambda_{\min}}{\lambda_r}\right)^{k_r}. \quad (4.17)$$

This form of c_k is not convenient for computation. We will derive a recursive relationship for c_k so that the c_k 's can conveniently be calculated.

Lemma 4.6 If a random variable W has a chi-square distribution with one degree of freedom, the n -th moment of W denoted by μ'_n is given by

$$\mu'_n = (2n-1)!!$$

where $(2n-1)!! = (2n-1)(2n-3)\dots 5.3.1$.

Proof Since $W \sim \chi_1^2$, the moment generating function of W is

$$M_W(t) = (1 - 2t)^{-1/2}.$$

Differentiating the above equation n times with respect to t , we obtain

$$M_W^{(n)}(t) = 1.3.5\dots(2n-1)(1-2t)^{-(2n+1)/2}.$$

Hence
$$M_W^{(n)}(0) = (2n-1)!!.$$

That is,
$$\mu_k' = E(W^n) = (2n-1)!!.$$

Lemma 4.7 Let $W_1, W_2, \dots, W_r$ be mutually stochastically independent random variables which are identically distributed as a chi-square distribution with one degree of freedom. Then the random variable W defined by

$$W = \frac{1}{2} \sum_{j=1}^r a_j W_j$$

has the cumulant generating function $Q(\theta, a_j)$ given by

$$Q(\theta, a_j) = \frac{1}{2} \sum_{i=1}^{\infty} \sum_{j=1}^r \frac{(a_j \theta)^i}{i}. \quad (4.18)$$

Furthermore, if κ_i represents the i -th cumulant of $Q(\theta, a_j)$, then

$$\kappa_i = \frac{(i-1)!}{2} \sum_{j=1}^r a_j^i. \quad (4.19)$$

Proof Since $W_j \sim \chi_1^2$ and the W_j 's are mutually stochastically independent, the moment generating function of W is

$$\begin{aligned} E(e^{\theta W}) &= E(e^{\frac{1}{2}\theta \sum_{j=1}^r a_j W_j}) \\ &= \prod_{j=1}^r E(e^{\frac{1}{2}\theta a_j W_j}) \\ &= \prod_{j=1}^r (1 - \theta a_j)^{-1/2}. \end{aligned}$$

Hence the cumulant generating function $Q(\theta, a_j)$ is

$$\begin{aligned} Q(\theta, a_j) &= \log E(e^{\theta W}) \\ &= -\frac{1}{2} \sum_{j=1}^r \log (1 - \theta a_j). \end{aligned} \quad (4.20)$$

Expanding $\log (1 - \theta a_j)$ as a Taylor series, equation (4.20) becomes

$$\begin{aligned} Q(\theta, a_j) &= \frac{1}{2} \sum_{j=1}^r \sum_{i=1}^{\infty} \frac{(\theta a_j)^i}{i} \\ &= \frac{1}{2} \sum_{i=1}^{\infty} \sum_{j=1}^r \frac{(a_j \theta)^i}{i} \end{aligned}$$

which is equation (4.18).

If the i -th cumulant of $Q(\theta, a_j)$ is denoted by κ_i , then

$$Q(\theta, a_j) = \sum_{i=1}^{\infty} \kappa_i \frac{\theta^i}{i!}. \quad (4.21)$$

Comparing the coefficients in equation (4.18) and (4.21), we have

$$\frac{\kappa_i}{i!} = \frac{1}{2} \sum_{j=1}^r \frac{a_j^i}{i}.$$

Hence

$$\kappa_i = \frac{1}{2} (i-1)! \sum_{j=1}^r a_j^i$$

which is equation (4.19).

The following lemma is a well-known formula [12].

Lemma 4.8

$$\mu_k' = \sum_{j=0}^{k-1} \binom{k-1}{j} \kappa_{k-j} \mu_j'.$$

Notation The trace of a matrix A is denoted by $\text{tr } A$.

The following theorem makes it possible to compute the values of c_k conveniently.

Theorem 4.4 The constants c_k in equation (4.17) satisfy the recursive relationship

$$c_0 = 1,$$

$$c_k = \frac{1}{2k} \sum_{j=0}^{k-1} \text{tr} [I - \lambda_{\min} A^{-1} (V_1 + V)^{-1}]^{k-j} c_j, \quad k \geq 1.$$

Proof It follows easily that $c_0 = 1$.

For $k \geq 1$, equation (4.17) can be written as

$$\begin{aligned}
c_k &= 2^{-2k} \sum_{k_1 + \dots + k_r = k} \frac{(2k_1)!}{k_1! k_1!} \dots \frac{(2k_r)!}{k_r! k_r!} \\
&\quad \cdot (1 - \frac{\lambda_{\min}}{\lambda_1})^{k_1} \dots (1 - \frac{\lambda_{\min}}{\lambda_r})^{k_r} \\
&= 2^{-2k} \frac{1}{k!} \sum_{k_1 + \dots + k_r = k} \frac{k!}{k_1! \dots k_r!} (1 - \frac{\lambda_{\min}}{\lambda_1})^{k_1} \frac{(2k_1)!}{k_1!} \\
&\quad \dots (1 - \frac{\lambda_{\min}}{\lambda_r})^{k_r} \frac{(2k_r)!}{k_r!} . \quad (4.22)
\end{aligned}$$

Observe that

$$\begin{aligned}
\frac{(2n)!}{n!} &= \frac{(2n)(2n-1)\dots 4.3.2.1}{n(n-1)\dots 4.3.2.1} \\
&= \frac{(2n)(2n-1)\dots 4.3.2.1}{(2n)(2n-2)\dots 8.6.4.2} \cdot 2^n \\
&= 2^n [(2n-1)!!] .
\end{aligned}$$

We can then write equation (4.22) as

$$\begin{aligned}
c_k &= 2^{-k} \frac{1}{k!} \sum_{k_1 + \dots + k_r = k} \binom{k}{k_1, \dots, k_r} (1 - \frac{\lambda_{\min}}{\lambda_1})^{k_1} (2k_1-1)!! \\
&\quad \dots (1 - \frac{\lambda_{\min}}{\lambda_r})^{k_r} (2k_r-1)!! \quad (4.23)
\end{aligned}$$

$$\text{where } \binom{k}{k_1, \dots, k_r} = \frac{k!}{k_1! \dots k_r!} . \quad (4.24)$$

Let W_j , $j=1, \dots, r$, be mutually stochastically independent random variables identically distributed as χ_1^2 and let

$$W = \frac{1}{2} \sum_{j=1}^r (1 - \lambda_{\min}/\lambda_j) W_j.$$

By lemma 4.6, equation (4.23) can be written as

$$\begin{aligned} c_k &= \frac{2^{-k}}{k!} \sum_{k_1 + \dots + k_r = k} \binom{k}{k_1, \dots, k_r} (1 - \lambda_{\min}/\lambda_1)^{k_1} E(W_1^{k_1}) \\ &\quad \dots (1 - \lambda_{\min}/\lambda_r)^{k_r} E(W_r^{k_r}) \\ &= \frac{2^{-k}}{k!} E \left\{ \sum_{k_1 + \dots + k_r = k} \binom{k}{k_1, \dots, k_r} (1 - \lambda_{\min}/\lambda_1)^{k_1} W_1^{k_1} \right. \\ &\quad \left. \dots (1 - \lambda_{\min}/\lambda_r)^{k_r} W_r^{k_r} \right\} \\ &= \frac{2^{-k}}{k!} E \left\{ \left[\sum_{j=1}^r (1 - \lambda_{\min}/\lambda_j) W_j \right]^k \right\} \\ &= \frac{1}{k!} E \left\{ \left[\frac{1}{2} \sum_{j=1}^r (1 - \lambda_{\min}/\lambda_j) W_j \right]^k \right\} \\ &= \frac{1}{k!} E(W^k) \\ &= \frac{1}{k!} \mu'_k. \end{aligned} \tag{4.25}$$

Hence, $\mu'_k = k! c_k.$ (4.26)

Substituting equation (4.26) into lemma 4.8, we have

$$k! c_k = \sum_{j=0}^{k-1} \binom{k-1}{j} \kappa_{k-j} j! c_j. \quad (4.27)$$

By lemma 4.7, we have

$$\kappa_{k-j} = \frac{(k-j-1)!}{2} \sum_{j=1}^r (1 - \lambda_{\min}/\lambda_j)^{k-j}. \quad (4.28)$$

Recall from theorem 4.2 that $\lambda_1, \lambda_2, \dots, \lambda_r$ are eigenvalues of the matrix $A(V_1+V)$, which are the same as the eigenvalues of $(V_1+V)A$. Equation (4.28) becomes

$$\kappa_{k-j} = \frac{(k-j-1)!}{2} \operatorname{tr} [I - \lambda_{\min} A^{-1}(V_1+V)^{-1}]^{k-j}. \quad (4.29)$$

Substituting equation (4.29) into equation (4.27), we obtain

$$\begin{aligned} k! c_k &= \sum_{j=0}^{k-1} \binom{k-1}{j} \frac{(k-j-1)!}{2} \operatorname{tr} [I - \lambda_{\min} A^{-1}(V_1+V)^{-1}]^{k-j} j! c_j \\ &= \frac{(k-1)!}{2} \sum_{j=0}^{k-1} \operatorname{tr} [I - \lambda_{\min} A^{-1}(V_1+V)^{-1}]^{k-j} c_j. \end{aligned}$$

Hence

$$c_k = \frac{1}{2k} \sum_{j=0}^{k-1} \operatorname{tr} [I - \lambda_{\min} A^{-1}(V_1+V)^{-1}]^{k-j} c_j, \quad k \geq 1.$$

This completes the proof of theorem 4.4.

4.4 An Additional Formula for c_k

We have obtained two expressions for c_k , namely equations (4.17) and (4.25). We shall now obtain an additional formula for c_k .

