ESTIMATING THE PROBABILITY OF MISCLASSIFICATION*

by

Marilyn J. Sorum
Technical Report No. 110

University of Minnesota
Minneapolis, Minnesota

* This research was supported by National Science Foundation Grant GP-6859
ACKNOWLEDGEMENT

The author wishes to express her gratitude to Professor Robert Buehler for providing guidance and continuing encouragement during the course of this research. This research was supported by National Science Foundation Grant GP-6859.
**TABLE OF CONTENTS**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>vi</td>
</tr>
<tr>
<td><strong>Chapter 1. Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1. The Two-group Classification Problem and the Various Probabilities</td>
<td>1</td>
</tr>
<tr>
<td>of Misclassification</td>
<td></td>
</tr>
<tr>
<td>2. Survey of the Literature Related to Estimating the Probabilities</td>
<td>6</td>
</tr>
<tr>
<td>of Misclassification</td>
<td></td>
</tr>
<tr>
<td>**Chapter 2. Description of Estimators of the Conditional Probability</td>
<td>10</td>
</tr>
<tr>
<td>of Misclassification</td>
<td></td>
</tr>
<tr>
<td>1. Estimators Not Using Distribution Assumptions</td>
<td>10</td>
</tr>
<tr>
<td>1.1 The reclassification estimator, $P_R$</td>
<td>10</td>
</tr>
<tr>
<td>1.2 The test sample estimator, $P_T$</td>
<td>11</td>
</tr>
<tr>
<td>1.3 The two-straight lines estimator, $P_{TSL}$</td>
<td>12</td>
</tr>
<tr>
<td>1.4 Lachenbruch's U method estimator, $P_U$</td>
<td>15</td>
</tr>
<tr>
<td>2. Estimators Which Make Use of Distribution Assumptions</td>
<td>16</td>
</tr>
<tr>
<td>2.1 The reclassification estimator adjusted for bias, $P_{Ra}$</td>
<td>16</td>
</tr>
<tr>
<td>2.2 The expression for the conditional probability of misclassification</td>
<td>17</td>
</tr>
<tr>
<td>with parameters replaced by estimates</td>
<td></td>
</tr>
<tr>
<td>2.3 The expression for the true unconditional probability of misclassi-</td>
<td>17</td>
</tr>
<tr>
<td>ification with parameters replaced by estimates</td>
<td></td>
</tr>
<tr>
<td>2.4 The expression for the true probability of misclassification when</td>
<td>17a</td>
</tr>
<tr>
<td>all parameters are known, with parameters replaced by estimates</td>
<td></td>
</tr>
<tr>
<td>2.5 Minimum variance unbiased estimators</td>
<td>17a</td>
</tr>
<tr>
<td>2.6 Estimators involving a prior distribution on $\theta$</td>
<td>17b</td>
</tr>
<tr>
<td>**Chapter 3. Estimating the Conditional Probability of Misclassification</td>
<td>21</td>
</tr>
<tr>
<td>in the Univariate Normal Case</td>
<td></td>
</tr>
<tr>
<td>1. The Classification Rule and the Expressions for the Various Probabi-</td>
<td>21</td>
</tr>
<tr>
<td>lities of Misclassification</td>
<td></td>
</tr>
<tr>
<td>2. The Estimators</td>
<td>25</td>
</tr>
<tr>
<td>2.1 The reclassification estimator adjusted for bias, $P_{Ra}$</td>
<td>25</td>
</tr>
</tbody>
</table>
2.2 The true conditional probability of misclassification with parameters replaced by estimates, \( Q_D \) and \( Q_T \)

2.3 The expression for the true unconditional probability of misclassification with parameters replaced by estimates, \( Q_D^* \)

2.4 The expression for the true probability of misclassification when all parameters are known, with parameters replaced by sample values, \( Q_D^{**} \)

2.5 Minimum variance unbiased estimator, \( \hat{\theta}_{RB} \)

2.6 Estimators involving a prior distribution on \( \theta \), \( \hat{\theta}_3 \) and \( \hat{\theta}_5 \)

3. Conditional Means and Variances

4. Asymptotic Expressions for Conditional Squared Bias, Variance, and Mean Square Error

4.1 Methods

4.2 The results on asymptotic conditional bias, variance, and mean square error

4.3 Conclusions

4.4 Remarks on the distinction between estimators of \( P_2 \), \( P_{2*} \), and \( P_{2**} \)

Table I. Notation Used in the Discussion of the Univariate Normal Case

Table II. Expressions for Estimators of \( P_2(R(x_1, x_2), \mu_2) \)

Table III. Expressions of Estimators for \( P_2(R(x_1, x_2), \mu_2) \)

Table IV. Conditional Means and Variances for Estimators of \( P_2 \) when \( \pi_1 \) is \( N(\mu_1, 1) \) and \( \pi_2 \) is \( N(\mu_2, 1) \)

Table V. Expressions and Expansions Needed to Get the Asymptotic Expressions for Conditional Bias, Variance, and Mean Square Error Given in Tables VI, VII, and VIII

Table VI. Squared Biases for Estimators of \( P_2(R(x_1, x_2), \mu_2) \)

Table VII. Conditional Variances for Estimators of \( P_2(R(x_1, x_2), \mu_2) \) when \( \pi_1 \) is \( N(\mu_1, 1) \) and \( \pi_2 \) is \( N(\mu_2, 1) \): Asymptotic Expressions
Table VIII. Conditional Asymptotic Mean Square Errors for Estimators of $P_2(R(x_1, x_2), \mu_2)$ when $\Pi_1$ is $N(\mu_1, 1)$ and $\Pi_2$ is $N(\mu_2, 1)$.

Table IX. Numerical Values for Terms of Order $o_p(N^{-1+\epsilon})$ in the AMSE's of Table VIII, for Selected Values of $\delta$ and $\sqrt{N} \alpha_2$.

Chapter 4. Estimating the Conditional Probability of Misclassification in the Multivariate Normal Case

1. The Classification Rule and the Various Probabilities of Misclassification

2. The Estimators
   2.1 The reclassification estimator adjusted for bias, $P_{Ra}$
   2.2 The expression for the conditional probability of misclassification with $\mu_2$ estimated, $Q_D$ and $Q_T$
   2.3 The expression for the unconditional probability of misclassification with parameters replaced by samples values, $Q^*_0$ and $Q^*_S$
   2.4 The expression for the true probability of misclassification when all parameters are known, with parameters replaced by sample values, $Q^{**S}$
   2.5 Minimum variance unbiased estimators, $\theta_{RB}$
   2.6 Estimators involving a prior distribution on $\theta$, $\theta_3$ and $\theta_5$

3. Conditional Means and Variances

4. Asymptotic Expressions for MSE and Discussion of Results

5. Conclusions on the Problem of Estimating $P_2(R(x_1, x_2), \mu_2)$ when the Populations are p-Dimensional Normal. Comparisons With Other Work on the Problem.
   5.1 Conclusions
   5.2 Comparisons with other work on the estimation of the conditional probability of misclassification for normal populations

Table X. Notation Used in the Discussion of the Multivariate Normal Case: $\Pi_1$ is $N(\mu_1, \Sigma)$ and $\Pi_2$ is $N(\mu_2, \Sigma)$.

Table XI. List of Estimators for $P_2(R(x_1, x_2), \mu_2)$ when $\Pi_1$ is $N_p(\mu_1, \Sigma)$ and $\Pi_2$ is $N_p(\mu_2, \Sigma)$. 

- iii -
Table XII. Conditional Means and Variances for Estimators of $P_2$ when $\pi_1$ is $N(\mu_1, \Sigma)$ and $\pi_2$ is $N(\mu_2, \Sigma)$

Table XIII. Asymptotic Expansions Needed to Get the Asymptotic Expressions for Conditional Squared Bias, Variance and Mean Square Error That are Given in Tables XIV, XV, and XVI.

Table XIV. Conditional Squared Biases for Estimators of $P_2(R(\bar{x}_1, \bar{x}_2), \mu_2) = 1 - \gamma(C)$ when $\pi_1$ is $N_p(\mu_1, \Sigma)$ and $\pi_2$ is $N_p(\mu_2, \Sigma)$: Exact and Asymptotic Expressions

Table XIV. Asymptotic Expressions for Conditional Variances of Estimators of $P_2(R(\bar{x}_1, \bar{x}_2), \mu_2)$ when $\pi_1$ is $N_p(\mu_1, \Sigma)$ and $\pi_2$ is $N_p(\mu_2, \Sigma)$

Table XVI. Asymptotic Expressions for Conditional Mean Square Errors of Estimators of $P_2(R(\bar{x}_1, \bar{x}_2), \mu_2)$ when $\pi_1$ is $N(\mu_1, \Sigma)$ and $\pi_2$ is $N(\mu_2, \Sigma)$

Chapter 5. Estimation of a Quantity of the Form $\phi(R(x), \theta)$

1. Situation
2. Cramer-Rao Type Bound
   2.1 Derivation of the bound
   2.2 Application of the bound to the case where $X_1$ is $N(\mu, 1)$ and $R(x) = \Sigma c_i x_i$
   2.3 Bound for estimators of the conditional probability of misclassification in the univariate normal classification problem
   2.4 Application of the bound in the normal case with $R(x) = \Sigma c_i x_i$ and $\phi = k_1 R + k_2 \mu$ or $\phi = k_3 \mu$
3. Chapman-Robbins Type Bound
   3.1 Derivation of a Chapman-Robbins type bound
   3.2 Chapman-Robbins type bound applied to a normal example
4. Rao-Blackwell Type Theory
   4.1 Definitions
   4.2 Relation between sufficiency and conditional sufficiency, and between complete sufficiency and conditional complete sufficiency
   4.3 Rao-Blackwell theory
   4.4 Application of the theory to the multivariate normal classification problem
Table XVII. Numerical Values for: i) the Cramer-Rao Bound on the Conditional Variance of Estimators of $P_2(R(x_1, x_2), \mu_2)$, ii) the Conditional Variance of the (Conditional) UMVU Estimator, $\mathcal{R}_B$ and iii) the Conditional Variance of the Test Sample Estimator, $P_T$

Appendix I. Derivations to Obtain the Estimators of $P_2(R(x_1, x_2), \mu_2)$ That Involve a Prior Distribution

Appendix II. Computations for the Conditional Means and Variances of Estimators of the Conditional Probability of Misclassification

Appendix III. Derivations Related to the Asymptotic Results in the Univariate Normal Case

Appendix IV. Derivations Related to the Asymptotic Results in the Multivariate Normal Case

Appendix V. Unconditional Mean Square Error Expressions Needed for the Discussion of the Distinction Between Estimators of $P_2$, $P^*$, $P^{**}$

Appendix VI. The Problem of a Bound on the Asymptotic Conditional Squared Bias

Bibliography
SUMMARY

An observation assumed to come from one of two populations, \( \Pi_1 \) or \( \Pi_2 \), is to be classified. The problem considered in this dissertation is that of estimating the conditional probability of misclassifying the observation, given a fixed classification rule based on samples from \( \Pi_1 \) and \( \Pi_2 \). In Chapter 1 the problem is introduced and in Chapter 2 various estimators are described. In Chapters 3 and 4 these estimators are studied in the cases where \( \Pi_1, \Pi_2 \) are univariate and multivariate normal with unknown means but known variances and covariances. Comparisons between estimators are based mainly on asymptotic conditional biases, variances and mean square errors. The results may be summarized roughly and briefly as follows:

i) among the estimators studied those which make use of the normality are better than those which do not;

ii) among the estimators studied those which require additional observations ("test samples") beyond those used to obtain the classification rule ("original samples") are no better than estimators based only on the original samples;

iii) all the estimators studied that are based only on original samples have equal leading terms in the asymptotic expression for the conditional squared bias. Since, as is shown, any estimator based only on original samples is biased (conditional), the equivalence of all such estimators studied (in terms of asymptotic conditional bias) leads to a conjecture on a lower bound on the asymptotic conditional squared bias. A more complete discussion of the asymptotic results is given in Chapter 3, Section 4 and in Chapter 4, Sections 4 and 5. The conjecture is discussed in Appendix VI.

In Chapter 5 it is noted that the conditional probability of misclassification has the form \( \varphi(R(x), \theta) \), where \( \theta \) is a parameter...
and \( R(\tilde{x})\) is a function of the observation \( \tilde{x} \). Cramér-Rao type bounds are derived for estimators of \( \varphi(\tilde{R}(x), \theta) \). A generalization of Rao-Blackwell theory is indicated and is used to get a conditionally UMVU estimator, based on test samples, for the conditional probability of misclassification in the normal case.

Remark on notation: Notation used in Chapters 3 and 4 (univariate and multivariate normal populations) is summarized in Tables I and X respectively. The equation \((u,v)\) is the \( v^{th} \) equation of Section \( u \) within a given chapter (or appendix). When an equation in a different chapter is referred to, the chapter number is explicitly given.
CHAPTER 1
INTRODUCTION

1. The Two-Group Classification Problem and the Various Probabilities of Misclassification.

Let \( x \) be a random observation known to come from one of two populations, \( \Pi_1, \Pi_2 \). The problem is to classify \( x \) as being from \( \Pi_1 \) or \( \Pi_2 \). Three situations, determined by the amount of information available on the distribution of observations from \( \Pi_i, i = 1, 2 \), are readily distinguished, and are listed below for the case in which densities exist:

(i) \( \Pi_1, \Pi_2 \) have known densities \( f(y; \theta_1), f(y; \theta_2) \) with \( \theta_1, \theta_2 \) known,

(ii) the form of the density \( f(y; \theta_1) \) is known with \( \theta_1, \theta_2 \) unknown,

(iii) the form of the density is unknown.

In the first situation, assuming \( P \{ f(x; \theta_1) \geq c \} = 0 \), the following rules form a complete class of admissible rules (Anderson, 1958, Chapter 6):

\[
\begin{align*}
\text{if } & \frac{f(x; \theta_1)}{f(x; \theta_2)} \geq c, \text{ classify } x \text{ as } \Pi_1 \\
\text{if } & \frac{f(x; \theta_1)}{f(x; \theta_2)} < c, \text{ classify } x \text{ as } \Pi_2
\end{align*}
\]

(1.1)

If the prior probabilities of drawing an observation from \( \Pi_1, \Pi_2 \) and the costs of misclassification are known, then \( c \) is a function

\footnote{Some writers call this problem the discrimination problem. Kendall (1966), for example, refers to this problem as discrimination and reserves the term classification for the situation in which there is no beforehand knowledge that the observations fall into groups. Hills (1966) uses the term allocation to mean what we mean by classification.}
of the prior probabilities and the costs, and the rule defined by (1.1) minimizes the expected cost. In the special case of equal prior probabilities and equal misclassification costs, \( c = 1 \). If no information is available on prior probabilities and costs, \( c \) can be chosen to meet some criterion on the probability of misclassification. For example, \( c \) could be chosen to make the probability of misclassifying an observation from \( \Pi_1 \) equal to the probability of misclassifying an observation from \( \Pi_2 \). If costs are known but prior probabilities are unknown, \( c \) might be chosen to make the two conditional expected costs equal. In any case, the classification problem is essentially solved for situation (i), the only problem being what criterion to use for selecting the constant \( c \).

In the situation of unknown parameters, the procedure given by (1.1) is often used with the parameters \( \theta_1, \theta_2 \) replaced by sample values, perhaps the maximum likelihood estimates. Other procedures have been suggested, such as the likelihood ratio procedure, given by Anderson (1958, p. 141) for multivariate normal populations, and the Bayes procedures involving a prior distribution on the unknown parameters as in the work of Geisser (1964), (1966). No attempt will be made here to survey all the relevant literature. Anderson (1958, Chapter 6) is a basic reference on the classification problem. Hodges (1950) gives a historical survey of the discrimination-classification problem and provides a bibliography, having more than 250 entries, including many papers on the application of the linear discriminant function. There is another survey article by Tatsuoka and Tiedeman (1954) and there is a bibliography of about 250 references by Posten (1962). A number of references from 1962 and later dates are included in a separate bibliography of this dissertation.
Some work on non-parametric classification has been done, providing methods suitable for situation (iii) in which nothing is known about the distributions. There are papers by Fix and Hodges (1951a), (1951b), (1952), Stoller (1954), Matusita (1956), Johns (1961), Hudimota (1964), and Kendall (1966), all of which appear in the statistical literature. In addition there is work on non-parametric classification appearing in the engineering literature under the names of pattern recognition or pattern classification. Some references are Kanal (1962), Nilsson (1965), and Rosen (1967). Van Ryzin (1965), (1966) has done work on non-parametric Bayesian pattern classification.

The present thesis is primarily concerned with estimation of the probability of misclassification in case (ii) and, to a lesser extent, in case (iii). In either case in order to classify an observation \( x \), some information must be obtained about the unknown parameters or about the unknown form of the distribution by sampling each population. A classification rule is based on the data in these samples. Thus, assume that random samples \( S_1, S_2 \) of sizes \( N_1, N_2 \) are taken from populations \( \Pi_1, \Pi_2 \) respectively and let \( R(S_1, S_2) \) be any classification rule based on these samples. Conceptually the simplest way to determine properties of \( R(S_1, S_2) \) is to use it on future observations whose true classification is known. Thus we will distinguish between "original samples" \( S_1, S_2 \) used to obtain the classification rule and "test samples" which are additional observations of known classification not so used. Some of the estimators we consider require test samples.

Conditional on an observation coming from a given population, there are actually three different kinds of probabilities of misclassification that might be considered. Even if densities are completely known (as in situation (i)), the rule as given in (1.1) will not in
general classify perfectly and there will be some positive probability of misclassification. This probability of misclassification will depend on \( f(y; \theta_1), f(y; \theta_2) \) and will be known. In situations (ii) and (iii), in which a classification rule based on observations is being used, there are two other probabilities of misclassification to consider. One is the conditional probability given a fixed rule\(^2\) \( R(S_1, S_2) \), and the other is the unconditional probability, which is the expectation of the conditional probability.

To be more specific, assume the parametric case applies and assume \( \sim \) is from \( \pi_2 \). Let \( \theta = (\theta_1, \theta_2) \) be the parameters. Then let

\[
(1.2) \quad P_2^{{**}}(\theta) = \text{probability of misclassifying } \sim \text{ from } \pi_2 \text{ as } \pi_1 \text{ using rule (1.1) with all parameters known,}
\]

\[
(1.3) \quad P_2(R(S_1, S_2), \theta) = \text{conditional probability of misclassifying } \sim \text{ from } \pi_2 \text{ as } \pi_1, \text{ given a rule } R(S_1, S_2) \text{ based on original samples of sizes } N_1, N_2 \text{ from } \pi_1, \pi_2.
\]

\[
(1.4) \quad P_2^{{*}}(R(\cdot, \cdot), \theta) = \text{unconditional probability of misclassifying } \sim \text{ from } \pi_2 \text{ as } \pi_1 \text{ when } R(S_1, S_2) \text{ is random.}
\]

Note that \( P_2^{{*}}(R(\cdot, \cdot), \theta) = E \, P_2(R(S_1, S_2), \theta) \) where \( E \) indicates expectation over random original samples. The probabilities \( P_2(R(S_1, S_2), \theta), P_2^{{*}}(R(\cdot, \cdot), \theta), \) and \( P_2^{{**}}(\theta) \) will be written in abbreviated form as \( P_2, P_2^{{*}}, \) and \( P_2^{{**}} \) when the dependence on \( \theta, R, S_1, S_2 \) is not being emphasized.

Given a fixed rule \( R(S_1, S_2) \), which will be used to classify new observations, the conditional probability of misclassification

\(^2\)The meaning of the expression "fixed rule" is discussed in detail in Chapter 3, Section 1, below.
using this rule, that is \( P_2(R(S_1, S_2), \theta) \), seems to us to be of primary interest. This conditional probability is the actual probability of misclassifying a new observation when the rule is fixed. The unconditional probability, on the other hand, is the expected value of the conditional probability as the original samples vary. Even if only one set of original samples is taken and one classification rule is determined and then used to classify all the new observations, the unconditional probability is of some interest as the mean of the conditional probability.

\( P_2^{**}(\theta) \) is also of interest because it indicates the best that can be done, as measured by the probability of misclassification, in classifying with the \( x \) variable when distributions are completely known. In situation (ii) the difference between \( P_2^*(R(\cdot, \cdot), \theta) \) and \( P_2^{**}(\theta) \) can be considered as a measure of the loss in accuracy of classification due to not knowing the parameter \( \theta \). Furthermore, in a given applied classification problem, there may be a question of what measurements to use to make the classification. In the mathematical formulation of the problem the question is what exactly are the \( \Pi_1, \Pi_2 \) populations that are to be distinguished. Estimates of \( P_2^{**}(\theta) \) can be used to compare various sets of measurements.

Note that the conditional probability considered here, \( P_2 \), is actually a double conditional probability because it is the probability given the rule and given that the observation \( x \) comes from a particular population (from \( \Pi_2 \) to be specific). The probabilities \( P_2^* \) and \( P_2^{**} \) are also conditional in this latter sense. When there are prior probabilities \( q_1, q_2 \) that an \( x \) drawn at random is from \( \Pi_1, \Pi_2 \) respectively, then \( q_1P_1^* + q_2P_2^* \) is the overall probability of drawing an observation and misclassifying it, given the fixed rule. This overall probability, as well as \( q_1P_1^{**} + q_2P_2^{**} \),
might be of interest in some classification situations. In this paper the prior probabilities \( q_1 \) and \( q_2 \) will not be used and only the three probabilities of misclassification, \( P_2, P^*_2, P^{**}_2 \) are considered below.

2. **Survey of the Literature Related to Estimating the Probabilities of Misclassification.**

As noted above, in the case where the distributions are completely known, the relevant probability of misclassification, \( P^*_2(\theta) \), is also known, and there is no estimation problem. In the case where parameters are unknown, all three of the probabilities of misclassification are also unknown because they are functions of the parameters.

Many papers on classification and discrimination contain some remarks on the errors of classification and some suggestions for estimating one or more of these three probabilities of misclassification. No attempt will be made to survey this literature. Several of these papers are referred to in Chapter 2 in connection with particular estimators. Some survey and bibliography papers on classification were listed above.

A few papers are devoted wholly to the problem of estimating probabilities of misclassification. Lachenbruch (1965) and Lachenbruch and Mickey (1968) considered the problem of estimating the conditional probability of misclassification for multivariate normal variables with unknown means and common unknown covariance matrix; the rule is that of (1.1) with unknown parameters estimated by sample means and pooled sample covariance matrix. They considered seven estimators, all of which are based only on data from the original samples, and evaluated them by a Monte Carlo sampling experiment. For each set of samples drawn from two multivariate normal populations with known parameters the actual conditional probability of misclassification
was compared with the various estimates. Unfortunately, in these papers, as has been pointed out by Hills (1966), the conditional and unconditional probabilities of misclassification are not always clearly distinguished. Also there are statements about expectation, bias, and variance, where it is not clear if the moments being considered are conditional or unconditional. For example, in Lachenbruch and Mickey (1968), in the discussion of Okamoto's approximation (p. 718, line 7ff) $P_1^*, P_2^*$ should be $P_1^*, P_2^*$ (where $P_1^*$, $P_2^*$ are conditional and unconditional probabilities of misclassification respectively); this change is necessary for consistency with the authors' own remarks on p. 724. On p. 718, line 24, in the discussion of the U method estimator (described in Section 1.4, Chapter 2, below), the use of "unbiased" is ambiguous since for each $1, n_1 - 1$ split and fixed rule a conditionally unbiased estimator can be obtained. Then taking all possible $1, n_1 - 1$ splits, the proportion of the $n_1$ observations misclassified is an unconditionally unbiased estimator of the unconditional probability of misclassification using rules based on $n_1 - 1, n_2$ observations. If, however, the conditional probability of misclassification is being estimated, the significance of this unconditional unbiasedness is not clear.

In connection with this distinction between conditional and unconditional probabilities of misclassification, it will be noted that several of the estimators studied by Lachenbruch and Mickey seem intuitively more sensible as estimators of the unconditional probability of misclassification. Oddly enough, the best estimator (one based on Okamoto's approximation for the unconditional probability of misclassification and denoted by them as the OS estimator) is one of these and all such estimators are relatively good. Lachenbruch and Mickey (1968, p. 724) point out this "oddity" themselves. This problem is discussed further in Chapter 3, Section 4.4 and the Lachenbruch
and Mickey results in general are discussed in Chapter 4, Section 5.2.

Another paper by Lachenbruch (1967) gives a confidence interval for $P_2^*$ based on his $U$ estimator (see Section 1.4, Chapter 2).

Hills (1966) has a paper on errors in classification in which he is extremely careful to distinguish between the probability of misclassification when all parameters are known, the conditional probability of misclassification given a fixed rule, and the unconditional probability when rule is random, or, in the notation of the previous section between $P_2^{**}(\vartheta)$, $P_2(R(s_1, s_2), \vartheta)$ and $P_2^*(R(\cdot, \cdot), \vartheta)$, or as abbreviated between $P_2^{**}$, $P_2$, $P_2^*$. He is concerned with estimating both conditional and unconditional probabilities of misclassification. There is some general discussion in the paper and then some material on the multinomial and normal distributions.

For estimating the conditional probability, Hills suggests two estimation procedures--reclassification of the observations in the original sample and substitution of estimates for unknown parameters in the expression for the conditional probability of misclassification. These estimators are more fully described in Chapter 2, Section 1.1 and Section 2.2.

Pogue's (1966) dissertation, though in large part concerned with the Monte Carlo evaluation of several classification procedures for normal populations, contains a substantial amount of material related to estimating $P_2^{**}(\vartheta)$. The estimators considered are based on reclassification of original samples and classification of test samples and thus are non-parametric estimators not making use of the normal distribution assumptions.

It is intended in this thesis to concentrate on estimating the conditional probability of misclassification given a fixed rule. It
will be assumed that the form of the distribution is known, but that there are unknown parameters. In the next chapter are described a number of estimators of $P_2(R(S_1, S_2), \vartheta)$, the probability of misclassifying an observation from $\Pi_2$ as coming from $\Pi_1$, given the rule $R(S_1, S_2)$ based on original samples $S_1, S_2$ from $\Pi_1, \Pi_2$ of sizes $N_1, N_2$. Some of the estimators make use of distribution assumptions and some do not; some are based only on original samples and some require test samples. Some of the estimators seem intuitively to be estimators for the unconditional rather than for the conditional probability of misclassification, but are included because of the Lachenbruch (1965) and Lachenbruch and Mickey (1968) results previously mentioned. (Among the estimators for $P_2$ studied by them the best estimators are some which seem to be estimating the unconditional probability, $P_2^*$. ) Those estimators which do not make use of distributional assumptions are described first.
CHAPTER 2
DESCRIPTION OF ESTIMATORS OF
THE CONDITIONAL PROBABILITY OF
MISCLASSIFICATION

The function to be estimated is \( P_2(R(S_1, S_2), \theta) \), the conditional probability of misclassifying an observation randomly selected from \( \Pi_2 \), given a fixed rule \( R(S_1, S_2) \) based on random samples \( S_1, S_2 \) of sizes \( N_1, N_2 \) from \( \Pi_1, \Pi_2 \), which have densities \( f(y; \theta_1) \), \( f(y; \theta_2) \) with \( f \) known and \( \theta = (\theta_1, \theta_2) \) unknown. Test samples of sizes \( M_1, M_2 \) are assumed to be available if needed.

Several estimators of \( P_2 \) are described below. In Chapters 3 and 4 these estimators are applied in the particular case of normally distributed variables.

1. Estimators Not Using Distribution Assumptions.

There are some methods for estimating the conditional probability of misclassification which make use of no distribution assumptions. Such methods can be used in situation (iii) where the form of densities is unknown. They can also be used in situation (ii) where the form of the density is known but one or more parameters are unknown. Intuitively, estimators which make use of knowledge on the form of the density will be better than those which do not. If however, the assumption made about the form of the density is wrong, an estimator based on this form may be poorer than an estimator not based on distributional assumptions. Hence even if situation (ii) applies, it is worthwhile to consider estimators not based on the distribution.

1.1 The reclassification estimator, \( P_R \).

After any classification rule \( R(S_1, S_2) \) has been computed, the \( N_2 \) observations of the original sample from \( \Pi_2 \) can be reclassified using \( R(S_1, S_2) \). Denote by \( P_0 \) the proportion of these observations
misclassified. Smith (1947) suggested \( p_0 \) as an estimator for \( p_2 \). It is one of the seven estimators compared by Lachenbruch (1965) and by Lachenbruch and Mickey (1968). Hills (1966) lists \( p_0 \) as a commonly used estimator for the conditional probability, and it is one of the estimators which he studies. Note that this estimator is based only on the original observations and does not require test samples.

The estimator \( p_0 \) will also be denoted by \( P_R \). In general, estimators not using distribution assumptions will be denoted by \( P \) with a subscript to identify the particular estimator. Thus in the symbol \( P_R \), \( R \) stands for "reclassification."

**1.2 The test sample estimator, \( p_t \).**

Let \( p_t \) denote the proportion of the \( M_2 \) observations in the test sample from \( \Pi_2 \) which are misclassified using the rule \( R(S_1, S_2) \). This test sample estimator will also be denoted as \( P_T \) in agreement with the general method of notation described above in Section 1.1. Since we are interested in the probability of misclassifying a new random observation, \( p_t \), which is based on such new observations would seem to be a more appropriate estimator than \( p_0 \). In some situations, however, the total number of known observations available from \( \Pi_1 \) and \( \Pi_2 \) is small, and it is desirable to use all of the observations to determine the rule, saving none for test samples. Hills (1966) and Lachenbruch (1965) list \( p_t \) as one possible estimator for \( p_2 \), but then confine their attention to the situation in which only original samples are available.

Note that the conditional distribution of \( M_2 p_t \) given the rule \( R(S_1, S_2) \) is binomial with mean \( M_2 p_2 (R(S_1, S_2), \theta) \). The unconditional distribution is also binomial with mean \( M_2 p_2^* (R(\cdot, \cdot), \theta) \).
1.3 The two straight lines estimator, $P_{\text{TSL}}$.

For this estimator, assume that original samples $S_1, S_2$ of sizes $N_1, N_2$ are obtained from $\pi_1, \pi_2$ and a rule $R(S_1, S_2)$ determined. The $N_2$ original observations from $\pi_2$ are reclassified, using $R(S_1, S_2)$, to obtain $p_0$, which will here be denoted $p_0(N_2)$. Then the original sample from $\pi_1$ is split into two samples of sizes $n_1, N_1-n_1$ and the sample from $\pi_2$ is split into two samples of sizes $n_2, N_2-n_2$. A new rule, $R(S'_1, S'_2)$ is obtained, where $S'_i$ indicates the sample of size $n_i$ from $\pi_i, i = 1, 2$. Thus the samples of sizes $n_1, n_2$ are the original samples with respect to the rule $R(S'_1, S'_2)$ and the samples of sizes $N_1-n_1, N_2-n_2$ are test samples. The $n_2$ observations from $\pi_2$ are reclassified using $R(S'_1, S'_2)$ to obtain $p_0(n_2)$ and the $N_2-n_2$ observations are classified using $R(S'_1, S'_2)$ to obtain $p_t(N_2-n_2)$.

The idea of the two straight lines estimator is based on work of Pogue (1966) in relation to estimating $P_2^{**}$. For this derivation, assume that $N_1 = N_2 = N$ and $n_1 = n_2 = N/2$. Let $E$ here denote expectation as test samples and original samples vary, i.e., unconditional expectation, and let $p_0(N), p_t(N)$ indicate respectively the proportions of original and test samples of size $N$ from $\pi_2$ misclassified when using a rule based on samples from $\pi_1, \pi_2$ of size $N$ each. The discussion below is illustrated by Figure 1. Note that the horizontal axis is $N^{-1}$, not $N$.

Pogue assumed that

1. $E p_0(N)$ is an increasing function of $N$ going to the limit $P_2^{**}$ as $N \to \infty$,

2. $E p_t(N)$ is a decreasing function of $N$ going to the limit $P_2^{**}$ as $N \to \infty$,

3. both $E p_0(N)$ and $E p_t(N)$ are (approximately) linear in $N^{-1}$. 

- 12 -
Figure 1
Sketch of Procedure for the Two Straight Lines Estimator

In Figure 1, the three points marked • are obtained by estimating the indicated quantities from the data. \( \hat{P}_2^{**} \), denoted \( \odot \), is found by drawing a straight line through the estimates of \( E_{p0}(N') \) and \( E_{p0}(\frac{3}{2}N') \). Then the point \( \Theta \), which can be considered as an estimate of \( P_2(R(\cdot, \cdot), \theta) \) or of \( P_2(R(S_1, S_2), \theta) \), is obtained by drawing a straight line between \( \Theta \) and the estimate of \( E_{p_0}(\frac{1}{2}N') \).
By assumptions (1) and (3), \( E_0(N) \) is linear in \( N^{-1} \), increasing to \( \hat{p}_2^{**} \) as \( N^{-1} \to 0 \). Thus if values of \( E_0(N) \) are known or estimated for two \( N \) values, say \( N' \) and \( N'/2 \), an estimate of \( \hat{p}_2^{**} \) can be obtained by drawing a straight line through the two points, \( E_0(N') \) and \( E_0(N'/2) \), plotted as ordinates with \( N^{-1} \) as abscissa. The point on the line corresponding to \( N^{-1} = 0 \) is the estimate, which will be denoted as \( \hat{p}_2^{**} \). See Figure 1.

By conditions (2) and (3), \( E_t(N) \) is also linear in \( N^{-1} \), decreasing to \( \hat{p}_2^{**} \) as \( N^{-1} \to 0 \). Furthermore \( E_t(N) \) is equal to the unconditional probability of misclassification for samples of size \( N \), which will here be denoted by \( \hat{p}_2^*(N) \). Therefore, if \( E_t(N'/2) \) is known or estimated, an estimate of \( \hat{p}_2^*(N) \) can be obtained by finding the point on the straight line between \( E_t(N'/2) \) and \( \hat{p}_2^{**} \) which corresponds to an abscissa of \( 1/N' \). The formula for the estimator is

\[
\hat{E}_0(N') = 0.5 \hat{E}_t(N'/2) - 0.5 \hat{E}_0(N'/2).
\]

This estimator for \( \hat{p}_2^*(N) \) given in (1.1) can also be considered as an estimator for \( \hat{p}_2(N) \). The quantities \( E_0(N) \), \( E_t(N/2) \), and \( E_0(N/2) \) can be estimated by the observed values \( p_0(N) \), \( p_t(N/2) \) and \( p_0(N/2) \). Hence we will take the formula for the "two straight lines" estimator, \( \hat{p}_{TSL} \), to be

\[
\hat{p}_{TSL} = p_0(N) - 0.5 p_0(N/2) + 0.5 p_t(N/2).
\]

The estimator of (1.2) is intuitively derived, but a more general estimator of this form could be considered, namely

\[
\alpha p_0(N_2) + \beta p_0(n_2) + \gamma p_t(N_2 - n_2)
\]

where \( \alpha, \beta, \) and \( \gamma \) are constants. The values of \( \alpha, \beta, \) and \( \gamma \)
could then be chosen to give the "best" estimator of this form. Determination of these constants would, however, usually require use of assumptions on the distribution of the observations.

1.4 Lachenbruch's U method estimator, $P_U$.

Lachenbruch (1965) and Lachenbruch and Mickey (1968) proposed a method of estimating the conditional probability of misclassification, denoted by them as the U method, in which all possible $N_2-1$, 1 splits are made of the observations from $\pi_2$. For each split a rule is determined based on $N_1 + N_2 - 1$ observations and then the one remaining observation from $\pi_2$ is classified using this rule. The proportion of the $N_2$ observations misclassified is used as an estimator of $P_2(R(s_1, s_2), \theta)$, where $R(s_1, s_2)$ is a rule based on samples $s_1, s_2$ of size $N_1, N_2$ from $\pi_1, \pi_2$. Let this proportion misclassified be taken as an estimator of $P_2$ and let it be denoted by $P_U$.

Actually, in view of the sample splitting and the $N_2$ different rules involved in getting the estimate, $P_U$ seems intuitively at least as reasonable as an estimator of $P_2^*$, and Lachenbruch (1967) does consider confidence intervals for $P_2^*$ based on $P_U$. In the Lachenbruch (1965) and Lachenbruch and Mickey (1968) results, however, $P_U$ is found to be a fairly good estimator of $P_2$.\(^1\)

If the $N_2$ observations of $\pi_2$ are to be split into two sub-samples, it is not necessary to have the split be 1, $N_2 - 1$. All possible splits into $n_2$ and $N_2 - n_2$ observations can be made, where $n_2$ is any fixed number such that $1 \leq n_2 \leq N_2/2$, and then the same general estimation procedure as described above can be used.

\(^1\)For other opinions on these questions see C.A.B. Smith's discussion of Hills (1966) and Hills' reply.
2. Estimators Which Make Use of Distribution Assumptions.

Several estimators that use information about the densities \( f(y; \theta_1) \) and \( f(y; \theta_2) \) are described in this section.

2.1 The reclassification estimator adjusted for bias, \( \hat{p}_{Ra} \) \(^2\)

Intuitively it appears that the reclassification estimator, \( p_0 \), will underestimate \( P_2 \) because the classification rule should work better on the observations on which it was based than on new observations. It has been suggested by Cochran and Hopkins (1961) and by Hills (1966, p. 6) that a correction factor be added to \( p_0 \) to adjust for the bias of \( p_0 \) as an estimator of the unconditional probability of misclassification. A similar adjustment can be made for estimating the conditional probability.

Let \( \text{E}(\cdot|R(S_1, S_2)) \) denote expectation for rule fixed, test samples random. Let \( p_0 \) and \( p_t \) be defined as in Sections 1.1 and 1.2. Observe that

\[
\text{E}[p_0 + \text{E}(p_t|R(S_1, S_2)) - \text{E}(p_0|R(S_1, S_2))|R(S_1, S_2)] = \text{E}(p_t|R(S_1, S_2)) = \frac{p_0}{p_0 + \text{E}(p_t|R(S_1, S_2)) - \text{E}(p_0|R(S_1, S_2))}.
\]

Hence the function \( p_0 + \text{E}(p_t|R(S_1, S_2)) - \text{E}(p_0|R(S_1, S_2)) \) would be a conditionally unbiased estimator of \( p_2 \), except that it is in general not a statistic because \( \text{E}(p_t|R(S_1, S_2)) \) and \( \text{E}(p_0|R(S_1, S_2)) \) are functions of population parameters. Let \( \hat{\text{E}}(p_t|R(S_1, S_2)) \) and \( \hat{\text{E}}(p_0|R(S_1, S_2)) \) be \( \text{E}(p_t|R(S_1, S_2)) \) and \( \text{E}(p_0|R(S_1, S_2)) \) with parameters replaced by sample values. Then the function \( \hat{p}_{Ra} \) defined by

\[
(2.1) \quad \hat{p}_{Ra} = p_0 + \hat{\text{E}}(p_t|R(S_1, S_2)) - \hat{\text{E}}(p_0|R(S_1, S_2))
\]

should be an approximately unbiased estimator of \( p_2(R(S_1, S_2), \theta) \).

\(^2\)Although in Section 1.1 it was stated that the symbol \( P \) with a subscript would denote an estimator not based on distribution assumptions, an exception is made here to indicate the close relationship between this estimator and the reclassification estimator, \( \hat{p}_R \). Thus "Ra" stands for "reclassification adjusted."
Observe that although $p_0$ can be computed without any assumptions about the distribution of the observations, the adjusted $p_0$, i.e., $p_{Ra}$, cannot, since without such assumptions $E(\cdot | R(S_1, S_2))$ will not be defined. Estimates of the parameters in $E(p_0 | R(S_1, S_2))$ and $E(p_t | R(S_1, S_2))$ can be based on the original samples. Furthermore, no observed value of $p_t$ is involved, but only $E(p_t | R(S_1, S_2))$, so that test samples are not necessary.

2.2 The expression for the true conditional probability of misclassification with parameters replaced by estimates.

If $f(y; \hat{\omega}_1)$ and $f(y; \hat{\omega}_2)$ are assumed known with $\hat{\omega} = (\hat{\omega}_1, \hat{\omega}_2)$ unknown, the true conditional probability of misclassification, $P_2(R(S_1, S_2), \hat{\omega})$, can be determined as a function of $\hat{\omega}$. Then sample estimates can be substituted for $\hat{\omega}$ to get an estimate of $P_2$. Let $P_2(R(S_1, S_2), \hat{\omega})$, or for abbreviation $P_2(\hat{\omega})$, denote a general estimator of this type. Various specific estimators can be obtained by using different estimators for $\hat{\omega}$, for example, the maximum likelihood estimator.

Estimators of the type, $P_2(\hat{\omega})$ are commonly used. Hills (1966) studies them in general and for some specific cases. One of the estimators of $P_2$ used by Lachenbruch (1965) and Lachenbruch and Mickey (1968) for the multivariate normal situation with the rule based on the discriminant function is of this type, specifically the estimator designated by them as the D method.

2.3 The expression for the true unconditional probability of misclassification with parameters replaced by estimates.

If the formula for the true unconditional probability of misclassification, $P_2^*(R(\cdot, \cdot), \hat{\omega})$ can be obtained as a function of parameters $\hat{\omega}$, then sample estimates can be substituted for $\hat{\omega}$ to obtain $P_2^*(R(\cdot, \cdot), \hat{\omega})$, which will be abbreviated to $P_2^*(\hat{\omega})$. $P_2^*(\hat{\omega})$ seems
naturally to be an estimator of $P_2^*$, but it can also be considered as
an estimator of $P_2$. For example, the O and OS estimators of
$P_2$ considered by Lachenbruch (1965) and by Lachenbruch and Mickey (1968)
are essentially estimators of the type $P_2^*(\hat{\theta})$. When the populations
are multivariate normal, which is the case they consider, there is no
simple expression for $P_2^*(R(\cdot, \cdot), \theta)$. They use an approximation
formula for $P_2^*$ due to Okamoto (1963). The O and OS estimators
are obtained by substituting two different sample estimates for
the unknown parameter in the Okamoto expression. As was mentioned
earlier (Chapter 1, Section 2) the OS estimator is generally the
best of the seven estimators of $P_2$ considered in the Lachenbruch
papers.

2.4 The expression for the true probability of misclassification when
all parameters are known, with parameters replaced by estimates.

When the form of the densities and all parameters are known, the
probability of misclassification is $P_{2^*}(\theta)$, which is also known. If
the form of the densities is known, but depends on an unknown parameter
$\theta$, $P_{2^*}(\theta)$ can be computed as a function of $\theta$ but will be unknown
because $\theta$ is unknown. Then $P_{2^*}(\hat{\theta})$, where $\hat{\theta}$ represents any
sample estimate of $\theta$, is naturally an estimator for $P_{2^*}(\theta)$, but it
can also be considered as an estimator for $P_2(R(S_1, S_2), \theta)$.

2.5 Minimum variance unbiased estimators.

Most of the estimators listed above are obvious kinds of estimators
or are derived by an intuitive argument. Given such an estimator, its
properties can then be determined. The adjusted $p_0$ estimator
described in Section 2.1, however, was derived in an attempt to get
an unbiased estimator. In general, attempts can be made to find
estimators with certain desirable or optimal properties, such as
minimum variance.
For example, suppose that in a given classification situation an unconditionally unbiased estimator for $P_2^*$ is known. Suppose further that there exists a sufficient statistic for $\Theta$. Then the Rao-Blackwell theory can be applied to find an unconditionally minimum variance unbiased estimator for $P_2^*(R(\cdot, \cdot), \Theta)$. When estimating $P_2$, the conditional probability of misclassification, one may desire to compare estimators on the basis of conditional rather than unconditional moments. If the Rao-Blackwell theory is adjusted to apply in this kind of conditional situation, a conditionally minimum variance estimator can be found. Such a conditional Rao-Blackwell type theory is derived in Chapter 5. A conditionally UMVU estimator for $P_2(R(S_1, S_2), \Theta)$ in the normal classification problem is also derived there.

2.6 Estimators involving a prior distribution on $\Theta$.

Let $\Theta = (\Theta_1, \Theta_2)$ be the vector of all unknown parameters for the two densities $f(y; \Theta_1), f(y; \Theta_2)$ which correspond to $\Pi_1, \Pi_2$. Suppose a prior distribution for $\Theta$, denoted $\mathcal{G}(\Theta)$, is available.

Under this supposition, a Bayesian classification procedure can be applied, as in Geisser (1964), (1966). To determine the procedure, samples are first obtained from $\Pi_1$ and $\Pi_2$. Then the posterior distribution of $\Theta$, given the two samples, is found. Denote this posterior distribution by $H(\Theta)$. Let $\sim x$ denote a new observation, from either $\Pi_1$ or $\Pi_2$, which is to be classified. The classification rule is based on the "predictive density" of $\sim x$, which is obtained for $\sim x$ from $\Pi_i$, $i = 1, 2$ by integrating $f(\sim x; \Theta_i), i = 1, 2$ with respect to $H(\Theta)$. Then the "predictive" probability of misclassifying $\sim x$ from $\Pi_2$, for example, is found by integrating the predictive density of $\sim x$ over the region which corresponds to classifying $\sim x$ as from $\Pi_1$. This probability will not be a function of $\Theta$, since $\Theta$ has been integrated out in finding the predictive density of $\sim x$. 

- 17b -
Thus there is no problem of estimating the probability of misclassification.

Alternatively, Bayesian theory can be used to invent estimators for the various probabilities of misclassification, when the rule used is non-Bayesian. For example, the rule might be that of (1.1), Chapter 1, parameters estimated, with the prior distribution, $G(\theta)$, entering into the estimation of the probabilities of misclassification. Geisser (1967) has done some work along these lines for multivariate normal variables.

