PARTIALLY BAYES ESTIMATES

by

Daniel L. Solomon

FSU Statistics Report M152

January, 1969
Department of Statistics
Florida State University
Tallahassee, Florida

Research supported by the Army, Navy and Air Force under Office of Naval Research Contract Number NONR 988(13), Task Order NR 042-003 with the Florida State University. Reproduction in whole or in part is permitted for any purpose of the United States Government. Also supported by training grant 5T1 GM-913 from the National Institute of General Medical Sciences.
ABSTRACT

Statistical decision problems are considered in which the decision maker is assumed to have prior information but cannot completely specify a prior distribution. The decision maker's prior knowledge is reflected in his willingness to specify a subset, \( \Lambda^* \) (called an incompleteness specification), of the class of all prior distributions \( \Lambda \). He is then recommended to select the decision rule \( \delta_0 \) to minimize the maximum over distributions in \( \Lambda^* \) of the Bayes risk. Such a rule \( \delta_0 \) is called partially Bayes with respect to \( \Lambda^* \), and reduces to a Bayes rule with respect to \( \lambda \) if \( \Lambda^* = \{ \lambda \} \) and a minimax rule if \( \Lambda^* = \Lambda \).

The problem of estimation of the mean \( \bar{\theta} \) of a p-variate random variable \( X \), with known covariance matrix \( \Sigma_1 \) is treated in detail. Suppose that \( \Lambda^* \) is the class of all prior distributions with given covariance matrix \( \Sigma_0 \) and it is only known that the mean \( \mu' \in U' \) where

\[
U' = \{ \mu' \in \mathbb{R}^p \mid |\Lambda_i - \mu_i'| \leq m_i, \ i = 1,2,\ldots,p \}
\]

and \( \Lambda_i \) and \( m_i \) are known. If the loss incurred in estimating \( \theta \) by \( \delta(x) \) when \( \bar{\theta} = \theta \) is

\[
(\delta(x) - \theta)^TK(\delta(x) - \theta),
\]

with \( K \) positive definite, then the linear (of the form \( \delta(x) = Bx + C \)) partially Bayes rule is related to the following maximization problem (where \( \bar{B} = I - B \)):
maximize $\operatorname{trace}[KB_\Pi(S \Sigma_0 + D_\Pi)]$

$\Pi \in \tilde{\mathcal{E}}$

where $\tilde{\mathcal{E}}$ is the set of all probability distributions $\Pi = (\pi_1, \pi_2, \ldots, \pi_u)$ on the set of extreme points $\{e_1, e_2, \ldots, e_u\}$ of

$$U = \{\mu \in \mathcal{E}_\Pi \mid |\mu_i| \leq m_i, \quad i = 1, 2, \ldots, p\},$$

$$B_\Pi = (\Sigma_0 + D_\Pi)(\Sigma_1 + \Sigma_0 + D_\Pi)^{-1}$$

and

$$D_\Pi = \sum_{i=1}^{u} \pi_i e_i e_i^T.$$

If the maximum occurs at $\Pi_0 \in \tilde{\mathcal{E}}$ then the rule defined by

$$\delta_0(x) = B_{\Pi_0} x + B_{\Pi_0} \Delta, \quad \Delta = (\Delta_1, \Delta_2, \ldots, \Delta_p)^T$$

is the linear partially Bayes rule with respect to $U'$. Some numerical examples are given for the case $p = 2$.

In the case when $\Sigma_0 = \sigma_0^0$, $\Sigma_1 = \sigma_1^1$, and $K$ are diagonal, the solution is

$$\delta_0(x) = B_0 x + B_0 \Delta$$

with

$$B_0 = \begin{bmatrix}
\frac{\sigma_0^0 + m_1^2}{\sigma_0^0 + \sigma_1^1 + m_1^2} \\
\frac{\sigma_1^1 + m_1^2}{\sigma_0^0 + \sigma_1^1 + m_1^2}
\end{bmatrix}.$$

For the univariate Normal case, the problem of estimation of the mean with variance unknown is treated using incomplete specification
of parameters of the natural conjugate prior distribution and the
linear partially Bayes rule is obtained.

The second major topic considered is estimation of a (univariate)
Normal scale parameter when the mean is 1) known and 2) unknown. Problem
2) is reduced to problem 1). Again it is assumed that the decision maker
is willing to use a natural conjugate prior distribution but can only
incompletely specify the parameters. The partially Bayes estimates are
determined and contours of the ratio of Bayes to partially Bayes risk
are given for various levels of: the incompleteness specification, a
parameter of the prior distribution, and the sample size.

The final topic discussed is the design problem, examples of
which are given for the estimation of (univariate) location and scale
parameters. A cost is postulated of obtaining an incompleteness specifica-
tion which increases as the specification tends toward full knowledge
of the prior distribution. The cost of sampling is assumed proportional
to the sample size, and the costs of sampling, prior specification, and
terminal action are assumed additive. The design problem consists of
selecting a sample size and an incompleteness specification to minimize
the total expected loss. The optimal sample size and incompleteness
specification for the two problems are given.

In what follows, E is the expectation operator and applies
to the joint distribution of any random variables which appear as its
argument. Conditional expectation is indicated by a vertical bar.
I. INTRODUCTION

In statistical decision problems, the decision maker may wish to choose between Bayes and minimax decision procedures. Menges [1966] asserts: "... the decision maker principally wants to adopt the Bayes type as an immediate expression of objective rationality."

The decision maker however, is often confronted with an incomplete specification of his prior distribution and thus cannot apply Bayesian techniques. This paper suggests an alternative which is other than minimax and recommends that the decision maker apply Bayes criteria as far as his prior knowledge permits. This will be made precise and some particular problems considered.

In the case of incomplete prior information, the decision maker can often improve his a priori knowledge (perhaps at some cost or utility loss) by, for example, introspection, consultation or an interviewing technique such as that due to Winkler [1967] for eliciting a prior distribution. It will be proposed that this possibility be included in the cost structure of the model (see section 1.3). To this author's knowledge, this aspect of the problem has not been investigated. Several authors have however, considered the terminal action problem for combinations of Bayes and minimax techniques, and some of these approaches will be discussed in section 1.2 after introducing some
notation. Examples of possible applications of incomplete prior techniques are given in section 1.4.

1.1 Bayes and Minimax Decision Rules

In any decision problem, the following are available to the decision maker:

1. a set $A$ of actions $a$, called the **action space**, 
2. a set $\Omega$ of states of nature $\theta$, called the **state space**, 
3. a random variable $X$ which takes values $x$, in a sample space $\mathcal{X}$ and the distribution function, $F_{X|\theta}$, which depends on $\theta$, 
4. the class $D$ of all non-randomized decision functions (rules) $\delta$, which are functions from $\mathcal{X}$ into $A$, and 
5. a loss function $L$, which is a real valued function on $\Omega \times A$.

In the special case in which $A = \Omega$, a rule $\delta \in D$ is called an **estimator** and $\delta(x) \in \Omega$, an **estimate**. (If $L$ is convex in its second argument, randomized rules need not be considered.)

Define the **risk function** $r$, a function on $\Omega \times D$ by

$$r(\theta, \delta) = E[L(\theta, \delta(X))|\theta] = \int L(\theta, \delta(x))dF_{X|\theta}(x).$$

If $\Lambda$ is the set of all probability measures $\lambda$ on $\Omega$, then $\Lambda$ will be called the class of prior distributions on $\Omega$. Finally, define the Bayes risk $r^*$, a function on $\Lambda \times D$ by

$$r^*(\lambda, \delta) = E_{\lambda} r(\theta, \delta) = \int r(\theta, \delta)d\lambda(\theta).$$

A decision function $\delta_0 \in D$ is called Bayes with respect to $\lambda \in \Lambda$ if
\[ r^*(\lambda, \delta_0) = \inf_{\delta \in D} r^*(\lambda, \delta), \]

and \( \delta_0 \in D \) is called minimax if

\[ \sup_{\theta \in \Omega} r(\theta, \delta_0) = \inf_{\delta \in D} \inf_{\theta \in \Omega} r(\theta, \delta). \]

Note that there is a relation between Bayes and minimax rules:

A minimax rule is Bayes with respect to a "least favorable" prior distribution, where \( \lambda_0 \in \Lambda \) is "least favorable" if

\[ \inf_{\delta \in D} r^*(\lambda_0, \delta) = \sup_{\lambda \in \Lambda} \inf_{\delta \in D} r^*(\lambda, \delta). \]

For a precise statement of this relationship see Lehmann [1949, 4-19].

To motivate the discussion which follows, consider the following example. Suppose a value \( x \) is observed of a random variable \( X \), which has mean \( \theta \in \Omega = (-\infty, \infty) \) and known variance \( \sigma^2 \), and an estimator \( \delta \) of \( \theta \) is sought such that \( \delta(x) = bx + c \). That is, the class of decision rules of interest is restricted to

\[ D_L = \{ \delta \in D | \delta(x) = bx + c \}. \]

Suppose further that the loss associated with this decision is \( (\delta(x) - \theta)^2 \) if \( \theta \) is the state of nature. In the absence of other information the minimax criterion might be applied; that is choose \( b_L, c_L \) such that

\[ \sup_{\theta \in \Omega} E[(b_L X + c_L - \theta)^2|\theta] = \inf_{b, c} \sup_{\theta \in \Omega} E[(bX + c - \theta)^2|\theta]. \]
It can be shown that \( b_1 = 1, c_1 = 0 \) so that \( \delta_1(x) = x \) is minimax among the class of linear rules, \( D_L \).

Consider now a situation in which the information is available that \( \theta \) is in some bounded interval, say \( |\theta - \Delta| \leq M \), where \( \Delta \) and \( M \) are known, \( M \geq 0 \). How should this information be used to modify the estimate \( \delta(x) \)? With the minimax criterion the maximization is restricted to the set \( \{ \theta \in \Omega \mid |\theta - \Delta| \leq M \} \). For \( \delta \in D_L \) one can show that the minimax rule is now

\[
\delta_2(x) = b_2x + c_2
\]

where

\[
(1) \quad b_2 = \frac{M^2}{\sigma^2 + M^2}, \quad c_2 = \frac{\sigma^2 \Delta}{\sigma^2 + M^2} = (1 - b_2)\Delta.
\]

Now the risk in the minimax rule \( \delta_1(x) = x \) is

\[
r(\theta, \delta) = E[(X - \theta)^2 \mid \theta] = \sigma^2,
\]

while from (1) and the fact that \( |\theta - \Delta| \leq M \)

\[
r(\theta, \delta_2) = E[(\delta_2(x) - \theta)^2 \mid \theta]
\]

\[
= E[(b_2x + c_2 - \theta)^2 \mid \theta]
\]

\[
= b_2^2 \sigma^2 + (1 - b_2)^2(\Delta - \theta)^2
\]

\[
= \left( \frac{M^2}{\sigma^2 + M^2} \right)^2 \sigma^2 + \left( \frac{\sigma^2}{\sigma^2 + M^2} \right)^2 (\Delta - \theta)^2
\]

\[
\leq \frac{M^2 \sigma^2}{\sigma^2 + M^2} \leq \sigma^2 = r(\theta, \delta_1),
\]
with strict inequality if $M < \infty$ and $\sigma^2 \neq 0$. Thus if it is known that $|\theta - \Delta| \leq M$, the risk in $\delta_2$ is no larger than that in $\delta_1$, and is smaller in most situations.

Assume a prior distribution, $\lambda$, is available for $\theta$ of the previous example with

$$E_\lambda \theta = \mu \quad \text{and} \quad \text{var}_\lambda \theta = \sigma_0^2.$$

The linear Bayes estimate of $\theta$ with respect to $\lambda$ is

$$\delta_0(x) = \frac{\frac{X}{\sigma^2} + \frac{\mu}{\sigma_0^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}},$$

with Bayes risk

$$\left[ \frac{1}{\sigma^2} + \frac{1}{\sigma_0^2} \right]^{-1}.$$

On the other hand, suppose that the prior information is incomplete, and the decision maker knows $\sigma^2$ but only that $|\Delta - \mu| \leq M$. Then it is reasonable to choose a linear decision rule $\delta_3$ such that

$$\sup_\lambda E_\lambda [E[(\delta_3(x) - \theta)^2|\theta]] = \inf_{\delta \in D_L} \sup_\lambda E_\lambda [E[(\delta(x) - \theta)^2|\theta]]$$

where the supremum extends over all $\lambda$ for which $|\Delta - \mu| \leq M$. It can be shown (corollary of section 3.1) that
\[
\delta_3(x) = \frac{x + \frac{\beta}{\sigma_0^2 + \nu^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2 + \nu^2}}
\]

and that the risk of \( \delta_3 \) is

\[
\sup_{\lambda} E_\lambda \{ E[ (\delta_3(x) - \theta)^2 | \theta] \} = \left[ \frac{1}{\sigma^2} + \frac{1}{\sigma_0^2 + \nu^2} \right] \leq \sigma^2.
\]

Defining

\[
r_0 = E_\lambda r(\theta, \delta_0) = \left[ \frac{1}{\sigma^2} + \frac{1}{\sigma_0^2} \right]^{-1},
\]

\[
r_1 = r(\theta, \delta_1) = \sigma^2,
\]

\[
r_2 = \text{supremum} \ r(\theta, \delta_2) = \left[ \frac{1}{\sigma^2} + \frac{1}{\nu^2} \right]^{-1},
\]

\[
r_3 = \text{supremum} \ E_\lambda r(\theta, \delta_3) = \left[ \frac{1}{\sigma^2} + \frac{1}{\sigma_0^2 + \nu^2} \right]^{-1},
\]

it can be shown that

\[
r_0 \leq r_2 \leq r_3 \leq r_1 \quad \text{if} \quad \sigma_0^2 \leq \nu^2
\]

while

\[
r_2 \leq r_0 \leq r_3 \leq r_1 \quad \text{if} \quad \sigma_0^2 \geq \nu^2.
\]

### 1.2 Incomplete Prior Techniques: Literature

In this section, several approaches to the use of partial prior information are discussed.

Suppose that the decision maker's prior information takes the form of a best guess \( \theta_0 \) of a parameter \( \theta \) which is to be estimated.
Thompson [1968] suggests "shrinking" the usual estimator \( \hat{\theta} \) (maximum likelihood estimate, minimum variance unbiased estimate, etc.) toward \( \theta_0 \) by using an estimate of the form

\[
\hat{\theta}_s = c(\hat{\theta} - \theta_0) + \theta_0 = c\hat{\theta} + (1 - c)\theta_0.
\]

Suppose \( \theta_0 = 0 \) so that \( \hat{\theta}_s = c\hat{\theta} \). Choose \( c \) to minimize \( E(c\hat{\theta} - \theta)^2 \). In general \( c = c(\theta) \), that is, involves the unknown \( \theta \), so replace \( \theta \) by \( \hat{\theta} \) in \( c(\theta) \). For example, suppose \( X \) has the Normal distribution with mean \( \theta \) and variance \( \sigma^2/n \), and \( \theta \) is to be estimated. Then \( E(c\bar{X} - \theta)^2 \) is minimized for

\[
c = \frac{\sigma^2}{\sigma^2 + n^{-1} \theta^2}.
\]

Thompson's estimator becomes

\[
\hat{\theta}_s = \frac{\bar{X}^2}{\bar{X}^2 + n^{-1} \sigma^2} \bar{X}
\]

if \( \sigma^2 \) is known and

\[
\frac{\bar{X}^2}{\bar{X}^2 + n^{-1} \sigma^2} \bar{X}
\]

if \( \sigma^2 \) is not known.

Stone [1963] suggests restricting the class of decision rules in such a way that only approximate prior information is required. Such procedures are termed non-ideal. If \( P_s \) is the "supposed" or approximate prior distribution and \( P_a \) the "actual" prior distribution, then Stone defines the robustness of the non-ideal procedure by

\[
u(P_a \mid P_a) - u(P_s \mid P_a), \]

the difference in mean utility between the Bayes
and non-ideal procedures when evaluated by the "actual" prior. He considers the examples of estimating a general mean and a Normal variance and obtains inequalities on the moments of $P_s$ and $P_a$ under which the mean square error of the non-ideal procedure does not exceed that of the classical (sample mean, sample variance) procedure. The non-ideal rule is chosen from the class of rules which are linear in the sufficient statistic.

Now consider a criterion for combining Bayes and minimax procedures due to Menges [1966]. The criterion is adopted for the analysis of this paper. For a given decision maker, assume he has a prior distribution, $\lambda$, but complete specification of $\lambda$ is unavailable to him. Assume that he does however have some prior information reflected in his willingness to assert that $\lambda \in \Lambda^* \subset \Lambda$. Menges calls $\Lambda^*$ the space of uncertainty, here called an incompleteness specification.