Observe that

$$\begin{aligned}
 \binom{2k_1}{k_1} \cdots \binom{2k_r}{k_r} &= \frac{(2k_1)!}{k_1! k_1!} \cdots \frac{(2k_r)!}{k_r! k_r!} \\
 &= \frac{(2k_1)! \cdots (2k_r)!}{(2k)!} \left(\frac{k!}{k_1! \cdots k_r!} \right)^2 \frac{(2k)!}{k! k!} \\
 &= \frac{1}{\binom{2k}{2k_1, \dots, 2k_r}} \binom{k}{k_1, \dots, k_r}^2 \binom{2k}{k}.
 \end{aligned}$$

Therefore equation (4.17) can be written as

$$c_k = 2^{-2k} \binom{2k}{k}_{k_1 + \dots + k_r = k} \left[\frac{\binom{k}{k_1, \dots, k_r}^2}{\binom{2k}{2k_1, \dots, 2k_r}} \left(1 - \frac{\lambda_{\min}}{\lambda_1}\right)^{k_1} \cdots \left(1 - \frac{\lambda_{\min}}{\lambda_r}\right)^{k_r} \right]. \quad (4.30)$$

We remark finally that the recursive formulae given in theorem 4.4 appear to be best adapted for purposes of computation.

4.5 The Case when $2V_1+V_2-V$ is Neither Positive Definite Nor Negative Definite

Denote

$$H = 2V_1+V_2-V$$

and

$$Z = \hat{\beta}_1 - \bar{\beta} \quad (4.31)$$

where H is an $r \times r$ matrix neither positive definite nor negative definite and Z is an $r \times 1$ random vector. Equation (4.1) then becomes

$$V_2 - \text{MSE}(\hat{\beta}_2 | Y_1) = (I + V_1 V_2^{-1})^{-1} (H - Z Z') (I + V_2^{-1} V_1)^{-1}. \quad (4.32)$$

Since H is symmetric, there exists an orthogonal matrix G such that

$$GHG' = \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix} \quad (4.33)$$

where H_1 is an $r_1 \times r$ positive definite matrix, H_2 is an $r_2 \times r$ negative semi-definite matrix and $r_1 + r_2 = r$.

Partition G as follows:

$$G = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \quad (4.34)$$

where G_1 has size $r_1 \times r$ and G_2 has size $r_2 \times r$, and

$r_1 + r_2 = r$. Equation (4.33) becomes

$$\begin{bmatrix} G_1 \\ G_2 \end{bmatrix} H \begin{bmatrix} G_1' & G_2' \end{bmatrix} = \begin{bmatrix} H_1 & 0 \\ 0 & H_2 \end{bmatrix},$$

which implies

$$G_1 H G_1' = H_1,$$

$$G_2 H G_2' = H_2,$$

and

$$G_1 H G_2' = G_2 H G_1' = 0. \quad (4.35)$$

With the conditions in equation (4.35) and using the property of the orthogonality of G , we obtain from equation (4.32) that

$$\begin{aligned} & V_2 - \text{MSE}(\hat{\beta}_2 | Y_1) \\ &= (I + V_1 V_2^{-1})^{-1} G' G (H - Z Z') G' G (I + V_2^{-1} V_1)^{-1} \\ &= (I + V_1 V_2^{-1})^{-1} G' \begin{bmatrix} G_1 (H - Z Z') G_1' & -G_1 Z Z' G_2' \\ -G_2 Z Z' G_1' & G_2 (H - Z Z') G_2' \end{bmatrix} G (I + V_2^{-1} V_1)^{-1} \\ &= (I + V_1 V_2^{-1})^{-1} G' \begin{bmatrix} H_1 - G_1 Z Z' G_1' & -G_1 Z Z' G_2' \\ -G_2 Z Z' G_1' & H_2 - G_2 Z Z' G_2' \end{bmatrix} G (I + V_2^{-1} V_1)^{-1}. \end{aligned} \quad (4.36)$$

Since H_2 is negative semi-definite, then $H_2 - G_2 Z Z' G_2'$

is negative semi-definite. This suggests that the use of prior information on estimating $G_2\beta_2$ is detrimental.

However, since H_1 is positive definite, the criterion to determine whether prior information should be used to estimate $G_1\beta_2$ depends on whether $H_1 - G_1ZZ'G_1'$ is positive definite. The probability that the use of prior information is beneficial in estimating $G_1\beta_2$ has been discussed in section 4.2 of this chapter.

Next, we shall derive an expression for the estimator of β_2 in this situation. The following lemma is well-known [10].

Lemma 4.9 Under the linear model given in equation (3.1), the best linear unbiased estimate of $G_1\beta_i$, $i=1,2$, is

$$\begin{aligned}\widehat{G_1\beta_i} &= G_1\hat{\beta}_i \\ &= G_1(X_i'\Sigma_i^{-1}X_i)^{-1}X_i'\Sigma_i^{-1}Y_i \\ &= G_1V_iX_i'\Sigma_i^{-1}Y_i.\end{aligned}$$

Theorem 4.5 Under the combined linear model given in equation (3.4), the best estimate of $G_1\beta_2$ is given by

$$\begin{aligned}\widehat{G_1\beta_2} &= [(G_1V_1G_1')^{-1} + (G_1V_2G_1')^{-1}]^{-1}[(G_1V_1G_1')^{-1}G_1\hat{\beta}_1 \\ &\quad + (G_1V_2G_1')^{-1}G_1\hat{\beta}_2].\end{aligned}$$

Proof The combined linear model in equation (3.4) is

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta_2 + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (4.37)$$

where e_1 and e_2 are independent.

Multiplying equation (4.37) on the left by

$$\begin{bmatrix} G_1 (X_1' \Sigma_1^{-1} X_1)^{-1} X_1' \Sigma_1^{-1} & 0 \\ 0 & G_1 (X_2' \Sigma_2^{-1} X_2)^{-1} X_2' \Sigma_2^{-1} \end{bmatrix}$$

we have

$$\begin{bmatrix} G_1 (X_1' \Sigma_1^{-1} X_1)^{-1} X_1' \Sigma_1^{-1} Y_1 \\ G_1 (X_2' \Sigma_2^{-1} X_2)^{-1} X_2' \Sigma_2^{-1} Y_2 \end{bmatrix} = \begin{bmatrix} I \\ I \end{bmatrix} G_1 \beta_2 + \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \quad (4.38)$$

where $\delta_i = G_1 (X_i' \Sigma_i^{-1} X_i)^{-1} X_i' \Sigma_i^{-1} e_i$, $i = 1, 2$.

Since $E(e_i e_i') = \Sigma_i$ and $E(e_i e_j') = 0$, we obtain

$$\begin{aligned} & \text{cov}(\delta_i) \\ &= E(\delta_i \delta_i') \\ &= G_1 (X_i' \Sigma_i^{-1} X_i)^{-1} X_i' \Sigma_i^{-1} E(e_i e_i') \Sigma_i^{-1} X_i (X_i' \Sigma_i^{-1} X_i)^{-1} G_1' \\ &= G_1 (X_i' \Sigma_i^{-1} X_i)^{-1} G_1' \\ &= G_1 V_i G_1'. \end{aligned} \quad (4.39)$$

Thus,

$$\begin{aligned}
 \text{cov} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} &= E \left\{ \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} [\delta_1' \quad \delta_2'] \right\} \\
 &= E \begin{bmatrix} \delta_1 \delta_1' & \delta_1 \delta_2' \\ \delta_2 \delta_1' & \delta_2 \delta_2' \end{bmatrix} \\
 &= \begin{bmatrix} G_1 V_1 G_1' & 0 \\ 0 & G_1 V_2 G_1' \end{bmatrix} \quad (4.40)
 \end{aligned}$$

by equation (4.39).

Applying the Gauss-Markoff theorem to equation (4.38), we have

$$\begin{aligned}
 \hat{\hat{G}}_1 \beta_2 &= \left\{ \begin{bmatrix} I \\ I \end{bmatrix} \begin{bmatrix} G_1 V_1 G_1' & 0 \\ 0 & G_1 V_2 G_1' \end{bmatrix}^{-1} \begin{bmatrix} I \\ I \end{bmatrix} \right\}^{-1} \begin{bmatrix} I \\ I \end{bmatrix} \begin{bmatrix} G_1 V_1 G_1' & 0 \\ 0 & G_1 V_2 G_1' \end{bmatrix}^{-1} \\
 &\quad \cdot \begin{bmatrix} G_1 (X_1' \Sigma_1^{-1} X_1)^{-1} X_1' \Sigma_1^{-1} Y_1 \\ G_1 (X_2' \Sigma_2^{-1} X_2)^{-1} X_2' \Sigma_2^{-1} Y_2 \end{bmatrix} \\
 &= [(G_1 V_1 G_1')^{-1} + (G_1 V_2 G_1')^{-1}]^{-1} [(G_1 V_1 G_1')^{-1} G_1 \hat{\beta}_1 \\
 &\quad + (G_1 V_2 G_1')^{-1} G_1 \hat{\beta}_2]
 \end{aligned}$$

which is the required result.

Note that $G_1 V_1 G_1'$ is the covariance matrix of $G_1 \hat{\beta}_1$ and $G_1 V_2 G_1'$ is the covariance matrix of $G_1 \hat{\beta}_2$.