As stated above, we are concerned here with estimating the conditional probability of misclassification, $P_2(R(s_1, s_2), \theta)$. Starting from $G(\theta)$, estimators for $P_2$ can be generated by various methods. Several such methods are enumerated below. An estimator which is obtained, as these below are, by a Bayesian or quasi-Bayesian argument can, of course, be evaluated by non-Bayesian criteria if desired.

In the descriptions below, $H(\theta)$ denotes the posterior distribution of $\theta$ given the original samples, $s_1$ and $s_2$.

Method 1. Use the mean of the posterior distribution of $\theta$ as an estimator for $\theta$ and substitute this estimator in the expression for $P_2(R(s_1, s_2), \theta)$.

Method 2. Use that $\theta$ value which minimizes the posterior expected loss (i.e., the loss in estimating $\theta$) as an estimator of $\theta$ and substitute this estimator in $P_2(R(s_1, s_2), \theta)$.

Note that if the loss function is $(\theta - \hat{\theta})^2$, procedure 2 yields the same estimator of $\theta$, and hence the same estimator of $P_2$, as does procedure 1. In other words, method 1 is a special case of method 2.

- 18 -
Method 3. Use the mean of the posterior distribution of $P_2(R(s_1, s_2), \sim)$, with $R$ fixed, as the estimator of $P_2$. This estimator can be computed either by taking the expectation of $P_2(R(s_1, s_2), \sim)$ with respect to $H(\theta)$, or by actually finding the posterior distribution of $P_2$ from the posterior distribution of $\sim$ and then taking the expectation of $P_2(R(s_1, s_2), \sim)$ with respect to the posterior distribution of $P_2$.

Method 3 is suggested by Geisser (1967, pp 815-816) as a way of getting a point estimate of $P_2$.

Method 4. Find the posterior distribution of $P_2(R(s_1, s_2), \sim)$ and use as an estimator for $P_2$ that value which minimizes the posterior expected loss (loss in estimating $P_2$).

The same comment applies here about the relation between methods 3 and 4 as applies to the relation between 1 and 2, namely, if the loss function is $(P_2 - \hat{P}_2)^2$, method 4 leads to the same estimator as 3.

The next methods are written out in terms of densities.

Method 5. Using $H(\theta)$, get the joint density of $(\theta, \sim, R)$, where $R = R(s_1, s_2)$ is the rule based on original samples $s_1, s_2$ from $\Pi_1, \Pi_2$ and $\sim$ is a new observation to be classified. (Note that "density of $(\theta, \sim, R)$" is a loose expression standing for the joint density of $\sim, \theta$, and some functions of $s_1, s_2$ that are involved in $R(s_1, s_2)$. Exactly what these functions of $s_1, s_2$ will be must be decided for each particular situation.) As before, $\sim$ will be assumed to be from $\Pi_2$. Integrate out $\theta$ from the joint density to get the marginal density of $(\sim, R)$ and then get the conditional density of $\sim$ given $R$. The estimator of $P_2$ is the function obtained by integrating the density of $\sim/R$ over the region where $\sim$ is classified as $\Pi_1$.

Note that the conditional density of $\sim/R$ is a kind of "conditional predictive" density.
Method 6. Using $H(\theta)$, find the joint density of $(\theta, x)$ and integrate out $\theta$ to get the marginal density of $x$. Then take as an estimator of $P_2$ the value of the integral of the marginal density over the region where $x$ is classified as $\Pi_1$.

Note that this marginal density of $x$ is the "predictive" density of $x$, and the estimator is the "predictive probability of misclassification" that arise in the Bayes classification procedure described at the beginning of this section.

Some estimators similar to those above, but making use of the prior distribution of $\theta$ instead of the posterior distribution might be considered. For example:

Method 7. Take the expectation of $P_2(R(s_1, s_2), \theta)$ with respect to $G(\theta)$.

Method 7 is the same as Method 3 with $H(\theta)$ replaced by $G(\theta)$. No sample information about $\theta$ enters into the Method 7 estimator. Therefore it is a peculiar kind of estimator and may not have good properties.

Method 8. Use Method 5, but get the joint density of $(\theta, x, R)$ by using $G(\theta)$.

Note that even though $G(\theta)$ rather than $H(\theta)$ is used in this estimator, sample information about $\theta$ does enter through $R(s_1, s_2)$. And indeed under certain conditions Method 8 yields the same estimator as Method 3.

Method 9. Use Method 6, but get the joint density of $(\theta, x)$ by using $G(\theta)$. 
CHAPTER 3

ESTIMATING THE CONDITIONAL PROBABILITY
OF MISCLASSIFICATION IN THE
UNIVARIATE NORMAL CASE

In this chapter, a number of estimators for $P_2(R(S_1, S_2), \theta)$, which were described in Chapter 2, are considered in the case where the populations are univariate normal with unknown means $\mu_1, \mu_2$ and common known variance $\sigma^2$; the rule used is the rule of (1.1) Chapter 1 with parameters estimated. Since $\mu_1$ and $\mu_2$ are the only unknown parameters, we take $\theta = (\mu_1, \mu_2)$. Furthermore when $\sigma^2$ is known, the variables can be transformed to make the variance equal to one, and so, for simplicity, $\sigma^2$ will be taken equal to one.

In Section 1 of the chapter the expressions for the classification rule and for the probabilities of misclassification, $P_2$, $P^*_2$ and $P^{**}_2$, are given. In Section 2 each of the estimation methods based on distribution assumptions that were described in general in Section 2 of Chapter 2 are discussed in the particular case of univariate normal populations and expressions for each one of such estimators to be studied are given. In Sections 3 and 4 all the estimators studied, which are listed in Table III, are compared on the basis of conditional moments given $\bar{x}_1$, $\bar{x}_2$ fixed. Section 3 relates to exact expressions for the conditional means and variances, while in Section 4, asymptotic expressions for the conditional bias, variance, and mean square error are given. Section 4 also contains a discussion of the results and conclusions.

1. The Classification Rule and the Expressions for the Various Probabilities of Misclassification.

Let $\Pi_1$ be $N(\mu_1, 1)$ and $\Pi_2$ be $N(\mu_2, 1)$. Let $\bar{x}_1$, $\bar{x}_2$ be sample means based on original samples of sizes $N_1$, $N_2$ from $\Pi_1$, $\Pi_2$.
respectively. Let \( x \) be a new observation to be classified. The usual classification rule, and the only one considered in this chapter, is the rule of (1.1), Chapter 1, with \( c = 1 \) and with \( \mu_1, \mu_2 \) estimated by \( \bar{x}_1, \bar{x}_2 \). This rule reduces to:

\[
\begin{cases}
\text{classify } x \text{ as } \Pi_1 & \text{if } x(x_1 - x_2) > \frac{1}{2}(x_1 - x_2)(x_1 + x_2) \\
\text{classify } x \text{ as } \Pi_2 & \text{if } x(x_1 - x_2) < \frac{1}{2}(x_1 - x_2)(x_1 + x_2). \\
\end{cases}
\]  

(1.1)

An equivalent expression of the rule is the following:

\[
\begin{cases}
\text{classify } x \text{ as } \Pi_1 & \text{if } \bar{x}_1 > \bar{x}_2 \text{ and } x > \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \\
\text{or if } \bar{x}_1 < \bar{x}_2 \text{ and } x < \frac{1}{2}(\bar{x}_1 + \bar{x}_2), \\
\text{otherwise classify } x \text{ as } \Pi_2. \\
\end{cases}
\]  

(1.2)

We now digress to discuss the question of what is meant by conditioning on the rule. Note that if \( x_1 \) and \( x_2 \) are fixed, the rule is fixed. Furthermore, if the rule is assumed given exactly in the form (1.1), fixing the rule also fixes \( \bar{x}_1 \) and \( \bar{x}_2 \). On the other hand it is obvious from (1.2) that the rule can be given simply in terms of the two values, \( \bar{x}_1 + \bar{x}_2 \) and \( \text{sgn}(\bar{x}_1 - \bar{x}_2) \). Hence if the expression "given \( R(S_1, S_2) \)" is taken to mean "given the least amount of information from the original samples needed to determine \( R(S_1, S_2) \)," in this normal case "given R" means "given \( (\bar{x}_1 + \bar{x}_2) \) and \( \text{sgn}(\bar{x}_1 - \bar{x}_2) \)." Since different values of \( \bar{x}_1, \bar{x}_2 \) can lead to the same values for \( (\bar{x}_1 + \bar{x}_2) \) and \( \text{sgn}(\bar{x}_1 - \bar{x}_2) \), fixing \( R(S_1, S_2) \) is not equivalent to fixing \( \bar{x}_1, \bar{x}_2 \). Nevertheless we will condition on \( \bar{x}_1, \bar{x}_2 \) rather than precisely on \( R(S_1, S_2) \). We assume that \( \bar{x}_1, \bar{x}_2 \) will generally be known and fixing \( \bar{x}_1 \) and \( \bar{x}_2 \) is less awkward than fixing \( \bar{x}_1 + \bar{x}_2 \) and \( \text{sgn}(\bar{x}_1 - \bar{x}_2) \). Also other workers who have studied the conditional probability of misclassification for the normal case take \( \bar{x}_1, \bar{x}_2 \) (and estimated covariance, if covariance
is unknown) as fixed (see, for example, Lachenbruch and Mickey (1968, p. 716)). This distinction is indeed irrelevant for some purposes. The conditional probability of misclassification, derived below in this section, is the same for either conditioning. However, when we later consider conditional moments the distinction is relevant and we in fact condition on \( x_1 \) and \( x_2 \).

Now to derive the expressions for the various probabilities of misclassification, let

\[
F(y) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{y} e^{-\frac{1}{2}z^2} dz
\]

(1.3)

\[
F(u, v; \rho) = \frac{1}{2\pi} \frac{1}{\sqrt{1-\rho^2}} \int_{-\infty}^{u} \int_{-\infty}^{v} g(x, y; \rho) dy dx
\]

(1.4)

\[
g(x, y; \rho) = \exp \left[ -\frac{1}{2(1-\rho^2)} (x^2 - 2\rho xy + y^2) \right].
\]

(1.5)

Assume \( x \) is from \( \mu_2 \). Then using (1.2) the conditional probability of misclassification is computed as follows:

if \( \bar{x}_1 - \bar{x}_2 > 0 \),

\[
P_2(R(S_1, S_2), \theta) = P\{x > \frac{1}{2}(\bar{x}_1 + \bar{x}_2) | x \text{ from } \mu_2; \bar{x}_1, \bar{x}_2 \text{ fixed} \}
\]

\[
= 1 - F(\frac{1}{2}\bar{x}_1 + \frac{1}{2}\bar{x}_2 - \mu_2),
\]

(1.6)

and if \( \bar{x}_1 - \bar{x}_2 < 0 \),

\[
P_2(R(S_1, S_2), \theta) = P\{x < \frac{1}{2}(\bar{x}_1 + \bar{x}_2) | x \text{ from } \mu_2; \bar{x}_1, \bar{x}_2 \text{ fixed} \}
\]

\[
= F(\frac{1}{2}\bar{x}_1 + \frac{1}{2}\bar{x}_2 - \mu_2).
\]

Since the expressions of (1.6) do not involve \( \mu_1 \), and since \( \bar{x}_1, \bar{x}_2 \) determine the rule, let \( P_2(R(S_1, S_2), \theta) \) be denoted by \( P_2(\bar{x}_1, \bar{x}_2, \mu_2) \). As before, this expression will be abbreviated to \( P_2 \), when the dependence on \( \bar{x}_1, \bar{x}_2 \) and \( \mu_2 \) is not being emphasized.
The unconditional probability of misclassification is

\[
P^*_2(R(\cdot, \cdot), \theta) = \left\{ \begin{array}{l}
\mathbb{P}(x_1 - x_2 > 0 \text{ and } x - \frac{1}{2}(x_1 + x_2) > 0 \mid x \text{ from } \pi_2) \\
+ \mathbb{P}(x_1 - x_2 < 0 \text{ and } x - \frac{1}{2}(x_1 + x_2) < 0 \mid x \text{ from } \pi_2) \end{array} \right\}
\]

\[
= \left\{ \begin{array}{l}
F(\delta M^{-\frac{1}{2}}, -\frac{1}{2}\delta(1 + \frac{1}{4} M)^{-\frac{1}{2}}; \rho) \\
+ F(-\delta M^{-\frac{1}{2}}, \frac{1}{2}\delta(1 + \frac{1}{4} M)^{-\frac{1}{2}}; \rho) \end{array} \right\}
\]

where \( \delta = \mu_1 - \mu_2, M = N_1^{-1} + N_2^{-1}, \) and

\[
\rho = \rho(\overline{x}_1 - \overline{x}_2, x - \frac{1}{2}(\overline{x}_1 + \overline{x}_2)) = \frac{-N_1^{-1} + N_2^{-1}}{2\sqrt{N} \sqrt{1 + \frac{1}{4} M}}.
\]

If \( N_1 = N_2 = N, \) then \( \rho = 0 \) and \( P^*_2(R(\cdot, \cdot), \theta) \) can be written in terms of univariate normal distribution functions as

\[
P^*_2(R(\cdot, \cdot), \theta) = \left\{ \begin{array}{l}
F(\frac{1}{2}\delta(2N)^{\frac{1}{2}})F(-\frac{1}{2}\delta(1 + \frac{1}{2N})^{-\frac{1}{2}}) \\
+ F(-\frac{1}{2}\delta(2N)^{\frac{1}{2}})F(\frac{1}{2}\delta(1 + \frac{1}{2N})^{-\frac{1}{2}}) \end{array} \right\}.
\]

Note that in both (1.7) and (1.8), \( \theta = (\mu_1, \mu_2) \) enters in only in the form \( \delta = \mu_1 - \mu_2, \) so that \( P^*_2(R(\cdot, \cdot), \theta) \) can be denoted as \( P^*_2(R(\cdot, \cdot), \delta). \)

The above expressions for the conditional and unconditional probabilities of misclassification agree with those of John (1961).

When all the parameters are known, that is \( \mu_1, \mu_2 \) known, rule (1.1), Chapter 1, with \( c = 1, \) is the same as the rule of (1.1) or (1.2) with \( \overline{x}_i \) replaced by \( \mu_i, i = 1, 2. \) Thus the true probability of misclassification when all parameters are known is computed as follows:

\[
P^*_2(\theta) = \mathbb{P} \{x(\mu_1 - \mu_2) \geq \frac{1}{2}(\mu_1 - \mu_2)(\mu_1 + \mu_2)\}
\]

\[
= 1 - F(\frac{1}{2}|\delta|),
\]

- 24 -
or equivalently

$$P_2^*(\theta) = \begin{cases} 1 - F(\frac{\delta}{2}), & \delta > 0 \\ \frac{1}{2}F(\delta), & \delta < 0. \end{cases}$$

Note that like $P_2^*$, $P_2^{**}$ depends on $\theta$ only through $\mu_1 - \mu_2 = \delta$ and hence $P_2^{**(\delta)}$ can be denoted by $P_2^{**}(\delta)$.

We now consider various estimators for $P_2$, the conditional probability of misclassification. Notation used in the discussion is given in Table I; some of this notation has already been introduced and is summarized in the table for convenience.

2. The Estimators.

The four estimators described in Section 1 of Chapter 2 are not based on distribution assumptions and thus are defined the same way whatever the distribution of the observations. For the estimators described in Section 2 of Chapter 2, however, the exact form of the estimator depends on the distribution. Given that $\pi_1$ is $N(\mu_1, 1)$ and $\pi_2$ is $N(\mu_2, 1)$, these estimators have the forms derived below.

Note that the sub-section numbers below correspond to those of Chapter 2, Section 2; the appropriate sub-sections there can be consulted for more information on each of the estimators.

The expression and notation for each of the estimators of $P_2$ that will be studied is given in Table III, along with some descriptive material (such as whether or not the estimator uses test samples) and some remarks on who has previously studied the estimator.

Notation not defined below is defined in Table I. At the end of each sub-section title, the notation for estimators derived in that sub-section is given in parentheses.

2.1 The reclassification estimator adjusted for bias, $(P_{Ra})$.

By definition (2.1) of Chapter 2, the adjusted reclassification
estimator, \( P_{Ra} \), is equal to \( P_0 + \hat{E}(P_t | R(S_1, S_2)) - \hat{E}(P_0 | R(S_1, S_2)) \).

From Appendix II, for univariate normal populations,

\[
E(p_t | \bar{x}_1, \bar{x}_2) = \begin{cases} 
1 - F(c), & d > 0 \\
F(c), & d < 0
\end{cases}
\]

and

\[
E(p_0 | \bar{x}_1, \bar{x}_2) = \begin{cases} 
1 - F(b), & d > 0 \\
F(b), & d < 0
\end{cases},
\]

where \( d = \bar{x}_1 - \bar{x}_2 \), \( b = \frac{1}{2d}(1 - \frac{1}{N})^{-\frac{1}{2}} \) and \( c = \frac{1}{2}\bar{x}_1 + \frac{1}{2}\bar{x}_2 - \mu_2 \).

Estimating \( \mu_2 \) by its usual estimator, \( \bar{x}_2 \), we get

\[
\hat{E}(p_t | \bar{x}_1, \bar{x}_2) = \begin{cases} 
1 - F(\frac{1}{2}d), & d > 0 \\
F(\frac{1}{2}d), & d < 0
\end{cases},
\]

while \( \hat{E}(p_0 | \bar{x}_1, \bar{x}_2) = E(p_0 | \bar{x}_1, \bar{x}_2) \) because \( b \) does not depend on \( \mu_1 \) or \( \mu_2 \). Thus we get the estimator

\[
P_{Ra} = \begin{cases} 
P_0 + F(b) - F(\frac{1}{2}d), & d > 0 \\
P_0 + F(\frac{1}{2}d) - F(b), & d < 0
\end{cases}
\]

2.2 The true conditional probability of misclassification with parameters replaced by estimators, \( (Q_D \text{ and } Q_R) \).

This type of estimator is equal to \( P_2(R(\bar{x}_1, \bar{x}_2), \mu_2) \) with \( \mu_2 \) replaced by an estimate. From (1.6), \( P_2 = 1 - F(c) \) if \( d > 0 \), and \( P_2 = F(c) \) if \( d < 0 \). Hence, if \( \bar{x}_2 \) is used as the estimator of \( \mu_2 \), we have the estimator

\[
P_2(R(\bar{x}_1, \bar{x}_2), \hat{\mu}_2 = \bar{x}_2) = \begin{cases} 
1 - F(\frac{1}{2}d), & d > 0 \\
F(\frac{1}{2}d), & d < 0
\end{cases}
\]

Let this estimator be called \( Q_D \). In general estimators obtained by estimating unknown parameters in the expressions for \( P_2, P_2^*, \) and \( P_2^{**} \) will be denoted by \( Q, Q^*, Q^{**} \) respectively with a subscript referring to the particular estimator. This one has subscript \( D \).
because it corresponds to Lachenbruch's (1965) and Lachenbruch and Mickey's (1968) D estimator. In the multivariate normal case it can be obtained by estimating the unknown parameter in $P^*_2$, by

\[ D = (\bar{x}_1 - \bar{x}_2)' \Sigma^{-1}(\bar{x}_1 - \bar{x}_2) \]  
(see Chapter 4, Section 2.4).

Other estimators of $\mu_2$ can be used, leading to other estimators for $P_2$. For example, let $\bar{r}_2$ be the sample mean of a test sample of size $M_2$ from $\pi_2$. If $\bar{r}_2$ is substituted for $\mu_2$ in the expression for $P_2$, we get the estimator

\[ (2.3) \quad P_2(R(\bar{x}_1, \bar{x}_2), \hat{\mu}_2 = \bar{r}_2) = \begin{cases} 1 - F(\frac{1}{2}d + \bar{x}_2 - \bar{r}_2), & d > 0 \\ F(\frac{1}{2}d + \bar{x}_2 - \bar{r}_2), & d < 0. \end{cases} \]

Let this estimator be denoted by $Q_T (T$ denotes "test").

2.3 The expression for the true unconditional probability of misclassification with parameters replaced by estimates, $(Q_{D}^*)$.

This estimator is equal to $P_2^*(R(\cdot, \cdot), \theta)$ with $\theta$ replaced by a sample estimate. From (1.7), the parameter involved in $P_2^*$ is $\delta = \mu_1 - \mu_2$, which will be estimated by $d = \bar{x}_1 - \bar{x}_2$. Making the substitution of $d$ for $\delta$ in the expression for $P_2^*$ given in (1.7), we have the estimator

\[ (2.4) \quad P_2^*(R(\cdot, \cdot), \hat{\delta} = d) = \begin{cases} F(dM^{-\frac{1}{2}}, -\frac{1}{2}d(1 + \frac{1}{4}M)^{-\frac{1}{2}}; \rho) \\ + F(-dM^{-\frac{1}{2}}, \frac{1}{2}dM^{-\frac{1}{2}}; \rho) \end{cases} \]

where, as in Section 1,

\[ M = N_1^{-1} + N_2^{-1} \]

and

\[ \rho = \frac{1}{2}(-N_1^{-1} + N_2^{-1}M^{-\frac{1}{2}}(1 + \frac{1}{4}M)^{-\frac{1}{2}}) \]

In the case that $N_1 = N_2 = N$, $\rho = 0$ and the expression of (2.4) reduces to
This estimator will be denoted by $Q_d^*$.

More estimators can be obtained, as was done in Section 2.2, by using estimators of $\delta$ other than $d$ and substituting these in the expression of (1.7). No other such estimators will, however, be considered here.

2.4 The expression for the true probability of misclassification when all parameters are known, with parameters replaced by sample values, $(Q_{DS}^{**})$.

From (1.10), $P_2^{**}(\delta)$ is equal to $F(-\frac{1}{2}\delta)$ when $\delta > 0$, and equal to $F(\frac{1}{2}\delta)$ when $\delta < 0$. Now when parameters are unknown, the sign of $\delta$ is also unknown. However, $d \overset{a.s.}{\rightarrow} \delta$ as $N_1 \rightarrow \infty$, $N_2 \rightarrow \infty$, and so for large $N_1, N_2$ the probability that $d$ has the same sign as $\delta$ is close to 1. Hence for large sample sizes we can take as estimator

\[
P_2^{**}(|\delta| = |d|) = \begin{cases} 
1 - F(\frac{1}{2}d), & d > 0 \\
F(\frac{1}{2}d), & d < 0.
\end{cases}
\]

Another line of reasoning leading to this estimator is to write $P_2^{**}(\delta) = 1 - F(\frac{1}{2}|\delta|)$, as in (1.9), and estimate $|\delta|$ by $|d|$.

The estimator of (2.6) is the same as the estimator $P_2(R(\bar{x}_1, \bar{x}_2), \hat{\mu}_2 = \bar{x}_2)$, denoted $Q_d$, already derived in Section 2.2, so no new estimator is obtained by taking the expression for $P_2^{**}(\delta)$ and estimating $|\delta|$ by $|d|$. Let $(d')^2 = \left(\frac{N_1 + N_2 - 4}{N_1 + N_2 - 2}\right) d^2$. If $|\delta|$ is estimated by $|d'|$, a new estimator

\[
P_2^{**}(|\delta| = |d'|) = 1 - F(\frac{1}{2} \sqrt{\frac{N_1 + N_2 - 4}{N_1 + N_2 - 2}} |d|).
\]

is obtained, which is the DS estimator of Lachenbruch (1965) and
Lachenbruch and Mickey (1968). We will call this estimator $Q_{DS}$.

2.5 Minimum variance unbiased estimators, $(P_{RB})$.

Assume that the data available consists of original samples of sizes $N_1, N_2$ from $\Pi_1, \Pi_2$ and independent test samples of sizes $M_1, M_2$ from $\Pi_1, \Pi_2$. Let $\bar{x}_1, \bar{x}_2$ be the sample means for the original samples and let $\bar{t}_1, \bar{t}_2$ be the sample means for the test samples.

Let $P_{RB}$ denote the conditional UMVU estimator of $P_2(R(\bar{x}_1, \bar{x}_2), \mu_2)$. Then

$$P_{RB} = \begin{cases} 1 - F((\frac{1}{2}\bar{x}_1 + \frac{1}{2}\bar{x}_2 - \bar{t}_2)(1 - M_2^{-1})^{-\frac{1}{2}}), & d > 0 \\ F((\frac{1}{2}\bar{x}_1 + \frac{1}{2}\bar{x}_2 - \bar{t}_2)(1 - M_2^{-1})^{-\frac{1}{2}}), & d < 0. \end{cases}$$

The derivation of this estimator by conditional Rao-Blackwell theory is given in Chapter 5, Section 4.4 for the multivariate normal case. Note that because $P_2$ is a function of $\mu_2$ but not a function of $\mu_1$, the conditional UMVU estimator will be the same as that given in (2.8) if the data available is assumed to consist only of original samples plus an independent test sample from $\Pi_2$, that is, no test sample from $\Pi_1$.  

2.6 Estimators involving a prior distribution on $\theta$, $(P_3$ and $P_5$).

For the normal classification problem considered here, the parameter $\theta$ is equal to $(\mu_1, \mu_2)$. For the prior distribution, $G(\theta)$, we will take $\mu_1$ as $N(\gamma_1, \sigma_1^2)$, $\mu_2$ as $N(\gamma_2, \sigma_2^2)$ with $\mu_1, \mu_2$ independent. With this prior, the posterior distribution of $\theta$ given $N_1$ observations from $\Pi_1$ and $N_2$ observations from $\Pi_2$, which is denoted $H(\theta)$, is as follows:

---

1Estimators using normality but not derived by substituting estimators for unknown parameters in $P_2, P_3$ or $P_{RB}$ will be denoted by $P$ with a subscript. The "RB" here stands for "Rao-Blackwell."
(2.9) \[ H(\theta): \mu_1 \sim N(\nu_1, \tau_1^2), \mu_2 \sim N(\nu_2, \tau_2^2); \mu_1, \mu_2 \text{ independent,} \]

where

\[ \nu_1 = \frac{\gamma_1 + \sigma_1^2 N_{1-1}}{1 + N_1 \sigma_1^2}, \quad \tau_1^2 = \frac{\sigma_1^2}{1 + N_1 \sigma_1^2}, \quad i = 1, 2. \]

Using the above prior and subsequent posterior distributions, several estimators for \( P_2 \) were derived by methods listed in Section 2.6, Chapter 2. The estimators are given in Table II and the derivations are in Appendix I. Table II also gives, for each estimator, the limiting form of the estimator as \( \sigma_1^2 \to \infty, \sigma_2^2 \to \infty. \) This limit is the limit of no prior information on \( \theta. \) Since none of the estimators actually involve \( \sigma_1^2, \) the limits are simply limits as \( \sigma_2^2 \to \infty. \) The limiting forms of these estimators will be denoted \( \phi_j \), where \( j \) is the method number.

No estimators were derived by methods 2 or 4, except for the special case of quadratic loss, in which case methods 2 and 4 reduce to methods 1 and 3 respectively. The limiting form of the method 1 estimator is equal to the estimator \( Q_0 \) already derived in Section 2.2.

Method 3, on the other hand, does give a new estimator in the limit, namely

\[ (2.10) \quad \phi_3 = \begin{cases} 1 - F(\frac{1}{2d}(1 + N_2^{-1})^{-\frac{1}{2}}), & d > 0 \\ F(\frac{1}{2d}(1 + N_2^{-1})^{-\frac{1}{2}}), & d < 0. \end{cases} \]

Method 5 also leads to a new estimator,

\[ (2.11) \quad \phi_5 = \begin{cases} 1 - F(\frac{1}{2d}(1 + \frac{1}{2N_2})^{-\frac{1}{2}}), & d > 0 \\ F(\frac{1}{2d}(1 + \frac{1}{2N_2})^{-\frac{1}{2}}), & d < 0. \end{cases} \]

Method 7 yields in the limit as \( \sigma_2^2 \to \infty \) an estimator which is identically equal to \( \frac{1}{2} \) and which will therefore have a bias.

\(^2\text{See footnote on notation in Section 2.5.}\)
which does not go to zero as \( N \to \infty \). Method 8 yields for this univariate normal situation the same estimator as Method 3. Methods 6 and 9 lead to complicated expressions and will not be considered further.

In summary, the use of a prior distribution on \((\mu_1, \mu_2)\) has yielded two estimators, \( \hat{\theta}_3 \) and \( \hat{\theta}_5 \), which will be considered as estimators for \( \theta_2 \). As noted at the beginning of the chapter, all of the estimators to be studied are listed in Table III.


The conditional means and variances given \( x_1, x_2 \) fixed,\(^3\) but with test samples assumed to be random, were computed for the estimators listed in Table III. The results are given in Table IV using notation from Table I, and the computations are given in Appendix II.

The conditional mean square error (MSE) for each estimator can be obtained from Table IV. By definition, if \( \theta \) is the parameter to be estimated and \( \hat{\theta} \) is an estimator, the mean square error of \( \hat{\theta} \) is

\[
\text{MSE}(\hat{\theta}) = (E(\hat{\theta}) - \theta)^2 + \text{Var}(\hat{\theta}) = (\text{bias} \ \hat{\theta})^2 + \text{Var} \ \hat{\theta}.
\]

In this case, since it is the conditional MSE that is wanted, the expectation and variance will be conditional moments. The MSE will be useful for comparing estimators of \( \theta_2 \) since, as will be seen, for some estimators the bias is important and for others the variance. The MSE takes both bias and variance into account.

Before examining the results we discuss some problems that arise in comparing the estimators on the basis of conditional moments. One problem arises because the estimators are not all based on the same total number of observations. Some estimators are based only

\(^3\)As pointed out in Section 1, fixing \( x_1, x_2 \) is not equivalent to fixing \( R(s_1, s_2) \), but the conditional moments are moments for \( x_1, x_2 \) fixed.
on observations from original samples. Using such an estimator the rule plus estimates of both conditional probabilities of misclassification--$P_2(R(\bar{x}_1, \bar{x}_2), \mu_2)$ and $P_1(R(\bar{x}_1, \bar{x}_2), \mu_1)$--can be obtained from the $N_1 + N_2$ observations of the original samples. On the other hand, if independent test samples are involved in the estimator, $N_1 + N_2$ observations are needed to get the rule, and an additional $M_1 + M_2$ observations are needed to estimate $P_1$ and $P_2$. Thus any decreases in bias or variance brought about by using estimators requiring test samples must be balanced against the cost of the additional observations.

For Lachenbruch's $U$ estimator, $P_U$, we have computed the conditional moments given $\bar{x}_1, \bar{x}_2$ fixed, in accordance with the discussion in Section 1. The $N_2$ rules that are used in the calculation of $P_U$ then have a joint distribution arising from the conditional distributions of the samples $S_1, S_2$ given fixed values of $\bar{x}_1, \bar{x}_2$. This joint distribution determines the conditional behavior of $P_U$ which we have studied.

Consider next the estimator $P_{TSL}$. This estimator involves two sets of original samples and the corresponding two rules. The one set of original samples are of size $N$ each, while the second set of original samples and the test samples, assumed to be obtained by splitting the samples of size $N$, are of size $\frac{1}{2}N$. The sample means based on original samples of size $N$ and $\frac{1}{2}N$ are denoted respectively $\bar{x}_1, \bar{x}_2$ and $\bar{y}_1, \bar{y}_2$.

We believe that the most relevant computations for making comparisons with our results for other estimators require $\bar{x}_1$ and $\bar{x}_2$ to be fixed but $\bar{y}_1$ and $\bar{y}_2$ to be random, and of course the $\bar{x}$'s and $\bar{y}$'s are not independent since the latter are based on subsamples of the observations determining the former. For this model (which
we will call Assumption Set 1) we have obtained the expectation of $P_{TSL}'$ given in Table IV. The calculation of the conditional variance is, however, extremely tedious and hence this calculation was made under the simpler assumption that all four means $\bar{x}_1, \bar{x}_2, \bar{y}_1, \bar{y}_2$ are fixed and moreover that all six samples (one of size $N$ and two of size $\frac{1}{2}N$ from each population) are independent. We call this Assumption Set 2. For completeness we also give the expectation and MSE of $P_{TSL}$ under Assumption Set 2. The simplifying assumption of independence allows us to use values already obtained for conditional means and variances of $p_0$ and $p_t$.

Now looking at the results in Table IV, some facts are immediately obvious. The test sample estimator, $P_T$, and the minimum variance estimator, $\hat{\rho}_{RB}$, are the only unbiased\textsuperscript{4} estimators, although $Q_T$ appears to have a small bias when $M_2$ is large. All three of these estimators are based on independent test samples. In fact, no unbiased estimator of $P_2(R(\bar{x}_1, \bar{x}_2), \mu_2)$ based only on the original samples exists, because the conditional distribution of the original observations from $\pi_2$, when $\bar{x}_2$ is fixed, is independent of $\mu_2$ and thus no function of the original observations can have a conditional expectation which is a function of $\mu_2$. Another result is that the adjusted reclassification estimator, $P_{Ra}$, and the estimator $Q_D$ have the same expected value and hence the same bias.

Several estimators, namely $Q_D$, $Q^*_D$, $Q^{**}_D$, $\rho_3$ and $\rho_5$, have conditional variances equal to zero, because these estimators are constant when $\bar{x}_1$ and $\bar{x}_2$ are fixed. Note that all of these estimators are based only on original observations and all of them make use of normality assumptions. The reclassification estimator,

\textsuperscript{4}When not specified otherwise, moments referred to will be conditional moments given $\bar{x}_1, \bar{x}_2$ fixed.
and the adjusted reclassification estimator, \( P_{Ra} \), have the same variance because \( P_{Ra} = P_R + \text{constant} \). The variance of \( P_R \) is smaller than the variance of \( P_{TSL} \) (when \( N_2 = N \)) since \( \text{Var} P_{TSL} = \text{Var} P_R + \text{two terms both greater than zero} \). (As noted above \( \text{Var} P_{TSL} \) is computed under Assumption Set 2.) \( P_{RB} \) must have a smaller variance than any other unbiased estimator based on an additional independent test sample from \( \pi_2 \), since \( P_{RB} \) is the conditionally UMVU estimator under these conditions. Estimators based only on original samples can, however, have a smaller variance than \( P_{RB} \), and those estimators mentioned above as having variance equal to zero obviously do.

From Table IV, it can be seen that no estimator with zero variance is unbiased, so all estimators have a positive MSE. It follows from the above remarks, that the MSE of \( P_{Ra} \) is greater than the MSE of \( Q_D \), because the two estimators have the same bias whereas \( \text{Var} Q_D = 0 \) and \( \text{Var} P_{Ra} > 0 \). Also \( \text{MSE}(Q_T) \geq \text{MSE}(P_{RB}) \) because both are unbiased and \( P_{RB} \) is the minimum variance unbiased estimator in the situation where data consists of original samples from \( \pi_1, \pi_2 \) plus a test sample from \( \pi_2 \). In general, however, the MSE expressions for finite sample sizes are not readily comparable. Note also that most of the comparisons made above are comparisons within the groups of estimators based only on original samples or within the group of estimators requiring test samples. It is true that the unbiased estimators are based on test samples and zero variance estimators are among those using only original samples, but if estimators are to be compared on MSE, there is no convenient way to tell if taking test samples is worthwhile. Therefore in order to better compare the estimators, we now consider asymptotic expressions for the squared bias, the variance, and the mean square error.

- 34 -
4. Asymptotic Expressions for Conditional Squared Bias, Variance, and Mean Square Error (MSE).

In Section 4.1 the method of obtaining the asymptotic expressions is discussed. The results are presented in Section 4.2 and the conclusions suggested by the results are discussed in Section 4.3. Section 4.4 contains some remarks on the distinction between estimators of \( \bar{P}_2, \bar{P}_2^*, \) and \( \bar{P}_2^{**} \).

4.1 Methods.

The conditional bias, variance and MSE were computed for \( \bar{x}_1, \bar{x}_2 \) fixed and hence depend on \( \bar{x}_1 \) and \( \bar{x}_2 \), which are random variables. Thus an asymptotic comparison of these quantities is essentially a comparison of their distributions. All become small in a stochastic sense as sample sizes become large. Hence the comparisons are made by looking at smallness as measured by the \( o_p(\cdot) \) function of Mann and Wald (1943), where \( W_N = o_p(f_N) \) if \( W_N/f_N \to 0 \) in probability as \( N \to \infty \).

To simplify the discussion of the asymptotic results we will assume here that \( N_1 = N_2 = M_1 = M_2 \) with the common value denoted by \( N \). Thus there is only one sample size to go to infinity. The simplification is, however, achieved at the cost of some loss of information, since with all sample sizes equal to \( N \), it is not possible to keep separate those terms which go to zero as \( N_1, N_2 \to \infty \), and those which go to zero as \( M_1, M_2 \to \infty \). From Table IV, of course, it can easily be determined if a particular bias or variance goes to zero as a specified sample size goes to infinity, but the rate of convergence is unknown.

To derive the asymptotic values, which are given in Tables VI and VII, from the conditional means and conditional variances of Table IV, we require Taylor series expansions of functions of the
forms $F(K), F(K,K; \rho) - F(K,K; 0)$, and $F(K,K; \rho)$, where $K$ stands for various functions, for example $c, \frac{1}{2}d$, or $\frac{1}{2}d(1 + 1/N)^{-\frac{3}{2}}$. To get the expansions, the symbols for which $K$ stands are written in the form $a + t$, where $a$ is constant and independent of $N$ and $t$ is variable (when the rule is variable) or a function of $N$.

The needed $a + t$ expressions are given in Table V. The constant $a$ is always equal to $\frac{1}{2}6, \delta = \mu_1 - \mu_2$, while the various $t$'s have terms which are functions of $x_1 - \mu_1, x_2 - \mu_2, N^{-1}$, etc.

These terms are grouped according to their order in the $\tilde{6}_p(\cdot)$ sense. Let $\alpha_j = x_j - \mu_j$ and $\alpha_j^* = y_j - \mu_j^*$. By the lemma in Appendix IIIB, $\alpha_j$ and $\alpha_j^*$, $j = 1, 2$ are $\tilde{6}_p(N^{-\frac{3}{2}+\varepsilon})$ as $N \to \infty$ for any $\varepsilon > 0$.

Then $\alpha_j^2$ and $(\alpha_j^*)^2$ are $\tilde{6}_p(N^{-1+\varepsilon})$. (See Mann and Wald (1943)).

A term which is equal to $kN^{-\alpha}$ where $k$ and $\alpha > 0$ are constants is $\tilde{o}(N^{-\alpha+\varepsilon})$ as $N \to \infty$ for any $\varepsilon > 0$. If a term is $\tilde{o}(N^{-\alpha+\varepsilon})$, it is also $\tilde{6}_p(N^{-\alpha+\varepsilon})$, and so in the asymptotic expressions the terms which are $\tilde{o}(N^{-\alpha+\varepsilon})$ and $\tilde{o}(N^{-\alpha+\varepsilon})$ are grouped together and denoted as of order $\tilde{6}_p(N^{-\alpha+\varepsilon})$.

With $F(x)$ and $f(x)$ denoting the standard normal distribution and density functions, respectively, and using the fact that $f'(x) = -xf(x), f''(x) = (x^2 - 1)f(x)$, we get the Taylor expansion

$$F(a + t) = F(a) + F'(a)t + F''(a)\frac{1}{2}t^2 + F'''(a)\frac{t^3}{6} + \ldots$$

$$= F(a) + f(a)t - \frac{1}{2}a f(a)t^2 + \frac{1}{6}(a^2 - 1)f(a)t^3 + \ldots$$

This formula is used to get most of the $F(K)$ expansions listed in Table V. The expansions in Table V involving the bivariate normal distribution are derived in Appendix IIIC. The functions in Table V which have forms such as $F(-\frac{1}{2}d\sqrt{2N}), F(-d\sqrt{N(N-1)}, -d\sqrt{N(N-1)}; \rho)$, and $F(-\frac{1}{2}d\sqrt{2N}, \frac{1}{2}d(1 - \frac{3}{2N})^{-\frac{1}{2}})$ are special cases where, for $d > 0$, one or more arguments go to $-\infty$ as $N \to \infty$. Using the inequality,
1 - F(x) \leq \frac{1}{x} f(x) \text{ for all } x > 0, \text{ (see, for example, Feller (1957, p. 166) it is easily shown that these functions are } \delta_p(N^{-\alpha}), \alpha > 0.

Table VI gives the squared biases and their asymptotic forms obtained from the results in Table V. Asymptotic expressions for the variances of the estimators are given in Table VII. Asymptotic expressions for the MSE can be obtained from the results given in Tables VI and VII. These MSE expressions are given in Table VIII. Some numerical results are given in Table IX for selected parameter values and for all estimators except \( P_{TSL} \).

4.2 The results on asymptotic conditional bias, variance, and MSE.

Looking at Table VI we see, as previously noted, that the estimators \( P_T \) and \( P_{RB} \) are unbiased estimators of \( P_2(R(\bar{x}_1, \bar{x}_2), \mu_2) \). The estimator \( Q_T \) has a relatively small bias with the leading term in the squared bias of the order \( o_p(N^{-2+\epsilon}) \). For all the other estimators considered, the leading term in the squared bias is equal to \( o_p(N^{-1+\epsilon}) \) and, in fact, for all these estimators (except \( P_{TSL} \) when the expectation is computed under Assumption Set 2) the leading term is equal to \( g^2\alpha_2^2 \), where \( g = f(\frac{1}{2}\delta) \) and \( \alpha_2 = \bar{x}_2 - \mu_2 \). Since from the exact results it was known which estimators had no bias or small bias (see Section 3), the main information gained about the bias from the asymptotic results is this fact that the estimators \( P_R, P_{TSL} \) (Assumption Set 1), \( P_U, P_{Ra}, Q_D, Q^*_D, Q^{**}_D, P_3 \), and \( P_5 \) all have squared biases equal to \( g^2\alpha_2^2 + o_p(N^{-3/2 + \epsilon}) \). In particular note that the biases of \( P_R \) and \( P_{Ra} \) are equal, to this order of approximation, even though \( P_{Ra} \) was specifically designed to decrease the bias of \( P_R \). Note also that the \( P_{TSL} \) mean computed under Assumption Set 1 is the one most comparable to the other conditional means in terms of assumptions.

\(^5\text{In the discussion, moments are conditional moments given } \bar{x}_1 \text{ and } \bar{x}_2, \text{ unless otherwise specified.}\)
For each of this group of estimators, with leading term in the squared bias equal to $g^2 \alpha_2^2$, the term of order $\hat{\sigma}_p(N^{-3/2} + \varepsilon)$ is given in Table VI separated from the term of order $o_p(N^{-2+\varepsilon})$. This $o_p(N^{-3/2} + \varepsilon)$ term is of the form $-\frac{1}{2} \delta g^2 \alpha_2^2 (\alpha_1 \alpha_2 + kN^{-1})$. For the estimators $P_U$, $P_{Ra}$, and $Q_D$, $k = 0$; for $P_R$, $k = 1$; for $P_{TSL}$ (Assumption Set 1), $Q_{DS}^{**}$, and $P_3$, $k = -1$; and for $Q_D^*$ and $Q_5$, $k = -\frac{1}{2}$. In the case that $\delta \alpha_2 > 0$, this $o_p(N^{-3/2} + \varepsilon)$ term in the squared bias decreases as $k$ increases, and in the case that $\delta \alpha_2 < 0$, the squared bias decreases as $k$ decreases. Thus which of the estimators has larger squared bias when terms of order $o_p(N^{-1+\varepsilon})$ and $o_p(N^{-3/2} + \varepsilon)$ are considered depends on the unknown parameter, $\text{sgn}(\delta \alpha_2)$. However, the expected value of the $o_p(N^{-3/2} + \varepsilon)$ term when $\alpha_1$ and $\alpha_2$ are considered random is zero, so that on the average this term is equal for all of this group of estimators.

Note that this group of estimators with squared bias equal to $g^2 \alpha_2^2 + o_p(N^{-3/2} + \varepsilon)$ consists of some estimators which use normality assumptions and some which do not. All of them, however, are based only on original observations.

Looking at Table VII, we see that several estimators have conditional variances equal to zero, as has already been discussed in Section 3. For all those estimators with non-zero variance the leading term in the variance is of the order $o_p(N^{-1+\varepsilon})$, and is of the form $[k_1 G(1 - G) + k_2 g^2]N^{-1}$, where $G = F(\frac{1}{2} \delta)$, $g = f(\frac{1}{2} \delta)$, and $k_1$, $k_2$ are constants. For the reclassification estimator, $P_R$, and for Lachenbruch's $U$ estimator, $P_U$, $k_1 = 1$ and $k_2 = -1$. For the test sample estimator, $P_T$, $k_1 = 1$ and $k_2 = 0$. Hence the leading term in the variance expansion for $P_R$ and $P_U$ is smaller than the corresponding term for $P_T$. The variance of $P_{Ra}$ is equal to that for $P_R$, as was pointed out in the discussion of the exact variance.
expressions. For both $Q_T$ and the Blackwellized estimator, $\mathcal{O}_{RB}$, the leading term in the variance is equal to $g^2N^{-1}$, that is $k_1 = 0$ and $k_2 = 1$, and the next term is equal to $-\frac{1}{2}g^2(\alpha_1 + \alpha_2)N^{-1}$. The leading term in the variance of $P_T$, which is the other estimator besides $\mathcal{O}_{RB}$ and $Q_T$ based on independent test samples, is equal to $G(1 - G)N^{-1}$, which is greater than $g^2N^{-1}$. Note that $P_T$ does not use the normality assumption, while $\mathcal{O}_{RB}$ and $Q_T$ do and hence, using the normality decreases the variance, as one would expect.