Menges recommends that the decision maker is to choose $\delta_0 \in D$ such that

$$(1.2.1) \quad \sup_{\lambda \in \Lambda^*} r^*(\lambda, \delta_0) = \inf_{\delta \in D} \sup_{\lambda \in \Lambda^*} r^*(\lambda, \delta).$$

He calls such a rule $\delta_0$, an extended Bayes rule, here called a partially Bayes rule with respect to the incompleteness specification $\Lambda^*$. The Bayesian character of the criterion is in the minimization of the Bayes risk $r^*$, while its minimax nature is in choosing $\delta_0 \in D$ so that the maximum Bayes risk is as small as possible regardless of the prior distribution $\lambda \in \Lambda^*$. Should $\Lambda^*$ contain only one distribution $\lambda$, then a partially Bayes rule is a Bayes rule with respect to that
\( \lambda \). On the other hand, if \( \Lambda^\infty = \Lambda \), the class of all prior distributions, then the maximum is attained at that \( \lambda \) which assigns probability one to the values of \( \theta \) which maximize the risk, \( r \). Thus

\[
\sup_{\lambda \in \Lambda} r^*(\lambda, \delta) = \sup_{\theta \in \Omega} r(\theta, \delta),
\]

and a partially Bayes rule is a minimax rule.

Menges uses the specification \( \Lambda^\infty \) as follows: Let \( \{\Omega_1, \Omega_2, \ldots, \Omega_k\} \) be a partition of the state space \( \Omega \). If the decision maker knows that \( \Omega_i \) occurs with probability \( p_i \), where \( \sum_{i=1}^{k} p_i = 1 \), then \( \Lambda^\infty \) is taken to be the class of prior distributions \( \lambda \) such that \( p_\lambda(\Omega_i) = p_i \), \( i = 1, 2, \ldots, k \).

A partially Bayes solution then reduces to choosing \( \delta_0 \in D \) so that

\[
\sum_{i=1}^{k} p_i \sup_{\theta \in \Omega_i} r(\theta, \delta_0) = \inf_{\delta \in D} \sum_{i=1}^{k} p_i \sup_{\theta \in \Omega_i} r(\theta, \delta).
\]

It is clear that this procedure gives a minimax rule when \( k = 1 \).

Other approaches to combining Bayes and minimax procedures are given by Hodges and Lehmann [1952], Skibinsky and Cote [1963], Schneeweiss [1964], and Blum and Rosenblatt [1967]. This paper applies Menges' approach to the particular problems of estimation of location and scale parameters. A formal statement of the first of these problems follows.

1.3 Statement of the Location Problem

Suppose that \( X \) and \( \bar{Y} \) are \( (p \times 1) \) vector random variables (not necessarily Normal), that \( \mu' = (\mu_1', \mu_2', \ldots, \mu_p')^T \) and \( \Delta = (\Delta_1, \Delta_2, \ldots, \Delta_p)^T \).
are \((p \times 1)\) vectors of constants, \(\theta = (m_1, m_2, \ldots, m_p)^T\) is a \((p \times 1)\) vector of non-negative constants, and that \(\Sigma_0 = (\sigma_{ij}^0)\), \(\Sigma_1 = (\sigma_{ij}^1)\) and \(K = (k_{ij})\) are \((p \times p)\) matrices of constants. Assume further that the distribution of \(X\) is such that

\[
E(X|\theta) = \theta \quad \text{and} \quad \text{Cov}(X|\theta) = \Sigma_1
\]

where \(\Sigma_1\) is known.

Define the spaces \(\Omega = \mathcal{X} = \mathbb{R}^p\) (\(p\)-dimensional Euclidean space) and \(D\) as in section 1.1 but now require that the action space \(A\), coincide with the state space \(\Omega\). Thus a decision function \(\delta \in D\), is a function from the sample space into the state space, that is, an "estimator" of the location "parameter" \(\theta\). Suppose that if \(X\) is observed to have the value \(x\), and \(\theta\) is estimated by \(\delta(x)\), when \(\theta = \theta\), then the loss incurred is

\[
(\delta(x) - \theta)^T K (\delta(x) - \theta).
\]

In Menge's examples, \(\mathcal{I}^k\), the incompleteness specification, is characterized by attaching probabilities to the elements of a partition of \(\Omega\); here \(\mathcal{I}^k\) is characterized as follows: Suppose that the decision maker's prior knowledge is such that his prior distribution \(\lambda\) on \(\Omega\) has

\[
(1.3.1) \quad E_{\lambda}\theta = \mu' \quad \text{and} \quad \text{Cov}_{\lambda}\theta = \Sigma_0,
\]

where \(\Sigma_0\) is known, and that he has learned at cost \(\alpha(M)\) that

\[
(1.3.2) \quad \mu' \in U' = \{\mu' \in \mathbb{R}^p | |A - \mu'| \leq M\},
\]
where for $p$-vectors $Y$ and $Z$, $|Y| \leq Z$ means $|y_i| \leq z_i$, $i = 1, 2, \ldots, p$.

Then take for $\Lambda^\circ$, that class of prior distributions which satisfy (1.3.1) with $\mu' \in U'$. In section 4.1 it is noted that the results of chapter II apply for a more general formulation of $U'$. In particular, the results hold whenever $U'$ is the convex hull of a finite number of points. We choose to work with the "box" for simplicity of presentation.

Let $c_s(x)$ be the sampling cost associated with observing $x$, and assume that all costs are additive so that the overall loss function for the problem is

$$L(\Delta, \hat{\mu}, \delta, x, \theta) = c(\hat{\mu}) + c_s(x) + (\delta(x) - \theta)^T K(\delta(x) - \theta).$$

$\Delta$ is an argument of the loss function since the decision rule $\delta$, may depend on $\Delta$.

We make some definitions and find an optimality criterion for the estimator.

For $\delta \in D^\circ \subset D$, define the Bayes risk of $\delta$ for given $\mu' \in U'$ by

$$B(\mu', \delta) = B(\mu', \delta, \hat{\mu}, \delta, D^\circ) = E \{ E[(\delta(x) - \hat{\delta})^T K(\delta(x) - \hat{\delta})] \},$$

the maximum Bayes risk of $\delta$ by

$$B(\delta) = \sup_{\mu' \in U'} B(\mu', \delta),$$

and the partially Bayes risk by

$$B^* (U') = \inf_{\delta \in D^\circ} B(\delta).$$
Any rule \( \delta_0 \in D^\infty \) for which

\[
B(\delta_0) = B^*(U'),
\]

or equivalently for which

\[
\sup_{\mu'} B(\mu', \delta_0) = \inf_{\delta \in D^\infty} \sup_{\mu' \in U'} B(\mu', \delta),
\]

is then said to be partially Bayes (with respect to the incompleteness specification \( U' \)) in \( D^\infty \). The incompleteness specification is omitted if it is clear from the context.

Thus for each \( \delta \in D^\infty \), determine a worst (in terms of Bayes risk) \( \mu' \) and choose a \( \delta_0 \) for which the maximum risk is a minimum.

Let us illustrate how this criterion is related to the usual minimax criterion.

\[
B^*(U') = \inf_{\delta \in D^\infty} \sup_{\mu' \in U'} \mathbb{E}[\mathbb{E}[(\delta(X) - \overline{\delta})^T K(\delta(X) - \overline{\delta}) | \overline{\delta}]]
\]

\[
= \inf_{\delta \in D^\infty} \sup_{\mu' \in U'} \mathbb{E}[\mathbb{E}[(\delta(X) - \overline{\delta})^T K(\delta(X) - \overline{\delta}) | X, \mu']] \]

\[
= \inf_{\delta \in D^\infty} \sup_{\mu' \in U'} \mathbb{E} C(\mu', \delta(X)),
\]

where

\[
C(\mu', \delta(X)) = \mathbb{E}[(\delta(X) - \overline{\delta})^T K(\delta(X) - \overline{\delta}) | X, \mu'].
\]

This is the minimax criterion for loss function \( C \) and state space \( U' \).

### 1.4 Applications

It is perhaps difficult to conceive of a situation in which the experimenter would know \( \Sigma_0 \) and \( \Sigma_1 \) but neither \( \mu' \) nor \( \theta \). The
object of this study, however, is to develop techniques which may be useful in more realistic problems as well as to solve the particular problem just formulated. Results for several univariate problems are obtained in chapters IV and V and two hypothetical applications of the location problem just described are outlined now.

The National Bureau of Standards "standard meter" bar has actual length \( \mu' \), say. Occasionally, a batch of prototypes is made with length \( \bar{\theta} \sim N(\mu', \sigma_0^2) \). These prototypes have been made for years and so \( \sigma_0^2 \) can be assumed known. The prototypes are used to manufacture meter sticks of length \( X \sim N(\theta, \sigma_1^2) \). For a particular manufacturer, \( \sigma_1^2 \), being a characteristic of his production process, can also be assumed known.

It has been observed that \( \mu' \) varies unexplainedly in time. By costly spectroscopic techniques, however, bounds can be placed on \( \mu' \), say \( \mu' \in U' = \{ \mu' \mid |\Delta - \mu'| \leq M \} \). The manufacturer wishes to estimate the \( \theta \) for his prototype. His class of prior distributions for \( \bar{\theta} \) is then \( N(\mu', \sigma_0^2), \mu' \in U' \).

As a second application and in a different context, suppose that observations of a particular individual's serum cholesterol at a given time are \( X \sim N(\theta, \sigma_1^2) \). The parameter \( \theta \in \Omega \) varies from person to person and suppose that \( \bar{\theta} \sim N(\mu', \sigma_0^2) \), where \( \mu' \) is the mean cholesterol level for the population at large. Now at various costs, an individual's age, height, weight, diet, medical history, etc., can be determined and thus the individual can be placed in a sub-population. If it is now supposed that for this sub-population it is known that the mean cholesterol level \( \mu' \) is in some set \( U' \), then to estimate \( \theta \), the class of prior distributions is \( N(\mu', \sigma_0^2), \mu' \in U' \).
The matrix and convex function theory required in the sequel is in Rao [1965, chapter I] and Blackwell and Girshick [1954, section 2.2], respectively. Some additional required results concerning convex functions are given in the appendix and labelled A.1 and A.2. No claim to originality of these results is made, but they are stated and proved for the convenience of the reader.
II. THE GENERAL LOCATION PROBLEM

For the problem of section 1.3, let $D_L$ be that subset of the space $D$, of all decision rules defined by

$$D_L = \{\delta: x \rightarrow \Omega | \delta(x) = Bx + C\}$$

where $B (p \times p)$ and $C (p \times 1)$ are matrices of real constants which may depend on $\Delta$ and $\mathcal{M}$, that is, $D_L$ is the class of linear functions from $E^p$ into $E^p$. A partially Bayes rule in $D_L$ is a rule

$$\delta_0(x) = B_0x + C_0$$

such that

$$\sup_{\mu' \in U'} B(\mu', \delta_0) = \inf_{\delta \in D_L} \sup_{\mu' \in U'} B(\mu', \delta),$$

see (1.3.3).

It is assumed that: $\Sigma_0$ and $\Sigma_1$ are symmetric positive definite, and hence the corresponding distributions are non-degenerate. The matrix $K$, is symmetric positive definite and hence the loss is positive whenever $\delta(x) \neq \theta$ and zero if $\delta(x) = \theta$.

2.1 Game Theoretic Formulation

First, the Bayes risk of $\delta$ for given $\mu' \in U'$ is computed.

**Lemma 2.1.1:** If $X$ is a $p$-variate random variable with $E(X) = \theta$ and $\text{Cov}(X) = \Sigma$, and if $C$ is a $(p \times p)$ matrix of constants, then
\[ E[(X - \theta)^T G(X - \theta)] = \text{trace}(G\Sigma). \]

**Proof:** With \( \Sigma = (\sigma_{ij}) \) and \( G = (g_{ij}) \)

\[
E[(X - \theta)^T G(X - \theta)] = \sum_{i=1}^{p} \sum_{j=1}^{p} g_{ij} \sigma_{ji} = \text{tr}(G\Sigma).
\]

Theorem 2.3.2 asserts that in the class \( D_L \) of linear decision rules, any partially Bayes rule is in \( L \subset D_L \) defined by

\[(2.1.1) \quad L = \{ \delta \in D_L | \delta(x) = Bx + \bar{\theta} \Delta \},
\]

where

\[ \bar{B} = (I - B). \]

**In sections 2.1 and 2.2, only rules in L are considered.**

**Theorem 2.1.2:** If \( \delta \in L \), then the Bayes risk of \( \delta \) for given \( \mu' \) is

\[(2.1.2) \quad B(\mu', \delta) = \text{tr}B^T KB\Sigma L + \text{tr}B^T K\bar{B}\Sigma O + (\Delta - \mu')^T B^T KB(\Delta - \mu'). \]

**Proof:**

\[
B(\mu', \delta) = E(\delta(X) - \bar{\delta})^T K(\delta(X) - \bar{\delta}) = E(BX + \bar{\theta} \Delta - \bar{\delta})^T K(BX + \bar{\theta} \Delta - \bar{\delta})
\]

\[
= E[B(X - \bar{\delta}) + \bar{\theta} (\Delta - \bar{\delta})]^T K[B(X - \bar{\delta}) + \bar{\theta} (\Delta - \bar{\delta})]
\]

\[
= E(X - \bar{\delta})^T B^T KB(X - \bar{\delta}) + E(\Delta - \bar{\delta})^T B^T KB(\Delta - \bar{\delta})
\]

\[
= \text{tr} B^T KB\Sigma L + E[(\Delta - \mu') + (\mu' - \bar{\delta})] B^T KB[(\Delta - \mu') + (\mu' - \bar{\delta})]
\]

\[
= \text{tr} B^T KB\Sigma L + \text{tr} B^T KB\Sigma O + (\Delta - \mu')^T B^T KB(\Delta - \mu'). \quad \text{Q.E.D.}
\]

Write

\[ \mu = \Delta - \mu'. \]
and note that, since $\Delta$ is fixed, $\delta$ is determined by $B$ and $\mu'$ is determined by $\mu$. Thus identify

$$R(\mu, B) = B(\mu', \delta),$$

and (2.1.2) becomes

$$(2.1.3) \quad R(\mu, B) = \text{tr} \ B^T K B \Sigma_1^{-1} + \text{tr} \ \Sigma_0^{-1} B^T K B \Sigma_0 + \mu B^T K B \mu.$$

Similarly, the set $U'$ of (1.3.2) becomes

$$U = \{ \mu \in E^P \mid |\mu| \leq M \}.$$

By definition, if $\delta_0(x) = B_0 x + \bar{B}_0 \Delta$ is a partially Bayes rule in $L$, then with $E^P$ the space of real $p \times p$ matrices ($= E^{P^2}$),

$$\sup_{\mu \in U} R(\mu, B_0) = \inf_{B \in U} \sup_{\mu \in U} R(\mu, B),$$

in which case, if there is no risk of confusion, $B_0$ will be called partially Bayes in $L$.

For reasons made explicit following theorem 2.1.6 we wish to show that a minimax theorem holds. To do so, it will first be shown that in searching for partially Bayes rules, attention may be restricted to sets $\beta \subset \bigcup_{\mathcal{P}} \Sigma$ and $\beta \subset U$ for which the following minimax theorem applies. See Stein [1963, unpublished, pp.1.3.1-1.3.2].

**Theorem:** (A Minimax Theorem) Let $\mathcal{E}$ be a finite set, $\beta$ an arbitrary convex set, and $R$ a bounded real valued function on $\mathcal{E} \times \beta$ which is convex in its second argument. Let $\tilde{\mathcal{E}}$ be the set of all probability functions on $\mathcal{E}$ and extend the definition of $R$ to $\tilde{\mathcal{E}} \times \beta$ by
(2.1.4) \[ R(\Pi, B) = \sum_{e \in E} R(e, B) \Pi(e). \]

Then

\[ \sup_{\Pi \in \mathcal{F}} \inf_{B \in \mathcal{B}} R(\Pi, B) = \inf_{B \in \mathcal{B}} \sup_{\Pi \in \mathcal{F}} R(\Pi, B). \]

From (2.1.3) note that \( R(\mu, I) = \text{tr } K_{\mu} I = \tau \) say. Define

(2.1.5) \[ \mathcal{B} = \{ B \in \mathcal{B} | \sup_{\mu \in U} R(\mu, B) \leq \tau \}. \]

Now \( I \in \mathcal{B} \) so \( \mathcal{B} \) is not empty.

**Theorem 2.1.3:** If \( B^x \notin \mathcal{B} \), then \( B^{x^*} \) is not partially Bayes in \( L \).

**Proof:** Since \( B^x \notin \mathcal{B} \), there exists a \( \mu^x \in U \) such that \( R(\mu^x, B^x) > \tau \).

Thus

\[ \inf_{B \in \mathcal{B}} \sup_{\mu \in U} R(\mu, B) \leq \sup_{\mu \in U} R(\mu, I) = \tau < R(\mu^x, B^x) \leq \sup_{\mu \in U} R(\mu, B^{x^*}). \]

**Corollary 2.1.4:**

\[ \inf_{B \in \mathcal{B}} \sup_{\mu \in U} R(\mu, B) = \inf_{B \in \mathcal{B}} \sup_{\mu \in U} R(\mu, B). \]

Thus only \( B \in \mathcal{B} \) need be considered.