Theorem 4.6 Under the combined linear model given in equation (3.4), the best estimate of β_2 when $2V_1+V_2-V$ is neither positive definite nor negative definite is given by

$$\tilde{\beta}_2 = G_1' \widehat{G_1 \beta_2} + G_2' \hat{G_2 \beta_2}.$$

Proof When $2V_1+V_2-V$ is neither positive definite nor negative definite, prior information on estimating $G_2\beta_2$ should not be used. The estimator for $G\beta_2$ in this case is

$$\widetilde{G\beta_2} = \begin{bmatrix} \widehat{G_1 \beta_2} \\ G_2 \hat{\beta_2} \end{bmatrix}.$$

Hence, the estimator for β_2 is

$$\begin{aligned} \tilde{\beta}_2 &= G^{-1} \begin{bmatrix} \widehat{G_1 \beta_2} \\ G_2 \hat{\beta_2} \end{bmatrix} \\ &= [G_1' \quad G_2'] \begin{bmatrix} \widehat{G_1 \beta_2} \\ G_2 \hat{\beta_2} \end{bmatrix} \\ &= G_1' \widehat{G_1 \beta_2} + G_2' \hat{G_2 \beta_2} \end{aligned}$$

which is the required result.

CHAPTER V

INTERVAL ON PROBABILITY CONTENT

5.1 Bounds for c_k

We see from theorems 4.2 and 4.3 that the probability of using prior information is, in general, complicated. In addition to the recursive formulae for c_k given in theorem 4.4, we shall derive an upper and lower bound for c_k , hence some bounds for the probability.

The following lemma is due to Watkins [25].

Lemma 5.1

$$\sum_{k_1 + \dots + k_r = k} \binom{2k_1}{k_1} \dots \binom{2k_r}{k_r} = \frac{2^{2k} \Gamma(k+r/2)}{k! \Gamma(r/2)} .$$

Theorem 5.1 The c_k given in theorem 4.4 satisfy the inequality

$$\frac{\Gamma(k+r/2)}{\Gamma(r/2)} \frac{(1 - 1/a_{\min})^k}{k!} \leq c_k \leq \frac{\Gamma(k+r/2)}{\Gamma(r/2)} \frac{(1 - 1/a_{\max})^k}{k!}$$

where $a_i = \lambda_i / \lambda_{\min}$, $i=1, \dots, r$;

$$a_{\min} = \min \{a_1, a_2, \dots, a_r\};$$

and $a_{\max} = \max \{a_1, a_2, \dots, a_r\}$.

Proof The c_k 's can be written as

$$c_k = 2^{-2k} \sum_{k_1 + \dots + k_r = k} \binom{2k_1}{k_1} \dots \binom{2k_r}{k_r} (1-1/a_1)^{k_1} \dots (1-1/a_r)^{k_r}. \quad (5.1)$$

Clearly, $a_j \geq 1$ for $j = 1, 2, \dots, r$, we have

$$0 \leq 1 - 1/a_{\min} \leq 1 - 1/a_j \leq 1 - 1/a_{\max} \leq 1. \quad (5.2)$$

Applying equation (5.2) to equation (5.1), we obtain

$$\begin{aligned} & 2^{-2k} (1 - 1/a_{\min})^k \sum_{k_1 + \dots + k_r = k} \binom{2k_1}{k_1} \dots \binom{2k_r}{k_r} \\ & \leq c_k \\ & \leq 2^{-2k} (1 - 1/a_{\max})^k \sum_{k_1 + \dots + k_r = k} \binom{2k_1}{k_1} \dots \binom{2k_r}{k_r}. \end{aligned} \quad (5.3)$$

By lemma 5.1, equation (5.3) becomes

$$\frac{\Gamma(k+r/2)}{\Gamma(r/2)} \frac{(1-1/a_{\min})^k}{k!} \leq c_k \leq \frac{\Gamma(k+r/2)}{\Gamma(r/2)} \frac{(1-1/a_{\max})^k}{k!}$$

which is the required result.

5.2 An Upper Bound for the Error in using a Finite Number of Terms to Approximate the Probability Content

In theorem 5.2, the probability that prior information is beneficial is an infinite sum of a series. It is possible to find an upper bound for the error in using a finite (N) number of terms to approximate the infinite sum. That is, an upper bound for

$$\frac{\lambda_{\min}^{r/2}}{\sqrt{|A(V_1+V)|}} \sum_{k=N}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}})$$

which is the remainder after N-1 terms.

The following lemma is a consequence of Taylor's series.

Lemma 5.2 For any number x such that $0 \leq x \leq 1$, we have

$$\sum_{k=0}^{\infty} \frac{\Gamma(k+r/2)}{\Gamma(r/2)} \frac{x^k}{k!} = (1-x)^{-r/2}.$$

Theorem 5.2 The series $\sum_{k=0}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}})$ is

uniformly convergent.

Proof Since $\Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \leq 1$, we have

$$\sum_{k=0}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \leq \sum_{k=0}^{\infty} c_k. \quad (5.4)$$

By theorem 5.1, equation (5.4) becomes

$$\sum_{k=0}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \leq \sum_{k=0}^{\infty} \frac{\Gamma(k+r/2)}{\Gamma(r/2)} \frac{(1-1/a_{\max})^k}{k!}. \quad (5.5)$$

Applying lemma 5.2, equation (5.5) reduces to

$$\sum_{k=0}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \leq [1 - (1-1/a_{\max})]^{-r/2} = a_{\max}^{r/2},$$

since $0 \leq 1-1/a_{\max} \leq 1$. This proves the theorem.

Theorem 5.3

$$\begin{aligned} & \left| \sum_{k=N}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \right| \\ & \leq \frac{\Gamma(N+r/2)}{N! \Gamma(r/2)} (\lambda_{\max}/\lambda_{\min})^{N+r/2} (1-\lambda_{\min}/\lambda_{\max})^N \Pr(\chi_{r+2N}^2 \leq \frac{1}{\lambda_{\max}}). \end{aligned}$$

Proof From theorem 5.1, we have

$$c_k \leq \frac{\Gamma(k+r/2)}{\Gamma(r/2)} \frac{(1-1/a_{\max})^k}{k!}.$$

Recalling that

$$\Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) = \frac{1}{2^{k+r/2} \Gamma(k+r/2)} \int_0^{1/\lambda_{\min}} e^{-z/2} z^{k+r/2-1} dz,$$

we have

$$\begin{aligned}
& \left| \sum_{k=N}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \right| \\
&= \sum_{k=N}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\
&\leq \sum_{k=N}^{\infty} \frac{\Gamma(k + \frac{r}{2})}{\Gamma(\frac{r}{2})} \frac{(1 - \frac{1}{a_{\max}})^k}{k!} \frac{1}{2^{k+r/2} \Gamma(k + \frac{r}{2})} \int_0^{1/\lambda_{\min}} e^{-z/2} z^{k+r/2-1} dz \\
&= \sum_{k=N}^{\infty} \frac{1}{\Gamma(r/2)} \frac{(1 - 1/a_{\max})^k}{k!} \frac{1}{2^{k+r/2}} \int_0^{1/\lambda_{\min}} e^{-z/2} z^{k+r/2-1} dz.
\end{aligned} \tag{5.6}$$

Letting $z = \frac{2y}{(1 - 1/a_{\max})}$, we have, from equation (5.6) that

$$\begin{aligned}
& \sum_{k=N}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\
&\leq \frac{1}{\Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2}} \sum_{k=N}^{\infty} \frac{1}{k!} \int_0^{\frac{1}{2\lambda_{\min}} (1 - \frac{1}{a_{\max}})} e^{-\frac{y}{(1 - 1/a_{\max})}} y^{k+r/2-1} dy \\
&= \frac{1}{\Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2}} \frac{1}{2\lambda_{\min}} \left(1 - \frac{1}{a_{\max}}\right) \int_0^{\frac{1}{2\lambda_{\min}} (1 - \frac{1}{a_{\max}})} e^{-y/(1 - 1/a_{\max})} y^{r/2-1} dy \\
&\quad \cdot \sum_{k=N}^{\infty} \frac{y^k}{k!} dy
\end{aligned} \tag{5.7}$$

where the interchange of the integral and summation is allowed by the uniform convergence of the series. Observe that

$$\begin{aligned}
 \sum_{k=N}^{\infty} \frac{y^k}{k!} &= \frac{y^N}{N!} + \frac{y^{N+1}}{(N+1)!} + \frac{y^{N+2}}{(N+2)!} + \dots \\
 &= \frac{y^N}{N!} \left[1 + \frac{y}{N+1} + \frac{y^2}{(N+1)(N+2)} + \frac{y^3}{(N+1)(N+2)(N+3)} + \dots \right] \\
 &\leq \frac{y^N}{N!} \left[1 + \frac{y}{1!} + \frac{y^2}{2!} + \frac{y^3}{3!} + \dots \right] \\
 &= \frac{y^N}{N!} e^y.
 \end{aligned} \tag{5.8}$$

With equation (5.8), we reduce equation (5.7) to

$$\begin{aligned}
 &\sum_{k=N}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\
 &\leq \frac{1}{\Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2}} \int_0^{\frac{1}{2\lambda_{\min}} (1 - \frac{1}{a_{\max}})} e^{-y/(1-1/a_{\max})} \cdot y^{r/2-1} \frac{y^N}{N!} e^y dy \\
 &= \frac{1}{N! \Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2}} \int_0^{\frac{1-1/a_{\max}}{2\lambda_{\min}}} e^{-y/(a_{\max}-1)} y^{N+r/2-1} dy.
 \end{aligned} \tag{5.9}$$

Letting $y = (a_{\max}-1)/2$ in equation (5.9), we write

$$\begin{aligned}
& \sum_{k=N}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\
& \leq \frac{1}{N! \Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2}} \left(\frac{a_{\max} - 1}{2} \right)^{N+r/2} \int_0^{\frac{1}{\lambda_{\min} a_{\max}}} e^{-w/2} w^{N+r/2-1} dw.
\end{aligned}
\tag{5.10}$$

