ESTIMATION ON RESTRICTED PARAMETER SPACES

by

William Franklin Watson, Jr.

Institute of Statistics 33
Mimeograph Series No. 10~~25~~
Raleigh, N. C.

ABSTRACT

WATSON, WILLIAM FRANKLIN, JR. Estimation on Restricted Parameter Spaces. (Under the direction of HUBERTUS ROBERT VAN DER VAART and BENEE FRANK SWINDEL.)

The problem of finding point estimates of parameters when the feasible parameter space is a proper and convex subset of Euclidean m-space was studied. The algorithms of maximum likelihood estimation for the parameters of linear models, restricted in such a manner, were reviewed for the case in which the elements of the error vector have a normal distribution. These estimators were shown to be biased, to possess a type of consistency, and, in the univariate case, to have a mean square error no larger than the unrestricted maximum likelihood estimator. Also, these estimators were shown to map all unrestricted estimates which are not in the feasible parameter space to the boundary of the feasible parameter space. It would be difficult to believe that the parameter is on the boundary so often.

The Bayesian estimators, the median and mean of the posterior distribution, were shown to have different unpleasant properties when the parameter space is a proper, convex subset in Euclidean m-space. The median of the posterior distribution was found to take on points on the boundary of the feasible parameter space only if a supporting hyperplane of the posterior contained at least half of the probability mass of the posterior distribution. Similarly, the mean of the posterior distribution would never take on some of the points in the feasible parameter space as estimates unless the posterior

distribution tended to a dengenerate distribution at these points for some point in the sample space.

However, the mean of the univariate and a bivariate truncated normal posterior distribution, were, shown to take on every point in the support of the posterior for some value of the random variable. Assuming the prior density to be proportional to either a uniform, exponential, or truncated normal density over the feasible space, zero elsewhere, lead to a truncated normal posterior when the random variable was distributed normally.

A detailed examination was made of the estimators for the mean parameter of a univariate normal distribution for the situation in which the parameter was known to be contained in a half-line. Neither the mean of appropriate truncated normal posteriors using any of the priors mentioned above nor the restricted maximum likelihood estimators had uniformly smaller mean square error over the feasible parameter space. The regret function was then introduced and was defined to be the difference in the mean square error of an estimator at a point in parameter space and the smallest mean square error of the candidate estimators of that point. The strategy chosen was to find an estimator which would minimize, among the candidate estimators, the maximum regret over the sample space. Joined estimation procedures were proposed in which the mean of a posterior (exponential prior) was used over a portion of the sample space and maximum likelihood procedures were used over the remainder of the sample space. An optimal joined estimator was found to give an 18% reduction in maximum regret over the best of the classical estimators.

To extend the technique, optimal Bayesian estimators of this type were found for several subsets of the sample space. The resulting estimator gave a 48% reduction in the maximum regret over what was found for the best of the classical estimators. Similar results were found for a bivariate example.

ESTIMATION ON RESTRICTED PARAMETER SPACES

by

WILLIAM FRANKLIN WATSON JR.

A thesis submitted to the Graduate Faculty of
North Carolina State University at Raleigh
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

DEPARTMENTS OF FORESTRY AND STATISTICS

RALEIGH

1 9 7 4

APPROVED BY:

Co-Chairman of Advisory Committee

Co-Chairman of Advisory Committee

BIOGRAPHY

William F. Watson Jr. was born September 18, 1945, in Tifton, Georgia, and was raised in the farming community of Eldorado which is near Tifton. He received his elementary and secondary education in the Tift County, Georgia, school system and was graduated from Tift County High School in 1963.

He attended Abraham Baldwin Agricultural College, Auburn University, and the University of Georgia. From the latter, he received a Bachelor of Science degree in forestry in 1967 and the Master of Science degree in 1969.

In 1969 he was inducted into the U.S. Army where he served as a computer programmer and systems analyst for the 4th and 23rd Infantry Divisions in the Republic of Viet Nam. Upon his release from active duty in 1971, he entered North Carolina State University to pursue the Doctor of Philosophy degree. In 1974, he assumed a research position with the Forestry Department at Mississippi State University.

The author is married to the former Linda Diane Hamlin and they have one son, Hank.

ACKNOWLEDGMENTS

Any expression of gratitude would be insufficient for the Co-Chairmen of this author's Advisory Committee, Professors' H. R. van der Vaart and B. F. Swindel. These gentlemen were extremely generous with their time, and provided the counsel which inspired many of the significant portions of this study. A special word of thanks is due Professor W. L. Hafley who served as guide through the administrative hurdles while serving as Co-Chairman of the Advisory Committee. The author wishes to also thank the other members of the Graduate Faculty who served on his committee, T. M. Gerig, T. E. Maki, and T. O. Perry; all of whom made significant contributions to his educational experience. Professor J. M. Danby's suggestions dealing with the problems of numerical integration encountered in this study were also appreciated.

The author was supported during his graduate study by the North Carolina Agricultural Experiment Station.

Sincere gratitude is extended to the author's wife and son for their sacrifices, and especially to his wife for her efforts in the completion of this paper.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

1. INTRODUCTION

1.1 The Problem

The statistician is often confronted with the problem of estimating the parameters for the linear model

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon} \,. \tag{1.1.1}$$

In this problem, $\underline{y}$ is an n element vector of responses, X is an $n \times m$ design matrix, $\underline{\beta}$ is an m element vector containing the unknown parameters, and $\underline{\varepsilon}$ is an n element vector of the random components. There are many situations where the true value of the parameter vector $\underline{\beta}$ is known to lie in a proper subset of R^m (Euclidean m-space). Often such information can be written as linear inequality constraints in terms of the parameter vector $\underline{\beta}$. An example of such restrictions is

$$C\underline{\beta} \geq \underline{d} \,, \tag{1.1.2}$$

where C is a matrix of order $k \times m$, and $\underline{d}$ is a vector of k elements. Not all restrictions are of this simple form: Hudson (1969) cites a case of polynomial regression where the derivative of the polynomial must be positive over an interval.

Modelers of growth in biological population can often rule out subsets of the parameter space, R^m, because values of the parameters in these sets would violate known biological laws. An example of such a violation would be a model for the amount of wood fiber accumulated in the bole of a tree at various ages with parameters which give decreasing predictions of fiber accumulation over age.

It would be desirable if statisticians could prescribe a uniform set of rules for the modeler who has prior information that the true value of the parameters are certain to be found in a proper subset of R^m. Unfortunately, such a set of rules has not been forthcoming. Searle (1971) has listed the alternatives that have been proposed to resolve the problem of negative variance components. Many of these alternatives are applicable to the problem of estimation when the true values of the parameters of a linear model are known to be in a subset of R^m.

Some statisticians view estimates which violate known constraints as being indicative of a failure of the model to represent the true situation, and investigate alternative formulations of the model. Others choose to ignore occasional violations of the constraints if the unrestricted estimates possess good properties otherwise. Another group prefers to incorporate the restrictions into the estimating process. They realize that infeasible estimates are otherwise possible even when the model is correct due to the randomness of sampling and the construction of the estimators.

Even those who agree to incorporate the restrictions on the parameter space in the estimation procedure do not agree on which estimation procedure best takes advantage of this additional information. Many statisticians feel that maximizing the likelihood function over the set of feasible parameters is the most desirable alternative. The Bayesians, however, suggest that prior probabilities should be assigned to the elements of the feasible parameter space, and classical Bayesian techniques be invoked. Actually, we will see that each of these alternatives has discouraging properties.

1.2 Terminology

Several expressions have been used to describe the subset of R^m in which the true value of the parameter is known to lie. This subset will usually be called the *restricted* or *feasible parameter space*, occasionally simply the *parameter space*.

The unrestricted least squares estimator is also the unrestricted maximum likelihood estimator when the likelihood is proportional to a normal distribution. In convex programming literature, the term *basic solution* refers to the value $\underline{x}_o$ of $\underline{x}$ which gives the global minimum for a convex objective function $F(\underline{x})$. This paper will deal with the normal likelihood function and for our purposes the terms *unrestricted least squares estimate*, *unrestricted maximum likelihood estimate*, and *basic solution* or *basic estimate* will be considered synonomous. Similarly, *restricted least squares estimate*, *restricted maximum likelihood estimate*, and *minimum feasible solution* are interchanged.

The term *Bayesian estimate* is often used for the mean of the posterior distribution in discussions of Bayesian techniques. Should any other meaning be intended, it will be made clear by the text.

1.3 Review of Literature

Many statisticians have taken an interest in the problem of finding estimators for the parameters of the linear model (1.1.1) where the parameters can be restricted as in (1.1.2). Most of the earlier work has simply attempted to find the least squares estimator which satisfies the restrictions (*cf*. *e*.*g*., Judge and Takayama (1966), Mantel (1969), Lovell and Prescott (1970), Zellner

(1961), and Malinvaud (1966), page 317. Finding the restricted least squares solution is an application of quadratic programming which is covered in most convex or nonlinear programming texts (*e.g.*, Boot (1964), Hadley (1964), and Kunzi, Krelle, and Oettli (1966)).

A particular class of restricted least squares estimators, *viz*., those in isotonic regression, received much attention. (A brief discussion of the problems for which isotonic regression is appropriate, is contained in Section 2.3.) Early work in this area was performed by Ayer *et al*. (1955) and Brunk (1958) and a recent text by Barlow and others (1972) was devoted entirely to the subject. The book contains a nearly complete bibliography.

Bayesian procedures for the unrestricted regression problem have been discussed for example by Raiffa and Schlaifer (1961) and by Zellner (1971). However, Bayesian estimators for $\underline{\beta}$ in a restricted parameter space have not received much attention. Bartholomew (1965) mentioned the topic and Barlow *et al*. (1972, p. 95) discussed the mode of the posterior as a possible estimate under the conditions considered here.

1.4 Scope, Objectives, and Organization of This Paper

This paper will concentrate on point estimators for the parameters of the linear model which satisfy constraints (1.1.2). Attention will be restricted to situations in which the vector $\underline{\varepsilon}$ in (1.1.1) has an n-variate normal distribution. This paper will consider the case of full rank design matrices only.

This paper will have two objectives. The first will be detailing the properties of maximum likelihood estimation on restricted parameter spaces for the normal likelihood function. The second will be to determine if Bayesian techniques or some other estimation procedure will give properties superior to the maximum likelihood estimates.

The maximum likelihood estimation procedure for restricted parameter spaces and normal likelihood functions will be considered in Chapter 2. A simple quadratic programming algorithm will be reviewed there to give readers unfamilar with quadratic programming an understanding of the mappings onto feasible parameter space carried out by maximum likelihood estimation under restrictions.

Chapter 3 will deal with Bayesian estimation on restricted parameter spaces and illuminate some seemingly unknown differences in Bayesian estimation on restricted parameter spaces as compared to estimation on parameter spaces which include all the elements of R^m.

Chapter 4 will be devoted to incorporating these findings into a situation where the likelihood function is the normal distribution function. The Bayesian estimators for a flexible class of prior distributions will be presented. Properties of the means of the resulting posterior distributions will also be discussed.

Finally, in Chapter 5 the possibility of combining some of the previously presented estimators will be explored. The aim will be to profit from the improvements in the mean square error made by some estimators over certain sets of the feasible parameter space while minimizing inflation of the mean square error.

2. MAXIMUM LIKELIHOOD ESTIMATION ON RESTRICTED PARAMETER SPACES

2.1 General Discussion

When the vector $\underline{\varepsilon}$ in (1.1.1) has a normal distribution with mean zero and covariance matrix $\sigma^2 I$, the parameter $\underline{\beta}$ has a likelihood function

$$L(\underline{\beta}|\underline{y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp(-\frac{1}{2}\sigma^2(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta})) . \qquad (2.1.1)$$

To maximize this function with respect to $\underline{\beta}$, it is necessary to minimize the residual sum of squares:

$$\Psi(\underline{\beta}) = (\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta}) . \qquad (2.1.2)$$

When X is of full column rank, the value of $\underline{\beta}$ which minimizes (2.1.2) over R^m is the least squares estimator

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y} .$$

By the Gauss-Markov theorem this is the best linear unbiased estimator.

One approach to finding an estimator for $\underline{\beta}$ which satisfies constraints such as (1.1.2) is to maximize (2.1.1) for $\underline{\beta}$ on the appropriate subset of R^m. The execution of the maximization of a likelihood function on a proper subset of R^m is not easy, but it is not altogether new to the statistician. For example, any problem which includes constructing a likelihood ratio test of the hypothesis

$$H: \; C'\underline{\beta} \geq \underline{d}$$

conceptually involves finding the point at which (2.1.1) is maximized on a proper subset of R^m. Most introductory statistical texts give a method for finding the estimators of $\underline{\beta}$ when elements of $\underline{\beta}$ are known *a priori*. This also is an example of estimation on a convex subset of R^m.

Now consider a normal likelihood function of a mean vector, $\underline{\mu}$, with more than one element. If one wishes to maximize this likelihood with respect to $\underline{\mu}$ over a proper subset, S, of Euclidean space and the global maximum of the likelihood is not contained in S, classical analysis procedures are not adequate to find the solutions. One can utilize, however, some of the techniques of nonlinear (in fact, quadratic) programming to obtain such maximum likelihood estimators.

2.2 Quadratic Programming

The algorithms of quadratic programming provide methods for minimizing a convex quadratic function

$$Q(\underline{u}) = \underline{p}'\underline{u} + \underline{u}'D\underline{u} \qquad (2.2.1)$$

subject to the restrictions that

$$C\underline{u} \geq \underline{b}\,. \qquad (2.2.2)$$

For $Q(\underline{u})$ to be strictly convex, it is necessary for D to be positive definite (see Kunzi, Krelle, and Oettli (1966), page 39). For the function $\Psi(\underline{\beta})$ in (2.1.2) to be strictly convex, the matrix $X'X$ must be positive definite. This is assured by X having full column rank.

Boot (1964) notes that if the restrictions in (2.2.2) were equalities the desired solution could be found by the standard application of Lagrangian multipliers. But, the restrictions are inequalities and there exist situations where the basic solution to (2.2.1) satisfies the restrictions. In situations where this is not the case, some or all of the restrictions must be invoked to obtain the required solution. The restricted solution will then lie on the boundry of the feasible parameter space, that is, it will satisfy some of the inequalities in (2.2.2) as equalities; see Theorem 1 in Section 8.1. These equalities will here be called "binding" constraints.

Illustrations would perhaps clarify the situation. Consider a two parameter linear model of full rank, where the parameters to be estimated are restricted to the first quadrant, that is, $\underline{\beta} \geq 0$. The contours of the objective function (2.1.2) are then ellipses. Figures 2.1, 2.2, and 2.3 give examples of the optimum feasible solutions that can be obtained when the basic solution is infeasible. The ellipse shown in each figure is the contour corresponding to the minimum of the criterion function on the feasible sample space.

In Figure 2.1 the basic solution violates the constraint $\beta_2 \geq 0$ and the optimal feasible solution lies on the line $\beta_2 = 0$. Thus, $\beta_2 = 0$ is the binding constraint. In Figure 2.2, the basic solution violates both the constraints, but the optimal feasible solution is on the line $\beta_1 = 0$ and $\beta_1 = 0$ is the binding constraint. Figure 2.3 illustrates a situation where only one constraint is violated by the basic estimate, but the optimal feasible solution lies on $\beta_1 = 0$ and on $\beta_2 = 0$; $\beta_1 = 0$ and $\beta_2 = 0$ are the binding constraints.

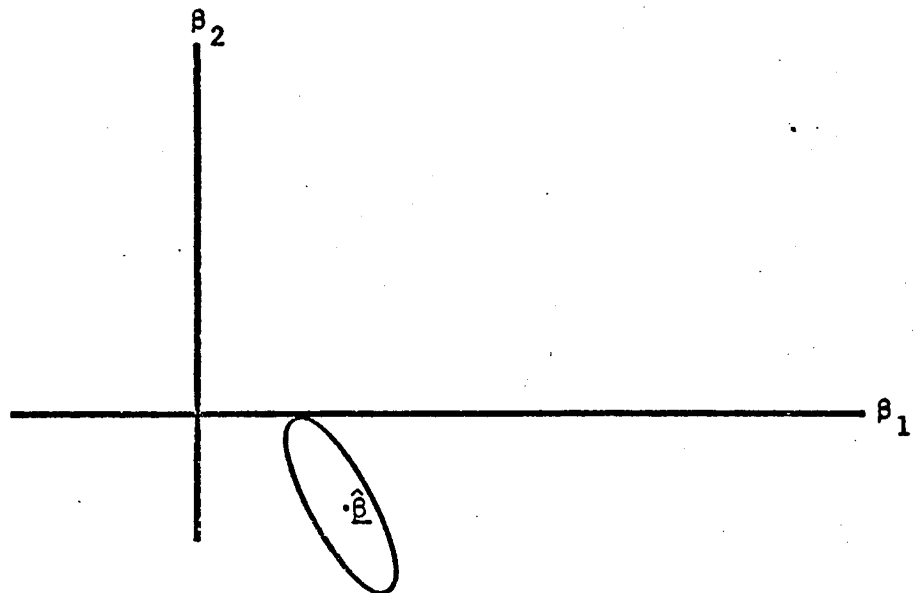


Figure 2.1 An example of a solution to a quadratic programming problem in which the basic estimate violates the constraint $\beta_2 \geq 0$, and the same constraint is binding

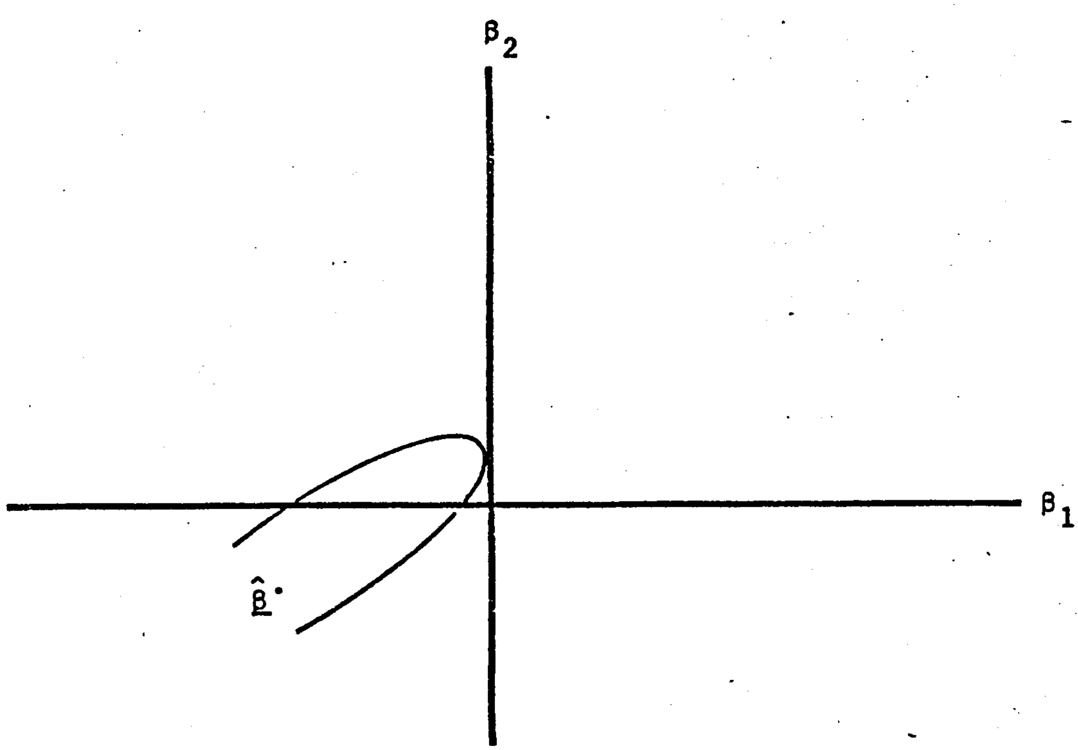


Figure 2.2 An example of a solution to a quadratic programming problem in which the basic estimate violates both constraints, and the constraint $\beta_1 \geq 0$ is binding

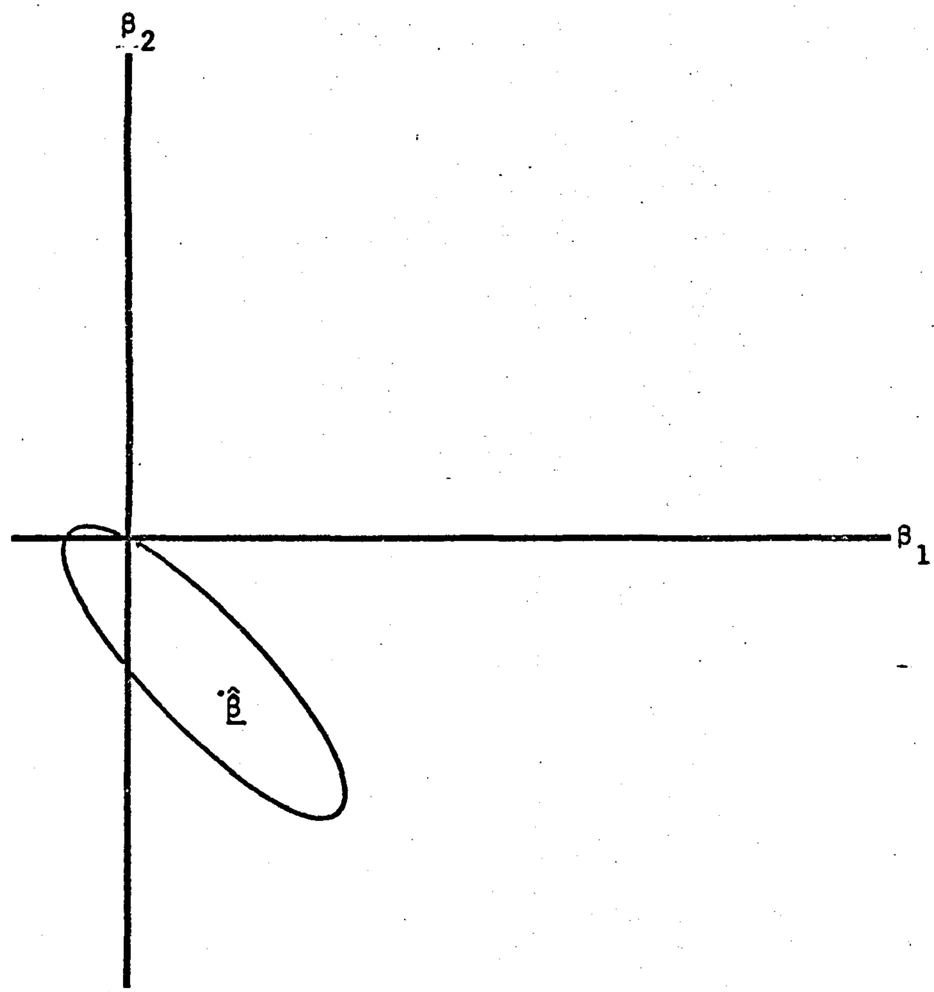


Figure 2.3 An example of a solution to a quadratic programming problem in which the constraint $\beta_2 \geq 0$ is violated by the basic estimate, and both constraints are binding

From these examples it is apparent that a significant problem in quadratic programming is that of deciding which constraints are binding. An algorithm due to Theil and Van de Panne (1960) will be outlined for finding the binding constraints and the optimum feasible solution.

If the restrictions which are binding are known, then optimum feasible estimates would be found by the straight forward use of Lagrangian multipliers. For example, if only a subset, S , of the original restrictions stated as equalities are binding, then the minimum of (2.2.1) under the restrictions (2.2.2) could be found by minimizing

$$(\underline{y} - X\underline{\beta})'(\underline{y} - X\underline{\beta}) + \underline{\lambda}'(C_s\underline{\beta} - \underline{d}_s) \tag{2.2.3}$$

where $C_s\underline{\beta} = \underline{d}_s$ describes the binding constraints in the set S .

Taking the derivative of (2.2.3) with respect to $\underline{\beta}$ and setting it equal to zero, one finds $\hat{\underline{\beta}}_s$, the optimum feasible solution under S , *viz.*,

$$\underline{\beta}_s + (X'X)^{-1}X'\underline{y} - \frac{1}{2}(X'X)^{-1}C_s'\underline{\lambda} . \tag{2.2.4}$$

Premultiply (2.2.4) by C_s and substitute $\underline{d}_s$ for $C_s\underline{\beta}_s$. Then

$$\frac{1}{2} C_s(X'X)^{-1}C_s'\underline{\lambda} = C_s(X'X)^{-1}X'\underline{y} - \underline{d}_s .$$

The matrix C_s always has full row rank, (see Boot (1964), page 99). Thus,

$$\underline{\lambda} = 2(C_s(X'X)^{-1}C'_s)^{-1}(- \underline{d}_s + C_s(X'X)^{-1}X'\underline{y}) .$$

Substituting this expression for $\underline{\lambda}$ in (2.2.4) gives

$$\hat{\underline{\beta}}_s = (X'X)^{-1}X'\underline{y} - (X'X)^{-1}C_s'[C_s(X'X)C_s']^{-1}[-\underline{d}_s + C_s(X'X)^{-1}X'\underline{y}] \ . \tag{2.2.5}$$

To discover the binding constraints, S , which give in turn the minimum feasible solution by (2.2.5), Theil and van de Panne (1960) recommended the following procedure. In this discussion S_k refers to a collection of k constraints which are being imposed as equality constraints.

1) Find the basic solution, i.e., the unrestricted solution, $\hat{\underline{\beta}}_o = (X'X)^{-1}X'\underline{y}$. If $\hat{\underline{\beta}}_o$ satisfies the constraints, it is the minimum feasible solution.

2) If $\hat{\underline{\beta}}_o$ violates any of the constraints, the sets S_1 will be formed by taking one at a a time each of the constraints which are violated by $\hat{\underline{\beta}}_o$. The restricted estimate, $\hat{\underline{\beta}}_{s_1}$, is then found by (2.2.5) for each set, S_1 . If any $\hat{\underline{\beta}}_{s_1}$ satisfies all the constraints it is the desired solution, and the corresponding set S_1 is the set of binding constraints.

3) If the optimal solution was not found in Step 2, sets of constraints S_2 are found by adding one at a time to each S_1 , each of the constraints violated by the corresponding

$\hat{\underline{\beta}}_{s_1}$. The restricted estimate $\hat{\underline{\beta}}_{s_2}$ is then found for each unique set of constraints S_2 . If an estimate $\hat{\underline{\beta}}_{s_2'}$ violates none of the constraints, it is the optimal solution if and only if the estimates found by eliminating either of the constraints in S_2' violate the omitted constraint.

4) If the optimal solution is not found in Step 3, sets of constraints, S_3 , are constructed by adding one at a time to each of the sets S_2 , the constraints found to be violated by the corresponding $\hat{\underline{\beta}}_{s_2}$. Sets, S_2' , which fail to satisfy the final condition given in Step 3 are not considered in this step. The restricted estimates, $\hat{\underline{\beta}}_{s_3}$, are then found for each unique sets of constraints S_3 . If an estimate, $\hat{\underline{\beta}}_{s_3'}$, violates no constraints and the three estimates found by eliminating one of the constraints in S_3' violates the constraint omitted, then $\underline{\beta}_{s_3'}$ is the optimal solution. If a feasible estimate fails to satisfy the last condition, the corresponding set, S_3' , is not considered in subsequent steps.

5) The process is continued as in Step 4 by considering successively larger sets of constraints

S_k . A given feasible estimate $\hat{\underline{\beta}}_{S_k'}$ is optimal if each of k estimates found by eliminating one of the constraints in S_k' violates the constraint omitted. The algorithm is continued until such an optimal feasible solution is found.

Kunzi, Krelle, and Oettli (1966) give proofs that the preceding algorithm will lead to the solution of the quadratic programming problem. Their proofs are based on the saddle point theorem given by Kuhn and Tucker (1951).

The optimal solution $\hat{\underline{\beta}}_S$ will be unique, although it is possible to reach the same solution with different sets of "binding" constraints. This can occur if the point of tangency of the criterion function (2.1.2) and one of the restrictions forming the boundary of the feasible parameter space is also the point of intersection of other restrictions. Such is the case in Figure 2.4.

There are many algorithms for finding the solutions to the quadratic programming problem. One of the more recent contributions is Mantel's (1969) paper in which he gives a procedure which can simplify calculations in some instances. However, the Theil-van de Panne algorithm has a geometrical interpretation which is somewhat easier to grasp.