Let \( \mathcal{C} \subset U \) be the collection of extreme points \( e_i \) of the convex, compact set \( U \).

That is

\[ \mathcal{C} = \{ e_1, e_2, \ldots, e_u \} = \{ \mu \in E^p | \| \mu \| = M \}. \]

Note that the number of elements in \( \mathcal{C} \) is \( u \leq 2^p \), with equality if all \( m_j > 0 \). By application of corollary A.2, observe that for any \( B \), since \( B^T K_B B \) is non-negative definite,
\[
\sup_{\mu \in \mathcal{U}} \mu^T B^T K \mu = \sup_{e^T e \leq 2} e^T B^T K e
\]
and thus from (2.13)

(2.1.6) \[
\sup_{\mu \in \mathcal{U}} R(\mu, B) = \sup_{e^T e \leq 2} R(e, B).
\]

**Lemma 2.1.5:** The set \( \mathcal{B} \) is convex, and for each \( e \in \mathcal{C} \), \( R(e, B) \) is convex in \( B \).

**Proof:** Given \( B_0, B_1 \in \mathcal{B} \) and \( 0 < \alpha < 1 \), let \( B_{\alpha} = (1 - \alpha)B_0 + \alpha B_1 \). It must be shown that for each \( e \in \mathcal{C} \)

\[
R(e, B_{\alpha}) \leq (1 - \alpha)R(e, B_0) + \alpha R(e, B_1).
\]

It follows from (2.1.3) and properties of the trace that

\[
R(e, B) = \text{tr} B^T K B e + \text{tr} B_0^T K B_0 e + e^T e.
\]

We wish to show that \( T \leq 0 \) where

\[
T = R(e, B_{\alpha}) - (1 - \alpha)R(e, B_0) - \alpha R(e, B_1).
\]

Let \( \Sigma = \Sigma_0 + e e^T \). Then it can be shown that

\[
T = \text{tr}[B_{\alpha}^T K B_{\alpha} \Sigma_1 + B_0^T K B_0 \Sigma] - (1 - \alpha)\text{tr}[B_0^T K B_0 \Sigma_1 + B_0^T K B_0 \Sigma] - \alpha \text{tr}[B_1^T K B_1 \Sigma_1 + B_1^T K B_1 \Sigma]
\]
\[
= -\alpha(1 - \alpha)\text{tr}(B_0 - B_1)^T K (B_0 - B_1)(\Sigma_1 + \Sigma).
\]

It will suffice to show that

\[
S = \text{tr}[(B_0 - B_1)^T K (B_0 - B_1)](\Sigma_1 + \Sigma) \geq 0.
\]
Now $\Sigma_1 + \Sigma = \Sigma_1 + \Sigma_0 + ee^T$ is symmetric positive definite, 
$(\beta_0 - B_1)^T K (\beta_0 - B_1)$ is symmetric non-negative definite, and thus all the characteristic roots of each are non-negative. See Rao [1965, p.35]. It follows from a theorem of Anderson and Gupta [1963, corollary 2.2.1] that all the characteristic roots of their product are non-negative. Since the trace of a matrix is the sum of its characteristic roots (Rao [1965, problem 1.8, p.54]), $S \geq 0$. Hence for all $e \in \mathbb{S}$

\[(2.1.7)\] \[R(e, B^\alpha) \leq (1 - \alpha)R(e, B_0) + \alpha R(e, B_1).\]

It remains to show that $B^\alpha \in \beta$, that is $\beta$ is convex. By

\[(2.1.6)\]

\[\beta = \{B \in \mathbb{S} \mid \sup_{\mu \in U} R(\mu, B) \leq \tau\} = \{B \in \mathbb{S} \mid \sup_{e \in \mathbb{S}} R(e, B) \leq \tau\}.

By (2.1.7) with $B_0, B_1 \in \beta$, it follows that

\[
\sup_{e \in \mathbb{S}} R(e, B^\alpha) \leq \sup_{e \in \mathbb{S}} [(1 - \alpha)R(e, B_0) + \alpha R(e, B_1)]
\]

\[
\leq (1 - \alpha) \sup_{e \in \mathbb{S}} R(e, B_0) + \alpha \sup_{e \in \mathbb{S}} R(e, B_1)
\]

\[
\leq (1 - \alpha) \tau + \alpha \tau = \tau.
\]

That is, $B^\alpha \in \beta$. Thus, the lemma.

Also, for all $e \in \mathbb{S}$, $B \in \beta$,

\[(2.1.8)\]

\[R(e, B) \leq \tau < \infty\]

and so $R(\cdot, \cdot)$ is bounded on $\mathbb{S} \times \beta$. 
Define \( \mathcal{Z} \) to be the convex set of all probability distributions on \( \mathcal{F} \):

\[
(2.1.9) \quad \mathcal{Z} = \left\{ \pi = (\pi_1, \pi_2, \ldots, \pi_u) \in \mathbb{R}^u \mid \pi_i \geq 0, \ i = 1, 2, \ldots, u; \sum_{i=1}^{u} \pi_i = 1 \right\},
\]

and extend the definition of \( R \) to \( \mathcal{Z} \times \beta \) by

\[
R(\pi, B) = \sum_{i=1}^{u} R(e_i, B) \pi_i.
\]

By definition, \( \mathcal{Z} \) is finite, by (2.1.8) \( R \) is bounded on \( \mathcal{Z} \times \beta \), and by lemma 2.1.5, \( R \) is convex in its second argument. The hypotheses of the minimax theorem are thus satisfied and so

**Theorem 2.1.6:**

\[
\sup_{\pi \in \mathcal{Z}} \inf_{B \in \beta} R(\pi, B) = \inf_{B \in \beta} \sup_{\pi \in \mathcal{Z}} R(\pi, B).
\]

**Theorem 2.1.6** guarantees the existence of a saddle point of \( R \), that is, a pair \((\pi_0, B_0)\) \( \in \mathcal{Z} \times \beta \) such that for all \((\pi, B) \in \mathcal{Z} \times \beta \),

\[
R(\pi, B_0) \leq R(\pi_0, B_0) \leq R(\pi_0, B).
\]

See Karlin [1959, Vol.II, p.9]. The theorem thus guarantees the existence of a rule \( B_0 \), which is partially Bayes (with respect to \( \mathcal{Z} \)) in \( \beta \); that is, such that (see display (1.2.1) of section 1.2)

\[
\sup_{\pi \in \mathcal{Z}} R(\pi, B_0) = \inf_{B \in \beta} \sup_{\pi \in \mathcal{Z}} R(\pi, B).
\]

**Lemma 2.1.7** shows that for any \( B \in \mathcal{U}_p' \),
\[ \sup_{\Pi \in \mathcal{B}} R(\Pi, \mathcal{B}) = \sup_{\mu \in \mathcal{U}} R(\mu, \mathcal{B}), \]

so that

\[ \sup_{\mu \in \mathcal{U}} R(\mu, \mathcal{B}_0) = \inf_{\mathcal{B} \in \mathcal{B}_0} \sup_{\mu \in \mathcal{U}} R(\mu, \mathcal{B}), \]

and \( \mathcal{B}_0 \) is partially Bayes (with respect to \( \mathcal{U} \)) in \( \mathcal{B} \). Finally, by corollary 2.1.4,

\[ \sup_{\mu \in \mathcal{U}} R(\mu, \mathcal{B}_0) = \inf_{\mathcal{B} \in \mathcal{B}_0} \sup_{\mu \in \mathcal{U}} R(\mu, \mathcal{B}), \]

that is, \( \mathcal{B}_0 \) is partially Bayes (with respect to \( \mathcal{U} \)) in \( \mathcal{L} \).

Note too that theorem 2.1.6 simplifies the calculation of the partially Bayes rule, for to compute the partially Bayes rule directly is to seek \( \mathcal{B}_0 \) such that

\[ \sup_{\mu \in \mathcal{U}} R(\mu, \mathcal{B}_0) = \inf_{\mathcal{B} \in \mathcal{B}_0} \sup_{\mu \in \mathcal{U}} R(\mu, \mathcal{B}). \]

For each \( \mathcal{B} \in \mathcal{B}_0 \), \( R(\cdot, \mathcal{B}) \) is convex and continuous on \( \mathcal{U} \), and \( \mathcal{U} \) is compact and convex. Thus by theorem A.1, for each \( \mathcal{B} \in \mathcal{B}_0 \), the supremum, \( \sup_{\mu \in \mathcal{U}} R(\mu, \mathcal{B}) \) is attained at an extreme point, \( e(\mathcal{B}) \) of \( \mathcal{U} \); that is, at an element of \( \mathcal{E} = \{e_1, e_2, \ldots, e_u\} \).

Let

\[ \beta_j = \{\mathcal{B} \in \mathcal{B}_0 | R(e_j, \mathcal{B}) = \sup_{\mathcal{E}} R(e, \mathcal{B})\}, \]

so that \( \mathcal{B}_0 = \bigcup_{j=1}^{u} \beta_j \). Now determine \( \inf_{\mathcal{B} \in \beta_j} R(e_j, \mathcal{B}) \) and then \( \inf_{1 \leq j \leq u} \inf_{\mathcal{B} \in \beta_j} R(e_j, \mathcal{B}) \) to obtain the risk of the partially Bayes rule. But the minimization over \( \beta_j \) is a constrained minimization, namely
minimize \( R(e_j, B) \)

subject to

\[ R(e_j, B) = \sup_{e \in \mathcal{E}} R(e, B) = R(e(B), B). \]

The constraint is complicated in that the relation between \( e(B) \) and \( B \) is not a simple one.

As will be seen (Lemma 2.2.1), Theorem 2.1.6 allows the minimization (for each \( \Pi \in \mathcal{P} \)) of \( R(\Pi, B) \) over all of \( \chi_p \). This minimization is unconstrained and affords an analytical solution for the minimum by standard techniques of the calculus. Then it remains only to maximize \( \inf_{\Pi, B} R(\Pi, B) \) over the convex, compact set \( \mathcal{C} \); a task which is accomplished numerically with relative ease. Some examples of this computation are given in section 2.4 for the case \( p = 2 \).

We now return to the derivation of the partially Bayes rule.

**Lemma 2.1.7:** For each \( B \in \mathcal{P} \),

\[ \sup_{\mu \in U} R(\mu, B) = \sup_{e \in \mathcal{E}} R(e, B) = \sup_{\Pi \in \mathcal{C}} R(\Pi, B). \]

**Proof:** Recall that \( \mathcal{C} = \{ e_1, e_2, \ldots, e_u \} \). Since the distribution \( \pi_i \), which assigns probability one to \( e_i \), is in \( \mathcal{C} \) for \( i = 1, 2, \ldots, u \),

\[ \sup_{e \in \mathcal{E}} R(e, B) \leq \sup_{\Pi \in \mathcal{C}} R(\Pi, B). \]

Also, for any \( \Pi \in \mathcal{C} \),

\[ R(\Pi, B) = \sum_{i=1}^{u} R(e_i, B) \pi_i \leq \sum_{i=1}^{u} \sup_{e \in \mathcal{E}} R(e, B) \pi_i = \sup_{e \in \mathcal{C}} R(e, B), \]
and thus
\[
\sup_{\Pi \in \mathcal{F}} R(\Pi, B) \leq \sup_{e \in \mathcal{E}} R(e, B).
\]
The first equality is (2.1.6).

2.2 Reduction to a Mathematical Programming Problem

In this section the search for a partially Bayes rule will be reduced to a problem of maximizing a continuous function on a compact, convex set -- a problem which can be solved by numerical techniques.

Recall that the objective is to find a matrix $B_0$ such that
\[
\sup_{\mu \in U} R(\mu, B_0) = \inf_{B \in \mathcal{B}} \sup_{\mu \in U} R(\mu, B).
\]

**Lemma 2.2.1:** The partially Bayes risk (in $L$) satisfies
\[
\inf_{B \in \mathcal{B}} \sup_{\mu \in U} R(\mu, B) = \sup_{\Pi \in \mathcal{F}} \inf_{B \in \mathcal{B}} R(\Pi, B).
\]

**Proof:**
\[
\begin{align*}
\inf_{B \in \mathcal{B}} \sup_{\mu \in U} R(\mu, B) &= \inf_{B \in \mathcal{B}} \sup_{\mu \in U} R(\mu, B) \\
&= \inf_{B \in \mathcal{B}} \sup_{\Pi \in \mathcal{F}} R(\Pi, B) \\
&= \sup_{\Pi \in \mathcal{F}} \inf_{B \in \mathcal{B}} R(\Pi, B)
\end{align*}
\]
by corollary 2.1.4

by lemma 2.1.7

(2.2.1)

by theorem 2.1.6.

Suppose there exists a $B^* \in \mathcal{B}$ such that...
(2.2.2) \[ \sup_\pi \inf_{\beta} R(\pi, \beta) = \sup_\pi R(\pi, B^\pi). \]

By theorem 2.1.6, there exists a partially Bayes rule \( B_0 \in \beta \) so

\[ \sup_{\mu \in \mathcal{U}} R(\mu, B^\pi) = \sup_\pi R(\pi, B^\pi) \]

by lemma 2.1.7

= \sup_{\pi \in \beta} \inf_{\beta} R(\pi, \beta) \quad \text{by (2.2.2)}

\leq \sup_{\pi \in \beta} \inf_{\beta} R(\pi, \beta) \quad \text{since } \beta \subseteq \beta_p

= \inf_{\beta} \sup_{\pi \in \beta} R(\pi, \beta) \quad \text{by theorem 2.1.6}

= \sup_{\pi \in \beta} R(\pi, B_0) \quad \text{since } B_0 \text{ is partially Bayes}

\leq \tau \quad \text{since } B_0 \in \beta.

Thus \( B^\pi \in \beta \). It has been shown that if \( B^\pi \in \beta_p \) and \( B^\pi \) satisfies (2.2.2) then \( B^\pi \in \beta \). Thus

\[ \sup_{\pi \in \beta} \inf_{\beta} R(\pi, \beta) = \sup_{\pi \in \beta} \inf_{\beta} R(\pi, \beta), \]

\[ \sup_{\pi \in \beta} \inf_{\beta} R(\pi, \beta), \quad \text{and the result follows from (2.2.1).} \]

The next theorem accomplishes the reduction to a maximization problem and is the main result of this chapter.

**Theorem 2.2.2:** In \( L \), the partially Bayes risk

\[ \inf_{\beta} \sup_{\mu \in \mathcal{U}} R(\mu, \beta) = \sup_{\pi \in \mathcal{E}} R(\pi, B^\pi) \]

= \sup_{\pi \in \mathcal{E}} \text{tr } K_\pi^\pi (\Sigma_0 + D_\pi)
where

\[ B_{\Pi} = (\Sigma_0 + D_{\Pi})(\Sigma_1 + \Sigma_0 + D_{\Pi})^{-1}, \]

and

\[ D_{\Pi} = \sum_{i=1}^{u} \pi_i e_i e_i^T. \]

Furthermore, there is a \( \Pi_0 \in \mathcal{E} \) such that

\[ R(\Pi_0, B_{\Pi_0}) = \sup_{\Pi \in \mathcal{E}} R(\Pi, B_{\Pi}), \]

and the rule \( e_0 \) defined by

\[ e_0(x) = B_{\Pi_0} x + B_{\Pi_0} \Delta \]

is partially Bayes in \( L \).