Recalling from theorem 5.1 that $a_j = \lambda_j / \lambda_{\min}$, we have

$$\lambda_{\min} a_{\max} = \lambda_{\max}. \tag{5.11}$$

Equation (5.10) is now reduced to

$$\begin{aligned}
& \sum_{k=N}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\
& \leq \frac{1}{N! \Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2}} \left(\frac{a_{\max} - 1}{2} \right)^{N+r/2} \int_0^{\frac{1}{\lambda_{\max}}} e^{-w/2} w^{N+r/2-1} dw \\
& = \frac{a_{\max}^{r/2} (a_{\max} - 1)^N}{N! \Gamma(\frac{r}{2}) 2^{N+r/2}} \int_0^{\frac{1}{\lambda_{\max}}} e^{-w/2} w^{N+r/2-1} dw \\
& = \frac{\Gamma(N + \frac{r}{2})}{\Gamma(\frac{r}{2})} \frac{a_{\max}^{N+r/2}}{N!} (1 - \frac{1}{a_{\max}})^N \frac{1}{2^{N+r/2} \Gamma(N + \frac{r}{2})} \int_0^{\frac{1}{\lambda_{\max}}} e^{-\frac{w}{2}} w^{N + \frac{r}{2} - 1} dw
\end{aligned}$$

$$= \frac{\Gamma(N+\frac{r}{2})}{\Gamma(\frac{r}{2})} \frac{a_{\max}^{N+r/2}}{N!} \left(1 - \frac{1}{a_{\max}}\right)^N \Pr(\chi_{r+2N}^2 \leq \frac{1}{\lambda_{\max}}). \quad (5.12)$$

Using equation (5.11), equation (5.12) becomes

$$\begin{aligned} & \sum_{k=N}^{\infty} c_k \Pr(\chi_{r+2N}^2 \leq \frac{1}{\lambda_{\min}}) \\ & \leq \frac{\Gamma(N+r/2)}{N! \Gamma(r/2)} (\lambda_{\max}/\lambda_{\min})^{N+r/2} (1 - \lambda_{\min}/\lambda_{\max})^N \Pr(\chi_{r+2N}^2 \leq \frac{1}{\lambda_{\max}}) \end{aligned}$$

which proves the theorem.

Theorem 5.4 An upper bound for the error in using a finite (N) number of terms to approximate the probability P is given by

$$\begin{aligned} & \left| \frac{\lambda_{\min}^{r/2}}{\sqrt{|A(V_1+V)|}} \sum_{k=N}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \right| \\ & \leq \frac{\Gamma(N+r/2)}{N! \Gamma(r/2)} \frac{\lambda_{\max}^{N+r/2}}{\lambda_{\min}^N \sqrt{|A(V_1+V)|}} (1 - \lambda_{\min}/\lambda_{\max})^N \Pr(\chi_{r+2N}^2 \leq \frac{1}{\lambda_{\max}}). \end{aligned}$$

Proof This follows immediately from theorem 5.3.

5.3 Interval for the Probability Content

Theorem 4.3 gives the exact probability that the use of prior information is beneficial. This expression is complicated. We now set up an upper bound and a lower bound for this probability.

Theorem 5.5 The probability P that the use of prior information is beneficial satisfies the inequality

$$\frac{\lambda_{\min}^{r/2}}{\sqrt{|A(V_1+V)|}} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\min}}) \leq P \leq \frac{\lambda_{\max}^{r/2}}{\sqrt{|A(V_1+V)|}} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\max}}).$$

Proof First, we find the upper bound for P . From theorem 5.1, we have

$$\begin{aligned} & \sum_{k=0}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\ & \leq \sum_{k=0}^{\infty} \frac{\Gamma(k+r/2)}{\Gamma(r/2)} \frac{(1-1/a_{\max})^k}{k!} \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\ & = \sum_{k=0}^{\infty} \frac{\Gamma(k+\frac{r}{2})}{\Gamma(\frac{r}{2})} \frac{(1-\frac{1}{a_{\max}})^k}{k!} \frac{1}{2^{k+r/2} \Gamma(k+r/2)} \int_0^{\frac{1}{\lambda_{\min}}} e^{-z/2} z^{k+r/2-1} dz \\ & = \sum_{k=0}^{\infty} \frac{(1-1/a_{\max})^k}{k!} \frac{1}{2^{k+r/2} \Gamma(r/2)} \int_0^{\frac{1}{\lambda_{\min}}} e^{-z/2} z^{k+r/2-1} dz. \end{aligned}$$

(5.13)

Let $z = \frac{2y}{(1-1/a_{\max})}$. Equation (5.13) becomes

$$\begin{aligned}
 & \sum_{k=0}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\
 & \leq \sum_{k=0}^{\infty} \frac{(1 - \frac{1}{a_{\max}})^k}{k!} \frac{1}{2^{k+r/2} \Gamma(\frac{r}{2})} \int_0^{\frac{1}{2\lambda_{\min}} (1 - \frac{1}{a_{\max}})} e^{-y/(1-1/a_{\max})} \\
 & \quad \cdot \left(\frac{2}{1-1/a_{\max}}\right)^{k+r/2} y^{k+r/2-1} dy \\
 & = \sum_{k=0}^{\infty} \frac{1}{\Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2} k!} \int_0^{\frac{1}{2\lambda_{\min}} (1 - \frac{1}{a_{\max}})} e^{-y/(1-1/a_{\max})} \\
 & \quad \cdot y^{k+r/2-1} dy \\
 & = \frac{1}{\Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2}} \int_0^{\frac{1}{2\lambda_{\min}} (1 - \frac{1}{a_{\max}})} e^{-y/(1-1/a_{\max})} \\
 & \quad \cdot y^{r/2-1} \sum_{k=0}^{\infty} \frac{y^k}{k!} dy
 \end{aligned} \tag{5.14}$$

where the interchange of summation and integration is allowed by the uniformity of convergence.

Equation (5.14) then becomes

$$\begin{aligned}
& \sum_{k=0}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\
& \leq \frac{1}{\Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2}} \int_0^{\frac{1}{2\lambda_{\min}} (1 - \frac{1}{a_{\max}})} e^{-y/(1-1/a_{\max})} y^{r/2-1} e^y dy \\
& = \frac{1}{\Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2}} \int_0^{\frac{1}{2\lambda_{\min}} (1 - \frac{1}{a_{\max}})} e^{-y/(a_{\max}^{-1})} y^{r/2-1} dy.
\end{aligned} \tag{5.15}$$

Let $y = (a_{\max}^{-1})w/2$. Equation (5.15) becomes

$$\begin{aligned}
& \sum_{k=0}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\
& \leq \frac{1}{\Gamma(\frac{r}{2}) (1 - \frac{1}{a_{\max}})^{r/2}} \int_0^{1/(\lambda_{\min} a_{\max})} e^{-w/2} \left(\frac{a_{\max}^{-1}}{2} \right)^{r/2} w^{r/2-1} dw \\
& = \frac{a_{\max}^{r/2}}{\Gamma(r/2) 2^{r/2}} \int_0^{1/\lambda_{\max}} e^{-w/2} w^{r/2-1} dw \\
& = a_{\max}^{r/2} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\max}})
\end{aligned} \tag{5.16}$$

where we have used equation (5.11), $\lambda_{\min} a_{\max} = \lambda_{\max}$.

Hence,

$$\begin{aligned}
P &= |A(V_1+V)|^{-1/2} \lambda_{\min}^{r/2} \sum_{k=0}^{\infty} c_k \Pr(\chi_{r+2k}^2 \leq \frac{1}{\lambda_{\min}}) \\
&\leq |A(V_1+V)|^{-1/2} \lambda_{\min}^{r/2} a_{\max}^{r/2} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\max}}) \\
&= |A(V_1+V)|^{-1/2} \lambda_{\max}^{r/2} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\max}})
\end{aligned}$$

where we have used equations (5.16) and (5.11).

Similarly we can obtain

$$P \geq |A(V_1+V)|^{-1/2} \lambda_{\min}^{r/2} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\min}})$$

by changing $a_{\max}$ to $a_{\min}$ in the proof. This completes the proof of theorem 5.5.

Remark In theorem 5.5, if $\lambda_{\max} = \lambda_{\min} = \lambda$, then

$$P = \Pr(\chi_r^2 \leq \frac{1}{\lambda}).$$

This agrees with corollary 4.2.

Theorem 5.6 In theorem 5.5, let I be the length of the interval between the upper bound and the lower bound, then

$$I = \frac{1}{\sqrt{|A(V_1+V)|}} \frac{1}{2^{r/2} \Gamma(\frac{r}{2})} \int_0^1 [e^{-z/(2\lambda_{\max})} - e^{-z/(2\lambda_{\min})}] z^{r/2-1} dz.$$

Proof From theorem 5.5,

$$\begin{aligned}
I &= \frac{\lambda_{\max}^{r/2}}{\sqrt{|A(V_1+V)|}} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\max}}) - \frac{\lambda_{\min}^{r/2}}{\sqrt{|A(V_1+V)|}} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\min}}) \\
&= \frac{1}{\sqrt{|A(V_1+V)|}} \frac{1}{2^{r/2} \Gamma(\frac{r}{2})} \left[\lambda_{\max}^{r/2} \int_0^{\frac{1}{\lambda_{\max}}} e^{-w/2} w^{r/2-1} dw \right. \\
&\quad \left. - \lambda_{\min}^{r/2} \int_0^{\frac{1}{\lambda_{\min}}} e^{-w/2} w^{r/2-1} dw \right].
\end{aligned} \tag{5.17}$$

Letting $w = z/\lambda_{\max}$ and $w = z/\lambda_{\min}$ in the two integrals in equation (5.17), respectively, we obtain

$$I = \frac{1}{\sqrt{|A(V_1+V)|}} \frac{1}{2^{r/2} \Gamma(\frac{r}{2})} \int_0^1 [e^{-z/(2\lambda_{\max})} - e^{-z/(2\lambda_{\min})}] z^{r/2-1} dz$$

which proves this theorem.

Remark In theorem 5.6, let $\lambda_{\max}/\lambda_{\min} = \delta$. Clearly $\delta \geq 1$.