2.3 Isotonic Regression

In many cases the expectation of an observation can be expressed as $E(y_{ij}) = \beta_j$. This is equivalent to having a design matrix with

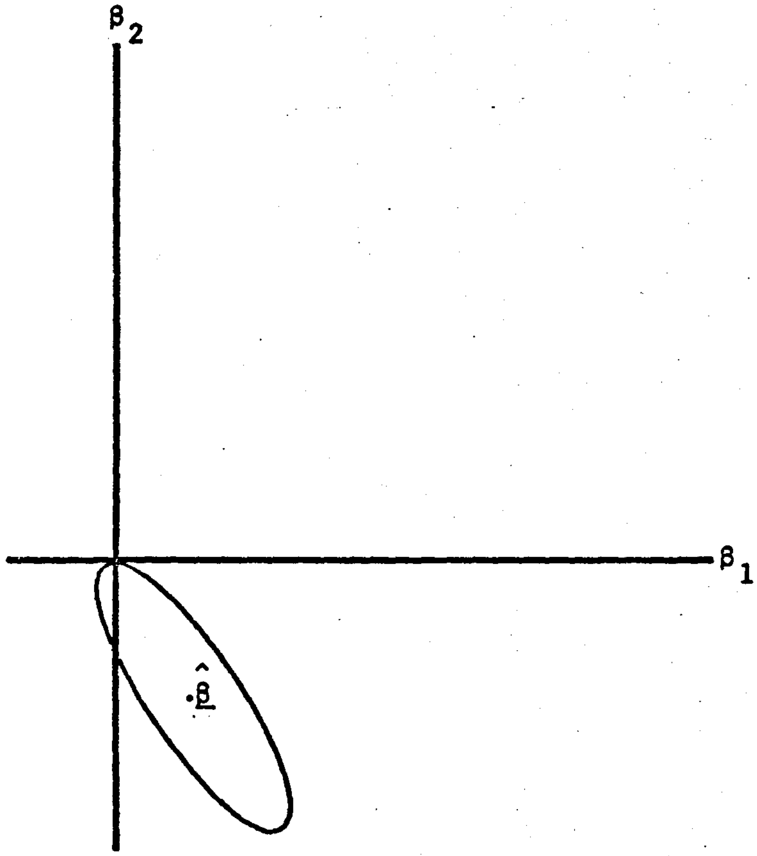


Figure 2.4 An example of a feasible solution to a quadratic programming problem found for the binding constraint $\beta_2 = 0$ where $\beta_1 = 0$ is also satisfied by the feasible solution

0 or 1 for each element, and only one 1 per row. Then the β_j which would maximize (2.1.1) subject to a given set of order restrictions on the parameters would be called the isotonic regression with respect to the particular restrictions. Maximizing (2.1.1) for this particular problem has sufficient applications and has generated enough interest to warrant a book devoted entirely to the subject (see Barlow *et al*. (1972)).

As an estimating procedure isotonic regression could prove extremely useful for the growth model problem in which observations have been made over several time intervals. For many biological phenomena the growth model should be restricted to being monotonically increasing. In the absence of further knowledge of the functional form of the growth process, the maximum likelihood estimates under the assumption of normality would be the isotonic regression with respect to restrictions that

$$\beta_1 \le \beta_2 \le \ldots \le \beta_m , \tag{2.3.1}$$

a simple ordering of the β's.

The restriction given in (2.3.1) can be expressed in the form (1.1.2) by letting $\underline{d}$ equal the null vector and by defining C as follows:

$$C = \begin{bmatrix} -1 & 1 & 0 & . & . & 0 & 0 \\ 0 & -1 & 1 & . & . & 0 & 0 \\ 0 & 0 & -1 & . & . & 0 & 0 \\ . & . & . & & & & \\ . & . & . & & & & \\ 0 & 0 & 0 & . & . & -1 & 1 \end{bmatrix}_{(m-1) \times m} . \quad (2.3.2)$$

A solution satisfying the simple ordering restriction indicated in (2.3.1) and (2.3.2) consists of first finding the basic estimates by either unrestricted least squares or unrestricted weighted least squares. Then an algorithm called pooling adjacent violaters is applied. This procedure involves taking the weighted average of adjacent estimates (e.g., $\hat{\beta}_i > \hat{\beta}_{i+1}$) which violate the restrictions. This pooled estimate is then assigned as the isotonic estimates for each parameter. Any pooled estimates will be considered as a block if further pooling is required to obtain the desired order. The weight of the block will be the sum of the weights for the basic estimates. Pooling is continued until the isotonic estimates satisfy the ordering imposed on the parameters. A method of steepest descent gives a strategy for choosing violators which would be the most efficient in many cases (see Kruskal (1964)).

2.4 Properties of the Restricted Maximum Likelihood Estimates

A property of an estimator which is usually considered desirable is consistency. For if one has a consistent estimator, the true value of the parameter can be bought. That is with a sufficiently large sample size the estimate has a nearly degenerate distribution at the true value of the parameter.

The basic estimates are consistent in a certain, reasonable sense when $\underset{\sim}{\varepsilon}$ in (1.1.1) is distributed normally with mean zero and variance $\sigma^2 I$. In particular, if the experiment represented by (1.1.1) is repeated k times, yielding k independent vectors

$$\underset{\sim}{y}_i \sim MVN(X\underset{\sim}{\beta}, \sigma^2 I) ,$$

then the design matrix for the entire experiment is

$$X_k = \begin{bmatrix} X \\ X \\ \vdots \\ X \end{bmatrix}$$

where there are k submatrices all identical to X . Then

$$X_k'X_k = k(X'X)$$

and

$$(X_k'X_k)^{-1} = (X'X)^{-1}/ k .$$

Now the basic estimates are

$$\hat{\underline{\beta}}_k = (X_k'X_k)^{-1}X_k'\underline{y}$$

$$= (X'X)^{-1}X'(\sum_{j=1}^{k} \underline{y}_j)/k..$$

where the vector $\underline{y}_j$ is the observation for the $j^{\underline{th}}$ repetition of the experiment (1.1.1). Recall that $\underline{y}_j$ has an n-variate normal distribution with mean $X\underline{\beta}$, covariance matrix $\sigma^2 I$. Then $\sum_{j=1}^{k} \underline{y}_j$ would have an n-variate normal distribution with mean $kX\underline{\beta}$ and variance $k\sigma^2 I$. Thus, $\hat{\underline{\beta}}_k$ would be distributed as an m variate normal with mean $\underline{\beta}$ and variance $\sigma^2(X'X)^{-1}/k$. Thus, as k becomes large, the covariance matrix becomes a zero matrix, and the distribution of $\hat{\underline{\beta}}_k$ is degenerate at $\underline{\beta}$.

To show that the restricted estimates satisfying (1.1.2) are consistent in the same sense, observe that

$$\lim_{k\to\infty} \Pr(C'\hat{\underline{\beta}}_k < \underline{d}) \to 0 .$$

This is a consequence of the convergence of the distribution of the basic estimator to a degenerate distribution at the true value of $\underline{\beta}$ which is known to satisfy (1.1.2). This implies that as the sample size increases, the basic estimates will violate the restrictions on the model with a probability of zero. If the basic estimates fail to violate the restrictions, then the basic estimates are the restricted maximum likelihood estimates. (Barlow et al. (1972) gives an equivalent proof for the case in which the restricted estimates are an isotonic regression with respect to a quasi-ordering.)

The restricted maximum likelihood estimators are, in general, biased. Mantel (1969) gives the following example which illustrates the reason. Consider the one parameter model in which the one parameter cannot be negative. The restricted maximum likelihood estimate $\tilde{\beta}$, is zero when the basic estimate, $\hat{\beta}$, is less than zero, and is equal to the basic estimate when the basic estimate satisfies the restriction. The expected value of $\tilde{\beta}$ is

$$E(\tilde{\beta}) = \int_{-\infty}^{0} 0p(\hat{\beta})d\hat{\beta} + \int_{0}^{\infty} \hat{\beta}p(\hat{\beta})d\hat{\beta}$$

where $p(\hat{\beta})$ is the probability density of the basic estimate. The basic estimate is unbiased and its expectation is

$$E(\hat{\beta}) = \int_{-\infty}^{0} \hat{\beta}p(\hat{\beta})d\hat{\beta} + \int_{0}^{\infty} \hat{\beta}p(\hat{\beta})d\hat{\beta} = \beta .$$

Note that $\hat{\beta}$ is less than zero on the interval $[-\infty, 0]$ so

$$\int_{[-\infty, 0)} \hat{\beta}p(\hat{\beta})d\hat{\beta} < \int_{[-\infty, 0)} 0\ p(\hat{\beta})d\hat{\beta}$$

if $p(\hat{\beta}) > 0$ anywhere on $[-\infty, 0)$; so

$$\beta = E(\hat{\beta}) < E(\tilde{\beta}) .$$

If the basic estimates lie outside the feasible parameter space with probability near zero, the restricted maximum likelihood estimators can have little bias. This property might encourage the modeler to be conservative in formulating the restrictions, and he might include elements in the "feasible" parameter space which are not feasible. This type of "bordering" of the feasible parameter space would be paid for by negating some of the gains made in reducing mean square error by restricting the parameter space.

Barlow et al. (1972), page 64, gives a theorem for the isotonic regression estimates, $\tilde{\underline{\beta}}$, of ordered parameters which shows that these estimates have a smaller mean square error than the unrestricted least squares estimates. The theorem states that

$$\sum_{i=1}^{m} (\beta_i - \tilde{\beta}_i)^2 w_i \leq \sum_{i=1}^{m} (\beta_i - \hat{\beta}_i)^2 w_i \tag{2.4.1}$$

where $\underline{\beta}$ is the true value of the parameter, $\hat{\underline{\beta}}$ the least squares estimate of $\underline{\beta}$, and $\underline{w}$ is a vector of weights. Taking the expectation of (2.4.1) with the w_i equal shows that the mean square error of the isotonic regression estimator is less than the mean square error for the unrestricted estimator. (This is a straightforward application of the comparison theorem.) This result of the inequality in (2.4.1) is that for ordered $\underline{\beta}$, isotonic estimates reduce the mean square error although for $\underline{\beta}$ not 'near' a boundary, the reduction would be expected to be small.

In this example, it is possible to show that the mean square error can be decreased because the isotonic regression estimates are the nearest points in the restricted parameter space to the basic estimates. The general restricted maximum likelihood estimate or restricted least squares estimate is not usually the nearest point in the restricted parameter space to the basic estimate; so the same proof would not hold.

Mantel (1969) states, without proof, that the mean square error for any unrestricted least squares estimator is larger than for the restricted estimator. Judge and Takayama (1966) concluded that for a broad class of problems, Mantel's contention is true. Their conclusions were based on Zellner's (1961) work in which a mapping similar to the one given in the proof for the isotonic regression example above was considered.

A final property to consider is that the restricted estimates will always be boundary points of the restricted parameter space whenever the basic solution is infeasible. This property is shown in Section 2.2 above and more formally by Theorem 1 in Section 8.1. Thus, restricted maximum likelihood estimates will pile up on boundary points - points which barely satisfy the restrictions. The property is unappealing because the same restricted estimate could be obtained for the case where the basic estimate satisfies the restrictions exactly and when the basic estimates grossly violate the restrictions.

To summarize, some properties of the restricted maximum likelihood estimator are: it is consistent in a reasonable sense, but it is biased; it can have small mean square error, but the sampling distribution is somewhat unattractive.

3. BAYESIAN ESTIMATION ON RESTRICTED PARAMETER SPACES

3.1 Properties and Problems

The problem described in the introduction lends itself well to the Bayesian philosophy of estimation. One wishes to find estimators for certain parameters where the true values cannot possibly belong to a certain subset of Euclidean space. The Bayesian approach would define a "prior" desnity which assigns probability zero to this impossible subset.

The next step in the Bayesian approach would be to specify the prior density on the feasible parameter space. In the situation where little additional prior information is available with regard to the true value of the true value of the parameters being estimated, a uniform prior is often chosen. The uniform prior has the interpretation that no greater degree of *a priori* belief is placed in any one point in the feasible parameter space than in any other. The final step in the Bayesian approach would be to compute the "posterior" distribution, that is the conditional distribution of the parameter given the observations, and to estimate the parameter by some measure of the central tendency of this posterior distribution: its mean (most frequently), its median, or its mode.

The Bayesian approach does seem to be appropriate for finding estimates in the situations described here, but few publications have addressed this problem. Bartholomew (1965) discussed the special problems of constructing interval estimates when the parameter space is restricted. Barlow *et al*. (1972) discussed the use of the mode of the posterior as an estimator when the parameter space is restricted. These are the only references that were found which dealt with Bayesian estimation on restricted parameter spaces.

The mode of the posterior density is the same as the traditional maximum likelihood estimator when a uniform prior distribution is used. This is true whether the feasible parameter space is a proper subset of Euclidean space or not. In case the feasible part of parameter space is a proper subset of the Euclidean space, this estimator will be bunched up at the boundary of the feasible space. This is an unpleasant property of this estimator mentioned earlier.

The Bayesian estimator most often used is the mean of the posterior distribution. Now the mean of any distribution will be contained in the convex hull of the support of that distribution. Since the support of the posterior distribution is a proper or improper subset of the support of the prior distribution, this is a somewhat attractive property. However, this Bayesian estimator also has an unpleasant property: it can assume values on the finite part of the boundary of the convex hull of the support of the posterior distribution if and only if the posterior distribution is degenerate at (a flat subset of) this finite part (i.e., the part with finite coordinates); see Theorem 8.3. In fact the mean of the posterior distribution will always be bounded away from that boundary unless the posterior distribution is degenerate at it.

This property is particularly easy to observe when a beta prior with parameters α and β is assumed for the parameter θ of a binomial distribution with density

$$p(x|\theta) = \binom{n}{x}\theta^x(1 - \theta)^{n-x}; \quad x = 1, 2, \ldots, n \; ; \quad 0 \le \theta \le 1 \; .$$ [1/]

The posterior distribution for θ is a beta distribution with parameters $(\alpha + x)$ and $(\beta + n-x)$. The mean of the posterior is

$$\hat{\theta} = (\alpha + x)/(\alpha + \beta + n) \; .$$

The parameters α and β are greater than zero, so for a given value of n , $\hat{\theta}$ could never take on the value of 0 or 1 . In fact, it is easy to see that $\hat{\theta}$ cannot take any values between 0 and $\alpha/(\alpha + \beta + n)$, nor between $(\alpha + n)/(\alpha + \beta + n)$ and 1 , (the reader will see this by finding the value of $\hat{\theta}$ for $x = 0$ and for $x = n$).

The mean of the posterior distributions for the continuous conjugate distributions given by Raiffa and Schlaifer (1961) show the same property. As an example, consider the rectangular distribution with density

$$f(x|\theta) = 1/\theta \; , \quad 0 \le x \le \theta \; , \quad \theta > 0 \; ,$$

where the real life problem indicates that the feasible part of parameter space is given by $\theta \in [\gamma, \infty)$, $0 < \gamma$. The joint density for a sample $(X_1, X_2, \ldots, X_n)$ of size n is

$$f(\underline{x}|\theta) \begin{cases} = \theta^{-n} , & X_{(n)} \le \theta < \infty \\ = 0 & \text{otherwise} \end{cases}$$

[1/] Bennee F. Swindel suggested this example in personal communications where $X_{(n)}$ is the largest order statistic. The conjugate prior density is

$$p(\theta) \propto \begin{cases} \theta^{-n'} , & n' > 1 , \quad \gamma \leq \theta \leq \infty \\ 0 & \text{otherwise} \end{cases}$$

Then the mean of the posterior, since $n + n' > 2$, is

$$\hat{\theta} = \frac{M^{n''-1}}{M^{n''-2}} \frac{n''-1}{n''-2} = \frac{n''-1}{n''-2} M ,$$

where $M = \text{Max}(X_{(n)}, \gamma)$, $n'' = n + n'$. Since $\gamma > 0$, M also is strictly positive, and $\hat{\theta}$ has a minimum distance of $\gamma/(n''-2)$ from γ , the finite boundary of the feasible parameter space.

Thus, the Bayesian estimator (in the sense of the mean of the posterior distribution) seems to be as unappealing as the maximum likelihood estimator, (i.e., the mode of the posterior distribution for a uniform prior), since legitimate values of the parameters, i.e., values carrying positive probability density in both the prior and the posterior distributions, will be ignored by the estimation process.

3.2 Alternative Bayesian Procedures

The mean of the posterior distribution is the appropriate estimator for a parameter θ when the loss function is the squared error

$$\ell_1(\hat{\theta},\theta) = (\hat{\theta} - \theta)^2 . \tag{3.2.1}$$

Tiao and Box (1973) have suggested that other loss functions not be overlooked. For example the loss function

$$\ell_2(\hat{\theta},\theta) = \begin{cases} 1 & \text{for } |\hat{\theta} - \theta| > \varepsilon \\ 0 & \text{for } |\hat{\theta} - \theta| < \varepsilon \end{cases}, \quad \varepsilon \text{ small and positive,}$$

gives rise to the mode of the posterior distribution as an estimator of the parameter. The expected value of the loss function

$$\ell_3(\hat{\theta},\theta) = |\hat{\theta} - \theta|$$

is minimized by the median of the posterior distribution.

These two additional loss functions and the corresponding estimators seem inadequate for the particular problem under consideration. For example, when a uniform prior is used, it was observed before that the mode of the posterior is also the maximum likelihood estimate, see also Barlow _et al_. (1972), page 95.

The median is similar to the mean of the posterior in that it, too, excludes feasible values of the estimator. A point in a one-dimensional parameter space can be the median of a continuous posterior distribution if and only if the cumulative posterior distribution is equal $\frac{1}{2}$ at the point. Thus, for absolutely continuous posterior distributions, finite boundary points on the convex hull of the support would again be excluded as estimates. In fact, any posterior distribution which does not assign a probability of $\frac{1}{2}$ or more to a flat boundary point will fail to give boundary points as medians of the posterior.

In the search for an estimator of a restricted parameter with appealing sampling properties, estimators which accumulate at boundary points and estimators which never include certain neighborhoods of boundary points have so far been encountered. It would seem that it should be possible to find an estimation process that would avoid both extremes. For example, one might think that the problem can be solved by choosing a prior such that the mean of the resulting posterior can assume any value in the feasible parameter space. This can be done by assigning positive prior probability to points outside the feasible parameter space. This would be analogous to "averaging" the feasible parameter space and the unrestricted parameter space.

A specific example of how this could be carried out can be constructed using the rectangular process with a hyperbolic prior cited earlier in the section. Recall that in this example the feasible parameter space is $[\gamma, \infty)$, $0 < \gamma$, and the mean of the posterior is $(n''-1)M/(n''-2)$, where M is the larger of γ and the largest order statistic of the sample. Thus, the mean of the posterior will never fall in the interval $[\gamma, (n''-1)\gamma/(n''-2)]$, which is a non-empty subset of the feasible parameter space. If, however, the chosen prior assigns positive value over the interval $[(n''-2)\gamma/(n''-1), \infty]$ according to the hyperbolic distribution, then the minimum value of the mean of the posterior is γ. Thus, the mean of the posterior can be forced to cover the entire feasible parameter space by assigning positive prior probability to elements not contained in the feasible parameter space.

This example illustrates that the prior can be manipulated to achieve estimates that exhaust the feasible parameter space. It should be noted, however, that for the combination of prior and likelihood chosen, the resulting estimator now has the same shortcoming as the maximum likelihood estimator, i.e., for all samples in which the maximum order statistic is less than $(n''-2)\gamma/(n''-1)$, the estimate of θ is γ.

The evidence presented thus far indicates that for the problem presented in the introduction, Bayesian procedures can rectify the undesirable accumulation of estimates on the boundary of the feasible parameter space. The cost of this rectification seems to be that the mean or median of the posterior will not approach the finite boundaries of the feasible parameter space unless the mass of the posterior accumulates at these boundaries. The mode of the posterior distribution represents a distinct alternative to maximum likelihood estimators only if the prior chosen is not uniform.

Other general properties of Bayesian estimation on a restricted parameter space will not be explored here. In the remainder of this paper, attention will be restricted to some specific estimation problems regarding the m-variate normal distribution where inequality constraints can be placed on the elements of the mean vector. In these specific situations, specific Bayesian estimators will be explored.

4. BAYESIAN ESTIMATORS DERIVED FROM TRUNCATED NORMAL POSTERIOR DISTRIBUTIONS

4.1 Mean of a Truncated Normal Distribution

Raiffa and Schlaifer (1961) listed as a desirable trait of a prior density that it should lead to analytically tractable posterior distributions. In Section 4.3, several reasonable priors will be listed which yield a truncated normal posterior distribution for the situation in which the observations are a sample from a normal distribution. All of these reasonable priors assign positive probability only to the feasible parameter space, such as the space defined by (1.1.2). The truncated normal distribution does have a mean and mode which can be expressed with some degree of convenience. Of course, the mode is that point which gives the maximum of the normal density function on the restricted parameter space. Properties of the mode (i.e., the restricted maximum likelihood estimator) and algorithms for finding it were discussed at length in the second chapter of this paper.

The mean of the posterior distribution is the value of the parameter minimizing the loss function (3.2.1). To explore tractability of means of truncated normal distributions, we will first consider the univariate truncated normal distribution

$$g(x) \propto e^{-(x-\mu)^2/2\sigma^2} \quad \text{for} \quad x \geq a \, , \tag{4.1.1}$$
$$= 0 \qquad \text{for} \quad x < a \, .$$

Cramér (1951), page 248, gives the first moment of (4.1.1) as

$$E(x) = \mu + \sigma f(\tfrac{a-\mu}{\sigma})/(1 - F(\tfrac{a-\mu}{\sigma})) \tag{4.1.2}$$

$$= \mu + \sigma f(\tfrac{\mu-a}{\sigma})/F(\tfrac{\mu-a}{\sigma}) \; .$$

Here $f(x)$ and $F(x)$ are the density and distribution function respectively for a normal random variable with mean zero and variance one. Equation (4.1.2) involves only standardized normal density and distribution functions and is easily evaluated for values of $(a-\mu)/\sigma$ less than 3 . However, as $(a-\mu)$ goes to ∞, both the functions $f(\frac{a-\mu}{\sigma})$ and $(1 - F(\frac{a-\mu}{\sigma}))$ approach zero rapidly. The attendent computational problems are easily taken care of by using a continued fraction expression for $(1 - F(x))$ (cf. Abramowitz and Stegum (1964), page 932), namely

$$1 - F(x) = f(x)\left[\frac{1}{x+}\,\frac{1}{x+}\,\frac{2}{x+}\,\frac{3}{x+}\,\frac{4}{x+}\,\cdots\right] \quad \text{for} \quad x > 0$$

$$= f(x)\;CF(x) \; . \tag{4.1.3}$$

Substituting (4.1.3) into (4.1.2) gives

$$E(x) = \mu + \sigma f(\tfrac{a-\mu}{\sigma})/(f(\tfrac{a-\mu}{\sigma})CF(\tfrac{a-\mu}{\sigma})) \quad \text{for} \quad \frac{a-\mu}{\sigma} > 0$$

$$= \mu + \sigma/CF(\tfrac{a-\mu}{\sigma}) \; . \tag{4.1.4}$$

Section 8.2 contains a table of the values we computed of $f(x)/F(x) = 1/CF(-x)$ for $-10 \le x \le 5$.

Thus, in univariate truncated normal posterior distributions finding the mean of the posterior is not difficult for any given set of values of μ, σ, and a. In our applications μ will be seen to be a function of the observations. The obvious next question is what happens for multivariate truncated normal distributions? This problem is dealt with in the Appendix, Section 8.3. Equation (8.3.11) gives the mean for a broad class of posterior distributions, and (8.3.13) gives Cramér's result for the univariate case as derived from (8.3.11).

4.2 Priors Producing a Truncated Normal Posterior Distribution for the Problem of Isotonic Regression

In Chapter 3 the uniform prior was said to be applicable when the modeler can not specify in advance that any one point in the parameter space is more likely than any other. It was also noted in that chapter that a uniform prior yields a posterior which is proportional to the likelihood for those points in the parameter space that belong to the support of the prior.

Consider the case of one observation taken from each of m normal populations. The populations have means μ_i, respectively, $i = 1, 2, \ldots, m$, and all have variance, σ^2, known. Let the uniform prior be assigned over the feasible parameter space, A, defined by the simple ordering

$$\mu_1 \leq \mu_2 \leq \mu_3 \leq \ldots \leq \mu_m \quad .$$

Note that the feasible parameter spaces defined by general linear inequalities $C\underline{\beta} \geq \underline{d}$ are considered in the Appendix, Section 8.3. In this case the joint density of $y_1, y_2, \ldots, y_m$ is

$$f(\underline{y}|\underline{\mu}) = (\sqrt{2\pi}\,\sigma)^{-m}\exp(-(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})/(2\sigma^2))$$

where $\underline{y}$ is the vector with components $y_1, y_2, \ldots, y_m$, and with support as given in (8.3.14). This posterior density is a special case of (8.3.3), and the Bayesian estimator of $\underline{\mu}$ is a special case of (8.3.17).

If n observations are taken from each population, the density function is

$$f(\underline{y}|\underline{\mu}) = (\sqrt{2\pi}\,\sigma)^{-nm}\exp(-\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij} - \mu_i)^2/(2\sigma^2))$$

$$= (\sqrt{2\pi}\,\sigma)^{-nm}\exp(-(\sum_{i=1}^{m}\sum_{j=1}^{n}y_{ij}^2 - \sum_{i=1}^{m}n\bar{y}_i^2)/(2\sigma^2))$$

$$\cdot \exp(-\sum_{i=1}^{m}n(\bar{y}_i - \mu_i)^2/(2\sigma^2)).$$

Here, y_{ij} refers to the $j^{\underline{th}}$ observation on the $i^{\underline{th}}$ population, and $\bar{y}_i$ is the mean of the observations from the $i^{\underline{th}}$ population. Then, the posterior density of $\underline{\mu}$ on A is

$$p(\underline{\mu}|\underline{y},\sigma) = (\sqrt{2\pi}\,\sigma)^{-nm}\exp(-(\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij}^2 - n\sum_{i=1}^{m}\bar{y}_i^2)/(2\sigma^2))$$

$$\cdot\exp(-n(\bar{\underline{y}} - \underline{\mu})'(\bar{\underline{y}} - \underline{\mu})/(2\sigma^2))/$$

$$[\int_A(\sqrt{2\pi}\,\sigma)^{-nm}\exp(-(\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij}^2 - n\sum_{i=1}^{m}\bar{y}_i^2)/(2\sigma^2))$$

$$\cdot\exp(-n(\bar{\underline{y}} - \underline{\mu})'(\bar{\underline{y}} - \underline{\mu})/(2\sigma^2)d\underline{\mu}]$$

$$= \frac{\exp(-n(\bar{\underline{y}} - \underline{\mu})'(\bar{\underline{y}} - \underline{\mu})/(2\sigma^2))}{\int_A \exp(-n(\bar{\underline{y}} - \underline{\mu})'(\bar{\underline{y}} - \underline{\mu})/(2\sigma^2))d\underline{\mu}} \quad . \tag{4.2.1}$$

Therefore (4.2.1) is a truncated normal posterior distribution of the form (8.3.3). (The case where the number of observations differs for the m populations is considered in the Appendix, Section 8.4).

An <u>exponential prior</u> and a normal joint density also yield a truncated normal posterior. The exponential prior is appealing to those modelers who know that the true value of the parameter is more likely to barely satisfy the restrictions than to be far away from the boundary of the feasible parameter space.

Again, the case of one observation per population for m normal populations will be considered first. Again, the populations have mean μ_i respectively, $i = 1, 2, \ldots, m$, and all have variance σ^2 known. An exponential prior for a simple ordering is:

$$p(\underline{\mu}|\underline{\theta}) = K\exp(-(\mu_2 - \mu_1)/\theta_1 - (\mu_3 - \mu_2)/\theta_2 -$$

$$- \cdots - (\mu_m - \mu_{m-1})/\theta_{m-1})$$

$$= K \exp(-\underline{\mu}'C\underline{\theta}) \qquad \text{for} \quad \underline{\mu} \in A , \tag{4.2.2}$$

and

$$p(\underline{\mu}|\underline{\theta}) = 0 \qquad \text{otherwise} .$$

Here, K is a scalar constant, $\underline{\theta}$ is the vector whose elements are $1/\theta_i$, and the matrix C is given in (2.3.2). The resulting posterior density is:

$$p(\underline{\mu}|\underline{y},\underline{\theta},\sigma) = (\sqrt{2\pi}\,\sigma)^{-m}\exp(-(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})/(2\sigma^2))$$

$$\cdot\ K \exp(- \underline{\theta}'C\underline{\mu})/$$

$$[\int_A (\sqrt{2\pi}\,\sigma)^{-m}\exp(-(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})/(2\sigma^2))$$

$$\cdot\ K \exp(- \underline{\theta}'C\underline{\mu})d\underline{\mu}]$$

$$= \frac{\exp(-((\underline{y} - \sigma^2C'\underline{\theta})-\underline{\mu})'((\underline{y} - \sigma^2C'\underline{\theta}) - \underline{\mu})/(2\sigma^2)}{\int_A \exp(-((\underline{y}-\sigma^2C'\underline{\theta}) - \underline{\mu})'((\underline{y}-\sigma^2C'\underline{\theta}) - \underline{\mu})/(2\sigma^2)d\underline{\mu}} \ . \tag{4.2.3}$$

The posterior (4.2.3) is again of the form (8.3.3) with support given in (8.3.14).

The last prior considered here is a <u>truncated multivariate normal</u>. This prior gives a truncated normal posterior when the observations are taken from a random variable normally distributed. The truncated normal is a candidate prior when the modeler knows that some interior points in the feasible parameter space are more likely to be the true value of the parameter than any other points.

The truncated normal prior for m mean parameters μ_i of the densities of the observed variables considered here is

$$p(\underline{\mu}|\underline{\zeta},\alpha) = K \exp(-(\underline{\mu}-\underline{\zeta})'(\underline{\mu}-\underline{\zeta})/(2\alpha^2)) \quad \text{for } \underline{\mu} \in A ,$$

$$= 0 \qquad\qquad \text{elsewhere} .$$

As before, K is the normalizing constant. Then, the posterior density on A is

$$p(\underline{\mu}\underline{y},\underline{\zeta},\alpha,\sigma) = \exp[-(\alpha^2+\sigma^2)(\underline{\mu} - \frac{\alpha^2\underline{y}+\sigma^2\underline{\zeta}}{\alpha^2+\sigma^2})'$$

$$(\underline{\mu} - \frac{\alpha^2\underline{y}+\sigma^2\underline{\zeta}}{\alpha^2+\sigma^2})/(2\alpha^2\sigma^2)]/$$

$$[\int_A \exp[-(\alpha^2+\sigma^2)(\underline{\mu} - \frac{\alpha^2\underline{y}+\sigma^2\underline{\zeta}}{\alpha^2+\sigma^2})'$$

$$(\underline{\mu} - \frac{\alpha^2\underline{y}+\sigma^2\underline{\zeta}}{\alpha^2+\sigma^2})]/2\sigma^2\alpha^2)d\underline{\mu}] . \qquad (4.2.4)$$

4.3 Construction of Several Bayesian Estimators and Comparison With the Restricted Maximum Likelihood Estimators

In Section 3.1 concern was expressed that the mean of the posterior distribution might not yield all the points in the feasible parameter space as estimates. Example distributions were given which showed this property for all possible values of the observations on the random variable. A theorem is given in the Appendix, Section 8.1, which shows that all points in the convex hull of the support could be obtained as Bayesian estimates only if the posterior became degenerate at the boundary points.