Let AMSE stand for "asymptotic mean square error." It is understood that the moments involved are conditional. Looking at the AMSE's in Table VIII, the most striking fact is that for every estimator considered the leading term is of the order $o_p(N^{-1+\epsilon})$. For the estimators $Q_D$, $Q^*_D$, $Q^{**}_D$, $P_3$, and $\mathcal{O}_5$, all of which make use of the normality assumption and are based only on original samples, this leading term is equal to $g^2\alpha_2^2$. For the estimators $Q_T$ and $\mathcal{O}_{RB}$, which make use of the normality assumption and test samples, the leading term in the AMSE is equal to $g^2N^{-1}$. Note that $g^2N^{-1}$ is the expected value of $g^2\alpha_2^2$ when $\alpha_2$ is considered random.

For the reclassification estimator, $P_R$, the adjusted reclassification estimator, $P_{Ra}$, and Lachenbruch's $U$ estimator, $P_U$, all of which use only original samples, the leading term in the AMSE is equal to $g^2\alpha_2^2 + [G(1 - G) - g^2]N^{-1}$ which is greater than $g^2\alpha_2^2$ because $G(1 - G) > g^2$. The leading term for the test sample estimator, $P_T$, is $G(1 - G)N^{-1}$, which is the expected value of $g^2\alpha_2^2 + [G(1 - G) - g^2]N^{-1}$ when $\alpha_2$ is random. Note that none of the estimators $P_R$, $P_T$, $P_U$ make use of distribution assumptions. The adjusted reclassification estimator does use the normality assumption to get the adjustment to $P_0$, but the estimator is asymptotically equivalent to the ordinary reclassification estimator and might be
considered as an estimator which does not essentially use the normality assumption.

Leaving the two straight lines, estimator $R_{TSL}$ aside, the results on AMSE can be summarized as follows:

i) For all those estimators based only on original samples and using the normality assumption, the leading term in the AMSE is the same.

ii) The estimators $P_R$ and $P_U$, (and $P_{Ra}$), which are based only on original samples and do not use the normality assumption, have the same leading term in the AMSE. This term is greater than the common leading term for those estimators using original samples and normality.

iii) For the two estimators based on test samples and normality, that is $Q_T$ and $P_{RB}$, the leading term in the AMSE is the same.

iv) The leading term in the AMSE for the estimator $P_T$, which uses test samples but does not make use of the normality, is greater than the common leading term in the AMSE of $Q_T$ and $P_{RB}$.

v) Among the estimators using the normality, the two estimators based on test samples have leading term in the AMSE equal to the unconditional expectation of the corresponding term for the estimators based on original samples.

vi) The leading term in the AMSE of $P_T$ is equal to the unconditional expectation of the corresponding term for $P_R$ and $P_U$.

4.3 Conclusions.

There are three major conclusions suggested by the results presented above. First, within the group of estimators using only original observations and within the group using test samples, the estimators of $P_2(R(x_1, x_2), \mu_2)$ which make use of the normality are better, when the criterion is size of the conditional AMSE with the
moments computed under the normality assumptions, than estimators not using the normality. This conclusion is not surprising, since if the observations are known to have a particular distribution, estimators of $P_2$ which make use of this distribution information should, intuitively, be better than those which do not.

Secondly, within the group of estimators using the normality and within the group (three only, or four if REa is included) not using the normality, the estimators based only on original samples have AMSE's which on the average (unconditional) are equal to the AMSE for those which use test samples. Hence, since under the sample size assumption used here ($N_1 = N_2 = M_1 = M_2 = N$), estimators based on independent test samples require twice as many observations to get the rule and estimates of $P_1$, $P_2$ as do the estimators based only on original observations, their use is of questionable value. Such estimators might, however, be used to protect against extreme values. Note that if $\alpha_2$ is large, the AMSE of estimators based on normality and original samples cannot only be larger than the AMSE of estimators based on normality and test samples, but can be larger than the AMSE of the test sample estimator, which does not make use of the normality assumption.

The third conclusion is really several suggested conclusions, theoretical and practical, based on the observed result that the estimators of a given kind are equivalent in terms of AMSE. All the estimators considered here which are based on normality and original samples have the same first order term in the AMSE (except the estimator $P_{Ra}$, which uses the normality only to adjust the basic estimator $P_R$). Similarly the two estimators based on normality and test samples, $Q_T$ and $Q_{RB}$, are equivalent in terms of AMSE, as are the estimators $P_R$ and $P_U$ (and $P_{Ra}$) based on original samples and
not using normality. The MSE for the estimator $P_{TSL}$, which belongs in the latter group, was not computed under the assumptions needed for appropriate comparison, as has been discussed above in Section 3. In the class of estimators based on independent test samples and not using normality only the one estimator, $P_T$, was studied.

Looking first at the group of estimators based on normality and original samples, the equivalence of all the estimators ($Q_D$, $Q_D^*$, $Q_{DS}$, $Q_3$, and $Q_5$) raises the question as to how much "information" there is in the original samples for estimating the probability of misclassification. Can an estimator based on normality and original samples be found which has a smaller AMSE than those studied here, or is the observed AMSE, $g^2a^2 + o_p(N^{-3/2 + \epsilon})$, a lower bound? Note that all of these estimators have conditional variance equal to zero, and hence reducing the bias is equivalent to reducing the MSE.

Thus consider the problem of reducing the bias. All of the estimators, $Q_D$, $Q_D^*$, $Q_{DS}$, $Q_3$, and $Q_5$, that are based on normality and original samples and that have AMSE (and asymptotic squared bias) equal to $g^2a^2 + o_p(N^{-3/2 + \epsilon})$ can be written as $P_2$, $P_2^*$, or $P_2^{**}$ with parameters estimated (although $Q_3$ and $Q_5$ were not derived in this way). Hence they are a restricted class of original sample estimators. However, the three estimators $P_R$, $P_{TSL}$, and $P_U$, which do not use normality at all, and the estimator $P_{Ra}$, which uses the normality only to get an adjustment to $P_R$, also have asymptotic squared bias equal to $g^2a^2 + o_p(N^{-3/2 + \epsilon})$, (expectation of $P_{TSL}$ computed under Assumption Set 1). Taken all together these results suggest that the observed value is a lower bound on the conditional asymptotic squared bias for estimators based on original samples. If this is so, then the estimators with conditional variance equal to zero have minimum conditional AMSE. In Appendix VI, we discuss a
conjecture that no estimator exists based on original samples having an asymptotic conditional squared bias less than \( g^2 \alpha^2 + o_p(N^{-3/2} + \epsilon) \).

The asymptotic equivalence of the estimators \( Q_D, Q_D^*, Q_{DS}^*, \) and \( \mathcal{Q}_5 \) also suggests the practical conclusion that if an estimator of the \( Q, Q^*, \) or \( Q^{**} \) type is to be used, it doesn't matter which one. For finite sample sizes, however, the estimators may differ. (See the discussion of the Lachenbruch and Mickey (1968) results in Chapter 4, Section 5.) Now consider again the group of estimators based on original samples and not using normality--\( P_R, P_{TSL}, P_U, \) and \( P_{Ra} \). The results on \( P_{TSL} \) are unfortunately incomplete, but the asymptotic equality of all the biases and the asymptotic equality of the MSE's of \( P_R, P_U, \) and \( P_{Ra} \) suggest that it is not worthwhile to use an estimator such as \( P_U \) or \( P_{TSL} \) in place of \( P_R \), which is conceptually and computationally simpler. Despite asymptotic equalities which we have obtained, Lachenbruch and Mickey (1968) found that \( P_R \) performs poorly for finite sample sizes when the distributions are multivariate normal (see the discussion in Chapter 4, Section 5). Of course, in a practical situation, if an estimator not based on normality is being used because the normality assumption is suspect, neither the asymptotic nor the small sample performance of these estimators under normality is of much interest.

As far as the estimators based on normality and test samples are concerned, it is known that \( \mathcal{Q}_{RB} \) is the conditional UMVU estimator in this class. It is not known if an estimator with smaller conditional MSE exists. The estimator \( Q_T \), which is slightly biased, has an AMSE with the first two terms equal to the AMSE for \( \mathcal{Q}_{RB} \).

Summarizing briefly, the first two conclusions on the problem of estimating the conditional probability of misclassification in the univariate normal case, are that using the normality helps and using
independent test samples does not. The third conclusion is related to the observed asymptotic equivalence of estimators of a given kind.

4.4 Remarks on the distinction between estimators of $P_2$, $P_2^*$, and $P_2^{**}$.

Before taking up the problem of estimating $P_2$ when the distributions are multivariate normal (Chapter 4), we will make some remarks on the problem of distinguishing between estimators of $P_2$, $P_2^*$, and $P_2^{**}$. As noted above the estimator $Q_0^*$, which is a "natural" estimator for $P_2^*$, and the estimators $Q_0^{**}$, $P_3$, and $Q_5$, which are "natural" estimators for $P_2^{**}$ are as good (in terms of size of AMSE) as estimators of $P_2$ as is $Q_0^*$, which is a "natural" and commonly used estimator of $P_2$. Of course, any function of the observations can be considered as an estimator for any parameter. Also since $P_2^*$ is the expected value of $P_2$, and $P_2^{**}$ is the limiting value of $P_2$ as sample sizes go to infinity, estimating $P_2$ by estimating parameters in the expression for $P_2^*$ or $P_2^{**}$ is not so odd. Also (as shown in Section 2.4) the estimator $Q_0^*$ can itself be obtained by estimating the unknown parameter in $P_2^{**}$ and thus is a $Q^{**}$ type estimator. However, since a number of workers have commented on this problem (see for example Lachenbruch and Mickey (1968, p. 724), Hills (1966), and Section 1.4, Chapter 2), we will examine the situation.

It is known that for random variables $u$, $v$ the quantity

$$[E(u - v)^2]^{1/2}$$

is a metric and so may be considered to be the "distance" between $u$ and $v$. Thus the conditional mean square error

$$E[(P_2 - P_2)^2|\bar{x}_1, \bar{x}_2]$$

measures the squared distance between any estimator $P_2$ and $P_2$. When estimating $P_2^*$ (or $P_2^{**}$) it is natural to use the unconditional MSE $E(P_2^* - P_2^2)^2$ for any $P_2$. Similarly one can consider the "distances" between the three unknowns $P_2$, $P_2^*$ and $P_2^{**}$ (for simplicity $E(P_2 - P_2^2)^2$, etc., will also be called a MSE). The asymptotic forms of these quantities are revealing in a discussion
Figure 2

Distances between $P_2$, $P_2^*$, $P_2^{**}$ and between these probabilities and some estimators.

The "distances" marked in the diagram are the leading terms in the AMSE.
of the appropriateness of any possible estimator of each of the unknowns.

Figure 2 shows the leading terms in several unconditional MSE's whose derivations are given in Appendix V. Consider first the three unknowns \( P_2, \hat{P}_2^*, \) and \( \hat{P}_2^{**} \). \( P_2 \) has the same squared distance from both \( \hat{P}_2^* \) and \( \hat{P}_2^{**} \), namely, \( \frac{1}{28}N^{-1} \), this value being the leading term in the asymptotic expansion for both of the expressions

\[
E(P_2 - P_2^*)^2 = EP_2^2 - (P_2^*)^2
\]
and

\[
E(P_2 - P_2^{**})^2 = EP_2^2 - 2P_2^*P_2^{**} + (P_2^{**})^2
\]
where

\[
EP_2^2 = F(\alpha)F(-\beta, -\beta; \frac{1}{2N+1}) + F(-\alpha)F(\beta, \beta; \frac{1}{2N+1})
\]

\[
P_2^* = F(\alpha)F(-\beta) + F(-\alpha)F(\beta)
\]

\[
P_2^{**} = F(-\frac{1}{2}|\delta|)
\]

\[
\alpha = \frac{1}{2}\delta(2N)^{\frac{1}{2}}
\]

\[
\beta = \frac{1}{2}\delta(1 + \frac{1}{2N})^{-\frac{1}{2}}.
\]
On the other hand \( \hat{P}_2^* \) and \( \hat{P}_2^{**} \) are closer since asymptotically

\[
E(P_2 - P_2^{**})^2 = \frac{1}{64} \delta^2 gN^{-2} + o(N^{-3+c}),
\]
which is of order \( N^{-2} \) rather than \( N^{-1} \).

Next let us consider the two estimators

\[
Q_D^* = P_2^*(R(\cdot, \cdot), \hat{\delta} = d)
\]
and

\[
Q_D = P_2(R(\overline{x}_1, \overline{x}_2), \hat{\mu}_2 = \overline{x}_2)
\]
\[
= \hat{P}_2^{**}(\delta = d).
\]
The definition of \( Q_D^* \) justifies the assertion that \( Q_D^* \) is a natural estimator of \( P_2^* \). Indeed, Figure 2 shows that \( Q_D^* \) estimates \( P_2^* \) better than it estimates \( P_2 \), since the respective unconditional AMSE's are \( \frac{1}{2}g_2n^{-1} \) and \( g_2n^{-1} \) (see Appendix V and Table VIII). Turning to \( Q_D \), the equivalent definitions displayed above show that \( Q_D \) can be regarded as a natural estimator of either \( P_2 \) or \( P_2^{**} \). It is perhaps surprising to find from Figure 2 that (like \( Q_D^* \)) \( Q_D \) actually estimates \( P_2^* \) better than it estimates \( P_2 \), the unconditional AMSE's being exactly the same as those just given for \( Q_D^* \). A possible explanation of this lies in the second definition of \( Q_D \) as an estimator of \( P_2^{**} \) and in the closeness of \( P_2^* \) and \( P_2^{**} \).

One more conclusion we can draw from Figure 2 is that both \( P_2^* \) and \( P_2^{**} \) are closer to \( P_2 \) than is either of the typical estimators \( Q_D^* \) and \( Q_D \). Thus one might expect that a natural estimator of \( P_2 \) would also be a reasonable estimator of \( P_2^* \) and vice versa.
Table I
Notation Used in the Discussion of the Univariate Normal Case

$\pi_1, \pi_2$ are populations distributed $N(\mu_1, \sigma^2_1), N(\mu_2, \sigma^2_2)$ respectively.

$\bar{x}_1, \bar{x}_2$ are sample means based on original samples of sizes $n_1, n_2$
from $\pi_1, \pi_2$.

$\bar{y}_1, \bar{y}_2$ are sample means based on original samples from $\pi_1, \pi_2$ of size $\frac{n_1}{2}$ each.

$x$ is a new observation to be classified.

$\bar{t}_1, \bar{t}_2$ are sample means based on test samples of sizes $m_1, m_2$ from $\pi_1, \pi_2$.

$d = \bar{x}_1 - \bar{x}_2$
$d^* = \bar{y}_1 - \bar{y}_2$
$\delta = \mu_1 - \mu_2$

$\alpha_j = \bar{x}_j - \mu_j$
$\alpha_j^* = \bar{y}_j - \mu_j$, $j = 1, 2$

$b = \frac{1}{2}d(1 - 1/n_2)^{-\frac{1}{2}}$
$b^* = \frac{1}{2}d^*(1 - 2/n_2)^{-\frac{1}{2}}$

$c = \frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_2 = \frac{1}{2}d + \alpha_2$
$c^* = \frac{1}{2}d^* + \alpha_2^*$

$F(u), F(u, v; \rho)$ denote standard normal univariate and bivariate distribution functions as defined in (1.3) and (1.4), Chapter 3.

$H(u, v; \rho) = F(-u, -v; \rho)$
$f(u) = F'(u)$
$G = F(\frac{1}{2}\delta)$
$g = f(\frac{1}{2}\delta)$

$E(u|R(S_1, S_2))$ denotes expectation of $u$ with $R(S_1, S_2)$ fixed but test samples random. We actually fix $\bar{x}_1$ and $\bar{x}_2$ rather than $R(S_1, S_2)$ and write $E(u|\bar{x}_1, \bar{x}_2)$.

$p_0 = \text{proportion of original sample of size } n_2 \text{ from } \pi_2 \text{ misclassified by } R(S_1, S_2)$.

$p_t = \text{proportion of test sample of size } m_2 \text{ from } \pi_2 \text{ misclassified by } R(S_1, S_2)$.
Table I (cont.)

Probabilities of Misclassification:

$P_2$ and $P_2^*$ denote conditional and unconditional probabilities of
misclassification as given in (1.6) and (1.7) of Chapter 3.

$P_2^{**}$ is the probability of misclassification when all parameters
are known as given in (1.10), Chapter 3.

Estimators:

$P$ with a letter(s) subscript denotes an estimator of $P_2$
not based on normality.

$Q, Q^*, Q^{**}$ with subscripts denote estimators of $P_2$
obtained by estimating unknown parameters in $P_2, P_2^*, P_2^{**}$.

$\phi$ with a subscript denotes other estimators based on normality.
Table II

Expressions for Estimators of $P_2(R(x_1, x_2), \mu_2)$

Derived by Using a Prior Distribution on $\theta$

Populations: $\Pi_1$ is $N(\mu_1, 1), \Pi_2$ is $N(\mu_2, 1); \theta = (\mu_1, \mu_2)$

Prior Distribution: $\mu_i$ is $N(\eta_i, \sigma_i^2), i = 1, 2$ with $\mu_1, \mu_2$ independent

Posterior Distribution: $\mu_i$ is $N(\eta_i, \tau_i^2), i = 1, 2$ with $\mu_1, \mu_2$ independent;

\[
\eta_i = \frac{\nu_i + N_i \sigma_i^2 x_i}{1 + N_i \sigma_i^2}, \quad \tau_i^2 = \frac{\sigma_i^2}{1 + N_i \sigma_i^2}
\]

$\nu_i \to x_i$ as $\sigma_i^2 \to \infty$, $\tau_i^2 \to 1/N_i$ as $\sigma_i^2 \to \infty$.

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimator</th>
<th>Limit of Estimator as $\sigma_i^2, \sigma_2^2 \to \infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1: Estimate $\mu_2$ by mean of posterior distribution of $\mu_2$</td>
<td>$1 - F(\frac{1}{2d} + x_2 - \nu_2), d &gt; 0$</td>
<td>$1 - F(\frac{1}{2d}), d &gt; 0$</td>
</tr>
<tr>
<td></td>
<td>$F(\frac{1}{2d} + x_2 - \nu_2), d &lt; 0$</td>
<td>$F(\frac{1}{2d}), d &lt; 0$</td>
</tr>
<tr>
<td>Method 2: Use that estimate of $\mu_2$ which minimizes posterior expected loss (in estimating $\mu_2$)</td>
<td>Estimator derived only for loss function $(\mu_2 - \hat{\mu}_2)^2$, in which case method 2 is equivalent to method 1.</td>
<td></td>
</tr>
<tr>
<td>Method 3: Take expectation of $P_2$ with respect to posterior distribution of $(\mu_1, \mu_2)$</td>
<td>$1 - F(\frac{\frac{1}{2d} + x_2 - \nu_2}{\sqrt{1 + \tau_2^2}}, d &gt; 0$</td>
<td>$1 - F(\frac{\frac{1}{2d}(1+N_2)^{-1} - \frac{1}{2}}{\sqrt{1 + \tau_2^2}}), d &gt; 0$</td>
</tr>
<tr>
<td></td>
<td>$F(\frac{\frac{1}{2d} + x_2 - \nu_2}{\sqrt{1 + \tau_2^2}}, d &lt; 0$</td>
<td>$F(\frac{\frac{1}{2d}(1+N_2)^{-1} - \frac{1}{2}}{\sqrt{1 + \tau_2^2}}), d &lt; 0$</td>
</tr>
<tr>
<td>Method 4: Use that estimate of $P_2$ which minimizes posterior expected loss (in estimating $P_2$)</td>
<td>Estimator derived only for loss function $(P_2 - \hat{P}_2)^2$, in which case Method 4 is equivalent to Method 3.</td>
<td></td>
</tr>
</tbody>
</table>
Table II (cont.)

<table>
<thead>
<tr>
<th>Method 5:</th>
<th>( 1 - F\left(\frac{\frac{1}{2}d + \bar{x}_2 - \mu_X}{\sigma_X}\right), d &gt; 0 )</th>
<th>( 1 - F\left(\frac{\frac{1}{2}d}{\sqrt{1 + \frac{1}{2N_2}}}\right), d &gt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( F\left(\frac{\frac{1}{2}d + \bar{x}_2 - \mu_X}{\sigma_X}\right), d &lt; 0 )</td>
<td>( F\left(\frac{\frac{1}{2}d}{\sqrt{1 + \frac{1}{2N_2}}}\right), d &lt; 0 )</td>
</tr>
<tr>
<td></td>
<td>Where ( \mu_X = \frac{Y_2 + 2N_2\sigma_z^2x_2}{1 + 2N_2\sigma_z^2} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \sigma_X^2 = \frac{1 + \sigma_z^2 + 2N_2\sigma_z^2}{1 + 2N_2\sigma_z^2} )</td>
<td></td>
</tr>
<tr>
<td>Method 6:</td>
<td>No estimator derived by this method.</td>
<td></td>
</tr>
<tr>
<td>Method 7:</td>
<td>Take the expectation of ( P_2 ) with prior distribution of ( (\mu_1, \mu_2) )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( 1 - F\left(\frac{\frac{1}{2}d + \bar{x}_2 - \gamma_2}{\sqrt{1 + \sigma_z^2}}\right), d &gt; 0 )</td>
<td>( \frac{1}{2}, d &gt; 0 )</td>
</tr>
<tr>
<td></td>
<td>( F\left(\frac{\frac{1}{2}d + \bar{x}_2 - \gamma_2}{\sqrt{1 + \sigma_z^2}}\right), d &lt; 0 )</td>
<td>( \frac{1}{2}, d &lt; 0 )</td>
</tr>
<tr>
<td>Method 8:</td>
<td>For normal populations Method 8 yields the same estimator as Method 3.</td>
<td></td>
</tr>
<tr>
<td>Method 9:</td>
<td>No estimator derived by this method.</td>
<td></td>
</tr>
</tbody>
</table>
Table III
Expressions of Estimators for $P_2(R(\bar{x}_1, \bar{x}_2), \mu_2)$ when $\Pi_1$ is $N(\mu_1, 1)$ and $\Pi_2$ is $N(\mu_2, 1)$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Expression for Estimator</th>
<th>Name or Definition</th>
<th>D_1</th>
<th>S_1</th>
<th>Studied by^3</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_R</td>
<td>$p_0 \equiv$ proportion of $N_2$ original observations from $\Pi_2$ misclassified by $R(\bar{x}_1, \bar{x}_2)$</td>
<td>Reclassification estimator</td>
<td>No</td>
<td>0</td>
<td>L, H</td>
</tr>
<tr>
<td>P_T</td>
<td>$p_t \equiv$ proportion of $M_2$ test observations from $\Pi_2$ misclassified by $R(\bar{x}_1, \bar{x}_2)$</td>
<td>Test sample estimator</td>
<td>No</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>P_TSL</td>
<td>$p_0(N) - \frac{1}{2}p_0(\frac{1}{2}N) + \frac{3}{2}p_t(\frac{1}{2}N)$</td>
<td>Two straight lines estimator</td>
<td>No</td>
<td>0</td>
<td>P (as estimator of $P^*_2$)</td>
</tr>
<tr>
<td>P_U</td>
<td>$\frac{1}{N_2} \sum_{i=1}^{N_2} q_i$</td>
<td>Lachenbruch's U method</td>
<td>No</td>
<td>0</td>
<td>L, H, L_0 (as estimator of $P^*_2$)</td>
</tr>
<tr>
<td>P_Ra</td>
<td>$p_0 + F(b) - F\left(\frac{1}{2}d\right)$ if $d &gt; 0$</td>
<td>Reclassification estimator adjusted for bias</td>
<td>Yes-</td>
<td>0</td>
<td>H, C, E</td>
</tr>
<tr>
<td></td>
<td>$p_0 + F\left(\frac{1}{2}d\right) - F(b)$ if $d &lt; 0$</td>
<td></td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q_D</td>
<td>$1 - F\left(\frac{1}{2}d\right)$ if $d &gt; 0$</td>
<td></td>
<td>Yes</td>
<td>0</td>
<td>L, H</td>
</tr>
<tr>
<td></td>
<td>$F\left(\frac{1}{2}d\right)$ if $d &lt; 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table III (cont.)

<table>
<thead>
<tr>
<th>$Q_T$</th>
<th>$1 - F(\frac{d}{2} + \overline{x}_2 - \overline{t}_2) \ , \ d &gt; 0$</th>
<th>$P_2(R(\overline{x}_1, \overline{x}_2), \hat{\mu}_2 = \overline{t}_2)$</th>
<th>Yes</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F(\frac{d}{2} + \overline{x}_2 - \overline{t}_2) \ , \ d &lt; 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$Q^*_D$

$F(\alpha, -\beta; \rho) + F(-\alpha, \beta; \rho)$

where $\alpha = dM^{-\frac{1}{2}}, \beta = \frac{1}{2}d(1 + \frac{1}{4}M)^{-\frac{1}{2}},$

$M = N_1^{-1} + N_2^{-1}$

$\rho = \frac{1}{2}(-N_1^{-1} + N_2^{-1})M^{-\frac{1}{2}}(1 + \frac{1}{4}M)^{-\frac{1}{2}}$

Note: $\alpha, \beta$ are temporary notation

$Q^{**}_{DS}$

$1 - F(\frac{N_1 + N_2 - 4}{N_1 + N_2 - 2} |d|)$

Lachenbruch's DS estimator or $P^*_2(\delta)$

$\hat{\delta} = |d|(\frac{N_1 + N_2 - 4}{N_1 + N_2 - 2})^{\frac{1}{2}}$

<table>
<thead>
<tr>
<th>$P_{RB}$</th>
<th>$1 - F(\frac{\frac{1}{2}(\overline{x}_1 + \overline{x}_2) - \overline{t}_2}{(1 - M_2^{-1})^{\frac{1}{2}}} \ ) \ , \ d &gt; 0$</th>
<th>Conditionally UMVU estimator obtained by conditional Rao-Blackwell theory</th>
<th>Yes</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F(\frac{\frac{1}{2}(\overline{x}_1 + \overline{x}_2) - \overline{t}_2}{(1 - M_2^{-1})^{\frac{1}{2}}} \ ) \ , \ d &lt; 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table III (cont.)

<table>
<thead>
<tr>
<th>( \xi )</th>
<th>( 1 - F(\frac{\frac{1}{2}d}{(1 + N_2^{-1})^{\frac{1}{2}}} ), d &gt; 0 )</th>
<th>Method 3 prior distribution estimator.</th>
<th>Yes</th>
<th>0</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \xi )</td>
<td>( F(\frac{\frac{1}{2}d}{(1 + N_2^{-1})^{\frac{1}{2}}} ), d &lt; 0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( 1 - F(\frac{\frac{1}{2}d}{\sqrt{1 + \frac{1}{2N_2}}} \), d > 0 \) Method 5 prior distribution estimator. Yes 0

\( F(\frac{\frac{1}{2}d}{\sqrt{1 + \frac{1}{2N_2}}} \), d < 0 \)

---

1 Yes: Estimator uses distribution assumptions; No: Estimator doesn't use distribution assumptions.

2 O: Estimator uses only original samples; T: Estimator requires independent test samples.

3 Lachenbruch (1965) and Lachenbruch and Mickey (1968); all work for multivariate normal distributions.

Cochran and Hopkins (1961).

Elashoff (undated).

G: Geisser (1967).


Table IV

Conditional Means and Variances for Estimators of $P_2$

when $\pi_1$ is $N(\mu_1, 1)$ and $\pi_2$ is $N(\mu_2, 1)$

$$P_2 = P_2(R(\bar{x}_1, \bar{x}_2), \mu_2) = \frac{1-F(c)}{F(c)}, \quad d < 0$$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Conditional Expectation</th>
<th>Conditional Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_R$</td>
<td>$1 - F(b), \ d &gt; 0$</td>
<td>$-\frac{1}{N^2}[1-F(b)] + \frac{N^2-1}{N^2} F(-b,-b; -\frac{1}{N^2-1}), \ d &gt; 0$</td>
</tr>
<tr>
<td></td>
<td>$F(b), \ d &lt; 0$</td>
<td>$-F^2(b) + \frac{1}{N^2} F(b)$ + $\frac{N^2-1}{N^2} F(b;b; -\frac{1}{N^2-1}), \ d &lt; 0$</td>
</tr>
<tr>
<td>$P_T$</td>
<td>$P_2 = \begin{cases} 1 - F(c), &amp; d &gt; 0 \ F(c), &amp; d &lt; 0 \end{cases}$</td>
<td>$\frac{1}{N^2} F(c)[1 - F(c)]$</td>
</tr>
<tr>
<td>$P_{TSL}$</td>
<td>$F(-</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>$F(-</td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>$-\frac{1}{2}F(-\alpha,\gamma; \rho_1)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$+\frac{1}{2}F(\alpha,-\beta; \rho_2)$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$+\frac{1}{2}F(-\alpha,-\beta; \rho_2)$,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha = \frac{1}{2}d\sqrt{2N}, \quad \beta = \frac{1}{2}d(1 + \frac{1}{2N})^{-\frac{1}{2}}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\gamma = \frac{1}{2}d(1 - \frac{3}{2N})^{-\frac{1}{2}}$,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_1 = -(2N-3)^{-\frac{1}{2}}, \quad \rho_2 = (2N+1)^{-\frac{1}{2}}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Assumption Set 1)</td>
<td></td>
</tr>
</tbody>
</table>
| \( P_U \) | \( F(-\alpha) + F(-\beta), \ d > 0 \)  
  | \( F(\alpha) + F(\beta), \ d < 0 \)  
where \( \alpha = \frac{d}{2A}, \ \beta = \frac{d}{B} \)  
and  
\[
A = \frac{(N_2^{\frac{1}{2}})}{N_2(N_2-1)}, \quad B = N_2^{\frac{1}{2}}(N_2-1)^{-\frac{1}{2}}
\]  
| \( F(-|\beta|) + F(-|\alpha|) \)  
  | \( \frac{-F^2(-|\beta|) - F^2(-|\alpha|)}{2F(-|\beta|)F(-|\alpha|)} \)  
  | \( \frac{-F^2(-|\beta|)}{2F(-|\beta|)F(-|\alpha|)} \)  
  | \( \frac{-2F(-|\beta|)F(-|\alpha|)}{2F(-|\beta|)F(-|\alpha|)} \)  
  | \( \frac{-2F(-|\beta|)F(-|\alpha|)}{2F(-|\beta|)F(-|\alpha|)} \)  
  | \( \frac{-F^2(-|\alpha|)}{2F(-|\beta|)F(-|\alpha|)} \)  
  | \( \frac{-F^2(-|\alpha|)}{2F(-|\beta|)F(-|\alpha|)} \)  
where \( \rho = -(N_2-1)^{-1} \). |

| \( P_{Ra} \) | \( 1 - F(\frac{1}{2}d), \ d > 0 \)  
  | \( F(\frac{1}{2}d), \ d < 0 \)  
| Same as for \( P_U \). |

| \( Q_D \) | \( 1 - F(\frac{1}{2}d), \ d > 0 \)  
  | \( F(\frac{1}{2}d), \ d < 0 \)  
| Variance = 0 |

| \( Q_T \) | \( 1 - F(\frac{c}{\sqrt{1+M_2^{-1}}}), \ d > 0 \)  
  | \( F(\frac{c}{\sqrt{1+M_2^{-1}}}, \frac{c}{\sqrt{1+M_2^{-1}}}; \frac{1}{M_2+1}) \)  
  | \( F(\frac{c}{\sqrt{1+M_2^{-1}}}, \frac{c}{\sqrt{1+M_2^{-1}}}; 0) \)  
| Variance = 0 |

| \( Q_D^* \) | \( F(\alpha, -\beta; \rho) + F(-\alpha, \beta; \rho) \)  
where  
\[
\alpha = dM^{-\frac{1}{2}}, \ \beta = \frac{d}{2}(1 + \frac{1}{4M})^{-\frac{1}{2}} 
\]  
\[
M = N_1^{-1} + N_2^{-1} 
\]  
\[
\rho = \frac{1}{2}(-N_1^{-1} + N_2^{-1})M^{-\frac{1}{2}}(1 + \frac{1}{4M})^{-\frac{1}{2}} 
\]  
| Variance = 0 |
Table IV (cont.)

| $Q^{**}_{DS}$ | $1 - F\left(\frac{1}{2}\left(\frac{N_1 + N_2 - 4}{N_1 + N_2 - 2}\right)^{\frac{1}{2}}|d|\right)$ | Variance = 0 |
|---------------|-----------------------------------------------|--------------|
| $\frac{1}{N_1+N_2}$ | $1 - F\left(\frac{1}{2}|d|\left(1 + \frac{1}{N-2}\right)^{-\frac{1}{2}}\right)$ |              |

| $P_{RB}$ | $P_2 = \begin{cases} 1 - F(c), & d > 0 \\ F(c), & d < 0 \end{cases}$ | $F(c, c; \frac{1}{M_2}) - F(c, c; 0)$ |

| $P_3$ | $1 - F\left(\frac{\frac{1}{2}d}{\sqrt{1+N_2^{-1}}}ight)$, $d > 0$ | Variance = 0 |
|       | $F\left(\frac{\frac{1}{2}d}{\sqrt{1+N_2^{-1}}}ight)$, $d < 0$ |              |

| $P_5$ | $1 - F\left(\frac{\frac{1}{2}d}{\sqrt{1+\frac{1}{2N_2}}}ight)$, $d > 0$ | Variance = 0 |
|       | $F\left(\frac{\frac{1}{2}d}{\sqrt{1+\frac{1}{2N_2}}}ight)$, $d < 0$ |              |
Table V

Expressions and Expansions Needed to Get
the Asymptotic Expressions for Conditional
Bias, Variance, and Mean Square Error
Given in Tables VI, VII, and VIII

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Expansion: Term of order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Constant</td>
</tr>
<tr>
<td>$({1-N^{-1}})^{-\frac{1}{2}}$</td>
<td>1</td>
</tr>
<tr>
<td>$({1+N^{-1}})^{-\frac{1}{2}}$</td>
<td>1</td>
</tr>
<tr>
<td>$(1+\frac{1}{2}N^{-1})^{-\frac{1}{2}}$</td>
<td>1</td>
</tr>
<tr>
<td>$(1-\frac{1}{N-2})^{-\frac{1}{2}}$</td>
<td>1</td>
</tr>
<tr>
<td>$(1-N^{-1})^\frac{1}{2}(1-\frac{1}{2}N^{-1})^{-1}$</td>
<td>1</td>
</tr>
<tr>
<td>$(1-\frac{3}{2}N^{-1})^{-\frac{1}{2}}$</td>
<td>1</td>
</tr>
<tr>
<td>$\frac{1}{2}d$</td>
<td>$\frac{1}{2}d$</td>
</tr>
<tr>
<td>$b$</td>
<td>$\frac{1}{2}b$</td>
</tr>
<tr>
<td>$b^*$</td>
<td>$\frac{1}{2}b$</td>
</tr>
<tr>
<td>$c$</td>
<td>$\frac{1}{2}c$</td>
</tr>
<tr>
<td>$c^*$</td>
<td>$\frac{1}{2}c$</td>
</tr>
<tr>
<td>$c(1+N^{-1})^{-\frac{1}{2}}$</td>
<td>$\frac{1}{2}c$</td>
</tr>
<tr>
<td>$\frac{1}{2}d(1+\frac{1}{2}N^{-1})^{-\frac{1}{2}}$</td>
<td>$\frac{1}{2}d$</td>
</tr>
<tr>
<td>$\frac{1}{2}d(1+\frac{1}{N-2})^{-\frac{1}{2}}$</td>
<td>$\frac{1}{2}d$</td>
</tr>
<tr>
<td>$\frac{1}{2}d(1+N^{-1})^{-\frac{1}{2}}$</td>
<td>$\frac{1}{2}d$</td>
</tr>
<tr>
<td>$\frac{1}{2}d(1-N^{-1})^\frac{1}{2}(1-\frac{1}{2}N^{-1})^{-1}$</td>
<td>$\frac{1}{2}d$</td>
</tr>
<tr>
<td>$\frac{1}{2}d(1-\frac{3}{2}N^{-1})^{-\frac{1}{2}}$</td>
<td>$\frac{1}{2}d$</td>
</tr>
</tbody>
</table>
| \( F(\frac{1}{2}d) \) | \( G + \frac{1}{2}g \) | \(- (16)^{-1} \delta_0 u^2 \) | +++
| \( F(b) \) | \( G + \frac{1}{2}g \) | \(- (16)^{-1} \delta_0 u^2 \) | +++
| \( F(c) \) | \( G + \frac{1}{2}g \) | \(- (16)^{-1} \delta_0 v^2 \) | +++
| \( F(b^*) \) | \( G + \frac{1}{2}g \) | \(- (16)^{-1} \delta_0 (u^*)^2 \) | +++
| \( F(c^*) \) | \( G + \frac{1}{2}g \) | \(- (16)^{-1} \delta_0 (v^*)^2 \) | +++
| \( F(c(1+N^{-1})^{-\frac{1}{2}}) \) | \( G + \frac{1}{2}g \) | \(- (16)^{-1} \delta_0 u^2 \) | +++
| \( F(\frac{1}{2}d(1+N^{-1})^{-\frac{3}{2}}) \) | \( G + \frac{1}{2}g \) | \(- (16)^{-1} \delta_0 u^2 \) | +++
| \( F(\frac{1}{2}d(1-N^{-1})^{-\frac{1}{2}}) \) | \( G + \frac{1}{2}g \) | \(- (16)^{-1} \delta_0 u^2 \) | +++
| \( F(\frac{1}{2}d(1-N^{-1})^{-\frac{3}{2}}) \) | \( G + \frac{1}{2}g \) | \(- (16)^{-1} \delta_0 u^2 \) | +++
| \( F(-\frac{1}{2}d\sqrt{2N}) \) | \( (d>0) \) | \+++
| \( F(-d\sqrt{N(N-1)}) \) | \( (d<0) \) | \+++
| \( F(b,b;-(N-1)^{-1}) \) | \( F(b,b;0) \) | \(- g^2 N^{-1} \) | +++
| \( H(b,b;-(N-1)^{-1}) \) | \( H(b,b;0) \) | \(- g^2 N^{-1} \) | +++

---

| 59 |
Table V (cont.)

<table>
<thead>
<tr>
<th>Expression</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H(b,b;0)$</td>
<td>$H(\frac{1}{2}\delta,\frac{1}{2}\delta;0) - g(1-G)u$</td>
</tr>
<tr>
<td>$F(c',c';\frac{1}{N+1})$</td>
<td>$g^2N^{-1} - \frac{1}{2}\delta g^2vN^{-1}$</td>
</tr>
<tr>
<td>$F(c,c;N^{-1}) - F(c,c;0)$</td>
<td>$g^2N^{-1} - \frac{1}{2}\delta g^2vN^{-1}$</td>
</tr>
<tr>
<td>$H(a,a;\frac{1}{N-1})$</td>
<td>$- g^2N^{-1} + \frac{1}{2}\delta g^2uN^{-1}$</td>
</tr>
<tr>
<td>$H(a,a;0)$</td>
<td>$- g^2N^{-1} + \frac{1}{2}\delta g^2uN^{-1}$</td>
</tr>
</tbody>
</table>

where $c'$

$= c(1+N^{-1})^{-\frac{1}{2}}$

$F(a,a;\frac{1}{N-1}) - F(a,a;0)$

$F(-a,-a;\frac{1}{N-1}) = (d>0)$

$\delta = \sqrt{d(N-1)}$

$F(-a,-\delta;\frac{1}{N-1}) = (d>0)$

$F(-\frac{1}{2d}\sqrt{2N}, K; \rho)$

$F(\frac{1}{2d}\sqrt{2N}, -\frac{1}{2d}(1+\frac{1}{2N})^{-\frac{1}{2}}; \rho_1)$

$= (d>0)$

$\frac{1}{2}\delta gN^{-1}$

$= (d>0)$

$\rho_1 = (2N+1)^{-\frac{1}{2}}$

$\rho_2 = -(2N+3)^{-\frac{1}{2}}$

$\delta = \mu_1 - \mu_2$, $G = F(\frac{1}{2}\delta)$, $g = f(\frac{1}{2}\delta)$, $u = \alpha_1 - \alpha_2$, $v = \alpha_1 + \alpha_2$, $u^* = \alpha^*_1 - \alpha^*_2$, $v^* = (\alpha^*_1 + \alpha^*_2)$. Other notation is defined in Table I.

-60-
Table VI

Squared Biases for Estimators of $P_2(R(x_1, x_2), \mu_2)$ when $\Pi_1$ is $N(\mu_1, 1)$ and $\Pi_2$ is $N(\mu_2, 2)$:

Exact and Asymptotic Expressions.

$$P_2(R(x_1, x_2), \mu_2) = \begin{cases} 1 - F(c), & d > 0 \\ \frac{1}{F(c)}, & d < 0 \end{cases}$$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Squared Bias</th>
<th>Asymptotic Squared Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_R$</td>
<td>$[F(b) - F(c)]^2$</td>
<td>$o_p(N^{-1+e})$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$o_p(N^{-3/2+e})$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$o_p(N^{-2+e})$</td>
</tr>
<tr>
<td>$P_T$</td>
<td>Zero</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{TSL}$</td>
<td>$[F(b) - F(c)]^2 + \frac{1}{4}[F(b^<em>) - F(c^</em>)]^2$</td>
<td>$g^2\alpha_2^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$-\frac{1}{2}g^2\alpha_2^2(\alpha^2 + N^{-1})$</td>
</tr>
<tr>
<td>Assumption Set 2</td>
<td>$\Delta_{dd^<em>}[F(b) - F(c)][F(b^</em>) - F(c^*)]$</td>
<td>$+$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{TSL}$</td>
<td>$[F(b) - F(c) + \frac{1}{2}F(\alpha, -\gamma; \rho_1) + \frac{1}{2}F(-\alpha, \gamma; \rho_1)]^2$</td>
<td>$g^2\alpha_2^2$</td>
</tr>
<tr>
<td>Assumption Set 1</td>
<td>$-\frac{1}{2}F(\alpha, -\beta; \rho_2) - \frac{1}{2}F(-\alpha, \beta; \rho_2)$. See $P_{TSL}$ on Table IV for $\alpha, \beta, \gamma, \rho_1, \rho_2$ notation.</td>
<td>$-\frac{1}{2}g^2\alpha_2^2(\alpha^2 + N^{-1})$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$+$</td>
</tr>
</tbody>
</table>
### Table VI (cont.)

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Expression</th>
<th>$o_p(N^{-1+\varepsilon})$</th>
<th>$o_p(N^{-3/2+\varepsilon})$</th>
<th>$o_p(N^{-2+\varepsilon})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_U$</td>
<td>$[1 - F(c) - F(\frac{d}{2A}) - (1 - F(\frac{d}{B}))]^2$, $d &gt; 0$</td>
<td>$g^2\alpha_2^2$</td>
<td>$-\frac{1}{2}g^2\alpha_1\alpha_2^2$</td>
<td>$+++            $</td>
</tr>
<tr>
<td></td>
<td>$[F(c) - F(\frac{d}{B}) - F(\frac{1}{A})]^2$, $d &lt; 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$A = \frac{N - \frac{1}{2}}{\sqrt{N(N-1)}} = \frac{(1-\frac{1}{2N})}{\sqrt{1-\frac{1}{N}}} = B = \frac{1}{\sqrt{N(N-1)}}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{Ra}$</td>
<td>$[F(\frac{1}{2}d) - F(c)]^2$</td>
<td>$g^2\alpha_2^2$</td>
<td>$-\frac{1}{2}g^2\alpha_1\alpha_2^2$</td>
<td>$+++$</td>
</tr>
<tr>
<td>$Q_D$</td>
<td>$[F(\frac{1}{2}d) - F(c)]^2$</td>
<td>$g^2\alpha_2^2$</td>
<td>$-\frac{1}{2}g^2\alpha_1\alpha_2^2$</td>
<td>$+++$</td>
</tr>
<tr>
<td>$Q_T$</td>
<td>$[F(\frac{c}{\sqrt{1 + N^{-1}}}) - F(c)]^2$</td>
<td>$0$</td>
<td>$0$</td>
<td>$+++$</td>
</tr>
<tr>
<td>$Q_D^*$</td>
<td>$[\alpha - 1 + F(c)]^2$, $d &gt; 0$</td>
<td>$g^2\alpha_2^2$</td>
<td>$-\frac{1}{2}g^2\alpha_2(\alpha_1\alpha_2^{-1}N^{-1})$</td>
<td>$+++           $</td>
</tr>
<tr>
<td></td>
<td>$[\alpha - F(c)]^2$, $d &lt; 0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>where</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\alpha = F(\frac{1}{2}(2N)^{1/2})F(-\frac{1}{2}(1 + \frac{1}{2N})^{-1/2})$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$+ F((-\frac{1}{2}(2N)^{1/2})F(\frac{1}{2}(1 + \frac{1}{2N})^{-3/2})$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table VI (cont.)

<table>
<thead>
<tr>
<th>( Q^{**}_{DS} )</th>
<th>( \left[ F\left(\frac{\frac{3d}{2}d\left(1 + \frac{1}{N - 2}\right)^{-\frac{1}{2}}\right) - F(c)\right]^2 )</th>
<th>( o_p(N^{-1+\varepsilon}) )</th>
<th>( o_p(N^{-3/2 +\varepsilon}) )</th>
<th>( o_p(N^{-2+\varepsilon}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{RB} )</td>
<td>Zero</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_3 )</td>
<td>( \left[ F\left(\frac{\frac{3d}{2}d}{\sqrt{1 + N^{-1}}}\right) - F(c)\right]^2 )</td>
<td>( g^2\alpha_2^2 )</td>
<td>( -\frac{1}{2}\delta g^2\alpha_2^2(\alpha_1\alpha_2 - N^{-1}) )</td>
<td>***</td>
</tr>
<tr>
<td>( P_5 )</td>
<td>( \left[ F\left(\frac{\frac{3d}{2}d}{\sqrt{1 + \frac{1}{2N}}}\right) - F(c)\right]^2 )</td>
<td>( g^2\alpha_2^2 )</td>
<td>( -\frac{1}{2}\delta g^2\alpha_2^2(\alpha_1\alpha_2 - 2N^{-1}) )</td>
<td>***</td>
</tr>
</tbody>
</table>

+++ indicates a term of the order given in the column heading.