We have the following three cases:

Case I. If $\delta = 1$, then $I = 0$ and

$$P = \Pr(\chi_r^2 \leq \frac{1}{\lambda})$$

where $\lambda = \lambda_{\max} = \lambda_{\min}$. This agrees with corollary 4.2.

Case II. If $\delta \approx 1$, then $I \approx 0$. This means that if all the eigenvalues of $A(V_1+V)$ are about equal, the length of the interval for the probability content is small.

Case III. If $\delta \gg 1$, then $\lambda_{\max} \gg \lambda_{\min}$. This implies that

$$e^{-z/(2\lambda_{\max})} \gg e^{-z/(2\lambda_{\min})}.$$

We can write

$$I \approx \frac{1}{\sqrt{|A(V_1+V)|}} \frac{1}{2^{r/2} \Gamma(\frac{r}{2})} \int_0^1 e^{-z/(2\lambda_{\max})} z^{r/2-1} dz.$$

Letting $z = \lambda_{\max} w$, we have

$$\begin{aligned} I &\approx \frac{1}{\sqrt{|A(V_1+V)|}} \frac{\lambda_{\max}^{r/2}}{2^{r/2} \Gamma(\frac{r}{2})} \int_0^{1/\lambda_{\max}} e^{-w/2} w^{r/2-1} dw \\ &= \frac{\lambda_{\max}^{r/2}}{\sqrt{|A(V_1+V)|}} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\max}}) \end{aligned}$$

which shows that when $\lambda_{\max} \gg \lambda_{\min}$, the length of the interval for the probability content is about the same as the upper bound in theorem 5.5.

In theorem 5.5, recalling that the matrix $A(V_1+V)$ has eigenvalues $\lambda_1, \dots, \lambda_r$, we can write

$$|A(V_1+V)| = \prod_{j=1}^r \lambda_j. \quad (5.18)$$

Hence, the upper bound varies inversely as $\lambda_{\min}^{1/2}$ and the lower bound varies inversely as $\lambda_{\max}^{1/2}$. We then have the following theorem.

Theorem 5.7 The upper bound in theorem 5.5 is a decreasing function of $\lambda_{\min}$ and the lower bound is a decreasing function of $\lambda_{\max}$.

Proof In theorem 5.5, the bounds are

$$\frac{\lambda_{\min}^{r/2}}{\sqrt{|A(V_1+V)|}} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\min}}) \leq P \leq \frac{\lambda_{\max}^{r/2}}{\sqrt{|A(V_1+V)|}} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\max}}).$$

Consider the lower bound. Let

$$g(\lambda_j) = \frac{\lambda_{\min}^{r/2}}{\sqrt{|A(V_1+V)|}} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\min}}), \quad j=1, \dots, r.$$

Then

$$\begin{aligned} & \frac{\partial}{\partial \lambda_{\max}} g(\lambda_j) \\ &= \lambda_{\min}^{r/2} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\min}}) \frac{\partial}{\partial \lambda_{\max}} \frac{1}{\sqrt{|A(V_1+V)|}} \\ &= -\frac{1}{2} \lambda_{\min}^{r/2} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\min}}) |A(V_1+V)|^{-3/2} \frac{\partial}{\partial \lambda_{\max}} |A(V_1+V)| \\ &= -\frac{1}{2} \lambda_{\min}^{r/2} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\min}}) |A(V_1+V)|^{-3/2} \frac{|A(V_1+V)|}{\lambda_{\max}} \\ &= -\frac{1}{2} \lambda_{\min}^{r/2} \Pr(\chi_r^2 \leq \frac{1}{\lambda_{\min}}) \frac{1}{\lambda_{\max} \sqrt{|A(V_1+V)|}} \\ &< 0 \end{aligned}$$

which proves that the lower bound is a decreasing function of $\lambda_{\max}$. The proof for the other part of this theorem is similar.

Remark From theorem 5.7, if $\lambda_{\max}$ decreases, then the lower bound increases; if $\lambda_{\min}$ increases, then the upper bound decreases. If $\lambda_{\max} = \lambda_{\min} = \lambda$, then

$$P = \Pr(\chi_r^2 \leq \frac{1}{\lambda}).$$

CHAPTER VI

OPTIMAL DESIGN PROCEDURES

6.1 Introduction

We now return to the design procedures for the model that we have considered in chapter III. The models are given in equations (3.1) and (3.4). There are two cases.

Case I. Prior data is taken.

In this case, the design matrix X_1 is known. We consider the problem of selecting positions in restricted regions for reading observations Y_2 to minimize the determinant of V_2 , the covariance matrix of the Gauss-Markoff estimate of β_2 in the present experiment. Optimal procedures for this design are discussed in sections 6.2, 6.3 and 6.5.

Case II. Prior data has not been taken.

In this case, we are able to design both prior and present experiments. The following two steps should be taken:

Step 1. Choose the design matrix X_2 such that $|V_2|$ is minimized. That is to say, we design the present experiment so that we will do as good as we can.

Step 2. With the optimal design matrix X_2 in step

1, we choose the design matrix X_1 such that the determinant of the mean square error given in equation (3.13) is minimized. That is, we will design the prior experiment so that prior information will help as as much as it can.

Optimal procedures for case II are discussed in sections 6.4 and 6.5.

We now consider the choice of design matrix in an ellipsoidal region.

6.2 Choice of Design Matrix in Ellipsoidal Regions

Consider the linear model

$$Y = X\beta + e \quad (6.1)$$

where $E(e) = 0$ and $E(ee') = \Sigma$.

If we let $X = (x_{ij})$, $i = 1, \dots, n$ and $j = 1, \dots, p$ and denote the i -th column of X by X_i , we will limit the range of the x_{ij} to the restricted regions of interest. We will derive a method to choose X such that $|\Sigma|$ is minimized.

The following lemma is well-known [9].

Lemma 6.1 Let B be an $n \times n$ matrix that is partitioned as follows:

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where B_{ij} has size $n_i \times n_j$, $i, j = 1, 2$, and $n_1 + n_2 = n$.

Then

$$|B| = |B_{11}| \cdot |B_{22} - B_{21}B_{11}^{-1}B_{12}|.$$

If we make a restriction of the form

$$x_i' \Sigma^{-1} x_i \leq d_i, \quad i = 1, 2, \dots, p \quad (6.2)$$

where d_i is given, then we can prove the following theorem.

Theorem 6.1 Let X be a design matrix and $\hat{\beta}$ the Gauss-Markoff minimum variance linear unbiased estimate of β . Then under the restriction in equation (6.2),

$$a) \quad |V(\hat{\beta})| \geq \frac{1}{\prod_{i=1}^p d_i}$$

b) the minimum is obtained when $x_i' \Sigma^{-1} x_j = 0$ for $i \neq j$ and $x_i' \Sigma^{-1} x_i = d_i$. That is, $|V(\hat{\beta})|$ is a minimum when

$$X' \Sigma^{-1} X = \begin{bmatrix} d_1 & & & \\ & \cdot & \phi & \\ & \phi & \cdot & \\ & & & d_p \end{bmatrix}.$$

Proof Consider the case where $X = [X_1, X_2]$, i.e., X has two columns, then $X' \Sigma^{-1} X$ can be written as

$$X' \Sigma^{-1} X = \begin{bmatrix} x_1' \Sigma^{-1} x_1 & x_1' \Sigma^{-1} x_2 \\ x_2' \Sigma^{-1} x_1 & x_2' \Sigma^{-1} x_2 \end{bmatrix}. \quad (6.3)$$

By lemma 6.1, we have

$$\begin{aligned} & |X' \Sigma^{-1} X| \\ &= |x_1' \Sigma^{-1} x_1| \quad |x_2' \Sigma^{-1} x_2 - x_2' \Sigma^{-1} x_1 (x_1' \Sigma^{-1} x_1)^{-1} x_1' \Sigma^{-1} x_2|. \end{aligned} \quad (6.4)$$

Therefore,

$$\begin{aligned} & |V(\hat{\beta})| \\ &= |X' \Sigma^{-1} X|^{-1} \\ &= \frac{1}{x_1' \Sigma^{-1} x_1 [x_2' \Sigma^{-1} x_2 - x_2' \Sigma^{-1} x_1 (x_1' \Sigma^{-1} x_1)^{-1} x_1' \Sigma^{-1} x_2]} \\ &\geq \frac{1}{x_1' \Sigma^{-1} x_1 \cdot x_2' \Sigma^{-1} x_2} \\ &\geq \frac{1}{d_1 d_2} \end{aligned} \quad (6.5)$$

since $x_2' \Sigma^{-1} x_1 (x_1' \Sigma^{-1} x_1)^{-1} x_1' \Sigma^{-1} x_2 \geq 0$,

where equality holds only if $x_1' \Sigma^{-1} x_2 = x_2' \Sigma^{-1} x_1 = 0$

and $x_1' \Sigma^{-1} x_1 = d_1$, $x_2' \Sigma^{-1} x_2 = d_2$.