**Proof:** The existence of the partially Bayes rule is demonstrated in the comments following theorem 2.1.6. Now by lemma 2.2.1,

\[ \inf_{B \in \mathcal{E}} \sup_{\mu \in \mathcal{U}} R(\mu, B) = \sup_{\Pi \in \mathcal{E}} \inf_{B \in \mathcal{E}} R(\Pi, B), \]

and from (2.1.3)

\[ R(\mu, B) = \text{tr} \, B^T K \Sigma_1 + \text{tr} \, B^T K \Sigma_0 + \mu^T B^T K \mu \]

so that

\[ R(\Pi, B) = \sum_{i=1}^{u} R(e_i, B) \pi_i \]
\[ = \text{tr} B^T K B \Sigma_1 + \text{tr} B^T K B \Sigma_0 + \sum_{i=1}^{u} e_i^T B^T K B e_i \pi_i. \]

Now \( e_i^T B^T K B e_i \) is a scalar, and so

\[ e_i^T B^T K B e_i = \text{tr} e_i^T B^T K B e_i = \text{tr} B^T K B e_i e_i^T. \]

Thus

\[ (2.2.3) \quad R(\Pi, B) = \text{tr} [B^T K B \Sigma_1 + B^T K B (\Sigma_0 + D \Pi)] \]

\[ = \text{tr} K [B \Sigma_1 B^T + B \Sigma B^T] \]

by the cyclical property of the trace, where

\[ \Sigma = \Sigma_0 + D \Pi. \]

To obtain \( B \in \mathcal{B}_P \) to minimize \( R(\Pi, B) \), let

\[ B = \Sigma (\Sigma_1 + \Sigma)^{-1} + G. \]

Then from (2.2.3)

\[ R(\Pi, B) = \text{tr} K [B \Sigma_1 B^T + \Sigma + B \Sigma B^T - \Sigma B^T - B \Sigma] \]

\[ = \text{tr} K [B (\Sigma_1 + \Sigma) B^T + \Sigma - \Sigma B^T - B \Sigma] \]

\[ = \text{tr} K [(\Sigma (\Sigma_1 + \Sigma)^{-1} + G)(\Sigma_1 + \Sigma)[(\Sigma_1 + \Sigma)^{-1} \Sigma + G^T] + \Sigma \]

\[ - \Sigma (\Sigma_1 + \Sigma)^{-1} \Sigma + G^T - \Sigma (\Sigma_1 + \Sigma)^{-1} + G] \Sigma] \]

\[ = \text{tr} K \left[ (\Sigma_1 + \Sigma)^{-1} \Sigma + G^T + G \Sigma + G (\Sigma_1 + \Sigma) G^T + \Sigma \]

\[ - \Sigma (\Sigma_1 + \Sigma)^{-1} \Sigma - G^T - \Sigma (\Sigma_1 + \Sigma)^{-1} \Sigma - G \Sigma \right]. \]
\[ = \text{tr} \ K [ \Sigma - \Sigma (\Sigma_1 + \Sigma)^{-1} \Sigma ] + \text{tr} \ K G (\Sigma_1 + \Sigma) G^T. \]

Since \( K \) and \( G (\Sigma_1 + \Sigma) G^T \) are at least non-negative definite, all their characteristic roots are non-negative, and appealing again to the result of Anderson and Gupta [1963, corollary 2.2.1], all the characteristic roots of \( K G (\Sigma_1 + \Sigma) G^T \) are non-negative. Thus

\[ \text{tr} \ K G (\Sigma_1 + \Sigma) G^T \geq 0 \]

with equality if \( G = 0 \). It follows that for each \( \Pi \in \mathcal{A} \)

\[ B_{\Pi} = \Sigma (\Sigma_1 + \Sigma)^{-1} \]

(2.2.4)

\[ = (\Sigma_0 + D_{\Pi})(\Sigma_1 + \Sigma_0 + D_{\Pi})^{-1} \]

minimizes \( \mathcal{R}(\Pi, B) \) over \( \mathcal{A}_\Pi \). Finally

\[ \inf_{B \in \mathcal{A}} \mathcal{R}(\Pi, B) \]

\[ = \mathcal{R}(\Pi, B_{\Pi}) \]

\[ = \text{tr} \ B_{\Pi}^T K B_{\Pi} - \text{tr} \ B_{\Pi}^T K B_{\Pi} (\Sigma_0 + D_{\Pi}) \]

\[ = \text{tr} \ K [ B_{\Pi} (\Sigma_1 + \Sigma_0 + D_{\Pi}) B_{\Pi}^T + (\Sigma_0 + D_{\Pi}) - B_{\Pi} (\Sigma_0 + D_{\Pi}) - (\Sigma_0 + D_{\Pi}) B_{\Pi}^T ] \]

\[ = \text{tr} \ K [ B_{\Pi} (\Sigma_0 + D_{\Pi}) B_{\Pi}^T + (\Sigma_0 + D_{\Pi}) B_{\Pi}^T ] \]

\[ = \text{tr} \ K B_{\Pi} (\Sigma_0 + D_{\Pi}). \]
2.3 Linear Estimators

In section 2.2, a partially Bayes rule in the set \( L \) (2.1.1) was derived. It was asserted in section 2.1 that only such rules need be considered, that is, in the class \( D_L = \{ \delta : x \rightarrow \Theta(\delta(x) = Bx + C) \} \), of all linear rules, any partially Bayes rule must be in \( L \). Theorem 2.3.2 justifies that assertion.

**Lemma 2.3.1:** For \( p \)-vectors \( G, \mu \), and \( M \), \( p \times p \) matrix \( \bar{B} \) and positive definite \( p \times p \) matrix \( K \), if \( G \neq 0 \) and \( M > 0 \),

\[
(2.3.1) \quad \begin{align*}
\sup_{|\mu| \leq M} (\bar{B} \mu + G)^T K (\bar{B} \mu + G) > \sup_{|\mu| \leq M} (\bar{B} \mu)^T K (\bar{B} \mu).
\end{align*}
\]

**Proof:** There is a \( \mu_0 \) such that \( |\mu_0| \leq M \) and the supremum on the right hand side of (2.3.1) is attained at \( \mu_0 \). Now

\[
(\bar{B} \mu_0 + G)^T K (\bar{B} \mu_0 + G) = \mu_0^T \bar{B}^T K \bar{B} \mu_0 + 2 \mu_0^T \bar{B}^T K G + G^T K G,
\]

and since \( |-\mu_0| \leq M \)

\[
\sup_{|\mu| \leq M} (\bar{B} \mu + G)^T K (\bar{B} \mu + G) \geq \mu_0^T \bar{B}^T K \bar{B} \mu_0 + 2 |\mu_0^T \bar{B}^T K G| + G^T K G.
\]

But \( K \) is positive definite and \( G \neq 0 \), so \( G^T K G > 0 \) and thus

\[
\sup_{|\mu| \leq M} (\bar{B} \mu + G)^T K (\bar{B} \mu + G) > \mu_0^T \bar{B}^T K \bar{B} \mu_0 = \sup_{|\mu| \leq M} (\bar{B} \mu)^T K (\bar{B} \mu).
\]

**Theorem 2.3.2:** If the rule \( \delta(x) = Bx + C \) is partially Bayes among all linear rules, then \( C = \bar{B} \Delta \), so that partially Bayes rules in \( D_L \) are of the form \( \delta_0(x) = Bx + \bar{B} \Delta \), that is \( \delta \in L \).
Proof: Without loss of generality, write \( \delta(x) = Bx + \bar{B}\Delta + G \), where \( G \) is the \( p \)-vector \( C - \bar{B}\Delta \). The Bayes risk of \( \delta \) for given \( \mu \) is

\[
B(\mu, \delta) = E(\delta(x) - \bar{\delta})^T K (\delta(x) - \bar{\delta})
\]

\[
= E[B(x-\bar{\delta}) + \bar{B}(\mu-\bar{\delta}) + \bar{B}(\Delta-\mu) + G]^T K [B(x-\bar{\delta}) + \bar{B}(\mu-\bar{\delta}) + \bar{B}(\Delta-\mu) + G]
\]

\[
= \text{tr} \bar{B}^T K B \Sigma_1 + \text{tr} \bar{B}^T K B \Sigma_0 + [\bar{B}(\Delta-\mu) + G]^T K [\bar{B}(\Delta-\mu) + G].
\]

Since \( \delta_0(x) \) is \( \delta(x) \) with \( G = 0 \),

\[
B(\mu, \delta_0) = \text{tr} \bar{B}^T K B \Sigma_1 + \text{tr} \bar{B}^T K B \Sigma_0 + [\bar{B}(\Delta-\mu)]^T K [\bar{B}(\Delta-\mu)].
\]

It must be shown that if \( G \neq 0 \),

\[
(2.3.2) \quad \sup_{\mu \in U} B(\mu, \delta) > \sup_{\mu \in U} B(\mu, \delta_0).
\]

But this is an immediate consequence of lemma 2.3.1.

Notice that in sections 2.1 and 2.2 the non-negative definite character of \( K \) was used (lemma 2.1.5, theorem 2.2.2), but that only in this section (lemma 2.3.1) was it necessary to require \( K \) to be positive definite. If it is assumed that \( K \) is non-negative definite and \( \delta_1 \) is partially Bayes in \( D_L \), there exists a rule \( \delta_0 \) in \( L \) with partially Bayes risk not exceeding that of \( \delta_1 \). This is seen by noting that lemma 2.3.1 holds for \( K \) non-negative definite if "\( > \)" is replaced by "\( \geq \)", and thus (2.3.2) holds subject to the same substitution.
2.4 Example: Estimation of a Bivariate Location Parameter

This section presents numerical examples of the location problem when \( p = 2 \). With \( m_1, m_2 > 0 \), the notation becomes:

\[
K = \begin{bmatrix} k_{11} & k_{12} \\ k_{12} & k_{22} \end{bmatrix}, \quad \Sigma_0 = \begin{bmatrix} \sigma_{11}^0 & \sigma_{12}^0 \\ \sigma_{12}^0 & \sigma_{22}^0 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} \sigma_{11}^1 & \sigma_{12}^1 \\ \sigma_{12}^1 & \sigma_{22}^1 \end{bmatrix},
\]

\[
M = (m_1, m_2)^T, \quad \Pi = (\pi_1, \pi_2, \pi_3, \pi_4)^T.
\]

Also,

\[
e_1 = M, \quad e_2 = (-m_1, m_2)^T, \quad e_3 = -e_2, \quad e_4 = -e_1,
\]

and

\[
D_q = \sum_{i=1}^{l} \pi_i e_i e_i^T = \begin{bmatrix} m_1^2 & qm_1 m_2 \\ qm_1 m_2 & m_2^2 \end{bmatrix},
\]

where \( q = (\pi_1 + \pi_4) - (\pi_2 + \pi_3) \) and \( q \in [-1, 1] \). Note that the maximizing \( \Pi \) depends only on the difference, \( q \), in weights placed on the diagonals of the box. The computations are based on theorem 2.2.2 which becomes:

**Corollary:** If \( p = 2 \), the partially Bayes risk of the partially Bayes rule is given by

\[
\mathcal{R}(B_0) = \sup_{q \in [-1, 1]} \text{tr} K \Sigma_1^{-1}(\Sigma_1 + \Sigma_0 + D_q)^{-1}(\Sigma_0 + D_q),
\]

and if the supremum is attained at \( q = q_0 \), then the rule defined by

\[
\delta_0(x) = B_0 x + \overline{B}_0 \Delta,
\]
with

\[ B_0 = (\Sigma_0 + D_{q_0})(\Sigma_1 + \Sigma_0 + D_{q_0})^{-1} \]

is partially Bayes.

The computations were performed on the CDC 6400 computer at The Florida State University Computing Center. For various choices of \( K, \Sigma_0, \Sigma_1 \), and \( \lambda \), Table 1 gives the partially Bayes rule, its partially Bayes risk, and the maximizing \( q_0 \), together with the corresponding (linear) Bayes rule and its risk. A (linear) Bayes rule (with respect to the prior distribution \( \lambda \)) is a rule \( \delta_1 \) such that

\[
E_{\lambda}\{E[(\delta_1(x) - \bar{\theta})^T K (\delta_1(x) - \bar{\theta}) | \bar{\theta}]\}
\]

\[
= \inf_{\delta \in D_L} E_{\lambda}\{E[(\delta(x) - \bar{\theta})^T K (\delta(x) - \bar{\theta}) | \bar{\theta}]\}.
\]

In what follows, \( B_1 \) is the coefficient matrix of \( x \) in the (linear) Bayes rule

\[ \delta_1(x) = B_1 x + \bar{B}_1 \mu. \]

That is, \( B_1 = \Sigma_0 (\Sigma_1 + \Sigma_0)^{-1} \). The Bayes risk of the (linear) Bayes rule is

\[ R(B_1) = tr KE_1 B_1^T. \]
Table 1.—Selected results for the location problem with $p = 2$

<table>
<thead>
<tr>
<th>Example</th>
<th>$K$</th>
<th>$\Sigma_0$</th>
<th>$\Sigma_1$</th>
<th>$M$</th>
<th>$q_0$</th>
<th>$B_0$</th>
<th>$R(B_0)$</th>
<th>$B_1$</th>
<th>$R(B_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>0.00</td>
<td>$\begin{bmatrix} 0.667 &amp; 0 \end{bmatrix}$</td>
<td>1.333</td>
<td>$\begin{bmatrix} 0.500 &amp; 0 \end{bmatrix}$</td>
<td>1.000</td>
</tr>
<tr>
<td>2</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 2 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>0.80</td>
<td>$\begin{bmatrix} 0.641 &amp; 0.096 \end{bmatrix}$</td>
<td>2.756</td>
<td>$\begin{bmatrix} 0.500 &amp; 0 \end{bmatrix}$</td>
<td>2.000</td>
</tr>
<tr>
<td>3</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 3 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>0.63</td>
<td>$\begin{bmatrix} 0.651 &amp; 0.073 \end{bmatrix}$</td>
<td>3.403</td>
<td>$\begin{bmatrix} 0.500 &amp; 0 \end{bmatrix}$</td>
<td>2.500</td>
</tr>
<tr>
<td>4</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 2 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 5 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 3 &amp; 0 \ 0 &amp; 2 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>0.58</td>
<td>$\begin{bmatrix} 0.473 &amp; 0.104 \end{bmatrix}$</td>
<td>6.197</td>
<td>$\begin{bmatrix} 0.382 &amp; 0.088 \end{bmatrix}$</td>
<td>5.471</td>
</tr>
<tr>
<td>5</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 2 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 5 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 3 &amp; 0 \ 0 &amp; 2 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>0.56</td>
<td>$\begin{bmatrix} 0.641 &amp; 0.094 \end{bmatrix}$</td>
<td>7.177</td>
<td>$\begin{bmatrix} 0.382 &amp; 0.088 \end{bmatrix}$</td>
<td>5.471</td>
</tr>
<tr>
<td>6</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 5 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 3 &amp; 0 \ 0 &amp; 2 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>-1.00</td>
<td>$\begin{bmatrix} 0.500 &amp; 0 \end{bmatrix}$</td>
<td>3.000</td>
<td>$\begin{bmatrix} 0.382 &amp; 0.088 \end{bmatrix}$</td>
<td>2.559</td>
</tr>
<tr>
<td>7</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 5 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 3 &amp; 0 \ 0 &amp; 2 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>-1.50</td>
<td>$\begin{bmatrix} 0.667 &amp; 0 \end{bmatrix}$</td>
<td>3.500</td>
<td>$\begin{bmatrix} 0.382 &amp; 0.088 \end{bmatrix}$</td>
<td>2.559</td>
</tr>
<tr>
<td>8</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 5 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 3 &amp; 0 \ 0 &amp; 2 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>-1.00</td>
<td>$\begin{bmatrix} 0.384 &amp; 0.087 \end{bmatrix}$</td>
<td>2.565</td>
<td>$\begin{bmatrix} 0.382 &amp; 0.088 \end{bmatrix}$</td>
<td>2.559</td>
</tr>
<tr>
<td>9</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 5 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 3 &amp; 2 \ 0 &amp; 2 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>1.00</td>
<td>$\begin{bmatrix} 0.500 &amp; 0 \end{bmatrix}$</td>
<td>2.750</td>
<td>$\begin{bmatrix} 0.423 &amp; -0.038 \end{bmatrix}$</td>
<td>2.269</td>
</tr>
<tr>
<td>10</td>
<td>$\begin{bmatrix} 1 &amp; 0 \ 0 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 2 &amp; 1 \ 1 &amp; 1 \end{bmatrix}$</td>
<td>$\begin{bmatrix} 1 \end{bmatrix}$</td>
<td>0.63</td>
<td>$\begin{bmatrix} 0.583 &amp; 0.032 \end{bmatrix}$</td>
<td>1.859</td>
<td>$\begin{bmatrix} 0.500 &amp; 0 \end{bmatrix}$</td>
<td>1.500</td>
</tr>
<tr>
<td>Example</td>
<td>$K$</td>
<td>$\Sigma_0$</td>
<td>$\Sigma_1$</td>
<td>$M$</td>
<td>$q_0$</td>
<td>$B_0$</td>
<td>$R(B_0)$</td>
<td>$B_1$</td>
<td>$R(B_1)$</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>------------</td>
<td>------------</td>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>---------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>-1.00</td>
<td>0.522</td>
<td>0.1307</td>
<td>2.391</td>
<td>0.400</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>2</td>
<td>0.087</td>
<td>0.522</td>
<td></td>
<td>2.391</td>
<td>0.200</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>-1.00</td>
<td>0.500</td>
<td>0.00</td>
<td>2.500</td>
<td>0.357</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0.00</td>
<td>0.500</td>
<td></td>
<td>2.500</td>
<td>0.143</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0.00</td>
<td>0.500</td>
<td>0.714</td>
<td>2.929</td>
<td>0.400</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0.00</td>
<td>0.714</td>
<td></td>
<td>2.929</td>
<td>0.065</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>-1.00</td>
<td>0.667</td>
<td>0.00</td>
<td>3.200</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.00</td>
<td>0.600</td>
<td></td>
<td>3.200</td>
<td>0.065</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>-0.50</td>
<td>0.750</td>
<td>0.00</td>
<td>3.450</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.00</td>
<td>0.600</td>
<td></td>
<td>3.450</td>
<td>0.065</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>-0.25</td>
<td>0.750</td>
<td>0.00</td>
<td>3.750</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0.00</td>
<td>0.750</td>
<td></td>
<td>3.750</td>
<td>0.065</td>
</tr>
</tbody>
</table>
The first three examples illustrate the effect of changing $K$. There is no effect on the Bayes rule (only on its Bayes risk), but the partially Bayes rule does change. Examples 4 and 5; 6, 7, and 8; and 14, 15, and 16, show the effect of changing $M$. Examples 6 and 9; and 10, 11, and 12, have different $F_1$, while 6, 12, 13, and 14, differ in $\Sigma_0$. Examples 6, 12, and 14 show that $B_0$ can be diagonal when $B_1$ is not. Notice that the examples show many values of $q_0$. 
III. SOME SPECIAL CASES OF THE LOCATION PROBLEM

The first section of this chapter presents, for use in chapter VI, the partially Bayes rule and its partially Bayes risk for the case $p = 1$. The remaining sections treat other special cases.