\( \alpha_1 = \bar{x}_1 - \mu_1 \), \( g = f(\frac{1}{2\delta}) \). See Table I for other notation not defined here.

The exact bias expressions are given for the case \( N_1 = N_2 = N \).
Table VII

Conditional Variances for Estimators of $P_2(R(x_1, x_2), \mu_2)$
when $\pi_1$ is $N(\mu_1, 1)$ and $\pi_2$ is $N(\mu_2, 1)$:

Asymptotic Expressions

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Variance: Term of order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$o_p(N^{-1+\epsilon})$</td>
</tr>
<tr>
<td>$P_R$</td>
<td>$[G(1-G)-g^2]N^{-1}$</td>
</tr>
<tr>
<td>$P_T$</td>
<td>$G(1-G)N^{-1}$</td>
</tr>
<tr>
<td>$P_{TS}$</td>
<td>$\frac{1}{2}[G(1-G)-3g^2]N^{-1}$</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_U$</td>
<td>$[G(1-G)-g^2]N^{-1}$</td>
</tr>
<tr>
<td>$P_{R^a}$</td>
<td>Same as $\text{Var} P_R$</td>
</tr>
<tr>
<td>$Q_D$</td>
<td>Variance = 0</td>
</tr>
<tr>
<td>$Q_T$</td>
<td>$g^2N^{-1}$</td>
</tr>
<tr>
<td>$Q^*$</td>
<td>Variance = 0</td>
</tr>
<tr>
<td>$Q_{DS}$</td>
<td>Variance = 0</td>
</tr>
<tr>
<td>$P_{RB}$</td>
<td>$g^2N^{-1}$</td>
</tr>
<tr>
<td>$P_3$</td>
<td>Variance = 0</td>
</tr>
<tr>
<td>$P_5$</td>
<td>Variance = 0</td>
</tr>
</tbody>
</table>

+++ indicates a term of the order given in the column heading.

$G = G(\frac{1}{2}G)$, $g = g(\frac{1}{2}G)$, $u = \alpha_1 - \alpha_2$, $u^* = \alpha_1^* - \alpha_2^*$, $v = \alpha_1 + \alpha_2$, $v^* = \alpha_1^* + \alpha_2^*$.
Table VIII

Conditional Asymptotic Mean Square Errors for

Estimators of \( p_2(R(x_1, x_2), \mu_2) \) when \( \Pi_1 \) is

\( N(\mu_1, 1) \) and \( \Pi_2 \) is \( N(\mu_2, 1) \)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>MSE: Term of order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \sigma_p(N^{-1+\varepsilon}) )</td>
</tr>
<tr>
<td>( P_R )</td>
<td>( g^2\alpha^2 + [G(1-G) - g^2]N^{-1} )</td>
</tr>
<tr>
<td>( P_T )</td>
<td>( G(1-G)N^{-1} )</td>
</tr>
<tr>
<td>( P_{TSL} )</td>
<td>( g^2\alpha^2 + \frac{1}{4}g^2(\alpha_2^*)^2 )</td>
</tr>
<tr>
<td>(Assumption</td>
<td>( + \frac{1}{2}[4G(1-G) - g^2]N^{-1} )</td>
</tr>
<tr>
<td>Set 2)</td>
<td></td>
</tr>
<tr>
<td>( P_U )</td>
<td>( g^2\alpha^2 + [G(1-G) - g^2]N^{-1} )</td>
</tr>
<tr>
<td>( P_{Ra} )</td>
<td>( g^2\alpha^2 + [G(1-G) - g^2]N^{-1} )</td>
</tr>
<tr>
<td>( Q_D )</td>
<td>( g^2\alpha^2 )</td>
</tr>
<tr>
<td>( Q_T )</td>
<td>( g^2\alpha^2 )</td>
</tr>
<tr>
<td>( Q_D^* )</td>
<td>( g^2\alpha^2 )</td>
</tr>
<tr>
<td>( Q_{DS} )</td>
<td>( g^2\alpha^2 )</td>
</tr>
<tr>
<td>( P_{RB} )</td>
<td>( g^2\alpha^2 )</td>
</tr>
</tbody>
</table>
Table VIII (cont.)

<table>
<thead>
<tr>
<th>$P_3$</th>
<th>$g^2\alpha_2^2$</th>
<th>$-\frac{1}{2}\delta g^2\alpha_2(\alpha_1\alpha_2 - N^{-1})$</th>
<th>+++</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_5$</td>
<td>$g^2\alpha_2^2$</td>
<td>$-\frac{1}{2}\delta g^2\alpha_2(\alpha_1\alpha_2 - \frac{1}{2}N^{-1})$</td>
<td>+++</td>
</tr>
</tbody>
</table>

+++ indicates a term of the order given in column heading.

$G = F(\frac{1}{2}\delta), g = f(\frac{1}{2}\delta), u = \alpha_1 - \alpha_2, v = \alpha_1 + \alpha_2.$

$\Delta_{dd^*} = 1$ if $d\cdot d^* > 0$, $-1$ if $d\cdot d^* < 0$.

See Table I for other notation.
Table IX

Numerical Values for Terms of Order $o_p(N^{-1}+\epsilon)$ in the AMSE's of Table VIII, for Selected Values of $\delta$ and $\sqrt{N}\alpha_2$.

| Estimator | Value of $\sqrt{N}\alpha_2$ | $|\delta| = 0$ | $|\delta| = 1$ | $|\delta| = 2$ | $|\delta| = 3$ |
|-----------|---------------------------|----------------|----------------|----------------|----------------|
| $P_T$, $P_U$, $P_{Ra}$ | 0 | 0.091N$^{-1}$ | 0.089N$^{-1}$ | 0.075N$^{-1}$ | 0.045N$^{-1}$ |
| $Q_D$, $Q_D^*$, $Q_{DS}^*$, $\rho_3$, $\rho_5$ | | | | | |
| $Q_T$, $\rho_{RB}$ | 0.159 | 0.124 | 0.059 | 0.017 |
| $P_T$ | 1 | 0.250N$^{-1}$ | 0.213N$^{-1}$ | 0.134N$^{-1}$ | 0.062N$^{-1}$ |
| $Q_D$, $Q_D^*$, $Q_{DS}^*$, $\rho_3$, $\rho_5$ | | | | | |
| $Q_T$, $\rho_{RB}$ | 0.159 | 0.124 | 0.059 | 0.017 |
| $P_T$ | 2 | 0.727N$^{-1}$ | 0.585N$^{-1}$ | 0.309N$^{-1}$ | 0.112N$^{-1}$ |
| $Q_D$, $Q_D^*$, $Q_{DS}^*$, $\rho_3$, $\rho_5$ | | | | | |
| $Q_T$, $\rho_{RB}$ | | | | | |
| $P_T$ | 3 | 1.523N$^{-1}$ | 1.205N$^{-1}$ | 0.602N$^{-1}$ | 0.196N$^{-1}$ |
| $Q_D$, $Q_D^*$, $Q_{DS}^*$, $\rho_3$, $\rho_5$ | | | | | |
| $Q_T$, $\rho_{RB}$ | 0.159 | 0.124 | 0.059 | 0.017 |

- 67 -
CHAPTER 4
ESTIMATING THE CONDITIONAL PROBABILITY
OF MISCLASSIFICATION IN THE
MULTIVARIATE NORMAL CASE

In this chapter, estimators for $P_2(R(S_1, S_2), \theta)$ are considered in the multivariate normal case with unknown means $\mu_1, \mu_2$ and common known covariance matrix $\Sigma$. As in the univariate normal case, when $\Sigma$ is known, the variables can be transformed to reduce the problem to the simpler case $\Sigma = I$. Here, however, the variables will be kept in the original form and the $\Sigma$ will appear in the formulas.

1. The Classification Rule and the Various Probabilities of Misclassification.

Let $\pi_1$ be p-dimensional $N(\mu_1, \Sigma)$ and $\pi_2$ be p-dimensional $N(\mu_2, \Sigma)$ with $\mu_1, \mu_2$ unknown and $\Sigma$ known. Thus $\theta = (\mu_1, \mu_2)$. As before, original samples of sizes $N_1, N_2$ and independent test samples of sizes $M_1, M_2$ are available from $\pi_1, \pi_2; \omega$, which is assumed to be from $\pi_2$, is a new random observation to be classified.

Some notation used in discussing the classification rule and the estimators for $P_2$ is given in Table X. Observe that this notation is related to the notation used in the univariate normal discussion, which is given in Table I. The form which some of the symbols of Table X take when $p = 1$ and $\sigma^2 = 1$, i.e., the situation of Chapter 3, is given at the bottom of the table. Furthermore $D^*, B^*, C^*$ are to $d^*, b^*, c^*$ as $D, B, C$ are to $d, b, c$. The symbols $F(u), F(u, v; \rho), H(u, v; \rho)$ are the same as in Table I. Symbols such as $\overline{x}_j, \alpha_j$, for example, are the same except that in this chapter they represent p-dimensional vectors and in Chapter 3, scalars.

The classification rule to be used is that of (1.1), Chapter 1, with $c = 1$ and with the unknown means estimated by the sample means.
This rule reduces to

\[
\begin{align*}
\text{classify } \mathbf{x} \text{ as } \pi_1 & \text{ if } \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)'\Sigma^{-1}(\mathbf{x}_1 + \mathbf{x}_2) \\
\text{classify } \mathbf{x} \text{ as } \pi_2 & \text{ if } \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)'\Sigma^{-1}(\mathbf{x}_1 + \mathbf{x}_2) < 0
\end{align*}
\]

(1.1)

As in the univariate case, fixing the rule is not equivalent to fixing \( \bar{x}_1, \bar{x}_2 \); nevertheless we will condition on \( \bar{x}_1, \bar{x}_2 \). (See discussion in Chapter 3, Section 1.) Again this distinction is irrelevant for the derivation of the expression for \( P_2(R(S_1, S_2), \theta) \), the conditional probability of misclassifying \( \mathbf{x} \) from \( \pi_2 \) as \( \pi_1 \), but makes a difference in the conditional moments.

To compute \( P_2(R(S_1, S_2), \theta) \) let \( y \) denote \( (\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}\mathbf{x} \). Then

\[
E(y|\bar{x}_1, \bar{x}_2) = (\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}\bar{x}_2,
\]

\[
\text{Var}(y|\bar{x}_1, \bar{x}_2) = (\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}\Sigma^{-1}(\bar{x}_1 - \bar{x}_2) = D^2,
\]

and

\[
P_2(R(S_1, S_2), \theta) = P\{y > \frac{1}{2}(\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}(\bar{x}_1 + \bar{x}_2) | \mathbf{x} \text{ from } \pi_2; \bar{x}_1, \bar{x}_2 \text{ fixed}\}
\]

\[
= P\{z > \frac{1}{2}D + D^{-1}(\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}\bar{x}_2\}
\]

\[
= 1 - F(C),
\]

where \( z \) is a univariate standard normal variate and \( C \) is defined in Table X.

Since fixing \( \bar{x}_1, \bar{x}_2 \) fixes \( R(S_1, S_2) \) and since \( \mu_2 \) is the only parameter that enters into the expression for the conditional probability of misclassification, we will denote \( P_2(R(S_1, S_2), \theta) \) by \( P_2(R(\bar{x}_1, \bar{x}_2), \mu_2) \), which will as usual be abbreviated to \( P_2 \) when the dependence on \( \bar{x}_1, \bar{x}_2, \mu_2 \) is not being emphasized.

This formula for the conditional probability of misclassification matches that given in John (1961).
The unconditional probability of misclassification has no simple expression when the variables are multivariate normal, even when \( \Sigma \) is known. John (1961, p. 1137ff) has considered this case and gives an exact and an approximate result. Okamoto (1963) gives an approximation to \( P^*_2 \) in the more general situation where \( \mu_1, \mu_2 \) and \( \Sigma \) are all unknown. In the special case that \( N_1 = N_2 = N \) and the number of degrees of freedom associated with the estimate of the covariance matrix is infinite, the approximation is

\[
(1.3) \quad P^*_2 = \frac{1}{N^2} \left[ \frac{1}{2} + \frac{1}{2} (p-1) N^{-1} \right] + o(N^{-2+\epsilon}).
\]

The true probability of misclassification when all parameters are known and the rule of (1.1), Chapter 1 is used, is denoted by \( P^{**}_2(\theta) \) and is computed as follows:

\[
(1.4) \quad P^{**}_2(\theta) = P \left[ \frac{1}{2} \langle \tilde{\mu} - \mu \rangle' \Sigma^{-1} \tilde{\mu} > \frac{1}{2} \langle \mu_1 - \mu_2 \rangle' \Sigma^{-1} (\mu_1 + \mu_2) \right] = 1 - F(\frac{1}{2} \Delta).
\]

2. The Estimators.

The estimators described in Section 2 of Chapter 2 were derived for the special case where populations are univariate normal in Section 2 of Chapter 3. For \( p \)-dimensional normal populations, these estimators are described below with the sub-section numbers kept the same as in Chapter 2, Section 2 and Chapter 3, Section 2. Some of the derivations are very similar to the univariate normal derivations and no details are given below. The sub-section numbers are maintained, despite the very short sub-sections, to facilitate cross-reference to Sections 2 of Chapters 2 and 3.

Table XI gives all of the estimators studied in the multivariate

\[\text{Okamoto gives explicitly the terms of order } o(N^{-2+\epsilon}).\]
normal case, including those described below and the estimators not based on distribution assumptions, which are described in Chapter 2, Section 1.

2.1 The reclassification estimator adjusted for bias, \( \widehat{P_{Ra}} \).

Using the results on expectations of \( p_0 \) and \( p_t \) from Appendix II, and estimating \( \mu_2 \) by \( \bar{x}_2 \) we get the expression for the adjusted reclassification estimator,

\[
\widehat{P_{Ra}} = p_0 \cdot \Phi(b) \cdot \Phi(\frac{1}{2}d).
\]

2.2 The expression for conditional probability of misclassification with \( \mu_2 \) estimated, \( Q_D \) and \( Q_T \).

For p-dimensional multivariate normal populations we have from

\[
P_2(R(\bar{x}_1, \bar{x}_2), \mu_2) = 1 - F(c), \quad \text{where} \quad c = \frac{1}{2}D + D^{-1}(\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}(\bar{x}_2 - \mu_2).
\]

Hence if \( \mu_2 \) is estimated by \( \bar{x}_2 \) we get the estimator

\[
P_2(R(\bar{x}_1, \bar{x}_2), \mu_2 = \bar{x}_2) = 1 - F(\frac{1}{2}d).
\]

If \( \mu_2 \) is estimated by \( \bar{c}_2 \), we get the estimator

\[
P_2(R(\bar{x}_1, \bar{x}_2), \mu_2 = \bar{c}_2) = 1 - F(\frac{1}{2}D + D^{-1}(\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}(\bar{x}_2 - \bar{c}_2)).
\]

In correspondence with the univariate case, these two estimators will be denoted \( Q_D \) and \( Q_T \) respectively.

2.3 The expression for the unconditional probability of misclassification with parameters replaced by sample values, \( Q_0^* \) and \( Q_0^* \).

As discussed in Section 1, there is no simple expression for \( P_2^*(R(\cdot, \cdot), \mu) \) when populations are multivariate normal. Lachenbruch (1965) and Lachenbruch and Mickey (1968) obtained two estimators for \( P_2 \) by estimating the parameter \( \Delta^2 \) in the expansion for \( P_2^* \) due to Okamoto (1963). The two estimators used for \( \Delta^2 \) were
\[(\bar{x}_1 - \bar{x}_2)'S^{-1}(\bar{x}_1 - \bar{x}_2) = \frac{N_1+N_2-p-3}{N_1+N_2-p-2} S^{-1}(\bar{x}_1 - \bar{x}_2)'S^{-1}(\bar{x}_1 - \bar{x}_2),\]

where \(S\) is the usual sample estimate of \(\Sigma\). The resulting estimators for \(p_2\) were called the O and OS estimators, respectively. We get two estimators of the same type, which we will call \(Q_0^*\) and \(Q_{OS}^*\), by using the Okamoto expansion for the special case of \(N_1 = N_2 = N\) and \(\Sigma\) known, as it is given in (1.3). Estimating \(\Delta^2\) by \(D^2\) (or \(\Delta\) by \(D\)) leads to the estimator

\[(2.4) \quad Q_0^* = F(-\frac{1}{2}D) + (\frac{1}{2}N)f(\frac{1}{2}D)[D + 4(p-1)D^{-1}]\]

which corresponds to Lachenbruch and Mickey's O estimator. Estimating \(\Delta^2\) by \(D')^2 = \frac{2N-p-3}{2N-2} D^2\) leads to the estimator corresponding to Lachenbruch and Mickey's OS estimator,

\[(2.5) \quad Q_{OS}^* = F(-\frac{1}{2}D') + (\frac{1}{2}N)f(\frac{1}{2}D')[D' + 4(p-1)/D']\].

2.4 The expansion for the true probability of misclassification when all parameters are known, with parameters replaced by sample values, \((Q_{DS}^{**})\).

From (1.4), \(P_{2}^{**}(\Delta = D) = 1 - F(\Delta D)\). If \(\Delta\) is estimated by \(D\), no new estimator is obtained, since \(P_2^{**}(\Delta = D) = 1 - F(\Delta D)\), which is equal to \(P_2(R(\bar{x}_1, \bar{x}_2), \hat{\mu}_2 = \bar{x}_2)\). Estimating \(\Delta\) by \(\frac{N_1+N_2-p-3}{N_1+N_2-p-2} \frac{1}{2}D\) does lead to a new estimator, which is Lachenbruch's DS estimator and which will be denoted by \(Q_{DS}^{**}\) here. Thus

\[(2.6) \quad Q_{DS}^{**} = 1 - F(\frac{1}{2}D(\frac{N_1+N_2-p-3}{N_1+N_2-p-2} \frac{1}{2})).\]

2.5 Minimum variance unbiased estimators, \((Q_{RB}^{*})\).

The conditional UMVU estimator given that the data consists of original samples from \(\Pi_1\) and \(\Pi_2\) plus a test sample of size \(M_2\) from \(\Pi_2\) (a test sample from \(\Pi_1\) can be included or not) is derived in Chapter 5, Section 4, using the conditional Rao-Blackwell theory.
developed there. When the distributions are p-dimensional normal, the estimator, which as in the univariate case will be denoted \( \mathcal{P}_{RB} \), is, according to (4.8), Chapter 5,

\[
\mathcal{P}_{RB} = 1 - F\left(\frac{3}{2}(1 - M_2^{-1})^{-\frac{3}{2}}D + (1 - M_2^{-1})^{-\frac{3}{2}}D^{-1}(\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}(\bar{x}_2 - \bar{x}_2)\right).
\]

2.6 Estimators involving a prior distribution on \( \theta \), \( \mathcal{P}_3 \) and \( \mathcal{P}_5 \).

No estimators were derived for multivariate normal populations by putting a prior distribution on \( \hat{\theta} = (\hat{\mu}_1, \hat{\mu}_2) \), but two estimators were invented by using the univariate results. For the univariate normal case, among the estimators derived by using a prior list on \( \hat{\theta} \), only two seemed sensible in the limit of "no prior information"— \( \mathcal{P}_3 \) and \( \mathcal{P}_5 \), where \( \mathcal{P}_3 = 1 - F\left(\frac{3}{2}|d|(1 + \frac{1}{N_2})^{-\frac{3}{2}}\right) \) and \( \mathcal{P}_5 = 1 - F\left(\frac{3}{2}|d|(1 + \frac{1}{2N_2})^{-\frac{3}{2}}\right) \).

Using \( D \) in place of \( |d| \) in these expressions we get two estimators for \( P_2 \) in the multivariate case,

\[
\begin{align*}
\mathcal{P}_3 &= 1 - F\left(\frac{3}{2}D(1 + N_2^{-1})^{-\frac{3}{2}}\right) \\
\mathcal{P}_5 &= 1 - F\left(\frac{3}{2}D(1 + \frac{1}{2N_2})^{-\frac{3}{2}}\right).
\end{align*}
\]

Note that both of these estimators can be considered as estimators of the type considered in Section 2.4, i.e., \( P_{2}^{**}(\theta) \) with an estimator substituted for \( \theta \). From (1.4), \( P_{2}^{**}(\theta) = 1 - F\left(\frac{3}{2}D\right) \) and so \( \mathcal{P}_3 \) is equivalent to \( P_{2}^{**}(\theta) \) with \( \Delta \) estimated by \( D(1 + N_2^{-1})^{-\frac{3}{2}} \) and \( \mathcal{P}_5 \) is equivalent to \( P_{2}^{**}(\theta) \) with \( \Delta \) estimated by \( D(1 + \frac{1}{2N_2})^{-\frac{3}{2}} \).


The conditional means and variances for the multivariate normal case are given in Table XII, and the computations are in Appendix II. The moments for \( P_{TSL} \) were computed only under Assumption Set 2 (see Chapter 3, Section 3). The estimator \( P_U \) was not studied in the multivariate case.
The results on means and variances parallel those for the univariate situation and the obvious conclusions are the same, namely,

(i) $P_T$ and $P_{RB}$ are the only unbiased estimators.

(ii) $P_{Ra}$ and $Q_D$ have the same expectation and hence the same bias.

(iii) The estimators $Q_D$, $Q_{DS}^*$, $Q_{DS}^{**}$, $P_3$, and $P_5$ have the smallest possible variance, zero. Note that the estimators based on the Okamoto expansion, $Q_0^*$ and $Q_{OS}^*$, are the counterpart to the estimator $Q_D^*$ of the univariate case.

(iv) $P_R$ and $P_{Ra}$ have the same variance.

(v) $\text{Var } P_{TSL} \geq \text{Var } P_R$ when $N_2 \geq N$. Note, however, that $\text{Var } P_{TSL}$ was computed under Assumption Set 2, rather than Assumption Set 1. (see Chapter 3, Section 3).

(vi) Although it is not obvious from the variance expressions, $\text{Var } P_{RB}$ must be smaller than the variance of $P_T$ because $P_{RB}$ is the minimum variance unbiased estimator among those based on test samples.

Note that the unbiased estimators require test observations, while all the estimators with zero variance are based on original samples. There are no estimators with zero MSE and furthermore, as in the univariate case, it is difficult to compare the sizes of the MSE's using the exact expressions. Hence asymptotic expressions will now be given.

4. **Asymptotic Expressions for MSE and Discussion of Results.**

Our asymptotic expressions for conditional squared bias, variance and mean square error can be derived by the same methods used in the univariate normal case (see Chapter 3, Section 4.1). The needed "a + t" expressions for the various symbols and the Taylor series expansions for the functions are given in Table XIII with $\mathcal{Y} = F(\frac{1}{2} \Delta)$, $\mathcal{G} = f(\frac{1}{2} \Delta)$ and other symbols as defined there. The computations are
in Appendix IV. Note that the "a" used in these expressions is $\frac{1}{2}A$, which is non-negative and reduces to $\frac{1}{2}|\delta|$ when $p = 1$ and $\sigma^2 = 1$, whereas in the expressions used for the univariate normal problem, the corresponding "a" was $\frac{1}{2}b$. Note also that the classification rule of (1.1), Chapter 3, rather than the form given in (1.2), Chapter 3, is the univariate form of the multivariate classification rule of (1.1). There are some superficial differences in appearance between the univariate results on MSE given in Chapter 3 and the multivariate results of Chapter 4 specialized to $p = 1$ which arise from using the rule of (1.2) (rather than (1.1)) Chapter 3 in getting the univariate results.

The asymptotic expressions for squared bias and for variances are in Tables XIV and XV. Given the exact formula for squared bias in Table XIV, the asymptotic expressions are easily obtained from the expansions given in Table XIII. The asymptotic variance expansions are also obtainable from the expressions in Table XIII. The asymptotic mean square error (AMSE) expressions are obtained by adding the expansions for squared bias and variance and are given in Table XVI.

The results on conditional asymptotic squared bias, variance, and MSE are essentially the same as for the univariate case, and the discussion of those results given in Section 4.2 of Chapter 3 applies here with the following adjustments.

(i) Actual values for leading terms in the MSE must be replaced by the multivariate values; for example, $g^2\alpha^2_2$ should be replaced by $\frac{1}{2}A^2\int z^2$ and $[k_1G(1-G) + k_2g^2]$ by $[k_1G(1-G) + \int z^2]$.

(ii) The discussion relating to terms in the AMSE other than the leading term should be ignored.

(iii) The two estimators based on the Okamoto expansion for $P^*_2$, $Q^*_0$ and $Q^*_0$, take the place of the univariate estimator $Q^*_D$. 
Since the results are parallel, for \( Q^* \) read \( Q_0^* \) and \( Q_{OS}^* \).

(iv) The remarks on \( P_U \) should be ignored, since this estimator was not studied in the multivariate normal case. Note that the omission of \( P_U \) has reduced the number of estimators studied from among the class of estimators based on original samples and not using normality.

(v) The moments for \( P_{TSL} \) were computed only under Assumption Set 2, so the remarks related to the expectation of \( P_{TSL} \) computed under Assumption Set 1 do not apply.

In the next section, the overall conclusions on the problem of estimating the conditional probability of misclassification in the p-dimensional normal case, \( p \geq 1 \) are discussed and are compared to other work on this problem.

5. Conclusions on the Problem of Estimating \( P_2(\bar{x}_1, \bar{x}_2, \mu_2) \) When the Populations are p-Dimensional Normal. Comparisons with Other Work on the Problem.

The major asymptotic results and the conclusions on the problem of estimating \( P_2(\bar{x}_1, \bar{x}_2, \mu_2) \) when the distributions are univariate normal were discussed in detail in Section 4 of Chapter 3. The asymptotic results for multivariate normal populations are discussed in Section 4 above. The overall conclusions for p-dimensional normal populations, \( p \geq 1 \), are summarized below in Section 5.1. Since, as pointed out in Section 4, the multivariate results are essentially the same as the univariate results, these conclusions differ only in detail from the conclusions already given in Section 4.3 of Chapter 3. They are restated here for completeness.

In Section 5.2, our results and conclusions are compared to other work on the problem, in particular that of Lachenbruch (1965) and Lachenbruch and Mickey (1968).
5.1 Conclusions.

Several estimators of $P_2(R(\bar{x}_1, \bar{x}_2), \mu_2)$ were compared on asymptotic conditional mean square error, with moments computed under normality. The major conclusions, which follow from the results presented in Tables VIII and XVI and discussed in Chapter 3, Section 4 and Chapter 4, Section 4, are listed below. The criterion of the comparisons is mainly the leading term in the AMSE.

1. The first conclusion is that, among the estimators studied, those making use of the normality are better than estimators not using distribution assumptions.

2. The second conclusion is that, among the estimators studied, those based only on original samples are as good, on the (unconditional) average, as those requiring independent test samples.

3. The third conclusion is that the several estimators studied that are based on normality and original samples are asymptotically equivalent. This equivalence suggests a theoretical result related to a lower bound on the conditional AMSE (and on the asymptotic conditional squared bias, since these estimators have conditional variance equal to zero). This conclusion and the problem of the bound are discussed in Chapter 3, Section 4 and in Appendix VI. Note that in the univariate results the four estimators $P_R, P_U, P_{TSL}$ (mean computed under Assumption Set 1) and $P_{Ra}$, all had the same leading term in the asymptotic squared bias as the group of estimators based on normality and original samples, suggesting that a bound on the bias holds for the whole class of estimators based on original samples. In the multivariate case there is less evidence for this conjecture because the expectations of the estimators $P_U$ and $P_{TSL}$ (Assumption Set 1) are not available.

Furthermore, because $P_U$ was not studied in the multivariate case and
no moments were obtained for $P_{TSL}$ under Assumption Set 1, the group of estimators based on original samples, but not using normality, has been reduced to one, $P_R$, or to two, if $P_{Ra}$ is put in this class. Thus the suggested conclusion (that one might as well use $P_R$ rather than a more complicated estimator such as $P_U$ or $P_{TSL}$) based on the equivalence of the group of estimators using original samples but not using normality, which is discussed as part of the third conclusion of Chapter 3, Section 4.3, is not applicable.

5.2 Comparisons with other work on the estimation of the conditional probability of misclassification for normal populations.

Hills' (1966) paper, mentioned in Chapter 1, Section 2, contains a section on normal distributions in which he considers the various probabilities of misclassification and some possible estimators. The section is in part new material and in part a summary of work that has been done previously. There is no extensive material on the problem of estimating the conditional probability, $P_2$. Hills does suggest an approximate 90 percent (unconditional) confidence interval for $P_2$ in the univariate normal case, based on the point estimate $Q_D = P_2(R(\bar{x}_1, \bar{x}_2), \mu_2 = \bar{x}_2)$. The interval is

$$Q_D \pm 1.64 \times f(\frac{3}{2}d)N_{-\frac{1}{2}}^{-1},$$

which is arrived at by assuming the unconditional distribution of $P_2(R(\bar{x}_1, \bar{x}_2), \mu_2) - P_2(R(\bar{x}_1, \bar{x}_2), \hat{\mu}_2 = \bar{x}_2)$ to be approximately $N(0, f^2(\frac{3}{2}d)N_{-\frac{1}{2}}^{-1})$.

The work most comparable to ours is that of Lachenbruch (1965) and Lachenbruch and Mickey (1968)(see Chapter 1, Section 2). Considering the multivariate normal case, of the seven estimators of $P_2$ evaluated by them, using a Monte Carlo technique, five are included among the estimators compared here on conditional AMSE. The five estimators are
the ones designated by them as the R, D, DS, O, and OS estimators and designated here as $\text{P}_R$, $Q_D$, $Q_{DS}$, $Q_O$, and $Q_{OS}$. For the univariate normal case, we also studied their U estimator, $P_U$. For the univariate case, however, we did not study the estimators $Q_0$ and $Q_{OS}$ (O and OS), which are based on an approximation to $P_2^*$, but did study the estimator $Q_D^*$, based on the exact expression for $P_2^*$.

In the Lachenbruch and Mickey results, the estimators $P_R$ and $Q_D$ (R and D) are not good and are especially poor for large $p$ ($p = 20$); the estimators $Q_{DS}$, $Q_O$, and $P_U$ (DS, O, U) are better; and the estimator $Q_{OS}$ (OS) is the best of all seven estimators studied. In the overall results based on 288 samples from normal populations, with various values for samples sizes, $p$, and $\Delta^2$, the absolute value of the discrepancy between $P_2$ and $Q_{OS}$ (OS) was less than .05 in 118 samples (41 percent) and greater than .20 in 21 samples (7 percent). The corresponding numbers for $P_R$ (R) are 64 samples (22 percent) and 88 samples (31 percent).

In our multivariate results, the leading term in the AMSE is the same for all four estimators, $Q_D$, $Q_{DS}$, $Q_O$, and $Q_{OS}$ (D, DS, O, and OS). The leading term in the AMSE for the reclassification estimator, $P_R$ (R), is larger. Thus our comparison of estimators on conditional AMSE separates out $P_R$ (R) as a poorer estimator, but does not distinguish between $Q_D$, $Q_{DS}$, $Q_O$, and $Q_{OS}$ (D, DS, O, OS). Similarly, in our univariate results, the estimators $Q_D$, $Q_{DS}$, and $Q_D$ (D, DS, and the univariate estimator comparable to O and OS) have the same leading term in the conditional AMSE. The estimators $P_R$ and $P_U$ (R and U) also have equal leading terms and again this term is larger. Hence for the univariate case, the AMSE criterion groups the estimators

---

2 In the ensuing discussion the estimators are referred to in our notation with the Lachenbruch and Mickey notation in parentheses.
Q_0, Q_{OS}^{**} (D, DS) and the estimators P_R, P_U, (R, U) and indicates that the estimators Q_0, Q_{DS}^{**} are better than P_R, P_U. In contrast the multivariate Lachenbruch and Mickey results group (roughly speaking) P_R and Q_0 (R and D) as poor and Q_{DS}^{**} and P_U (DS and U) as better.

In regard to our 1st and 3rd conclusions, we note that, the Lachenbruch and Mickey results do not show all the estimators based on normality and original samples as equivalent. Their results also do not indicate that using the normality always makes a better estimator, since the estimator P_U (U) is better than the estimator Q_0 (D).

Neither Hills nor Lachenbruch and Mickey studied any estimators based on independent test samples, so there are no results to compare with ours on that class of estimators and no evidence for or against our second conclusion.

In connection with the discussion in Chapter 3, Section 4.4 on the distinction between estimators of P_2, P_{D}^{*} and P_{D}^{**}, note that the best estimator of P_2 in the Lachenbruch and Mickey results is Q_{OS}^{*} (OS), which is a "natural" estimator of P_{D}^{*}. In particular Q_{OS}^{*} (OS), which is obtained by estimating \Delta^2 in the Okamoto expansion for P_2 \text{ by } \frac{N_1 + N_2 - P - 3}{N_1 + N_2 - 2} \cdot D^2, is better than Q_{DS}^{**} (DS), which is obtained by estimating \Delta^2 in the expression for P_{D}^{**} by the same statistic. Furthermore, comparing Q_0^{*} and Q_0 (D and D), which are obtained in the same way as Q_{OS}^{*} and Q_{DS}^{**} (OS and DS) respectively, except that \Delta^2 is estimated by D^2, we find that Q_0^{*}(O) is better. Q_0(D) can also be arrived at by estimating \mu_2 in the expression for P_2 by \bar{x}_2 and our notation for the estimator reflects this derivation.
TABLE X
Notation Used in the Discussion of the Multivariate Normal Case; \( \pi_1 \)
is \( N(\mu_1, \Sigma) \) and \( \pi_2 \) is \( N(\mu_2, \Sigma) \).

\( \bar{x}_1, \bar{x}_2 \) are sample mean vectors based on original samples of sizes \( N_1, N_2 \) from \( \pi_1, \pi_2 \).

\( \bar{y}_1, \bar{y}_2 \) are sample mean vectors based on original samples from \( \pi_1, \pi_2 \) of size \( N/2 \) each.

\( \bar{x}_1, \bar{x}_2 \) are sample mean vectors based on test samples of sizes \( M_1, M_2 \) from \( \pi_1, \pi_2 \).

\( x \) is a new vector observation to be classified, assumed to be from \( \pi_2 \).

\( \alpha_j = \bar{x}_j - \mu_j, j = 1, 2 \)

\( \alpha^*_j = \bar{y}_j - \mu_j, j = 1, 2 \)

\( \Delta^2 = (\bar{\mu}_1 - \bar{\mu}_2)'\Sigma^{-1}(\bar{\mu}_1 - \bar{\mu}_2) \)

\( d^2 = (\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}(\bar{x}_1 - \bar{x}_2) \)

\( (d^*)^2 = (\bar{y}_1 - \bar{y}_2)'\Sigma^{-1}(\bar{y}_1 - \bar{y}_2) \)

\( \Delta, D \) and \( d^* \) are positive square roots of \( \Delta^2, D^2 \), and \( (d^*)^2 \).

\( B = \frac{1}{2}D(1-N_2^{-1})^{-\frac{1}{2}} \)

\( B^* = \frac{1}{2}D^*(1-2N_2^{-1})^{-\frac{1}{2}} \)

\( C = \frac{1}{2}D + D^{-1}(\bar{x}_1 - \bar{x}_2)'\Sigma^{-1} \alpha_{\bar{x}_2} \)

\( C^* = \frac{1}{2}D^* + (D^*)^{-1}(\bar{y}_1 - \bar{y}_2)'\Sigma^{-1} \alpha^*_{\bar{y}_2} \)

\( F(u), F(u, v; \rho) \) are standard normal distribution functions as defined in (1.3) and (1.4) of Chapter 3.

\( H(u, v; \rho) = F(-u, -v; \rho) \)

\( f(u) = F'(u) \)

\( G = F(\frac{1}{2}\Delta) \)

\( G = f(\frac{1}{2}\Delta) \)

\( p_0 = \text{proportion of original sample of size } N_2 \text{ from } \pi_2 \text{ misclassified by } R(S_1, S_2) \).
\( p_t = \) proportion of test sample of size \( n_2 \) from \( \Pi_2 \) misclassified by \( R(S_1, S_2) \).

\( P_2, P^* \) and \( P^{**} \) denote probabilities of misclassification: Conditional, unconditional and the value when all parameters are known.

**Estimators:**

\[ p \] with a letter(s) subscript denotes an estimator of \( P_2 \) not based on normality.

\[ Q, Q^*, Q^{**} \] with subscripts denote estimators of \( P_2 \) obtained by estimating unknown parameters in \( P_2 \), Okamoto expansion for \( P^{**}, P^{**} \).

\( \hat{p} \) with a subscript denotes other estimators based on normality.

**Relation of notation to univariate notation given in Table I:**

When \( p = 1 \) and \( \sigma^2 = 1 \)

\[ \Delta^2 \text{ reduces to } \delta^2, \quad \Delta \text{ to } |\delta| \]

\[ D^2 \text{ red. to } d^2, \quad D \text{ to } |d| \]

\[ B \] red. to \( |b| \)

\[ C \] red. to \( \frac{1}{2}|d| + \frac{d}{|d|} \alpha_2 = \begin{cases} c & \text{if } d > 0 \\ -c & \text{if } d < 0 \end{cases} \]
TABLE XI

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Expression for Estimator</th>
<th>Name or Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_R$</td>
<td>$P_0 = \text{proportion of } N_2 \text{ original observations from } \Pi_2 \text{ misclassified by } R(\bar{x}_1, \bar{x}_2)$.</td>
<td>Reclassification estimator</td>
</tr>
<tr>
<td>$P_T$</td>
<td>$P_t = \text{proportion of } N_2 \text{ test observations from } \Pi_2 \text{ misclassified by } R(\bar{x}_1, \bar{x}_2)$.</td>
<td>Test sample estimator</td>
</tr>
<tr>
<td>$P_{TSL}$</td>
<td>$P_0(N) - \frac{1}{2}P_0\left(\frac{1}{2}N\right) + \frac{1}{2}P_t\left(\frac{1}{2}N\right)$</td>
<td>Two straight lines estimator</td>
</tr>
<tr>
<td>$P_{Ra}$</td>
<td>$P_0 + F(B) - F\left(\frac{1}{2}D\right)$</td>
<td>Reclassification estimator adjusted for bias</td>
</tr>
<tr>
<td>$Q_D$</td>
<td>$1 - F\left(\frac{1}{2}D\right)$</td>
<td>$P_2(R(\bar{x}_1, \bar{x}<em>2), \mu</em>{x_2} = \bar{x}_2)$</td>
</tr>
<tr>
<td>$Q_T$</td>
<td>$1 - F\left(\frac{1}{2}D + D^{-1}(\bar{x}_1 - \bar{x}_2)\right)\Sigma^{-1}\left(\bar{x}_2 - \bar{x}_2\right)$</td>
<td>$P_2(R(\bar{x}_1, \bar{x}<em>2), \mu</em>{x_2} = \bar{x}_2)$</td>
</tr>
<tr>
<td>$Q_0^*$</td>
<td>$1 - F\left(\frac{1}{2}D\right) + \phi\left(\frac{1}{2}D\right)\left[\frac{1}{2}(p-1)D^{-1} + \frac{1}{2}D\right]N^{-1}$</td>
<td>Okamoto expansion for $P_2$ with $\Delta^2$ estimated by $D^2$; $N_1 = N_2 = N$</td>
</tr>
<tr>
<td>$Q_{OS}$</td>
<td>$1 - F\left(\frac{1}{2}D'\right) + \phi\left(\frac{1}{2}D'\right)\left[\frac{1}{2}(p+1)(D')^{-1} + \frac{1}{2}D\right]N^{-1}$</td>
<td>Okamoto expansion for $P_2$ with $\Delta^2$ estimated by $(D')^2$; $N_1 = N_2 = N$</td>
</tr>
<tr>
<td>$Q_{DS}$</td>
<td>$1 - F\left(\frac{1}{2}\left(\frac{N_1+N_2-3-p}{N_1+N_2-2}\right)^2D\right)$</td>
<td>Lachenbruch’s DS estimator or $P_{**}(\hat{\Delta})$ with $\hat{\Delta}^2 = \frac{N_1+N_2-3-p}{N_1+N_2-2}D^2$</td>
</tr>
</tbody>
</table>
TABLE XI (cont.)

<table>
<thead>
<tr>
<th>( \hat{\theta}_{RB} )</th>
<th>( 1 - F(A - W) ) where ( A = \frac{1}{2}(1-N_2^{-1})^{-\frac{3}{2}}D ) + ( (1-N_2^{-1})^{-\frac{1}{2}}D^{-1}(\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}x_2 ) ( W = (1-N_2^{-1})^{-\frac{1}{2}}D^{-1}(\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}x_2 )</th>
<th>Conditionally UMVU estimator obtained by conditional Rao-Blackwell theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\theta}_3 )</td>
<td>( 1 - F(\frac{3}{2}D(1+N_2^{-1})^{-\frac{1}{2}}) )</td>
<td>( \hat{\theta}_3 ) and ( \hat{\theta}_5 ) were invented by analogy to the ( \hat{\theta}_3 ) and ( \hat{\theta}_5 ) estimators derived in univariate case by using a prior distribution on the parameters.</td>
</tr>
<tr>
<td>( \hat{\theta}_5 )</td>
<td>( 1 - F(\frac{3}{2}D(1 + \frac{1}{2N_2})^{-\frac{1}{2}}) )</td>
<td></td>
</tr>
</tbody>
</table>

† Information on who has previously studied the estimators was given in Table III. In addition to the work referred to there, Lachenbruch (1965) and Lachenbruch and Mickey (1968) studied the estimators \( \hat{\theta}_0^* \) and \( \hat{\theta}_0^* \).
TABLE XII
Conditional Means and Variances for Estimators of $P_2$
When $\pi_1$ is $N(\mu_1, \Sigma)$ and $\pi_2$ is $N(\mu_2, \Sigma)$.

$$P_2(R(\bar{x}_1, \bar{x}_2), \mu_2) = 1-F(C)$$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Conditional Mean</th>
<th>Conditional Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{Ra}$</td>
<td>$1 - F(\frac{1}{2}D)$</td>
<td>Same as Var $P_R$.</td>
</tr>
<tr>
<td>$Q_D$</td>
<td>$1 - F(\frac{1}{2}D)$</td>
<td>Variance = 0</td>
</tr>
<tr>
<td>$Q_T$</td>
<td>$1 - F(C(1+M_2^{-1})^{-\frac{1}{2}})$</td>
<td>$F(C(1+M_2^{-1})^{-\frac{1}{2}}, C(1+M_2^{-1})^{-\frac{1}{2}}; \frac{1}{1+M_2})$</td>
</tr>
<tr>
<td>$Q^*$</td>
<td>$1 - F(\frac{1}{2}D) + \frac{1}{3}N^{-1}f(\frac{1}{2}D)[D+4(p-1)D^{-1}]$</td>
<td>Variance = 0</td>
</tr>
<tr>
<td>$Q_{OS}$</td>
<td>Same as for $Q^*$ with $D$ replaced by $D' = (\frac{2N-3-\beta}{2N-2})^{\frac{1}{2}}D$</td>
<td>Variance = 0</td>
</tr>
<tr>
<td>$Q_{DS}$</td>
<td>$1 - F(\frac{1}{2}\left(\frac{N_1+N_2-3-\beta}{N_1+N_2-2}\right)^{\frac{1}{2}}D)$</td>
<td>Variance = 0</td>
</tr>
<tr>
<td>$P_{RB}$</td>
<td>$P_2 = 1 - F(C)$</td>
<td>$F(C, C; M_2^{-1}) - F(C, C; 0)$</td>
</tr>
<tr>
<td>$P_3$</td>
<td>$1 - F(\frac{1}{2}D(1+N_2^{-1})^{-\frac{1}{2}})$</td>
<td>Variance = 0</td>
</tr>
<tr>
<td>$P_5$</td>
<td>$1 - F(\frac{1}{2}D(1+\frac{1}{2N_2})^{-\frac{1}{2}})$</td>
<td>Variance = 0</td>
</tr>
</tbody>
</table>
Table XIII

Asymptotic Expansions Needed to Get the Asymptotic Expressions for Conditional Squared Bias, Variance and Mean Square Error That are Given in Tables XIV, XV, and XVI.