Now assuming the theorem is true for

$$X = [X_1, X_2, \dots, X_k],$$

we can show by induction it is true for

$$X = [X_1, X_2, \dots, X_{k+1}].$$

$$\text{Let } (\beta)_k = (\beta_1, \beta_2, \dots, \beta_k)'$$

$$\text{Denote } V(\hat{\beta})_k = \text{covariance matrix of } (\hat{\beta})_k$$

$$\text{for } X = [X_1, X_2, \dots, X_k]$$

$$\text{and } V(\hat{\beta})_{k+1} = \text{covariance matrix of } (\hat{\beta})_{k+1}$$

$$\text{for } X = [X_1, X_2, \dots, X_{k+1}].$$

We assume

$$\text{a) } |V(\hat{\beta})_k| \geq \frac{1}{\prod_{i=1}^k d_i}$$

$$\text{b) the minimum is obtained when } X_i' \Sigma^{-1} X_j = 0 \text{ for } i \neq j$$

$$\text{and } X_i' \Sigma^{-1} X_i = d_i, \quad i = 1, 2, \dots, k.$$

We want to show that the above is true for

$$i = 1, 2, \dots, k+1.$$

$$V(\hat{\beta})_{k+1} = \{[X_1, \dots, X_{k+1}]' \Sigma^{-1} [X_1, \dots, X_{k+1}]\}^{-1}$$

$$= \begin{bmatrix} x_1' \Sigma^{-1} x_1 & \dots & x_1' \Sigma^{-1} x_k & x_1' \Sigma^{-1} x_{k+1} \\ \vdots & & \vdots & \vdots \\ x_k' \Sigma^{-1} x_1 & \dots & x_k' \Sigma^{-1} x_k & x_k' \Sigma^{-1} x_{k+1} \\ x_{k+1}' \Sigma^{-1} x_1 & \dots & x_{k+1}' \Sigma^{-1} x_k & x_{k+1}' \Sigma^{-1} x_{k+1} \end{bmatrix}^{-1} \quad (6.6)$$

Let

$$A_{11} = \begin{bmatrix} x_1' \Sigma^{-1} x_1 & \dots & x_1' \Sigma^{-1} x_k \\ \vdots & & \vdots \\ x_k' \Sigma^{-1} x_1 & \dots & x_k' \Sigma^{-1} x_k \end{bmatrix}$$

$$a_{12} = a_{21}' = \begin{bmatrix} x_1' \Sigma^{-1} x_{k+1} \\ \vdots \\ x_k' \Sigma^{-1} x_{k+1} \end{bmatrix}$$

and $a_{22} = x_{k+1}' \Sigma^{-1} x_{k+1}$. Equation (6.6) becomes

$$V(\hat{\beta})_{k+1} = \begin{bmatrix} A_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} \quad (6.7)$$

By lemma 6.1,

$$|V(\hat{\beta})_{k+1}| = \frac{1}{|A_{11}| |a_{22} - a_{21}A_{11}^{-1}a_{12}|}.$$

Since $A_{11} = V^{-1}(\hat{\beta})_k$ is positive definite, it follows that A_{11}^{-1} is positive definite, which implies that

$a_{21}A_{11}^{-1}a_{12} \geq 0$ with equality holding only if

$a_{12} = a_{21}' = 0$, in which case $X_i' \Sigma^{-1} X_{k+1} = 0$ for $i = 1, \dots, k$.

Thus it follows that equation (6.7) becomes

$$V(\hat{\beta})_{k+1} = \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & (X_{k+1}' \Sigma^{-1} X_{k+1})^{-1} \end{bmatrix},$$

$$|V(\hat{\beta})_{k+1}| = \frac{1}{|A_{11}| |X_{k+1}' \Sigma^{-1} X_{k+1}|}$$

$$= \frac{|V(\hat{\beta})_k|}{X_{k+1}' \Sigma^{-1} X_{k+1}}$$

$$\geq \frac{1}{\prod_{i=1}^k d_i} \cdot \frac{1}{d_{k+1}}$$

$$= \frac{1}{\prod_{i=1}^{k+1} d_i}.$$

This completes the proof.

Corollary 6.1 If $\Sigma = I$, then under the hypothesis of theorem 6.1, $|V(\hat{\beta})|$ is minimum when $X_i'X_j = 0$ for $i \neq j$ and $X_i'X_i = d_i$.

Proof This follows immediately from theorem 6.1.

6.3 On Selecting Sample Points from an Interval

Let us now consider a model such as

$$y_i = \beta_1 + \beta_2 x_i + e_i, \quad i=1,2,\dots,n \quad (6.8)$$

or in matrix form

$$Y = X\beta + e$$

where

$$Y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{bmatrix}$$

Furthermore, we assume $E(e) = 0$, $E(ee') = \sigma^2 I$ and that $x_i \in [-a, a]$.

By taking n observations in $[-a, a]$ we estimate β using the Gauss-Markoff estimator $\hat{\beta}$. However, we would like to choose the x_i 's so that $|V(\hat{\beta})|$ is a minimum.

Theorem 6.2 Under the model given in equation (6.8), if $x_i \in [-a, a]$, $i = 1, 2, \dots, n$, (n even), then in order to minimize $|V(\hat{\beta})|$, $n/2$ of the x_i 's must be equal to a , and $n/2$ of them must be equal to $-a$.

Proof If

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

then

$$X'X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

which implies

$$V(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \begin{bmatrix} \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2} & \frac{-\sigma^2 \sum x_i}{n \sum (x_i - \bar{x})^2} \\ \frac{-\sigma^2 \sum x_i}{n \sum (x_i - \bar{x})^2} & \frac{\sigma^2 n}{n \sum (x_i - \bar{x})^2} \end{bmatrix}$$

and

$$|V(\hat{\beta})| = \frac{\sigma^2}{|X'X|} = \frac{\sigma^2}{n \sum (x_i - \bar{x})^2}.$$

Hence $|V(\hat{\beta})|$ is minimized when the x_i 's are chosen so that $\bar{x} = 0$ and x_i^2 's as large as possible for each i . This condition implies that the observations are taken at the endpoints of the interval $[-a, a]$, but for each $x_i = a$ we must have one equal to $-a$ since $\bar{x} = 0$. Hence we readily see that $n/2$ of the observations are taken at each endpoint.

Corollary 6.2 If in the previous theorem n is odd, then $\frac{n-1}{2}$ of the observations are taken at $x_i = -a$, $\frac{n-1}{2}$ taken at $x_i = a$, and one taken at $x_i = 0$.

Proof Assume $\sum_{i=1}^n x_i = 0$, where $|x_i| \leq a$ for each i , and

that $\sum_{i=1}^n x_i^2$ is a maximum where at least two of the

observations are different from the endpoints, say x_1 and x_2 . Without loss of generality we assume $0 \leq x_1 \leq x_2 \leq a$.

Let $d = |a - x_2|$.

Define $x_1' = x_1 - d$

$x_2' = x_2 + d$

$x_i' = x_i$ for $i > 2$.

Now note that $\sum_{i=1}^n x_i' = \sum_{i=1}^n x_i + d - d = 0$

and that
$$\sum_{i=1}^n (x_i')^2 = \sum_{i=1}^n x_i^2 + \delta$$

where $\delta = 2d^2 + (2x_2d - 2x_1d) > 0$.

Thus it follows that
$$\sum_{i=1}^n (x_i')^2 > \sum_{i=1}^n x_i^2 .$$
 However $\sum x_i^2$

was assumed to be a maximum. Therefore we have a contradiction to the fact that at least two of the x_i 's are different from the endpoints. Obviously, if $x_i = \pm a$ for every i , then $\sum x_i \neq 0$ since n is odd. Hence we conclude that exactly one of the x_i 's must differ from $\pm a$ and, in fact must be equal to zero since we require $\sum x_i = 0$, i.e., $\bar{x} = 0$. Thus we have one $x_i = 0$ and from theorem 6.2, $\frac{n-1}{2}$ of the $x_i = a$ and $\frac{n-1}{2}$ of the $x_i = -a$.

The following theorem, the statement of which is due to Graybill [10], will be needed for the proofs of some succeeding theorems.

Theorem 6.3 Using the linear model $Y = X\beta + e$ with $E(ee') = \sigma^2 I$, the best linear unbiased estimate of any linear combination of the β_i 's is the same linear combination of the best linear unbiased estimate of the β_i 's, i.e., the best linear unbiased estimate of $t'\beta$ (t is a $p \times 1$ known vector of constants) is $t'\hat{\beta}$ where $\hat{\beta}$ is the best linear unbiased estimate of β , i.e.,

$$t'\hat{\beta} = t'S^{-1}X'Y$$

where $S^{-1} = (X'X)^{-1}$.

A question that now arises is "What if the x_i 's are located in an interval $[a, b]$?". The following theorem will answer this question.

Theorem 6.4 (Translation) Suppose we have the model

$$y_i = \beta_1 + \beta_2 z_i + e_i$$

where $z_i \in [a, b]$ for $i = 1, 2, \dots, n$, (n even), then the observations are taken just as they were in theorem 6.2, i.e., $n/2$ of them are taken at $z_i = a$ and $n/2$ of them are taken at $z_i = b$.

Proof Consider the model

$$y_i = \beta_1 + \beta_2 z_i + e_i, \quad z_i \in [a, b].$$

Now let $z_i = x_i + (a+b)/2$ where $x_i \in [-(b-a)/2, (b-a)/2]$ and therefore

$$\begin{aligned} y_i &= \beta_1 + \beta_2 z_i + e_i \\ &= \beta_1 + \beta_2 [x_i + (a+b)/2] + e_i \\ &= \beta_1 + \beta_2 (a+b)/2 + \beta_2 x_i + e_i \\ &= \beta_3 + \beta_2 x_i + e_i \end{aligned}$$

where $\beta_3 = \beta_1 + \frac{a+b}{2} \beta_2$ and $x_2 \in [-(b-a)/2, (b-a)/2]$.

Note that $\beta_1 = \beta_3 - \frac{a+b}{2} \beta_2$ and that from theorem 6.3, the best linear unbiased estimate of β_1 , say $\hat{\beta}_1$ is

$$\hat{\beta}_1 = \hat{\beta}_3 - \frac{a+b}{2} \hat{\beta}_2$$

where $\hat{\beta}_3$ and $\hat{\beta}_2$ are the best linear unbiased estimate of β_3 and β_2 , respectively. However from theorem 6.2, $|V(\hat{\beta})|$

where $\hat{\beta} = \begin{bmatrix} \hat{\beta}_3 \\ \hat{\beta}_2 \end{bmatrix}$, is minimized when $n/2$ of x_i 's are equal

to $-\frac{b-a}{2}$, and $n/2$ of the x_i 's are equal to $\frac{b-a}{2}$. Thus

$|V(\hat{\beta})|$ where $\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$ takes on its minimum whenever $n/2$

of the z_i 's are equal to a and $n/2$ of them equal to b .