It can be shown that the truncated normal posterior does become degenerate at boundary points for some observed values of the random variable. To illustrate, consider the univariate truncated normal posterior:

$$p(\mu|y) = \frac{\frac{1}{\sqrt{2\pi}\,\sigma} e^{-(\mu-y)^2/2\sigma^2}}{\int_S^\infty \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(\mu-y)^2/2\sigma^2} d\mu} \quad \text{for } \mu \in [S, \infty)$$

$$= 0 \qquad\qquad \text{otherwise} .$$

Here S is a finite number.

The posterior probability that μ lies in an interval $[a,b] \subset [S, \infty)$ is:

$$\Pr(\mu \in [a,b]) = \frac{\int_a^b \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(\mu-y)^2/2\sigma^2} d\mu}{\int_S^\infty \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(\mu-y)2\sigma^2} d\mu} = \frac{F(\frac{y-a}{\sigma}) - F(\frac{y-b}{\sigma})}{F(\frac{y-S}{\sigma})}$$

where $F(x)$ is the distribution function for the normally distributed random variable with mean zero, and variance one.

First, the case in which $S = a < b$ will be considered. Then by (4.3.1),

$$\Pr(\mu \in [S,b]) = 1 - F((y-b)/\sigma)/F((y-S)/\sigma) .$$

As y goes to $-\infty$, $\Pr(\mu \in [S,b])$ becomes 1. This can be seen by application of L'Hospital's rule to

$$F(\frac{y-b}{\sigma})/F(\frac{y-S}{\sigma}) ,$$

i.e.,

$$\lim_{y\to-\infty} F((y-b)/\sigma)/F((y-S)/\sigma)$$

$$= \lim_{y\to-\infty} f((y-b)/\sigma)/f((y-S)/\sigma) . \tag{4.3.2}$$

Here $f(x)$ is the density function for a random variable normally distributed with mean zero and variance one. Then

$$\lim_{y\to-\infty} \frac{f((y-b)/\sigma)}{f((y-S)/\sigma)} = \lim_{y\to-\infty} e^{y(b-S)/(2\sigma^2)} e^{(S^2-b^2)/(2\sigma^2)} . \tag{4.3.3}$$

Since $b > S$,

$$\lim_{y\to-\infty} e^{y(\frac{b-S}{\sigma})} = 0$$

and (4.3.3) is equal to zero. Thus,

$$\lim_{y\to-\infty} \Pr(\mu\in[S,b]) = 1 .$$

In the case $S < a \leq b$,

$$\Pr(\mu\in[a,b]) = \frac{F(\frac{y-a}{\sigma})-F(\frac{y-b}{\sigma})}{F(\frac{y-S}{\sigma})}$$

$$= \frac{F(\frac{y-a}{\sigma})}{F(\frac{y-S}{\sigma})} - \frac{F(\frac{y-b}{\sigma})}{F(\frac{y-S}{\sigma})}$$

and both

$$F(\frac{y-a}{\sigma})/F(\frac{y-S}{\sigma})$$

and

$$F(\frac{y-b}{\sigma})/F(\frac{y-S}{\sigma})$$

have limits of zero as y goes to $-\infty$ from the results found for (4.3.2). Thus, the probability that μ is in any interval entirely in the interior of $[S,\infty) \to 0$ as $y \to -\infty$, and the mass of the probability is accumulated at the boundary point, S. Therefore, this truncated normal posterior tends to a degenerate distribution at S as y goes to $-\infty$. This implies that for $y \to -\infty$ the mean of this posterior approaches S. Thus, the mean of a truncated normal posterior can take on all the values in the support of the posterior provided the observed values of the random variable can take on all negative, real values. We will see examples of this in the discussion below.

The case where one observation is made on a population distribution normally with mean $\mu \geq 0$ and variance one will now be considered. A uniform prior is assumed over the feasible parameter space $[0,+\infty)$. Then according to (4.2.1), the resulting posterior density on $0 \leq \mu < +\infty$ is

$$p(\mu|y) = \frac{\exp(-(y-\mu)^2/2)}{\int_A \exp(-(y-\mu)^2/2)d\mu} ,$$

so that by (4.1.2), the Bayesian estimator (*i.e.*, the mean of the posterior density) is

$$\hat{\mu} = y + f(y)/F(y) \ . \tag{4.3.4}$$

As y approaches $+\infty$, $f(y)$ tends to zero and $F(y)$ approaches unity very rapidly. Thus, $\hat{\mu}$ approaches y very rapidly as y becomes large and positive. For $y < 0$, $\hat{\mu}$ can be expressed in terms of the continued fraction given in (4.1.3) and (4.1.4), *i.e.*,

$$\hat{\mu} = y + 1/CF(-y) \ .$$

The value of a continued fraction,

$$\frac{a_1}{b_1}\,\frac{a_2}{b_2}\,\frac{a_3}{b_3}\,\cdots ,$$

lies between two successive convergents if, for every term a_i/b_i of the continued fraction, a_i and b_i are positive (Abramowitz and Stegun (1964), page 19). The terms of $CF(x)$ have positive integers for a_i, b_i is equal to $-y$, while negative values of y are being examined, so the value of $CF(-y)$ would be in the interval

$$\left(\frac{1}{-y}, \frac{-y}{y^2 + 1}\right) ,$$

i.e., between the first two convergents. Thus, $\hat{\mu}$ must be between L and U, where:

$$L = y + \frac{1}{\frac{1}{-y}} = 0$$

and

$$U = y + \frac{1}{\frac{-y}{y^2+1}} = 1/-y .$$

As y tends to $-\infty$, U approaches zero, as $\hat{\mu}$ approaches zero as y goes to $-\infty$. Figure 4.1 contains a plotting of $\hat{\mu}$ as a function of y for the interval $[-10,10]$.

Assuming an exponential prior over the feasible parameter space for the same density of the observed variable gives somewhat different Bayes estimators. For this case, the mean of the posterior distribution is:

$$\hat{\mu} = \frac{\int_0^{\infty} \mu (\sqrt{2\pi})^{-1} e^{-(y-\mu)^2/2} e^{-\mu/\theta} d\mu}{\int_0^{\infty} (\sqrt{2\pi})^{-1} e^{-(y-\mu)^2/2} e^{-\mu/\theta} d\mu}$$

$$= \frac{\int_0^{\infty} \mu e^{-((y-1/\theta)-\mu)^2/2} d\mu}{\int_0^{\infty} e^{-((y-1/\theta)-\mu)^2/2} d\mu}$$

$$= y - 1/\theta + f(y-1/\theta)/F(y-1/\theta) \qquad (4.3.5)$$

as follows from (4.1.2), since the posterior distribution here is a normal distribution with mean $y - 1/\theta$ truncated at zero.

Figure 4.2 gives a plot of the estimators, $\hat{\mu}$, for several values of θ. Note that as y becomes large and positive, $\hat{\mu}$ tends to $y - 1/\theta$, which for θ near zero is quite different from the maximum likelihood estimator. No point in the feasible parameter space is excluded as an estimate of μ.

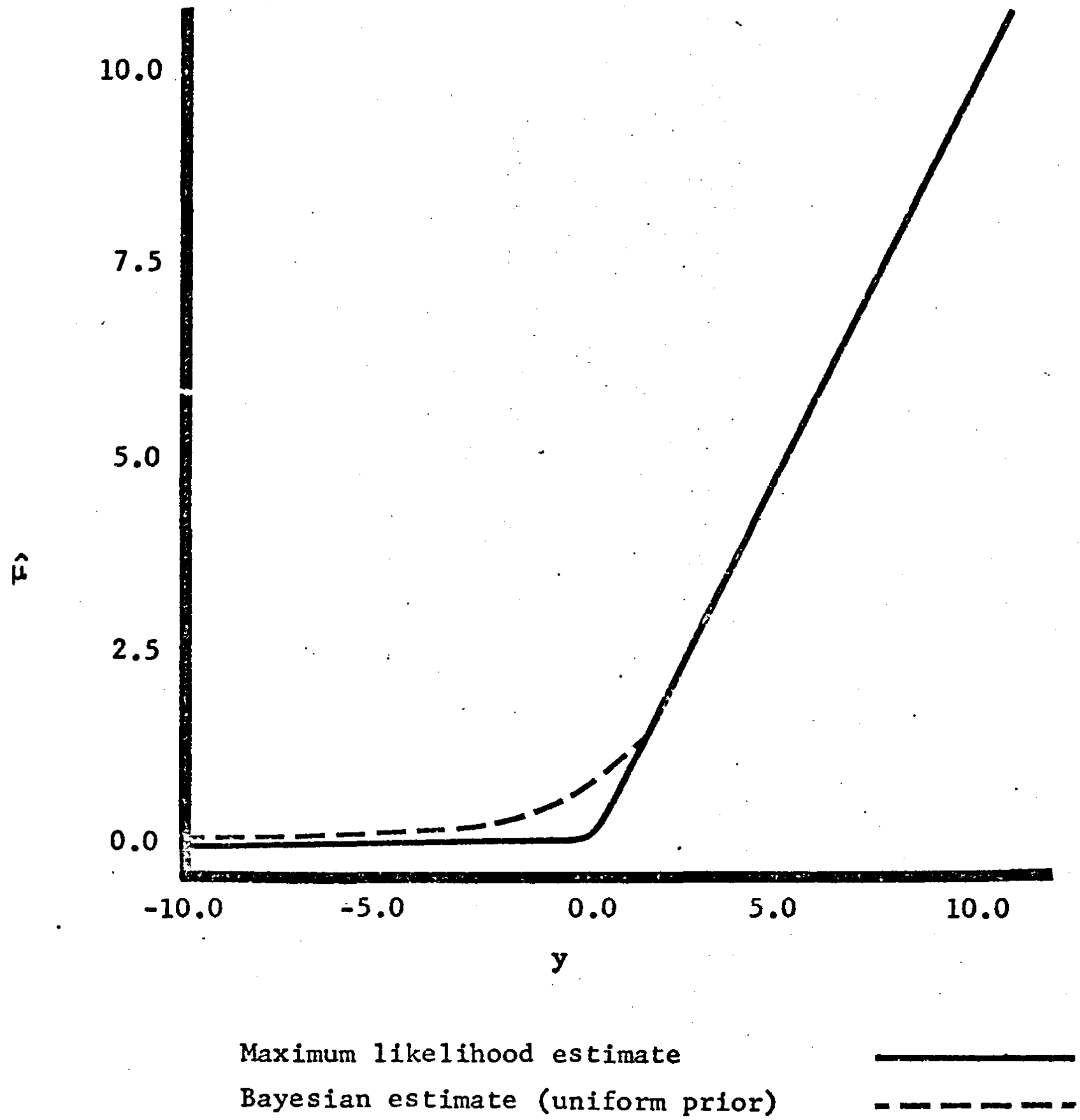


Figure 4.1 Bayes estimates (uniform prior) and maximum likelihood estimates of the mean, $\mu \geq 0$, of a normal distribution when the observation is y

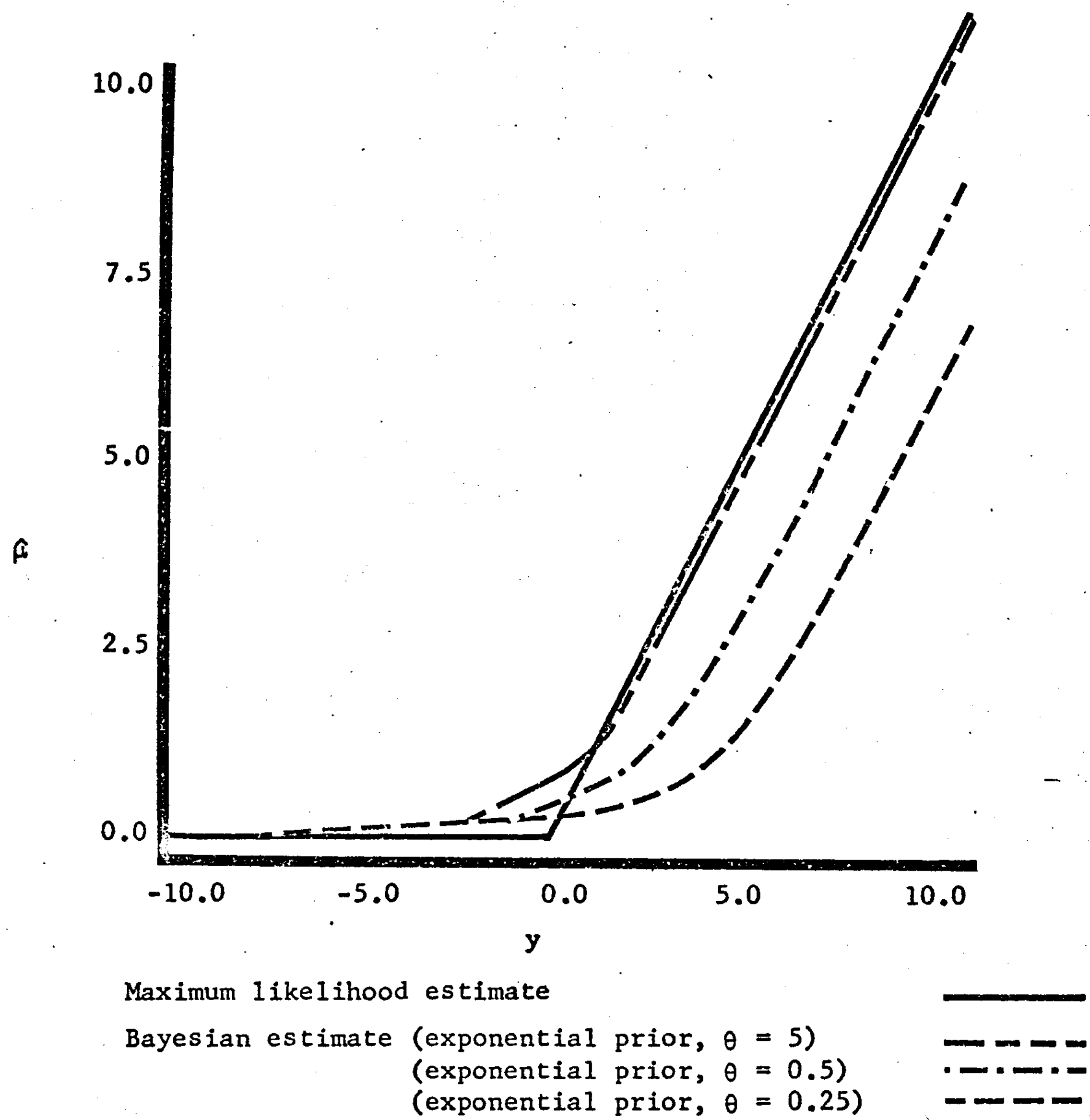


Figure 4.2 Bayes estimates (exponential priors) and maximum likelihood estimates of the mean, $\mu \geq 0$, of a normal distribution when the observation is y

For the same density of the observed variables and a <u>truncated normal</u> prior with positive probability over $[0, \infty)$, mean θ, and variance one, the mean of the posterior is:

$$\hat{\mu} = \frac{\int_0^\infty \mu e^{-(y-\mu)^2/2} e^{-(\theta-\mu)^2/2} d\mu}{\int_0^\infty e^{-(y-\mu)^2/2} e^{-(\theta-\mu)^2/2} d\mu}$$

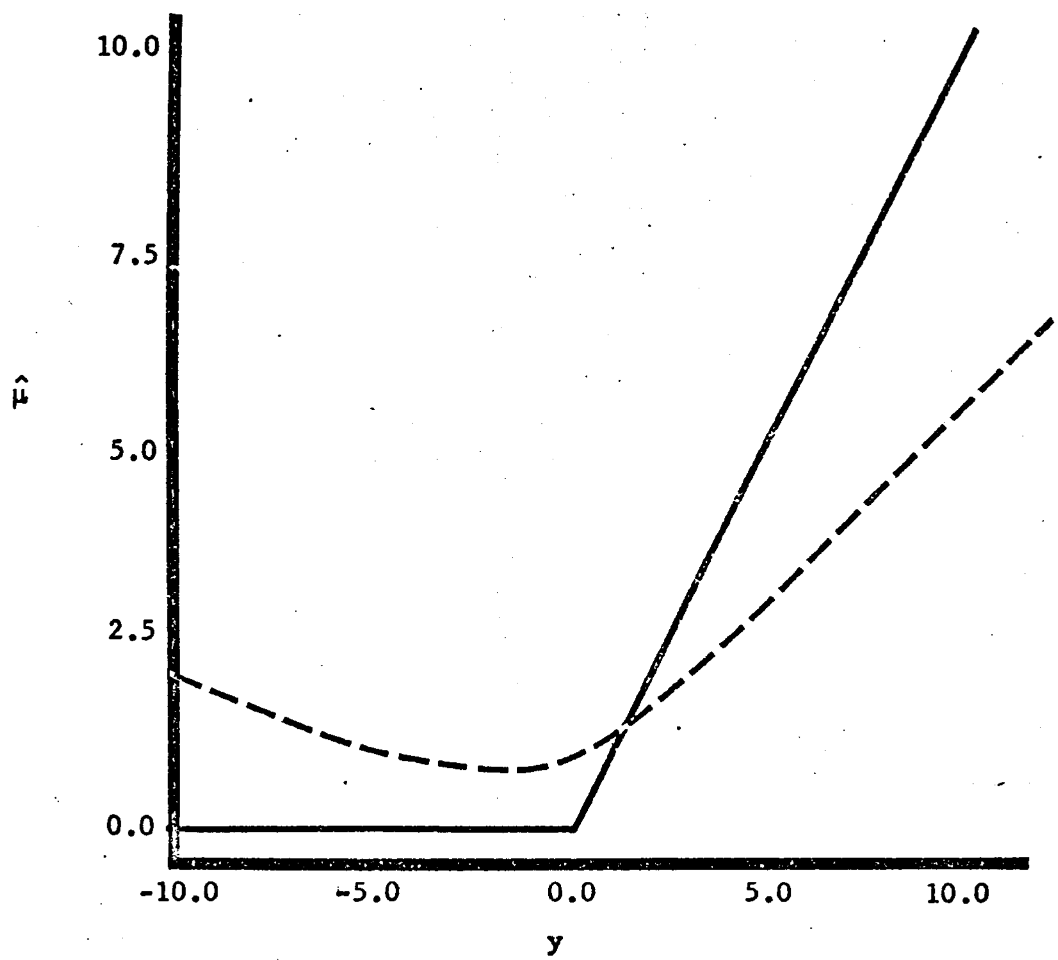
$$= \frac{y+\theta}{2} + \frac{f((y+\theta)/\sqrt{2})}{F((y+\theta)/\sqrt{2})} .$$

The values for $\hat{\mu}$ when $y \in [-10, 10]$, and $\theta = 1$ are shown in Figure 4.3. It should be noted that for y negative, $\hat{\mu}$ approaches zero. For y positive and large, $\hat{\mu}$ approaches $(y+1)/2$. Again, all points in the feasible parameter space are found as estimates of μ for some value of y.

For the case of the bivariate truncated normal, as is given in (8.3.18), with a simple ordering on the parameters μ_i, again all points in the feasible parameter space can be found as estimates, for some value of $\underline{y}$. The expression in (8.3.19) gives the expected value of this posterior. When $(y_2 - y_1)$ is negative, μ_1 and μ_2 can be written as continued fraction as was shown in (4.1.3). Thus,

$$\left.\begin{aligned} &\hat{\mu}_1 = y_1 - \sigma/(\sqrt{2}\ CF((y_2-y_1)/\sqrt{2}\sigma)) \\ \text{and} & \\ &\hat{\mu}_2 = y_2 + \sigma/(\sqrt{2}\ CF((y_2-y_1)/\sqrt{2}\sigma)) \end{aligned}\right\} . \qquad (4.3.6)$$

Then $\hat{\mu}_1$ would be between L and U where:



Maximum likelihood estimate

Bayesian estimate (prior, a normal (1,1) truncated at zero)

Figure 4.3 Bayes estimates (truncated normal prior) and maximum likelihood estimates of the mean, $\mu \geq 0$, of a normal distribution when the observation is y

$$L = y_1 + \frac{\sigma}{\sqrt{2}} \frac{\left(\frac{y_2-y_1}{\sqrt{2}\,\sigma}\right)^2}{\left(\frac{y_2-y_1}{\sqrt{2}\,\sigma}\right)} + 1 = \frac{y_2+y_1}{2} + \frac{\sigma^2}{y_2-y_1}$$

$$U = y_1 + \frac{\sigma}{\sqrt{2}} \frac{y_2-y_1}{\sqrt{2}\,\sigma} = \frac{y_2+y_1}{2} .$$

The values of L and U are found by substituting the first two convergents for $CF(x)$ in μ_1. As (y_2-y_1) goes to $-\infty$, L approaches $(y_2+y_1)/2$; so $\mu_1 - (y_2+y_1)/2$ tends to zero as (y_2-y_1) approaches $-\infty$.

Similarly, it can be shown that $\mu_2 - (y_2+y_1)/2$ tends to zero as (y_2-y_1) approaches $-\infty$. This expression, $(y_2+y_1)/2$, is the same as would be found by isotonic regression for a simple ordering on μ_1 and μ_2 when the basic estimates violate the ordering on the parameters. (See the discussion following (2.3.2).) Thus, as (y_2-y_1) approaches $-\infty$, the Bayesian estimates $\hat{\mu}_1$ and $\hat{\mu}_2$ tend to the maximum likelihood estimates. As (y_2-y_1) becomes large and positive, $f((y_2-y_1)/\sigma\sqrt{2})$ approaches zero rapidly; $F((y_2-y_1)/\sigma\sqrt{2})$ tends to one; and $\hat{\mu}_1$ becomes y_1 and $\hat{\mu}_2$ becomes y_2. These limits are again the isotonic regression estimates when the basic estimates satisfy the ordering on the parameters. Since the isotonic regression estimates are on the boundary of the feasible parameter space when the basic solution violates the ordering, these Bayesian estimates will take on values in the neighborhood of the boundary with a positive probability.

It would be desirable to determine if, in general, the Bayesian estimates are close to the maximum likelihood estimates for some observations. This is not possible analytically because of the complexity of the multinormal integral, which Kendall and Stuart (1969, pp. 350-353) have pointed out. These difficulties would not arise when the n-variate density function is a product of n independent univariate normal density functions. It is difficult to conceive of practical situations in which a truncated normal posterior would have this property. For this reason, the remaining discussion will be limited to the univariate and bivariate truncated normal posteriors.

The univariate and bivariate Bayesian estimators discussed here are usually consistent in the case of a uniform prior. In both cases, the Bayesian estimator (cf. (4.3.4) and (4.3.6)) consists of the unrestricted maximum likelihood estimator with an additional term of the form

$$\pm \lambda f(b/\lambda)/F(b/\lambda) . \tag{4.3.7}$$

Here λ is the standard deviation of the unrestricted maximum likelihood estimator times a positive constant, c . Thus, if the random variable is distributed normally with variance σ^2 , the unrestricted maximum likelihood estimator has variance σ^2/n for a sample of size n . By application of (4.2.1) and (4.1.2), the Bayesian estimator is

$$\hat{\mu} = \bar{y} + \frac{\sigma}{n} f((\bar{y}-a)(\sqrt{n}/\sigma))/F((\bar{y}-a)(\sqrt{n}/\sigma)) .$$

Therefore, (4.3.6) is

$$\frac{c\sigma}{n} f(b\sqrt{n}/(c\sigma))/F(b\sqrt{n}/(c\sigma)) \ . \tag{4.3.8}$$

The value of b in the univariate case is $(\bar{y}-a)$ for the uniform prior in which $\mu > a$; see (4.1.2). Then as was shown in (2.4.1),

$$\Pr(\bar{y} - a \leq 0) = 0$$

as sample size goes to ∞. Then b will be positive with probability one, $b\sqrt{n}/(c\sigma)$ will approach ∞, and

$$f(b\sqrt{n}/(c\sigma))/F(b\sqrt{n}/(c\sigma))$$

becomes zero, since $f(b\sqrt{n}/(c\sigma))$ becomes zero and $F(b\sqrt{n}/(c\sigma))$ approaches one.

In the bivariate case b is $(\bar{y}_2-\bar{y}_1)$ where $\mu_1 < \mu_2$. Again by (2.4.1)

$$\Pr(\bar{y}_2 - \bar{y}_1 \leq 0) = 0$$

as n goes to ∞. So by the same argument, (4.3.8) becomes zero. Thus, as n becomes large, these Bayesian estimators approach the unrestricted maximum likelihood estimator which is consistent.

Notice that these estimators are not consistent when the feasible parameter space is a closed set. For example when $\mu \geq a$, the argument for $f(x)$ and $F(x)$ in (4.3.8) would approach zero when $\mu = a$. Then (4.3.8) would approach a positive quantity as n increased, and therefore, the estimator would not be consistent.

For the normal and exponential priors discussed here, it is possible that b is negative as n goes to ∞. For example b is equal to $y - \sigma^2/\theta$ for the exponential prior with $\mu \geq 0$. If the true value of μ were less than σ^2/θ, b would be negative with probability one as n became large. Then since the argument of $f(x)/F(x)$ would approach $-\infty$, the estimate would become zero. Thus, if μ were larger than zero, the Bayesian estimator would not be consistent. The same situation exists for the truncated normal prior distribution.

4.4 Comparison of Mean Square Errors of Restricted Maximum Likelihood Estimators and Bayesian Estimators

The mean of a truncated normal posterior does seem to solve the problem of accumulating estimates on the boundary points that the restricted maximum likelihood estimator presented. This gain is made without incurring the problems anticipated in Section 3.1. That is, estimates are found in any neighborhood of the boundary for the univariate and bivariate truncated normal posterior.

The restricted maximum likelihood estimators are consistent. Bayesian estimators for a uniform prior with support on an open set are also consistent; however, other priors do not necessarily lead to consistent estimators.

Another fitting comparison of the Bayes and maximum likelihood estimators is with respect to mean square error. Does one of these estimators have mean square error uniformly smaller than any other? This question will be studied in depth for the univariate case.

Without loss of generality, numerical examples will only be given for prior distributions with support $[0, \infty)$.

The restricted likelihood estimator of one observation from the univariate normal, with mean known to be larger than a , is

$$\hat{\mu}_{ML} = a \quad \text{when} \quad y < a ,$$

$$= y \quad \text{when} \quad y \geq a .$$

The mean square error for this estimator is

$$MSE(\hat{\mu}_{ML}) = \int_{-\infty}^{a} \frac{(a-\mu)^2}{\sqrt{2}\,\pi\sigma} e^{-(y-\mu)^2/(2\sigma^2)} dy$$

$$+ \int_{a}^{\infty} \frac{(y-\mu)^2}{\sqrt{2}\,\pi\sigma} e^{-(y-\mu)^2/(2\sigma^2)} dy .$$

Integrating by parts, the last integral becomes

$$\int_{a}^{\infty} \frac{(y-\mu)^2}{\sqrt{2\pi}\,\sigma} e^{-(y-\mu)^2/(2\sigma^2)} dy = -\sigma \frac{(y-\mu)}{\sqrt{2}\,\pi} e^{-(y-\mu)^2/(2\sigma^2)} \Big|_{a}^{\infty}$$

$$+ \int_{a}^{\infty} \frac{\sigma}{\sqrt{2}\,\pi} e^{-(y-\mu)^2/(2\sigma^2)} dy .$$

Then,

$$MSE(\hat{\mu}_{ML}) = (a-\mu)^2 F((a-\mu)/\sigma) + (a-\mu)\sigma f((a-\mu)/\sigma)$$

$$+ \sigma^2 [1 - F((a-\mu)/\sigma)] . \qquad (4.4.1)$$

For the same sampling density and a uniform prior for μ greater than a , the mean square error of the Bayes estimator is

$$MSE(\hat{\mu}_U) = \int_{-\infty}^{\infty} (y-\sigma f((y-a)/\sigma)/F((y-a)/\sigma)-\mu)^2$$

$$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-\mu)^2/(2\sigma^2)} dy \; . \qquad (4.4.2)$$

This expression does not lend itself well to analytical examination. However, numerical approximations of this formula can be found by applying Gaussian-Hermite quadrature formulas. An explanation of the technique is contained in Ghizzetti and Ossicini (1970). The computer programs used to evaluate (4.4.2) were DQH32 and DQH64 in the System/360 Scientific Subroutine Package (1970). Figure 4.4 gives a plotting of (4.4.1) and (4.4.2) for a equal zero, σ^2 equal to one, and $\mu \varepsilon [0,8]$. (The values of these functions from which Figure 4.4 was made are given in Table 4.1.) Neither estimator has uniformly smaller mean square error than the other.