3.1 The Univariate Case

Consider in the univariate case, rules of the form $b(x) = bx + c$, and terminal loss function $k(b(x) - \theta)^2$. Identify $\Sigma_1 = \sigma_1^2$ and $\Sigma_0 = \sigma_0^2$. Other changes of notation should be clear.

**Corollary:** If $p = 1$ and $D_L = \{b: x \rightarrow \Omega | b(x) = bx + c\}$, then the partially Bayes risk in $D_L$ is

$$k \left[ \frac{1}{\sigma_1^2} + \frac{1}{\sigma_0^2 + M^2} \right]^{-1}$$

and the unique partially Bayes rule is

$$b_0(x) = \frac{x + \Delta}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_0^2 + M^2}} \cdot \frac{1}{\sigma_1^2} \cdot \frac{1}{\sigma_0^2 + M^2}$$

**Proof:** Since $p = 1$, $\Sigma = \{M, -M\}$ has at most two elements, while each element of $\Sigma$ may be characterized by one parameter $\Pi$ and $\Pi = 1 - \Pi$, $0 \leq \Pi \leq 1$. Applying theorem 2.2.2, note that for any $\Pi$,
\[ D_\Pi = \Pi M^2 + \overline{\Pi}(-4)^2 = \lambda^2, \text{ independent of } \Pi. \] It also follows from the proof of theorem 2.2.2 that for each \( \Pi \) there is a unique \( b_\Pi \) such that

\[ R(\Pi, b_\Pi) = \inf_b R(\Pi, b) \]

and that this is

\[ b_\Pi = \frac{\sigma_0^2 + \lambda^2}{\sigma_1^2 + \sigma_0^2 + \lambda^2}, \]

independent of \( \Pi \).

Thus the partially Bayes risk of the partially Bayes rule is

\[ \sup_{\Pi \in \mathcal{E}} k b_\Pi (\sigma_0^2 + \lambda^2) = k \frac{\sigma_0^2 (\sigma_0^2 + \lambda^2)}{\sigma_1^2 + \sigma_0^2 + \lambda^2} = k \left[ \frac{1}{\sigma_1^2} + \frac{1}{\sigma_0^2 + \lambda^2} \right]^{-1}, \]

and the unique partially Bayes rule is obtained by substituting \( b_\Pi \) into \( \delta_0(x) = b_\Pi x + \overline{\delta}_\Pi \lambda \).

### 3.2 The Case \( M = 0 \)

If \( M = (0, 0, \ldots, 0)^T \), so that \( \mu = \lambda \) is known, the partially Bayes rule in \( D_L \) reduces to the (linear) Bayes rule. For if \( M = 0 \), the set \( \mathcal{E} \) consists of only the \( p \)-vector zero, and \( \overline{\mathcal{E}} = \{ \Pi \} \) where \( \Pi \) assigns probability one to the vector zero. Thus \( D_\Pi \) of theorem 2.2.2 is zero, and the partially Bayes rule is

\[ \delta_0(x) = B_0 x + \overline{\delta}_0 \mu, \]

where

\[ B_0 = \Sigma_0 (\Sigma_1 + \Sigma_0)^{-1}. \]
This can be shown to be the (linear) Bayes rule. This result does not depend on the loss matrix $K$.

### 3.3 The Case of Diagonal $B$

Suppose that the loss matrix is diagonal, $K = \text{diag}(k_i)$. Assume furthermore that the matrix $B$ is required to be diagonal, so that only rules of the form

$$\hat{s}(x) = (\hat{s}_1(x), \hat{s}_2(x), \ldots, \hat{s}_p(x))^T,$$

where

$$\hat{s}_i(x) = b_i x_i + \Delta_i, \quad \Delta_i = l - b_i$$

are considered. Let

$$L' = \{ \hat{s}: x \rightarrow \Omega | \hat{s}(x) = Bx + \Delta, \text{with } B \text{ diagonal} \}.$$

Then $L' \subset L \subset D_L$.

**Theorem:** If $K = \text{diag}(k_i)$, then the partially Bayes rule in $L'$ is

$$\hat{s}'(x) = (\hat{s}'_1(x), \hat{s}'_2(x), \ldots, \hat{s}'_p(x))^T$$

where

$$\hat{s}'_i(x) = \frac{x_i}{\sigma_{ii}} + \frac{\Delta_i}{\sigma_{ii}^2 + m_i^2}$$

and the partially Bayes risk is

$$\sum_{i=1}^{p} \frac{k_i}{\frac{1}{\sigma_{ii}} + \frac{1}{\sigma_{ii}^2 + m_i^2}}^{-1}.$$
\textbf{Proof:} In \( L' \), the Bayes risk of \( \delta \) for given \( \mu \) is

\[
B(\mu, \delta) = \mathbb{E}(\delta(X) - \bar{\delta})^2 \mathbb{E}(\delta(X) - \bar{\delta})
\]

\[
= \mathbb{E} \sum_{i=1}^{p} k_i [b_i X_i + \bar{b}_i \Delta_i - \bar{\delta}]^2
\]

\[
= \sum_{i=1}^{p} k_i \mathbb{E}[b_i (X_i - \bar{\mu}_i) + \bar{b}_i (\Delta_i - \bar{\delta})]^2
\]

\[
= \sum_{i=1}^{p} k_i [b_i \sigma^2_{ii} + \bar{b}_i \sigma^2] + \bar{b}_i (\Delta_i - \mu_i)^2,
\]

and the maximum Bayes risk of \( \delta \) is

\[
B(\delta) = \sup_{\mu \in U} B(\mu, \delta)
\]

(3.3.2)

\[
= \sum_{i=1}^{p} k_i [b_i \sigma^2_{ii} + \bar{b}_i (\sigma^2 + m^2)].
\]

By differentiating with respect to \( b_i \), \( i = 1, 2, \ldots, p \), and setting each derivative equal to zero, it follows that \( B(\delta) \) is minimized at

\[
b_i^* = \frac{m_i^2 + \sigma^2_{ii}}{\sigma^2_{ii} + \sigma^2 + m^2_i}.
\]

The result is obtained by substitution in (3.3.1) and (3.3.2).

In considering diagonal rules, note from section 3.2 that the (linear) Bayes rule has the coefficient matrix \( B_0 \) diagonal whenever \( \Sigma_0 \) and \( \Sigma_1 \) are diagonal. This seems reasonable since if the \( X_i \) and \( \bar{\delta}_j \) are
uncorrelated, one would not expect to obtain information about $\mathbf{\tilde{y}}_j$ from $X_i$ ($i \neq j$). If the loss matrix $K$ is diagonal, the partially Bayes rule in $D_L$ also has this property as shall be seen in the next section.

3.4 The Case of Diagonal $\Sigma_0, \Sigma_1, K$

In this section an analytic expression is obtained for the partially Bayes rule in $D_L$ for the case in which $\Sigma_0$, $\Sigma_1$, and $K$ are all diagonal. The approach is similar to that used in deriving the relationship between Bayes and minimax rules mentioned in chapter I.

Suppose $\Sigma_1 = \text{diag}(\sigma_1^1)$, $\Sigma_0 = \text{diag}(\sigma_0^1)$, and $K = \text{diag}(k_i)$. Recall that $B_0$ is sought such that

$$\sup_{\mu \in U} R(\mu, B_0) = \inf_{B \in \mathcal{B}} \sup_{\mu \in U} R(\mu, B),$$

or equivalently, by (2.1.6), such that

$$\sup_{e \in \mathcal{E}} R(e, B_0) = \inf_{B \in \mathcal{B}} \sup_{e \in \mathcal{E}} R(e, B).$$

Define

$$B_0 = \text{diag.} \left[ \frac{\sigma_i^0 + m_i^2}{\sigma_i^1 + \sigma_i^0 + m_i^2} \right].$$

(3.4.1)

Needing only to consider rules in $L$, a distribution $\lambda$ on $\mathcal{E}$ will be found such that $B_0$ minimizes $E_\lambda R(e, B)$ and it will be shown that $R(e, B_0)$ is constant on $\mathcal{E}$. It will follow that $\delta_0$ defined by

$$\delta_0(x) = B_0x + \tilde{B}_0\Delta$$
with $B_0$ as in (3.4.1), is partially Bayes in $D_L$ (theorem 3.4.3).

**Lemma 3.4.1:** With $B_0$ diagonal and with $\Sigma_0$, $\Sigma_1$, and $K$ diagonal, $R(e, B_0) = R$ is constant on $E$.

**Proof:** Notice that for an arbitrary element $e_i \in E$,

$$R(e_i, B_0) = \text{tr} \ B_0^T K B_0 \Sigma_1 + \text{tr} B_0^T K \Sigma_0 e_i e_i^T + \text{tr} B_0^T K B_0 e_i e_i^T$$

$$= \text{tr} B_0^T K B_0 \Sigma_1 + \text{tr} B_0^T K \Sigma_0 e_i e_i^T + \sum_{j=1}^{p} (B_0^T K B_0)_{jj} m_j^2$$

$$= \text{R \ say,}$$

since $B_0^T K B_0$ is diagonal, and the diagonal elements of $e_i e_i^T$ are $m_1^2, m_2^2, \ldots, m_u^2$ for every $i = 1, 2, \ldots, u$.

**Lemma 3.4.2:** With $B_0$ as in (3.4.1), there exists a distribution $\lambda_0$ on $E$ such that

$$R = E_{\lambda_0} R(e, B_0) = \inf_{B \in E} R(e, B).$$

**Proof:** For $e_i \in E$ let

$$\lambda_0(e_i) = \frac{1}{u}, \quad i = 1, 2, \ldots, u.$$

Then for $B \in E$

$$E_{\lambda_0} R(e, B) = \sum_{i=1}^{u} R(e_i, B) \lambda_0(e_i)$$

$$= \text{tr} B^T K B \Sigma_1 + \text{tr} B^T K \Sigma_0 + \frac{1}{u} \sum_{i=1}^{u} e_i B^T K B e_i$$
= \text{tr} \, \mathbf{B}^T \mathbf{K} \mathbf{B} \Sigma_1 + \text{tr} \, \mathbf{B}^T \mathbf{K} \mathbf{B} \Sigma_0 + \frac{1}{u} \text{tr} \, \mathbf{B}^T \mathbf{K} \mathbf{B} \sum_{i=1}^{u} \mathbf{e}_i \mathbf{e}_i^T.

With some computation it can be shown that

\[ \sum_{i=1}^{u} \mathbf{e}_i \mathbf{e}_i^T = u \hat{\mathbf{M}}^2, \]

where \( \hat{\mathbf{M}} \) is the \( p \times p \) diagonal matrix with entries \( m_1, m_2, \ldots, m_p \). Thus

\[ E_{\lambda_0^*} R(e, B) = \text{tr} \, \mathbf{B}^T \mathbf{K} \mathbf{B} \Sigma_1 + \text{tr} \, \mathbf{B}^T \mathbf{K} \mathbf{B} (\Sigma_0 + \hat{\mathbf{M}}^2). \]

The calculus may be applied directly to determine the minimizing \( B \), but note that the minimization is identical to that in the proof of theorem 2.2.2 if \( D_\Pi \) is replaced by \( \hat{\mathbf{M}}^2 \) in display (2.2.3). Thus from (2.2.4), the minimizing \( B \) is

\[ (\Sigma_0 + \hat{\mathbf{M}}^2)(\Sigma_1 + \Sigma_0 + \hat{\mathbf{M}}^2)^{-1} = \text{diag} \left[ \frac{\sigma_1^0 + m_1^2}{\sigma_1^0 + \sigma_1^0 + m_1^2} \right] = B_0. \]

**Theorem 3.4.3:** With \( \Sigma_0 = \text{diag} (\sigma_1^0) \), \( \Sigma_1 = \text{diag} (\sigma_1^0) \), and \( \mathbf{K} \) diagonal, the partially Bayes rule in \( D_L \) is given by

\[ \delta_0(x) = B_0 x + \overline{B}_0 \Delta, \]

where \( B_0 \) is defined in (3.4.1).

**Proof:** With \( \lambda_0 \) as in lemma 3.4.2, for any \( B \in \mathcal{G}_p \) observe that
\[
\sup_{e \in \mathcal{E}} R(e, B_0) = \sum_{i=1}^u \lambda_0(e_i) \quad \text{by lemma 3.4.1}
\]
\[
= \sum_{i=1}^u R(e_i, B_0) \lambda_0(e_i)
\]
\[
\leq \sum_{i=1}^u R(e_i, B) \lambda_0(e_i) \quad \text{by lemma 3.4.2}
\]
\[
\leq \sup_{e \in \mathcal{E}} R(e, B).
\]

Thus
\[
\sup_{e \in \mathcal{E}} R(e, B_0) = \inf_{B \in \mathcal{D}} \sup_{e \in \mathcal{E}} R(e, B)
\]

so that \(\delta_0\) is partially Bayes in \(D_L\).
IV. SOME RELATED RESULTS FOR THE LOCATION PROBLEM

This chapter will make suggestions for further research, give indications of the results one might obtain, and solve some problems related to the location problem.

4.1 Other Incomplete Specifications of the Prior Distribution

With the sampling distribution, prior distribution, and loss structure as in sections 1.3 and chapter II, there is a need for investigation of other forms of the incompleteness specification for the prior mean. For example, consider an arbitrary decreasing collection \( \{ C_m \} \) of sets, for which \( \mu = \cap C_m \), and suppose that it is learned that \( \mu \in C_m \) at some cost \( a(m) \). Note that the results of chapter II apply if the \( C_m \) are arbitrary closed convex polyhedra with extreme points \( \{ e_i \} \) since the crucial assumption in the minimax theorem of section 2.1 is the finiteness of the number of extreme points. For any compact set \( S_m \), an approximation to the partially Bayes rule with respect to \( S_m \) is obtained by using the partially Bayes rule with respect to a convex polyhedron \( C_m \) containing \( S_m \). In this case, the maximum Bayes risk on \( C_m \) is an upper bound for the maximum on \( S_m \), and thus the partially Bayes risk with respect to \( C_m \) is an upper bound for that with respect to \( S_m \). This suggests an approach for obtaining a rule which is partially Bayes with respect to the arbitrary convex compact set \( S \), by considering the limit of a sequence of rules, partially Bayes with respect to a sequence \( \{ C_m \} \) of convex polyhedra decreasing to \( S \).
Observe too, that the techniques of chapter II are applicable to compact convex sets other than convex polyhedra. For example, suppose

\[ U = \{ \mu \in \mathbb{E}^p \mid (\Delta - \mu)^T G (\Delta - \mu) \leq m \} \]

is an ellipsoid in p-dimensional Euclidean space. Then by theorem A.1 it still follows that the maximum over U, of the Bayes risk occurs at an extreme point of U. A complication arises in that the set of extreme points,

\[ \mathcal{E} = \{ \mu \in \mathbb{E}^p \mid (\Delta - \mu)^T G (\Delta - \mu) = m \} \]

is the boundary of the ellipsoid and is not finite. Thus the minimax theorem of section 2.1 does not apply. However a more general version of the theorem only requires the set of extreme points to be compact in the Wald topology. For a precise statement see Stein [1963, p.I.3.7].

### 4.2 Other Distribution Assumptions

In considering quadratic form loss functions and linear rules, the first two moments sufficiently characterize the prior and sampling distributions. Subject to this restriction, the assumptions of chapter II may be weakened by assuming not that \( \text{Cov}(X|\bar{y}) = \Sigma \), but only that \( \text{Cov}(X|\bar{y}) = \Sigma_1(\bar{y}) \) where

\[ E\{ \Sigma_1(\bar{y}) | \mu \} = \Sigma_1 \]

independent of \( \mu \). Phrasing this somewhat differently, suppose that both the sampling distribution's mean \( \bar{y} \) and covariance \( \Sigma \) are unknown, and that the decision maker has a joint prior distribution on \( \bar{y} \) and \( \Sigma \),
such that under this prior, \( E(\tilde{\sigma}) = \mu \), \( \text{Cov}(\tilde{\sigma}) = \Sigma_0 \), and \( E(\Sigma) = \Sigma_1 \), where \( \Sigma_0 \) and \( \Sigma_1 \) are known, and it has been learned that \( \mu \in U \). The results are identical to those of chapter II. This shall be seen in a slightly different context in section 4.4.