<table>
<thead>
<tr>
<th>Symbol or Function</th>
<th>Constant</th>
<th>Term of order $\mathcal{O}(N^{-1+\varepsilon})$</th>
<th>Term of order $\mathcal{O}(N^{-3/2+\varepsilon})$</th>
<th>Term of order $\mathcal{O}(N^{-2+\varepsilon})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>$\Delta$</td>
<td>$+\Delta^{-1}v$</td>
<td>$+\frac{1}{2}\Delta^{-1}u-\frac{3}{2}\Delta^{-3}v^2$</td>
<td>$-\frac{1}{2}\Delta^{-3}uv+\frac{1}{2}\Delta^{-5}v^3$</td>
</tr>
<tr>
<td>$D^{-1}$</td>
<td>$\Delta^{-1}$</td>
<td>$-\Delta^{-3}v$</td>
<td>$-\frac{1}{2}\Delta^{-3}u+\frac{3}{2}\Delta^{-5}v^2$</td>
<td>+++</td>
</tr>
<tr>
<td>$C$</td>
<td>$\frac{1}{2}\Delta$</td>
<td>$+\Delta^{-1}(\frac{3}{2}v+z)$</td>
<td>$+\Delta^{-1}(\frac{3}{2}u+w)-\Delta^{-3}(\frac{1}{2}v^2+vz)$</td>
<td>+++</td>
</tr>
<tr>
<td>$B$</td>
<td>$\frac{1}{2}\Delta$</td>
<td>$+\Delta^{-1}(\frac{3}{2}v+z)$</td>
<td>$+\Delta^{-1}(\frac{3}{2}u+w)-\Delta^{-3}(\frac{1}{2}v^2+vz)$</td>
<td>+++</td>
</tr>
<tr>
<td>$\frac{1}{2}D(1+N^{-1})^{-\frac{1}{2}}$</td>
<td>$\frac{1}{2}\Delta$</td>
<td>$+\Delta^{-1}v$</td>
<td>$+\frac{1}{2}\Delta^{-1}u-\frac{3}{2}\Delta^{-3}v^2$</td>
<td>+++</td>
</tr>
<tr>
<td>$\frac{1}{2}D(1+\frac{1}{2N})^{-\frac{1}{2}}$</td>
<td>$\frac{1}{2}\Delta$</td>
<td>$+\Delta^{-1}v$</td>
<td>$+\frac{1}{2}(\Delta^{-1}u-\Delta^{-3}v^2+\Delta N^{-1})$</td>
<td>$+\frac{1}{2}\Delta^{-1}v(-\Delta^{-2}u+\Delta^{-4}v^2+N^{-1})$</td>
</tr>
<tr>
<td>$(1+\frac{1+p}{2N-3-p})^{-\frac{1}{2}}$</td>
<td>$1$</td>
<td>$+0$</td>
<td>$-\frac{1}{2}(1+p)(2N-3-p)^{-1}$</td>
<td>$+0$</td>
</tr>
<tr>
<td>$\frac{1}{2}D'\frac{1}{2}D'(1+\frac{1+p}{2N-3-p})^{-\frac{1}{2}}$</td>
<td>$\frac{1}{2}\Delta$</td>
<td>$+\Delta^{-1}v$</td>
<td>$+\Delta^{-1}u-\Delta^{-3}v^2$</td>
<td>$-\frac{1}{2}\Delta(1+p)(2N-3-p)^{-1}$</td>
</tr>
<tr>
<td>$(D')^{-1}$</td>
<td>$\Delta^{-1}$</td>
<td>$-\Delta^{-3}v$</td>
<td>$+$</td>
<td>$++$</td>
</tr>
<tr>
<td>$C(1+N^{-1})^{-\frac{1}{2}}$</td>
<td>$\frac{1}{2}\Delta$</td>
<td>$+\Delta^{-1}(\frac{3}{2}v+z)$</td>
<td>$+\Delta^{-1}(\Delta^{-1}u-\Delta^{-3}v^2-\Delta N^{-1})$</td>
<td>$+\Delta^{-1}w-\Delta^{-3}vz$</td>
</tr>
</tbody>
</table>
\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Table XIII (cont.) & Constant & $o_p(N^{-\frac{3}{2}+\varepsilon})$ & $o_p(N^{-1+\varepsilon})$ & $o_p(N^{-3/2+\varepsilon})$ & $o_p(N^{-2+\varepsilon})$ \\
\hline
$f(\frac{1}{2}D)$ & $g$ & $-\frac{1}{2}g\nu$ & & & \\
$f(\frac{1}{2}D')$ & $g$ & $-\frac{1}{2}g\nu$ & & & \\
\hline
$F(\frac{1}{2}D)$ & $g$ & $+\frac{1}{2}\Delta^{-1}g\nu$ & $+\frac{1}{2}\Delta^{-1}g(U-k\nu^2)-k\Delta^{-\frac{3}{2}}g\nu^2$ & & \\
$F(B)$ & $g$ & $+\frac{1}{2}\Delta^{-1}g\nu$ & $+\frac{1}{2}\Delta^{-1}g(U-k\nu^2)-k\Delta^{-\frac{3}{2}}g\nu^2$ & & \\
$F(C)$ & $g$ & $+\Delta^{-1}g(\frac{1}{2}\nu+z)$ & $+\frac{1}{2}\Delta^{-1}g(U-k\nu^2+4w-vz-z^2)$ & & \\
$F(B^*)$ & $g$ & & & & \\
$F(C^*)$ & $g$ & & & & \\
$F(\frac{1}{2}D(1+N^{-1})^{-\frac{3}{2}})$ & $g$ & $+\frac{1}{2}\Delta^{-1}g\nu$ & $+\frac{1}{2}\Delta^{-1}g(U-k\nu^2)-k\Delta^{-\frac{3}{2}}g\nu^2$ & & \\
$F(\frac{1}{2}D(1+\frac{1}{2N})^{-\frac{3}{2}})$ & $g$ & $+\frac{1}{2}\Delta^{-1}g\nu$ & $+\frac{1}{2}\Delta^{-1}g(U-k\nu^2)-k\Delta^{-\frac{3}{2}}g\nu^2$ & & \\
$F(\frac{1}{2}D')$ & $g$ & $+\frac{1}{2}\Delta^{-1}g\nu$ & $+\frac{1}{2}\Delta^{-1}g(U-k\nu^2)-k\Delta^{-\frac{3}{2}}g\nu^2$ & & \\
$F(C(1+N^{-1})^{-\frac{3}{2}})$ & $g$ & $+\Delta^{-1}g(\frac{1}{2}\nu+z)$ & $+\frac{1}{2}\Delta^{-1}g(U-k\nu^2+4w-vz-z^2)$ & & \\
\hline
\end{tabular}
\end{table}
Table XIII (cont.)

<table>
<thead>
<tr>
<th>Constant</th>
<th>$o_p(N^{-3/2}+\varepsilon)$</th>
<th>$o_p(N^{-1}+\varepsilon)$</th>
<th>$o_p(N^{-3/2}+\varepsilon)$</th>
<th>$o_p(N^{-2}+\varepsilon)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H(B,B;\frac{1}{N-1})-H(B,B;0)$</td>
<td>$= 0$</td>
<td>$= 0$</td>
<td>$\mathcal{J}_{2N}^{-1}$</td>
<td>$+ \frac{1}{2} \mathcal{J}_{2N}^{-1}$</td>
</tr>
<tr>
<td>$H(B,B;0)$</td>
<td>$= H(\frac{1}{2}\Delta,\frac{1}{2}\Delta;0)$</td>
<td>$= \Delta^{-1} \mathcal{J}(1-\mathcal{J})V$</td>
<td>$+++$</td>
<td>$+++$</td>
</tr>
<tr>
<td>$F(C,C;N^{-1})-F(C,C;0)$</td>
<td>$= 0$</td>
<td>$= 0$</td>
<td>$\mathcal{J}_{2N}^{-1}$</td>
<td>$-\mathcal{J}_{2N}^{-1}$</td>
</tr>
<tr>
<td>$F(C',C';\frac{1}{N+1})-F(C',C';0)$</td>
<td>$= 0$</td>
<td>$= 0$</td>
<td>$\mathcal{J}_{2N}^{-1}$</td>
<td>$-\mathcal{J}_{2N}^{-1}$</td>
</tr>
</tbody>
</table>

Expansions for terms such as $(1+N^{-1})^{-\frac{1}{2}}$ are given in Table V.

$+++$ indicates a non-zero term of the order given in the column heading.

$\Delta^2 = (\mu_1 - \mu_2)' \Sigma^{-1}(\mu_1 - \mu_2)$, $\mathcal{J} = F(\frac{1}{2}\Delta)$, $\mathcal{J} = f(\frac{1}{2}\Delta)$, $D' = \frac{3}{2} D(\frac{2N-3-P}{2N-2})$, $U = (\alpha_1 - \alpha_2)' \Sigma^{-1}(\alpha_1 - \alpha_2)$, $V = (\mu_1 - \mu_2)' \Sigma^{-1}(\alpha_1 - \alpha_2)$, $W = (\alpha_1 - \alpha_2)' \Sigma^{-1} \alpha_2$, $Z = (\mu_1 - \mu_2)' \Sigma^{-1} \alpha_2$.

$U^*, V^*, W^*, Z^*$ are the same as $U, V, W, Z$ with $\alpha_1, \alpha_2$ replaced by $\alpha_1^*, \alpha_2^*$. 
Table XIV

Conditional Squared Biases for Estimators of $P_2(R(x_1, x_2); \mu_2) = 1 - F(C)$ when $\Pi_1$ is $N_p(\mu_1, \Sigma)$ and $\Pi_2$ is $N_p(\mu_2, \Sigma)$: Exact and Asymptotic Expressions.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Squared Bias</th>
<th>Asymptotic Expression for Squared Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_R$</td>
<td>$[F(B)-F(C)]^2$</td>
<td>$\Delta^2 g^{2z^2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$+ A - \frac{1}{2} g^{2z^2 N^{-1}}$</td>
</tr>
<tr>
<td>$P_T$</td>
<td>Zero</td>
<td></td>
</tr>
</tbody>
</table>
| $P_{TSL}$  | $[F(B)-F(C)]^2 + \frac{1}{\Delta}[F(B^*)-F(C^*)]^2$ \[
|            | - $[F(B)-F(C)][F(B^*)-F(C^*)]$      | $\Delta^2 g^{2z^2 + \frac{1}{2} \Delta^2 g^2(z^*)^2}$  \[
|            |                                       | $- \Delta^2 g^{2z^2}$                   \[
|            |                                       | $+ A$                                   |
| $P_{Ra}$   | $[F(\frac{3}{2}D)-F(C)]^2$          | $\Delta^2 g^{2z^2}$                    |
|            |                                     | $+ A$                                   |
| $Q_D$      | $[F(\frac{1}{2}D)-F(C)]^2$          | $\Delta^2 g^{2z^2}$                    |
|            |                                     | $+ A$                                   |
| $Q_T$      | $[F(C(1+N^{-1})-\frac{1}{2})-F(C)]^2$| 0                                       |
|            |                                     | $+ O$                                   |
| $Q_0^*$    | $[F(\frac{1}{2}D)-F(C)-\frac{1}{8}N^{-1}f(\frac{1}{2}D)[D+4(p-1)D^{-1}]]^2$ \[
|            |                                       | $\Delta^2 g^{2z^2}$                    |
|            |                                       | $+ A + \frac{1}{2} g^2[2(p-1)\Delta^{-2}+1]ZN^{-1}$ |
| $Q_0^{*}$  | Same as $Q_0^*$ with $D$ replaced by  | $\Delta^2 g^{2z^2}$                    |
|            | $D' = (\frac{2N-3-P}{2N-2})^{\frac{1}{2}}$ \[
|            |                                       | $+ A + \frac{1}{2} g^2[2(p-1)\Delta^{-2}+1]ZN^{-1}$ |
|            |                                       | $+ \frac{1}{2} g^2Z(\frac{1+p}{2N-3-p})$    |
Table XIV (cont.)

| \( q_{DS}^{**} \) | \( [F(\frac{1}{2}D(1+a)^{-\frac{1}{2}})-F(C)]^2 \) | \( \Delta^{-2}g^2z^2 \) | \(+ A + \frac{1}{2} g^2z^{\frac{1+p}{2N-3-p}} \) | +++ |
|---|---|---|---|
| \( \rho_{RB} \) | Zero | | |
| \( \rho_3 \) | \( [F(\frac{1}{2}D(1+\frac{1}{N}z^2))^{-\frac{1}{2}})-F(C)]^2 \) | \( \Delta^{-2}g^2z^2 \) | \(+ A + \frac{1}{2} g^2z^{N^{-1}} \) | +++ |
| \( \rho_5 \) | \( [F(\frac{1}{2}D(1+\frac{1}{2N})^{-\frac{1}{2}})-F(C)]^2 \) | \( \Delta^{-2}g^2z^2 \) | \(+ A + \frac{1}{2} g^2z^{N^{-1}} \) | +++ |

+++ indicates a non-zero term of the order given in the column heading.

\[ U = (\alpha_1 - \alpha_2)'\Sigma^{-1}(\alpha_1 - \alpha_2), \quad V = (\mu_1 - \mu_2)'\Sigma^{-1}(\alpha_1 - \alpha_2), \quad W = (\alpha_1 - \alpha_2)'\Sigma^{-1}\alpha_2, \quad Z = (\mu_1 - \mu_2)'\Sigma^{-1}\alpha_2. \]

\[ A = \frac{1}{2}A^{-2}g^{2(4wz-vz^2-z^3)-2A^{-1}g^{2vz^2}}. \quad g = f(\frac{1}{2}A), \quad p \text{ is the dimension of the normal variable}. \]

For other notation see Table X.
**Table XV**

Asymptotic Expressions for Conditional Variances of Estimators of $P_2(R(X_1, X_2), \mu_2)$ When $\Pi_1$ is $N_p(\mu_1, \Sigma)$ and $\Pi_2$ is $N_p(\mu_2, \Sigma)$

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Conditional Variance: Term of order $o_p(N^{-1+\epsilon})$</th>
<th>$o_p(N^{-3/2+\epsilon})$</th>
<th>$o_p(N^{-2+\epsilon})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_R$</td>
<td>$(\mathcal{Y}(1-\mathcal{Y}) - \mathcal{G}^2)N^{-1}$ + $\frac{1}{2}\Delta^{-1}\mathcal{G}(\Delta \mathcal{G} + 1 - 2\mathcal{Y})VN^{-1}$</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>$P_T$</td>
<td>$\mathcal{G}(1-\mathcal{Y})N^{-1}$ + $\Delta^{-1}\mathcal{G}(1-2\mathcal{Y})(\frac{1}{2}V + Z)N^{-1}$</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>$P_{TSL}$</td>
<td>$\frac{1}{2}[4\mathcal{Y}(1-\mathcal{Y}) - 3\mathcal{G}^2]N^{-1}$</td>
<td>+++</td>
<td></td>
</tr>
</tbody>
</table>

(Assumption Set 2)

<table>
<thead>
<tr>
<th>$P_{Ra}$</th>
<th>Same as Var $P_R$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_D$</td>
<td>Variance = 0</td>
<td></td>
</tr>
<tr>
<td>$Q_T$</td>
<td>$\mathcal{G}^2N^{-1}$</td>
<td>$-\mathcal{G}^2(\frac{1}{2}V + Z)N^{-1}$</td>
</tr>
<tr>
<td>$Q^*_0$</td>
<td>Variance = 0</td>
<td></td>
</tr>
<tr>
<td>$Q^*_0$</td>
<td>Variance = 0</td>
<td></td>
</tr>
<tr>
<td>$Q_{OS}$</td>
<td>Variance = 0</td>
<td></td>
</tr>
<tr>
<td>$Q_{DS}$</td>
<td>Variance = 0</td>
<td></td>
</tr>
<tr>
<td>$P_{SR}$</td>
<td>$\mathcal{G}^2N^{-1}$</td>
<td>$-\mathcal{G}^2(\frac{1}{2}V + Z)N^{-1}$</td>
</tr>
<tr>
<td>$P_{3}$</td>
<td>Variance = 0</td>
<td></td>
</tr>
<tr>
<td>$P_{5}$</td>
<td>Variance = 0</td>
<td></td>
</tr>
</tbody>
</table>

+++ indicates a non-zero term of the order given in the column heading.

$\mathcal{Y} = F(\frac{1}{2}\Delta), \mathcal{G} = f(\frac{1}{2}\Delta), \Delta^2 = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$

$V = (\mu_1 - \mu_2)'\Sigma^{-1}(\alpha_1 - \alpha_2), Z = (\mu_1 - \mu_2)'\Sigma^{-1}\alpha_2$
Table XVI

Asymptotic Expressions for Conditional Mean Square Errors of Estimators of \( p_2(R_1, \overline{R_2}), \mu_2 \) When \( \Pi_1 \) is \( N_p(\mu_1, \Sigma) \) and \( \Pi_2 \) is \( N_p(\mu_2, \Sigma) \)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Mean Square Error: Term of order ( o_p(N^{-1+\varepsilon}) )</th>
<th>( o_p(N^{-3/2+\varepsilon}) )</th>
<th>( o_p(N^{-2+\varepsilon}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_R )</td>
<td>( \Delta^2 g^2 z^2 + [G(1-\bar{G})-g^2]N^{-1} )</td>
<td>( A + \frac{1}{2} G[G + \Delta^{-1}(1-2\bar{G})]VN^{-1} )</td>
<td>( \Delta^2 g^2 z^2 + [G(1-\bar{G})-g^2]N^{-1} )</td>
</tr>
<tr>
<td>( P_T )</td>
<td>( \bar{G}(1-\bar{G})N^{-1} )</td>
<td>( \Delta^{-1} g(1-2\bar{G})(\frac{1}{2}V+Z)N^{-1} )</td>
<td>( \Delta^{-1} g(1-2\bar{G})(\frac{1}{2}V+Z)N^{-1} )</td>
</tr>
<tr>
<td>( P_{TS} ) (Assumption Set 2)</td>
<td>( \Delta^2 g^2 z^2 + \frac{1}{2}[4G(1-\bar{G})-3g^2]N^{-1} )</td>
<td>( \Delta^2 g^2 z^2 + \frac{1}{2}[4G(1-\bar{G})-3g^2]N^{-1} )</td>
<td>( \Delta^2 g^2 z^2 + \frac{1}{2}[4G(1-\bar{G})-3g^2]N^{-1} )</td>
</tr>
<tr>
<td>( Q_a )</td>
<td>( \Delta^2 g^2 z^2 + [G(1-\bar{G})-g^2]N^{-1} )</td>
<td>( A + \frac{1}{2} G[G + \Delta^{-1}(1-2\bar{G})]VN^{-1} )</td>
<td>( \Delta^2 g^2 z^2 + [G(1-\bar{G})-g^2]N^{-1} )</td>
</tr>
<tr>
<td>( Q_o )</td>
<td>( \Delta^2 g^2 z^2 )</td>
<td>( -g^2 \left( \frac{1}{2}V+Z \right)N^{-1} )</td>
<td>( -g^2 \left( \frac{1}{2}V+Z \right)N^{-1} )</td>
</tr>
<tr>
<td>( Q_o^{*} )</td>
<td>( \Delta^2 g^2 z^2 )</td>
<td>( A + \frac{1}{2} g^2 \left[ 2(p-1)\Delta^{-2}+1 \right]ZN^{-1} )</td>
<td>( A + \frac{1}{2} g^2 \left[ 2(p-1)\Delta^{-2}+1 \right]ZN^{-1} )</td>
</tr>
<tr>
<td>( Q_{OS} )</td>
<td>( \Delta^2 g^2 z^2 )</td>
<td>( A + \frac{1}{2} g^2 \left[ 2(p-1)\Delta^{-2}+1 \right]ZN^{-1} )</td>
<td>( A + \frac{1}{2} g^2 \left[ 2(p-1)\Delta^{-2}+1 \right]ZN^{-1} )</td>
</tr>
<tr>
<td>( Q_{DS}^{**} )</td>
<td>( \Delta^2 g^2 z^2 )</td>
<td>( A + \frac{1}{2} g^2 \left( \frac{1+p}{2N-3-p} \right)Z )</td>
<td>( A + \frac{1}{2} g^2 \left( \frac{1+p}{2N-3-p} \right)Z )</td>
</tr>
<tr>
<td>( R_{RB} )</td>
<td>( \Delta^2 g^2 z^2 )</td>
<td>( -g^2 \left( \frac{1}{2}V+Z \right)N^{-1} )</td>
<td>( -g^2 \left( \frac{1}{2}V+Z \right)N^{-1} )</td>
</tr>
<tr>
<td>( R_{3} )</td>
<td>( \Delta^2 g^2 z^2 )</td>
<td>( A + \frac{1}{2} g^2 ZN^{-1} )</td>
<td>( A + \frac{1}{2} g^2 ZN^{-1} )</td>
</tr>
<tr>
<td>( R_{5} )</td>
<td>( \Delta^2 g^2 z^2 )</td>
<td>( A + \frac{1}{2} g^2 ZN^{-1} )</td>
<td>( A + \frac{1}{2} g^2 ZN^{-1} )</td>
</tr>
</tbody>
</table>

\( U = (\mu_1 - \mu_2)' \Sigma^{-1} (\alpha_1 - \alpha_2), \ V = (\mu_1 - \mu_2)' \Sigma^{-1} (\alpha_1 - \alpha_2), \ W = (\alpha_1 - \alpha_2)' \Sigma^{-1} \alpha_2, \ Z = (\mu_1 - \mu_2)' \Sigma^{-1} \alpha_2, \ A = \frac{1}{2} g^2 (4W^2 - VZ^2 - Z^3) - 2\Delta g^2 VZ, \ \beta = f(\frac{3}{2} \Delta), \ \delta = f(\frac{3}{2} \Delta), \ p = \text{dimension of the multivariate normal variables. } +++ \text{ is same as in preceding tables.} \)
CHAPTER 5

ESTIMATION OF A QUANTITY OF THE FORM $\varphi(R(x), \theta)$.

The conditional probability of misclassification, given a fixed rule based on observations from the populations, will, in general, be a function of both sample statistics and population parameters. For example, in the normal case considered in Chapters 3 and 4, the conditional probability of misclassifying an observation from $\Pi_i$, $i = 1, 2$ is a function of the statistics $\bar{x}_1, \bar{x}_2$ and of the parameter $\mu_i$. Let $\bar{x}$ denote a (vector-valued) random variable with probability distribution indexed by $\theta$ and let $x$ be an observation of $\bar{x}$. Let $R(x)$, where $R$ is a function of $\bar{x}$, be fixed and let $\varphi = \varphi(R(x), \theta)$ be a function of $R(x)$ and $\theta$. Then the conditional probability of misclassifying an observation from $\Pi_2$ as $\Pi_1$, which was denoted by $P_2(R(S_1, S_2), \theta)$ in Chapter 1, with $R(S_1, S_2)$ being the rule based on samples $S_1, S_2$ from $\Pi_1, \Pi_2$, is a quantity of the form $\varphi(R(x), \theta)$.

In the following sections some theory related to the general problem of estimating $\varphi(R(x), \theta)$ is given. The situation in stated in Section 1. Section 2 is on Cramér-Rao type bounds. Section 3 is on Chapman-Robbins bounds, that is, Cramér-Rao type bounds using differences instead of derivatives and eliminating regularity conditions. Section 4 gives Rao-Blackwell type theory involving sufficient statistics. The general theory is applied to some particular cases, including the normal classification problem.

1. Situation.

Let $(\Omega, \mathcal{F})$ be a measurable space with probability measures $P_\theta$, $\theta \in \Theta$ defined on $\mathcal{F}$. Let $X$ be a (vector-valued) random variable defined on $\Omega$ with range space $(\mathcal{X}, \mathcal{Q})$ and for each $\theta \in \Theta$ let $P^X_\theta$ be a probability measure on $\mathcal{Q}$ defined in terms of $P_\theta$ by the relationship...
Let $x$ denote an observation of $X$. Let $R = R(x)$ be fixed and assume
the conditional distribution of $X$, given $R$, denoted $P_{\theta}^{X|R}$, is a
probability distribution. Let $\phi = \phi(R(x), \theta)$ be a quantity to be
estimated. Let $\hat{\phi} = \hat{\phi}(x)$ denote an estimator.

2. Cramér-Rao Type Bound.

In this section we consider a Cramér-Rao type bound for estimation
of a function of observations and parameters. Typically we estimate
$\phi(R(x), \theta)$ where $X = (X_1, \ldots, X_n)'$ and the $X_i$ are independent
random variables. The theorem will, however, be stated for transformed
variables $R, Y$ where $R = R(x)$ and $Y = (Y_1(x), \ldots, Y_{n-1}(x))'$.
Since the bound is derived with the assumption that $\theta$ is a scalar,
the $X_i$ and subsequently the $R$ and $Y_j$ will usually be univariate.
Note, however, that if a vector valued $R$ is needed we can take
$R = (R_1, \ldots, R_m)'$ and then take $Y = (Y_1, \ldots, Y_{n-m})'$.

2.1 Derivation of the bound.

Theorem 2.1. Let $R, Y$ be random variables with a joint distribution
depending on $\theta \in \Theta$, where $\Theta$ is an open interval of the real line.
Let the conditional probability of $Y$ given $R = r$ be denoted by
$P_{\theta}^{Y|R}$ and let $Y(\mathcal{B})$ denote the range space of $Y$. Assume that for
each $\theta \in \Theta$ there exists a function $g_{\theta}(y|r)$ and a $\sigma$-finite measure
$\mu$ on $(Y, \mathcal{B})$ such that
$$
\int_{B} g_{\theta}(y|r) \, d\mu(y) = P_{\theta}^{Y|R} (B), \ B \in \mathcal{B}.
$$
Let $\phi = \phi(R, \theta)$ be a function of $R, \theta$ such that $\frac{\partial \phi}{\partial \theta}$ is defined.
Suppose there exists a conditionally unbiased estimator of $\phi$ given
$R$, that is a function $\hat{\phi} = \hat{\phi}(R, Y)$ satisfying
$\int \hat{\phi}(r, y) g_{\theta}(y|r) \, d\mu(y) = \phi(r, \theta)$
for each $r$ and $\theta$. Let $S$ be defined as
Assume that \( g_\theta(y|x) \) satisfies the following regularity conditions, which parallel those for the usual Cramér-Rao bound.

1) For almost all \((\mu)_x, \frac{\partial g_\theta(y|x)}{\partial \theta}\) exists for all \(\theta\),

2) \[ \frac{\partial}{\partial \theta} \int g_\theta(y|x) du(y) = \int \frac{\partial}{\partial \theta} g_\theta(y|x) du(y), \]

3) \[ \frac{\partial}{\partial \theta} \int \Phi(x, y) g_\theta(y|x) du(y) = \int \frac{\partial}{\partial \theta} \Phi(x, y) g_\theta(y|x) du(y), \]

4) \[ E_\theta(S^2|R=r) = \int S^2(x, y) g_\theta(y|x) du(y) > 0, \quad \text{all } \theta \in \Theta. \]

Then

\[
(2.2) \quad \text{Var}(\Phi|R=r) \geq \frac{1}{E_\theta(S^2|R=r)} \left[ \frac{\partial \log g_\theta(y|x)}{\partial \theta} \right]^2 g_\theta(y|x) du(y).
\]

**Proof:** (The proof as given here parallels that for ordinary Cramér-Rao bound.) From the properties of a density function, we have

\[
(2.3) \quad \int g_\theta(y|x) du(y) = 1
\]

and from the unbiasedness condition on \( \Phi \) we have

\[
(2.4) \quad \int \Phi(x, y) g_\theta(y|x) du(y) = \Phi(x, \theta).
\]

Taking derivatives of both sides of the equations with respect to \(\theta\), we get from (2.3) and (2.4) respectively,

\[ E_\theta(S|R) = 0, \quad E_\theta(\Phi S|R) = 1. \]

Then by the Cauchy-Schwarz inequality,

\[ \text{Var}(\Phi|R) \geq \frac{1}{E_\theta(S^2|R)}, \]

and the theorem is proved.

Suppose, going back to the situation of Section 1, that the given variable is \( X = (X_1, \ldots, X_n)' \) with distribution \( P^X_\theta \) and that the
function to be estimated is \( \varphi = \varphi(R(x), \theta) \), where \( R(x) \) is some function of \( x \). Let \( \hat{\varphi} = \hat{\varphi}(x) \) be an estimator. Then we are interested in obtaining a bound on the conditional variance of \( \hat{\varphi} \), given \( R = r \).

Assume there exists a 1-1 transformation of \( x \) to
\[(R, y) \equiv (R(x), Y_1(x), \ldots, Y_{n-1}(x))'.\] Then if the conditions of the hypothesis are satisfied, Theorem 2.1 gives a bound for the variance of \( \hat{\varphi} \) in terms of the conditional distribution of \( y \) given \( R \). The question arises as to whether, for given \( x \), \( P^X_\theta \), and \( R \), the bound will depend on the particular \( Y_i \) involved in the transformation. The theorem given below (Theorem 2.2), which applies to the case where there are densities with respect to Lebesgue measure, shows that with reasonable assumptions the bound is independent of the particular transformation used.

**Theorem 2.2.** Let \( \vec{x} = (x_1, \ldots, x_n)' \) be a random variable with distribution indexed by \( \theta \in \Theta \), where \( \Theta \) is an open interval of the real line. Let \( R(x) \) be a function of \( x \) and let \( \varphi = \varphi(R(x), \theta) \) be a function to be estimated. Assume that \( \frac{\partial \varphi}{\partial \theta} \) is defined. Assume that there exist two 1-1 transformations of \( x \) to
\[(R(x), Y_1(x), \ldots, Y_{n-1}(x))' \equiv (R, y)' \] and of \( \vec{x} \) to
\[(R(x), Z_1(x), \ldots, Z_{n-1}(x))' \equiv (R, z)' \], such that the two conditional densities (with respect to Lebesgue measure) of \( y \) given \( R \) and of \( z \) given \( R \) are defined and satisfy the regularity conditions of Theorem 2.1. Assume also that the Jacobians of both transformations are well defined and never zero. Suppose there exists an estimator \( \hat{\varphi} = \hat{\varphi}(r, y) = \hat{\varphi}(r, z) \), such that \( \hat{\varphi} \) is conditionally unbiased with respect to both conditional distributions. Then two bounds on the conditional variance of \( \hat{\varphi} \) can be computed according to Theorem 2.1, using the two different sets of transformed variables, and these two bounds are equal.
Proof: Let \( M_\theta(r, \chi) \) and \( N_\theta(r, z) \) denote the joint densities of \((R, \chi)\) and of \((R, z)\), respectively. The existence of the two 1-1 transformations from \( \chi \) to \((R, \chi)\) and from \( z \) to \((R, z)\) implies the existence of a 1-1 transformation from \((R, \chi)\) to \((R, z)\), which will be of the form \( R \mapsto R, \chi \mapsto z \). Thus, with the Jacobians defined and non-zero,

\[
N_\theta(r, z) = M_\theta(r, \chi(z))|J|,
\]

where \( \chi(z) \) denotes \( \chi \) as a function of \( z \) and where

\[
J = \begin{vmatrix}
\frac{\partial(r, y_1, \ldots, y_{n-1})}{\partial(r, z_1, \ldots, z_{n-1})}
\end{vmatrix}.
\]

Because of the form of the transformation the Jacobian will reduce to

\[
J = \begin{vmatrix}
\frac{\partial(y_1, \ldots, y_{n-1})}{\partial(z_1, \ldots, z_{n-1})}
\end{vmatrix}.
\]

Hence, denoting the marginal density of \( R \) by \( f_\theta(r) \),

\[
f_\theta(r) = \int M_\theta(r, \chi) d\chi = \int M_\theta(r, \chi(z))|J|dz = \int N_\theta(r, z)dz.
\]

The two conditional densities of \( \chi|r \) and \( z|r \) are then defined as

\[
(2.5) \quad g_\theta(\chi|r) = \frac{M_\theta(r, \chi)}{f_\theta(r)},
\]

\[
(2.5) \quad h_\theta(z|r) = \frac{M_\theta(r, \chi(z))}{f_\theta(r)} |J|.
\]

Now let

\[
(2.6) \quad s_1(r, \chi) = \frac{\partial \log g_\theta(\chi|r)}{\partial \phi} = \frac{\partial \log g_\theta(\chi|r)}{\partial \theta} \left( \frac{\partial \phi}{\partial \theta} \right)^{-1}
\]

and

\[
(2.7) \quad s_2(r, z) = \frac{\partial \log h_\theta(z|r)}{\partial \phi} = \frac{\partial \log h_\theta(z|r)}{\partial \theta} \left( \frac{\partial \phi}{\partial \theta} \right)^{-1}.
\]

Then

\[
(2.8) \quad s_1(r, y) = \left( \frac{\partial \phi}{\partial \theta} \right)^{-1} \left[ \frac{\partial}{\partial \theta} \log M_\theta(r, \chi) - \frac{\partial}{\partial \theta} \log f_\theta(r) \right]
\]
and
\[(2.9) \quad S_2(r, z) = (\frac{\partial}{\partial \theta})^{-1} \left[ \frac{\partial}{\partial \theta} \log M_\theta(r, \chi(z)) + \frac{\partial}{\partial \theta} \log |J| \right. - \left. \frac{\partial}{\partial \theta} \log f_\theta(r) \right].\]

Since \(\frac{\partial}{\partial \theta} \log |J| = 0\) and \(\frac{\partial}{\partial \theta} \log M_\theta(r, \chi(z)) = \left[ \frac{\partial}{\partial \theta} \log M_\theta(r, \chi) \right]_{\chi=\chi(z)}\),

we have the result
\[(2.10) \quad S_2(r, z) = S_1(r, \chi) \bigg|_{\chi=\chi(z)}.\]

Then using (2.5) and (2.10),
\[(2.11) \quad E(S_2^2(r, z) | R=r) = \int S_2^2(r, z) h_\theta(z | r) dz \]
\[= \int S_1^2(r, \chi(z)) \frac{M_\theta(r, \chi(z))}{f_\theta(r)} |J| dz \]
\[= \int S_1^2(r, \chi) g_\theta(\chi | r) d\chi \]
\[= E(S_1^2(r, \chi) | R=r).\]

Thus the two bounds are equal.

2.2 Application of the Cramér-Rao bound to the case where \(X_i\) is
\[N(\mu, 1)\] and \(R(x) = \Sigma c_i x_i.\]

We will now compute the Cramér-Rao bound derived in Section 2.1 for a special case.

Let \(x' = (x_1, \ldots, x_n)\), where the \(x_i\) are independent and identically distributed\(^1\) \(N(\mu, 1)\), and let \(x\) be an observation of \(x\). Let \(R(x) = \Sigma c_i x_i, c_1 \neq 0\). If \(R\) is sufficient for \(\mu\) (equivalently if all \(c_i\) are equal), the theory will not apply; hence we consider any other weighted sum \(\Sigma c_i x_i\). Let \(\varphi = \varphi(R, \mu)\) be a quantity to be estimated. In this section \(\varphi\) is arbitrary; in Sections 2.3 and 2.4 below, we exhibit particular cases.

\(^{1}\)The symbol \(\mu\) is used here to denote the mean of the normal variate and is not related to the measure \(\mu\) used in Section 2.1.
Let \( X' \) be transformed to \((R, Y_1', \ldots, Y_{n-1}') \equiv (R, Y)\) by the transformation
\[
\begin{align*}
(X'_1, \ldots, X'_{n}) &\leftrightarrow (R, X_2', \ldots, X_n'), \\
\text{that is } Y_i &= X_{i+1}', i = 1, \ldots, n-1.
\end{align*}
\]
Define
\[
\begin{align*}
b &= \sum_{i=1}^{n} c_i, \quad d = \sum_{i=1}^{n} c_i^2, \quad c' = (c_2', \ldots, c_n) \\
\mathbf{1}' &= (1, \ldots, 1) \quad X' = (Y_1', \ldots, Y_{n-1}') \\
I_{n-1} &= \text{the identity matrix of order } n-1.
\end{align*}
\]
\( Y \mid R = r \) is \((n-1)\) variate normal with expectation
\[
(2.13) \quad \mu \sim \mu + d^{-1} c (r - \mu) = d^{-1} r c + \mu (1 - d^{-1} b c)
\]
and covariance
\[
(2.14) \quad C = I_{n-1} - d^{-1} c c'.
\]
Hence the conditional density is
\[
(2.15) \quad g_{\mu}(y \mid r) = k e^{-\frac{1}{2} [(y - \mu)' c^{-1} (y - \mu)]},
\]
where \( k \) denotes a constant, and the derivative of the log is
\[
(2.16) \quad \frac{\partial \log g_{\mu}(y \mid r)}{\partial \mu} = -\left[ \frac{\partial}{\partial \mu} (y - \mu) \right]' c^{-1} (y - \mu) = (1 - d^{-1} bc)' c^{-1} (y - \mu).
\]
Since \((1 - d^{-1} bc)' c^{-1}\) is constant, \( \frac{\partial \log g_{\mu}(y \mid r)}{\partial \mu} \) is univariate normal with mean zero and variance \( V \), where
\[
(2.17) \quad V = (1 - d^{-1} bc)' c^{-1} (1 - d^{-1} bc).
\]
It follows that \( \frac{\partial \log g_\mu(y|x)}{\partial \mu} \)² is distributed as \( V \) times a \( \chi^2 \) variable with one degree of freedom. From (2.1), \( S = \frac{\partial \log g_\mu(y|x)}{\partial \mu} (\frac{\partial \phi}{\partial \theta})^{-1} \) and hence \( E_\mu(S^2|x) = V(\frac{\partial \phi}{\partial \theta})^{-2} \). Thus, if \( \hat{\phi} \) is any conditionally unbiased estimator of \( \phi \), the bound on the variance of \( \hat{\phi} \) as given by (2.2) is

\[
(2.18) \quad \text{Var} (\hat{\phi}|x) \leq \frac{(\frac{\partial \phi}{\partial \theta})^2}{V},
\]

where \( V \) is defined by (2.17) and (2.14).

Observe that \( V \) does not depend on \( r \), the observed value of \( R \), but only on the form of \( R \) as determined by the \( c_i, i = 1, \ldots, n \). Hence the bound is a function of \( r \), only if \( \frac{\partial \phi}{\partial \mu} \) is a function of \( r \).

Following are three simple examples to illustrate the bound.

**Example 1:** Let \( x' = (x_1, x_2, x_3) \) and let \( R(x) = x_1 + x_2 + 2x_3 \). Then \( b = \Sigma c_i = 4, d = \Sigma c_i^2 = 6, c' = (1, 2), \) and \( (1 - d^{-1}bc)' = \frac{1}{3}(1, -1) \).

Making the transformation as in (2.12), we have \( y_1 = x_2, y_2 = x_3 \).

From (2.13) and (2.14), we get the mean, covariance, and covariance inverse,\n\[
\nu = \frac{1}{6}x(1) + \frac{1}{3}y(-1),
\]
\[
c = I_2 - \frac{1}{6}c'c' = \frac{1}{6}(-2, -2),
\]
\[
c^{-1} = (2, 2).
\]

Using (2.17), \( V = (1 - d^{-1}bc)'c^{-1}(1 - d^{-1}bc) = \frac{1}{3} \). Then from (2.18)
\[
\text{Var} (\hat{\phi}|x) \leq 3(\frac{\partial \phi}{\partial \theta})^2.
\]

**Example 2:** Let \( x' = (x_1, x_2, x_3) \) and let \( R(x) = x_1 + x_2 \). Then \( b = 2, d = 2, c' = (1, 0), \) and \( (1 - d^{-1}bc)' = (0, 1) \). Again we use the transformation of (2.12). Then from (2.14),
\[
c = I_2 - \frac{1}{2}c'c' = \frac{1}{2}(1, 0)
\]

- 100 -
and
\[ c^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}. \]

Using (2.17), \( V = 1 \) and then from (2.18), the bound is
\[ \text{Var} (\hat{\varphi}|r) \geq \frac{1}{2} (\frac{3\varphi}{\varphi})^2. \]

**Example 3:** Let \( \mathbf{x}' = (x_1, x_2, x_3) \) and let \( R(x) = x_3 - x_1 \). Then \( b = 0, d = 2, c' = (0, 1), \) and \( (1 - d^{-1}bc)' = (1, 1). \) With the transformation of (2.12), we get from (2.13) and (2.14)
\[ v = \frac{1}{2} r(0, 1)' + \mu(1, 1)', \]
\[ C = I_2 - \frac{1}{2} (0, 1)' = \frac{1}{2} (2, 0, 1), \]
\[ c^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}. \]

Using (2.17), \( V = 3 \), and then from (2.18) the bound is
\[ \text{Var} (\hat{\varphi}|r) \geq \frac{1}{3} (\frac{3\varphi}{\varphi})^2. \]

It may be noted that examples 2 and 3 correspond to special conditions on the \( c_i \) in the expression \( R(x) = \sum c_i x_i \), which may be of interest. In example 3, \( \sum c_i = b = 0 \). In the case that \( b = 0, V = 1', C^{-1} = \text{sum of entries in } C^{-1}, \) and the bound is
\[ \text{Var} (\hat{\varphi}|r) \geq \frac{(\frac{3\varphi}{\varphi})^2}{\text{Sum of entries in } C^{-1}}. \]

Example 2 is an example of the special case, \( R(x) = \sum_{i=1}^{m} x_i, m < n. \) In this case, \( b = \sum c_i = m, d = \sum c_i^2 = m, \)
\[ c' = (1, \ldots, 1, 0, \ldots, 0) \]
\[ m-1 \quad n-m \]

and
\[
(1 - d^{-1}bc)' = (0, \ldots, 0, 1, \ldots, 1).
\]

Further, from (2.14),

\[
C = I_{n-1-m-1} \begin{pmatrix}
1 & \cdots & 1 \\
\vdots & \ddots & \vdots \\
1 & \cdots & 1 \\
0 & \cdots & 0
\end{pmatrix} = (M \circ O)I_{n-m},
\]

where

\[
M_{(m-1)(m-1)} = I_{m-1-m-1}^{-1} = (1, \ldots, 1).
\]

Then

\[
C^{-1} = \begin{pmatrix}
M^{-1} & 0 \\
0 & I_{n-m}
\end{pmatrix}
\]

and

\[
V = (1 - d^{-1}bc)'C^{-1}(1 - d^{-1}bc) = n - m.
\]

Hence from (2.18), for any \( \hat{\phi} \) which is a conditionally unbiased estimator of \( \phi \),

\[
(2.20) \quad \text{Var}(\hat{\phi}|r) \geq \frac{(\phi)^2}{n-m}.
\]

Note the relationship between this bound and the usual bound on the variance of an unbiased estimator of \( \mu \) based on a sample of size \( N \), which is \( \frac{1}{N} \). In the situation considered here, fixing \( R(x) = \sum_{i=1}^{m} x_i \) leaves \( n-m \) independent observations for estimating \( \mu \).

2.3 Bound for estimators of the conditional probability of misclassification in the univariate normal classification problem.

The situation for estimating the conditional probability of misclassification when independent test samples are available is a slightly generalized version of the situation discussed in the preceding section, where \( X' = (X_1, \ldots, X_n) \), \( X_i \overset{d}{=} N(\mu, 1) \) and \( R(x) = \sum_{i=1}^{m} x_i \), \( m < n \). Let

\[
(2.21) \quad \tilde{X}' = (x_{11}, \ldots, x_{1N_1}, x_{21}, \ldots, x_{2N_2}, t_{21}, \ldots, t_{2M_2}),
\]

where \( x_{i1} \overset{d}{=} N(\mu_1, 1), i = 1, \ldots, N_1 \) and \( x_{2j}, t_{2k} \overset{d}{=} N(\mu_2, 1) \),
j = 1, ..., N_2; k = 1, ..., M_2. The \( x_{1i}, x_{2j} \) would be the original samples and the \( t_{2k} \) would be the test sample observations from \( \Pi_2 \).

A test sample from \( \Pi_1 \) can be included in \( \sim \) but will have no effect on the results, because the quantity to be estimated (see (2.22) below) is a function of \( \mu_2 \) but not of \( \mu_1 \).

Let \( \bar{x}_1 = \frac{1}{N_1} \sum x_{1i}, \bar{x}_2 = \frac{1}{N_2} \sum x_{2j} \). Let \( R(x) = (R_1(x), R_2(x)) \) where \( R_1(x) = N_1 \bar{x}_1 \) and \( R_2(x) = N_2 \bar{x}_2 \). The quantity to be estimated is the conditional probability of misclassifying an observation from \( \Pi_2 \) as \( \Pi_1 \), using the rule (1.2) of Chapter 3.

Hence using (1.6), Chapter 3, we will take

\[
(2.22) \quad \varphi = \varphi(R_1, R_2, \mu_2) = F\left(\frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_2\right),
\]

where \( F \) is the distribution function of a standard normal variable.

The bound on the \( \text{Var}(\hat{\varphi}|r) \), where \( \hat{\varphi} \) is a conditionally unbiased estimator of \( \varphi \) will be derived by the general method of Section 2.1.

Let \( \tilde{x} \) be transformed as follows:

\[
(2.23) \quad \tilde{x} = (R_1(\bar{x}), R_2(\bar{x}), \bar{x}(\bar{x}))
\]

where, with \( n_1 \) denoting \( N_1 - 1 \) and \( n_2 \) denoting \( N_2 - 1 \),

\[
\bar{x}(\bar{x}) = (x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{2n_2}, t_{21}, \ldots, t_{2M_2}).
\]

Then \( \tilde{x}|R_1, R_2 \) is distributed as a \( (N_1 + N_2 + M_2 - 2) \) dimensional normal variable with mean \( \sim \) say, and covariance matrix \( C \). Let \( I_N \) denote the identity matrix of dimension \( N \) and let \( K_M \) be the \( M \times M \) matrix with all entries equal to 1. Let \( A = I_{N_1-1} - \frac{1}{N_1} K_{N_1-1} \) and let \( B = I_{N_2-1} - \frac{1}{N_2} K_{N_2-1} \). Then

\[
(2.24) \quad \tilde{x}' = \left(\frac{\bar{x}_1, \ldots, \bar{x}_1}{N_1-1}, \frac{\bar{x}_2, \ldots, \bar{x}_2}{N_2-1}, \frac{\mu_2, \ldots, \mu_2}{M_2}\right)
\]

and
Thus
\[ C = \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & I_{M_2} \end{pmatrix}. \]

Thus
\[ g_{\mu_2}(y|\Pi_1, \Pi_2) = K \exp\left[ -\frac{1}{2} \sum_{1}^{M_2} (t_{2i} - \mu_2)^2 \right] \]
where \( K \) is a constant with respect to \( \mu_2 \). Taking the log and then the derivative,
\[ \frac{\partial \log g_{\mu_2}(y|\Pi_1, \Pi_2)}{\partial \mu_2} = \frac{M_2}{1} (t_{2i} - \mu_2). \]

Following (2.1) we let
\[ s = \frac{\partial \log g_{\mu_2}(y|\Pi_1, \Pi_2)}{\partial \mu_2} (\frac{\partial \phi}{\partial \mu_2})^{-1}. \]

Then, because \( E(t_{2i} - \mu_2)^2 = 1 \) and \( E(t_{2i} - \mu_2)(t_{2j} - \mu_2) = 0, \ i \neq j \),
for conditional (and unconditional) expectations,
\[ E(s^2|\Pi_1, \Pi_2) = M_2 (\frac{\partial \phi}{\partial \mu_2})^{-2}. \]

Further, using (2.22)
\[ \frac{\partial \phi}{\partial \mu_2} = -f(\frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_2), \]
where \( f \) is the density of a standard normal variable. Hence, if
\( \hat{\phi} \) is any conditionally unbiased estimator of \( \phi \), the bound on the variance is, according to (2.2),
\[ \text{Var}(\hat{\phi}|\bar{x}_1, \bar{x}_2) \geq \frac{1}{M_2} f^2(\frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_2), \]
where \( M_2 \) is the size of the test sample from \( \Pi_2 \).