The Bayesian estimator from an exponential prior and with the same density of the observed variable has the following mean square error; (in this case a is set equal to zero)

$$MSE(\hat{\mu}_E) = \int_{-\infty}^{\infty} [y-\sigma^2/\theta+\sigma f((y-\sigma^2/\theta)/\sigma)/F((y-\sigma^2/\theta)/\sigma)$$

$$-\mu]^2 \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y-\mu)^2/(2\sigma^2)} dy \; . \qquad (4.4.3)$$

Figure 4.5 was found by evaluating (4.4.3) by program DQH64 , see above. Again, σ^2 was set equal to one. As can be seen in Figure 4.5, the estimates found from exponential priors do not give uniformly smaller mean square errors than the restricted maximum likelihood estimates either. In fact, the Bayesian estimator which gives the

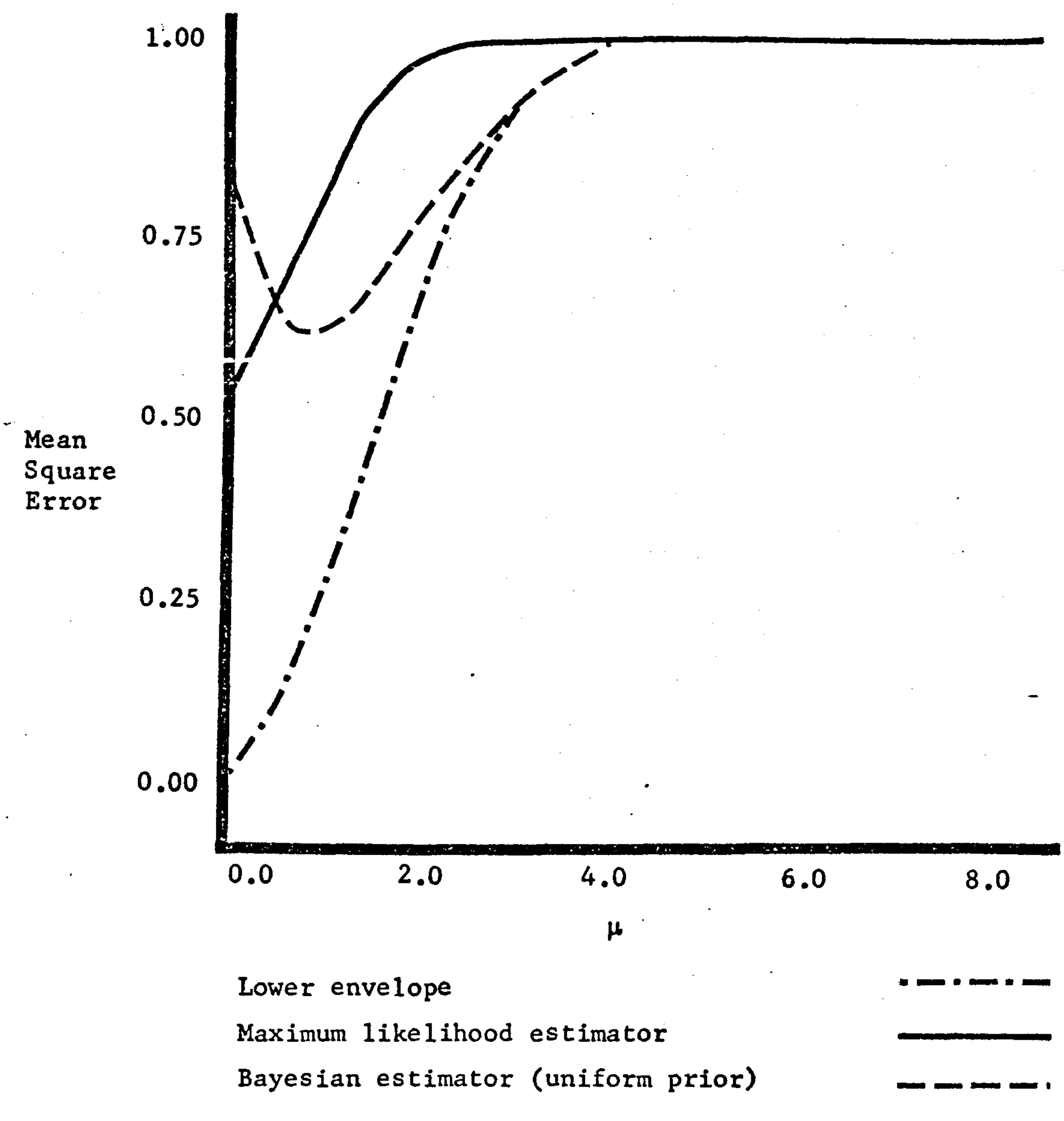


Figure 4.4 Plots of the lower envelope[a] and the mean square error for the maximum likelihood estimator and a Bayesian estimator (uniform prior)

---

[a]The term lower envelope will be introduced in Chapter 5

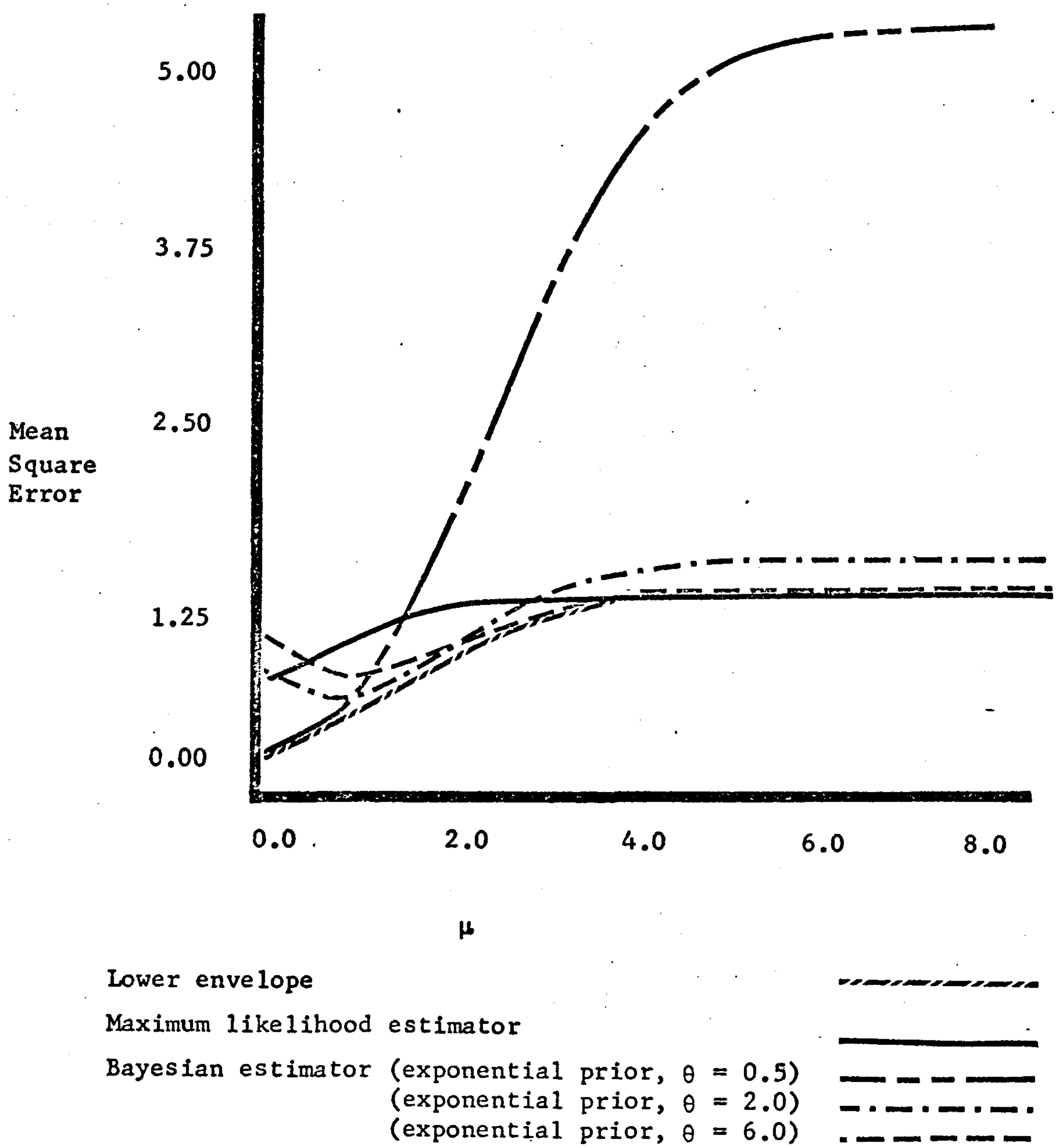


Figure 4.5 Plots of the lower envelope and mean square error for the maximum likelihood estimator and several Bayesian estimators (exponential priors)

Table 4.1 Mean square error for the maximum likelihood estimator and a Bayesian estimator (uniform prior)

μ	MSE Bayesian estimator	MSE ML estimator	μ	MSE Bayesian estimator	MSE ML estimator
0.1	0.91554	0.50473	4.1	0.98171	0.99996
0.2	0.84237	0.51788	4.2	0.98485	0.99997
0.3	0.77987	0.53788	4.3	0.98751	0.99998
0.4	0.72739	0.56325	4.4	0.98977	0.99999
0.5	0.68427	0.59256	4.5	0.99165	0.99999
0.6	0.64983	0.62454	4.6	0.99323	1.00000
0.7	0.62339	0.65802	4.7	0.99453	1.00000
0.8	0.60427	0.69198	4.8	0.99561	1.00000
0.9	0.59179	0.72555	4.9	0.99649	1.00000
1.0	0.58528	0.75803	5.0	0.99721	1.00000
1.1	0.58408	0.78885	5.1	0.99780	1.00000
1.2	0.58754	0.81761	5.2	0.99827	1.00000
1.3	0.59504	0.84401	5.3	0.99864	1.00000
1.4	0.60598	0.86791	5.4	0.99894	1.00000
1.5	0.61979	0.88923	5.5	0.99918	1.00000
1.6	0.63592	0.90801	5.6	0.99937	1.00000
1.7	0.65388	0.92435	5.7	0.99952	1.00000
1.8	0.67320	0.93837	5.8	0.99963	1.00000
1.9	0.69346	0.95028	5.9	0.99972	1.00000
2.0	0.71426	0.96027	6.0	0.99979	1.00000
2.1	0.73526	0.96855	6.1	0.99984	1.00000
2.2	0.75617	0.97535	6.2	0.99988	1.00000
2.3	0.77672	0.98085	6.3	0.99991	1.00000
2.4	0.79669	0.98527	6.4	0.99994	1.00000
2.5	0.81590	0.98878	6.5	0.99995	1.00000
2.6	0.83421	0.99153	6.6	0.99997	1.00000
2.7	0.85151	0.99367	6.7	0.99997	1.00000
2.8	0.86771	0.99531	6.8	0.99998	1.00000
2.9	0.88277	0.99656	6.9	0.99999	1.00000
3.0	0.89667	0.99750	7.0	0.99999	1.00000
3.1	0.90939	0.99820	7.1	0.99999	1.00000
3.2	0.92096	0.99872	7.2	1.00000	1.00000
3.3	0.93142	0.99910	7.3	1.00000	1.00000
3.4	0.94080	0.99937	7.4	1.00000	1.00000
3.5	0.94916	0.99956	7.5	1.00000	1.00000
3.6	0.95656	0.99970	7.6	1.00000	1.00000
3.7	0.96308	0.99980	7.7	1.00000	1.00000
3.8	0.96878	0.99986	7.8	1.00000	1.00000
3.9	4.97374	0.99991	7.9	1.00000	1.00000
4.0	0.97803	0.99994	8.0	1.00000	1.00000

greatest improvement for smaller values of μ, performs the poorest as μ increases.

Now the case of the truncated normal prior will be considered. The prior examined here will be proportional to a normal density with mean parameter λ and variance δ over the interval $[0,\infty)$, and the prior will be zero elsewhere. The observations again have a univariate normal density with mean μ and variance σ^2. Then the posterior is the univariate case of (4.2.4) and using (4.1.2) the Bayesian estimator is found to be:

$$\hat{\mu} = \frac{\delta y+\sigma^2\lambda}{\delta+\sigma^2} + \sqrt{\frac{\delta\sigma^2}{\delta+\sigma^2}}\,\frac{f((\delta y+\sigma^2\lambda)/(\delta+\sigma^2))}{F((\delta y+\sigma^2\lambda)/(\delta+\sigma^2))}\;.$$

The mean square error of this estimator is

$$\mathrm{MSE}(\hat{\mu}_N) = \int_{-\infty}^{\infty} (\hat{\mu}-\mu)^2 \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(y-\mu)^2/(2\sigma^2)}dy\;.$$

This function was also evaluated by Gaussian-Hermite quadrature for σ^2, λ, and δ equal to one. A plot of the values of this function are shown in Figure 4.6 for μ in the interval $[0,8]$. On this interval, the mean square error for the Bayesian estimator was smaller only in the neighborhood of λ. The same conclusion can be drawn from Figure 4.7 in which the mean square error is plotted for the same example with λ set equal to 3.

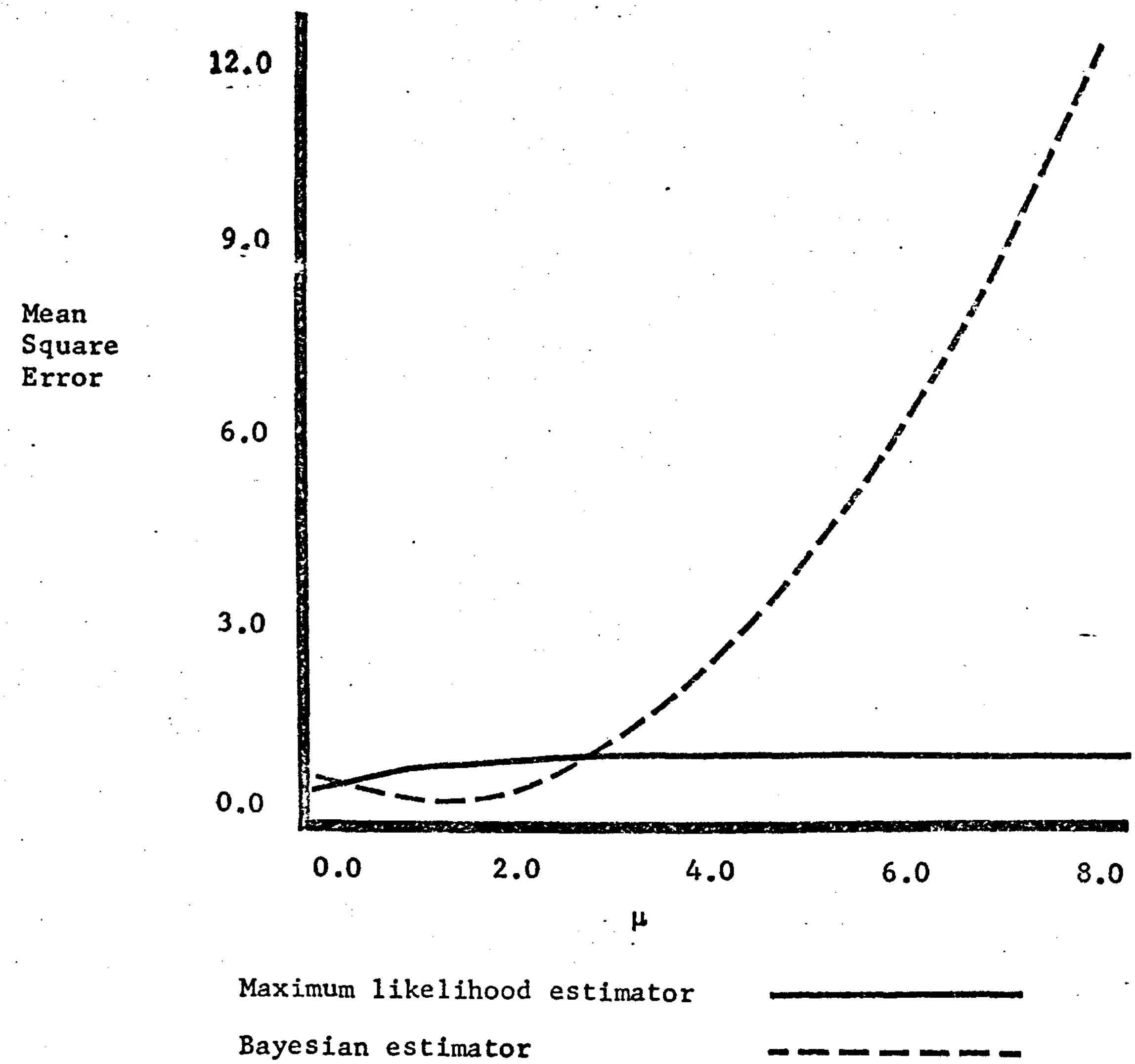


Figure 4.6 Plots of the mean square error for the maximum likelihood estimator and a Bayesian estimator (prior, normal (1,1) truncated at zero)

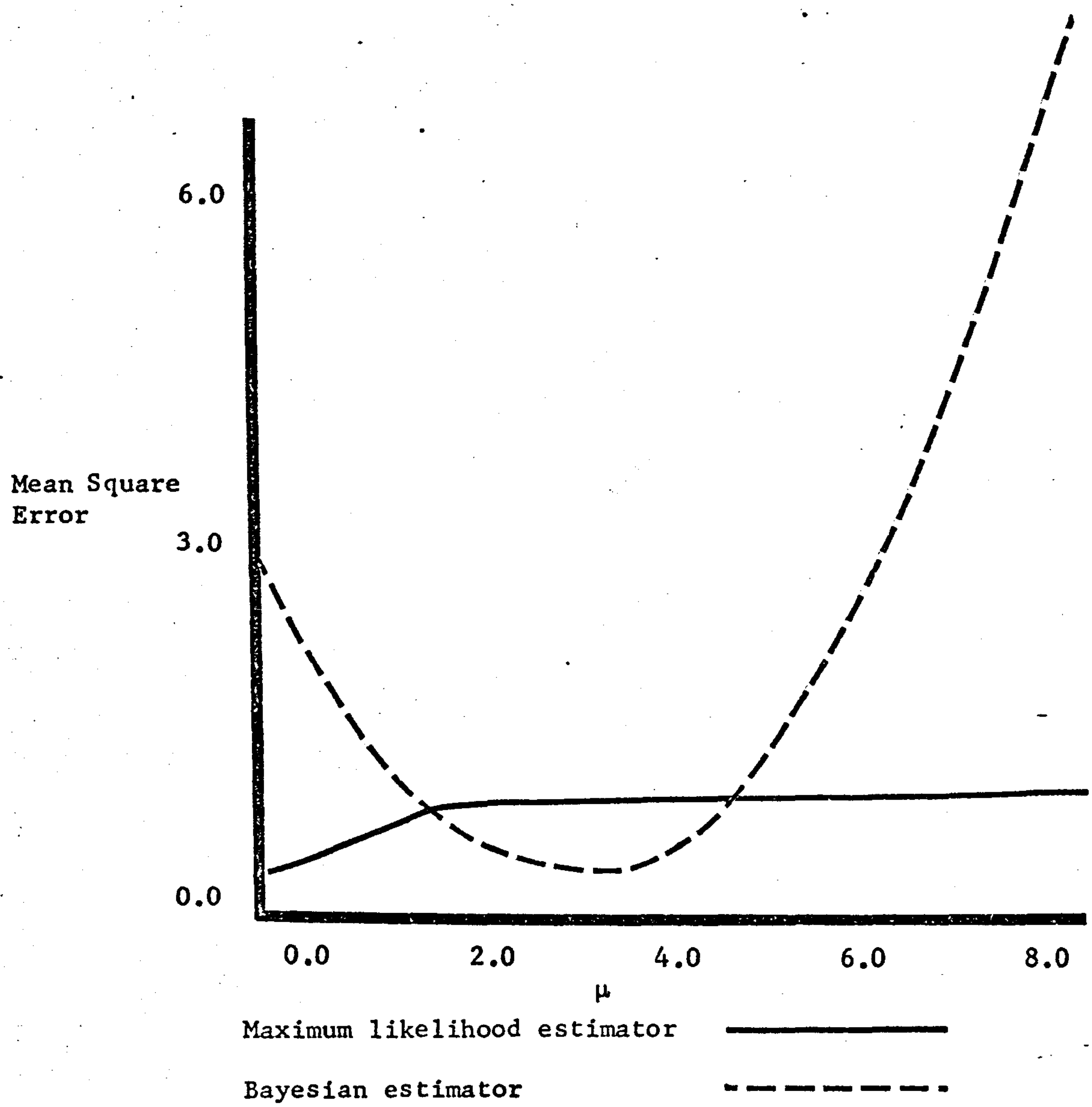


Figure 4.7 Plots of the mean square error for the maximum likelihood estimator and a Bayesian estimator (prior, normal (3,1) truncated at zero)

5. IMPROVED ESTIMATORS

5.1 Joining Estimators

It was shown in Section 4.4 that none of the Bayesian estimators presented have uniformly smaller mean square error than do restricted maximum likelihood estimators and *vice versa*. However, if the true value of the parameter μ happened to be near the boundary of the feasible parameter space, an exponential prior has been found which gave a smaller mean square error of the resulting statistic for the values of μ near the boundary (see Figure 4.5). This improvement in mean square error for values of μ near the boundary corresponds to sacrifices in mean square error for values of μ away from the boundary. The restricted maximum likelihood estimator had larger mean square error near the boundary, but is vastly superior to the Bayesian estimators found from exponential priors at points farther from the boundary. The Bayesian estimators found from a uniform prior had a mean square error which was smaller than that of the restricted maximum likelihood estimator for values of μ in the feasible parameter space away from the boundary, and larger near the boundary. (The uniform prior will not be considered separately in the remainder of this paper since it can be derived as a limiting case of exponential priors.)

All this suggests that a modeler having information only that the mean, μ, of some normal density function belongs to a certain half-line might try to combine the better properties of both types of estimators.

Combined estimators are not foreigh to statisticians. In fact, the restricted maximum likelihood procedures mentioned in Chapter 2 are essentially combined estimators. If the unrestricted estimates are points in the feasible parameter space, they are the restricted maximum likelihood estimates. If the unrestricted estimates are not points in the feasible parameter space, another algorithm is employed to produce the restricted estimates.

Other combined estimators have been considered for entirely different situations. Bancroft (1944), Mosteller (1948), and Gun (1965) have studied estimation procedures with a preliminary significance test. Their estimators are found by first testing to determine if the estimates from several populations are significantly different. If significance is found, individual population estimates are used. Otherwise, the estimates from the various populations are pooled; note that the significance levels recommended for these situations are larger than the significance levels normally used.

Consider the univariate normal density with a mean known to be not less than d. An estimator with generally smaller mean square error could hopefully be created by using a Bayesian estimator derived from the exponential prior when the unrestricted maximum likelihood estimate is near the boundary or falls outside the feasible parameter space. The unrestricted maximum likelihood estimate would be taken as the estimate in all other situations. Finding such an estimator which does give a reduction in mean square error is a formidable task. A good value θ of the parameter of the exponential prior must be found, and the part of sample space in which the maximum likelihood estimator is to be used must be determined. Of course, a criterion of goodness must be established to dictate the choices.

5.2 The Criterion of Regret

In the framework of statistical decision theory, the mean square error of any estimator is often regarded as the expected value of the (quadratic) loss suffered when using that estimator (the loss being a consequence of the fact that the estimator is not equal to the value of the parameter being estimated). The expected loss is a function of μ; its value also depends on the estimation procedure used: thus, in the case of Bayesian estimators, it depends on the θ characterizing a particular prior within a family of priors; more basically, it depends on the family of priors. Similarly, it depends on whether one uses a (restricted or unrestricted) maximum likelihood estimator or a Bayesian estimator. The task of somehow combining several estimators; each of which is 'good' for some μ-values, 'poor' for others; must be confronted. Now, for each point μ in the feasible parameter space we can determine the infimum of the expected loss corresponding to all competing estimators; the value of this infimum will, of course, depend on the class of competing estimators. Thus, a function of μ which will be called the lower envelope (of the expected loss function) will be defined. This lower envelope indicates the best we can possibly do with the available estimators if for each experiment the true μ-value is known. Since this is not known, the expected loss can be no smaller than the lower envelope; no matter how the estimators previously studied are combined. Thus, it must be accepted that the mean square error of the combined estimators for most μ-values will exceed the value of the lower envelope. The difference between the two will be called regret (cf. Savage (1954), Sections 9.4, 9.5, and 9.8, and Savage (1968)). A combined estimator will be sought which minimizes this regret, which again, depends on the class of competing estimators and, of course, on μ.

The plan is to define a real-valued criterion summarizing the behavior of this regret function over the feasible parameter space, then to select such a 'combination' of the above estimators as to make this real number as small as possible. There are many such criteria available. Gun (1965) suggested using the L_1-norm of the regret function in the situation he studied. Other L_n-norms are also candidates. Of course, the computation of such norms requires selection of a measure over the feasible parameter space. A criterion which can be implemented with less difficulty is maximizing the regret function over the feasible parameter space and minimizing this over the competing estimators. Thus, the criterion would be minimax regret. Minimax procedures are described by Wald (1950) and Savage (1954). As Wald has stated, minimax is applicable when a particular prior cannot be justified. This is more in line with the situation proposed in this section. The minimax criterion is a pessimistic approach, but it does protect against large losses.

5.3 The Application of Minimax Regret to the Construction of a Joined Estimator

Consider again a sample of size one, y, from a normal distribution with unknown mean $\mu \geq d$ and known variance σ^2. The objective now is to investigate joined estimators of the form

$$\hat{\mu}_J = \hat{\mu}_\theta \quad \text{for} \quad y < \alpha ,$$

$$= \hat{\mu}_{ML} \quad \text{for} \quad y \geq \alpha ,$$

where $\alpha > d$, $\hat{\mu}_\theta$ denotes the Bayesian estimator corresponding to an exponential prior with parameter θ, and $\hat{\mu}_{ML}$ denotes the (unrestricted) maximum likelihood estimator. The regret function for such an estimator thus depends on μ, θ and α will be denoted by $R(\mu, \theta, \alpha)$. The objective is to choose θ and α so as to minimize

$$\max_{\mu \geq d} R(\mu, \theta, \alpha) . \tag{5.3.1}$$

The pair (θ', α') which minimizes (5.3.1) characterizes the optimum combined estimator, i.e., one chooses the Bayesian estimator corresponding to the exponential prior with parameter θ' when the unrestricted maximum likelihood estimate is less than α', and chooses the unrestricted maximum likelihood estimate otherwise.

To find the values of α and θ which minimize (5.3.1), one first must determine the lower envelope of the family of mean square error curves. The initial step is to determine the lower envelope of the mean square error (see (4.3.5)) of all Bayesian estimators corresponding to an exponential prior with $\theta \in (0, \infty)$. Then it will turn out that for no value of μ the mean square error of the restricted maximum likelihood estimator or the mean square error of the Bayesian estimator corresponding to the uniform prior is less than the constructed lower envelope. Therefore, this lower envelope is the lower envelope for the class of competing estimators mentioned in Section 5.2. An approximation for it was found by numerical methods. This will be done first for the case of d=0 and $\sigma^2 = 1$. Table 5.1 gives the approximation to the lower envelope that was found as follows. For μ equal 0.1 or 0.2, candidate values for θ were found by increasing θ by steps of length 0.1. For values for $\mu \in [0.3, 8.0]$ such candidate values were found by incrementing θ either by 0.1 or by half the difference in the optimizing θ for the two preceding values of μ, whichever was larger. By comparing the values in Table 4.1 with the values in Table 5.1 the reader will easily convince himself that the function tabulated in Table 5.1 gives the sought after lower envelope.

The next step was finding the mean square error for the joined estimator. This mean square error is (cf. equation (4.3.4))

$$\mathrm{MSE}(\hat{\mu}_J) = \int_{-\infty}^{\alpha} [y - 1/\theta + f(y - 1/\theta)/F(y - 1/\theta) - \mu]^2 \frac{1}{\sqrt{2\pi}} e^{-(y-\mu)^2/2} \, dy + \int_{\alpha}^{\infty} (y-\mu)^2 \frac{1}{\sqrt{2\pi}} e^{-(y-\mu)^2/2} \, dy \qquad (5.3.2)$$

for any given values of α and μ. The second term in this expression reduces to (cf. equation (4.4.1))

$$\int_{\alpha}^{\infty} (y-\mu)^2 \frac{1}{\sqrt{2\pi}} e^{-(y-\mu)^2/2} \, dy = -(\alpha-\mu)f(\alpha-\mu) + 1 - F(\alpha-\mu) \qquad (5.3.3.)$$

where f(x) and F(x) are normal density and distribution functions respectively for the univariate normal distribution with mean zero and variance one.