As a further example of changing the distribution assumptions, suppose \( X \) is a (univariate) positive random variable with mean \( \overline{\sigma} \) and variance \( \text{var}(X|\overline{\sigma}) = \nu \overline{\sigma} \), where \( \nu > 0 \) is known. Again suppose that the prior distribution is such that \( E(\overline{\sigma}) = \mu \), \( \text{var}(\overline{\sigma}) = \sigma_o^2 \), where \( \sigma_o^2 \) is known and

\[
\mu \in U = \{ \mu \in \mathbb{R}^1 \mid |\Delta - \mu| \leq M \}.
\]

Considering rules of the form \( \delta(x) = bx + \overline{b}\Delta \) and quadratic loss,

\[
B(\mu, \delta) = E(bX + \overline{b}\Delta - \overline{\delta})^2
\]

\[
= E[b(X - \overline{\delta}) + \overline{b}(\Delta - \overline{\delta})]^2
\]

\[
= b^2\nu E(\overline{\delta}) + \overline{b}^2 E[(\Delta - \overline{\delta})^2] + \nu\mu + \overline{b}^2[\Delta^2 + \sigma_o^2].
\]

Now

\[
\frac{n^2B(\mu, \delta)}{\mu^2} = 2\overline{b}^2 \geq 0
\]

for all \( \mu \), so \( B(\mu, \delta) \) is convex. It is clearly continuous, so by theorem A.1 it assumes its maximum at an extreme point \( \mu = \Delta \pm M \).

In particular

\[
\sup_{\mu \in U} B(\mu, \delta) = b^2\nu(\Delta + M) + \overline{b}^2(M^2 + \sigma_o^2).
\]
The minimizing $b$ is

$$b = \frac{M^2 + \sigma_0^2}{\nu(\Delta + M) + M^2 + \sigma_0^2}$$

and the partially Bayes risk is

$$R = \left[ \frac{1}{\nu(\Delta + M)} + \frac{1}{M^2 + \sigma_0^2} \right]^{-1}$$

Thus the partially Bayes rule is

$$\delta(x) = \left[ \frac{x}{\nu(\Delta + M)} + \frac{\Delta}{M^2 + \sigma_0^2} \right] R.$$

Note that $R$ involves $\Delta$. More will be said of this in section 4.3.

### 4.3 $\mu$ Known, $\Sigma_0$ Purchased: Univariate

Finally consider a univariate extension which involves changing both distribution assumptions and the manner in which the prior information is improved. Suppose now that the moments of the prior and sampling distributions are as in the original problem (section 3.1), but that $\mu$ is known and $\Sigma_0 = \sigma^2_c$ is not. Assume however, that the decision maker may learn about $\sigma^2_0$ as follows: at some cost, say $\alpha$, he learns that

$$0 \leq \sigma^2_0 \leq \tau_1 \leq \sigma^2_2.$$

Considering rules of the form
\[ \delta(x) = bx + \bar{b}\mu, \]

maintaining the quadratic loss structure, and performing the now familiar computation, with obvious change of notation, it follows that

\[
B(\sigma^2_o, \delta) = E(bX + \bar{b}\mu - \bar{\delta})^2
\]

\[
= E[b(X - \bar{\delta}) + \bar{b}(\mu - \bar{\delta})]^2
\]

\[
= b^2\sigma^2 + \bar{b}^2\sigma^2_o.
\]

Clearly

\[
\sup_{\delta_1 \leq \sigma^2_o \leq \delta_2} B(\sigma^2_o, \delta) = b^2\sigma^2_1 + \bar{b}^2\tau_2.
\]

The minimizing rule is therefore

\[
\delta(x) = \frac{x + \mu}{\frac{\sigma^2}{\sigma^2_1} + \frac{\tau_2}{\tau_2}},
\]

and its risk is

\[
\left[ \frac{1}{\sigma^2} + \frac{1}{\sigma^2_1 \tau_2} \right]^{-1}
\]

Since the procedure is minimax, it is not surprising that \( \tau_2 \), the largest permissible value of \( \sigma^2_o \) appears as the prior variance. Consider a particular choice of \( \tau_1 \) and \( \tau_2 \) to illustrate a point. Suppose that
\[ \tau_1(\Delta, M) = \Delta e^{-M} \text{ and } \tau_2(\Delta, M) = \Delta e^{M}. \]

The risk becomes

\[
\left[ \frac{1}{\sigma^2_0} + \frac{1}{\Delta e^M} \right]^{-1}
\]

Now the decision maker pays to make \( M \) small, thus reducing the size of the interval about \( \sigma^2_0 \), and after payment, receives \( \Delta \). At the worst, \( \Delta \) is taken large to make the risk large. But the risk is bounded by \( \sigma^2_1 \) as \( \Delta \) tends to infinity (and in which case \( \delta(x) \to x \)), and so even in a malicious world, the procedure is workable. The same considerations apply to the last example of section 4.2. Recall that for the problem of chapter II, the partially Bayes risk does not involve \( \Delta \).

4.4 Estimation of a Location Parameter When the Scale Parameter is Unknown

As noted at the end of chapter I, a deterrent to the applicability of the partially Bayes estimate of location as derived in chapter II is the assumption that the variance of the sampling distribution is known. In this section a univariate Normal problem is discussed in which this variance is unknown, but a prior distribution on it is known. The necessary distribution theory can be found in Raiffa and Schlaifer [1961, chapters 7 and 11].

Let \( X_1, X_2, \ldots, X_n \) be independent identically Normally distributed random variables with mean \( \theta \) and variance \( \frac{1}{h} \) (precision \( h \)).
Let
\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,
\quad V = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2,
\]
with the convention that \( V = 0 \) if \( n \leq 1 \). The joint probability density function of \( \bar{X} \) and \( V \) is then
\[
f(\bar{X}, V | \theta, h) = (\text{const.}) \left[ e^{-\frac{1}{2} h n (\bar{X} - \theta)^2 \frac{1}{h}} \right] \left[ e^{-\frac{1}{2} h V \frac{1}{r}} \right]
\]
where \( r = n - 1 \).

The natural conjugate prior distribution for \( \theta \) and \( h \) is the Normal- Gamma, with density
\[
f_N(\theta, h | \mu, v', n^*, n') = f_N(\theta | \mu, h n^*) f_{\nu_2}(h | v', n')
\]
\[
= (\text{const.}) \left[ e^{-\frac{1}{2} h n^* (\theta - \mu)^2 \frac{1}{h}} \right] \left[ e^{-\frac{1}{2} h v' \frac{1}{h} n' - 1} \right]
\]
for
\[-\infty < \theta, \mu < \infty, \quad h \geq 0, \quad \text{and} \quad v', n^* > 0, \quad n' > 2.\]

The following expectations will be needed:
\[
E(\bar{X} | \bar{\theta}, \bar{h}) = \bar{\theta}, \quad E(\theta) = \mu, \quad E \left( \frac{1}{\bar{h}} \right) = \frac{n' v'}{n' - 2} = \psi
\]
\[
\text{var}(\bar{X} | \bar{\theta}, \bar{h}) = \frac{1}{\bar{h} n}, \quad \text{var}(\theta) = \frac{v' n'}{n^*(n' - 2)} = n^{-1} \psi
\]

Suppose that \( \theta \) is to be estimated with quadratic loss, but as before the mean, \( \mu \) of the prior distribution on \( \theta \) is unknown. Again assume that the decision maker has learned
\[ \mu \in U = \{ \mu \mid |\Delta - \mu| \leq M \}. \]

Considering only rules linear in \( \overline{x} \), without loss of generality write

\[ b(\overline{x}) = b\overline{x} + c\Delta, \]

where \( b, c \) do not depend on \( v \), as an estimate of \( \overline{\theta} \) when \( \overline{x} \) is the observed value of \( \overline{x} \). The partially Bayes rule requires \( b, c \) to minimize

\[ \sup_{\mu \in U} E[b\overline{x} + c\Delta - \overline{\theta}]^2. \]

Now

\[ E[b\overline{x} + c\Delta - \overline{\theta}]^2 = E[b(\overline{x} - \overline{\theta}) + \overline{b}(\Delta - \overline{\theta}) + (c - \overline{b})\Delta]^2 \]

\[ = b^2E(\overline{x} - \overline{\theta})^2 + \overline{b}^2E[(\Delta - \mu) + (\mu - \overline{\theta})]^2 + (c - \overline{b})^2\Delta^2 + 2\overline{b}(c - \overline{b})\Delta E(\Delta - \overline{\theta}) \]

\[ (4.4.1) = b^2E \frac{1}{hn} + \overline{b}^2[(\Delta - \mu)^2 + E(\overline{\theta} - \mu)^2] \]

\[ + (c - \overline{b})^2\Delta^2 + 2\overline{b}(c - \overline{b})\Delta(\Delta - \mu) \]

\[ = n^{-1}b^2\psi + [\overline{b}(\Delta - \mu) + (c - \overline{b})\Delta]^2 + n^{-1}b^{-2}\psi \]

\[ = (n^{-1}b^2 + n^{-1}b^{-2})\psi + \overline{b}^2[(\Delta - \mu) + \frac{c - \overline{b}}{\overline{b}}\Delta]^2. \]

The maximum of this expression for \( |\Delta - \mu| \leq M \) is

\[ (n^{-1}b^2 + n^{-1}b^{-2})\psi + \overline{b}^2 \left[ M + \frac{(c - \overline{b})\Delta}{\overline{b}} \right]^2, \]

which in turn is minimized for \( c = \overline{b} \). (Even if \( \Delta = 0 \), so that any \( c \) gives the same value of the expression, the rule
\[ \delta(\bar{x}) = b\bar{x} + \bar{b}\Delta = b\bar{x} + c\Delta = b\bar{x} \]

for all \( c \).

It remains to minimize

\[ (n^{-1}b^2 + n^*-1\bar{b}^2)\psi + \bar{b}^2M^2 \]

over values of \( b \). Applying the calculus, the result is

\[ b = \frac{n^*-1\psi + M^2}{(n^{-1} + n^*-1)\psi + M^2} \]

and the partially Bayes rule is

\[ \delta(\bar{x}) = b\bar{x} + c\Delta \]

\[(4.4.2)\]

\[ = \frac{\bar{x}}{n^{-1}\psi} + \frac{\Delta}{n^*-1\psi + M^2} \]

\[ = \frac{\bar{x}}{\frac{1}{n} + \frac{1}{\text{var}\bar{x} + M^2}} \]

\[ = \frac{\bar{x}}{\frac{1}{n} + \frac{1}{\text{var}\bar{x} + M^2}} \]

Recall that when \( h \) is non-random and known, \((\sigma^2_1 = \frac{1}{nh}\) and \(\text{var}\bar{x} = \sigma^2_o\) as in section 3.1), the partially Bayes rule is

\[ \frac{\bar{x}}{\sigma^2_1} + \frac{\Delta}{\sigma^2_o + M^2} \]

\[ = \frac{1}{\sigma^2_1} + \frac{1}{\sigma^2_o + M^2} \]
Observe that this is the rule of display (4.4.2) with \( \frac{1}{h} \) replaced by \( \frac{1}{h} \).

Knowledge of the parameter \( n' \) is not required if the loss function \( h(\theta(x) - \theta)^2 \) is used. The partially Bayes rule becomes

\[
\frac{\bar{x}}{n^{-1}v'} + \frac{\Delta}{n^{-1}v' + M^2}.
\]

\[
\frac{1}{n^{-1}v'} + \frac{1}{n^{*2}v' + M^2}.
\]

In (4.4.1) the expected loss is a decreasing function of \( n^* \). Thus if the incompleteness specification is to extend to \( n^* \), it must be of the form \( n^* \geq n_0 \) say, in which case the maximum risk occurs at \( n_0 \). The partially Bayes rule with respect to this incompleteness specification would be the rule (4.4.2) with \( n^* \) replaced by \( n_0 \). Since the precision of the prior distribution for \( \theta \) is \( hn^* \), and \( h \) is the precision of the sampling process, this incompleteness specification treats the prior information as equivalent to at least \( n_0 \) observations on the process.

For further comment on equivalent prior samples, see Winkler [1967].

### 4.5 Estimation of a Linear Function

With notation and assumptions as in chapter II, let

\[ W = (w_1, w_2, \ldots, w_p)^T \]

be a vector of non-zero constants, and suppose

\[ W^T\theta = \sum_{i=1}^{p} w_i \theta_i \]

is to be estimated. It is proposed to estimate \( W^T\theta \) by \( W^T\hat{\theta} \) where \( \hat{\theta} \in D_L \), and \( W^* = (w_1^*, w_2^*, \ldots, w_p^*)^T \).

Taking as loss function

\[ h(\theta(x) - \theta)^2 \]
\[ L(W^*, \delta, \theta, x, \mu) = \frac{(W^T \delta(x) - W^T \theta)^2}{W^* W} \]

the Bayes risk is

\[ E(\mu, W^*, \delta) = E L(W^*, \delta, \theta, x, \mu) \]

\[ = \frac{1}{W^* W} \sum_{i=1}^{P} \sum_{j=1}^{P} (w^*_i \delta_i(x) - w^*_i \delta_i) \sum_{i=1}^{P} \sum_{j=1}^{P} \left[ w^*_i \delta_i(x) - w^*_i \delta_i \right] \left[ w^*_j \delta_j(x) - w^*_j \delta_j \right] \]

\[ = \frac{1}{W^* W} \sum_{i=1}^{P} \sum_{j=1}^{P} \left[ \frac{w^*_i \delta_i(x) - \delta_i}{w^*_i} \right] w_i w_j \left[ \frac{w^*_j \delta_j(x) - \delta_j}{w^*_j} \right] \]

\[ = E \sum_{i=1}^{P} \sum_{j=1}^{P} \left[ \frac{w^*_i \delta_i(x) - \delta_i}{w^*_i} \right] k_{ij} \left[ \frac{w^*_j \delta_j(x) - \delta_j}{w^*_j} \right], \]

where \( K = (k_{ij}) = \frac{W^* W}{W^* W} \) is non-negative definite and symmetric.

Now recall that \( \delta(x) = Bx + C \), so that if the \( p \times p \) matrix \( Q \) is defined by

\[ q_{ij} = \frac{w^*_i}{w^*_i} \]

and the \( p \)-vector \( R \) by

\[ r_i = \frac{w^*_i}{w^*_i} \]

and if
\[ \gamma(x) = Qx + R, \]

then the Bayes risk, \( B(\mu, \gamma, \delta) \) becomes

\[ E[\gamma(x) - \bar{\gamma}]^TK[\gamma(x) - \bar{\gamma}], \]

where \( K \) is non-negative definite symmetric, and \( \gamma \in D_L \). Next find \( \gamma \in D_L \) which is partially Bayes for loss matrix \( K \); that is to determine the partially Bayes values of \( Q \) and \( R \). Given \( Q \) and \( R \), what is \( \gamma^T \delta \), the estimator of \( \gamma^T \delta \)? Observe that

\[ \gamma^T \delta(x) = \gamma^T(Bx + C) = W^T Qx + W^T R = W^T \gamma(x). \]

Thus to estimate the contrast \( \gamma^T \delta \), first estimate \( \bar{\gamma} \) by \( \gamma \in D_L \) which is partially Bayes for the loss matrix \( K = \frac{WW^T}{W^TW} \), and then estimate \( \gamma^T \delta \) by \( W^T \gamma(x) \).

It is appropriate to recall here the comment relating to the non-negative definiteness of \( K \) which appears at the end of section 2.3.
V. THE SCALE PROBLEM

In this chapter the problem of partially Bayes estimation of a univariate scale parameter is considered. Two cases are distinguished: the case of known mean and the case of a prior distribution on the mean. Again, Raiffa and Schlaifer [1961, chapters 7 and 11], supplies the distribution theory.

5.1 Mean Known

For the location problem, the distribution assumptions involved only moments and in particular did not assume Normality. Now, suppose that $X_1, X_2, \ldots, X_n$ are independent random variables with identical Normal distributions with mean 0 and variance $h^{-1} = \sigma^2$. Define

$$W = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

with the convention that $W = 0$ for $n = 0$. Then $(nh)W$ has the $\chi^2$ distribution with $n$ degrees of freedom.

Suppose $\sigma^2$ is to be estimated with loss function

$$\left[ \frac{\ell(w) - \sigma^2}{\sigma^2} \right]^2 = h^2(\ell(w) - h^{-1})^2$$
where \( w, \sigma^2, \) and \( h \) are realizations of the random variables \( W, \sigma^2, \) and \( \tilde{h} \). Thus a mis-estimation of small values of \( \sigma^2 \) is more costly than an equal mis-estimation of large values.