As noted at the beginning, this section applies to the problem of estimating the conditional probability of misclassification, only when test samples are available. This limitation is indicated by the definition of \( \bar{x} \) to include original samples and a test sample. When
only original samples are available, \((R_1 = N_1 \bar{x}_1, R_2 = N_2 \bar{x}_2)\) is sufficient for \((\mu_1, \mu_2)\) and the derivation of the bound by the methods of Section 2.1 cannot be done. The conditional distribution is not a function of the parameter \(\theta \equiv (\mu_1, \mu_2)\). Thus there is no conditionally unbiased estimator of \(\varphi = F(\frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_2)\) based only on original samples, because if the conditional distribution of the original samples given \(\bar{x}_1, \bar{x}_2\) is independent of \(\mu_2\), the conditional expectation of any function of the observations cannot be equal to \(\varphi\) which is a function of \(\mu_2\). Hence if \(\hat{\varphi}(x)\) is to be an unbiased estimator of \(\varphi\), it must depend on at least one test observation.

In Section 4.4, we derive the (conditional) UMVU estimator of the conditional probability of misclassification for the situation where test samples are available. Some numerical results are given there in Table XVII comparing the variance of the minimum variance estimator to the Cramér-Rao bound given by (2.29).

2.4 Application of the Cramér-Rao bound in the normal case with

\[ R = \sum c_i x_i \quad \text{and} \quad \varphi = k_1 R + k_2 \mu \quad \text{or} \quad \varphi = k_3 R \mu. \]

In Section 2.2 we considered the normal case with \(R = \sum c_i x_i\) and with \(\varphi\) arbitrary. Now we will consider special forms of \(\varphi\) to which the theory applies, in particular \(\varphi(r, \mu) = k_1 r + k_2 \mu\) and \(\varphi(r, \mu) = k_3 r \mu\), where \(k_1, k_2,\) and \(k_3\) are constants. In each case a conditionally unbiased minimum variance estimator of \(\varphi\) is obtained by putting \(\hat{\varphi}(x) = \varphi(r, \hat{\mu})\), where \(\hat{\mu}\) is the conditionally unbiased minimum variance estimator of \(\mu\) in the class of estimators that are linear combinations of the \(x_i\). The proof of this fact is given below. First, we find the desired linear conditionally unbiased minimum variance estimator \(\hat{\mu}\), and then we show that \(\text{Var} \varphi(r, \hat{\mu})\) achieves the Cramér-Rao bound for estimators of \(\varphi\). (The conditional unbiasedness of \(\varphi(r, \hat{\mu})\) is an obvious consequence of the conditional unbiasedness of \(\hat{\mu}\).)
Thus, as in Section 2.2, assume that $X = (X_1, \ldots, X_n)'$ where the $X_i$ are independent and identically distributed $N(\mu, 1)$. Let

$R = \Sigma c_i X_i, c_i \neq 0$ and not all $c_i$ equal. We want to find $\beta$ such that if $\hat{\mu} = \beta' X$, then $E(\hat{\mu} | R = r) = \mu$ and $\text{Var}(\hat{\mu} | r)$ is a minimum.

We will first show that the problem is simplified by considering linearly transformed variates. Let $Z = AX$, $|A| \neq 0$, be any non-singular linear transformation of $X$. Then $\hat{\mu}$ can be expressed also as a linear combination of the $Z_i$, namely $\hat{\mu} = (\beta' A^{-1}) Z = [(A^{-1})' \beta]' Z$.

Furthermore, letting $\hat{\mu}_X$ and $\hat{\mu}_Z$ denote $\hat{\mu}$ expressed, respectively, in terms of $X$ and in terms of $Z$, $E(\hat{\mu}_Z | r) = \beta' A^{-1} E(Z | r) = \beta' A^{-1} A E(X | r) = E(\hat{\mu}_X | r)$. Also $\text{Var}(\hat{\mu}_Z | r) = \beta' A^{-1} \text{Cov}(Z | r)(A^{-1})' \beta = \beta' A^{-1} A \text{Cov}(X | r) A'(A')^{-1} \beta = \text{Var}(\hat{\mu}_X | r)$. Hence we can find the conditional unbiased minimum variance linear estimator of $\mu$ by working with the original variable $X$ or by working with a variable which is a non-singular linear transformation of $X$.

Therefore let $X$ be transformed to $Y \equiv (R, Y_1, \ldots, Y_{n-1})' \equiv (R, Y)'$ with $Y_i = X_{i+1}, i = 1, \ldots, n-1$; that is $Y = AX$, where

$$(2.29a) \quad A = \begin{pmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ \mathbf{0} & \cdots & \cdots & \cdots \end{pmatrix}.$$

Note that this transformation is the same one used in Section 2.2.

The result on the conditionally unbiased minimum variance estimator of $\mu$ is given below in Lemma 2.3 essentially in terms of the $Y$ variable. The use of the transformed variable makes the statement of the lemma a bit awkward. On the other hand, because the mean and variance are conditional moments given $R(X) = r$, the conditions necessary for $\hat{\mu}$ to be unbiased and minimum variance are simpler when $\hat{\mu}$ is taken as a linear combination of $r$ and $Y_i$.

Lemma 2.3. Let $X = (X_1, \ldots, X_n)'$ with $X_i$ independent, identically distributed $N(\mu, 1)$. Let $R = \Sigma c_i X_i, c_i \neq 0$ and not all $c_i$ equal. Then

$$\text{c}$$
Let $A$ be as in (2.29a). . . Let $\mathbf{g}' = (\alpha_1, \ldots, \alpha_n)$ and $\mathbf{q}' = (\alpha_2, \ldots, \alpha_n)$, where the $\alpha_j$ are constants. Let $\beta' = g^A$. (Note, then, that $\beta'X = g'A$, when $A = AX$). Let $b = \Sigma c_i$, $d = \Sigma c_i^2$, $\mathbf{c}' = (c_2, \ldots, c_n)$, $\mathbf{1}' = (1, \ldots, 1)$, $a = \mathbf{1}' - \mathbf{d}'\mathbf{c}$ and $C = I_{n-1} - \mathbf{d}'\mathbf{c}'$. Then

1) $\beta'X$ is a conditionally unbiased estimator of $\mu$ given $R = r$ if (i) $\alpha_1 + d^{-1}\alpha'c = 0$ and (ii) $\alpha'a = 1$;

2) given conditions (i) and (ii), $\beta'X$ has minimum conditional variance given $R = r$ if $\alpha = \frac{c}{\mathbf{a}'C^{-1}\mathbf{a}}$;

3) and this minimum conditional variance is equal to $\frac{1}{\mathbf{a}'C^{-1}\mathbf{a}}$.

Proof: Let $\mathbf{y} = AX$. As indicated in the discussion preceding the statement of the lemma, the conditionally minimum variance unbiased estimator of $\mu$, i.e., $\hat{\mu}$, will be found with $\hat{\mu}$ expressed as a linear combination of $\mathbf{y} = (R, Y_1, \ldots, Y_{n-1})' \in (R, \mathbf{1})'$. Thus let $\hat{\mu} = \mathbf{a}^*\mathbf{y}$. By definition, $E(\mathbf{a}^*\mathbf{y}|r) = \alpha_1 r + \alpha' E(Y|r)$. Then using the result for $E(Y|r)$ from (2.13),

$$E(\mathbf{a}^*\mathbf{y}|r) = \alpha_1 r + \alpha'[d^{-1}\mathbf{c} + \mu\mathbf{a}]$$

$$= r(\alpha_1 + d^{-1}\alpha'c) + \mu\alpha'a.$$

This quantity is equal to $\mu$ if and only if the coefficient of $r$ is zero and the coefficient of $\mu$ is one, that is, $E(\mathbf{a}^*\mathbf{y}|r) = \mu$ if and only if $\alpha_1 + d^{-1}\alpha'c = 0$ and $\alpha'a = 1$. Thus statement 1) is proved.

Looking at the variance, by definition

$$\text{Var}(\mathbf{a}^*\mathbf{y}|r) = \text{Var}(\alpha_1 r + \alpha'Y|r)$$

$$= \alpha'Cov(Y, Y'|r) \alpha.$$
Thus, to get the minimum variance unbiased linear estimator we need to minimize $\alpha' C \alpha$ subject to the unbiasedness conditions: i) $\alpha_1 + d^{-1} \alpha' \widetilde{c} = 0$ and ii) $\alpha' [\widetilde{1} - d^{-1} bc] = 1$. However, the function to be minimized does not involve $\alpha_1$. Furthermore, using condition i), the $\alpha_1$ needed for unbiasedness can be determined in terms of $\alpha' = (\alpha_2, \ldots, \alpha_n)$, once an $\alpha$ which minimizes $\alpha' C \alpha$ subject to condition ii) has been found. The condition i) puts no constraint on the $\alpha$ needed to minimize $\alpha' C \alpha$. Therefore, the problem reduces to minimizing $\alpha' C \alpha$ subject to $\alpha' a = 1$. Using the method of Lagrange multipliers, the solution is found to be $\alpha = \frac{C^{-1} A}{\widetilde{V}}$, where as in (2.17) $\widetilde{V} = \alpha' C^{-1} \alpha = (\widetilde{1} - d^{-1} bc) C^{-1} (\widetilde{1} - d^{-1} bc)$. Then substituting this value for $\alpha$ in $\alpha' C \alpha$ the minimum variance is equal to $\widetilde{V}^{-1}$. Thus statements 2) and 3) are proved.

Using condition i), $\alpha_1 = -d^{-1} \alpha' \widetilde{c}$, and $\mu$ in terms of $\widetilde{r}$ and $Y$ is equal to $\hat{\mu} = \frac{\alpha' C^{-1}}{\widetilde{V}} [\frac{r}{d} \widetilde{c} + \widetilde{y}]$. $\mu$ can also be expressed as $\hat{\mu} = (\alpha_1, \alpha)' A \widetilde{x}$.

Theorem 2.4. Let $X = (X_1, \ldots, X_n)'$, where the $X_i$ are independent, identically distributed $N(\mu, 1)$. Let $R = \Sigma c_i X_i$, $c_1 \neq 0$, not all $c_i$ equal. Let $\varphi_1 = k_1 R + k_2 \mu$ and $\varphi_2 = k_3 R \mu$ where $k_1, k_2, k_3$ are constants. Let $\hat{\mu}$ denote the estimator of $\mu$ given by Lemma 2.3. Let $\hat{\varphi}_1 = k_1 \hat{r} + k_2 \hat{\mu}$ and let $\hat{\varphi}_2 = k_3 \hat{\mu}$. Then $\hat{\varphi}_1$, $i = 1, 2$ is the conditionally unbiased minimum variance estimator of $\varphi_i$, $i = 1, 2$, given $R = r$.

Proof: The conditional unbiasedness of $\hat{\varphi}_1$ and $\hat{\varphi}_2$ follow directly from the conditional unbiasedness of $\hat{\mu}$.

From part 3 of Lemma 2.3, the conditional variance of $\hat{\mu}$ is

$$\frac{1}{\alpha' C^{-1} \alpha} = \widetilde{V}^{-1}$$

and hence $\text{Var}(\hat{\varphi}_1 | r) = k_2 \widetilde{V}^{-1}$ and $\text{Var}(\hat{\varphi}_2 | r) = k_3 \widetilde{r} \widetilde{V}^{-1}$. 

- 108 -
From (2.18) the Cramér-Rao bound on the variance of \( \hat{\varphi}_1 \) is 
\[
\left( \frac{\partial \varphi_1}{\partial \mu} \right)^2 \sigma^2 = k_2 \sigma^2 \]
and the bound on the variance of \( \hat{\varphi}_2 \) is 
\[
\left( \frac{\partial \varphi_2}{\partial \mu} \right)^2 \sigma^2 = (k_3 \sigma)^2 \sigma^2 = k_3 \sigma^3.
\]
Thus both estimators satisfy the Cramér-Rao bound and hence are conditionally minimum variance unbiased estimators.

3. Chapman-Robbins Type Bound.

Chapman and Robbins (1951) obtained a lower bound for the variance of unbiased estimators of a given parameter. Their derivation is much like that leading to the Cramér-Rao bound, but involves use of differences instead of derivatives and does not require the regularity conditions needed for the Cramér-Rao derivation. In Section 3.1 below a Chapman-Robbins type bound is obtained for the conditional variance of estimators \( \hat{\varphi} \), where \( \hat{\varphi} \) is a conditionally unbiased estimator of \( \varphi = \varphi(R(x), \theta) \).

The derivation follows that of Chapman and Robbins (1951) with minor changes. As in the case of the Cramér-Rao bound (Section 2.1), the result is given in terms of the transformed variables \( (R, Y_1, \ldots, Y_{n-1}) \) instead of in terms of the basic variable \( X \). In Section 3.2 the bound is applied to a special normal situation.

3.1 Derivation of a Chapman-Robbins type bound.

Theorem 3.1. Let \( (R, Y), Y = (Y_1, \ldots, Y_{n-1}) \) be random variables with a joint distribution indexed by \( \theta \in \Theta \). Assume there exists a function \( g_\theta(y | r) \) such that \( g_\theta(y | r) \) is the conditional density of \( Y | R = r \) with respect to a measure \( \mu \). Assume that the range of the variables is independent of \( \theta \).

Let \( \theta, \theta + h, h \neq 0 \) be two distinct values in \( \Theta \). Define

\[
A = A(\varphi, \theta, h, r) = \frac{1}{h} \left[ \varphi(r, \theta + h) - \varphi(r, \theta) \right]
\]

and

This condition, which is one of the regularity conditions for the Cramér-Rao bound, is replaced by a less restrictive assumption in the Chapman-Robbins paper. The stronger assumption is used here to eliminate some of the detail in the derivation.
(3.2) \[ J = J(\varphi, \theta, h, r) = \frac{A^{-2}}{h^2} \left\{ \left[ \frac{g_{\theta+h}(y|x)}{g_\theta(y|x)} \right]^2 - 1 \right\} \]

Assume there exists a function \( \hat{\varphi} = \hat{\varphi}(x) = \hat{\varphi}(r, y) \) such that

(3.3) \[ \int \hat{\varphi}(r, y) g_\theta(y|r) \, d\mu(y) = \varphi(r, \theta), \text{ all } r \text{ and } \theta, \]
i.e., \( \hat{\varphi} \) is a conditionally unbiased estimator of \( \varphi \). Then for any such \( \hat{\varphi} \),

(3.4) \[ \text{Var}(\hat{\varphi}|r) \geq \frac{1}{\inf E[J(\varphi, \theta, h, r)|r]} \]
where the infimum is taken over the set of all \( h \neq 0 \) such that \( \theta, \theta + h \) are two distinct elements of \( \Theta \).

Proof: From the properties of a density,

(3.5) \[ \int g_\theta(y|x) \, d\mu(y) = 1, \text{ all } \theta \in \Theta. \]

Let \( U = [\hat{\varphi}(r, y) - \varphi(r, \theta)] \sqrt{g_\theta(y|x)} \) and let

\[ V = \left[ \frac{g_{\theta+h}(y|x) - g_\theta(y|x)}{h g_\theta(y|x)} \right] \sqrt{g_\theta(y|x)} A^{-1}. \]

Then using (3.3) and (3.5)

(3.6) \[ \int UV \, d\mu(y) = 1. \]

Hence by the Schwarz inequality,

(3.7) \[ 1 \leq \int U^2 \, d\mu(y) \int V^2 \, d\mu(y) = \text{Var}(\hat{\varphi}|r) \frac{A^{-2}}{h^2} \left\{ \left[ \frac{g_{\theta+h}(y|x)}{g_\theta(y|x)} \right]^2 g_\theta(y|x) \, d\mu(y) - 1 \right\}. \]

With \( J \) as defined in (3.2), (3.7) can be written as

(3.8) \[ \text{Var}(\hat{\varphi}|r) \geq \frac{1}{E[J(\varphi, \theta, h, r)|r]}. \]

Since (3.8) holds for all \( h \neq 0 \) such that \( \theta, \theta + h \) are distinct
elements of $\Theta$, we can take the infimum over this set of $h$'s to get the inequality of (3.4). Thus the theorem is proved.

Since the theorem is stated in terms of the transformed variable $(R, Y)$ instead of in terms of $X$, the question arises as it does in the derivation of the Cramér-Rao bound in Section 2, whether or not for a given $X$, the bound depends on the particular transformation used.

As in Theorem 2.2, Section 2.1, assume there are two transformations, $X \rightarrow (R, Y)$ and $X \rightarrow (R, Z)$ with two joint densities $M_\theta(r, y)$ and $N_\theta(r, z)$ respectively, and two conditional densities

$$g_\theta(y|r) = \frac{M_\theta(r, y)}{f_\theta(r)}, \quad h_\theta(z|r) = \frac{M_\theta(r, y(z))/(f_\theta(r))}{f_\theta(r)}$$

as given in (2.5).

From (3.1) and (3.2), we see that, for given $\varphi$, when $X \rightarrow (R, Y)$, $E(J|r)$ depends on

$$E\left\{ \left[ \frac{g_{\theta+h}(y|r)}{g_\theta(y|r)} \right]^2 \left| r \right. \right\} = E\left\{ \left[ \frac{M_{\theta+h}(r, y)}{M_\theta(r, y)} \frac{\varphi(r)}{f_{\theta+h}(r)} \right]^2 \left| r \right. \right\},$$

and when $X \rightarrow (R, Z)$, $E(J|r)$ depends on

$$E\left\{ \left[ \frac{M_{\theta+h}(r, y(z))}{M_\theta(r, y(z))} \frac{|J|}{f_{\theta+h}(r)} \right]^2 \left| r \right. \right\}.$$

These two expectations are equal and hence the bounds will be equal.

3.2 Chapman-Robbins type bound applied to a normal example.

Consider the example as discussed with respect to the Cramér-Rao bound in Section 2.2. In that example $X = (x_1, \ldots, x_n)$ with $x_i$ independent and identically distributed $N(\mu, 1)$; $R(x) = \Sigma c_i x_i$, $c_i \neq 0$, all $c_i$ not equal; $x_{n}^{1-1}(r, x_2, \ldots, x_n) \Xi (r, y)$; and

$$(3.9) \quad g_\mu(y|r) = \frac{1}{(2\pi)^{n-1}|C|^{1/2}} \exp\left\{-\frac{1}{2}(y-\varphi)'c^{-1}(y-\varphi)\right\}.$$
Letting \( a = (1 - d^{-1}bc) \), where \( (1 - d^{-1}bc) \) is defined in Section 2.2, we get the ratio of densities

\[
\frac{g_{\mu+h}(y|x)}{g_{\mu}(y|x)} = \exp\left\{ -\frac{1}{2} [-2h(y - \gamma)'c^{-1}\gamma + h^2a'c^{-1}a] \right\}.
\]

Then letting \( u = y - \gamma \)

\[
E\left[ \left( \frac{g_{\mu+h}(y|x)}{g_{\mu}(y|x)} \right)^2 \right] = (2\pi)^{-\frac{1}{2}} |c|^{-\frac{1}{2}} \int \exp\{2hu'c^{-1}\gamma - h^2a'c^{-1}a - \frac{1}{2}u'c^{-1}u\} du
\]

\[
= (2\pi)^{-\frac{1}{2}} |c|^{-\frac{1}{2}} \exp\{h^2a'c^{-1}a\}
\]

\[
\cdot \int \exp\{-\frac{1}{2}[(u-2ha)'c^{-1}(u-2ha)]\} du
\]

\[
= \exp\{h^2a'c^{-1}a\}.
\]

Note that \( a'c^{-1}a \) is the quantity defined in (2.17) to be \( V \); let it be denoted again by \( V \). Thus using the definition of \( J \) from (3.2),

\[
E(J|x) = \frac{A^{-2}}{h^2} (e^{h^2V} - 1),
\]

where \( A = \frac{1}{h}[\varphi(r, \mu+h) - \varphi(r, \mu)] \). Then given a particular \( \varphi \), for any \( \hat{\varphi} \) which is a conditionally unbiased estimator of \( \varphi \), we have

\[
\text{Var}(\hat{\varphi}|x) \geq \frac{1}{\inf_{h \neq 0} E(J|x)}.
\]

Below are examples for particular functions \( \varphi \).

**Example 1:** \( \varphi(R, \mu) = \alpha R + \beta \mu \) where \( \alpha, \beta \) are constants.

For this example, from (3.1), \( A = \frac{1}{h} [\alpha R + \beta(\mu + h) - \alpha R - \beta \mu] = \beta \).

From (3.12),

\[
\inf_{h \neq 0} E(J|x) = \inf_{h \neq 0} \beta^{-2} \frac{e^{h^2V} - 1}{h^2}
\]

\[
= \lim_{h \to 0} \beta^{-2} \frac{e^{h^2V} - 1}{h^2} = \beta^{-2} V,
\]

\[ -112 - \]
where $V = \tilde{a}'C^{-1}\tilde{a} = (1 - d^{-1}bc)'C^{-1}(1 - d^{-1}bc)$ as in (2.17). Then the Chapman-Robbins bound is

$$\text{Var}(\hat{r}) \geq \frac{\beta^2}{V},$$

and this bound is the same as the Cramér-Rao bound given by (2.18).

Example 2: $\varphi(R, \mu) = \alpha R^k \mu$. For this $\varphi$, $A = \alpha r^k$, $\inf E(\mathcal{J} | r) = (\alpha r^k)^{-2}V$ and $\text{Var}(\hat{r}) \geq \frac{\alpha^2 r^{2k}}{V}$, which is the same bound as given by the Cramér-Rao theory, (2.18).

Observe that in both of the above examples, $A$ is not a function of $h$, which simplifies the problem of finding the $\inf E(\mathcal{J} | r)$. Also, when $A$ is independent of $h$, finding the Chapman-Robbins bound in the conditional situation is more like the ordinary problem of finding the bound for the variance of an estimator of $\mu$ when the observations are normal. In this latter situation the Chapman-Robbins and the Cramér-Rao bounds are equal. See Chapman-Robbins (1951). Example 3, in which $\varphi$ is closely related to the conditional probability of misclassification in the normal situation, is a case where $A$ does depend on $h$.

Example 3: $\varphi(R, \mu) = F(R - \mu)$, where $F$ is standard normal distribution function.

Here $A = \frac{1}{h} [F(r - \mu - h) - F(r - \mu)]$ and

$$\inf E(\mathcal{J} | r) = \inf [F(r - \mu - h) - F(r - \mu)]^{-2}(e^{h^2V} - 1).$$

It can be shown that if $f^2(r - \mu) < V$ (where $f$ is the standard normal density), the inf is achieved in the limit as $h \to 0$ and equals $V/f^2(r - \mu)$. Otherwise the inf is smaller and occurs for $h \neq 0$.

4. Rao-Blackwell Type Theory.

In this section some Rao-Blackwell type theory related to the estimation of $\varphi(R, \theta)$ is developed and then is applied to get a
conditionally minimum variance unbiased estimator of the conditional probability of misclassification in the normal case (see Chapters 3 and 4).

Assume the situation described in Section 1, where \( X \) is a random variable with distribution \( P^X_\theta \) indexed by \( \theta \in \Theta \) and where \( R(x) \) is a function of \( X \). Further let \( T = T(x) \) be a function of \( X \). Let \( P^X|R_\theta \) and \( P^T|R_\theta \) denote the conditional distributions, respectively, of \( X \) and \( T \) given \( R \) fixed. Let \( E^X|R_\theta \) and \( E^T|R_\theta \) be the corresponding expectations. The concepts of sufficiency and complete sufficiency are defined as usual. In the following sub-sections, these concepts, and subsequently the Rao-Blackwell theory, are extended to the conditional situation. Then the conditional Rao-Blackwell theory is applied to the problem of estimating the conditional probability of misclassification in the multivariate (and univariate) normal case.

4.1 Definitions.

Definition 4.1: \( T \) is said to be **conditionally sufficient** for \( \theta \) given \( R \) if \( T \) is sufficient for \( \theta \) in the conditional distribution of \( X \) for each given value \( r \) of \( R \), i.e., if there exists a determination of the conditional distribution of \( X \) given \( R = r \) and \( T = t \) which is independent of \( \theta \) for all \( r, t \).

Thus in slightly different phrasing the definition is:

**Definition 4.1a:** \( T \) is **conditionally sufficient** for \( \theta \) given \( R \) if \((T, R)\) is sufficient for \( \theta \).

The definition of conditionally complete sufficient will be given for discrete and for absolutely continuous random variables. In each case assume that the ranges of \( X, T, \) and \( R \) do not depend on \( \theta \).

**Definition 4.2:** Let \( X \) be discrete. \( T \) is said to be **conditionally complete sufficient** given \( R \) if for each value \( r \) of \( R \)
\[ \sum_{t} f(t) \mathbb{P}_{\theta}(T = t \mid R = r) = 0 \text{ for all } \theta \in \Theta \Rightarrow \]

\[ f(t) = 0 \text{ for all } t \text{ such that } \mathbb{P}_{\theta}(T = t \mid R = r) > 0. \]

**Definition 4.3:** Let \( X \) be an absolutely continuous random variable. Let \( h_{\theta}(\cdot) \) be the marginal density of \( R \). Let \( g_{\theta}(\cdot \mid r) \) be the conditional density of \( T \) given \( R = r \), which is defined for \( r \) values such that \( h_{\theta}(r) > 0 \). \( T \) is said to be **conditionally complete sufficient given \( R \)** if for each value \( r \) of \( R \) such that \( h_{\theta}(r) > 0 \)

\[ \int f(t) g_{\theta}(t \mid r) dt = 0 \text{ for all } \theta \in \Theta \Rightarrow f(T) = 0 \text{ a.e. } \mathbb{P}_{\theta}^{T \mid R}. \]

### 4.2 Relation between sufficiency and conditional sufficiency, and between complete sufficiency and conditional complete sufficiency.

It is intuitively obvious from Definition 4.1a that if \( T \) is sufficient for \( \theta \), then \( T \) is conditionally sufficient given \( R \), because if all the "information" in the sample relative to \( \theta \) is contained in \( T \), it must be contained in \( (T, R) \). Hence \( (T, R) \) is sufficient. Bahadur (1954, p. 440) gives a general proof of the fact that \( T \) sufficient implies \( (T, R) \) is sufficient. A special proof for the discrete case is given below after the statement of the theorem.

**Theorem 4.1.** If \( T \) is sufficient for \( \theta \), then \( T \) is conditionally sufficient for \( \theta \) given \( R \).

**Proof for \( X \) discrete:**

For \( r, t \) such that \( \mathbb{P}_{\theta}(R = r, T = t) > 0 \), we have the conditional probability of \( X \) defined as

\[ P_{\theta}(X = x \mid R = r, T = t) = \frac{P_{\theta}(X = x, R = r, T = t)}{P_{\theta}(R = r, T = t)} = \frac{P_{\theta}(X = x, R = r \mid T = t)P_{\theta}(T = t)}{P_{\theta}(R = r \mid T = t)P_{\theta}(T = t)}. \]
By the sufficiency of $T$, both $P_\theta[X = x, R = r|T = t]$ and $P_\theta[R = r|T = t]$ are independent of $\theta$. Hence $P_\theta[X = x|R = r, T = t] = c(x, r, t)$ independent of $\theta$.

**Theorem 4.2.** Let the situation of Section 1 hold and let $T = T(x)$ be a function of $X$. Assume the ranges of $X$, $T$, and $R$ do not depend on $\theta$. Then $T$ complete sufficient implies $T$ is conditionally complete sufficient (see Definition 4.3).

**Proof:** The proof is given here for $X$ absolutely continuous and with the assumption that all the densities listed below exist. The same proof holds for discrete variables with the densities replaced by the corresponding probability functions and with integrals interpreted as summations.

Let the various densities which are needed in the proof be denoted by the following:

- $h_\theta(r, t)$ is the joint density of $R, T$
- $g_\theta(t|r)$ is the conditional density of $T$ given $R = r$
- $k_\theta(r|t)$ is the conditional density of $R$ given $T = t$
- $\ell_\theta(r)$ is the marginal density of $R$
- $m_\theta(t)$ is the marginal density of $T$.

Then, for $r$ values such that $\ell_\theta(r) > 0$,

\begin{equation}
E_\theta^T|R=r f(T) = \int f(t)g_\theta(t|r)dt = \frac{1}{\ell_\theta(r)} \int f(t)h_\theta(r, t)dt
\end{equation}

$$= \frac{1}{\ell_\theta(r)} \int f(t)k_\theta(r|t)m_\theta(t)dt.$$

By the sufficiency of $T$, $k_\theta(r|t)$ is independent of $\theta$ and can be written as $k(r|t)$. Then for $r$ values such that $\ell_\theta(r) > 0$,

$$E_\theta^T|R=r f(T) = 0 \text{ all } \theta \iff \int f(t)k(r|t)m_\theta(t)dt = 0 \text{ all } \theta.$$

Equivalently, letting $F(t) = f(t)k(r|t)$,
By $T$ complete sufficient

(4.3) $E_{\theta}^{T}f(T) = 0 \text{ all } \theta \Rightarrow F(T) = 0 \text{ a.e. } F_{\theta}^{T}$.

Since $k(r|t) > 0$ when $g_{\theta}(t|r) > 0$,

(4.4) $F(T) = 0 \text{ a.e. } F_{\theta}^{T} \Rightarrow f(T) = 0 \text{ a.e. } F_{\theta}^{T|R}$.

Hence following the chain of reasoning of (4.2), (4.3) and (4.4), we get that $E_{\theta}^{T|R}f(T) = 0 \text{ all } \theta$ implies $f(T) = 0 \text{ a.e. } F_{\theta}^{T|R}$ and therefore $T$ is conditionally complete sufficient.

4.3 Rao-Blackwell theory.

The usual Rao-Blackwell theory leading to UMVU estimators holds also in this conditional case. The situation is assumed to be that described in Section 1. The Rao-Blackwell theorems modified to fit the conditional case are stated below without the proofs which parallel those for the usual Rao-Blackwell theory, as given, for example, in Lehmann (1950).

**Theorem 4.3.** Let $\hat{\varphi}$ be any estimator for $\varphi$ such that $E_{\theta}(\hat{\varphi}|R=r) = \varphi(r, \theta)$. Let $T$ be conditionally sufficient for $\theta$ given $R$. Define $B(\varphi) = E_{\theta}(\hat{\varphi}|R, T)$. Then $B(\varphi)$ is also a conditionally unbiased estimator of $\varphi$ given $R$, and $\text{Var}(\hat{\varphi}|R) \geq \text{Var}(B(\varphi)|R)$.

**Lemma 4.4.** Let $\hat{\varphi}(x)$ and $\hat{\psi}(x)$ be two conditionally unbiased estimators of $\varphi(R, \theta)$ based on a conditionally complete sufficient statistic. Then $\hat{\varphi}(x) = \hat{\psi}(x)$ a.e. $F_{\theta}^{X|R}$.

**Theorem 4.5.** Let $T$ be conditionally complete sufficient for $\theta$. Let $\varphi(R, \theta)$ be any quantity for which a conditionally unbiased estimator given $R$ exists. Then $\varphi(R, \theta)$ has a unique (a.e.) conditionally UMVU estimator which is a function of $T$. 

- 117 -
The conditionally UMVU estimator can be obtained, as usual, by finding a conditionally unbiased estimator and "Blackwellizing" it.

4.4 Application of the theory to the multivariate normal classification problem. (See Chapter 4)

Let $\Pi_1$ and $\Pi_2$ be $p$-dimensional normal with unknown means $\mu_1$ and $\mu_2$, respectively, and with common known covariance matrix $\Sigma$.

Let

$$x = (x_{11}, \ldots, x_{1N_1}, x_{21}, \ldots, x_{2N_2}, t_{11}, \ldots, t_{1M_1}, t_{21}, \ldots, t_{2M_2})',$$

where $x_{1i}$, $t_{1k}$ are $N_p(\mu_1, \Sigma)$ and $x_{2j}$, $t_{2l}$ are $N_p(\mu_2, \Sigma)$, $(i = 1, \ldots, N_1; k = 1, \ldots, M_1; j = 1, \ldots, N_2; l = 1, \ldots, M_2)$. That is, the observations consist of original samples from $\Pi_1$, $\Pi_2$ and test samples from $\Pi_1$, $\Pi_2$. (Note that the estimator for $p_2$ derived below will be the same if the test sample from $\Pi_1$ is omitted.) Let

$$\bar{x}_1 = N_1^{-1} \sum_{i=1}^{N_1} x_{1i}, \quad \bar{x}_2 = N_2^{-1} \sum_{j=1}^{N_2} x_{2j}$$

and let $R = (R_1, R_2)$ with $R_1 = \bar{x}_1$, $R_2 = \bar{x}_2$. The classification rule is the one given in (1.1), Chapter 4, which is a function of $R_1$ and $R_2$.

The quantity $\varphi$ to be estimated here is the conditional probability of misclassifying an observation from $\Pi_2$ as $\Pi_1$ given the rule.

From (1.2), Chapter 4

$$(4.5) \quad \varphi = \varphi(R_1, R_2, \mu_2) = 1 - F(C)$$

with $C = \frac{3D}{2} + D^{-1}(\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}(\bar{x}_2 - \mu_2)$ and $D^2 = (\bar{x}_1 - \bar{x}_2)'\Sigma^{-1}(\bar{x}_1 - \bar{x}_2)$.

Note that in Chapter 4 this function $\varphi(R_1, R_2, \mu_2)$ is denoted by $p_2(R(\bar{x}_1, \bar{x}_2), \mu_2)$.

Let $T_1 = \frac{1}{M_1} \sum_{i=1}^{M_1} t_{1i}$, $T_2 = \frac{1}{M_2} \sum_{j=1}^{M_2} t_{2j}$, and let $T = (T_1, T_2)$. Then $T$ is conditionally complete sufficient for $(\mu_1^\prime, \mu_2^\prime)$ given $\bar{x}$, and the Rao-Blackwell theory can be applied to get the conditionally UMVU estimator based on $\bar{x}$. (Because $\varphi$ is not a function of $\mu_1^\prime$, the
test sample from \( \Pi_1 \) can be omitted and the Rao-Blackwell estimator can be based on \( T_{\Sigma_2} \), which is then conditionally complete sufficient for \( \mu_2 \) given \( R \). To apply the conditional Rao-Blackwell theory, a conditionally unbiased estimator of \( \varphi \) is needed. Let

\[
\hat{\varphi} = \begin{cases} 
1 & \text{if the rule (1.1), Chapter 4, classifies } t_{21} \text{ as } \Pi_1 \\
0 & \text{otherwise.}
\end{cases}
\]

Then the conditional expectation of \( \hat{\varphi} \) given \( R_1, R_2 \) is the conditional probability of misclassifying an observation from \( \Pi_2 \) as \( \Pi_1 \); in other words, \( \hat{\varphi} \) is a conditionally unbiased estimator of \( \varphi \).

Now let \( B(\hat{\varphi}) \) be the estimator obtained by "Blackwellizing" \( \hat{\varphi} \) with respect to the conditional distribution given \( \sim \) and \( \sim' \), i.e.,

\[
(4.6) \quad B(\hat{\varphi}) = E(\hat{\varphi} | R, T).
\]

Thus from the classification rule of (1.1), Chapter 4,

\[
(4.7) \quad B(\hat{\varphi}) = P\{(x_{11} - \bar{x}_{2})'\Sigma^{-1} t_{21} > 0 (x_{11} - \bar{x}_{2})'\Sigma^{-1}(\bar{x}_1 + \bar{x}_2) | \bar{x}_1, \bar{x}_2, \bar{t}_1, \bar{t}_2\}.
\]

Hence, to get \( B(\hat{\varphi}) \) the conditional distribution of \( t_{21} \) given \( \bar{x}_1, \bar{x}_2, \bar{t}_1, \bar{t}_2 \) is needed. Because the test sample from \( \Pi_2 \) is independent of the original samples and of the test sample from \( \Pi_1 \), it suffices to find the conditional distribution of \( t_{21} \) given \( \bar{x}_2 \). To get this conditional distribution, observe that \( (t_{21}, \bar{x}_2) \overset{d}{=} (\mu_2^*, \Sigma^*) \) where

\[
\Sigma^* = \begin{pmatrix} \Sigma & M_2^{-1} \Sigma \\ M_2^{-1} \Sigma & M_2^{-1} \Sigma \end{pmatrix}.
\]

Hence (Anderson (1958, p. 28)), \( t_{21} | \bar{x}_2 \overset{d}{=} N_p (\mu_2^*, \Sigma_{11.2}^*) \) where

\[
\mu_2^* = \mu_2 + (M_2^{-1} \Sigma)(M_2 \Sigma^{-1})(\bar{x}_2 - \mu_2) = \bar{x}_2
\]

and

\[
\Sigma_{11.2}^* = \Sigma - (M_2^{-1} \Sigma)(M_2 \Sigma^{-1})(M_2^{-1} \Sigma) = \Sigma - \frac{M_2 - 1}{M_2}.
\]
Then $(\overline{x}_1 - \overline{x}_2)' \Sigma^{-1} \overline{x}_2 | \overline{x}_1, \overline{x}_2, \overline{\xi}_1, \overline{\xi}_2$ is univariate normal with expectation $(\overline{x}_1 - \overline{x}_2)' \Sigma^{-1} \overline{\xi}_2$ and variance $\frac{M_2 - 1}{M_2} D^2$, where 

\[ D^2 = (\overline{x}_1 - \overline{x}_2)' \Sigma^{-1}(\overline{x}_1 - \overline{x}_2). \]

Hence letting $z$ denote a standard univariate normal variable and $F(u)$ denote the standard normal distribution function, and using (4.7),

\[ (4.8) \quad B(\varphi) = P\{z > \frac{1}{\sqrt{2}}(\overline{x}_2 - 1)^{1/2} \left[ \frac{1}{2} D + F^{-1}(\overline{x}_1 - \overline{x}_2)' \Sigma^{-1}(\overline{x}_2 - \overline{\xi}_2) \right] \} \]

\[ = 1 - F((1 - M_2^{-1})^{1/2} \left[ \frac{1}{2} D + F^{-1}(\overline{x}_1 - \overline{\xi}_2)' \Sigma^{-1}(\overline{x}_2 - \overline{\xi}_2) \right] ). \]

Setting $p = 1$ and $\Sigma = \sigma^2 = 1$, we find that $D^2$ reduces to $(\overline{x}_1 - \overline{x}_2)^2$ and $D$, which is taken positive, is then equal to $|d|$, where $d$ denotes $\overline{x}_1 - \overline{x}_2$. Hence in the univariate normal case considered in Chapter 3

\[ B(\varphi) = 1 - F(M_2^{-1/2}(1 - 1 - 1)^{1/2} [1/2 |d| + d/|d| (\overline{x}_2 - \overline{\xi}_2)] ) \]

or

\[ (4.9) \quad B(\varphi) = \begin{cases} 
1 - F(\frac{1}{2} d + \overline{x}_2 - \overline{\xi}_2)(1 - M_2^{-1})^{1/2} & , \quad d > 0 \\
F(\frac{1}{2} d + \overline{x}_2 - \overline{\xi}_2)(1 - M_2^{-1})^{1/2} & , \quad d < 0.
\end{cases} \]

$B(\varphi)$, as an estimator for $P_2(R(\overline{x}_1, \overline{x}_2), \mu_e)$, the conditional probability of misclassification, is studied in Chapter 4 in the multivariate normal case and in Chapter 3 in the univariate normal case. (In those chapters it is denoted by $\rho_{RB}$. The conditional mean and variance of the estimator are computed in Appendix II.) By the Rao-Blackwell type theory given above, $B(\varphi)$ is the conditional UMVU estimator of the conditional probability of misclassifying an observation from $\pi_2$ as $\pi_1$ when observations available consist of original samples and test samples. (Note that as in the derivation of the Cramer-Rao bound (Section 2.3), the test sample from $\pi_1$ can be
put in or left out without changing the results.) Since the Rao-Blackwell theory can be applied to the problem of estimating $P_2$ when a test sample from $\pi_2$ is available, the Cramér-Rao bound is not of so much interest. Also, the Rao-Blackwell estimator was derived for the multivariate case, while the Cramér-Rao bound applies only to the univariate problem, since $\theta = \mu_2$ must be a scalar. For the univariate case, the variance of $\hat{B}(\phi)$ can, however, be compared to the Cramér-Rao bound to see how tight the bound is.

Assume $\pi_1$ is univariate $N(\mu_1, 1)$ and $\pi_2$ is univariate $N(\mu_2, 1)$. Let $F(u), F(u, v; \rho)$ be univariate and bivariate normal distribution functions. Let $H(u) = F(-u), H(u, v; \rho) = F(-u, -v; \rho)$ and, $f(u) = F'(u)$. Then by (2.29), if $\hat{\phi}$ is any (conditional) unbiased estimator of $P_2$, $\text{Var}(\hat{\phi} | x_1, x_2) \geq \sigma_2^{-1} f^2(c)$, where $c = \frac{1}{2}(x_1 + x_2) - \mu_2$.

From Table IV, the conditional variance of $\hat{\phi} \equiv \rho_{RB}$ is $F(c, c; \sigma_2^{-1}) - F(c, c; 0)$. Values for the bound and for the conditional variance of $\rho_{RB}$ are given in Table XVII for several values of $c$ and $\sigma_2$. For further comparison, values of the conditional variance of the test sample estimator, $P_T$, (see Chapter 2, Section 1.2), are also given, where from Table IV, $\text{Var} P_T = \sigma_2^{-1} F(c)[1 - F(c)]$. Since $f^2(c) = f^2(-c), F(c, c; \rho) - F(c, c; 0) = F(-c, -c; \rho) - F(-c, -c; 0)$, and $F(c)[1 - F(c)] = F(-c)[1 - F(-c)]$, only positive $c$ values are included in the table. The bivariate values were obtained from the tables put out by the Bureau of Standards (Tables of the Bivariate Normal Distribution Function and Related Functions (1959)).

Looking at Table XVII, we see that the values of the conditional variance of $\rho_{RB}$ are close to the values of the Cramér-Rao bound. The values of $\text{Var} P_T$ are, of course, larger than the values of $\text{Var} \rho_{RB}$ because $P_T$ is an (conditional) unbiased estimator based on an independent test sample from $\pi_2$ and $\rho_{RB}$ is the minimum variance estimator in this class of estimators.
Table XVII

Numerical Values for: i) the Cramer-Rao Bound on the Conditional Variance of Estimators of
\[ P_2(R(\bar{x}_1, \bar{x}_2), \mu_2) , \]
i) the Conditional Variance of the (Conditional) UMVU Estimator, \( \hat{\theta}_{RB} \) and
iii) the Conditional Variance of the Test Sample Estimator, \( P_T \).

\[ \pi_1 \text{ is } N(\mu_1, 1), \pi_2 \text{ is } N(\mu_2, 1). \text{ C-R Bound } = M_2^{-1}f^2(c). \]

\[ \text{Var} \hat{\theta}_{RB} = F(c, c; \rho) - F(c, c; 0), \rho = M_2^{-1}. \]
\[ c = \frac{1}{\bar{x}_1} + \frac{1}{\bar{x}_2} - \mu_2. \]

\[ \text{Var} P_T = M_2^{-1}f(c)[1 - F(c)]. \]

<table>
<thead>
<tr>
<th>c</th>
<th>Var ( P_T )</th>
<th>Var ( \hat{\theta}_{RB} )</th>
<th>C-R Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M_2 = 5, \rho = .20 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>.0500</td>
<td>.0320</td>
<td>.0320</td>
</tr>
<tr>
<td>0.5</td>
<td>.0427</td>
<td>.0255</td>
<td>.0248</td>
</tr>
<tr>
<td>1.0</td>
<td>.0267</td>
<td>.0129</td>
<td>.0117</td>
</tr>
<tr>
<td>1.5</td>
<td>.0125</td>
<td>.0041</td>
<td>.0034</td>
</tr>
<tr>
<td>2.0</td>
<td>.0044</td>
<td>.0009</td>
<td>.0006</td>
</tr>
<tr>
<td></td>
<td>( M_2 = 10, \rho = .10 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>.0250</td>
<td>.0159</td>
<td>.0159</td>
</tr>
<tr>
<td>0.5</td>
<td>.0213</td>
<td>.0126</td>
<td>.0124</td>
</tr>
<tr>
<td>1.0</td>
<td>.0134</td>
<td>.0061</td>
<td>.0059</td>
</tr>
<tr>
<td>1.5</td>
<td>.0062</td>
<td>.0018</td>
<td>.0017</td>
</tr>
<tr>
<td>2.0</td>
<td>.0022</td>
<td>.0004</td>
<td>.0003</td>
</tr>
<tr>
<td></td>
<td>( M_2 = 20, \rho = .05 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>.0125</td>
<td>.0080</td>
<td>.0080</td>
</tr>
<tr>
<td>0.5</td>
<td>.0107</td>
<td>.0062</td>
<td>.0062</td>
</tr>
<tr>
<td>1.0</td>
<td>.0067</td>
<td>.0030</td>
<td>.0029</td>
</tr>
<tr>
<td>1.5</td>
<td>.0031</td>
<td>.0009</td>
<td>.0008</td>
</tr>
<tr>
<td>2.0</td>
<td>.0011</td>
<td>.0002</td>
<td>.0001</td>
</tr>
</tbody>
</table>
APPENDIX I

DERIVATIONS TO OBTAIN THE ESTIMATORS OF $P_2(R(x_1, x_2), \mu_2)$

... THAT INVOLVE A PRIOR DISTRIBUTION ON $\theta$ ...