Table 5.1 Approximation for the lower envelope of the mean square errors for the estimators derived from the truncated normal posterior formed with exponential priors

μ	θ giving the minimum MSE	Minimum MSE	μ	θ giving the minimum MSE	Minimum MSE
0.1	0.1	0.00010	4.1	43.5	0.98122
0.2	0.2	0.00156	4.2	52.2	0.98451
0.3	0.3	0.00617	4.3	65.1	0.98728
0.4	0.4	0.01551	4.4	78.1	0.98960
0.5	0.4	0.03034	4.5	91.1	0.99154
0.6	0.5	0.05150	4.6	110.5	0.99315
0.7	0.6	0.07704	4.7	139.7	0.99448
0.8	0.6	0.10812	4.8	168.9	0.99558
0.9	0.7	0.14293	4.9	212.7	0.99647
1.0	0.8	0.18153	5.0	256.5	0.99720
1.1	0.9	0.22284	5.1	322.2	0.99779
1.2	1.0	0.26607	5.2	420.7	0.99826
1.3	1.1	0.31062	5.3	519.2	0.99864
1.4	1.2	0.35597	5.4	667.0	0.99894
1.5	1.4	0.40132	5.5	814.8	0.99918
1.6	1.5	0.44628	5.6	1036.5	0.99937
1.7	1.7	0.49040	5.7	1369.0	0.99952
1.8	1.9	0.53311	5.8	1701.6	0.99963
1.9	2.1	0.57427	5.9	2366.6	0.99972
2.0	2.4	0.61357	6.0	3031.7	0.99979
2.1	2.7	0.65079	6.1	4029.2	0.99984
2.2	3.0	0.68582	6.2	5026.8	0.99988
2.3	3.4	0.71854	6.3	7022.0	0.99991
2.4	3.7	0.74895	6.4	9017.1	0.99994
2.5	4.2	0.77700	6.5	12010.0	0.99995
2.6	4.7	0.80276	6.6	16499.0	0.99997
2.7	5.5	0.82625	6.7	23232.6	0.99997
2.8	6.2	0.84761	6.8	29966.2	0.99998
2.9	7.0	0.86690	6.9	43433.9	0.99999
3.0	8.1	0.88423	7.0	56900.6	0.99999
3.1	9.3	0.89972	7.1	77101.7	0.99999
3.2	10.4	0.91352	7.2	107403.2	1.00000
3.3	12.1	0.92572	7.3	152855.3	1.00000
3.4	14.7	0.93648	7.4	221033.5	1.00000
3.5	17.2	0.94592	7.5	323300.6	1.00000
3.6	19.9	0.95415	7.6	425568.0	1.00000
3.7	22.4	0.96130	7.7	630102.3	1.00000
3.8	26.2	0.96748	7.8	936904.0	1.00000
3.9	32.0	0.97280	7.9	1243705.7	1.00000
4.0	37.7	0.97735	8.0	1857309.1	1.00000

The first term of (5.3.2) must be evaluated by numerical procedures. An algorithm very useful in the minization over α, evaluates this term by the Hermitian formula using the first derivative (see System 360/Scientific Subroutine Package (1970) subprogram DQHFE, or Hildebrand (1956)). This algorithm approximates the value of the integral at several equidistant points over the interval of integration as follows. Define

$$z_i = z_i(x_i) = \int_a^{x_i} y(x)\ dx$$

at equidistant points x_i which satisfy the following relationship

$$x_i = a + (i-1)h \ .$$

The value of z_1 is assigned to be zero, and all other values of z_i are found by the formula

$$z_i = z_{i-1} + h(y_{i-1}+y_i+h(y'_{i-1} - y'_i)/6)/2 \ .$$

where y'_i is the derivative of the function $y(x)$ at x_i and y_i is equal to $y(x_i)$. The maximum error will be less than

$$s\ h^4\ y^4(v)/720$$

where s is the total length of the interval and $y^4(v)$ is the fourth derivative of $y(x)$ evaluated at $v \in [x_1, x_n]$.

Therefore, the first term in (5.3.2)

$$\int_{-\infty}^{\alpha}[y - 1/\theta + f(y - 1/\theta)/F(y - 1/\theta) - \mu]^2 \ \frac{1}{\sqrt{2\pi}}$$

$$e^{-(y-\mu)^2/2}\ dy \tag{5.3.4}$$

could be evaluated at several values of α in one pass. The value of (5.3.4) is very near zero for α not greater than -10. In Figure 4.2 it can be seen that $\hat{\mu}_{ML}$, the unrestricted maximum likelihood estimator, would be less than $\hat{\mu}_E$ for y less than zero. For y less than 10 and μ greater than zero, it is easily seen that $(\hat{\mu}_E - \mu)^2 < (\hat{\mu}_{ML} - \mu)^2$. Thus,

$$\int_{-\infty}^{10} [y - 1/\theta + f(y - 1/\theta)/F(y - 1/\theta) - \mu]^2 \frac{1}{\sqrt{2\pi}} e^{-(y-\mu)^2/2} dy$$

$$< \int_{-\infty}^{10} (y-\mu)^2 \frac{1}{\sqrt{2\pi}} e^{-(y-\mu)^2/2} dy .$$

The last integral can be found to be

$$-(-10 - \mu)f(-10 - \mu) + F(-10 - \mu)$$

following the steps outlined in (4.4.1). Since $\mu \geq 0$

$$f(-10 - \mu) \leq f(-10)$$

and

$$F(-10 - \mu) \leq F(-10) ,$$

and both f(-10) and F(-10) are very near zero; therefore, (5.3.4) can be closely approximated by evaluating it over the interval $[-10, \alpha]$.

In this way regret was computed for values of μ, θ, and α, all greater than zero. Here, the values of $\mu \in [0.1, 8.0]$ were integer multiples of 0.1; joining points, α, were allowed to take on values between 0.25 and 5.0 which were integer multiples of 0.25; values of

θ were 0.25, 0.5, 0.75, 0.875, 1.0, 1.5, 2.0, and 2.5. Table 5.2 gives the maximum regret for each (θ,α) pair considered, where regret is the value of (5.3.2) for each θ, α, and μ considered minus the lower envelope given in Table 5.1 for that value of μ.

As can be seen in Table 5.2, the values θ' and α' which minimize (5.3.1) seem to lie in the intervals

$$0.75 \leq \theta' \leq 1.0$$

and

$$1.25 \leq \alpha' \leq 1.75$$

and the associated regret is at most 0.47991. The maximum likelihood estimator from the same likelihood function has a maximum regret of 0.58386 (when compared to the same lower envelope). The Bayesian estimator from the uniform prior and the same likelihood has a maximum regret of 0.91544. The optimal joined estimator is given by

$$\begin{aligned} \hat{\mu}_J &= y - 1/.875 + f(y - 1/.875)/F(y - 1/.875) && \text{for} \quad y < 1.50 \\ &= y && \text{for} \quad 1.50 \leq y \ . \end{aligned}$$

Note that $\hat{\mu}_J$ is discontinuous at y equal 1.50, which results in an interval of feasible μ-space being unattainable by this estimator. So this estimator which is the first result of an attempt to construct a small mean square error estimator onto the feasible parameter space, fails again to exhaust the feasible parameter space. The following section will discuss remedies for this.

Table 5.2   Maximum regrets for joined estimators

α \ θ	0.250	0.500	0.750	0.875	1.000	1.500	2.000	2.500
				Maximum Regret				
0.25	3.07098	0.53246	0.55373	0.56275	0.93493	1.18827	1.33428	1.42621
0.50	3.49415	0.53429	0.56364	0.57622	1.04432	1.34450	1.51355	1.61907
0.75	3.86061	0.52072	0.55959	0.57641	1.17629	1.51911	1.70844	1.82655
1.00	4.14993	0.49140	0.53712	0.55870	1.32916	1.71163	1.91584	2.04095
1.25	4.35675	0.53673	0.49792	0.52453	1.50075	1.91475	2.13144	2.26498
1.50	4.48888	0.61666	0.48414	0.47991	1.69316	2.12679	2.35314	2.48818
1.75	4.56236	0.73520	0.55142	0.49056	1.90199	2.34798	2.57464	2.70802
2.00	4.59578	0.89043	0.64532	0.45494	2.12449	2.57122	2.79291	2.92155
2.25	4.60572	1.08054	0.76166	0.65849	2.35616	2.79261	3.00438	3.12555
2.50	4.60440	1.30178	0.89494	0.76550	2.59129	3.00735	3.20498	3.31649
2.75	4.59934	1.54858	1.03829	0.87963	2.82360	3.21102	3.39130	3.49162
3.00	4.59428	1.81380	1.18555	0.99545	3.04708	3.39976	3.56061	3.64891
3.25	4.59054	2.08932	1.33086	1.10815	3.25652	3.57070	3.71118	3.78727
3.50	4.58822	2.36667	1.46924	1.21384	3.44794	3.72205	3.84223	3.90648
3.75	4.58694	2.63773	1.59648	1.30973	3.61877	3.85313	3.95389	4.00706
4.00	4.58630	2.89539	1.71115	1.39417	3.76780	3.96414	4.04697	4.09013
4.25	4.58602	3.13399	1.81087	1.46650	3.89493	4.05603	4.12281	4.15717
4.50	4.58590	3.34964	1.89621	1.52683	4.00099	4.13027	4.18309	4.20993
4.75	4.58585	3.54016	1.96708	1.57582	4.08747	4.18864	4.22963	4.25020
5.00	4.58583	3.70499	2.02414	1.61448	4.15626	4.23358	4.26433	4.27981

5.4 Other Joined Estimators

As a combination of two estimators, one Bayesian estimator and the maximum likelihood estimator, an estimator was created in Section 5.3 which had smaller maximum regret than any of the classical estimators previously considered in this paper. This suggests that maximum regret could be decreased further by combining several Bayesian estimators with the maximum likelihood estimator. To explore these possibilities, the case of a sample of size one from a univariate normal density will again be examined. This section will consider the case where the mean of this density is known to be non-negative (*i.e.*, d=0) and the variance is one.

Instead of attempting to find one optimum interval as was done in Section 5.3, the domain of the observation will now be divided into several fixed intervals. (The intervals considered in this case are given in Table 5.3.) A search was carried out to find the optimal Bayesian estimator (exponential prior) for each interval in Table 5.3; the maximum likelihood estimator will be used for $[5.00, \infty)$.

The mean square error for such a joined estimator where q+1 intervals are considered is

$$\mathrm{MSE}(\hat{\mu}_c) = \int_{-\infty}^{a_1} (\hat{\mu}_1 - \mu)^2 p(y)\,dy + \sum_{i=2}^{q} \int_{a_{i-1}}^{a_i} (\hat{\mu}_i - \mu)^2 p(y)\,dy$$

$$\int_{a_q}^{\infty} (\hat{\mu}_{ML} - \mu)^2 p(y)\,dy \ , \qquad (5.4.1)$$

where $a_1 = 0$, and $p(y)$ is the normal density function with mean μ and variance one. The estimator $\hat{\mu}_{ML}$ is the maximum likelihood estimator and is a function of y. The estimators $\hat{\mu}_i$ are the Bayesian

Table 5.3 Values of the parameters μ and θ and the intervals used in the stepwise optimizing process[a]

Values of θ used	Values of μ used	Intervals considered	Optimal value of θ on each interval
0.125	0.2	$(-\infty, 0.00)$	0.125
0.250	0.4	$[0.00, 0.25)$	0.250
0.375	0.6	$[0.25, 0.50)$	0.250
0.500	0.8	$[0.50, 0.75)$	0.375
0.625	1.0	$[0.75, 1.00)$	0.875
0.750	1.2	$[1.00, 1.25)$	1.250
0.875	1.4	$[1.25, 1.50)$	1.250
1.000	1.6	$[1.50, 1.75)$	1.250
1.250	1.8	$[1.75, 2.00)$	1.750
1.500	2.0	$[2.00, 2.25)$	1.750
1.750	2.5	$[2.25, 2.50)$	2.000
2.000	3.0	$[2.50, 2.75)$	2.500
2.500	3.5	$[2.75, 3.00)$	2.500
3.000	4.0	$[3.00, 3.25)$	2.500
3.500	4.5	$[3.25, 3.50)$	2.500
4.000	5.0	$[3.50, 3.75)$	4.000
4.500	5.5	$[3.75, 4.00)$	3.000
10.000	6.0	$[4.00, 4.25)$	4.000
11.000	6.5	$[4.25, 4.50)$	3.000
12.000	7.0	$[4.50, 4.75)$	3.500
13.000		$[4.75, 5.00)$	2.500
14.000			

[a]The meaning of the first two columns is explained in the text

estimators found by assigning different exponential priors, characterized by their parameter θ_i, on each μ-interval and therefore they are functions of y and θ_i. The problem is then to choose the parameters

$$\theta_1 \in [0,\infty),\ \theta_2 \in [0,\infty),\ \ldots,\ \theta_q \in [0,\infty)$$

so as to minimize

$$\max_{\mu \geq 0} \left\{ \int_{-\infty}^{0=a_1} [\hat{\mu}_1(y,\theta_1) - \mu]^2 p(y)dy + \sum_{i=2}^{q} \int_{a_{i-1}}^{a_i} [\hat{\mu}_i(y,\theta_i) - \mu]^2 p(y)dy \right.$$

$$\left. + \int_{a_q}^{\infty} [\hat{\mu}_{ML}(y) - \mu]^2 p(y)dy - LE(\mu) \right\}, \qquad (5.4.2)$$

where $LE(\mu)$ is the above-mentioned lower envelope.

This problem of finding an optimum estimator in each interval evokes memories of dynamic programming. (See Bellman and Dreyfus, (1962).) The intervals correspond to the stages in the dynamic programming problem, choosing the θ_i on each interval corresponds to the activities, and maximizing regret corresponds to the objective function which is to be minimized. However, the problem of finding the θ_i so as to minimize maximum regret cannot be restated in terms of a recursive relation since the choice of the θ_i in any one interval affects the maximum regret function as a whole. This property violates the underlying assumptions of dynamic programming.

Thus, to determine the choice of $(\theta_1, \theta_2, \ldots, \theta_q)$ which would truly be optimal would require the evaluation of (5.4.2) for all points in a q-dimensional space. Note that the evaluation of (5.4.2) is quite costly, even at one point $(\theta_1, \theta_2, \ldots, \theta_q)$.2/ To reduce computer costs, only relatively few alternative values for each coordinate θ_i were examined; these are given in the first column of Table 5.3. Also, in determining the maximum in (5.4.2) only a few μ-values were used; see the second column of Table 5.3.

Even so, the cost of the computational work is prohibitive. Therefore, an approximation was used which is similar to the stepwise inclusion method of regression analysis. A joined estimator, $\hat{\mu}_c$, was constructed sequentially as follows. First define $\hat{\mu}_{c0}$ as

$$\begin{aligned} \hat{\mu}_{c0}(y) &= \hat{\mu}_{\theta}(y) \quad \text{with} \quad \theta = 10 \quad \text{for} \quad y < 5 \ , \\ &= \hat{\mu}_{ML}(y) \qquad\qquad\qquad\quad \text{for} \quad y \geq 5 \ , \end{aligned}$$

where $\hat{\mu}_{\theta}(y)$ is the Bayesian estimator corresponding to the exponential prior on $[0, \infty)$ with θ equal 10, and $\hat{\mu}_{ML}(y)$ is the maximum likelihood estimator. Then define $\hat{\mu}_{c1}$ as

$$\begin{aligned} \hat{\mu}_{c1}(y) &= \hat{\mu}_{c0}(y) \quad \text{for} \quad y \geq 0 \ , \\ &= \hat{\mu}_{\theta}(y) \quad \text{for} \quad y < 0 \ , \end{aligned}$$

where θ is chosen from among the candidate values listed in Table 5.3

2/The first integral in (5.4.2) is evaluated as $\int_{-10}^{a_i}$, for reasons given with respect to (5.3.4). All but the last two terms are evaluated using the subprogram DQHFE in the System 360/Scientific Subroutine Package (1970). The next to last term is identical to (5.2.3) and is evaluated using the normal density and distribution functions given there.

in such a way that $\hat{\mu}_{c1}(y)$ will have the smallest possible maximum regret. Then define $\hat{\mu}_{c2}$ as

$$\begin{aligned} \hat{\mu}_{c2}(y) &= \hat{\mu}_{c1}(y) \qquad \text{for} \qquad y \in R^1/[0.0,\ 0.25)\ , \\ &= \hat{\mu}_{\theta}(y) \qquad \text{for} \qquad y \in [0.0,\ 0.25)\ , \end{aligned}$$

where θ is chosen from among the same candidate values so that $\hat{\mu}_{c2}(y)$ will have the smallest possible maximum regret. This process is continued so that $\hat{\mu}_{c3}(y)$ will be equal to $\hat{\mu}_{c2}(y)$ except on the y-interval $[0.25,\ 0.50)$, where

$$\hat{\mu}_{c3}(y) = \hat{\mu}_{\theta}(y)$$

and θ is again chosen so as to minimize maximum regret of $\hat{\mu}_{c3}(y)$, and so on. Eventually the entire y-interval $[0.0,\ 5.0)$ will be divided into intervals of length 0.25 and on each interval sequentially, the parameter θ will be chosen so as to minimize maximum regret of $\hat{\mu}_{ci}$ at the i^{th} step. The maximum regret of $\hat{\mu}_{c21}(y)$ is equal to 0.304207, which is indeed substantially less than the value found for the joined estimators presented in Section 5.3. The optimal values of θ chosen at each stage are shown in Table 5.3. It should be noted that this estimator is discontinuous at many of the 21 join points.

It would be desirable to refine the y-intervals and values of θ attempted in the preceding process, but this is too costly an operation. If the process could be continued a continuous function, $\theta(y)$, for θ in terms of the observations could be found. This function could then be substituted for θ in the expression for the Bayesian estimator and would yield an estimator which would give a maximum regret smaller than any of the other estimators used for this particular likelihood.

Using the values of θ listed in Table 5.3 as a basis some functions $\theta(y)$ were constructed. The maximum regret for the estimators found by substituting these functions of the observations for θ were then found by approximating the mean square errors by using the subprogram DQH64 in the System 360/Scientific Subroutine Package (1970), i.e., Gaussian-Hermite quadrature.

As for the construction of these functions $\theta(y)$, first consider Figure 5.1, which depicts the θ-values of Table 5.3. The variability of these θ-values after the twelfth interval could be ignored in searching for functions $\theta(y)$. When the observation y is large and θ is large, the Bayesian estimate tends to the maximum likelihood estimate as was shown in Section 4.2. Therefore, the variability of the θ's is most likely due to the Bayesian estimate for these values of θ or any larger θ not being significantly different from the maximum likelihood estimate for observations that are large.

Notice in Figure 5.1 that when the observation is larger than 0.5, when one ignores the variability of θ for observations larger than 2.75, a linear function of the observations (y) seems to fit the values of θ as they depend on y. The maximum regret was computed for several linear functions $\theta(y)$. Table 5.4 gives the functions that were considered, the interval on which the linear function was used, and the value used for θ on the remainder of the domain of the observation, in columns 1, 2, and 3 respectively. Note that these

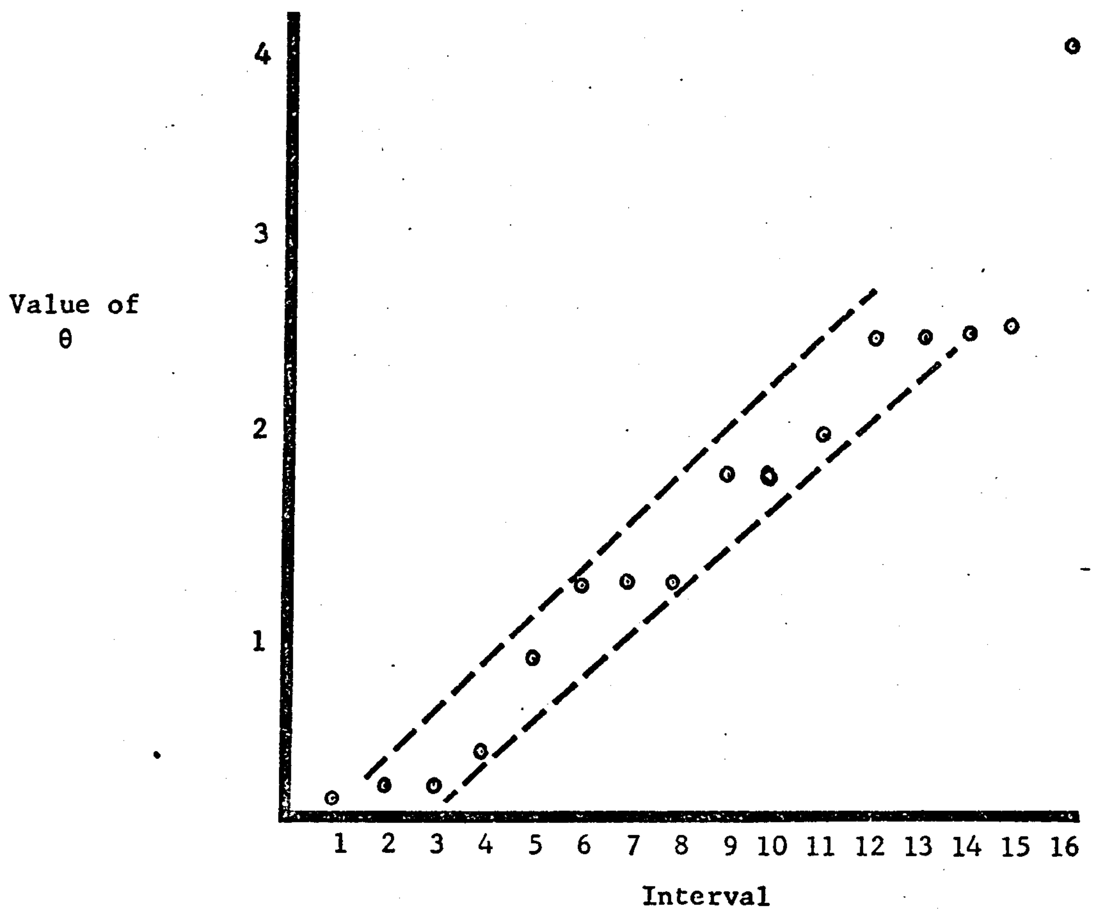


Figure 5.1 A plot of the optimal value of θ found by the stepwise optimizing procedure

Table 5.4 Linear functions of the observations used in place of θ in the exponential prior

Linear function	S, the set on which the linear function was used	Value of θ used on S^c	Maximum regret
$\theta = -0.05 + y$	$y \in [0.1, \infty]$	0.1	0.361671
$\theta = -0.25 + 0.75y$	$y \in [0.2, \infty]$	0.1	0.387803
$\theta = -0.10 + 1.50y$	$y \in [0.1, \infty]$	0.1	0.441211
$\theta = -0.70 + 1.50y$	$y \in [0.5, \infty]$	0.1	0.550220
$\theta = -1.13 + 1.14y$	$y \in [0.75, \infty]$	0.125	0.552206
$\theta = -1.38 + 1.63y$	$y \in [1.0, \infty]$	0.25	0.569923

functions gave some improvement in the maximum regret over the joining of two estimators. Figure 5.2 gives a plotting of the mean square error for the estimator found by using the linear function

$$\theta = -0.025 + 0.75y$$

when y is greater than or equal to 0.2, and letting $\theta = .1$ otherwise. This would yield the estimator

$$\begin{aligned} \hat{\mu} &= y - 1/(-0.025 + 0.75y) \\ &\quad + f[y - 1/(-0.025 + 0.75y)]/F[y - 1/(-0.025 + 0.75y)] \qquad \text{for } y \geq 0.2 \ , \\ &= y - 10 + f(y - 10)/F(y - 10) \qquad \text{for } y < 0.2 \ , \end{aligned}$$

5.5 Extending the Technique

All of the work so far reported in this chapter was concerned with a sample of size one from a univariate normal density with variance one and mean, μ, not less than zero. Analogous results hold for any sample of size $n \geq 1$ from any univariate normal density with a mean known to lie in the interval $[d, \infty)$. Given a sample size of n and mean $\bar{y}$ from a normal distribution with unknown expectation $\mu \geq d$ and known variance σ^2, the likelihood function for the mean, μ, is

$$L(\mu|\bar{y}) \propto e^{-n(\bar{y}-\mu)^2/2\sigma^2} \ .$$

By substituting $\bar{y}$ for y and σ^2/n for σ^2 in the following results, they are applicable to samples of any size n.

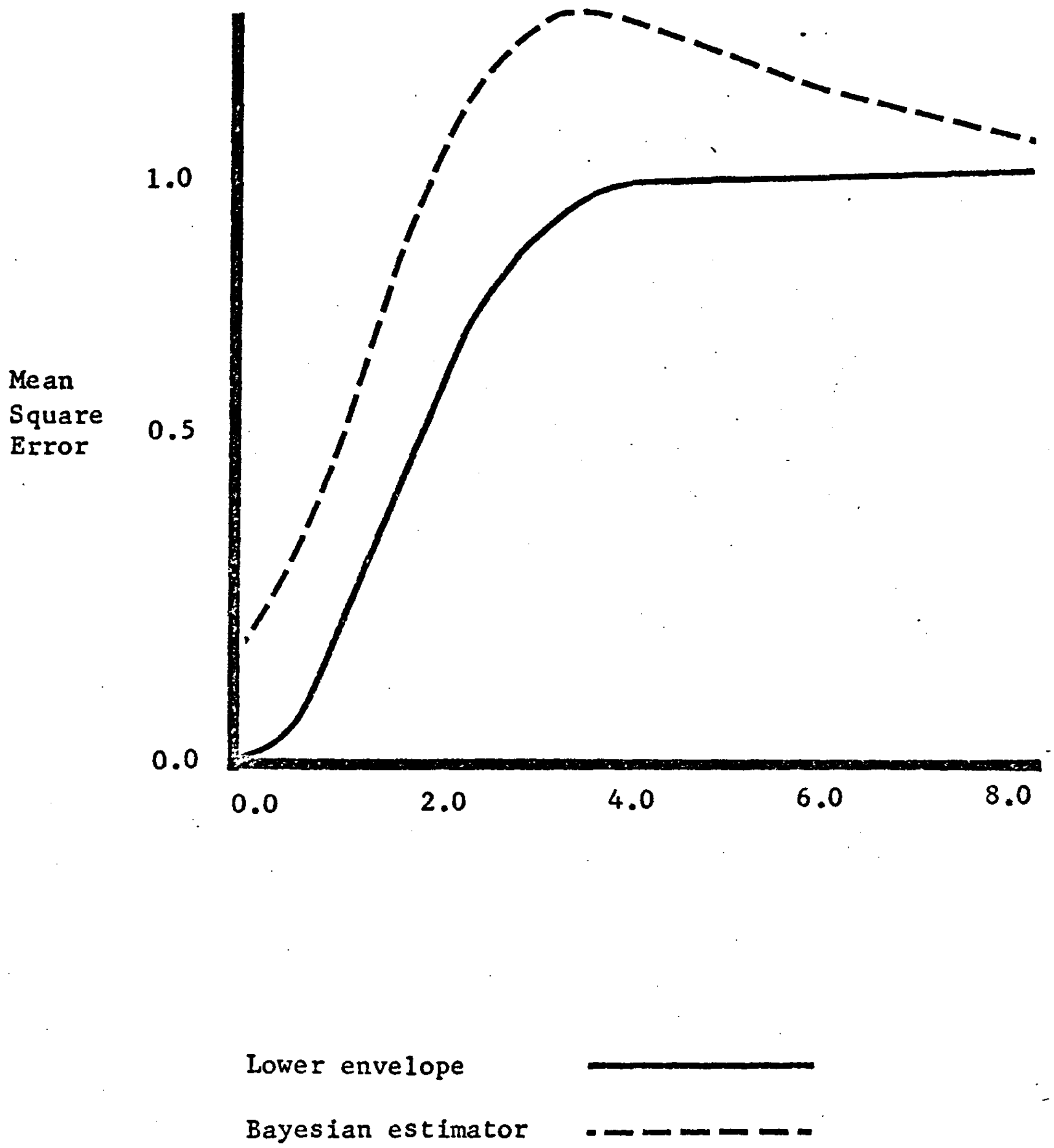


Figure 5.2 Lower envelope and mean square error for a Bayesian estimator using a continuous function of the observations for the parameter

The posterior using the exponential prior is

$$p(\mu|y,\theta,\sigma) = \frac{e^{-[(y-\sigma^2/\theta) - \mu]^2/2\sigma^2}}{\int_d^\infty e^{-[(y-\sigma^2/\theta) - \mu]^2/2\sigma^2} d\mu} , \quad \mu \geq d$$

$$= 0 \qquad \text{elsewhere} .$$

Using (4.1.2) the mean of this posterior is found to be

$$\hat{\mu} = y - \sigma^2/\theta + \frac{\sigma f((y - \sigma^2/\theta - d)/\sigma)}{F((y - \sigma^2/\theta - d)/\sigma)} . \tag{5.5.1}$$

The mean square error of this estimator is a function of μ, σ, θ, defined for $\theta > 0$, $\sigma > 0$, $\mu \geq d$:

$$\chi_d(\mu,\sigma,\theta) = \int_{R'}[y - \mu - \frac{\sigma^2}{\theta} + \sigma \frac{f(\frac{y-d}{\sigma} - \frac{\sigma}{\theta})}{F(\frac{y-d}{\sigma} - \frac{\sigma}{\theta})}]^2 e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}} \frac{dy}{\sigma\sqrt{2\pi}}$$

$$= \sigma^2 \int_{R'}[\mu - \frac{\sigma}{\theta} + \frac{f(\frac{\mu-d}{\sigma} + u - \frac{\sigma}{\theta})}{F(\frac{\mu-d}{\sigma} + u - \frac{\sigma}{\theta})}]^2 e^{-\frac{1}{2}u^2} \frac{du}{\sqrt{2\pi}} ,$$

where the transformation $y = \mu + \sigma u$ was used. Note that replacing d by 0, and σ by 1 yields a function $\chi^*(\mu,\theta)$, defined for $\theta > 0$, $\mu \geq 0$:

$$\chi_0^*(\mu,\theta) = \int_{R'}[u - \frac{1}{\theta} + \frac{f(\mu + u - \frac{1}{\theta})}{F(\mu + u - \frac{1}{\theta})}]^2 e^{-\frac{1}{2}\mu^2} \frac{du}{\sqrt{2\pi}}$$

and that

$$\chi(\mu,\sigma,\theta) = \sigma^2\chi^* (\frac{\mu-d}{\sigma} , \frac{\theta}{\sigma}) , \tag{5.5.2}$$

both functions defined on their appropriate domain. Thus the mean square error for the estimator given in (5.5.1) is expressed in terms of the mean square error for the Bayesian estimator found when the variance is one, the feasible parameter space is the positive half line, and the prior is an exponential density with parameter θ/σ, i.e., the situation discussed in Section 5.3. Therefore, the lower envelope for the mean square error for estimators of the form (5.5.1) and variance equal to σ^2 can be found in Table 5.1. The value of the minimum mean square error at the point μ' in the feasible parameter space equals σ^2 times the value of the minimum mean square error at the point $(\mu' - d)/\sigma$ in the feasible parameter space $[0, \infty)$, as found in Table 5.1. Likewise the values of the regret function in the general case can be found from the values of the regret function in the special case of the previous sections.