Thus the desired result is obtained.

Corollary 6.3 In theorem 6.4 if n is odd, then $(n-1)/2$ of the observations are taken at $z_i = a$, $(n-1)/2$ taken at $z_i = b$, and one taken at $z_i = (a+b)/2$.

Proof Similar to corollary 6.2.

6.4 Minimization of Mean Square Error

When prior data has not been taken, we choose X_2 so that $|V_2|$ is minimized. With this V_2 , we shall find the design matrix X_1 to minimize the mean square error given

in equation (3.13) in the sense that $|\text{MSE}(\hat{\beta}_2)|$ is a minimum.

The following lemma is a familiar result [9].

Lemma 6.2 If A is an $m \times n$ matrix and B is an $n \times m$ matrix such that $I + BA$ is nonsingular, then

(i) $I + AB$ is also nonsingular, and

(ii) $(I + AB)^{-1} = I - A(I + BA)^{-1}B$.

Theorem 6.5 $|\text{MSE}(\hat{\beta}_2)|$ is minimized if $V_1 = 4V - V_2$.

Proof From equation (3.13),

$$\text{MSE}(\hat{\beta}_2) = (I + V_1 V_2^{-1})^{-1} (2V + V_1 + V_1 V_2^{-1} V_1) (I + V_2^{-1} V_1)^{-1}. \quad (6.9)$$

Minimizing $|\text{MSE}(\hat{\beta}_2)|$ is equivalent to minimizing $\ln |\text{MSE}(\hat{\beta}_2)|$. From equation (6.9) we obtain

$$\ln |\text{MSE}(\hat{\beta}_2)| = -2 \ln |I + V_1 V_2^{-1}| + \ln |2V + V_1 + V_1 V_2^{-1} V_1| \quad (6.10)$$

since $|I + V_1 V_2^{-1}| = |I + V_2^{-1} V_1|$.

Differentiating equation (6.10) with respect to V_1 , we have

$$\begin{aligned} & \frac{\partial}{\partial V_1} \ln |\text{MSE}(\hat{\beta}_2)| \\ &= -2 \frac{\partial}{\partial V_1} \ln |I + V_1 V_2^{-1}| + \frac{\partial}{\partial V_1} \ln |2V + V_1 + V_1 V_2^{-1} V_1|. \end{aligned} \quad (6.11)$$

Equating $\frac{\partial}{\partial V_1} \ln |\text{MSE}(\hat{\beta}_2)| = 0$, we have

$$2 \frac{\partial}{\partial V_1} \ln |I + V_1 V_2^{-1}| = \frac{\partial}{\partial V_1} \ln |2V + V_1 + V_1 V_2^{-1} V_1|. \quad (6.12)$$

By theorems 2.6 and 2.10, equation (6.12) can be written as

$$\begin{aligned} & 2(I + V_2^{-1} V_1)^{-1} V_2^{-1} - \text{diag } (I + V_2^{-1} V_1)^{-1} V_2^{-1} \\ &= (2V + V_1 + V_1 V_2^{-1} V_1)^{-1} V_1 V_2^{-1} + V_2^{-1} V_1 (2V + V_1 + V_1 V_2^{-1} V_1)^{-1} \\ &\quad - \text{diag } (2V + V_1 + V_1 V_2^{-1} V_1)^{-1} V_1 V_2^{-1} + (2V + V_1 + V_1 V_2^{-1} V_1)^{-1} \\ &\quad - \frac{1}{2} \text{diag } (2V + V_1 V_2^{-1} V_1)^{-1}. \end{aligned} \quad (6.13)$$

Taking diagonals on both sides of equation (6.13), we obtain

$$\begin{aligned} & \text{diag } (I + V_2^{-1} V_1)^{-1} V_2^{-1} \\ &= \text{diag } (2V + V_1 + V_1 V_2^{-1} V_1)^{-1} V_1 V_2^{-1} + \frac{1}{2} \text{diag } (2V + V_1 + V_1 V_2^{-1} V_1)^{-1}. \end{aligned} \quad (6.14)$$

Adding equations (6.13) and (6.14), we have

$$\begin{aligned} & 2(I + V_2^{-1} V_1)^{-1} V_2^{-1} \\ &= (2V + V_1 + V_1 V_2^{-1} V_1)^{-1} V_1 V_2^{-1} + V_2^{-1} V_1 (2V + V_1 + V_1 V_2^{-1} V_1)^{-1} \\ &\quad + (2V + V_1 + V_1 V_2^{-1} V_1)^{-1}, \end{aligned}$$

or

$$\begin{aligned}
 & 2(V_1 + V_2)^{-1} \\
 &= (V_1 + V_2 + 2V_2 V_1^{-1} V)^{-1} + (V_1 + V_2 + 2V V_1^{-1} V_2)^{-1} \\
 &\quad + (2V + V_1 + V_1 V_2^{-1} V_1)^{-1}. \tag{6.15}
 \end{aligned}$$

The V_1 that minimizes $|\text{MSE}(\hat{\beta}_2)|$ is the V_1 that solves equation (6.15). We claim that the solution is

$$V_1 = 4V - V_2. \tag{6.16}$$

To show this, we substitute equation (6.16) into equation (6.15), the left hand side (L.H.S.) and the right hand side (R.H.S.) are, respectively,

$$\text{L.H.S.} = 2(4V)^{-1} = \frac{1}{2} V^{-1}, \text{ and}$$

$$\begin{aligned}
 \text{R.H.S.} &= [4V + 2V_2(4V - V_2)^{-1}V]^{-1} + [4V + 2V(4V - V_2)^{-1}V_2]^{-1} \\
 &\quad + [2V + 4V - V_2 + (4V - V_2)V_2^{-1}(4V - V_2)]^{-1} \\
 &= \frac{1}{4} [V + \frac{1}{2} V_2(4V - V_2)^{-1}V]^{-1} + \frac{1}{4} [V + \frac{1}{2} V(4V - V_2)^{-1}V_2]^{-1} \\
 &\quad + [16VV_2^{-1}V - 2V]^{-1}. \tag{6.17}
 \end{aligned}$$

We now pre- and post- multiply the R.H.S. by $V^{1/2}$,

$$\begin{aligned}
& V^{1/2} (\text{R.H.S.}) V^{1/2} \\
&= \frac{1}{4} [I + \frac{1}{2} V^{-1/2} V_2 (4V - V_2)^{-1} V^{1/2}]^{-1} \\
&\quad + \frac{1}{4} [I + \frac{1}{2} V^{1/2} (4V - V_2)^{-1} V_2 V^{-1/2}]^{-1} \\
&\quad + \frac{1}{2} [8V^{1/2} V_2^{-1} V^{1/2} - I]^{-1}.
\end{aligned} \tag{6.18}$$

Consider the first parenthesis in the right hand side of equation (6.18), by lemma 6.2 with $A \leftrightarrow \frac{1}{2} V^{-1/2} V_2$ and $B \leftrightarrow (4V - V_2)^{-1} V^{1/2}$, we have

$$\begin{aligned}
& [I + \frac{1}{2} V^{-1/2} V_2 (4V - V_2)^{-1} V^{1/2}]^{-1} \\
&= I - \frac{1}{2} V^{-1/2} V_2 [I + \frac{1}{2} (4V - V_2)^{-1} V_2]^{-1} (4V - V_2)^{-1} V^{1/2} \\
&= I - \frac{1}{2} V^{-1/2} V_2 [4V - \frac{1}{2} V_2]^{-1} V^{1/2} \\
&= I - \frac{1}{2} V^{-1/2} [4V V_2^{-1} - \frac{1}{2} I]^{-1} V^{1/2} \\
&= I - \frac{1}{2} [4V^{1/2} V_2^{-1} V^{1/2} - \frac{1}{2} I]^{-1} \\
&= I - [8V^{1/2} V_2^{-1} V^{1/2} - I]^{-1}
\end{aligned} \tag{6.19}$$

Similarly, the second parenthesis in the right hand side of equation (6.18) will reduce to

$$[I + \frac{1}{2} V^{1/2} (4V - V_2)^{-1} V_2 V^{-1/2}]^{-1} = I - [8V^{1/2} V_2^{-1} V^{1/2} - I]^{-1}. \quad (6.20)$$

Hence, combining equations (6.18), (6.19) and (6.20) we have

$$\begin{aligned} & V^{1/2} (\text{R.H.S.}) V^{1/2} \\ &= \frac{1}{4} [I - (8V^{1/2} V_2^{-1} V^{1/2} - I)^{-1}] \\ &\quad + \frac{1}{4} [I - (8V^{1/2} V_2^{-1} V^{1/2} - I)^{-1}] \\ &\quad + \frac{1}{2} [8V^{1/2} V_2^{-1} V^{1/2} - I]^{-1} \\ &= \frac{1}{2} I. \end{aligned}$$

Therefore, we conclude that

$$\text{R.H.S.} = \frac{1}{2} V^{-1} = \text{L.H.S.}$$

This completes the proof of the theorem.

6.5 Choice of Design Matrices X_1 and X_2

We conclude this chapter with the optimal design procedures for the linear models in equations (3.1) and (3.4).

Case I. When prior data has been taken, the choice of X_2 in ellipsoidal regions is

$$X_2' \Sigma_2^{-1} X_2 = \begin{bmatrix} d_1 & & & \phi \\ & \cdot & & \\ & & \cdot & \\ \phi & & & d_p \end{bmatrix} \quad (6.21)$$

where the d_i 's are given constants such that

$$X_{2i}' \Sigma_2^{-1} X_{2i} \leq d_i$$

with X_{2i} representing the i -th column of X_2 .