POPULATIONS: $\Pi_1$ is $N(\mu_1, 1)$ and $\Pi_2$ is $N(\mu_2, 1)$, $\theta = (\mu_1, \mu_2)$.

A. The Posterior Distribution of $(\mu_1, \mu_2)$.

If we assume $(\mu_1, \mu_2)$ have independent prior distributions $N(\gamma_1, \sigma_1^2)$ and $N(\gamma_2, \sigma_2^2)$, then the posterior distributions of $(\mu_1, \mu_2)$ given original samples $x_{11}, \ldots, x_{1N_1}$ from $\Pi_1$ and $x_{21}, \ldots, x_{2N_2}$ from $\Pi_2$ will be independent with $\mu_1 \sim N(\gamma_1, \tau_1^2)$ and $\mu_2 \sim N(\gamma_2, \tau_2^2)$, where

$$\gamma_1 = \frac{\gamma_1 + N_1\sigma_1^2 x_{1i}}{1 + N_1\sigma_1^2}, \quad \tau_1^2 = \frac{\sigma_1^2}{1 + N_1\sigma_1^2},$$

$$\bar{x}_i = N^{-1}_i \sum_{j=1}^{N_i} x_{ij}, \quad i = 1, 2.$$

B. Derivations for Method 3 and Method 7.

The estimator of Method 3 (Chapter 2, Section 2.6) is found by integrating $P_2(R(x_1, x_2), \mu_2)$ with respect to the posterior distribution of $(\mu_1, \mu_2)$, or equivalently with respect to the posterior distribution of $\mu_2$ only, since $\mu_1$ and $\mu_2$ are independent in the posterior distribution and $P_2$ is not a function of $\mu_1$.

Let $F(t), f(t)$ denote the standard normal distribution and density functions, respectively, and let PR-3, here denote the estimator derived by Method 3. Using the formula for $P_2$ from (1.6), Chapter 3, and the posterior distribution of $\mu_2$ given in Section A above, and letting $d = \bar{x}_1 - \bar{x}_2$,

$$\begin{cases} 
(2\pi)^{-\frac{1}{2}} \tau_2^{-1} \int_{-\infty}^{\infty} [1-F(\frac{1}{2}d + \bar{x}_2 - \mu_2)] \exp[-\frac{1}{2}\tau_2^{-2}(\mu_2 - \gamma_2)^2] d\mu_2, & d > 0 \\
(2\pi)^{-\frac{1}{2}} \tau_2^{-1} \int_{-\infty}^{\infty} F(\frac{1}{2}d + \bar{x}_2 - \mu_2) \exp[-\frac{1}{2}\tau_2^{-2}(\mu_2 - \gamma_2)^2] d\mu_2, & d < 0.
\end{cases}$$

(B.1) PR-3 =
Letting $z = (\mu_2 - \nu_2)/\tau_2$,

$$
(2\pi)^{-\frac{1}{2}} \tau_2^{-1} \int_{-\infty}^{\infty} F\left(\frac{1}{2}d + \bar{x}_2 - \mu_2\right) \exp\left\{-\frac{1}{2}\tau_2^2 (\mu_2 - \nu_2)^2\right\} \, d\mu_2
$$

$$
= \int_{-\infty}^{\infty} F\left(-\tau_2 z + \frac{1}{2}d + \bar{x}_2 - \nu_2\right) \, f(z) \, dz
$$

$$
= F\left(\frac{1}{2}d + \bar{x}_2 - \nu_2\right) (1 + \tau_2^2)^{-\frac{1}{2}}.
$$

Hence, using the integration result of (B.2) in (B.1),

$$
\text{PR-3} = \begin{cases} 
1 - F\left(\frac{1}{2}d + \bar{x}_2 - \nu_2\right)(1 + \tau_2^2)^{-\frac{1}{2}} & , \quad d > 0 \\
F\left(\frac{1}{2}d + \bar{x}_2 - \nu_2\right)(1 + \tau_2^2)^{-\frac{1}{2}} & , \quad d < 0.
\end{cases}
$$

The estimator of Method 7 (Chapter 2, Section 2.6) is equal to the expectation of $P_2(R(\bar{x}_1, \bar{x}_2), \mu_2)$ with respect to the prior distribution of $\mu_1, \mu_2$. Since, according to Section A, in both prior and posterior distributions, $\mu_1$ and $\mu_2$ are independent normal with the distributions differing only in mean and variance, the expression for the Method 7 estimator, PR-7, is the same as the expression in (B.3) with the prior mean and variance of $\mu_2$, which are $\gamma_2$ and $\sigma_2^2$, substituted for the posterior mean and variance, $\nu_2$ and $\tau_2^2$. Thus

$$
\text{PR-7} = \begin{cases} 
1 - F\left(\frac{1}{2}d + \bar{x}_2 - \gamma_2\right)(1 + \sigma_2^2)^{-\frac{1}{2}} & , \quad d > 0 \\
F\left(\frac{1}{2}d + \bar{x}_2 - \gamma_2\right)(1 + \sigma_2^2)^{-\frac{1}{2}} & , \quad d < 0.
\end{cases}
$$

C. Derivations for Method 5 and Method 8.

Let $f_{PR}(\mu_1, \mu_2, x, \bar{x}_1, \bar{x}_2)$ denote the joint density of $(\mu_1, \mu_2, x, \bar{x}_1, \bar{x}_2)$ obtained by using the prior density of $(\mu_1, \mu_2)$; let $f_{P0}(\mu_1, \mu_2, x, \bar{x}_1, \bar{x}_2)$ denote the corresponding joint density based on the posterior distribution of $(\mu_1, \mu_2)$. The observation $x$ is
assumed to be from \( \pi_2 \). Let \( f_5 \) and \( f_8 \) denote, respectively, the conditional densities of \( x \) associated with Method 5 and Method 8 (see Chapter 2, Section 2.6). Then

\[
(f.1) \quad f_5 = \frac{\iiint f_{P0}(\mu_1, \mu_2, x, \overline{x}_1, \overline{x}_2) d\mu_1 d\mu_2}{\iiint f_{P0}(\mu_1, \mu_2, x, \overline{x}_1, \overline{x}_2) d\mu_1 d\mu_2 dx}
\]

and \( f_8 \) is the same with \( f_{PR} \) substituted for \( f_{P0} \).

By straightforward calculations we find that the conditional distribution of \( x \) associated with Method 5 (density \( f_5 \)) is

\[
N\left( \frac{\nu_2 + N_2 \tau_2 \overline{x}_2}{1 + N_2 \tau_2^2}, \frac{1 + \tau_2^2 + N_2 \tau_2^2}{1 + N_2 \tau_2^2} \right).
\]

The conditional distribution of \( x \) associated with Method 8 (density \( f_8 \)), is the same except that \( \nu_2 \) and \( \tau_2^2 \) are substituted for \( \nu_2 \) and \( \tau_2^2 \), that is the distribution is \( N(\nu_2, 1 + \tau_2^2) \).

Finally, the Method 5 and Method 8 estimators, PR-5 and PR-8, are found by integrating, respectively, the densities \( f_5 \) and \( f_8 \) over the region where \( x \) is classified as \( \pi_1 \). Thus PR-5 and PR-8 are equal to

\[
P\left\{ z > \frac{1}{2}(\overline{x}_1 + \overline{x}_2) - \frac{\mu_x}{\sigma_x} \right\}, \quad d > 0
\]

\[
P\left\{ z < \frac{1}{2}(\overline{x}_1 + \overline{x}_2) - \frac{\mu_x}{\sigma_x} \right\}, \quad d < 0,
\]

where \( z \) is \( N(0, 1) \) and where for PR-5

\[
\mu_x = \frac{\nu_2 + N_2 \overline{x}_2 \tau_2^2}{1 + N_2 \tau_2^2}, \quad \sigma_x^2 = \frac{1 + N_2 \tau_2^2 + \tau_2^2}{1 + N_2 \tau_2^2},
\]

and for PR-8, \( \mu_x = \nu_2 \) and \( \sigma_x^2 = 1 + \tau_2^2 \).
APPENDIX II

COMPUTATIONS FOR THE CONDITIONAL MEANS AND VARIANCES
OF ESTIMATORS OF THE CONDITIONAL PROBABILITY
OF MISCLASSIFICATION.

The computations for the conditional means and variances are generally
given in detail for the case in which \( \pi_1 \) and \( \pi_2 \) are p-dimensional
normal populations with unknown mean vectors \( \mu_1, \mu_2 \) and common known
covariance matrix \( \Sigma \). Then the results for \( \pi_1 \) and \( \pi_2 \) univariate
normal with common variance one are obtained as a special case. The
conditional mean and variance of \( \pi_U \) and the conditional mean of \( \pi_{TSL} \)
(Assumption Set 1), however, are derived only in the univariate case.

The notation pertaining to the multivariate normal case is summarized
in Table X and the notation for the univariate normal case in Table I.
Since all of the computations are for conditional moments, however,
the expectation with \( \bar{x}_1, \bar{x}_2 \) fixed, test samples random, will usually
be denoted simply by \( E(\cdot) \), instead of by the symbol \( E(\cdot|\bar{x}_1, \bar{x}_2) \) that
is given in Table I.

The estimators \( Q_0, Q^*_0 \) (studied only in univariate case), \( Q^*_0 \)
(multivariate only), \( Q_{OS}^{**} \) (multivariate only), \( Q_{DS}^{**}, P_3 \), and \( P_5 \) are
constant when \( \bar{x}_1, \bar{x}_2 \) are fixed and thus the conditional expectation of
each of these estimators is equal to the constant value and the conditional
variance is zero.

A. The Reclassification Estimator (denoted \( p_0 \) or \( P_R \)).

The reclassification estimator is equal to \( p_0 \), the proportion
of the original sample of size \( N_2 \) from \( \pi_2 \) misclassified by the
rule \( R(\bar{x}_1, \bar{x}_2) \), given in (1.1), Chapter 4. To compute the conditional
moments of \( p_0 \), let \( x_{2i} \) \( (i \neq 1, \ldots, N_2) \) be the observations in
the original sample from \( \pi_2 \), with \( \bar{x}_2 = N_2^{-1}\Sigma x_{2i} \) (all summations in
this section are for \( i = 1, \ldots, N_2 \)). Let

- 126 -
and let

\[ t_1 = \begin{cases} 
1 & \text{if } \bar{x}_{21} \text{ is misclassified by rule } R(\bar{x}_1, \bar{x}_2) \\
0 & \text{otherwise}
\end{cases} \]

Then

\[ Q_1 = Q_1(\bar{x}_1, \bar{x}_2) = P\{x_{21} \text{ misclassified} | \bar{x}_1, \bar{x}_2 \text{ fixed}\} \]

\[ Q_{1,2} = Q_{1,2}(\bar{x}_1, \bar{x}_2) = P\{x_{21} \text{ and } x_{22} \text{ misclassified} | \bar{x}_1, \bar{x}_2 \text{ fixed}\}. \]

Hence

\[ E_{t_1} = E_{t_2} = Q_1(\bar{x}_1, \bar{x}_2), \quad E_{t_1 t_2} = Q_{1,2}(\bar{x}_1, \bar{x}_2), \quad \text{and} \quad \Sigma_{t_1} = N_2 p_0. \]

Because the \( t_1 \) are identically distributed, the variance formula reduces to

\[ \text{Var} (N_2 p_0) = N_2 \text{Var} t_1 + N_2 (N_2 - 1) \text{Cov}(t_1, t_2). \]

Substituting \( Q \)'s and dividing by \( N_2^2 \),

\[ \text{Var} \ p_0 = N_2^{-1} (Q_1 - Q_1^2) + \frac{N_2 - 1}{N_2} (Q_{1,2} - Q_1^2). \]

In order to compute \( Q_{1,2} \) and \( Q_1 \), and then \( E p_0 \) and \( \text{Var} p_0 \), two distributions are needed—the joint conditional distribution of \( x_{21}, x_{22} \) given \( x_1, x_2 \) and the conditional distribution of \( x_{21} \) given \( x_1, x_2 \). To get the conditional distribution of \( x_{21}, x_{22} \) given \( x_1, x_2 \) observe that \( (x_{21}, x_{22}, \bar{x}_1, \bar{x}_2)' \stackrel{d}{=} N(\mu, \Sigma^*) \), where

\[ \mu = \begin{pmatrix} \mu_2 \\ \mu_2 \\ \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma^* = \begin{pmatrix} \Sigma & 0 & 0 & N_1^{-1} \\ 0 & \Sigma & 0 & 0 \\ 0 & 0 & \Sigma N_2^{-1} \Sigma^* & 0 \\ N_2^{-1} \Sigma^* & 0 & 0 & N_2^{-1} \Sigma \end{pmatrix} \text{def} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \]

\[ ^1 \text{Recall that } E(\cdot) \text{ denotes expectation with } \bar{x}_1, \bar{x}_2 \text{ fixed, test samples random.} \]
The division of $\Sigma^*$ into the four submatrices $\Sigma_{11}$, $\Sigma_{12}$, $\Sigma_{21}$, $\Sigma_{22}$ is for purposes of deriving the conditional distribution. Note that all of the 16 smaller submatrices of $\Sigma^*$ are $p$ by $p$. Then (Anderson (1958), p. 28) the desired conditional distribution is

$$
(A.3) \quad \left( \begin{array}{c|c}
\bar{x}_{21} \\
\bar{x}_{22}
\end{array} \right) \overset{d}{=} N(\mu_c, \Sigma_{11,2})
$$

where

$$
(A.4) \quad \mu_c := \left( \begin{array}{c}
\bar{\mu}_1 \\
\bar{\mu}_2
\end{array} \right) + \Sigma_{12} \Sigma_{22}^{-1} \left( \begin{array}{c}
\bar{x}_1 - \mu_1 \\
\bar{x}_2 - \mu_2
\end{array} \right) = \left( \begin{array}{c}
\bar{x}_2 \\
\bar{x}_2
\end{array} \right)
$$

and

$$
(A.5) \quad \Sigma_{11,2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \left( \begin{array}{cc}
\Sigma & 0 \\
0 & \Sigma
\end{array} \right) - N_2^{-1} \left( \begin{array}{cc}
\Sigma & \Sigma \\
\Sigma & \Sigma
\end{array} \right) = \left( \begin{array}{cc}
\frac{N_2 - 1}{N_2} \Sigma & -\frac{1}{N_2} \Sigma \\
-\frac{1}{N_2} \Sigma & \frac{N_2 - 1}{N_2} \Sigma
\end{array} \right).
$$

Similarly, the conditional distribution of $x_{21}$ given $x_1$, $x_2$ is found and

$$
(A.6) \quad (x_{21}|x_1, x_2) \overset{d}{=} N(x_2, \frac{N_2 - 1}{N_2} \Sigma).
$$

Now, let $A = \frac{1}{2}((x_1 - x_2)'\Sigma^{-1}(x_1 + x_2))$. Then by the definition of $Q_1$ and by (1.1), Chapter 4, $Q_1 = P\{y > A|x_1, x_2\}$, where $y = (x_1 - x_2)'\Sigma^{-1}x_{21}$. Using (A.6) we find $y$ is normal with

$$
E(y|x_1, x_2) = (x_1 - x_2)'\Sigma^{-1}x_2
$$

and

$$
\text{Var}(y|x_1, x_2) = \frac{N_2 - 1}{N_2} D^2,
$$

where $D^2$ is defined in Table X. Hence, if $z$ denotes a standard normal variate,

$$
(A.7) \quad Q_1 = P\{z > [A - (x_1 - x_2)'\Sigma^{-1}x_2](N_2 - 1)^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} D^{-1}]\}
$$

$$
= P\{z > B\} = 1 - F(B),
$$

- 128 -
where \( B \) is defined in Table X.

Similarly by the definition of \( Q_{1,2} \) and by (1.1), Chapter 4

\[
Q_{1,2} = P[y_1 > A \text{ and } y_2 > A | \bar{x}_1, \bar{x}_2],
\]

where \( y_1 = (\bar{x}_1 - \bar{x}_2)' \Sigma^{-1} \bar{x}_2, \quad y_2 = (\bar{x}_1 - \bar{x}_2)' \Sigma^{-1} \bar{x}_2'. \) Using (A.3), (A.4), and (A.5) we find that

\[
E(y_1 | \bar{x}_1, \bar{x}_2) = E(y_2 | \bar{x}_1, \bar{x}_2) = (\bar{x}_1 - \bar{x}_2)' \Sigma^{-1} \bar{x}_2;
\]

\[
\text{Var}(y_1 | \bar{x}_1, \bar{x}_2) = \text{Var}(y_2 | \bar{x}_1, \bar{x}_2) = \frac{N_2 - 1}{N_2} D^2;
\]

\[
\text{Cov}(y_1, y_2 | \bar{x}_1, \bar{x}_2) = - \frac{N_2 - 1}{N_2} D^2;
\]

\[
\rho(y_1, y_2 | \bar{x}_1, \bar{x}_2) = - (\frac{N_2 - 1}{N_2})^{-1}.
\]

Thus letting \( z_1 \) and \( z_2 \) denote standard normal variables with

\[
\rho(z_1, z_2) = - \frac{1}{\frac{N_2 - 1}{N_2}};
\]

(A.8) \( Q_{1,2} = P[z_1 > B, \; z_2 > B] = F(-B, -B; - \frac{1}{\frac{N_2 - 1}{N_2}}) \)

(see Table X for notation).

Now the conditional mean and variance of \( p_0 \) can be computed.

From (A.1) and (A.7)

(A.9) \( E p_0 = 1 - F(B). \)

Using (A.2), (A.7) and (A.8)

(A.10) \( \text{Var} p_0 = N_2^{-1} [1 \div F(B)] - [1 - F(B)]^2 + \frac{N_2 - 1}{N_2} F(-B, -B; - \frac{1}{\frac{N_2 - 1}{N_2}}). \)

For the special case \( p = 1 \) we find \( B = |b|; \) \( F(B) = F(b) \) when \( d \) and hence \( b \) is positive, and \( F(B) = F(-b) \) when \( d \) and \( b \) are negative. (See Tables I and X for notation.) Therefore for univariate normal populations the conditional moments of \( p_0 \) given \( \bar{x}_1, \bar{x}_2 \) fixed are
\[(A.11) \quad \text{Ep}_0 = \begin{cases} 
1 - F(b), & d > 0 \\
F(b), & d < 0 
\end{cases} = F(-|b|), \]

\[(A.12) \quad \text{Var p}_0 = \begin{cases} 
N_2^{-1}[1-F(b)] - [1-F(b)]^2 + \frac{N_2-1}{N_2} F(-b, -b; -\frac{1}{N_2-1}), & d > 0 \\
N_2^{-1}F(b) - \mu^2(b) + \frac{N_2-1}{N_2} F(b, b; -\frac{1}{N_2-1}), & d < 0. 
\end{cases} \]

B. The Test Sample Estimator (denoted p_t or p_T).

The test sample estimator is equal to \( p_t \), the proportion of the \( M_2 \) test observations from \( \pi_2 \) misclassified by \( R(\bar{x}_1, \bar{x}_2) \). \( M_2p_t \) is binomially distributed with conditional expectation equal to 
\( M_2p_2(R(\bar{x}_1, \bar{x}_2), \mu_2) \), where \( p_2(R(\bar{x}_1, \bar{x}_2), \mu_2) \) is the conditional probability of misclassification. Hence in the \( p \)-dimensional normal case, using (1.2), Chapter 4, \( \text{Ep}_t = 1 - F(c) \) and \( \text{Var p}_t = M_2^{-1}F(C)[1-F(C)] \), where \( C \) is defined in Table X. In the univariate normal case, using (1.6), Chapter 3, \( \text{Ep}_t = 1 - F(c) \) when \( d > 0 \), \( \text{Ep}_t = F(c) \) when \( d < 0 \), and \( \text{Var p}_t = M_2^{-1}F(c)[1-F(c)] \), where \( d = \bar{x}_1 - \bar{x}_2 \) and \( c = \frac{1}{2}d + \bar{x}_2 - \mu_2 \).

C. The Two-Straight Lines Estimator, \( p_{TSL} \).

By definition, \( p_{TSL} = p_0(N) - \frac{1}{2}p_0(\frac{1}{2}N) + \frac{1}{2}p_t(\frac{1}{2}N) \). Let \( \bar{x}_1, \bar{x}_2 \) be the sample means based on the original samples of size \( N \) and let \( \bar{x}'_1, \bar{x}'_2 \) be the sample means based on the original samples of size \( \frac{1}{2}N \). Under Assumption Set 2 (\( \bar{x}_1, \bar{x}_2, \bar{x}'_1, \bar{x}'_2 \) fixed and all six samples independent), the mean and variance of \( p_{TSL} \) are easily obtained from the results on \( p_0 \) and \( p_t \) derived in Sections A and B above.

In the univariate case, the expectation of \( p_{TSL} \) was also computed under Assumption Set 1 (\( \bar{x}_1, \bar{x}_2 \) fixed, original and test samples of size \( \frac{1}{2}N \) obtained by randomly splitting the original samples of size \( N \)). The derivation of this expectation follows.
By the definition of $P_{TSL}$,

$$(c.1) \quad E(P_{TSL} | \overline{x}_1, \overline{x}_2) = E[p_0(N) - \frac{1}{2}p_0(\frac{1}{2}N) + \frac{1}{2}p_0(\frac{1}{2}N) | \overline{x}_1, \overline{x}_2].$$

From Section A, $E(p_0(N) | \overline{x}_1, \overline{x}_2) = P(-|b|)$, where $b = \frac{1}{2}d(1 - N^{-1})^{-\frac{1}{2}}$. The values of $E(p_0(\frac{1}{2}N) | \overline{x}_1, \overline{x}_2)$ and $E(p_0(\frac{1}{2}N) | \overline{x}_1, \overline{x}_2)$ cannot, however, be obtained from the results in Sections A and B because the means that are fixed, $\overline{x}_1$ and $\overline{x}_2$, are not the means based on the original samples of size $\frac{1}{2}N$ but those based on the total samples of size $N$. Also the test sample of size $\frac{1}{2}N$ is not independent of the original sample of size $N$.

To compute $E(p_0(\frac{1}{2}N) | \overline{x}_1, \overline{x}_2)$, let $y_{21}$ denote the first observation in the original sample of size $\frac{1}{2}N$ from $\Pi_2$. Then by the rule (1.2), Chapter 3 and the definition of $p_0$,

$$(c.2) \quad E(p_0(\frac{1}{2}N) | \overline{x}_1, \overline{x}_2) = P\{y_{21} \text{ misclassified by } R(\overline{y}_1, \overline{y}_2) | \overline{x}_1, \overline{x}_2\}$$

$$= P\{\overline{y}_1 - \overline{y}_2 > 0, y_{21} - \frac{1}{2}y_1 + \frac{1}{2}x_2 > 0 | \overline{x}_1, \overline{x}_2\}$$

$$+ P\{\overline{y}_1 - \overline{y}_2 < 0, y_{21} - \frac{1}{2}y_1 + \frac{1}{2}x_2 < 0 | \overline{x}_1, \overline{x}_2\}.$$ 

By the same procedure used in Section A to get the conditional distributions (see also Anderson (1958, p. 28)), it can be shown that the conditional distribution of $(\overline{y}_1, y_{21}, \overline{y}_2)$ given $\overline{x}_1, \overline{x}_2$ is $N(\mu_c, \Sigma_{11.2})$, where $\mu_c = (\overline{x}_1, \overline{x}_2, \overline{x}_2)'$ and

$$\Sigma_{11.2} = \begin{pmatrix} N^{-1} & 0 & 0 \\ 0 & (N-1)N^{-1} & N^{-1} \\ 0 & N^{-1} & N^{-1} \end{pmatrix}. $$

Then the conditional distribution of $\overline{y}_1 - \overline{y}_2$ given $\overline{x}_1, \overline{x}_2$ is $N(d, 2N^{-1})$ and the conditional distribution of $y_{21} - \frac{1}{2}(\overline{y}_1 + \overline{y}_2)$

Because of the two sets of sample means involved in $P_{TSL}$, we will here explicitly indicate the conditioning variables in writing the expectations in order to avoid confusion.
given \( \bar{x}_1, \bar{x}_2 \) is \( N(-\frac{1}{2}d, 1 - \frac{3}{2}N^{-1}) \), where \( d \) denotes \( \bar{x}_1 - \bar{x}_2 \) (as in Table I). Thus using (C.2)

\[
E(p_0(\frac{1}{2}N)|\bar{x}_1, \bar{x}_2) = \{P(\frac{1}{2}d(2N)^{\frac{1}{2}} - \frac{3}{2}d(1 - \frac{3}{2}N^{-1})^{\frac{1}{2}}; \rho) \\
+ P(-\frac{1}{2}d(2N)^{\frac{1}{2}}, \frac{3}{2}d(1 - \frac{3}{2}N^{-1})^{-\frac{1}{2}}; \rho)\},
\]

with \( \rho = -(2N-3)^{-\frac{1}{2}} \).

Next we will compute \( E(p_t(\frac{1}{2}N)|\bar{x}_1, \bar{x}_2) \). Let \( t_{21} \) denote the first observation in the test sample of size \( \frac{1}{2}N \) from \( \Pi_2 \). Then by the rule (1.2), Chapter 3, and the definition of \( p_t \),

\[
E(p_t(\frac{1}{2}N)|\bar{x}_1, \bar{x}_2) = P\{t_{21} \text{ misclassified by } R(\bar{y}_1, \bar{y}_2)|\bar{x}_1, \bar{x}_2\}
\]

\[
= P\{\bar{y}_1 - \bar{y}_2 > 0, t_{21} - \frac{1}{2}(\bar{y}_1 + \bar{y}_2) > 0|\bar{x}_1, \bar{x}_2\}
\]

\[
+ P\{\bar{y}_1 - \bar{y}_2 < 0, t_{21} - \frac{1}{2}(\bar{y}_1 + \bar{y}_2) < 0|\bar{x}_1, \bar{x}_2\}.
\]

It can be shown that the conditional distribution of \( (\bar{y}_1, t_{21}, \bar{y}_2) \) given \( \bar{x}_1, \bar{x}_2 \) is \( N(\mu, \psi_{11.2}) \), where \( \mu = (\bar{x}_1, \bar{x}_2, \bar{x}_2)' \) and

\[
\psi_{11.2} = \begin{pmatrix}
N^{-1} & 0 & 0 \\
0 & (N-1)N^{-1} & -N^{-1} \\
0 & -N^{-1} & N^{-1}
\end{pmatrix}.
\]

Hence \( \bar{y}_1 - \bar{y}_2|\bar{x}_1, \bar{x}_2 \) is \( N(d, 2N^{-1}) \) and \( (t_{21} - \frac{1}{2}(\bar{y}_1 + \bar{y}_2)|\bar{x}_1, \bar{x}_2) \) is \( N(-\frac{1}{2}d, 1 + \frac{1}{2N}). \) Thus using (C.4),

\[
E(p_t(\frac{1}{2}N)|\bar{x}_1, \bar{x}_2) = \{P(\frac{1}{2}d(2N)^{\frac{1}{2}} - \frac{1}{2}d(1 + \frac{1}{2N^{-1}})^{-\frac{1}{2}}; \rho) \\
+ F(-\frac{1}{2}d(2N)^{\frac{1}{2}}, \frac{1}{2}d(1 + \frac{1}{2N^{-1}})^{-\frac{1}{2}}; \rho)\}
\]

with \( \rho = (2N+1)^{-\frac{1}{2}} \).

Finally, using the formula given in (C.1) and the results of (C.3) and (C.5), the expression for \( E(P_{TSL}|ar{x}_1, \bar{x}_2) \) given in Table IV is obtained.
D. Lachenbruch's U Estimator, $P_U$.

The mean and variance of $P_U$ are computed only in the univariate case where $\Pi_1$ is $N(\mu_1, 1)$ and $\Pi_2$ is $N(\mu_2, 1)$. Let $x_{1i}(i=1, \ldots, N_1)$ and $x_{2j}(j=1, \ldots, N_2)$ be the observations from $\Pi_1$ and $\Pi_2$, respectively.

As usual, $\bar{x}_1 = N_1^{-1} \sum x_{1i}$ and $\bar{x}_2 = N_2^{-1} \sum x_{2j}$. Let $\bar{x}_2^{(k)} = (N_2-1)^{-1} \sum_{j=1}^{k} x_{2j}$ and let $R^{(k)} = R(x_1, \bar{x}_2^{(k)})$ denote the rule based on the $N_1 + N_2 - k$ observations with $x_{2k}$ deleted.

By definition, $P_U = N_2^{-1} \sum q_i$, where

$$q_i = \begin{cases} 
1 & \text{if } x_{2i} \text{ is misclassified by } R^{(i)} = R(x_1, \bar{x}_2^{(i)}) \\
0 & \text{otherwise.}
\end{cases}$$

Hence

(D.1) \hspace{1cm} \mathbb{E}(P_U) = \mathbb{E}(q_1)

and

(D.2) \hspace{1cm} \text{Var}(P_U) = N_2^{-1} \left[ \mathbb{E} q_1^2 - (\mathbb{E} q_1)^2 \right] + \frac{N_2-1}{N_2} \left[ \mathbb{E} q_2 - (\mathbb{E} q_1)^2 \right],

where

(D.3) \hspace{1cm} \mathbb{E} q_1 = \mathbb{E} q_1^2 = \mathbb{P}\{x_{21} \text{ misclassified by } R^{(1)} | \bar{x}_1, \bar{x}_2\}

and

(D.4) \hspace{1cm} \mathbb{E} q_2 = \mathbb{P}\left\{ \begin{array}{l} x_{21} \text{ misclassified by } R^{(1)}, \\
\quad \text{or} \\
x_{22} \text{ misclassified by } R^{(2)} | \bar{x}_1, \bar{x}_2 \end{array} \right\}.

Therefore to get the moments of $P_U$, we will first compute $\mathbb{E} q_1$. By (D.3) and the rule given by (1.2), Chapter 3,

(D.5) \hspace{1cm} \mathbb{E} q_1 = \mathbb{P}\{\bar{x}_1 - \bar{x}_2^{(1)} > 0, x_{21} - \frac{1}{2}(\bar{x}_1 + \bar{x}_2^{(1)}) > 0 | \bar{x}_1, \bar{x}_2 \} \\
+ \mathbb{P}\{\bar{x}_1 - \bar{x}_2^{(1)} < 0, x_{21} - \frac{1}{2}(\bar{x}_1 + \bar{x}_2^{(1)}) < 0 | \bar{x}_1, \bar{x}_2 \}.

\footnote{All the expectations are expectations given $\bar{x}_1, \bar{x}_2$.}
When $\bar{x}_2$ is fixed, there is a linear relation between $\bar{x}_2^{(1)}$ and $x_{21}$. Substituting $x_{21} = N_2 \bar{x}_2 - (N_2 - 1)\bar{x}_2^{(1)}$,

$$\text{Eq}_1 = P\{x_2^{(1)} < \bar{x}_1, \bar{x}_2^{(1)} < (N_2 - \frac{1}{2})^{-1}(N_2 \bar{x}_2 - \frac{1}{2}x_1) \}$$

$$+ P\{x_2^{(1)} > \bar{x}_1, \bar{x}_2^{(1)} > (N_2 - \frac{1}{2})^{-1}(N_2 \bar{x}_2 - \frac{1}{2}x_1) \},$$

or equivalently

(D.6) \[ \text{Eq}_1 = P\{x_2^{(1)} < a_{\min}\} + P\{x_2^{(1)} > a_{\max}\}, \]

where $a_{\min} = \min[\bar{x}_1, (N_2 - \frac{1}{2})^{-1}(N_2 \bar{x}_2 - \frac{1}{2}x_1)]$ and

$a_{\max} = \max[\bar{x}_1, (N_2 - \frac{1}{2})^{-1}(N_2 \bar{x}_2 - \frac{1}{2}x_1)]$. The values of $a_{\min}$ and $a_{\max}$ are as shown in the table:

<table>
<thead>
<tr>
<th>$d = \bar{x}_1 - \bar{x}_2 &gt; 0$</th>
<th>$(N_2 - \frac{1}{2})^{-1}(N_2 \bar{x}_2 - \frac{1}{2}x_1)$</th>
<th>$\bar{x}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d = \bar{x}_1 - \bar{x}_2 &lt; 0$</td>
<td>$\bar{x}_1$</td>
<td>$(N_2 - \frac{1}{2})^{-1}(N_2 \bar{x}_2 - \frac{1}{2}x_1)$</td>
</tr>
</tbody>
</table>

It can be shown (see Section A or Anderson (1958, p. 28)) that $\bar{x}_2^{(1)}|\bar{x}_2$ is $N(\bar{x}_2, N_2^{-1}(N_2 - 1)^{-1})$. Thus, letting $B = N_2^{-\frac{1}{2}}(N_2 - 1)^{-\frac{1}{2}}$ and $A = (N_2 - \frac{1}{2})B$ and using (D.6) and the appropriate values of $a_{\min}$ and $a_{\max}$,

(D.7) \[ \text{Eq}_1 = \begin{cases} 
F\left(-\frac{d}{2A}\right) + F\left(-\frac{d}{B}\right), & d > 0 \\
F\left(\frac{d}{2A}\right) + F\left(\frac{d}{B}\right), & d < 0.
\end{cases} \]

Next we will compute $\text{Eq}_1^{q_2}$. Let the sets $S, T, U, V$ be defined as

$S: \bar{x}_1 - \bar{x}_2^{(1)} > 0$ \hspace{1cm} $T: \bar{x}_{21} - \frac{1}{2}(\bar{x}_1 + \bar{x}_2^{(1)}) > 0$

$U: \bar{x}_1 - \bar{x}_2^{(2)} > 0$ \hspace{1cm} $V: \bar{x}_{22} - \frac{1}{2}(\bar{x}_1 + \bar{x}_2^{(2)}) > 0$

and let $S', T', U', V'$ denote the complementary sets. Then using
(D.4) and the definition of the rule (1.2), Chapter 3

\[ \text{Eq}_{1.2} = P_I + P_{II} + P_{III} + P_{IV}, \]

where

\[
\begin{align*}
P_I &= P\{STUV|\bar{x}_1, \bar{x}_2\} & P_{II} &= P\{STUV'|\bar{x}_1, \bar{x}_2\} \\
P_{III} &= P\{S'T'UV|\bar{x}_1, \bar{x}_2\} & P_{IV} &= P\{S'T'U'V'|\bar{x}_1, \bar{x}_2\}.
\end{align*}
\]

Substituting \( x_{21} = N_2\bar{x}_2 - (N_2 - 1)\bar{x}_{2(1)} \) and \( x_{22} = N_2\bar{x}_2 - (N_2 - 1)\bar{x}_{2(2)} \)

and defining \( a_{\min}, a_{\max} \) as above,

\[ \text{(D.9)} \]

\[
\begin{align*}
P_I &= P\{x_{2(1)} < a_{\min}, x_{2(2)} < a_{\min}\}, \\
P_{II} &= P\{x_{2(1)} < a_{\min}, x_{2(2)} > a_{\max}\}, \\
P_{III} &= P\{x_{2(1)} > a_{\max}, x_{2(2)} < a_{\min}\}, \\
P_{IV} &= P\{x_{2(1)} > a_{\max}, x_{2(2)} > a_{\max}\}.
\end{align*}
\]

As above let \( B = N_2^{1/2}(N_2 - 1)^{-1/2} \) and \( A = (N_2 - 1/2)B \). The joint distribution of \( x_{2(1)}, x_{2(2)} \) given \( \bar{x}_2 \) is \( N(\mu_c, \Sigma_{11,2}) \) with

\[
\mu_c = (\bar{x}_2, \bar{x}_2)',
\]

and

\[
\Sigma_{11,2} = \begin{pmatrix}
B^2 & -B^2(N_2 - 1)^{-1} \\
-B^2(N_2 - 1)^{-1} & B^2
\end{pmatrix}.
\]

Hence using (D.8) and (D.9) and the values of \( a_{\min} \) and \( a_{\max} \) given above

\[ \text{(D.10)} \]

\[
\text{Eq}_{1.2} = \begin{cases}
F\left(-\frac{d}{2A}, -\frac{d}{2A}; \rho\right) + 2F\left(-\frac{d}{2A}, -\frac{d}{B}; -\rho\right) + F\left(-\frac{d}{B}, -\frac{d}{B}; \rho\right), & d > 0 \\
F\left(\frac{d}{2A}, \frac{d}{2A}; \rho\right) + 2F\left(\frac{d}{2A}, \frac{d}{B}; -\rho\right) + F\left(\frac{d}{B}, \frac{d}{B}; \rho\right), & d < 0
\end{cases},
\]

where \( \rho = -(N_2 - 1)^{-1} \). Note that the \( P_{II} \) and \( P_{III} \) terms have been combined by using the equality, \( P\{X < h, Y > k; \rho\} = P\{X < h, Y < -k; -\rho\} \)

for \( X, Y \) standard normal variables with correlation \( \rho \).
The mean and variance of \( P_U \) can now easily be found from the formulas (D.1) and (D.2) by using the results of (D.7) and (D.10).

E. The Reclassification Estimator Adjusted for Bias, \( P_{R_a} \)

The adjusted reclassification estimator, \( P_{R_a} \), is equal to \( P_0 \) (see Section A) with a term added to adjust for bias. Specifically, \( P_{R_a} = P_0 + \hat{E}_{P_t} - \hat{E}_{P_0} \). From Section A above, \( \hat{E}_{P_0} = 1 - F(B) = \hat{E}_{P_0} \) and from Section B, \( \hat{E}_{P_t} = 1 - F(C) \mid_{\mu_2 = \bar{x}_2} = 1 - F(\frac{1}{2}D) \). Hence \( E(P_{R_a}) = E(\hat{E}_{P_t}) = 1 - F(\frac{1}{2}D) \). When \( p = 1 \) and \( \Sigma = \sigma^2 = 1, D = |d| \), and thus in the univariate case, \( E(P_{R_a}) = 1 - F(\frac{1}{2}|d|) \). Since \( \hat{E}_{P_t} \) and \( \hat{E}_{P_0} \) are constants when \( \bar{x}_1, \bar{x}_2 \) are fixed, \( p \geq 1 \), \( \text{Var}(P_{R_a}) = \text{Var} P_0 \) in both multivariate and univariate cases.

F. \( Q_T = P_2(R(\bar{x}_1, \bar{x}_2), \hat{\mu}_2 = \bar{\bar{x}}_2) \).

Let \( \bar{\bar{x}}_2 \) be the mean of the \( M_2 \) test observations from \( \Pi_2 \). The estimator \( Q_T \) is obtained by using \( \bar{\bar{x}}_2 \) as an estimate of \( \mu_2 \) in the expression for the conditional probability of misclassification, \( P_2(R(\bar{x}_1, \bar{x}_2), \mu_2) \), given in (1.2), Chapter 4. It is equal to \( 1 - F(A - W) \), where \( A = \frac{1}{2}D + D^{-1}(\bar{x}_1 - \bar{x}_2)' \Sigma^{-1} \bar{x}_2 \) and \( W = D^{-1}(\bar{x}_1 - \bar{x}_2)' \Sigma^{-1} \bar{\bar{x}}_2 \).

(A and \( W \) defined only for this derivation and the one in Section G.)

When \( \bar{x}_1, \bar{x}_2 \) are fixed and test samples are random, \( A \) is constant and \( W \) is univariate normal with expectation \( EW = D^{-1}(\bar{x}_1 - \bar{x}_2)' \Sigma^{-1} \mu_2 \) and variance \( \sigma_W^2 = \mu_2^{-1} \).

Let \( F_W \) denote the distribution function of \( W \). Then setting \( z = (W - EW)/\sigma_W \)

\[ EF(A - W) = \int_{-\infty}^{\infty} F(A - W) dF_W \]

\[ = \int_{-\infty}^{\infty} F(-\sigma_W z + A - EW) f(z) dz \]

\[ = F((A - EW)(1 + \sigma_W^2)^{-\frac{1}{2}}) \]

Further
\[(F.2)\quad \text{EF}^2(A-W) = \int_{-\infty}^{\infty} F^2(-\sigma_w z + A-EW)f(z)dz \]
\[
= F((A-EW)(1 + \sigma_w^2)^{-\frac{1}{2}}, (A-EW)(1 + \sigma_w^2)^{-\frac{1}{2}}; \rho)
\]

where \(\rho = \rho(x + \sigma_w z, y + \sigma_w z)\) with \(x, y, z\) independent \(N(0, 1)\).

Using the expressions for \(A\) and \(EW\) given above, \(A-EW = C\), with \(C\) defined in Table X. Furthermore, \(\rho = \sigma_w^2(1 + \sigma_w^2)^{-1} = \frac{1}{1 + M_2}\).

Hence letting \(C(1 + M_2^{-1})^{-\frac{1}{2}}\) be denoted by \(C'\),

\[(F.3)\quad \text{EQ}_T = 1-F(C')\]

and

\[(F.4)\quad \text{Var} Q_T = F(C', C'; (1+M_2)^{-1}) - F(C', C'; 0)\]

When the populations are univariate normal, \(Q_T\) is equal to

\(1-F(\frac{1}{2}d + \overline{x}_2 - \overline{t}_2)\) when \(d > 0\), and is equal to \(F(\frac{1}{2}d + \overline{x}_2 - \overline{t}_2)\) when \(d < 0\), \(d = \overline{x}_1 - \overline{x}_2\). The conditional mean and variance can be computed directly, or more simply, can be obtained from the multivariate results. When \(p = 1\) and \(\Sigma = \sigma^2 = 1\),

\[
C = \begin{cases} 
  c & \text{if } d > 0 \\
  -c & \text{if } d < 0 
\end{cases}
\]

where \(c = \frac{1}{2}d + \overline{x}_2 - \mu_2\). Thus letting \(c' = c(1 + M_2^{-1})^{-\frac{1}{2}}\), from (F.3)

\[(F.5)\quad \text{EQ}_T = \begin{cases} 
  1-F(c') & , d > 0 \\
  F(c') & , d < 0 
\end{cases}, \quad \text{and from (F.4)}
\]

\[(F.6)\quad \text{Var} Q_T = F(c', c'; (1+M_2)^{-1}) - F(c', c'; 0), \quad \text{both for } d > 0 \text{ and } d < 0.
\]

G. The (Conditional) Minimum Variance Unbiased Estimator, \(Q_{RB}^c\)

The conditional UMVU estimator, \(Q_{RB}^c\), is derived in Chapter 5, Section 4.4. Let \(A' = A(1 - M_2^{-1})^{-\frac{1}{2}}\) and \(W' = W(1 - M_2^{-1})^{-\frac{1}{2}}\), where \(A\) and \(W\) are as defined in Section F above. Then \(Q_{RB}^c = 1-F(A' - W')\).
Let \( EW' \) and \( \sigma_{w'} \) denote the conditional mean and variance of \( W' \).

Then from the results in (F.1) and (F.2)

\[(G.1) \quad EF(A' - w') = F((A' - EW')(1 + \sigma_{w'}^2)^{-\frac{1}{2}})\]

\[(G.2) \quad EF^2(A' - w') = F((A' - EW')(1 + \sigma_{w'}^2)^{-\frac{1}{2}}, (A' - EW')(1 + \sigma_{w'}^2)^{-\frac{1}{2}}; \rho)\]

where \( \rho = \sigma_{w'}^2(1 + \sigma_{w'}^2)^{-1} \).

From the definition of \( A', W' \) and the results given in Section F,

\[A' - EW' = (1 - M_2^{-1})^{-\frac{1}{2}}(A - EW) = (1 - M_2^{-1})^{-\frac{1}{2}}C\]

\[\sigma_{w'}^2 = (1 - M_2^{-1})^{-1}\sigma_w^2 = (M_2 - 1)^{-1}\]

\[1 + \sigma_{w'}^2 = (1 - M_2^{-1})^{-1}\]

\[(A - EW')(1 + \sigma_{w'}^2)^{-\frac{1}{2}} = C.\]

Hence substituting in (G.1) and (G.2),

\[(G.3) \quad E \rho_{RB} = 1 - F(C)\]

and

\[(G.4) \quad \text{Var} \ \rho_{RB} = F(C, C; M_2^{-1}) - F(C, C; 0).\]

For the univariate case, following the same method as in Section F, we find that

\[(G.5) \quad E \rho_{RB} = \begin{cases} 1 - F(c) & d > 0 \\ F(c) & d < 0 \end{cases}\]

and

\[\text{Var} \ \rho_{RB} = F(c, c; M_2^{-1}) - F(c, c; 0),\]

where \( c = \frac{1}{2}d + \bar{x}_2 - \mu_2.\)
APPENDIX III

DERIVATIONS RELATED TO THE ASYMPTOTIC RESULTS
IN THE UNIVARIATE NORMAL CASE

(see Chapter 3, Section 4 and Table V).