In Section 5.4 we considered estimators for $\mu \geq 0$ from samples of size 1 from $N(\mu,1)$, where these estimators were obtained from Bayesian estimators with exponential prior, $-\theta e^{-\mu/\theta}$, by replacing θ by a function $\theta(y)$. The mean square error for such an estimator is defined for $\mu \geq 0$ and equals

$$\int_{R'} [y - \mu - \frac{1}{\theta(y)} + \frac{f(y - \frac{1}{\theta(y)})}{F(y - \frac{1}{\theta(y)})}]^2 e^{-\frac{1}{2}(y-\mu)^2} \frac{dy}{\sqrt{2\pi}} .$$

In the more general case we consider estimators for $\mu' \geq d$ from samples of size one (y', say) from $N(\mu',\sigma^2)$, where these estimators are obtained from Bayesian estimators with exponential prior, $-\theta e^{-(\mu'-d)/\theta}$, by replacing θ by a function $\theta'(y')$. The mean squares error for such an estimator is defined for $\mu' \geq d$ and equals

$$\sigma^2 \int_{R'} \left[\frac{y'-\mu'}{\sigma} - \frac{\sigma}{\theta'(y')} + \frac{f\left(\frac{y'-d}{\sigma} - \frac{\sigma}{\theta'(y')}\right)}{F\left(\frac{y'-d}{\sigma} - \frac{\sigma}{\theta'(y')}\right)}\right]^2 e^{-\frac{1}{2}\frac{(y'-\mu')^2}{\sigma^2}} \frac{dy'}{\sigma\sqrt{2\pi}} .$$

This integral is easily seen to reduce to the above one by means of the substitutions

$$\frac{y'-d}{\sigma} = y$$

$$\frac{\theta'(y')}{\sigma} = \theta(y)$$

$$\frac{\mu'-d}{\sigma} = \mu .$$

Note that these substitutions are compatible with the reduction in equation (5.5.2). Therefore, it is possible to find the value at μ' of the regret function of any estimator for $\mu' \geq d$ of the form

$$y' - \frac{\sigma^2}{\theta'(y')} + \sigma \frac{f\left(\frac{y'-d}{\sigma} - \frac{\sigma}{\theta'(y')}\right)}{F\left(\frac{y'-d}{\sigma} - \frac{\sigma}{\theta'(y')}\right)} ,$$

where $y' \sim N(\mu',\sigma^2)$, from the value at μ of the regret function of the estimator for $\mu \geq 0$ of the form

$$y - \frac{1}{\theta(y)} + \frac{f\left(y - \frac{1}{\theta(y)}\right)}{F\left(y - \frac{1}{\theta(y)}\right)} ,$$

where $y \sim N(\mu,1)$, simply by taking σ^2 times the value of the latter regret function at $\mu = \frac{\mu'-d}{\sigma}$ and computed with $\theta(y) = \frac{\theta'(d+\sigma y)}{\sigma}$.

Similarly, when one has found the function $\theta(y)$ which minimizes maximum regret for the problem of estimating $\mu \geq 0$ from one observation of $y \sim N(\mu,1)$, one can immediately conclude that the function $\theta'(y')$ which minimizes maximum regret for the problem of estimating $\mu' \geq d$ from one observation of $y' \sim N(\mu',\sigma^2)$ is given by $\theta'(y') = \sigma \cdot \theta(y) = \sigma \cdot \theta(\frac{y'-d}{\sigma})$.

The mean square error for the joined estimator considered in Section 5.3 was given by (5.3.2). Making the transformation $u = y-\mu$, this expression becomes

$$\int_{-\infty}^{\alpha-\mu} (u - 1/\theta + \frac{f(u+\mu-1/\theta)}{F(u+\mu-1/\theta)} - \frac{1}{\sqrt{2\pi}}$$

$$e^{-u^2/2}\, du + \int_{\alpha-\mu}^{\infty} u^2 \; \frac{e^{-u^2/2}}{\sqrt{2\pi}}\, du \; . \qquad (5.5.3)$$

For the case of the normal density with $\mu' \in [d, \infty)$ and variance σ^2, the mean square error for the joined estimator would be

$$\int_{-\infty}^{\alpha'} [y - \sigma^2/\theta + \frac{f((y-\sigma^2/\theta'-d)/\sigma)}{F((y-\sigma^2/\theta'-d)/\sigma)} - \mu']^2$$

$$\frac{1}{\sqrt{2\pi}} \; e^{-(y-\mu')^2/2\sigma^2} \; dy$$

$$+ \int_{\alpha'}^{\infty} (y-\mu')^2 \; \frac{e^{-(y-\mu')^2/2\sigma^2}}{\sqrt{2\pi}} \; dy \; .$$

Making the transformation $u = \frac{y-\mu'}{\sigma}$, this expression becomes

$$\int_{-\infty}^{(\alpha'-\mu')/\sigma} [u - \sigma^2/\theta' + \sigma \frac{f(u - \sigma/\theta' + \frac{\mu'-d}{\sigma})}{F(u - \sigma/\theta' + \frac{\mu'-d}{\sigma})}]^2 \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

$$+ \int_{(\alpha'-\mu')/\sigma}^{\infty} u^2 \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du . \qquad (5.5.4)$$

Then (5.5.4) would equal σ^2 times (5.5.3) when in (5.5.3) the substitutions $(\mu'- d)/\sigma$ for μ, θ'/σ for θ and $(\alpha'- d)/\sigma$ for α are made.

The optimal values of α and θ found for the case $\mu \in [0, \infty)$ and variance one (i.e., the case given by (5.5.3)) were

$$0.75 \leq \theta \leq 1$$

and

$$1.25 \leq \alpha \leq 1.75 .$$

So the optimal θ' for the general case would be

$$0.75 \leq \theta'/\sigma \leq 1$$

or

$$0.75\ \sigma \leq \theta' \leq \sigma .$$

The optimal choice for α would be

$$1.25 \leq (\alpha'- d)/\sigma \leq 1.75$$

or

$$1.25\sigma + d \leq \alpha' \leq 1.75\sigma + d .$$

5.6 Estimating Two Ordered Parameters

The methods in this chapter can be applied in a similar manner to the two parameter model in which the two parameters are known to be ordered. Without loss of generality, say $\mu_1 \leq \mu_2$. The Bayesian estimators for the two parameters when a uniform prior is assumed are given in (8.3.19). For the case of a sample of size one from a bivariate normal distribution with covariance matrix $\sigma^2 I$, and for the exponential prior defined in (4.2.2), the posterior would be as in (4.2.3), which is of the form of 8.3.3), and so the Bayesian estimators would be

$$\hat{\underline{\mu}} = \underline{y} + \begin{bmatrix} 2\sigma^2/\theta \\ \\ -2\sigma^2/\theta \end{bmatrix} + \begin{bmatrix} -1 \\ \\ 1 \end{bmatrix} \frac{\sigma\lambda}{\sqrt{2}} \qquad (5.6.1)$$

where

$$\lambda = \frac{f((y_2 - y_1 - 2\sigma^2/\theta)/\sigma\sqrt{2})}{F((y_2 - y_1 - 2\sigma^2/\theta)/\sigma\sqrt{2})} .$$

As a measure analogous to mean square error for this vector-valued estimator, the following scalar expression will be used

$$\mathrm{MSE}(\hat{\underline{\mu}}) = E(\hat{\underline{\mu}} - \underline{\mu})'(\hat{\underline{\mu}} - \underline{\mu})$$

which in this problem equals

$$\mathrm{MSE}(\hat{\underline{\mu}}_E) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [(y_1 + 2\sigma^2/\theta - \lambda\sigma/\sqrt{2} - \mu_1)^2$$

$$+ (y_2 - 2\sigma^2/\theta + \lambda\sigma/\sqrt{2} - \mu_2)^2] \frac{1}{2\pi\sigma^2}$$

$$\exp(-(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})/2\sigma^2)\, dy_1 dy_2$$

$$= E(y_1 - \mu_1)^2 + E(y_2 - \mu_2)^2 + 8\sigma^4/\theta^2$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[2\sigma(y_2 - y_1 - \mu_2 + \mu_1)\lambda/\sqrt{2} - 8\sigma^3\lambda/\theta\sqrt{2} + \sigma^2\lambda^2$$

$$+ \frac{4\sigma^2}{\theta}(y_1 - y_2 - \mu_1 + \mu_2)]/(2\pi\sigma^2)$$

$$\exp[-(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})/2\sigma^2]\, dy_1 dy_2$$

The first two terms are the variances of y_1 and y_2, respectively. By making the transformation

$$u_1 = y_2 - y_1 \qquad u_2 = y_1 + y_2$$

$$y_1 = \frac{u_2 - u_1}{2} \qquad y_2 = \frac{u_1 + u_2}{2}$$

$$dy_1 dy_2 = -\tfrac{1}{2}\, du_1 du_2 ,$$

more tractable expressions can be found for the remaining terms. Thus,

$$\mathrm{MSE}(\hat{\underline{\mu}}_\theta) = 2\sigma^2 + \int_{-\infty}^{\infty}[-4\sigma^2 u_1/\theta + 2\sigma u_1\lambda/\sqrt{2} + 8\sigma^4/\theta^2$$

$$- 8\sigma^3\lambda/(\theta\sqrt{2}) + 4\sigma^2(\mu_2 - \mu_1)/\theta + \sigma^2\lambda^2$$

$$- 2\sigma\lambda(\mu_2 - \mu_1)/\sqrt{2}][\exp(-(u_1 - (\mu_2 - \mu_1))^2/4\sigma^2)/\sqrt{4\pi}\,\sigma]d\mu_1$$

$$\int_{-\infty}^{\infty}[\exp(-(u_2 - (\mu_2 + \mu_1))^2/4\sigma^2)/\sqrt{4\pi}\,\sigma]du_2 . \qquad (5.6.2)$$

In this case

$$\lambda = \frac{f((u_1 - 2\sigma^2/\theta)/(\sigma\sqrt{2}))}{F((u_1 - 2\sigma^2/\theta)/(\sigma\sqrt{2}))} \quad .$$

After integrating with respect to u_2 expression (5.6.2) is seen to involve only one variable, $u_1 = y_2 - y_1$. Note that the difference between μ_2 and μ_1 completely specifies the mean square error.

Expression (5.6.2) can be evaluated using the same numerical methods as were used for the one parameter case considered previously in this chapter. The lower envelope for values of $\mu_2 - \mu_1$ and $\sigma^2 = 1$ was approximated in exactly the same manner as the univariate case.(See Table 5.5, where $\Delta\mu = \mu_2 - \mu_1$.)

Section 2.3 discussed the restricted maximum likelihood estimator for (μ_1, μ_2) under the given conditions. The mean square error for this estimator is given by

$$\begin{aligned}
MSE(\hat{\underline{\mu}}_{ML}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{y_2} [(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2] \\
&\quad \frac{1}{2\pi}\exp(-(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})/2)dy_1 dy_2 \\
&\quad + \int_{-\infty}^{\infty}\int_{y_2}^{\infty} [((y_1 + y_2)/2 - \mu_1)^2 + ((y_1 + y_2)/2 - \mu_2)^2] \\
&\quad \frac{1}{2\pi}\exp(-(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})/2)dy_1 dy_2 \\
&= \tfrac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2] \\
&\quad \frac{1}{2\pi}\exp(-\tfrac{1}{2}(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu}))dy_1 dy_2
\end{aligned}$$

Table 5.5 Approximation for the lower envelope of the mean square error for estimators of ordered parameters

$\Delta\mu$	θ giving the minimum MSE	Minimum MSE	$\Delta\mu$	θ giving the minimum MSE	Minimum MSE
0.1	2.1	1.10346	4.1	34.4	1.87625
0.2	2.1	1.11865	4.2	37.3	1.88725
0.3	2.4	1.12878	4.3	40.3	1.89752
0.4	2.6	1.14301	4.4	44.8	1.90707
0.5	2.7	1.15793	4.5	49.2	1.91592
0.6	2.8	1.17420	4.6	55.9	1.92411
0.7	3.0	1.19115	4.7	62.5	1.93166
0.8	3.2	1.20917	4.8	69.2	1.93861
0.9	3.4	1.22811	4.9	75.9	1.94497
1.0	3.6	1.24787	5.0	82.6	1.95080
1.1	3.8	1.26842	5.1	92.6	1.95611
1.2	4.0	1.28972	5.2	102.6	1.96094
1.3	4.3	1.31166	5.3	117.6	1.96533
1.4	4.6	1.33421	5.4	132.6	1.96930
1.5	4.9	1.36723	5.5	147.6	1.97288
1.6	5.2	1.38063	5.6	170.2	1.97610
1.7	5.5	1.40435	5.7	192.7	1.97899
1.8	5.8	1.42832	5.8	215.2	1.98158
1.9	6.2	1.45243	5.9	249.0	1.98388
2.0	6.7	1.47659	6.0	282.8	1.98594
2.1	7.1	1.50071	6.1	316.6	1.98775
2.2	7.6	1.52469	6.2	367.3	1.98937
2.3	8.0	1.54848	6.3	418.0	1.99080
2.4	8.7	1.57196	6.4	468.6	1.99205
2.5	9.4	1.59509	6.5	544.7	1.99315
2.6	10.0	1.61778	6.6	620.7	1.99411
2.7	10.7	1.63997	6.7	734.7	1.99495
2.8	11.4	1.66161	6.8	848.7	1.99568
2.9	12.4	1.68261	6.9	1019.8	1.99632
3.0	13.4	1.70297	7.0	1190.9	1.99687
3.1	14.4	1.72263	7.1	1361.9	1.99734
3.2	15.9	1.74156	7.2	1618.5	1.99775
3.3	16.7	1.75973	7.3	1875.1	1.99810
3.4	18.6	1.77710	7.4	2259.9	1.99840
3.5	19.5	1.79369	7.5	2644.8	1.99865
3.6	21.9	1.80946	7.6	3029.7	1.99887
3.7	24.3	1.82443	7.7	3607.0	1.99906
3.8	25.5	1.83858	7.8	4472.9	1.99921
3.9	28.4	1.85193	7.9	5338.9	1.99935
4.0	31.4	1.86448	8.0	6204.8	1.99946

$$+ \tfrac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{y_2} [(y_1 - \mu_1)^2 + (y_2 - \mu_2)^2]$$

$$\frac{1}{2\pi} \exp(- \tfrac{1}{2}(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})) dy_1 dy_2$$

$$+ \int_{-\infty}^{\infty} \int_{y_2}^{\infty} [y_1 y_2 - y_2 \mu_1 - y_1 \mu_2 + (\mu_1^{\,2} + \mu_2^{\,2})/2]$$

$$\frac{1}{2\pi} \exp(- \tfrac{1}{2}(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})) d\underline{y}_1 dy_2 \quad . \tag{5.6.3}$$

The first integral is $\frac{1}{2}$ the sum of the variances of y_1 and y_2. By making the transformation $\underline{u} = \underline{y} - \underline{\mu}$, and then

$$\underline{v} = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \; \underline{u}$$

expression (5.6.3) becomes

$$\mathrm{MSE}(\hat{\underline{\mu}}_{ML}) = 1 + \tfrac{1}{2} \int_{-\infty}^{\infty} \int_{(\mu_1 - \mu_2)/\sqrt{2}}^{\infty} (v_1^2 + v_2^2) \frac{1}{2\pi} e^{-\underline{v}'\underline{v}/2} dv_1 dv_2$$

$$+ \int_{-\infty}^{\infty} \int_{-\infty}^{(\mu_1 - \mu_2)/\sqrt{2}} [\tfrac{1}{2}(v_2^2 - v_1^2) + (\mu_1 - \mu_2)^2/2]$$

$$\frac{1}{2\pi} \quad e^{-\underline{v}'\underline{v}/2} \; dv_1 dv_2$$

$$= 2 - F((\mu_1 - \mu_2)/2) + ((\mu_1 - \mu_2)/\sqrt{2}) f((\mu_1 - \mu_2)/\sqrt{2})$$

$$((\mu_1 - \mu_2)^2/2) F((\mu_1 - \mu_2)/\sqrt{2}) \quad . \tag{5.6.4}$$

Here the functions $f(x)$ and $F(x)$ are the density and distribution functions for the univariate normal distribution with mean zero and variance one. Thus, the mean square error for the restricted maximum likelihood estimator is a function of $(\mu_2 - \mu_1)$ also.

The maximum regret for the maximum likelihood estimator calculated for values of $\Delta \mu = \mu_2 - \mu_1$ given in Table 5.5 was found to be 0.420623. This compares favorably to the maximum regret of 0.835614 found for the Bayesian estimator using a uniform prior.

To determine if these maximum regrets could be decreased even more, the process of joining Bayesian estimators (for exponential priors) with maximum likelihood estimators was examined for this case.

Following the procedure outlined in Section 5.4, an attempt was made to find the optimal value of θ on each of several intervals. This was done for the case of $\sigma^2 = 1$ and the ordering $\mu_2 \geq \mu_1$. (The values of θ and $\Delta \mu$ that were used are given in Table 5.6.)

The mean square error for these combined estimators can be written as in (5.4.1) for the exhaustive sets I_i, $i = 1, 2, \ldots, q+1$,

$$\mathrm{MSE}(\hat{\underline{\mu}}_C) = \sum_{i=1}^{q} \int \int_{I_i} (\hat{\underline{\mu}}_i(\underline{y},\theta) - \underline{\mu})'(\hat{\underline{\mu}}_i(\underline{y},\theta) - \underline{\mu})\, p(\underline{y})dy_1dy_2$$

$$+ \iint_{I_{q+1}} (\hat{\underline{\mu}}_{ML} - \underline{\mu})'(\hat{\underline{\mu}}_{ML} - \underline{\mu})\, p(\underline{y})dy_1dy_2 \,. \qquad (5.6.5)$$

Here the functions $\underline{\mu}_i(\underline{y},\theta)$ are the Bayesian estimators from exponential priors, and $\hat{\underline{\mu}}_{ML}$ is the restricted maximum likelihood estimator. Since the interval I_{q+1} will be required to be in the feasible parameter space, $\hat{\underline{\mu}}_{ML}$ is the unrestricted maximum likelihood estimator, *i.e.*, $\hat{\underline{\mu}}_{ML} = \underline{y}$. The term $(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})$ is found to occur in every one of the q+1 integrands in (5.6.5). (See the derivation of (5.6.2.)) Thus, (5.6.5) can be expressed as

Table 5.6 Values of the parameters θ and μ and the intervals used in the stepwise optimizing process for two ordered parameters[a]

Values of θ used	Values of $\Delta\mu$ used	Intervals considered $(y_2 - y_1)$	Optimal value of θ on each interval
2.0	0.2	$[\infty$, 0.00)	4.0
2.5	0.4	[0.00, 0.25)	2.5
3.0	0.6	[0.25, 0.50)	2.5
4.0	0.8	[0.50, 0.75)	2.5
5.0	1.0	[0.75, 1.00)	2.5
10.0	1.2	[1.00, 1.25)	10.0
15.0	1.4	[1.25, 1.50)	10.0
20.0	1.6	[1.50, 1.75)	5.0
25.0	1.8	[1.75, 2.00)	10.0
30.0	2.0	[2.00, 2.25)	10.0
40.0	2.5	[2.25, 2.50)	5.0
50.0	3.0	[2.50, 2.75)	5.0
75.0	3.5	[2.75, 3.00)	10.0
100.0	4.0	[3.00, 3.25)	10.0
125.0	4.5	[3.25, 3.50)	10.0
150.0	5.0	[3.50, 3.75)	10.0
200.0	5.5	[3.75, 4.00)	10.0
250.0	6.0	[4.00, 4.25)	10.0
	6.5	[4.25, 4.50)	10.0
	7.0	[4.50, 4.75)	10.0
		[4.75, 5.00)	10.0

[a]See the text for an explanation of columns one and two

$$\text{MSE}(\hat{\mu}_C) = 2\sigma^2 + \sum_{i=1}^{q} \int \int_{I_i} [-4\sigma^2\ u_i/\theta_i + 2\sigma\ u_i\lambda_i/\sqrt{2}$$

$$+ 8\sigma^4/\theta_i^2 - 8\sigma^3\lambda_i/(\theta_i\sqrt{2}) + 4\sigma^2(\mu_2 - \mu_1)/\theta_i$$

$$+ \sigma^2\lambda_i^2 - 2\sigma\ \lambda_i(\mu_2 - \mu_1)/\sqrt{2}]$$

$$\exp[-(u_1 - (\mu_2 - \mu_1))^2/(4\sigma^2)]/\sqrt{4\pi}\ \sigma\ du_1\ .$$

Here

$$\lambda_i = \frac{f((u_1 - 2\sigma^2/\theta_i)/(\sigma\sqrt{2}))}{F((u_1 - 2\sigma^2/\theta_i)/(\sigma\sqrt{2}))}\ .$$

The variable u_1 is $y_2 - y_1$, so the intervals I_i are found by dividing the (y_1, y_2)-plane into disjoint sets based on the values of $y_2 - y_1$. (The intervals used for this example are also given in Table 5.6.)

The stepwise optimizing procedure described in Section 5.4 was utilized to gain some idea of a proper function to use for θ in minimizing maximum regret. The values of θ for each interval which optimized the minimax regret by this procedure are listed in Table 5.6. Using the Bayesian estimator corresponding to the listed θ on the appropriate interval and the maximum likelihood estimator for an observation in which $y_2 - y_1 \geq 5.0$ yielded an estimator whose maximum regret was 0.316849. (Figure 5.3 gives a plotting of the mean square error for this estimator.)

To determine if the stepwise optimizing algorithm could be improved upon, the maximum regret was found for several other combinations of the θ's for the various intervals. The Bayesian estimator for the exponential prior with one of the values of θ listed was used on the following intervals when

$$y_2 - y_1 \in [-\infty, 0) \quad , \quad \theta = 2.0, 2.5, 3.0, 4.0;$$

$$y_2 - y_1 \in [0.0, 0.25) \quad , \quad \theta = 2.0, 2.5, 3.0;$$

$$y_2 - y_1 \in [0.25, 0.50) \quad , \quad \theta = 2.0, 2.5, 3.0;$$

$$y_2 - y_1 \in [0.50, 0.75) \quad , \quad \theta = 2.5, 3.0, 4.0;$$

$$y_2 - y_1 \in [0.75, 1.00) \quad , \quad \theta = 2.5, 3.0, 4.0, 5.0, 10.0;$$

and when

$$y_2 - y_1 \in [1.00, 5.0) \quad , \quad \theta = 10.0 \quad .$$

For the interval $y \in [5.0, \infty)$ the maximum likelihood estimator was the assigned estimator. The maximum regret was evaluated for the combined estimators found by using all possible combinations of the candidate estimators listed for the various intervals. The same values of θ, as in Table 5.6, were found to be optimal by this procedure. Of the other combinations tested, it was found that replacing the optimal θ in the interval $(y_2 - y_1) \in [0.0, 0.25)$ by $\theta = 3.0$, gave a maximum regret of 0.316861. Thus, the optimal choice of θ must decrease from approximately 4.0 at $y = 0.0$ to a value of θ near 2.5 at $y = 0.50$. In the vicinity of $y = 1.0$, the value of θ must begin increasing to a value which yields an estimator that differs from the maximum likelihood estimator by a negligible amount.

From (5.6.2) it can be seen that the mean square error of $\hat{\underline{\mu}}_\theta$ is a function

$$\Phi(\Delta\mu, \theta, \sigma)$$

of $\Delta\mu$, θ and σ, such that

$$\Phi(\Delta\mu, \theta, \sigma) = \sigma^2 \cdot \Phi(\frac{\Delta\mu}{\sigma}, \frac{\theta}{\sigma}, 1) = \sigma^2 \cdot \Phi^*(\frac{\Delta\mu}{\sigma}, \frac{\theta}{\sigma}) \quad ,$$

say; which allows the computation of mean square errors of estimators in the more general case with arbitrary (but known) σ^2 from the above discussed special case with $\sigma^2 = 1$. Thus, as in Section 5.5, if θ' is the optimal choice for θ when $y_2 - y_1$ is in the interval $[a_1, a_2]$ and when σ^2 is equal to one, then $\sigma\theta'$ would be the optimal choice on the interval $[\sigma a_1, \sigma a_2]$ when σ^2 is not equal to one. Likewise, if $\theta(y_2 - y_1)$ is an optimal continuous function when σ^2 equals one, $\sigma\theta((y_2 - y_1)/\sigma)$ would be the optimal continuous function to use for θ when σ^2 is not equal to one. Therefore, any results found for the case in which the covariance matrix is the identity matrix, could be made to apply to the case where the covariance matrix is $\sigma^2 I$.

5.7 Estimators for m Ordered Parameters

Suppose that instead of estimating the mean vector for a bivariate normal distribution, a modeler needed to estimate the mean vector for a multivariate normal distribution where the components were known to satisfy a certain order. This modeler could construct estimators of the type illustrated in this chapter, but his problem would be considerably more complex than the cases dealt with thus far.

Consider, for example, the case when m is equal to 3 and the covariance matrix is $\sigma^2 I$. Assuming an exponential prior of the form given in (4.2.2), the posterior would be as in (4.2.3), and the Bayesian estimator of $\underline{\mu}$ would follow upon substitution of $\underline{y} - C'\underline{\theta}$ for $\underline{y}$ in (8.3.17). (The estimator for this example is given by expressions (8.3.22) and (8.3.23).

The mean square error for this case is

$$MSE(\hat{\underline{\mu}}_\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\hat{\underline{\mu}}_\theta - \underline{\mu})'(\hat{\underline{\mu}}_\theta - \underline{\mu})(2\pi\,\sigma^2)^{-3/2}$$

$$\exp(-(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})/2\sigma^2)d\underline{y}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu}) + 2(\underline{y} - \underline{\mu})'(C'\underline{\theta} + \sigma^2 H'\underline{1}/P)$$

$$+ (C'\underline{\theta} + \sigma^2 H'\underline{1}/P)'(C'\underline{\theta} + \sigma^2 H'\underline{1}/P)](2\pi\,\sigma^2)^{-3/2}$$

$$\exp(-(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})/(2\sigma^2)dy_1 dy_2 dy_3 \;. \qquad (5.7.1)$$

Note that

$$E(\underline{y} - \underline{\mu})'(\underline{y} - \underline{\mu})$$

is the sum of the variances of y_1, y_2, and y_3 and would equal $3\sigma^2$. Making the transformation

$$u_i = y_i - \mu_i$$

for i = 1, 2, 3, $MSE(\hat{\underline{\mu}}_\theta)$ becomes

$$\mathrm{MSE}(\hat{\underline{\mu}}_\theta) = 3\sigma^2 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[2(\underline{\mu} - \underline{\nu})'(C'\underline{\theta} + \sigma^2 H'\underline{1}^*/P^*)$$

$$+ (C'\underline{\theta} + \sigma^2 H'\underline{1}^*/P^*)'(C'\underline{\theta} + \sigma^2 H'\underline{1}^*/P^*)]$$

$$(2\pi\,\sigma^2)^{-3/2}\exp(\underline{u} - \underline{\nu})'(\underline{u} - \underline{\nu})/2\sigma^2)$$

$$du_1 du_2 du_3 \tag{5.7.2}$$

with

$$\underline{\nu} = \begin{bmatrix} 0 \\ \mu_2 - \mu_1 \\ \mu_3 - \mu_1 \end{bmatrix},$$

$$\underline{1}^* = \begin{bmatrix} \exp(-b_1{}^*/4\sigma^2)F(\frac{b_1{}^*}{2} - b_2{}^*,\ \infty,\ V_1)2\pi\ \frac{\sigma^2}{\sqrt{2}} \\ \exp(-b_2{}^*/4\sigma^2)F(\frac{b_2{}^*}{2} - b_1{}^*,\ \infty,\ V_2)2\pi\ \frac{\sigma^2}{\sqrt{2}} \\ 0 \end{bmatrix},$$

and

$$P^* = F(b_1{}^*, b_2{}^*, \infty, Q)(2\pi)^{+3/2}|Q|^{\frac{1}{2}} .$$

Substituting for the y_i in (8.3.20) $b_1{}^*$ and $b_2{}^*$ are found to be

$$b_1{}^* = u_2 - u_1 - 2\sigma^2/\theta_1 + \sigma^2/\theta_2 ,$$

$$b_2{}^* = u_3 - u_2 - 2\sigma^2/\theta_2 + \sigma^2/\theta_1 .$$

Thus, the expression is a function of σ^2, θ_1, θ_2, and differences in the elements of $\underline{\mu}$, i.e.,

$$\begin{bmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_1 \end{bmatrix} . \tag{5.7.3}$$

The lower envelope for the mean square error would be a function of the differences in the means also. However, approximating the minimum value of (5.7.2) for fixed values of the elements in (5.7.3) would require searching over the possible values of both θ_1 and θ_2. The search method used for the univariate and bivariate cases would not be applicable in this situation.

Based on the results obtained in the univariate and bivariate cases discussed earlier in this chapter, it would seem likely that functions of the observations could be constructed to use for the parameters θ_1 and θ_2 which would give a reduction in maximum regret in this case. The domain of the observations could be divided into a grid based on values of $y_2 - y_1$ and $y_3 - y_2$. The form of the functions optimal for θ_1 and θ_2 in each set formed by the grid and for several values for the elements of (5.7.3). Again, it would be necessary to find an appropriate multidimensional numerical integration algorithm for evaluating the integral on the various elements of the grid. Thus, the task of finding improved estimators would be considerably more involved than was the case in the univariate and bivariate situations.

6. SUMMARY

Quite often a modeler knows that the true values of the parameters in his model could not possibly be contained in certain sets of the parameter space. This paper has examined such a situation for a linear model whose errors are distributed normally with a known covariance matrix. Attention was restricted to the case where the moderler knows linear inequalities which define the feasible parameter space. Three alternative estimation techniques were presented which took into account these restrictions on the parameter space.