The choice of X_2 in rectangular regions is described in section 6.3.

Case II. When prior data has not been taken, we first choose X_2 as described in case I. With this X_2 , we choose X_1 such that $|\text{MSE}(\hat{\beta}_2)|$ is minimized.

By theorem 6.5, X_1 should be chosen so that

$$V_1 = 4V - V_2, \quad (6.22)$$

provided that $4V - V_2$ is positive definite.

Recalling from equation (3.3) that

$$V_1 = (X_1' \Sigma_1^{-1} X_1)^{-1},$$

equation (6.22) becomes

$$X_1' \Sigma_1^{-1} X_1 = (4V - V_2)^{-1}. \quad (6.23)$$

Writting $\Sigma_1 = \Sigma_1^{1/2} \Sigma_1^{1/2}$, equation (6.23) becomes

$$\underset{\text{rxp}}{X_1} \underset{\text{pxp}}{\Sigma_1^{-1/2}} \underset{\text{pxp}}{\Sigma_1^{-1/2}} \underset{\text{pxr}}{X_1} = \underset{\text{rxr}}{(4V - V_2)^{-1}}. \quad (6.24)$$

Assuming that X_1 is full column rank, i.e., $p \geq r$, a particular solution to equation (6.24) is

$$\underset{\text{pxr}}{X_1} = \underset{\text{pxp}}{\Sigma_1^{1/2}} \begin{bmatrix} (4V - V_2)^{-1/2} \\ 0 \end{bmatrix} \begin{matrix} \text{rxr} \\ (p-r)\text{xr} \end{matrix}.$$

CHAPTER VII

SIMULATION RESULTS AND SUMMARY

6.1 Examples

We will give two numerical examples to demonstrate the major results of the text.

Example 6.1

Consider the situation where

$$\bar{\beta} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad X_1 = X_2 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad \Sigma_1 = \Sigma_2 = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}.$$

Based on 500 Y_1 's, 250 Y_2 's and five different V 's, the simulation results are summarized in Table 6.1.

Note that in Table 6.1, the upper bound is a decreasing function of $\lambda_{\min}$, and the lower bound is a decreasing function of $\lambda_{\max}$. When $\lambda_{\min} \approx \lambda_{\max}$, we have $LB \approx P \approx UB$. The probability P is high when $\lambda_{\min}$ is small.

TABLE 6.1

SIMULATION RESULTS FOR FIVE $\lambda_{\min}$'s

V	$\lambda_{\min}$	$\lambda_{\max}$	LB	P	UB	EB	EP
$\begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$	.8462	7.0000	.1551	.1758	.1983	0	.296
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	.4118	.6000	.5284	.6310	.6825	$.55 \times 10^{-5}$	.666
$\begin{bmatrix} .5 & 0 \\ 0 & .5 \end{bmatrix}$	.3714	.4545	.6687	.7027	.7380	$.60 \times 10^{-6}$	.722
$\begin{bmatrix} .1 & 0 \\ 0 & .1 \end{bmatrix}$	.3408	.3559	.7529	.7620	.7712	$.58 \times 10^{-9}$	.762
$\begin{bmatrix} .01 & 0 \\ 0 & .01 \end{bmatrix}$	.3341	.3356	.7744	.7754	.7764	$.75 \times 10^{-14}$	.782

P is the probability that the use of prior information is beneficial.
 LB and UB are lower bound and upper bound for P, respectively.

EB is the upper bound for the error in using 5 terms to approximate P.

EP is the empirical probability defined as $\hat{a}$, the ratio of the number of positive definite matrices generated for V_2 -MSE($\beta_2|Y_1$) and the total number of

matrices generated for V_2 -MSE($\hat{\beta}_2|Y_1$).

Example 6.2

Consider the model

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta_2 + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

where Y_1 , Y_2 , e_1 and e_2 are 10×1 matrices, respectively;

and $\beta_2 = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$.

Assume that

$$E(e_1) = E(e_2) = 0,$$

$$E(e_1 e_1') = E(e_2 e_2') = 4I$$

and that e_1 , e_2 are independent.

For

$$\bar{\beta} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad V = 0.2I, \quad X_1 = X_2 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{bmatrix},$$

we have the following simulated results based on 500 Y_1 's and 250 Y_2 's.

$\lambda_{\min}$	$\lambda_{\max}$	LB	P	UB	EB	EP
.4591	.6270	.5676	.6041	.6422	$.14 \times 10^{-5}$	.992

where the terms LB, P, ..., EP are defined in Table 6.1.
This is a non-optimal procedure.

An optimal choice of X_2 such that $|E(e_2 e_2')|$ is a minimum is

$$X_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix}$$

(10x2)

and an optimal choice of X_1 such that $|MSE(\hat{\beta}_2)|$ is a minimum is

$$X_1 = \begin{bmatrix} 3.162 & 0 \\ 0 & 3.162 \\ \phi & \phi \end{bmatrix}$$

(10x2)

With this optimal choice, we have

$\lambda_{\min}$	$\lambda_{\max}$	LB	P	UB	EB	EP
.6	.6	.5654	.5654	.5654	0	.716

Note that the probability P decreases since we have designed X_2 to do the best, so the chance that the use of

Y_1 is beneficial becomes smaller.

6.2 Summary

For the linear model $Y = X\beta + e$ where $e \sim N(0, \Sigma)$ and $\beta \sim N(\bar{\beta}, V)$, we have considered the incorporation of prior data with present data to obtain a better estimate of the present value of β . In chapter III, a method was derived to combine two sets of data to obtain an estimate of the value of β in one of the experiments. This estimator is biased. Its mean square error is compared with the covariance matrix of the standard estimator.

In chapter IV, the probability that the use of prior information is beneficial was obtained. The computation is, in general, complicated. Therefore, in chapter V, both upper and lower bounds for this probability were found. Monotonic properties of the bounds were also discussed.

In chapter VI, optimal design procedures were developed. The experimenter should design the present experiment to do as well as possible. If prior data Y_1 are given, he should use the probability content as a criterion to determine whether or not Y_1 should be used. If Y_1 have not been taken, he can always design his prior experiment so that prior information is helpful.

Some simulation results are given in chapter VII.

LIST OF REFERENCES

- [1] Anderson, R. L., The use of prior information in regression analysis, Essays in Honor of Gerhard Tinter, Chapter 13, Springer-Verlag, New York, 1969.
- [2] Anderson, R. L., and Battiste, E. L., The use of prior information in linear regression analysis, Communications in Statistics, Vol. 4, No. 6, pp. 497-517, 1975.
- [3] Battiste, E. L., Prior information and ill-condition in regression system, Unpublished Ph.D. dissertation, North Carolina State University at Raleigh, Institute of Statistics, Mimeo series no. 534, 1967.
- [4] Cooper, Richard A., Optimal designs for multi-linear regression, Ph.D. dissertation, Texas Tech University, 1970.
- [5] Deemer, Walter L., and Olkin, Ingram, The Jacobians of certain matrix transformations useful in multivariate analysis, Biometrika, Vol. 38, pp. 345-367, 1951.
- [6] Dwyer, Paul S., and MacPhail, M. S., Symbolic matrix derivatives, Ann. Math. Statist., Vol. 19, No. 4, pp. 517-534, 1948.
- [7] Dwyer, Paul S., Some applications of matrix derivatives in multivariate analysis, J., Amer. Statist. Assoc., Vol. 62, pp. 607-625, 1967.
- [8] Gould, Henry W., Combinatorial identities, Morgantown Printing and Binding Co., Morgantown, W. Va., 1973.
- [9] Graybill, F. A., Introduction to matrices with applications in statistics, Wadsworth, Belmont, Calif., 1969.
- [10] Graybill, F. A., An introduction to linear statistical models, Vol. I, McGraw-Hill, New York, 1961.

- [11] Johnson, Norman L., and Kotz, Samuel, Distribution in statistics - continuous univariate distribution - 2, John Wiley & Sons, New York, 1970.
- [12] Kendall, M. G., The advanced theory of statistics, Vol. I, Hafner, New York, 1968.
- [13] Lewis, T. O., and Odell, P. L., Estimation in linear models, Prentice-Hall, Englewood Cliffs, New Jersey, 1971.
- [14] Liebelt, Paul B., An introduction to optimal estimation, Addison-Wesley, Reading, Massachusetts, 1967.
- [15] Lukacs, E., & Laha, R. G., Applications of characteristic functions, Hafner, New York, 1964.
- [16] Reynolds, Roy F. Jr., Selection of the design matrix in linear models, Master's thesis, Texas Tech University, 1969.
- [17] Riordan, John, Combinatorial identities, John Wiley & Sons, New York, 1968.
- [18] Robbins, H. E., The distribution of a definite quadratic form, Annals of Mathematical Statistics, 19, pp. 266-270, 1948.
- [19] Robbins, H. E., and Pitman, E. J. G., Application of the method of mixtures to quadratic forms in normal variates, Annals of Mathematical Statistics, 20, pp. 552-560, 1949.
- [20] Ruben, H., Probability content of regions under spherical normal distribution, I, Annals of mathematical statistics, 31, pp. 598-619, 1960.
- [21] Ruben, H., Probability content of regions under spherical normal distributions, IV, Annals of Mathematical Statistics, 33, pp. 542-570, 1962.
- [22] Ruben, H., A new result on the distribution of quadratic forms, Annals of Mathematical Statistics, 34, pp. 1582-1584, 1963.
- [23] Tracy, D. S., and Dwyer, P. S., Multivariate maxima and minima with matrix derivatives, J., Amer. Statist. Assoc., Vol. 64, pp. 1576-1594, 1969.

- [24] Watkins, T. A., On estimating a function of the mean of a multivariate normal distribution, to appear in Communications in Statistics.
- [25] Watkins, T. A., A note on a new combinatorial identity, to appear in Texas Journal of Science.