The natural conjugate prior distribution for \( \tilde{h} \) is the gamma-2 with probability density

\[
\mathcal{Y}_2(\tilde{h}|v', n') = (\text{const.})e^{-\frac{3}{2}hn'v'} \tilde{h}^{\frac{3}{2}n' - 1}.
\]

The following moments will be required:

\[
E(W|\tilde{h}) = \frac{1}{\tilde{h}} , \quad \text{var}(W|\tilde{h}) = \frac{2}{nh^2} ,
\]

(5.1.1)

\[
E(\tilde{h}) = \frac{1}{v'} , \quad \text{var}(\tilde{h}) = \frac{2}{n'v'^2} ,
\]

and thus

\[
E(\tilde{h}^2) = \left( \frac{1}{v'} \right)^2 \left( \frac{n' + 2}{n'} \right).
\]

Suppose that the prior distribution is known except for \( v' \), and that the decision maker has learned

\[
v' \in U = \{ v' \mid \frac{A}{\Delta} \leq v' \leq \Delta A \}
\]

where \( \Delta \geq 0, \ A \geq 1. \)

As before, considering only estimators which are linear in the sufficient statistic, without loss of generality write

\[
\hat{s}(w) = bw + \beta c w \Delta
\]

where \( \kappa = \frac{n'}{n' + 2} \) and put \( \bar{\kappa} = 1 - \kappa. \)

For a partially Bayes rule, \( b \) and \( c \) must minimize
\[ \sup_{v' \in U} E \tilde{h}^2(b W + b c x A - \tilde{h}^{-1})^2. \]

Using the moments (5.1.1), observe that

\[ E \tilde{h}^2(b W + b c x A - \tilde{h}^{-1})^2 = E \tilde{h}^2[b(W - \tilde{h}^{-1}) + b(c x A - \tilde{h}^{-1})]^2 \]

\[ = b^2 E \tilde{h}^2 \frac{2}{n \tilde{h}^2} + b^2 [(c x A)^2 \frac{1}{v^T} \tilde{h}^{-1} - 2c x A \frac{1}{v^T} + 1] \]

\[ = \frac{2b^2}{n} + \bar{\mu} \tilde{h}^2 [(c x A)^2 \frac{1}{v^T} - 2c x A \frac{1}{v^T} + 1] + \bar{\mu} \tilde{h}^2 \]

\[ = \frac{2b^2}{n} + \bar{\mu} \tilde{h}^2 + \bar{\mu} \tilde{h}^2 \max \{(\frac{c}{A} - 1)^2, (c A - 1)^2\}. \]

The last expression is a continuous convex function of \( \frac{1}{v^T} \), and

\[ U = \{ v' \mid \frac{A}{A} \leq v' \leq \Delta A \} = \{ v' \mid \frac{1}{\Delta A} \leq \frac{1}{v^T} \leq \frac{A}{A} \} \]

is convex and compact so

\[ B(b, c) = \sup_{v' \in U} E \tilde{h}^2(b W + b c x A - \tilde{h}^{-1})^2 \]

\[ = \frac{2b^2}{n} + \bar{\mu} \tilde{h}^2 + \bar{\mu} \tilde{h}^2 \max \{(\frac{c}{A} - 1)^2, (c A - 1)^2\}. \]

Constants \( b \) and \( c \) are sought to minimize \( B(b, c) \). Let

\[ \mathcal{B} = \{ c \mid (c A - 1)^2 \geq (\frac{c}{A} - 1)^2 \}. \]

Clearly, to minimize \( B(b, c) \) subject to \( c \in \mathcal{B} \), choose \( c = c_0 \) such that

\[ (c_0 A - 1)^2 = (\frac{c_0}{A} - 1)^2. \]

On the complement of \( \mathcal{B} \), no minimum is attained (the infimum is at \( c_0 \)). Thus \( B(b, c) \) is minimized over all \( b \) at the solution, \( c_0 \), of \( (c_0 A - 1)^2 = (\frac{c_0}{A} - 1)^2 \). Thus it remains to solve
for $c$. If $\frac{c}{A} > 1$, then since $A^2 \geq 1$, $cA = \frac{c}{A}A^2 > 1$ and (5.1.2) becomes

(5.1.3) $$cA - 1 = \frac{c}{A} - 1$$

which has no solution for $A > 1$. If $A = 1$, $v'$ is known, and the problem reduces to the Bayes problem. If $\frac{c}{A} \leq 1$ and $cA < 1$, (5.1.2) becomes (5.1.3) and the same comment applies. Finally, if $\frac{c}{A} \leq 1$ but $cA \geq 1$, then (5.1.2) becomes

$$cA - 1 = 1 - \frac{c}{A}$$

which has as solution

$$c_o = 2(A + \frac{1}{A})^{-1} = 2\frac{A}{A^2 + 1}.$$  

Thus

$$B(b, c_o) = \frac{2b^2}{n} + \bar{b}^2 + \frac{n}{2} \left[ \frac{2A^2}{A^2 + 1} - 1 \right]^2$$

(5.1.4)

$$= \frac{2b^2}{n} + \frac{\bar{b}^2}{n'} + \frac{2 + n'G^2}{2 + n'G^2}$$

where

$$G = \frac{A^2 - 1}{A^2 + 1}.$$  

Differentiating to minimize $B(b, c_o)$ over values of $b$, the minimizing $b$ is

(5.1.5) $$b_o = \frac{\frac{2 + n'G^2}{n' + 2}}{\frac{2}{n} + \frac{2 + n'G^2}{n' + 2}}.$$  

and the partially Bayes rule is
\[ \delta_\circ(w) = b_\circ w + \bar{b}_\circ \frac{2}{A^2 + 1} \left( \frac{n'}{n'} + \frac{1}{2n'} \right) \Delta \]

which is the Bayes rule (with no restriction on \( D \)),
\[ \delta(w) = \frac{w}{n'} + \frac{v'}{n} \cdot \frac{n'}{n'} + \frac{1}{n} \]

when \( A = 1 \). (The posterior distribution of \( h \) is gamma-2 with parameters
\[ n'' = n' + n \quad \text{and} \quad v'' = \frac{n w + n' v'}{n'} + \frac{1}{n} \]

The posterior mean is \( \frac{1}{v''} \), and the Bayes estimate of \( \sigma^2 \) for quadratic loss is \( v'' \). Recall that the loss function for this problem, \( h^2(\delta(w) - \frac{1}{h})^2 \) is not quadratic, which accounts for the apparent discrepancy.)

To determine the efficiency of the partially Bayes rule with respect to the Bayes rule, compute the risk of the partially Bayes rule by substituting \( b_\circ \) of (5.1.5) for \( b \) in \( B(b, c_\circ) \) of (5.1.4). The result is
\[ B(b_\circ, c_\circ) = \frac{2(2 + n' G^2)}{2(n' + 2) + n(2 + n' G^2)} \]

Note that this does not depend on \( \Delta \). The Bayes risk (of the Bayes rule) is obtained by substituting \( A = 1 \) (and therefore \( G = 0 \)) in \( B(b_\circ, c_\circ) \) giving
\[ \frac{2}{n + n' + 2} . \]

The efficiency of the partially Bayes rule with respect to the Bayes rule is taken to be the ratio of these two risks. So

\[
\text{Eff.} = \frac{\frac{2}{n + n' + 2}}{\frac{2(2 + n'g^2)}{2(n' + 2) + n(2 + n'g^2)}} = \frac{2(n' + 2) + n(2 + n'g^2)}{(n + n' + 2)(2 + n'g^2)}.
\]

Figures 1-4 give contours of the efficiency for sample sizes \( n = 2, 5, 10, 20 \). For given \( n \), the label of the contour at the point \((A, n')\) is the efficiency of the partially Bayes rule with respect to the Bayes rule when the degrees of freedom parameter of the prior distribution is \( n' \), and the incompleteness specification \( U \), is determined by \( A \). Only for large values of \( n' \), do increases in \( A \) significantly decrease the efficiency, and the importance of small values of \( A \) decreases as the sample size increases.

The graphs were obtained by evaluating the efficiency over a 2,250 point grid at \( A = 1(.2)9.8 \) and \( n' = 0(.2)9.8 \) on a CDC6400 computer, and having the contours smoothed and plotted on an EAI3500 Dataplotter.
Fig. 1.--Contours of: \( \frac{\text{risk of Bayes estimate}}{\text{risk of partially Bayes estimate}} \) for estimation of a scale parameter. The degrees of freedom in the prior distribution is \( n' \), the incompleteness specification is determined by \( A \), and the sample size is \( n = 2 \).
Fig. 2.--Contours of: \((\text{risk of Bayes estimate})/(\text{risk of partially Bayes estimate})\) for estimation of a scale parameter. The degrees of freedom in the prior distribution is \(n'\), the incompleteness specification is determined by \(A\), and the sample size is \(n = 5\).
Fig. 3.--Contours of: (risk of Bayes estimate)/(risk of partially Bayes estimate) for estimation of a scale parameter. The degrees of freedom in the prior distribution is $n'$, the incompleteness specification is determined by $A$, and the sample size is $n = 10$. 
Fig. 4.--Contours of: \( \frac{\text{risk of Bayes estimate}}{\text{risk of partially Bayes estimate}} \) for estimation of a scale parameter. The degrees of freedom in the prior distribution is \( n' \), the incompleteness specification is determined by \( A \), and the sample size is \( n = 20 \).
The results of this section extend immediately to the mean
non-zero. In section 5.2 the estimation of \( \sigma^2 \) when \( \theta \) is not known
is treated.

5.2 Mean Unknown

Recall the problem of section 4.4 in which \( X_1, X_2, \ldots, X_n \) are
independent, identically Normally distributed random variables with
mean \( \theta \) and variance \( h^{-1} \). Let \( \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \) and \( V = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \),
where \( V = 0 \) if \( n \leq 1 \). The joint density of \( \bar{X} \) and \( V \) is

\[
  f(\bar{X}, V | \theta, h) = \text{(const.)} \left[ e^{-\frac{3}{2} h n (\bar{X} - \theta)^2} \frac{1}{h^{\frac{3}{2}}} \right] \left[ e^{-\frac{3}{2} h v} \frac{1}{h^{\frac{3}{2}}} \right]
\]

where \( r = n - 1 \), and the natural conjugate prior for \( \bar{\theta} \) and \( \bar{\theta} \) is the
Normal-gamma with density

\[
  f_N(\theta, h | \mu, \nu, n^*, n') = f_N(\theta | \mu, h n^*)f_{\gamma_2}(h | \nu', n')
\]

\[
  = \text{(const.)} \left[ e^{-\frac{3}{2} h n^* (\theta - \mu)^2} \frac{1}{h^{\frac{3}{2}}} \right] \left[ e^{-\frac{3}{2} h \nu' v} \frac{1}{h^{\frac{3}{2}}} \right]
\]

for \( -\infty \leq \theta \leq \infty \), \( h \geq 0 \), and \( \nu', n', n' > 0 \).

The following moments will be required:

\[
  E(V | \bar{h}) = \bar{h}^{-1}, \quad \text{var}(V | \bar{h}) = \frac{2}{\bar{h}^{-2}},
\]

(5.2.1)

\[
  E(\bar{h}) = \frac{1}{V}, \quad \text{var}(\bar{h}) = \frac{2}{n' V^{-2}}.
\]

Suppose that \( h^{-1} \) is to be estimated by a function \( \delta \) of the
sufficient statistic, \( V \), with loss
\[ h^2(\delta(v) - n^{-1})^2 \]

as in the previous section. Assume that prior knowledge about \( v' \) is incomplete and that the decision maker has learned

\[ v' \in U = \{ v' \mid \frac{A}{A} \leq v' \leq \Delta A \}, \]

where \( \Delta \geq 0, A \geq 1. \)

Restricting the discussion to rules linear in \( V \), write

\[ \delta(v) = bv + \frac{bc}{n'} + 2\Delta \]

as in the previous section and determine \( b, c \) to minimize

\[ \sup_{v' \in U} E \, \tilde{h}^2(\delta(v) - \tilde{h}^{-1})^2 = \sup_{v' \in U} E \, \tilde{h}^2 \left[ bv + \frac{bc}{n'} + 2\Delta - \tilde{h}^{-1} \right]^2. \]

The expectations in display (5.2.1) are the same as those in (5.1.1) with \( W \) replaced by \( V \), and \( n \) in \( \text{var}(W|\tilde{h}) = \frac{2}{n\tilde{h}^2} \) by \( r = n - 1 \) in \( \text{var}(V|\tilde{h}) = \frac{2}{r\tilde{h}^2} \). Thus with these substitutions, the required computations are identical to those of section 5.1, and hence the partially Bayes rule when \( \tilde{h} \) is unknown is

\[ \delta(v) = b_1 v + \frac{b_1 2A}{A^2 + 1} \frac{n'}{n' + 2\Delta}, \]

where

\[ b_1 = \frac{2 + n'G^2}{n' + 2}, \]

and

\[ G = \frac{A^2 - 1}{A^2 + 1}. \]
Furthermore, the partially Bayes risk of the partially Bayes rule is, by comparison with (5.1.6)

\[
(5.2.2) \quad \frac{2(2 + n'G^2)}{2(n' + 2) + r(2 + n'G^2)}.
\]

It follows that the contours given in figures 1 - 4, of the efficiency of the partially Bayes rule with respect to the Bayes rule, can be used when \( \theta \) is unknown if \( n \) is replaced by \( r = n - 1 \).
VI. THE DESIGN PROBLEM

In previous chapters only terminal action problems have been considered. In this chapter, the design problem will be illustrated with examples from the univariate case.

6.1 The Univariate Location Problem

Consider the univariate location parameter problem with

\[ \begin{align*}
\mathbb{E}(X|\bar{y}) &= \bar{y}, & \text{var}(X|\bar{y}) &= n^{-1} \sigma^2, \\
\mathbb{E}(\bar{y}) &= \mu, & \text{var}(\bar{y}) &= \sigma^2.
\end{align*} \]

Considering only linear rules, \( \delta \in D_\mathcal{L} \), suppose that the terminal loss is \((\delta(x) - \theta)^2\) for taking action \( \delta(x) \) when \( \theta \) is the state of nature.

Now suppose that the cost of learning that \( |\Delta - \mu| \leq M \) is

\[ c(M) = \frac{\alpha^2}{M^2}, \quad \alpha \geq 0, \quad M \geq 0, \]

and that the sampling cost is proportional to the sample size, so that the cost of obtaining a sample of size \( n \) is

\[ c_s(X) = c_s n, \quad c_s \geq 0. \]

Assuming additivity of the various costs, the overall loss function for this problem is...
\[ L(\Delta, M, n, s, x, \theta) = \frac{\alpha^2}{M^2} + c_s n + (s(x) - \theta)^2, \]

and by the corollary of section 3.1, the partially Bayes expected loss of the partially Bayes rule is

\[ L''(M, n) = \inf \sup_{s \in D_L} \mathbb{E} L(\Delta, M, n, s, X, \theta) \]

\[ = \frac{\alpha^2}{M^2} + c_s n + \left[ \frac{n}{\sigma_1^2} + \frac{1}{\sigma_o^2 + M^2} \right]. \]

The design problem consists of finding \( M \) and \( n \) to make this a minimum.