The literature contains many treatises on maximizing the likelihood function with restrictions of this sort. Maximizing the normal likelihood function is equivalent to minimizing a quadratic function, and the algorithms of quadratic programming give solutions to the problems of minimizing a quadratic function. Special, simplified algorithms exist for certain design matrices and for the cases when the restrictions are orderings of the parameters. Estimates in these cases are called the isotonic regression with respect to the ordering.

The restricted maximum likelihood estimators were shown to have some desirable properties. They possess a type of consistency and give a smaller mean square error than the unrestricted estimators in some cases. A property of these estimators which is unappealing is that all of the unrestricted estimates which violate the restrictions will be mapped to a boundary of the feasible parameter space. The consequences of this property is that many unrestricted estimates which are quite different are mapped to the same point on the boundary of the feasible parameter space by these restricted maximum likelihood procedures, so that they pile up on the boundary. It is hard to believe that the true parameter values are so often right on the boundary.

Bayesian estimators are used frequently in situations where the modeler knows that some subsets of the parameter space are more likely to contain the true value of the parameters than are other subsets. However, there have been few publications which deal with assigning a zero prior probability to portions of the sample space. For this reason, Chapter 3 dealt with the basic properties of Bayesian estimation on restricted parameter spaces.

The mean of the posterior distribution is the Bayesian estimator most commonly used. In Chapter 3, it was shown that this Bayesian estimator would not take on some values of the feasible parameter space unless the posterior distribution became degenerate at the boundary for some set of observations. These Bayesian estimators, too, are unappealing. Other measures of the central tendency of the posterior distribution did not seem to yield viable alternatives for estimation on the restricted parameter space.

Several different types of priors were illustrated which would yield a truncated normal posterior for the case of normal likelihood function. The truncated normal posterior was shown to degenerate at the boundary for some observations in the univariate and bivariate case. Thus, the mean of these truncated normal posteriors could give estimates in every point of the feasible parameter space.

The expression for the expectation of the truncated multivariate normal posteriors was found to include multivariate normal distribution functions. These distribution functions are analytically tractable only for special cases and these special cases do not necessarily coincide with the common situations.

The problem of estimating the mean of a univariate distribution was examined in detail to determine if some of the Bayesian estimators proposed would give a uniformly smaller mean square error than was found for the restricted maximum likelihood estimator. No such estimator was found, but many of the Bayesian estimators would give a smaller mean square error over a portion of the parameter space.

The third estimator examined consisted of a Bayesian estimator over a portion of the sample space and the maximum likelihood estimator over the remainder. It was hoped that this would take advantage of smaller mean square errors found for Bayesian estimators near the boundary, without incurring the larger mean square errors that the Bayesian estimators had away from the boundary. As a measure of how well one was doing in reconciling these two goals, the regret function was introduced, and the criterion of goodness chosen was minimax regret. See Section 5.2. For the case in which the variance of the underlying density is one, an optimal estimator of this type was found to be one in which the mean of the posterior found from an exponential prior with θ equal to 0.875 is used for an observation less than 1.5 and the unrestricted maximum likelihood estimator is used for an observation greater than 1.5. This estimator had a maximum regret of 0.47991 which shows a decrease from the maximum regret of 0.58386 found for the restricted maximum likelihood estimator.

Next, a procedure was proposed in which a different estimator would be used on several preassigned intervals in the sample space. Using a stepwise procedure for optimizing the choice of estimators, an estimation procedure was found which would reduce the maximum regret to 0.304207. These results indicate that some continuous functions of the observations could be used for the parameter θ in the Bayesian estimator and should lead to an estimator giving smaller maximum regret. Based on the limited information obtained here, some linear functions to use for θ were examined; the best gave a maximum regret of 0.361671.

The examined choices of functions for θ were good for a univariate normal density with variance one, whose mean was greater than zero. In Section 5.5, a method was shown for choosing the optimal function of the observations for θ for other variance values, for other feasible parameter spaces, or for other sample sizes.

Section 5.6 showed that in the problem of estimating two ordered parameters maximum regret could again be reduced in the same manner given for the univariate case. This procedure was still found to be relatively simple since the mean square errors of the Bayesian and maximum likelihood estimators are a function of the difference in the ordered parameters. However, extending these algorithms to more than two parameters, with a simple ordering, was shown to be a problem of much greater magnitude. For these cases the mean square error of the estimators became functions with arguments of higher dimensionality. The pedestrian techniques of analysis used for the one-dimensional case were found to be no longer adequate.

This study has shown that point estimators can be constructed which use to a greater extent more precise information regarding the parameter space. The criterion minimizing maximum regret is particularly applicable in the situation in which it is difficult to specify a particular prior distribution for the parameter. However, optimizing this criterion function was found to be most difficult and computer costs were excessive even for the crudest of approximations. This author would suggest that those interested in extending this method give top priority to the development of algorithms for finding the parameter θ as a function of the observations which would give the optimum for this criterion. A better optimizing algorithm would make for a much simpler task of extending this technique to the case of m ordered parameters.

This study did not exhaust the Bayesian alternatives for estimation on restricted parameter spaces. This area of study has been virtually untapped thus far. The mode of the posterior was examined only for the uniform prior. Under a more intensive study of other priors the mode of the posterior might be found to yield estimators with more desirable properties than the estimators presented here.

7. LIST OF REFERENCES

Abramowitz, M., and I. A. Stegun. 1964. Handbook of Mathematical Functions. National Bureau of Standards, Washington, D. C.

Ayer, M., H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman. 1955. An emperical distribution function for sampling with incomplete information. Annals of Mathematical Statistics. 26:641-647.

Bancroft, T. A. 1944. On biases in estimation due to the use of preliminary test of significance. Annals of Mathematical Statistics. 15:190-204.

Barlow, R. E., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk. 1972. Statistical Inference Under Order Restrictions. John Wiley and Sons, Inc., New York City, New York.

Bartholomew, D. J. 1965. A comparison of some Bayesian and frequentist inferences. Biometrika. 52:19-35.

Bellman, R. E. and S. E. Dreyfus. 1962. Applied Dynamic Programming. Princeton University Press, Princeton, New Jersey.

Birnbaum, Z. W. and P. L. Meyer. 1953. On the effect of truncation in some or all co-ordinates of a multinormal population. Journal of the Indian Society of Agricultural Statistics. 5:17-27.

Boot, J. C. B. 1964. Quadratic Programming: Algorithms, Anomalies, Applications. North-Holland Publishing Company, Amsterdam.

Brunk, H. D. 1958. On the estimation of parameters restricted by inequalities. Annals of Mathematical Statistics. 29:437-453.

Cramér, H. 1951. Mathematical Methods of Statistics. Princeton University Press, Princeton, New Jersey.

Cunrow, R. N., and C. W. Dunnett. 1962. The numerical evaluation of multivariate normal integrals. Annals of Mathematical Statistics. 33:571-579.

Dutt, J. E. 1973. A representation of multivariate normal probability integrals by integral transforms. Biometrika. 60:637-645.

Ghizzetti, A., and A. Ossicini. 1970. Quadrature Formulae. Academic Press Inc., New York City, New York.

Gun, A. 1965. The use of a preliminary test for interactions in the estimation of factorial means. Institute of Statistics Mimeograph Series, Number 436. North Carolina State University, Raleigh, North Carolina.

Gupta, S. S. 1963. Probability integrals of multivariate normal and multivariate t. Annals of Mathematical Statistics. 34:792-828.

Hadley, G. F. 1964. Nonlinear and Dynamic Programming. Addison-Wesley Publishing Company, Reading, Massachusetts.

Hildebrand, F. B. 1956. Introduction to Numerical Analysis. McGraw-Hill Book Company, Hightstown, New Jersey.

Hudson, D. J. 1969. Least squares fitting of a polynomial constrained to be either non-negative, non-decreasing, or convex. Journal of the Royal Statistical Society. 31:113-118.

Judge, G. G., and T. Takayama. 1966. Inequality restrictions in regression analysis. Journal of the American Statistical Association. 61:116-181.

Kendall, M. G., and A. Stuart. 1969. The Advanced Theory of Statistics. Vol. I. 3rd ed. Hafner Publishing Company, Inc., New York City, New York.

Kruskal, J. B. 1964. Nonmetric multidimensional scaling: A numerical method. Psychometrika. 29:115-129.

Kunzi, H. P., W. Krelle, and W. Oettli. 1966. Nonlinear Programming. Translated by F. Levin, Blaisdell Publishing Company, Waltham, Massachusetts.

Lovell, M. C., and E. Prescott. 1970. Multiple regression with inequality constraints, pretesting bias, hypothesis testing and efficiency. Journal of the American Statistical Association. 65:913-925.

Malinvaud, E. 1966. Statistical Methods of Econometrics. Rand McNally and Company, Chicago, Illinois.

Mantel, N. 1969. Restricted least squares regression and quadratic programming. Technometrics. 11:763-773.

Milton, R. C. 1972. Computer evaluation of the multivariate normal integral. Technometrics. 14:881-887.

Mosteller, F. 1948. On pooling data. Journal of the American Statistical Association. 43:231-242.

Raiffa, H., and R. Schlaifer. 1961. Applied Statistical Decision Theory. Division of Research, Graduate School of Business Administration, Harvard University, Boston, Massachusetts.

Savage, I. R. 1968. Statistics: Uncertainty and Behavior. Houghton Mifflin Company, Boston, Massachusetts.

Savage, L. J. 1954. The Foundations of Statistics. John Wiley and Sons, Inc., New York City, New York.

Searle, S. R. 1971. Linear Models. John Wiley and Sons, Inc., New York City, New York.

System/360 Scientific Subroutine Package. 1970. International Business Machines Corporation, White Plains, New York.

Theil, H., and C. Van de Panne. 1961. Quadratic programming as an extension of conventional quadratic maximation. Journal of the Institute of Management Science. 7:1-20.

Tiao, C. C., and G. E. P. Box. 1973. Some comments on Bayes estimators. The American Statistician. 27:12-14.

Wald, A. 1950. Statistical Decision Functions. John Wiley and Sons, Inc., New York City, New York.

Zellner, A. 1961. Linear Regression with Inequality Constraints on the Coefficients. Mimeographed Report 6109 of the International Center for Management Science.

Zellner, A. 1971. An Introduction to Bayesian Inference in Econometrics. John Wiley and Sons, Inc., New York City, New York.

8.1 Theorems and Proofs

In this appendix conventional mathematical and statistical symbolism and the terminology of Section 1.2 will be used without further explanation.

Theorem 1

Suppose the matrix D is positive definite and one wishes to minimize the function $F(\underline{x}) = \underline{x}'D\underline{x}$ where $\underline{x}$ is restricted to an arbitrary, closed set B. If the basic estimate $\underline{x}_0$ is not in B and if $\underline{x}_s$ minimizes F among all the boundary points of B, then $\underline{x}_s$ is a minimal feasible solution.

Proof

Since D is positive definite, $F(x)$ is a strictly convex function on R^n (*cf.* *e.g.*, Kunzi *et al.*, 1966, p. 38). Therefore for any $\underline{x}_r \neq \underline{x}_0$

$$F(\lambda\underline{x}_r + (1-\lambda)\underline{x}_0) < \lambda F(\underline{x}_r) + (1-\lambda)F(\underline{x}_0)$$

for $0 < \lambda < 1$. Since $F(\underline{x}_0)$ is the global minimum of $F(\underline{x})$, $F(x_0) < F(\underline{x}_r)$, and then

$$F(\lambda\underline{x}_r + (1-\lambda)\underline{x}_0) < F(\underline{x}_r) .$$

If $\underline{x}_r$ is an interior point of B, choose λ so that $(\lambda\underline{x}_r + (1-\lambda)\underline{x}_0)$ lies on the boundary of B. Then for any point, $\underline{x}_r$, in B not on the boundary of B, there exists a point on the boundary, $\underline{x}_b$, such that $F(\underline{x}_r) > F(\underline{x}_b)$. Therefore, the $\underline{x} \epsilon B$ which minimizes $F(\underline{x})$ is a boundary point $\underline{x}_s$ which minimizes $F(\underline{x})$ for all the boundary points.

Theorem 2

A non-degenerate probability distribution with support within a half closed half line (closed interval) has a mean bounded away from the finite end point(s) of that half line (interval), i.e.,

1) $F(\mu)$ is a distribution function with support $B \subset D \subset R^1 = (-\infty, \infty)$ where $B \neq r \; \forall \; r \epsilon R^1$ and $D = [s, +\infty)$ for some $s \epsilon R^1$ or $D = (-\infty, t]$ for some $t \epsilon R^1$ or $D = [s,t]$. Without loss of generality take $D = [s, +\infty)$.

2) $\hat{\mu} \equiv \int_B \mu \, dF(\mu), = \int_D \mu \, dF(\mu) \Rightarrow \exists \; \epsilon > 0 \ni \hat{\mu} \geq s + \epsilon$.

Proof

Note that, by the definition of D ,

$$\lim_{\substack{\Delta > 0 \\ \Delta \to 0}} F(s-\Delta) = F(s-0) = 0 .$$

Then $\forall \; \delta > 0$,

$$\hat{\mu} = \int_{[s,s+\delta]} \mu \, dF(\mu) + \int_{(s+\delta,+\infty)} \mu \, dF(\mu)$$

$$\geq \int_{[s,s+\delta]} s \, dF(\mu) + \int_{(s+\delta,+\infty)} (s+\delta) \, dF(\mu)$$

$$= s\{\int_{[s,s+\delta]} dF(\mu) + \int_{(s+\delta,+\infty)} dF(\mu)\} + \delta\int_{(s+\delta,+\infty)} dF(\mu)$$

$$= s + \delta[1 - F(s+\delta)] = s + \delta h(\delta), \text{ say} .$$

Now, $\exists\ \delta > 0 \ni h(\delta) > 0$ (otherwise $F(s+\delta) = 1\ \forall\ \delta > 0$ $\Rightarrow F(s) = F(s+0) = 1 \Rightarrow dF(s) = F(s) - F(s-0) = 1 \Rightarrow s = B$, a contradiction of the first statement in the theorem). Choose such a $\delta > 0$. Then

$$\hat{\mu} \geq s + \epsilon$$

where

$$\epsilon \equiv \delta h(\delta) > 0 .$$

Theorem 3

A non-degenerate probability distribution function, $F_n(\underline{\upsilon})$, on R^n with support within a convex proper subset, D^n , of R^n closed with respect to boundary points with finite coordinates has a mean bounded away from every boundary point, $\underline{s}$, of D^n with finite coordinates, *i.e.*,

1) $F_n(\underline{\upsilon})$ is a distribution function with support $B^n \subset D^n \subset R^n$ where $B^n \neq \{\underline{\upsilon}: \underline{c}'\underline{\upsilon} = d\ \forall\ \underline{c} \in R^n, d \in R^1\}$,
2) $\hat{\underline{\upsilon}} \equiv \int_{B^n} \underline{\upsilon} dF(\underline{\upsilon}) = \int_{D^n} \underline{\upsilon} dF(\underline{\upsilon})$, and
3) $\underline{s}$ is a finite boundary point of D^n with finite coordinates $\Rightarrow \hat{\underline{\upsilon}}$ is bounded away from $\underline{s}$.

Proof

Let $\underline{c}'\underline{\upsilon} = s$, $\underline{c} \epsilon R^n$, $\underline{c}'\underline{c} = 1$, $s \epsilon R^1$ be a supporting hyperplane of D^n containing $\underline{s}$. Without loss of generality assume

$$D^n \subset \{\underline{\upsilon}:\ \underline{c}'\underline{\upsilon} \geq d\} \ . \qquad (8.1.1)$$

Consider a random variable, say,

$$\underline{\mathfrak{V}}_{n \times 1} \sim F_n(\underline{\upsilon}) \ .$$

Define a scalar random variable

$$\mathfrak{U} \equiv \underline{c}'\underline{\mathfrak{V}} \ .$$

Then,

$$\hat{\mu} = E(\mathfrak{U}) = E(\underline{c}'\underline{\mathfrak{V}}) = \underline{c}'\hat{\underline{\upsilon}} \ . \qquad (8.1.2)$$

Moreover, it is clear that $\mathfrak{U} \sim F(\mu)$ where F satisfies the hypotheses of Theorem 2. Thus by Theorem 2 , $\exists\ \epsilon > 0$ such that

$$\hat{\mu} \geq s + \epsilon \ . \qquad (8.1.3)$$

Define $C_{n \times n}$, orthonormal, with first row equal to $\underline{c}'$. Then, the squared distance between $\underline{\upsilon}$ and $\underline{s}$ is seen to satisfy

$$(\hat{\underline{\upsilon}}-\underline{s})'(\hat{\underline{\upsilon}}-\underline{s}) = (\hat{\underline{\upsilon}}-\underline{s})'C'C(\hat{\underline{\upsilon}}-\underline{s})$$

$$= (C\hat{\underline{\upsilon}}-C\underline{s})'(C\hat{\underline{\upsilon}}-C\underline{s})$$

$$\geq (\underline{c}'\hat{\underline{\upsilon}} - \underline{c}'\underline{s})^2$$

(i.e., the squared length of a vector is not less than the square of the first coordinate)

$$= (\hat{\mu}-s)^2$$

(by (8.1.2) and the fact that $\underline{s}$ is in the hyperplace $\underline{c}'\underline{\nu} = s$) . Therefore

$$(\hat{\underline{\nu}}-\underline{s})'(\hat{\underline{\nu}}-\underline{s}) \geq \epsilon^2 > 0$$

(by (8.1.3)), i.e., $\hat{\underline{\nu}}$ is not closer than $\epsilon^2 > 0$ to $\underline{s}$, an arbitrary boundary point of D^n with finite coordinates. Thus, $\underline{\nu}$ is not closer than $\epsilon^2 > 0$ to any boundary point of D^n with finite coordinates.

Table 8.1 Values of the function f(x)/F(x)

x	f(x)/F(x)	x	f(x)/F(x)	x	f(x)/F(x)
-10.0	10.0980930	-6.0	6.1584826	-2.0	2.3732147
- 9.9	9.9990463	-5.9	6.0609159	-1.9	2.2849464
- 9.8	9.9000187	-5.8	5.9634228	-1.8	2.1973124
- 9.7	9.8010092	-5.7	5.8660049	-1.7	2.1103573
- 9.6	9.7020197	-5.6	5.7686663	-1.6	2.0241289
- 9.5	9.6030493	-5.5	5.6714095	-1.5	1.9386768
- 9.4	9.5041008	-5.4	5.5742397	-1.4	1.8540564
- 9.3	9.4051723	-5.3	5.4771595	-1.3	1.7703276
- 9.2	9.3062668	-5.2	5.3801737	-1.2	1.6875515
- 9.1	9.2073832	-5.1	5.2832870	-1.1	1.6057968
- 9.0	9.1085224	-5.0	5.1865034	-1.0	1.5251350
- 8.9	9.0096865	-4.9	5.0898285	-0.9	1.4456425
- 8.8	8.9108744	-4.8	4.9932661	-0.8	1.3674021
- 8.7	8.8120880	-4.7	4.8968239	-0.7	1.2904987
- 8.6	8.7133284	-4.6	4.8005056	-0.6	1.2150249
- 8.5	8.6145945	-4.5	4.7043190	-0.5	1.1410770
- 8.4	8.5158901	-4.4	4.6082706	-0.4	1.0687561
- 8.3	8.4172134	-4.3	4.5123672	-0.3	0.9981660
- 8.2	8.3185673	-4.2	4.4166174	-0.2	0.9294158
- 8.1	8.2199516	-4.1	4.3210268	-0.1	0.8626174
- 8.0	8.1213675	-4.0	4.2256069	0.0	0.7978845
- 7.9	8.0228167	-3.9	4.1303644	0.1	0.7353317
- 7.8	7.9243002	-3.8	4.0353117	0.2	0.6750731
- 7.7	7.8258181	-3.7	3.9404573	0.3	0.6172208
- 7.6	7.7273731	-3.6	3.8458128	0.4	0.5618827
- 7.5	7.6289663	-3.5	3.7513905	0.5	0.5091604
- 7.4	7.5305977	-3.4	3.6572037	0.6	0.4591471
-.7.3	7.4322701	-3.3	3.5632658	0.7	0.4119247
- 7.2	7.3339844	-3.2	3.4695911	0.8	0.3675614
- 7.1	7.2357426	-3.1	3.3761969	0.9	0.3261089
- 7.0	7.1376456	-3.0	3.2830982	1.0	0.2875999
- 6.9	7.0393953	-2.9	3.1903143	1.1	0.2520463
- 6.8	6.9412937	-2.8	3.0978661	1.2	0.2194365
- 6.7	6.8432426	-2.7	3.0057716	1.3	0.1897350
- 6.6	6.7452450	-2.6	2.9140568	1.4	0.1628812
- 6.5	6.6473007	-2.5	2.8227444	1.5	0.1387897
- 6.4	6.5494137	-2.4	2.7318611	1.6	0.1173516
- 6.3	6.4515858	-2.3	2.6414347	1.7	0.0984359
- 6.2	6.3538198	-2.2	2.5514956	1,8	0.0818925
- 6.1	6.2561178	-2.1	2.4520771	1.9	0.0675557

Table 8.1 (Continued)

x	f(x)/F(x)	x	f(x)/F(x)	x	f(x)/F(x)
2.0	0.0552479	3.1	0.0032700	4.1	0.0000893
2.1	0.0447836	3.2	0.0023857	4.2	0.0000589
2.2	0.0359748	3.3	0.0017234	4.3	0.0000385
2.3	0.0286341	3.4	0.0012326	4.4	0.0000249
2.4	0.0225796	3.5	0.0008729	4.5	0.0000160
2.5	0.0176378	3.6	0.0006120	4.6	0.0000064
2.6	0.0136466	3.7	0.0004248	4.7	0.0000040
2.7	0.0104572	3.8	0.0002920	4.8	0.0000024
2.8	0.0079357	3.9	0.0001987	4.9	0.0000015
2.9	0.0059637	4.0	0.0001338	5.0	0.0000009
3.0	0.0044378				

8.3 The Mean of a Truncated Multivariate Normal Posterior Distribution

Consider the situation in which the posterior distribution has the following form:

$$p(\underline{\beta}|\underline{y}) = \frac{\exp(-(\underline{y}-X\underline{\beta})'V(\underline{y}-X\underline{\beta})/2)}{\int_A \exp(-(\underline{y}-X\underline{\beta})'V(\underline{y}-X\underline{\beta})/2)d\underline{\beta}} \qquad (8.3.1)$$

on the convex set

$$A = \{\underline{\beta}:\ C\underline{\beta} \geq \underline{d}\}$$

and

$$p(\underline{\beta}|\underline{y}) = 0$$

elsewhere. The mean of this posterior would be

$$E(\underline{\beta}) = \int_A \underline{\beta}\, p(\underline{\beta}|\underline{y})d\underline{\beta} \ . \qquad (8.3.2)$$

Evaluating (8.3.2) is no easy task in the multivariate case. Finding the normalizing constant of the probability density (the denominator of (8.3.1)) requires evaluating a multivariate normal probability integral. Kendall and Stuart (1969), pages 350-353, Curnow and Dunnett (1962), Gupta (1963), and Dutt (1973), to mention a few, have given solutions to this integral for special cases of the region of integration. Abramowitz and Stegun (1964), pages 956-957, give techniques which can be adapted to evaluating a bivariate normal probability integral on a convex set. Milton (1972) illustrated the use of multidimensional Simpson quadrature to evaluate multivariate normal probability integrals such as these.

For the cases considered by these authors their techniques provide relatively inexpensive methods of evaluating such integrals on computers. However, the technique which handles the more general situation, Simpson quadrature, becomes quite expensive as the dimensionality increases.

For many practical problems, the numerator of expression (8.3.2) (after substitution of (8.3.1) into (8.3.2)) can also be reduced to evaluating a multivariate normal probability integral. This occurs when (8.3.2) reduces to finding the mean of a posterior distribution

$$p(\underline{\mu}|\underline{y}) \propto \exp(-(\underline{y}-\underline{\mu})' V(\underline{y}-\underline{\mu})/2) \tag{8.3.3}$$

with V positive definite, on the set

$$A = \{\underline{\mu} | a_1 \le \mu_1 \le e_1,\ a_2 \le \mu_2 \le e_2, \ldots, a_m \le \mu_m \le e_m\} ,$$

and $p(\underline{\mu}|\underline{y}) = 0$ elsewhere. Later in this section, the Bayesian estimator for a simple ordering of the mean parameters will be derived by making such a transformation. This example should aid the reader in formulating other problems of this sort so that the reduction which follows can be utilized.

The mean of (8.3.3) would be

$$E(\underline{\mu}) = \frac{\int_{a_m}^{e_m} \cdots \int_{a_2}^{e_2} \int_{a_1}^{e_1} \underline{\mu} \exp(-(\underline{y}-\underline{\mu})' V(\underline{y} - \underline{\mu})/2) d\underline{\mu}}{\int_{a_m}^{e_m} \cdots \int_{a_2}^{e_2} \int_{a_1}^{e_1} \exp(-(\underline{y}-\underline{\mu})' V(\underline{y}-\underline{\mu})/2) d\underline{\mu}} .$$

Making the transformation $\underline{z} = \underline{y} - \underline{\mu}$, the mean can then be written

$$E(\underline{\mu}) = \underline{y} - \frac{\int_{y_m-e_m}^{y_m-a_m} \cdots \int_{y_2-e_2}^{y_2-a_2} \int_{y_1-e_1}^{y_1-a_1} \underline{z} \exp(-\underline{z}'V\underline{z}/2)d\underline{z}}{\int_{y_m-e_m}^{y_m-a_m} \cdots \int_{y_2-e_2}^{y_2-a_2} \int_{y_1-e_1}^{y_1-a_1} \exp(-\underline{z}'V\underline{z}/2)d\underline{z}}$$

$$= \underline{y} - D(\underline{z})/P \text{ , say.} \tag{8.3.4}$$

Following a method used by Birnbaum and Myer (1953), $D(\underline{z})$ can be simplified to an expression involving only normal probability integrals.

By expressing the quadratic form as a sum, the elements of $D(\underline{z})$ can be expressed in the same manner as the following expression for $D(z_1)$, i.e.,

$$D(z_1) = \int_{y_m-e_m}^{y_m-a_m} \cdots \int_{y_2-e_2}^{y_2-a_2} \int_{y_1-e_1}^{y_1-a_1} z_1 \exp[-(z_1^2 v_{11} + 2z_1 \sum_{j=2}^{m} v_{1j} z_j$$

$$+ \sum_{i=2}^{m} \sum_{j=2}^{m} v_{ij} z_i z_j)/2]d\underline{z} \text{ .}$$

Defining $S_1(z_1)$ as

$$S_1(z_1) = \int_{y_m-e_m}^{y_m-a_m} \cdots \int_{y_3-e_3}^{y_3-a_3} \int_{y_2-e_2}^{y_2-a_2} \exp[-(2z_1 \sum_{j=2}^{m} v_{1j} z_j$$

$$+ \sum_{i=2}^{m} \sum_{j=2}^{m} v_{ij} z_i z_j)/2]dz_2 dz_3 \ldots dz_m \text{ ,}$$

then

$$D(z_1) = \int_{y_1-e_1}^{y_1-a_1} z_1 S_1(z_1) e^{-z_1^2 v_{11}/2} dz_1 \text{ .}$$

Integrating by parts gives

$$D(z_1) = S_1(y_1-e_1)\frac{1}{v_{11}} e^{-(y_1-e_1)^2 v_{11}/2} - S_1(y_1-a_1)\frac{1}{v_{11}}$$

$$e^{-(y_1-a_1)^2 v_{11}/2} + \int_{y_1-e_1}^{y_1-a_1} S_1'(z_1)\frac{1}{v_{11}} e^{-z_1^2 v_{11}/2} dz_1 \ .$$

Here

$$S_1'(z_1) = \frac{d(S_1)}{dz_1} = \int_{y_m-e_m}^{y_m-a_m} \cdots \int_{y_3-e_3}^{y_3-a_3} \int_{y_2-e_2}^{y_2-a_2} [- \sum_{j=2}^{m} v_{1j} z_j$$

$$\exp(-(2z_1 \sum_{j=2}^{m} v_{1j} z_j + \sum_{i=2}^{m} \sum_{j=2}^{m} v_{1j} z_i z_j)/2)]$$

$$dz_2 dz_3 \ldots dz_m \ .$$

Thus,

$$v_{11} D(z_1) + \sum_{j=2}^{m} v_{1j} D(z_j) = S_1(y_1-e_1) e^{-(y_1-e_1)^2 v_{11}/2}$$

$$- S_1(y_1-a_1) e^{-(y_1-a_1)^2 v_{11}/2} \ ,$$

or

$$\underline{v}_1 D(\underline{z}) = S_1(y_1-e_1) e^{-(y_1-e_1)^2 v_{11}/2} - S_1(y_1-a_1)$$

$$e^{-(y_1-a_1)^2 v_{11}/2} \qquad (8.3.5)$$

where $\underline{v}_1$ is the first row of V .

Repeating this process for the $i^{\underline{th}}$ element of $D(\underline{z})$, one finds that

$$\underline{v}_i E(\underline{z}) = S_i(y_i-e_i)e^{-(y_i-e_i)^2 v_{ii}/2} - S_i(y_i-a_i)e^{-(y_i-a_i)^2 v_{ii}/2} \qquad (8.3.6)$$

where

$$S_i(z_i) = \int_{y_m-e_m}^{y_m-a_m} \cdots \int_{y_{i+1}-e_{i+1}}^{y_{i+1}-a_{i+1}} \int_{y_{i-1}-e_{i-1}}^{y_{i-1}-a_{i-1}} \cdots \int_{y_1-e_1}^{y_1-a_1} \exp(-(2z_i$$

$$\sum_{\substack{j=1\\ j\neq i}}^{m} v_{ij}z_j + \sum_{\substack{k=1\\ k\neq i}}^{m} \sum_{\substack{j=1\\ j\neq i}}^{m} v_{kj}z_k z_j)/2)$$

$$dz_1 \ldots dz_{i-1}\, dz_{i+1} \ldots dz_m \,. \qquad (8.3.7)$$

The vector $\underline{v}_i$ would be the $i^{\underline{th}}$ row of V . Thus, from (8.3.5) and (8.3.6) it can be seen that

$$VD(\underline{z}) = \underline{1} \qquad (8.3.8)$$

where the $i^{\underline{th}}$ element of $\underline{1}$ would be

$$1_i = S_i(y_i-e_i)e^{-(y_i-e_i)^2 v_{ii}/2} - S_i(y_i-a_i)e^{-(y_i-e_i)^2 v_{ii}/2} \,.$$

Call the matrix V with the $i^{\underline{th}}$ row and column deleted, V_i , and denote the vector whose elements form the $i^{\underline{th}}$ row of V with the $i^{\underline{th}}$ element deleted by $\underline{q}_i$. (Note that V_i would be positive definite if V is positive definite.) Let $\underline{z}_i$ be the vector of all the elements of $\underline{z}$ with the exception of z_i . Then

$$S_i(x) = \int_{y_m-e_m}^{y_m-a_m} \cdots \int_{y_{i+1}-e_{i+1}}^{y_{i+1}-a_{i+1}} \int_{y_{i-1}-e_{i-1}}^{y_{i-1}-a_{i-1}} \cdots \int_{y_1-e_1}^{y_1-a_1} \exp(-2\,X\,\underline{q}_i'\underline{z}_i$$

$$+ \underline{z}_i'V_i\underline{z}_i)/2)dz_i \ .$$

Completing the square, $S_i(x)$ can be expressed as

$$S_i(x) = \exp(x^2\underline{q}_i'V_i^{-1}\underline{q}_i)/2 \int_{y_m-e_m}^{y_m-a_m} \cdots \int_{y_{i+1}-e_{i+1}}^{y_{i+1}-a_{i+1}} \int_{y_{i-1}-e_{i-1}}^{y_{i-1}-a_{i-1}} \cdots$$

$$\int_{y_1-e_1}^{y_1-a_1} \exp(-(xV_i^{-1}\underline{q}_i + \underline{z}_i)'V_i(xV_i^{-1}\underline{q}_i + \underline{z}_i)/2)d\underline{z}_i \ .$$

Making the transformation

$$\underline{t} = xV_i^{-1}\underline{q}_i + \underline{z}_i \ ,$$

$S_i(x)$ becomes

$$S_i(x) = \exp(x^2\underline{q}_i'V_i^{-1}\underline{q}_i/2)\int_{c_{m-1}}^{d_{m-1}} \cdots \int_{c_1}^{d_1} \exp(-\underline{t}'V_i\underline{t}/2)d\underline{t} \ .$$

The vector $\underline{c}$ would be

$$\underline{c} = x(V_i^{-1}\underline{q}_i) + \begin{bmatrix} y_1 - e_1 \\ \cdots \\ y_{i-1} - e_{i-1} \\ y_{i+1} - e_{i+1} \\ \cdots \\ y_m - e_m \end{bmatrix}$$

and $\underline{d}$ would be

$$\underline{d} = x(V_i^{-1}\underline{q}_i) + \begin{bmatrix} y_1 - a_1 \\ \cdots \\ y_{i-1} - a_{i-1} \\ y_{i+1} - a_{i+1} \\ \cdots \\ y_m - a_m \end{bmatrix} .$$

Then

$$S_i(x) = \exp(x^2\underline{q}_i'V_i^{-1}\underline{q}_i/2)(2\pi)^{(m-1)/2}|V_i^{-1}|^{\frac{1}{2}}$$

$$\int_{c_{m-1}}^{d_{m-1}} \cdots \int_{c_1}^{d_1} (2\pi)^{-(m-1)/2}|V_i^{-1}|^{-\frac{1}{2}}$$

$$\exp(-\underline{t}'V_i\underline{t}/2)d\underline{t} , \qquad (8.3.9)$$

and the integral is a multivariate normal probability integral.

Thus, the elements of $\underline{1}$ consist of exponential functions, known constants, and multivariate normal probability integrals. From (8.3.8), it is easily seen that

$$D(\underline{z}) = V^{-1}\underline{1} \qquad (8.3.10)$$

and substituting this expression into (8.3.4) it can be seen that

$$E(\underline{\mu}) = \underline{y} - V^{-1}\underline{1}/P . \qquad (8.3.11)$$

So, for posterior distributions of the form of (8.3.3), finding the Bayesian estimates for an observation $\underline{y}$ consists of evaluating multivariate normal probability integrals.

In the univariate case, the function $S_1(z_i)$ would have the value one for all values of z_1. Then in (8.3.10), $\underline{1}$ would be a scalar and would equal

$$1 = \exp[-(y-e)^2 v_{11}/2] - \exp[-(y-a)^2 v_{11}/2] .$$

The covariance matrix V^{-1} would be the scalar σ^2, so

$$1 = \exp[-(y-e)^2/(2\sigma^2)] - \exp[-(y-a)^2/(2\sigma^2)] ,$$

and from (8.3.4) and (8.3.11)

$$E(\mu) = y - \frac{\sigma^2 \exp[-(y-e)^2/(2\sigma^2)] - \exp[-(y-a)^2/(2\sigma^2)]}{\int_{y-e}^{y-a} \exp[-z^2/(2\sigma^2)]dz}$$

$$= y - \frac{\sigma^2 \exp[-(y-e)^2/(2\sigma^2)] - \exp[-(y-a)^2/(2\sigma^2)]}{\sqrt{2\pi}\,\sigma[F((y-a)/\sigma) - F((y-e)/\sigma)]} \qquad (8.3.12)$$

where $F(x)$ is the distribution function for a normally distributed random variable with mean zero and variance one. Then for $e = \infty$, Equation (8.3.12) becomes

$$E(\mu) = y + \sigma f((y-a)/\sigma)/F((y-a)/\sigma) \qquad (8.3.13)$$

where $f(x)$ is the normal density function with mean zero and variance one. Notice that expression (8.3.13) is identical to Cramer's results quoted in (4.1.2).

For the next example, consider the posterior given in (8.3.3), but with support, A , defined as

$$A = \{\underline{\mu}: \mu_1 \le \mu_2 \le \dots \le \mu_m\} , \tag{8.3.14}$$

i.e., a simple ordering of the μ_i . The mean of this posterior would be

$$E(\underline{\mu}) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\mu_{m-1}} \cdots \int_{-\infty}^{\mu_3}\int_{-\infty}^{\mu_2} \underline{\mu} \exp(-(\underline{y}-\underline{\mu})' V(\underline{y}-\underline{\mu})/2) d\underline{\mu}}{\int_{-\infty}^{\infty}\int_{-\infty}^{\mu_{m-1}} \cdots \int_{-\infty}^{\mu_3}\int_{-\infty}^{\mu_2} \exp(-(\underline{y}-\underline{\mu})' V(\underline{y}-\underline{\mu})/2) d\underline{\mu}} .$$

Make the one-to-one transformation $\underline{z} = H(\underline{y} - \underline{\mu})$, where

$$H = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ 0 & 0 & -1 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & 0 & \dots & -1 & 1 \\ 0 & 0 & 0 & \dots & 0 & -1 \end{bmatrix}$$

or equivalently

$$\begin{cases} z_i = y_{i+1} - y_i - \mu_{i+1} + \mu_i , & i = 1, \dots, m-1 \\ z_m = -y_m + \mu_m . \end{cases} \tag{8.3.15}$$

Now the region of integration in terms of the μ-coordinates is defined by the inequalities

$$\begin{cases} -\infty < \mu_1 \\ 0 < \mu_{i+1} - \mu_i , \qquad i = 1, \ldots, m-1 \\ \mu_m < +\infty \end{cases}$$

which entail the following inequalities in terms of the z-coordinates:

$$\begin{cases} -\infty < z_i < y_{i+1} - y_i , \qquad i = 1, \ldots, m-1 \\ -\infty < z_m < +\infty . \end{cases} \tag{8.3.16}$$

On the other hand, for any set of z-values satisfying the second set of inequalities one can find a set of μ-values, according to equations (8.3.15), which satisfy the first set of inequalities. In fact, solving (8.3.15) one finds

$$\begin{cases} \mu_m = y_m + z_m \\ \mu_i = y_i + z_{i+1} + z_i \qquad (i = 1, \ldots, m-1) . \end{cases}$$

Therefore $z_m < +\infty$ implies $\mu_m < +\infty$ and $z_{m-1} < y_m - y_{m-1}$ implies

$$\mu_{m-1} = y_{m-1} + z_m + z_{m-1} < y_{m-1} + z_m + y_m - y_{m-1} = z_m + y_m = \mu_m ,$$

and so on for the indices $i < m-1$. This proves that the region of integration in terms of the z-coordinates is given by the inequalities (8.3.16).

Thus

$$E(\underline{\mu}) = \underline{y} - \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{b_{m-1}}\cdots\int_{-\infty}^{b_2}\int_{-\infty}^{b_1} H^{-1}\underline{z}\ \exp(-(\underline{z}'H^{-1}{}'VH^{-1}\underline{z})/2)d\underline{z}}{\int_{-\infty}^{\infty}\int_{-\infty}^{b_{m-1}}\cdots\int_{-\infty}^{b_2}\int_{-\infty}^{b_1} \exp(-(\underline{z}'H^{-1}{}'VH^{-1}\underline{z})/2)d\underline{z}}$$

where $b_i = y_{i+1} - y_i$. Let

$$Q = H^{-1}{}'\,VH^{-1} .$$

Notice that Q is positive definite since H^{-1} is non-singular and V is positive definite. The matrix Q is symmetric since V is symmetric. Thus,

$$E(\underline{\mu}) = \underline{y} - \frac{H^{-1}\int_{-\infty}^{\infty}\int_{-\infty}^{b_{m-1}}\cdots\int_{-\infty}^{b_2}\int_{-\infty}^{b_1} \underline{z}\ \exp(-\underline{z}'Q\underline{z}/2)d\underline{z}}{\int_{-\infty}^{\infty}\int_{-\infty}^{b_{m-1}}\cdots\int_{-\infty}^{b_2}\int_{-\infty}^{b_1} \exp(-\underline{z}'\,Q\underline{z}/2)d\underline{z}}$$

$$= \underline{y} - H^{-1}D(\underline{z})/P = \underline{y} - H^{-1}Q^{-1}\underline{1}/P \qquad (8.3.17)$$

by applying the argument which derives (8.3.11) from (8.3.4).

The term $D(\underline{z})$ in (8.3.17) is similar to the term $D(z)$ in (8.3.4). By substituting b_i for $(y_i - a_i)$, $i = 1, 2, \ldots, m-1$, ∞ for $(y_m - a_m)$, $-\infty$ for $(y_i - e_i)$, $i = 1, 2, \ldots, m$, and Q for V in the derivation of (8.3.10); (8.3.17) can be expressed in terms of multivariate normal integrals and other functions more easily evaluated. An example will now be given to show how this can be accomplished in case $m = 2$.

Consider the posterior

$$p(\underline{\mu}|\underline{y}) = \frac{\exp(-(\underline{y}-\underline{\mu})'I\sigma^{-2}(\underline{y}-\underline{\mu})/2)}{\int_A \exp(-(\underline{y}-\underline{\mu})'I\sigma^{-2}(\underline{y}-\underline{\mu})/2)d\underline{\mu}} \qquad \text{for} \quad \underline{\mu} \in A \ , \qquad (8.3.18)$$

$$= 0 \qquad\qquad \text{elsewhere}$$

where the set A is as follows:

$$A = \{\underline{\mu}: \mu_1 \le \mu_2\} \ .$$

Then

$$H = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix} \ ,$$

$$H^{-1} = \begin{bmatrix} -1 & -1 \\ 0 & -1 \end{bmatrix}$$

and

$$Q = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\sigma^{-2} \ .$$

Notice that

$$b_1 = (y_2-y_1) \ , \quad q_{11} = \sigma^{-2} \ , \quad \underline{q}_1 = \sigma^{-2} \ , \quad Q_1^{-1} = \sigma^2/2 \ ,$$

and

$$|Q_1^{-1}| = \sigma^2/2 \ .$$

For this example, the appropriate substitutions in (8.3.8) would be to set (y_1-e_1) and (y_2-e_2) equal to $-\infty$, (y_2-a_2) to ∞, and (y_1-a_1) equal to (y_2-y_1). Then solving for $S_1(y_2-y_1)$ and $S_2(\infty)$ in (8.3.9), one finds that

$$\underline{1} = \begin{bmatrix} \exp(-\tfrac{1}{2}(y_2-y_1)^2/(2\sigma^2))\sqrt{2\pi}\,\sigma^2/2 \\ \\ 0 \end{bmatrix} .$$

Then substituting for $\underline{1}$ in (8.3.17) gives

$$E(\underline{\mu}) = \underline{y} + [{}^{-1}_{\ 1}]\,\frac{\exp(-(y_2-y_1)^2/(4\sigma^2))\sigma^2\sqrt{2\pi}\,\sigma^2/2}{\int_{-\infty}^{\infty}\int_{-\infty}^{y_2-y_1}\exp(-\underline{z}'Q\underline{z}/2)d\underline{z}} .$$

Now

$$\int_{-\infty}^{y_2-y_1}\int_{-\infty}^{\infty}\exp(-2(z_2^2 + z_1z_2 + z_1^2/4)/(2\sigma^2))dz_2\,\exp(-z_1^2/4\sigma^2)dz_1$$

$$= F((y_2-y_1)/(\sigma\sqrt{2}))\cdot 2\pi\sigma^2 .$$

Thus

$$E(\underline{\mu}) = \underline{y} + [{}^{-1}_{\ 1}]\sigma f((y_2-y_1)/\sigma\sqrt{2})/[f((y_2-y_1)/\sigma\sqrt{2})\sqrt{2}] . \qquad (8.3.19)$$

Now consider an example in which a random sample of size one is taken from each of three populations, where the $i^{\underline{th}}$ population has normal density with mean μ_i $(i = 1, 2, 3)$ and known variance σ^2. Suppose the parameters μ_i are unknown, but are known to satisfy the ordering

$$\mu_1 \le \mu_2 \le \mu_3 .$$

The set in $\underline{\mu}$-space defined by this ordering will again be called A. The Bayesian estimators for the μ_i from an exponential prior of the form (4.2.2) can be derived by finding the expected value of a posterior distribution of the following form:

$$p(\underline{\mu}|\underline{y}) \propto \exp[-(\underline{y}-\underline{\mu})'\sigma^{-2}I(\underline{y}-\underline{\mu})/2 - (\mu_2-\mu_1)/\theta_1 - (\mu_3-\mu_2)/\theta_2]$$

$$\text{for} \quad \underline{\mu} \in A$$

and $p(\underline{\mu}|\underline{y})$ is zero elsewhere. Then on the set A, $p(\underline{\mu}|\underline{y})$ would be

$$p(\underline{\mu}|\underline{y}) \propto \exp[-((\underline{y}-\sigma^2 C'\underline{\theta})-\underline{\mu})'(\underline{y}-\sigma^2 C'\underline{\theta})/2\sigma^2]$$

as is shown in (4.2.3). Here the matrix C would be

$$C' = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \qquad \text{and} \qquad C'\underline{\theta} = \begin{bmatrix} -1/\theta_1 \\ +1/\theta_1 - 1/\theta_2 \\ 1/\theta_2 \end{bmatrix} .$$

Now this posterior is of the form of (8.3.3), where the vector $\underline{y}$ in (8.3.3) is replaced by the vector $\underline{y} - \sigma^2 C'\underline{\theta}$. The support of the posterior is of the form (8.3.14), so the appropriate Bayesian estimator is found by evaluating the expression given in (8.3.17).

The expression for the denominator, P, will be developed first. The limits of integration, b_i, are found by substituting the appropriate expressions for the y_i so that

$$\left\{\begin{aligned} b_1 &= y_2 - \sigma^2/\theta_1 + \sigma^2/\theta_2 - (y_1 + \sigma^2/\theta_1) \\ &= y_2 - y_1 - 2\sigma^2/\theta_1 + \sigma^2/\theta_2 \\ b_2 &= y_3 - \sigma^2/\theta_2 - (y_2 - \sigma^2/\theta_1 + \sigma^2/\theta_2) \\ &= y_3 - y_2 - 2\sigma^2/\theta_2 + \sigma^2/\theta_1 . \end{aligned}\right. \qquad (8.3.20)$$

The matrix Q is

$$Q = \sigma^{-2} \begin{bmatrix} -1 & 0 & 0 \\ -1 & -1 & 0 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} -1 & -1 & -1 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} \sigma^{-2} .$$

The expression for P would then be

$$P = \int_{-\infty}^{\infty} \int_{-\infty}^{y_3 - y_2 - 2\sigma^2/\theta_2 + \sigma^2/\theta_1} \int_{-\infty}^{y_2 - y_1 - 2\sigma^2/\theta_1 + \sigma^2/\theta_2}$$

$$\exp(-\underline{z}'Q\underline{z}/2)\,dz_1 dz_2 dz_3 .$$

The function given for P is sometimes denoted in a way similar to the univariate normal distribution function (cf. Birnbaum and Meyer (1953)). Adopting this notation

$$P = (2\pi)^{3/2} |Q^{-1}|^{\frac{1}{2}} F_3(y_2 - y_1 - 2\sigma^2/\theta_1 + \sigma^2/\theta_2 ,$$

$$y_3 - y_2 - 2\sigma^2/\theta_2 + \sigma^2/\theta_1 , \infty , Q) .$$

Evaluating the numerator of (8.3.17) a substitution should be made into (8.3.10) as was outlined in the paragraph prior to (8.3.18), i.e., $y_i - a_i = b_i$ $(i = 1, 2)$, $y_3 - a_3 = \infty$, $y_i - e_i = -\infty$ $(i = 1, 2, 3)$. Then the above mentioned Q will correspond to V in (8.3.8) and so the vector $\underline{1}$ given there would have values

$$\underline{1} = \begin{bmatrix} S_1(b_1)e^{-(b_1)^2/2\sigma^2} \\ \\ S_2(b_2)e^{-b_2^2/\sigma^2} \\ \\ 0 \end{bmatrix}. \qquad (8.3.21)$$

Recall that in expression (8.3.9) V_i is the matrix V with the $i^{\underline{th}}$ row and column deleted, and $\underline{q}_i$ is the $i^{\underline{th}}$ row of V with the $i^{\underline{th}}$ element deleted. Then

$$V_1 = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}\sigma^{-2}, \qquad V_2 = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}\sigma^{-2}, \qquad V_1^{-1} = \begin{bmatrix} 3 & -2 \\ -2 & 2 \end{bmatrix}\frac{1\sigma^2}{2}$$

$$\underline{q}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\sigma^{-2} \qquad \underline{q}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}\sigma^{-2} \qquad V_2^{-1} = \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix}\frac{1\sigma^2}{2}$$

and substituting these into (8.3.9) it is found that

$$S_1(b_1) = \exp(b_1^2\sigma^{-2}[1\ 1]\begin{bmatrix} 3 & -2 \\ -2 & 2 \end{bmatrix}\frac{\sigma^{+2}}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\frac{\sigma^{-2}}{2})$$

$$(2\pi)\frac{\sigma^2}{\sqrt{2}}\int_{c_2}^{d_2}\int_{c_1}^{d_1}(2\pi)^{-1}|V_1^{-1}|^{-\frac{1}{2}}\exp(-\underline{t}'V_1\underline{t}/2)dt_1dt_2$$

with

$$\underline{c} = b_1\begin{bmatrix} 3 & -2 \\ -2 & 2 \end{bmatrix}\frac{\sigma^2}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\sigma^{-2} + \begin{bmatrix} -\infty \\ -\infty \end{bmatrix} = \begin{bmatrix} -\infty \\ -\infty \end{bmatrix},$$

and

$$\underline{d} = b_1\begin{bmatrix} 3 & -2 \\ -2 & 2 \end{bmatrix}\frac{\sigma^2}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\sigma^{-2} + \begin{bmatrix} b_2 \\ \infty \end{bmatrix} = \begin{bmatrix} b_1/2 - b_2 \\ \infty \end{bmatrix}.$$

Then the integral is a bivariate normal distribution function and $S_1(b_1)$ is

$$S_1(b_1) = \exp(b_1^2/4\sigma^2) 2\pi \frac{\sigma^2}{\sqrt{2}} F_2(\frac{b_1}{2} - b_2, \infty, V_1) .$$

The function $S_2(b_2)$ would be (by (8.3.9))

$$S_2(b_2) = \exp(b_2^2\sigma^{-2}[1\ 2]\begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \frac{\sigma^2}{2} \begin{bmatrix} 1 \\ 2 \end{bmatrix}\sigma^{-2}/2)$$

$$2\pi \frac{\sigma^2}{\sqrt{2}} \int_{c_2}^{d_2} \int_{c_1}^{d_1} (2\pi)^{-1} |V_2^{-1}|^{-\frac{1}{2}} \exp(-\underline{t}' V_2 \underline{t}/2) dt_1 dt_2$$

with

$$\underline{c} = b_2\begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \frac{\sigma^2}{2} \begin{bmatrix} 1 \\ 2 \end{bmatrix}\sigma^{-2} + \begin{bmatrix} -\infty \\ -\infty \end{bmatrix} = \begin{bmatrix} -\infty \\ -\infty \end{bmatrix} ,$$

and

$$\underline{d} = b_2\begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \frac{\sigma^2}{2} \begin{bmatrix} 1 \\ 2 \end{bmatrix}\sigma^{-2} + \begin{bmatrix} b_1 \\ \infty \end{bmatrix} = \begin{bmatrix} b_2/2 - b_1 \\ \infty \end{bmatrix} .$$

Then the integral is a bivariate normal distribution function and $S_2(b_2)$ is

$$S_2(b_2) = \exp(3b_2^2/4\sigma^2) 2\pi \frac{\sigma^2}{\sqrt{2}} F_2(b_2/2 - b_1, \infty, V_2) .$$

Then in (8.3.20)

$$\underline{1} = \begin{bmatrix} \exp(-b_1^2/4\sigma^2) 2\pi \frac{\sigma^2}{\sqrt{2}} F_2(b_1/2 - b_2, \infty, V_1) \\ \exp(-b_2^2/4\sigma^2) 2\pi \frac{\sigma^2}{\sqrt{2}} F_2(b_2/2 - b_1, \infty, V_2) \\ 0 \end{bmatrix} .$$

Substituting $\underline{y} - \sigma^2 C'\underline{\theta}$ for $\underline{y}$, in (8.3.17) one finds

$$E(\underline{\mu}) = \underline{y} - \sigma^2 C'\underline{\theta} - H^{-1}Q^{-1}\underline{1}/P$$

$$= \underline{y} - \sigma^2 C'\underline{\theta} - H^{-1}(HV^{-1}H')\underline{1}/P$$

$$= \underline{y} - \sigma^2 C'\underline{\theta} - V^{-1}H'\underline{1}/P \ . \tag{8.3.22}$$

Recall $V^{-1} = \sigma^2 I$. Then, substituting the expressions found for $\underline{1}$ and P , one finds that

$$E(\underline{\mu}) = \underline{y} - \sigma^2 \begin{bmatrix} -1/\theta_1 \\ 1/\theta_1 - 1/\theta_2 \\ 1/\theta_2 \end{bmatrix}$$

$$- \frac{\sigma^4}{F_3(b_1, b_2, \infty, Q)\sqrt{2\pi}\ |Q^{-1}|^{+\frac{1}{2}}\sqrt{2}} \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} \exp(-b_1^2/4\sigma^2)F_2(b_1/2 - b_2, \infty, V_1) \\ \exp(-b_2^2/4\sigma^2)F_2(b_2/2 - b_1, \infty, V_2) \\ 0 \end{bmatrix} \ . \tag{8.3.23}$$

3.4 Truncated Normal Posteriors Arising From Unequal Samples From Several Populations

Suppose n_i , $i = 1, 2, \ldots, m$, independent observations are made on each of m populations, and that each population has a normal distribution with mean μ_i and variance σ_i^2 , known. Let the $j^{\underline{th}}$ observation from the $i^{\underline{th}}$ population be denoted by y_{ij} and then the joint density of the $\underline{y}$ would be denoted by

$$f(\underline{y}|\underline{\mu},\underline{\sigma}) = (\sqrt{2\pi})^{-n} \prod_{i=1}^{m} \sigma_i^{-n_i} \exp\left(-\sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij}-\mu_i)^2/(2\sigma_i^2)\right) ,$$

where

$$n = \sum_{i=1}^{m} n_i .$$

Assuming a uniform prior for $\underline{\mu}$ over a set A , and setting the prior equal to zero over the complement of A yields a posterior density which is 0 on A^c and $p(\underline{\mu}|\underline{y})$ on A :

$$p(\underline{\mu}|\underline{y}) = \frac{\exp\left(-\sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij}-\mu_i)^2/(2\sigma_i^2)\right)}{\int_A \exp\left(-\sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij}-\mu_i)^2/(2\sigma_i^2)\right) d\underline{\mu}}$$

$$= \frac{\exp\left(-\sum_{i=1}^{m} \left[\sum_{j=1}^{n_i} y_{ij}^2 - 2\mu_i \sum_{j=1}^{n_i} y_{ij} + n_i\mu_i^2\right]/(2\sigma_i^2)\right)}{\int_A \exp\left(-\sum_{i=1}^{m} \left[\sum_{j=1}^{n_i} y_{ij}^2 - 2\mu_i \sum_{j=1}^{n_i} y_{ij} + n_i\mu_i^2\right]/(2\sigma_i^2)\right) d\underline{\mu}}$$

$$= \frac{\exp(-\sum_{i=1}^{m}(-2\mu_i\bar{y}_i+\mu_i^2)n_i/(2\sigma_i^2))}{\int_A \exp(-\sum_{i=1}^{m}(-2\mu_i\bar{y}_i+\mu_i^2)n_i/(2\sigma_i^2))d\underline{\mu}} \;.$$

Completing the square, the posterior is found to be

$$p(\underline{\mu}|\underline{y}) = \frac{\exp(-(\bar{\underline{y}}-\underline{\mu})' D(\bar{\underline{y}}-\underline{\mu})/2)}{\int_A \exp(-(\bar{\underline{y}}-\underline{\mu})' D(\bar{\underline{y}}-\underline{\mu})/2)d\underline{\mu}} \;. \tag{8.4.1}$$

Here D is a diagonal matrix with elements n_i/σ_i^2 . Thus, (8.4.1) is a truncated normal posterior density.

If instead of a uniform prior, an <u>exponential</u> prior similar to (4.2.2) is assumed, the posterior density is proportional to

$$\exp(-\underline{\mu}'C\underline{\theta})f(\underline{y}|\underline{\mu},\underline{\sigma}) \;.$$

Then following the precise steps given for (8.4.1) the posterior is found to be

$$p(\underline{\mu}|\underline{y}) \propto \exp(-\sum_{i=1}^{m}[\sum_{j=1}^{n_i} y_{ij}^2-2\mu_i\sum_{j=1}^{n_i} y_{ij}+n_i\mu_i^2]/\sigma_i^2)$$

$$\exp(-(-\mu_1/\theta_1 + \mu_2(1/\theta_1 - 1/\theta_2)$$

$$+ \mu_3(1/\theta_2 - 1/\theta_3) + \dots + \mu_{m-1}(1/\theta_{m-2} - 1/\theta_{m-1}) \;.$$

$$+ \mu_m(1/\theta_{m-1} - 1/\theta_m))) \;.$$

The term $\exp(-(\sum_{i=1}^{m}\sum_{j=1}^{n_i} y_{ij}^2/\sigma_i^2)/2)$ would cancel with the same term in the normalizing constant. Then the posterior would be

$$p(\underline{\mu}|\underline{y}) \propto \exp\{-[-2\mu_1(\bar{y}_1-\sigma_1^2/(n_1\theta_1))n_1/\sigma_1^2$$

$$-\ 2\mu_2(\bar{y}_2-\sigma_2^2/(n_2\theta_2) + \sigma_2^2/(n_2\theta_1))n_2/\sigma_2^2$$

$$-\ 2\mu_3(\bar{y}_3-\sigma_3^2/(n_3\theta_3) + \sigma_3^2/(n_3\theta_2))n_3/\sigma_3^2$$

$$-\ \ldots\ -\ 2\mu_{m-1}(y_{m-1}-\sigma_{m-1}^2/(n_{m-1}\theta_{m-1})$$

$$-\ \sigma_{m-1}^2/(n_{m-1}\theta_{m-2}))n_{m-1}/\sigma_{m-1}^2$$

$$-\ 2\mu_m(\bar{y}_m + \sigma_m^2/n_m\theta_{m-1}))n_m/\sigma_m^2$$

$$+\ \sum_{i=1}^{m} n_i\mu_i^2/\sigma_i^2]/2\}\ . \tag{8.4.2}$$

Then defining the vector $\underline{z}$ to have elements $(y_1-\sigma_1^2/(n_1\theta_1))$, $(\bar{y}_2-\sigma_2^2/(n_2\theta_2) + \sigma_2^2/(n_2\theta_1))$, etc., (the $i^{\underline{th}}$ term will correspond to the term with μ_i in expression (8.4.2)). Completing the square the posterior is found to be

$$p(\underline{\mu}|\underline{y}) \propto \exp[-(\underline{z}-\underline{\mu})' D(\underline{z}-\underline{\mu})/2]$$

where D is the diagonal matrix with elements n_i/σ_i^2. This too is of the form of (8.3.3) which is a truncated normal posterior. A truncated normal prior could be handled similarly.