Since \( \frac{\partial^2 L''(M,n)}{\partial^2 n^2} > 0 \) for \( n \geq 0 \), \( n \) may be treated as a continuous variable in seeking the minimum. If the minimizing value, \( n_o \), is not an integer, the optimal sample size is whichever of \( [n_o] \) and \( [n_o] + 1 \) makes \( L'' \) smaller. Let

\[ S = \{(\sigma_o, \sigma_1, c_s, \alpha) \mid \sigma_o > 0, \sigma_1 > 0, c_s > 0, \alpha > 0\}, \]

\[ \beta = \frac{\sigma_1}{c_s} \sqrt{\sigma_o} - \alpha, \]

and \( \{S_1, S_2, S_3\} \) be the partition of \( S \) defined by

\[ S_1 = \{(\sigma_o, \sigma_1, c_s, \alpha) \mid \sigma_o^2 \equiv \beta > 0\} \]

\[ S_2 = \{(\sigma_o, \sigma_1, c_s, \alpha) \mid \beta > \sigma_o^2 > 0\} \]

\[ S_3 = \{(\sigma_o, \sigma_1, c_s, \alpha) \mid \sigma_o^2 > 0 \equiv \beta\} \].
Theorem 6.1.1: With notation and assumptions as above,

\[
L^* = \inf_{M,n} L^* (M,n) = \begin{cases} 
2\sigma_1 c_s - \frac{\beta^2}{\sigma_0^2} & \text{on } S_1 \\
2\alpha + \sigma_0^2 & \text{on } S_2 \\
2\sigma_1 c_s & \text{on } S_3,
\end{cases}
\]

and the infimum is attained at

\[
n_0 = \begin{cases} 
\frac{\sigma_1 (\sigma_0^2 - \beta)}{\sigma_0^2 c_s} & \text{on } S_1 \\
0 & \text{on } S_2 \\
\frac{\sigma_1}{c_s} & \text{on } S_3
\end{cases}
\]

and

\[
\bar{M}_0 = \begin{cases} 
\frac{\alpha \sigma_0^2}{\beta} & \text{on } S_1 \\
\alpha & \text{on } S_2 \\
\infty & \text{on } S_3.
\end{cases}
\]

The (linear) partially Bayes rule is then
\[
\delta_0(x) = \begin{cases} 
\frac{1}{\sigma_0^2} \left[ (\sigma_0^2 - \beta) x + \beta \Delta \right] & \text{on } S_1 \\
\Delta & \text{on } S_2 \\
x & \text{on } S_3.
\end{cases}
\]

**Proof:** \(M\) and \(n\) are sought to minimize

\[
L^m(M, n) = \frac{\sigma_0^2}{M^2} + c_0 n + \frac{\sigma_1^2 (\sigma_0^2 + M^2)}{\sigma_1^2 + n (\sigma_0^2 + M^2)}.
\]

Setting the partial derivatives with respect to \(M\) and \(n\) equal to zero, it follows after some simplification, that the critical equations are

(6.1.1) \[ M^2 = \frac{\alpha(n \sigma_0^2 + \sigma_1^2)}{\sigma_1^2 - \alpha n}, \quad n < \frac{\sigma_1^2}{\alpha} \]

and

(6.1.2) \[ (\sqrt{c_s} n - \sigma_1)(\sigma_0^2 + M^2) + \sigma_1^2 \sqrt{c_s} = 0. \]

Substituting (6.1.1) into (6.1.2) gives

(6.1.3) \[ n_0 = \frac{\sigma_1 (\sigma_0^2 - \beta)}{\sigma_0^2 \sqrt{c_s}}, \quad \beta \leq \sigma_0^2, \]

and substituting \(n_0\) back in (6.1.1) gives

(6.1.4) \[ M_0^2 = \frac{\alpha \sigma_0^2}{\beta}. \]

With some computation, it can be shown that \(n_0 < \frac{\sigma_1^2}{\alpha}\) if and only if \(\beta > 0\), and that the matrix of second derivatives is positive definite.
so that the solution is in fact a minimum. So in the case \( \sigma^2_o \geq \beta > 0 \), that is, in \( S_1 \), a solution is given by (6.1.3) and (6.1.4). By a continuity argument, take \( n_o = 0 \) if \( \beta > \sigma^2_o \) [see display (6.1.3)]; that is, in \( S_2 \). In this case, by substitution in (6.1.1) it follows that \( M^2_o = \alpha \). Finally, if \( \beta \leq 0 \) (in \( S_3 \)), take \( M^2_o = \infty \) [see display (6.1.4)], and so \( L^s(\alpha, n) = c_s n + \frac{1}{n} \), which is minimized for \( n_o = \frac{1}{\sqrt{c_s}} \).

The proof is completed by substitution of \( n_o \) and \( M^2_o \) into \( L^s(M, n) \) to compute \( L^s \) and into

\[
8_o(x) = \frac{nx + \frac{\Delta}{\sigma^2_o + M^2}}{\frac{1}{n} + \frac{1}{\sigma^2_o + M^2}}.
\]

A few remarks about the regions \( S_2 \) and \( S_3 \) are in order. First, note that \( \beta = \sigma_c \frac{1}{\sqrt{c_s}} - \alpha \) increases with the sampling variance and the sample size cost parameter \( c_s \), and decreases with \( \alpha \), the learning cost parameter. The region \( S_2 \) has \( \beta > \sigma^2_o \), and so with respect to the prior variance, sampling is imprecise or costly, or it is inexpensive to improve the specification of the prior distribution. In such a situation it seems reasonable not to sample, and indeed this is the case. In \( S_3 \), \( \beta \leq 0 < \sigma^2_o \), so with respect to \( \sigma^2_o \), sampling is inexpensive or precise, or it is costly to buy prior information. In this case it would be appropriate not to improve the prior and thus \( \hat{M} = \infty \). The corresponding decision rules are as one would expect. Finally:
Corollary 6.1.2: With notation and assumptions as in theorem 6.1.1, the sample size required for the partially Bayes procedure does not exceed the sample size required for the procedure, \( \delta(x) = x \).

Proof: For the procedure with rule \( \delta(x) = x \), the risk is

\[
c_s n + \mathbb{E}[(X - \theta)^2 \mid \theta] = c_s n + \frac{\sigma^2}{n}.
\]

Minimizing on \( n \), the optimal sample size \( n_1 \), is

\[
n_1 = \frac{\sigma}{\sqrt{c_s}}.
\]

From theorem 6.1.1, in each of \( S_1 \), \( S_2 \), and \( S_3 \)

\[
n_0 = \frac{\sigma}{c_s} \left[ 1 - \frac{\alpha}{M_e^0} \right] \leq n_1.
\]

6.2 The Scale Problem

Recall that in estimating a scale parameter with location parameter unknown and loss structure as in section 5.2, the partially Bayes risk of the partially Bayes rule is, from (5.2.2):

\[
\frac{2(2 + n'G^2)}{2(n' + 2) + r(2 + n'G^2)},
\]

where \( r = n - 1 \) and \( G = \frac{A^2 - 1}{A^2 + 1} \).

The design problem is that of determining \( n \), the sample size and \( A \), the determinant of the incompleteness specification. Postulate a cost \( \varphi(A) \) of learning \( \frac{A}{A} \leq v' \leq \Delta A \) and a cost \( c_s r = c_s (n - 1) \), \( c_s \geq 0 \),
of obtaining a sample of size \( n \). (This choice of sampling cost gives
the same minimizing \( n \) as \( c_s \cdot n \) but simplifies the computations.) Again
assume that all costs are additive so that the design problem is to
find \( r \) and \( A \) to minimize

\[
\varphi(A) + c_s r + \frac{2(2 + n'g^2)}{2(n' + 2) + r(2 + n'g^2)}
\]

For this example a convenient choice for \( \varphi \) is to set
\( t = \frac{2}{n' + 2} \),
\( \bar{t} = 1 - t \), and \( y = t + \bar{t}g^2 \), and choose

\[
\varphi(A) = \frac{\varphi}{y^2}, \quad \varphi \geq 0.
\]

The loss function becomes

\[
L^s(y,r) = \frac{\varphi}{y^2} + c_s r + \frac{2y}{2 + ry},
\]

and this is to be minimized over values of \( y \) and \( r \) such that \( t \leq y \leq 1 \),
\( r \geq -1 \). As in section 6.1, \( r \) may be treated as continuous. Let

\[
S = \{(t,\varphi,c_s) \mid 0 \leq t \leq 1, \varphi \geq 0, \text{ and } c_s \geq 0\},
\]

\[
S_1 = \{(t,\varphi,c_s) \in S \mid t \leq \frac{\varphi}{c_s} \leq 1\}
\]

\[
S_2 = \{(t,\varphi,c_s) \in S \mid \frac{\varphi}{c_s} < t\}
\]

\[
S_3 = \{(t,\varphi,c_s) \in S \mid \frac{\varphi}{c_s} > 1\}
\]

\[
T_1 = \{(t,\varphi,c_s) \in S \mid \sqrt{\frac{2}{c_s} - \frac{2c_s}{\varphi}} \geq -1\}
\]
\[ T_2 = \{ (t, \varphi, c_s) \in S \mid \sqrt{\frac{2}{c_s}} - \frac{2}{t} \geq -1 \} \]

\[ T_3 = \{ (t, \varphi, c_s) \in S \mid \sqrt{\frac{2}{c_s}} - 2 \geq -1 \} = \{ c_s \mid c_s \leq 2 \}. \]

Finally, let \( T_j \) be the complement (in \( S \)) of \( T_j \), \( j = 1, 2, 3 \), and note that \( S = S_1 \cup S_2 \cup S_3 \).

**Theorem 6.2.1:** With notation as above

\[
\inf_{y, r} L^*(y, r) = \begin{cases} 
\varphi + 2 - \frac{c_s}{t} & \text{on } S_3 \cap \overline{T}_3 \\
\varphi + 2/2c_s - 2c_s & \text{on } S_3 \cap T_3 \\
\phi + 2/2c_s - \frac{2c_s}{t} & \text{on } S_2 \cap T_2 \\
\phi + 2/2c_s - \frac{2c_s}{t} - c_s & \text{on } S_2 \cap \overline{T}_2 \\
\phi + \frac{2t}{t^2} - c_s & \text{on } S_2 \cap \overline{T}_2 \\
\frac{c_s^2}{\varphi} + \frac{2\varphi}{2c_s - \varphi} - c_s & \text{on } S_1 \cap \overline{T}_1 \\
\frac{2\sqrt{2c_s}}{\varphi} - \frac{c_s}{\varphi} & \text{on } S_1 \cap T_1 
\end{cases}
\]

and is attained at
\[ r_0 = \begin{cases} 
\frac{\sqrt{2}}{c_s} - \frac{2c_s}{\varphi} & \text{on } S_1 \cap T_1 \\
\frac{\sqrt{2}}{c_s} - \frac{2}{t} & \text{on } S_2 \cap T_2 \\
\sqrt{2} & \text{on } S_3 \cap T_3 \\
-1 & \text{elsewhere on } S 
\end{cases} \]

and

\[ y_0 = \begin{cases} 
\frac{\varphi}{c_s} & \text{on } S_1 \\
t(A = 1) & \text{on } S_2 \\
l(A = \infty) & \text{on } S_3.
\end{cases} \]

Proof: Differentiating \( L''(y, r) \) with respect to \( y \) and \( r \) and setting the derivatives equal to zero gives

\[(6.2.1) \quad \frac{y_0^3}{(2 + y_0 r_0)^2} = \frac{\varphi}{2}, \]

and

\[(6.2.2) \quad \frac{y_0^2}{(2 + y_0 r_0)^2} = \frac{c_s}{2}. \]

Substituting (6.2.2) into (6.2.1) gives

\[ \frac{c_s}{y_0 \frac{c_s}{2}} = \frac{\varphi}{2}, \]
and thus

(6.2.3) \[ y_o = \frac{\varphi}{c_s}. \]

Substituting (6.2.3) into (6.2.2) gives

\[ \frac{\varphi}{c_s} \frac{c_s}{2 + r_o \frac{\varphi}{c_s}} = \sqrt{\frac{c_s}{2}} \]

and thus

(6.2.4) \[ r_o = \frac{\varphi - 2}{\varphi \sqrt{\frac{c_s}{2}}} \left( \frac{c_s}{2} \right) = \sqrt{\frac{2}{c_s}} - \frac{c_s}{\varphi}. \]

With some computation it can be shown that the matrix of second derivatives is positive definite at the solution. Now it is required that \( t \leq y \leq 1 \) and \( r \geq 1 \), so that (6.2.3) and (6.2.4) apply on \( S_1 \cap T_1 \).

By continuity, on \( S_1 \cap T_1 \) take \( y_o = \frac{\varphi}{c_s} \) but \( r_o = -1 \). On \( S_2 \), take \( y_o = t \), and substitute in (6.2.2) to obtain

\[ \frac{t}{2 + r_0 t} = \sqrt{\frac{c_s}{2}} \]

and thus

\[ r_o = \sqrt{\frac{2}{c_s}} - \frac{2}{t}, \]

which applies on \( T_2 \). Take \( r_0 = -1 \) on \( S_2 \cap T_2 \). Finally, on \( S_3 \) take \( y_0 = 1 \), which gives

\[ r_o = \sqrt{\frac{2}{c_s}} - 2, \]
and which applies on $T_3$. Take $r_o = -1$ on $S_3 \cap \overline{T_3}$. The minimum values of $L^*(y,r)$ are obtained by substitution.

Conclusions similar to those of the previous section can be drawn concerning the optimal $y$ and $r$ in the several regions. Recall that $y = t$ implies that $A = 1$ and $y = l$ implies $A = \infty$.

A final observation is appropriate.

**Corollary 6.2.2:** The optimal sample size for the partially Bayes estimate does not exceed the optimal sample size for the estimate $s(v) = v$, where $v$ is an observed value of

$$v = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

**Proof:** Note that for the partially Bayes procedure

$$r_o = \begin{cases} \sqrt{\frac{2}{c_s}} \frac{2}{y_o} & \text{on } S_1 \cap T_1, S_2 \cap T_2, S_3 \cap T_3 \\ -1 & \text{elsewhere on } S, \end{cases}$$

and that $y_o \geq 0$.

For the procedure $s(v) = v$, the cost is

$$c_s n + Eh^2 (v - \frac{1}{n})^2 = c_s n + \frac{2}{n - 1}.$$

The minimizing $n$ is

$$n_1 = 1 + \sqrt{\frac{2}{c_s}},$$

or equivalently
\[ r_1 = n_1 - 1 = \sqrt{\frac{2}{c_s}}. \]

It follows that \( r_0 \leq r_1 \), and the corollary is proved.
Theorem A.1: If \( f \) is a convex and continuous function on a (non-empty) convex, compact set \( S \subseteq \mathbb{E}^P \), then \( f \) assumes its maximum on \( S \) at an extreme point of \( S \). If \( f \) is strictly convex, the only maxima occur at extreme points.

Proof: Since \( f \) is continuous and \( S \) is compact, \( f \) assumes a maximum on \( S \). Suppose that this maximum occurs at a point \( Y_o \in S \) which is not an extreme point. Then there exist extreme points \( Y_1, Y_2, \ldots, Y_n \) (\( n \leq p + 1 \)) and real numbers \( c_1, c_2, \ldots, c_n \) with \( c_i \geq 0 \) and \( \sum_{i=1}^{n} c_i = 1 \) such that \( Y_o = \sum_{i=1}^{n} c_i Y_i \) (Rao [1965, p. 42, proposition 1]). Without loss of generality we can assume that

\[
f(Y_1) = \max_{1 \leq i \leq n} f(Y_i).
\]

Then

\[
f(Y_0) \leq \sum_{i=1}^{n} c_i f(Y_i) \leq \sum_{i=1}^{n} c_i f(Y_1) = f(Y_1),
\]

the first inequality by the convexity of \( f \), the second by the maximality of \( f(Y_1) \). Since \( Y_0 \) maximizes \( f \), \( f(Y_0) \geq f(Y_1) \), so \( f(Y_0) = f(Y_1) \) and it has been shown that the maximum is attained at an extreme point.

If \( f \) is strictly convex, the first inequality in (1) is strict which contradicts the maximality of \( f(Y_0) \). Thus any maximum must occur at an extreme point.
Corollary A.2: A non-negative definite quadratic form \( f(X) = X^T B X \) on a (non-empty) convex, compact set \( S \subseteq \mathbb{R}^p \) attains its maximum (on \( S \)) at an extreme point of \( S \).

Proof: By theorem A.1, since \( f \) is continuous on \( \mathbb{R}^p \), it will suffice to show that \( f \) is convex on \( \mathbb{R}^p \). Observe that the matrix of second partial derivatives,

\[
\left( \frac{\partial^2 f(X)}{\partial x_i \partial x_j} \right) = 2B
\]

is non-negative definite. The result is immediate, (Saaty [1959, p. 131]).
REFERENCES


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


Stone, M. [1963]. Robustness of non-ideal decision procedures. 
J. Amer. Statist. Assoc. 58, 480-486.

Thompson, J. R. [1968]. Some shrinkage techniques for estimating 

Winkler, R. [1967]. The assessment of prior distributions in 
UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

1. Originating Activity
   The Florida State University
   Department of Statistics
   Tallahassee, Florida

   2a. Report Security Classification
       Unclassified

   2b. Group

3. Report Title
   Partially Bayes Estimates

4. Descriptive Notes

5. Author(s)
   Solomon, Daniel L.

6. Report Date
   January, 1969

   7a. Total No. of Pages
       84

   7b. No. of References
       16

8a. Contract or Grant No.
    NONR 988(13)

   b. Project No.
   c. Task No.
      NR 042-003

   d.

9a. Originator's Report Number(s)

   9b. Other Report No(s).
      FSU Statistics Report M152

10. Availability/Limitation Notices
    Releasable without limitations on dissemination.

11. Supplementary Notes

12. Sponsoring Military Activity
    Logistics and Mathematical Statistics
    Branch
    Office of Naval Research
    Washington, D. C. 20360

13. Abstract
    Statistical decision problems are considered in which the decision maker is
    assumed to have prior information but cannot completely specify a prior distribution.
    The decision maker's prior knowledge is reflected in his willingness to specify a subset, \( \Lambda^* \) (called an incompleteness specification) of the class of all prior distributions \( \Lambda \). He is then recommended to select the decision rule to minimize the maximum over
    distributions in \( \Lambda^* \) of the Bayes risk. Such a rule is called partially Bayes with
    respect to \( \Lambda^* \), and reduces to the Bayes rule with respect to \( \lambda \) if \( \Lambda^* = \{\lambda\} \) and the
    minimax rule if \( * = \). The particular problems of estimation of a general mean
    and a Normal variance are considered in detail. Examples of the determination of
    optimal sample size and incompleteness specification are given for the two problems.

14. Key Words

<table>
<thead>
<tr>
<th>Link A</th>
<th>Link B</th>
<th>Link C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role</td>
<td>Wt.</td>
<td>Role</td>
</tr>
<tr>
<td>Bayes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision theory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incomplete prior</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimax</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial prior</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>