Let $\Pi_1$ and $\Pi_2$ be univariate normal with unknown means $\mu_1, \mu_2$ and common variance one. Let $P_2(R(x_1, x_2), \mu_2)$ be the conditional probability of misclassifying an observation from $\Pi_2$. To get desired asymptotic expressions for the conditional bias and conditional variance of the estimators of $P_2(R(x_1, x_2), \mu_2)$, Taylor series are needed for various univariate and bivariate normal distribution functions. The univariate computations are straightforward and are not given. Needed expansions for bivariate distribution functions are derived in Section C below with some preliminaries in Sections A and B. All of the Taylor series expansions are given in Table V.

A. Definitions of Notation Used Below.

$$a(N) = o(f(N)) \text{ if } \frac{a(N)}{f(N)} \to 0 \text{ as } N \to \infty$$

$$b(N) = o_p(f(N)) \text{ if } \frac{b(N)}{f(N)} \to 0 \text{ in probability as } N \to \infty$$

Note that $a(N) = o(f(N))$ implies $a(N) = o_p(f(N))$.

B. Lemma: Let $y \sim N(0, \sigma^2 N^{-1})$. Then $y = o_p(N^{-\frac{1}{2}+\epsilon})$, any $\epsilon > 0$.

Proof: $y \sim N(0, \sigma^2 N^{-1})$ implies $N^\beta y \sim N(0, \sigma^2 N^{2\beta-1})$, and by the Chebyshev inequality $P[|N^\beta y| > c] \leq c^{-2} \text{Var}(N^\beta y) = c^{-2} \sigma^2 N^{2\beta-1}$. Therefore $N^\beta y$ goes to zero in probability as $N$ goes to infinity for $\beta < \frac{1}{2}$, or equivalently $y = o_p(N^{-\frac{1}{2}+\epsilon})$, any $\epsilon > 0$. This result cannot be improved because for $\beta = \frac{1}{2}$, $N^\beta y$ is $N(0, \sigma^2)$ and hence $N^{\frac{3}{2}} y \not\to 0$ in probability.
C. Taylor Series Expansions for the Bivariate Normal Distribution

Functions.

The notation used below and not defined there is defined in Table I. For the asymptotic results all sample sizes are equal to \( N \).

C.1. Expansion for \( F(b, b; -(N-1)^{-1}) - F(b, b; 0) \) and for \( H(b, b; -(N-1)^{-1}) - H(b, b; 0) \), \( b = \frac{1}{2d}(1-N^{-1})^{-\frac{1}{2}} \).

Let \( \mathcal{D} = F(b, b; \rho) - F(b, b; 0) \) with \( \rho = -(N-1)^{-1} \). Expanding first on \( \rho \) with \( b \) fixed,

\[
\mathcal{D} = F(b, b; 0) - F(b, b; 0) - \frac{1}{N-1} F_3(b, b; 0) + o_p(N^{-2+\epsilon})
\]

where \( F_3(x, y; 0) = \frac{\partial}{\partial \rho} F(x, y; \rho) \bigg|_{\rho=0} \). Now let \( u = \alpha_1 - \alpha_2 \), \( \gamma = \frac{1}{2} \delta \), \( g = f(\gamma) \), \( G = F(\gamma) \). Then writing \( b = \gamma + t \), where

\[
t = \frac{1}{2} u + \frac{1}{2} \gamma N^{-1} + \frac{1}{2} u N^{-1} + o_p(N^{-2+\epsilon}),
\]

and expanding \( F_3(b, b; 0) \) by bivariate Taylor series (using an obvious notation),

\[
\mathcal{D} = -F_3(b, b; 0)N^{-1} + o_p(N^{-2+\epsilon})
\]

From (C.11) and (C.12) below, \( F_3(\gamma, \gamma; 0) = g^2 \) and \( F_3(\gamma, \gamma; 0) = F_3(\gamma, \gamma; 0) = -g^2 \) and hence

\[
(N-1) \mathcal{D} = -g^2 N^{-1} + \frac{1}{2} g^2 u N^{-1} + o_p(N^{-2+\epsilon}).
\]

The expansion for \( H(b, b; -\frac{1}{N-1}) - H(b, b; 0) \) is obtained in the same way and because \( H_3(\gamma, \gamma; 0) = F_3(\gamma, \gamma; 0), H_3(\gamma, \gamma; 0) = H_3(\gamma, \gamma; 0) = F_3(\gamma, \gamma; 0) \) (see (C.1) and (C.12)), the expansion is identical to that for \( \mathcal{D} \) given in (C.1).

C.2. Expansions for \( N^{-1} F(b, b; \rho) \) and \( N^{-1} H(b, b; \rho), \rho = -(N-1)^{-1} \).

Observe that, using (C.1),

\[
N^{-1} F(b, b; -\frac{1}{N-1}) = N^{-1} \mathcal{D} + F(b, b; 0) = N^{-1} F(b, b; 0) + o_p(N^{-2+\epsilon})
\]

(C.2)

\[
N^{-1} H(b, b; -\frac{1}{N-1}) = N^{-1} H(b, b; 0) + o_p(N^{-2+\epsilon}).
\]
Using Taylor series for two variables with  \( b = \frac{1}{2} \delta + t \), where
\[ t = \frac{1}{2}(\alpha_1 - \alpha_2) + o_p(N^{-1+\varepsilon}), \]
and denoting \( \alpha_1 - \alpha_2 = u, \frac{1}{2} \delta = \gamma \)
as before, we get the expansions

\[
N^{-1}F(b, b; 0) = F(\gamma, \gamma; 0)N^{-1} + \frac{1}{2}[F_1(\gamma, \gamma; 0) + F_2(\gamma, \gamma; 0)]vN^{-1} + o_p(N^{-2+\varepsilon}).
\]

\[
(c.3)
\]
\[
N^{-1}H(b, b; 0) = H(\gamma, \gamma; 0)N^{-1} + \frac{1}{2}[H_1(\gamma, \gamma; 0) + H_2(\gamma, \gamma; 0)]uN^{-1} + o_p(N^{-2+\varepsilon}).
\]

From (c.13) and (c.14) below, \( F_1(\gamma, \gamma; 0) = F_2(\gamma, \gamma; 0) = gG \) and \( H_1(\gamma, \gamma; 0) = H_2(\gamma, \gamma; 0) = -g(1-G) \). Thus using (c.2) and (c.3)

\[
N^{-1}F(b, b; \frac{1}{N-1}) = G^2N^{-1} + gGuN^{-1} + o_p(N^{-2+\varepsilon}),
\]

and

\[
N^{-1}H(b, b; \frac{1}{N-1}) = (1-G)^2N^{-1} - g(1-G)uN^{-1} + o_p(N^{-2+\varepsilon}).
\]

C.3. Expansion for \( F(c, c; N^{-1}) - F(c, c; 0) \), \( c = \frac{1}{2}(x_1 + x_2) - u_2 \).

Let \( \mathcal{F} = F(c, c; \rho) - F(c, c; 0) \) with \( \rho = N^{-1} \). Expanding

first on \( \rho \) with \( c \) fixed,

\[
\mathcal{F} = F(c, c; 0) - F(c, c; 0) + F_3(c, c; 0)N^{-1} + o_p(N^{-2+\varepsilon})
\]

\[
= F_3(c, c; 0)N^{-1} + o_p(N^{-2+\varepsilon}).
\]

Letting \( \gamma = \frac{1}{2} \delta \) and \( v = \alpha_1 + \alpha_2 \), we can write \( c = \gamma + \frac{1}{2}v \). Then

expanding \( F_3(c, c; 0) \) by bivariate Taylor series,

\[
(c.4) \quad \mathcal{F} = F_3(\gamma, \gamma; 0)N^{-1} + \frac{1}{2}[F_{31}(\gamma, \gamma; 0) + F_{32}(\gamma, \gamma; 0)]vN^{-1} + o_p(N^{-2+\varepsilon}).
\]

From (c.11) and (c.12) below, \( F_3(\gamma, \gamma; 0) = g^2 \) and \( F_{31}(\gamma, \gamma; 0) = F_{32}(\gamma, \gamma; 0) = -g^2 \). Thus

\[
(c.5) \quad \mathcal{F} = g^2N^{-1} - \frac{1}{2}g^2vN^{-1} + o_p(N^{-2+\varepsilon}).
\]

- 141 -
C.4. Expansion for $F(c', c'; (N+1)^{-1}) - F(c', c'; 0)$, $c' = c(N^{-1})^{-1/2}$.

Let $\mathcal{A}' = F(c', c'; \rho) - F(c', c'; 0)$ with $\rho = (N+1)^{-1}$.

Expanding on $\rho$ with $c'$ fixed,

$$\mathcal{A} = F_3(c', c'; 0)N^{-1} + o_p(N^{-2} + \epsilon).$$

Now from Table V, $c' = \frac{1}{2} \delta + \frac{1}{2}(\alpha_1 + \alpha_2) - \frac{1}{2} \delta N^{-1} + o_p(N^{-3/2} + \epsilon)$, or using $\gamma = \frac{1}{2} \delta$ and $v = \alpha_1 + \alpha_2$, $c' = \gamma + \frac{1}{2} v - \frac{1}{2} \gamma N^{-1} + o_p(N^{-3/2} + \epsilon)$.

Hence expanding $F_3(c', c'; 0)N^{-1}$,

(C.6) $$\mathcal{A}' = F_3(\gamma, \gamma; 0)N^{-1} + \frac{1}{2}[F_{31}(\gamma, \gamma; 0) + F_{32}(\gamma, \gamma; 0)]vN^{-1} + o_p(N^{-2} + \epsilon).$$

This expression for $\mathcal{A}'$ is equal to the expression for $\mathcal{A}$ given in (C.4) and hence by (C.5),

(C.7) $$\mathcal{A}' = g^2N^{-1} - \frac{1}{2} g \gamma v N^{-1} + o_p(N^{-2} + \epsilon).$$


Note: $u$ and $v$ are dummy variables in this section and have no connection with the notation $u = \alpha_1 - \alpha_2$ and $v = \alpha_1 + \alpha_2$ used in the preceding sections.

Let

$$g(x, y; \rho) = \exp\left\{-\frac{1}{2} \rho \frac{1}{1-\rho^2} [x^2 - 2\rho xy + y^2]\right\}$$

$$I = I(u, v; \rho) = \int_{-\infty}^{u} \int_{-\infty}^{v} g(x, y; \rho) dy dx$$

$$I^* = I^*(u, v; \rho) = \int_{u}^{\infty} \int_{v}^{\infty} g(x, y; \rho) dy dx$$

$$A = A(u, v; \rho) = \frac{\partial}{\partial \rho} I(u, v; \rho)$$

$$B = B(u, v; \rho) = \frac{\partial}{\partial \rho} I^*(u, v; \rho)$$

$$k_\rho = (2\pi)^{-1}(1-\rho^2)^{-3/2}$$

$$k'_{\rho} = \frac{d}{d\rho} k_\rho$$

$$k''_{\rho} = \frac{d^2}{d\rho^2} k_\rho.$$
Then the derivatives needed are, as functions of \(u, v\) and \(\rho\), as follows:

\[
\begin{align*}
F(u, v; \rho) &= k_\rho I(u, v; \rho) \\
H(u, v; \rho) &= k_\rho I^*(u, v; \rho) \\
F_3(u, v; \rho) &= k_A + k'_I \rho \\
H_3(u, v; \rho) &= k_B + k'_I \rho^* \\
F_{31}(u, v; \rho) &= k_\rho \frac{\partial A}{\partial u} + k'_I \frac{\partial I}{\partial u} \\
H_{31}(u, v; \rho) &= k_\rho \frac{\partial B}{\partial u} + k'_I \frac{\partial I^*}{\partial u} \\
F_{32}(u, v; \rho) &= k_\rho \frac{\partial A}{\partial v} + k'_I \frac{\partial I}{\partial v} \\
H_{32}(u, v; \rho) &= k_\rho \frac{\partial B}{\partial v} + k'_I \frac{\partial I^*}{\partial v} \\
F_1(u, v; \rho) &= k_\rho \int_v^u g(u, y; \rho) \, dy \\
H_1(u, v; \rho) &= -k_\rho \int_v^\infty g(u, y; \rho) \, dy \\
F_2(u, v; \rho) &= k_\rho \int_u^v g(x, v; \rho) \, dx \\
H_2(u, v; \rho) &= -k_\rho \int_u^\infty g(x, v; \rho) \, dx \\
F_{33}(u, v; \rho) &= k_\rho \frac{\partial A}{\partial \rho} + 2k'_I \rho + k''_I \rho^*.
\end{align*}
\]

When \(\rho = 0\),

\[
\begin{align*}
k_\rho &= k''_\rho = (2\pi)^{-1} \\
k'_\rho &= 0
\end{align*}
\]

\[
\begin{align*}
F_1(u, v; 0) &= f(u)F(v) \\
H_1(u, v; 0) &= -f(u)[1-F(v)] \\
F_2(u, v; 0) &= f(v)F(u) \\
H_2(u, v; 0) &= -f(v)[1-F(u)].
\end{align*}
\]

Further, taking the derivatives and then substituting \(\rho = 0\),

\[
\begin{align*}
A(u, v; 0) &= (2\pi)[-f(u)][-f(v)]; \quad B(u, v; 0) = (2\pi)f(u)f(v). \\
\frac{\partial A}{\partial u} |_{\rho=0} &= \frac{\partial B}{\partial u} |_{\rho=0} = -(2\pi)u f(u)f(v) \\
\frac{\partial A}{\partial v} |_{\rho=0} &= \frac{\partial B}{\partial v} |_{\rho=0} = -(2\pi)v f(u)f(v) \\
\frac{\partial A}{\partial \rho} |_{\rho=0} &= (2\pi)[uv f(u)f(v) - F(u)F(v)].
\end{align*}
\]

Substituting the values from (C.9) and (C.10) in the derivative formulas of (C.8) with \(u\) and \(v\) set equal to \(\frac{1}{2}\delta\), and again letting \(\gamma = \frac{1}{2}\delta\), \(g = f(\gamma)\), \(G = F(\gamma)\), we get the following values for the

(c.11) \( F_3(y, y; 0) = H_3(y, y; 0) = g^2. \)

(c.12) \( F_{31}(y, y; 0) = F_{32}(y, y; 0) = H_{31}(y, y; 0) = H_{32}(y, y; 0) = -yg^2. \)

(c.13) \( F_1(y, y; 0) = F_2(y, y; 0) = gG. \)

(c.14) \( H_1(y, y; 0) = H_2(y, y; 0) = -g(1-g). \)

(c.15) \( F_{33}(y, y; 0) = y^2g^2. \)

C.6. **Expansion for** \( F(\frac{1}{2}d(2N)^{\frac{1}{2}}, \frac{1}{2}d(1+\frac{1}{2}N^{-1})^{-\frac{1}{2}}; (2N+1)^{-\frac{1}{2}}) \)

\[ - F(\frac{1}{2}d(2N)^{\frac{1}{2}}, \frac{1}{2}d(1-\frac{3}{2}N^{-1})^{-\frac{1}{2}}; -(2N+3)^{-\frac{1}{2}}), d > 0. \]

Let \( U = \frac{1}{2}d(2N)^{\frac{1}{2}}, V = \frac{1}{2}d(1+\frac{1}{2}N^{-1})^{-\frac{1}{2}}, W = \frac{1}{2}d(1-\frac{3}{2}N^{-1})^{-\frac{1}{2}}, \rho_1 = (2N+1)^{-\frac{1}{2}}, \)
and \( \rho_2 = -(2N-3)^{-\frac{1}{2}}. \) Then expanding on \( \rho_1 \) and \( \rho_2 \)

(c.16) \[ F(U, -V; \rho_1) - F(U, -W; \rho_2) \]

\[ = \{F(U, -V; 0) + F_3(U, -V; 0)\rho_1 + \frac{1}{2}F_{33}(U, -V; 0)\rho_1^2 + \ldots \]

\[ - F(U, -W; 0) - F_3(U, -W; 0)\rho_2 - \frac{1}{2}F_{33}(U, -W; 0)\rho_2^2 - \ldots \}. \]

From Table V, when \( d > 0 \) \( F(U) = 1 + o_p(N^{-2+\varepsilon}) \), and hence using the expansions of \( F(V) \) and \( F(W) \) also given in Table V,

(c.17) \[ F(U, -V; 0) = F(-V) + o_p(N^{-2+\varepsilon}) \]

\[ = 1 - G - \frac{1}{2}gu + \frac{1}{16}g^2u^2 + \frac{1}{6}gN^{-1} + o_p(N^{-3/2} + \varepsilon) \]
and

(c.18) \[ F(U, -W; 0) = F(-W) + o_p(N^{-2+\varepsilon}) \]

\[ = 1 - G - \frac{1}{2}gu + \frac{1}{16}g^2u^2 - \frac{3}{8}gN^{-1} + o_p(N^{-3/2} + \varepsilon), \]

where \( u = \alpha_1 - \alpha_2. \) From (c.8), (c.9) and (c.10) above,

\( F_3(x, y; 0) = f(x)f(y) \) and \( F_{33}(x, y; 0) = xy f(x)f(y). \) Hence the
four functions $F_3(U, -V; 0)$, $F_3(U, -W; 0)$, $F_{33}(U, -V; 0)$ and
$F_{33}(U, -W; 0)$ are all $o_p(N^{-2+\varepsilon})$ because for $d > 0$, $f(U) = o_p(N^{-2+\varepsilon})$
and $Uf(U) = o_p(N^{-2+\varepsilon})$.

Thus substituting in (C.16),

\[(C.19) \quad F(U, -V; \rho_1) - F(U, -W; \rho_2) = \frac{1}{2} \delta N^{-1} + o_p(N^{-3/2+\varepsilon}).\]

Similarly we can show that for $d < 0$,

\[F(-U, V; \rho_1) - F(-U, W; \rho_2) = \frac{1}{2} \delta N^{-1} + o_p(N^{-3/2+\varepsilon}).\]
APPENDIX IV

DERIVATIONS RELATED TO THE ASYMPTOTIC RESULTS IN THE
MULTIVARIATE NORMAL CASE

(see Chapter 4, Section 4 and Table III).

Let \( \Pi_1 \) and \( \Pi_2 \) be \( p \)-dimensional normal with unknown mean
vectors \( \mu_1, \mu_2 \) and common known covariance matrix \( \Sigma \). As in the
case of univariate normal populations, Taylor series expansions are
needed for various univariate and bivariate normal distribution functions
in order to get asymptotic expansions for conditional biases and
variances of estimators of conditional probability of misclassification.
The univariate expansions are considered in Section A below and the
bivariate expansions in Sections B, C, and D. Notation not defined
below is given in Table X. All of the expansions are given in Table
XIII.

A. Taylor Series Expansions for \( F(\frac{1}{2}D) \), \( F(C) \), and Similar Functions.

To obtain expansions for \( F(\frac{1}{2}D) \), \( F(C) \), \( F(B) \) and other such
functions, the arguments \( \frac{1}{2}D \), \( C \), \( B \) are first written in the form
\( a + t \), where \( a \) is constant and \( t \) is variable (when the rule is
variable) or a function of \( N \). Then the expansions for \( F(\frac{1}{2}D) \), \( F(C) \),
etc. can be easily obtained from the Taylor series formula given in
Chapter 3, Section 4.1.

To obtain the \( a + t \) expression for \( D \), let
\[
\begin{align*}
\mathbf{u} &= (\mathbf{y}_1 - \mathbf{y}_2)' \Sigma^{-1} (\mathbf{y}_1 - \mathbf{y}_2) \\
\mathbf{v} &= (\mu_1 - \mu_2)' \Sigma^{-1} (\mathbf{y}_1 - \mathbf{y}_2).
\end{align*}
\]
\( \text{Then}\)

\[
D^2 = (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)' \Sigma^{-1} (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2) = \Delta^2 + 2V + U.
\]
Therefore \( D = (\Delta^2)^{\frac{1}{2}} (1 + t)^{\frac{1}{2}} \), where \( t = \Delta^{-2} (2V + U) \) and since the
numerator of \( t \) goes to zero as \( N \) goes to infinity, for asymptotic
purposes we can assume \( |t| < 1 \) and use the expansion \( (1 + t)^r = \sum_{k=0}^{\infty} \binom{r}{k} t^k \).
The resulting expression is

\[(A.2) \quad D = \Delta + \Delta^{-1}V + \frac{1}{2}\Delta^{-1}U - \frac{1}{2}\Delta^{-3}V^2 + o_p(N^{-3/2} + \epsilon).\]

To obtain the \(a + t\) expansion for \(C\) the expansion for \(D^{-1}\) is needed. From (A.1), \(D^{-1} = (D^2)^{-\frac{1}{2}} = (\Delta^2)^{-\frac{1}{2}}(1 + t)^{-\frac{1}{2}},\) where \(t = \Delta^{-2}(2V + U)\). Using the same expansion for \((1 + t)^r\) as above, the expression for \(D^{-1}\) becomes

\[(A.3) \quad D^{-1} = \Delta^{-1} - \Delta^{-3}V - \frac{1}{2}\Delta^{-3}U + o_p(N^{-3/2} + \epsilon).\]

Let \(W = (\varphi_1 - \varphi_2)'\Sigma^{-1}\varphi_2\) and \(Z = (\mu_1 - \mu_2)'\Sigma^{-1}\varphi_2\). Then
\[C = \frac{1}{2}D + D^{-1}W + D^{-1}Z\] (see Table X). Substituting the expansions for \(D\) and \(D^{-1}\) from (A.2) and (A.3), the expansion for \(C\) is

\[(A.4) \quad C = \frac{1}{2}\Delta + \Delta^{-1}(\frac{1}{2}V + Z) + \Delta^{-1}(\frac{1}{2}U + W) - \Delta^{-3}(\frac{1}{2}V^2 + VW) + o_p(N^{-3/2} + \epsilon).\]

Now consider the \(a + t\) expansions for \(B\) and similar functions.

By definition \(B = \frac{1}{2}D(1 - N^{-1})^{-\frac{1}{2}}\). Hence using \((1 - N^{-1})^{-\frac{1}{2}} = 1 + \frac{1}{2}N^{-1} + o(N^{-2+\epsilon})\) and the expansion for \(D\) from (A.2),

\[(A.5) \quad B = \frac{1}{2}\Delta + \frac{1}{2}\Delta^{-1}V + \frac{1}{2}\Delta^{-1}U - \frac{1}{4}\Delta^{-3}V^2 + o_p(N^{-3/2} + \epsilon).\]

The expressions for \(C(1 + N^{-1})^{-\frac{1}{2}}, \frac{1}{2}D(1 + N^{-1})^{-\frac{1}{2}}\) and other such functions are found similarly by multiplying the expression for \(C\) or \(\frac{1}{2}D\) by the appropriate expansion for the coefficient involving \(N\) (these latter expansions are mainly in Table V).

B. **Taylor Series Expansions for** \(H(B, B; \frac{1}{N-1}) - H(B, B; 0)\) and \(N^{-1}H(B, B; -\frac{1}{N-1})\).

Let \(\mathcal{O}^* = H(B, B; -\frac{1}{N-1}) - H(B, B; 0)\) and write \(B = \frac{1}{2}\Delta + t,\) where \(t = \frac{1}{2}\Delta^{-1}V + o_p(N^{-1+\epsilon}).\) Expanding first on \(\rho\) with \(B\) fixed and then on \(B\) as in obtaining the expansion for \(\mathcal{O}\) in Appendix III C.1, we get the expansion
(B.1) \[-\mathcal{G}^* = H_3(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0)N^{-1} + \frac{1}{2}[H_1(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0)+H_2(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0)]\Delta^{-1}VN^{-1} + o_p(N^{-2+\varepsilon}).\]

Then using (C.11) and (C.12) from Appendix III and setting \( \mathcal{G} = f(\frac{1}{2}\Delta) \), we get the needed expansion for \( \mathcal{G}^* \),

(B.2) \[\mathcal{G}^* = -\mathcal{G}2N^{-1} + \frac{1}{2}\mathcal{G}2VN^{-1} + o_p(N^{-2+\varepsilon}).\]

To get the expansion for \( H(B, B; -\frac{1}{N-1}) \) observe that

(B.3) \[N^{-1}H(B, B; -\frac{1}{N-1}) = [\mathcal{G}^* + H(B, B; 0)]N^{-1} = H(B, B; 0)N^{-1} + o_p(N^{-2+\varepsilon}).\]

Further by Taylor series expansion, \( H(B, B; 0)N^{-1} = H(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0)N^{-1} + \frac{1}{2}[H_1(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0)+H_2(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0)]\Delta^{-1}VN^{-1} + o_p(N^{-2+\varepsilon}). \) Hence using (C.14) from Appendix III and setting \( \mathcal{G} = f(\frac{1}{2}\Delta), \mathcal{G} = f(\frac{1}{2}\Delta), \)

(B.4) \[H(B, B; 0)N^{-1} = (1 - \mathcal{G})2N^{-1}\Delta^{-1}g(1 - \mathcal{G})VN^{-1} + o_p(N^{-2+\varepsilon})\]

\[= N^{-1}H(B, B; -\frac{1}{N-1}).\]

C. Taylor Series Expansion for \( F(C, C; N^{-1}) - F(C, C; 0) \).

Let \( \mathcal{G} = F(C, C; \rho) - F(C, C; 0) \) where \( \rho = N^{-1} \). From Table XIII, \( C = \frac{1}{2}\Delta + \Delta^{-1}(\frac{1}{2}V + Z) + o_p(N^{-1+\varepsilon}). \) Hence expanding first on \( \rho \) and then on \( C \) as in the univariate case of Appendix III C.3,

(C.1) \[\mathcal{G} = F_3(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0)N^{-1}\]

\[+ [F_3(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0) + F_3(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0)]\Delta^{-1}(\frac{1}{2}V + Z)N^{-1} + o_p(N^{-2+\varepsilon}).\]

From (C.11) and (C.12), Appendix III, \( F_3(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0) = \mathcal{G}^2 \) and \( F_3(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0) = F_3(\frac{1}{2}\Delta, \frac{1}{2}\Delta; 0) = -\frac{1}{2}\Delta \mathcal{G}^2 \). Hence

(C.2) \[\mathcal{G} = \mathcal{G}2N^{-1} - \mathcal{G}2(\frac{1}{2}V + Z)N^{-1} + o_p(N^{-2+\varepsilon}).\]
D. Taylor Series Expansion for $F(C', C'; (N+1)^{-1}) - F(C', C'; 0)$

where $C' = C(1+N^{-1})^{-\frac{1}{2}}$.

Let $g' = F(C', C'; (N+1)^{-1}) - F(C', C'; 0)$. From Table XIII, $C' = \frac{1}{2}A + A^{-1}(\frac{1}{2}V + Z) + p(N^{-1+e})$, which is equal to $C$ to this order of approximation. Hence the expression for $g'$ given in (C.2) is also the expression for $g'$. 
APPENDIX V

UNCONDITIONAL MEAN SQUARE ERROR EXPRESSIONS NEEDED FOR
THE DISCUSSION OF THE DISTINCTION BETWEEN
ESTIMATORS OF $p_2$, $p_2^*$, AND $p_2^{**}$
(see Chapter 3, Section 4.4).

For the derivations below, $\Pi_1$ is $N(\mu_1, 1)$ and $\Pi_2$ is $N(\mu_2, 1)$.
Notation not defined below is defined in Table I.

A. $E(p_2 - p_2^*)^2$: Exact and Asymptotic Expressions for $N_1 = N_2 = N$.

Because $E p_2 = p_2^*$, $E(p_2 - p_2^*)^2 = E p_2^2 - (p_2^*)^2$. The value of $p_2^*$
is given by (1.8), Chapter 3. We need to compute $E p_2^2$.

By (1.6), Chapter 3, $p_2 = 1 - F(c)$ when $d > 0$, and $p_2 = F(c)$
when $d < 0$ ($d = \bar{x}_1 - \bar{x}_2$, $c = \frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_2$). Hence, letting $g(w)$
denote the density of $d$, and letting $u, v$ denote standard normal
variables, with $u, v, d$ mutually independent,

$$EP_2^2 = \int_{-\infty}^{\infty} [1-F(c)]^2 g(w)dw + \int_{-\infty}^{0} F^2(c)g(w)dw$$

$$= P\{u - \frac{1}{2}x_1 - \frac{1}{2}x_2 + \mu_2 > 0, v - \frac{1}{2}x_1 - \frac{1}{2}x_2 + \mu_2 > 0, \bar{x}_1 - \bar{x}_2 > 0\}$$

$$+ P\{ \quad < 0, \quad < 0, \quad < 0 \}.$$ 

The variables $u - \frac{1}{2}x_1 - \frac{1}{2}x_2 + \mu_2$ and $v - \frac{1}{2}x_1 - \frac{1}{2}x_2 + \mu_2$ are
identically distributed $N(-\frac{1}{2}\delta, 1 + \frac{1}{2}N^{-1})$, and $\bar{x}_1 - \bar{x}_2$ is $N(\delta, 2N^{-1})$.
Furthermore $\rho(u - \frac{1}{2}x_1 - \frac{1}{2}x_2 + \mu_2, v - \frac{1}{2}x_1 - \frac{1}{2}x_2 + \mu_2) = (1 + 2N)^{-1}$,
$\rho(u - \frac{1}{2}x_1 - \frac{1}{2}x_2 + \mu_2, \bar{x}_1 - \bar{x}_2) = 0$. Thus letting $\alpha = \frac{1}{2}\delta(2N)^{\frac{1}{2}}$ and
$\beta = \frac{1}{2}\delta(1 + \frac{1}{2}N^{-1})^{-\frac{1}{2}}$,

(A.1) $EP_2^2 = F(\alpha)F(-\beta, -\beta; (1+2N)^{-1}) + F(-\alpha)F(\beta, \beta; (1+2N)^{-1})$. 

- 150 -
Asymptotic expression:

To get the asymptotic expression for $E(P_2 - P_2^*)^2$, we need the asymptotic expressions for $E(P_2^2)$ and $(P_2^*)^2$. The exact expression for $E(P_2^2)$ is given in (A.1), and from (1.8), Chapter 3,


When $\delta > 0$, $F(-\alpha) = o(N^{-K})$, for any $K > 0$, and thus letting $H(u) = F(-u)$ and $H(u, v; \rho) = F(-u, -v; \rho)$, we can write

(A.3) $E(P_2^2) = \begin{cases} H(\beta, \beta; (1+2N)^{-1}) + o(N^{-2+\epsilon}), \delta > 0 \\ F(\beta, \beta; (1+2N)^{-1}) + o(N^{-2+\epsilon}), \delta < 0 \end{cases}$

and

(A.4) $P_2^* = \begin{cases} H(\beta) + o(N^{-2+\epsilon}), \delta > 0 \\ F(\beta) + o(N^{-2+\epsilon}), \delta < 0. \end{cases}$

Using the expansion of $(1 + \frac{1}{2N})^{-\frac{1}{2}}$ from Table V,

$\beta = \frac{1}{2\delta} - \frac{1}{3\delta}N^{-1} + o(N^{-2+\epsilon})$, and hence with $g = f(\frac{1}{2\delta}), G = F(\frac{1}{2\delta})$,

(A.5) $F(\beta) = G - \frac{1}{3\delta}NgN^{-1} + o(N^{-2+\epsilon})$.

Expanding the bivariate functions first on $\rho = (1 + 2N)^{-1}$ and then expanding the partial derivatives $F_3(\beta, \beta; 0)$ and $H_3(\beta, \beta; 0)$ on $\beta$,

(A.6) $F(\beta, \beta; \rho) = F(\beta, \beta; 0) + \frac{1}{2}F_3(\beta, \beta; 0)N^{-1} + o(N^{-2+\epsilon})$

$= F(\beta, \beta; 0) + \frac{1}{2}F_3(\frac{1}{2\delta}, \frac{1}{2\delta}; 0)N^{-1} + o(N^{-2+\epsilon})$,

and

(A.7) $H(\beta, \beta; \rho) = H(\beta, \beta; 0) + \frac{1}{2}H_3(\frac{1}{2\delta}, \frac{1}{2\delta}; 0)N^{-1} + o(N^{-2+\epsilon})$,

where from (C.11), Appendix III, $F_3(\frac{1}{2\delta}, \frac{1}{2\delta}; 0) = H_3(\frac{1}{2\delta}, \frac{1}{2\delta}; 0) = g^2$.

Hence using (A.3), (A.4), (A.5), (A.6), and (A.7),

(A.8) $E(P_2^2) = \begin{cases} (1-g)^2 + \frac{1}{2}g(1-g)N^{-1} + \frac{1}{2g}2N^{-1} + o(N^{-2+\epsilon}), \delta > 0 \\ G^2 - \frac{1}{2}gG N^{-1} + \frac{1}{2g}2N^{-1} + o(N^{-2+\epsilon}) \end{cases}$
(A.9) \( (p_2^*)^2 = \begin{cases} 
(1-G)^2 + \frac{1}{2} \delta g(1-G)N^{-1} + o(N^{-2+\epsilon}) & , \delta > 0 \\
G^2 - \frac{1}{2} \delta gN^{-1} + o(N^{-2+\epsilon}) & , \delta < 0, 
\end{cases} \)

and for \( \delta > 0 \) or \( \delta < 0, \)

(A.10) \( E(p_2 - p_2^*)^2 = \frac{1}{2} g^2 N^{-1} + o(N^{-2+\epsilon}). \)

B. \( E(p_2 - p_2^{**})^2: \) Exact and Asymptotic Expressions, \( N_1 = N_2 = N. \)

We can write \( E(p_2 - p_2^{**})^2 = EP_2^2 - 2p_2^{**}p_2 + (p_2^{**})^2. \) The exact expressions for \( EP_2^2 \) and for \( p_2^* \) are given in (A.1) and (A.2) above. From (1.9), Chapter 3, \( p_2^{**} = F(-\frac{1}{2} \mid \delta \mid). \)

The asymptotic expression for \( EP_2^2 \) is given in (A.8) and the asymptotic expression for \( p_2^* \) can be obtained from (A.4) and (A.5). Using these expressions we find that

(B.1) \( E(p_2 - p_2^{**})^2 = \frac{1}{2} g^2 N^{-1} + o(N^{-2+\epsilon}). \)

C. Asymptotic Expression for \( E(p_2^* - p_2^{**})^2 = (p_2^* - p_2^{**})^2, N_1 = N_2 = N. \)

From (A.4) and (A.5), \( p_2^* = 1 - G + \frac{1}{8} \delta gN^{-1} + o(N^{-2+\epsilon}) \) when \( \delta > 0, \)
and \( p_2^* = G - \frac{1}{8} \delta gN^{-1} + o(N^{-2+\epsilon}), \) when \( \delta < 0. \) From (1.9), Chapter 3, \( p_2^{**} = 1 - G \) when \( \delta > 0 \) and \( p_2^{**} = G \) when \( \delta < 0. \) Hence

(C.1) \( (p_2^* - p_2^{**})^2 = [\frac{1}{8} \delta gN^{-1} + o(N^{-2+\epsilon})]^2 = (\delta g)^{-1} \delta^2 g^2 N^{-2} + o(N^{-3+\epsilon}). \)

D. \( E(q_D^* - p_2^*)^2: \) Asymptotic Approximation, \( N_1 = N_2 = N. \)

\( q_D^* \) is the estimator obtained by estimating the unknown parameter \( \delta \) in \( p_2^*(\delta) \) by \( \delta = \bar{x}_1 - \bar{x}_2, \) i.e., \( q_D^* = p_2^*(\delta). \) By Taylor series

(D.1) \( p_2^*(\delta) - p_2^*(\delta) = -p_2^*(\delta) + p_2^*(\delta) + \frac{dp_2^*(\delta)}{d\delta}(\delta - \delta) + \frac{d^2p_2^*}{d\delta^2}(\delta - \delta)^2 + \ldots. \)

Thus

(D.2) \( E[p_2^*(\delta) - p_2^*(\delta)]^2 = \frac{d^2p_2^*}{d\delta^2} E(\delta - \delta)^2 + \left(\frac{dp_2^*}{d\delta}\right)^2 E(\delta - \delta)^3 \)
\[ + \frac{1}{2} \left(\frac{d^2p_2^*}{d\delta^2}\right)^2 E(\delta - \delta)^4 + \ldots. \]
When \( N_1 = N_2 = N \), \( \hat{\delta} - \delta \equiv N(0, 2N^{-1}) \) and hence \( E(\hat{\delta} - \delta)^2 = 2N^{-1} \), \( E(\hat{\delta} - \delta)^3 = 0 \), and \( E(\hat{\delta} - \delta)^4 = 3(2N^{-1})^2 \). Substituting in (D.2)

\[
(D.3) \quad E[\hat{P}_2^*(\delta) - P_2^*(\delta)]^2 = 2(b_0^2)2N^{-1} + o(N^{-2+\varepsilon}).
\]

Therefore we now need the expression for \( \frac{dP_2^*}{d\delta} \). By (1.8), Chapter 3, or by (A.2) above,

\[
P_2^*(\delta) = F(\alpha)F(-\beta) + F(-\alpha)F(\beta),
\]

where \( \alpha = \frac{1}{2}\delta(2N)^{1/2} \) and \( \beta = \frac{1}{2}\delta(1 + \frac{1}{2}N^{-1})^{-1/2} \). Taking the derivative

\[
(D.4) \quad \frac{dP_2^*}{d\delta} = \frac{1}{2}F(\alpha)f(\beta)(1 + \frac{1}{2}N^{-1})^{-1} + \frac{1}{2}f(-\beta)f(\alpha)(2N)^{1/2}
\]

\[
+ \frac{1}{2}F(-\alpha)f(\beta)(1 + \frac{1}{2}N^{-1})^{-1/2} - \frac{1}{2}f(\beta)f(-\alpha)(2N)^{1/2}.
\]

Now \( f(\alpha) = f(-\alpha) \) is \( o(N^{-K}) \), for any \( K > 0 \) as \( N \to \infty \). If \( \delta > 0 \), then (by the inequality \( 1 - F(x) \leq \frac{1}{x}f(x) \) for all \( x > 0 \)) \( F(-\alpha) = o(N^{-K}) \), for any \( K > 0 \). Hence, for \( \delta > 0 \) or \( \delta < 0 \), we can write

\[
(D.5) \quad \frac{dP_2^*}{d\delta} = \left[ \frac{1}{2}f(\beta)(1 + \frac{1}{2}N^{-1})^{-1/2} + o(\varepsilon) \right]^2.
\]

From Table V, \( (1 + \frac{1}{2}N^{-1})^{-1/2} = 1 - \frac{1}{4}N^{-1} + o(N^{-2+\varepsilon}) \) and thus

\[
\beta = \frac{1}{2}\delta - \frac{1}{8}N^{-1} + o(N^{-2+\varepsilon}).
\]

Then by Taylor series,

\[
f(\beta) = 1 + (16)^{-1}\delta^2 gN^{-1} + o(N^{-2+\varepsilon}).
\]

Thus

\[
\frac{dP_2^*}{d\delta} = \left[ \frac{1}{2} + o(N^{-1+\varepsilon}) \right]^2 = \frac{1}{4} + o(N^{-1+\varepsilon}),
\]

and using (D.3)

\[
(D.6) \quad E[\hat{P}_2^*(\delta) - P_2^*(\delta)]^2 = \frac{1}{2}2N^{-1} + o(N^{-2+\varepsilon}).
\]

E. \( E(\hat{Q}_D - P_2^*)^2 \): Asymptotic Expression, \( N_1 = N_2 = N \).

The estimator \( Q_D \) is equal to \( F(-\frac{1}{2}|d|) \), \( d = \bar{x}_1 - \bar{x}_2 \). When \( \bar{x}_1 = \bar{x}_2 \)
\[ N_1 = N_2 = N, \ d = N(\delta, 2N^{-1}). \] Thus setting \( z = (d - \delta)N^{2-\frac{1}{2}}, \)

\[ \mathbb{E} \mathbb{F}(\frac{1}{2}d) = \int F(z(2N)^{-\frac{1}{2}} + \frac{\lambda}{2})f(z)dz = F(\beta) \]

and

\[ \mathbb{E}^2 \mathbb{F}(\frac{1}{2}d) = F(\beta, \beta; (1 + 2N)^{-1}), \]

where \( \beta = \frac{1}{2}(1 + \frac{1}{2}N^{-1})^{-\frac{1}{2}}. \) From (1.9), Chapter 3, \( P_2^{**} = F(-\frac{1}{2}|\delta|). \)

Hence assuming \( d \) is of the same sign as \( \delta(d \text{~as~} N \rightarrow \infty), \) for \( \delta > 0 \) or \( \delta < 0 \)

\[ (E.1) \]

\[ \mathbb{E}(Q_D - P_2^{**})^2 = EQ_D^2 - 2P_2^{**}EQ_D + (P_2^{**})^2 \]

\[ = F(\beta, \beta; (1 + 2N)^{-1}) - 2GF(\beta) + G^2, \]

where \( G = F(\frac{1}{2}\delta). \) The expansion for \( F(\beta) \) is given in (A.5) and the expression given in (A.8) for \( \mathbb{E}P_2^{2}, \delta < 0, \) is the expansion for \( F(\beta, \beta; (1 + 2N)^{-1}). \) Hence

\[ (E.2) \]

\[ \mathbb{E}(Q_D - P_2^{**})^2 = \frac{1}{2}g^2N^{-1} + o(N^{-2+\epsilon}). \]
APPENDIX VI

THE PROBLEM OF A BOUND ON THE ASYMPTOTIC CONDITIONAL SQUARED BIAS.

In the discussion of the asymptotic results for the univariate normal case (Chapter 3, Section 4.3, especially the third conclusion) it was pointed out that the estimators $Q_D, Q_D^*, Q_D^{**}, Q_3, Q_5, P_R, P_{Ra}, P_{TSL}$ and $P_U$ all have asymptotic conditional squared bias equal to $\theta^2 \alpha_2^2 + o_p(N^{-3/2} + \epsilon)$ (see Table I for notation), suggesting that this value is a lower bound for estimators based on original samples. Also in the multivariate normal case, several estimators based on original samples had the same leading term in the asymptotic conditional squared bias (see Chapter 4, Section 5.1). Below we make some remarks on the problem of either finding an estimator with smaller asymptotic bias or alternatively proving that no such estimator exists. We confine our attention to the univariate case. (See Table I for notation not defined below.)

Let $\Pi_1$ be $N(\mu_1, 1)$ and $\Pi_2$ be $N(\mu_2, 1)$. Let samples $S_1, S_2$ each of size $N$ be taken from $\Pi_1, \Pi_2$ and let $\bar{x}_1, \bar{x}_2$ be the sample means. From (1.6), Chapter 3 the conditional probability of misclassification is

$$P_2 = \begin{cases} 1-F(c), & d > 0 \\ F(c), & d < 0 \end{cases}$$

where $d = \bar{x}_1 - \bar{x}_2$ and $c = \frac{1}{2}d + \bar{x}_2 - \mu_2$. From Table V, the asymptotic expansion for $F(c)$ is

$$F(c) = G + \frac{1}{2}gv - (16)^{-1}6gv^2 + o_p(N^{-3/2} + \epsilon),$$

where $g = f(\frac{1}{2}\delta), G = F(\frac{1}{2}\delta), v = \alpha_1 + \alpha_2, \alpha_i = \bar{x}_i - \mu_i, i = 1, 2$. Note that by the lemma of Appendix III $B, v = o_p(N^{-1+\epsilon})$ and thus $v^2 = o_p(N^{-1+\epsilon})$. 

- 155 -
From (1) and (2) we conclude that if an estimator \( \hat{P}_2 = P_2(S_1, S_2) \) is to have asymptotic conditional squared bias with leading term of order \( o_p(N^{-1}) \), instead of order \( o_p(N^{-1+\epsilon}) \) which is the order of \( g^2\alpha_2^2 \), it is necessary and sufficient that

\[
E(\hat{P}_2 | \bar{x}_1, \bar{x}_2) = \begin{cases} 
1 - G - \frac{1}{2}\sigma_v + o_p(N^{-\frac{1}{2}}), & d > 0 \\
G + \frac{1}{2}\sigma_v + o_p(N^{-\frac{1}{2}}), & d < 0.
\end{cases}
\]

We could also consider the conditions necessary for the conditional squared bias of \( \hat{P}_2 \) to have leading term of order \( o_p(N^{-1+\epsilon}) \) but smaller than \( g^2\alpha_2^2 \), but will here only look at the problem of reducing the order of the leading term.

Assume that there exists \( \hat{P}_2 = \hat{P}_2(S_1, S_2) \) such that (3) holds. We will try to show that this assumption leads to a contradiction, and hence we can conclude that no such \( \hat{P}_2(S_1, S_2) \) exists. If (3) holds then

\[
E(\hat{P}_2 | \bar{x}_1, \bar{x}_2) = \begin{cases} 
F(-\frac{1}{2}\delta - \frac{1}{2}\sigma_v + o_p(N^{-\frac{1}{2}})), & d > 0 \\
F(\frac{1}{2}\delta + \frac{1}{2}\sigma_v + o_p(N^{-\frac{1}{2}})), & d < 0.
\end{cases}
\]

Taking \( F^{-1} \) of both sides and substituting \( \frac{1}{2}\delta = \mu_1 - \mu_2, \sigma_v = \alpha_1 + \alpha_2 \)

\[
F^{-1}[E(\hat{P}_2 | \bar{x}_1, \bar{x}_2)] = \begin{cases} \frac{1}{2}(\bar{x}_1 + \bar{x}_2) + \mu_2 + o_p(N^{-\frac{1}{2}}), & d > 0 \\
\frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \mu_2 + o_p(N^{-\frac{1}{2}}), & d < 0.
\end{cases}
\]

By the sufficiency of \( (\bar{x}_1, \bar{x}_2) \) for \( (\mu_1, \mu_2) \), \( E(\hat{P}_2 | \bar{x}_1, \bar{x}_2) \) is independent of \( \mu_1 \) and \( \mu_2 \) and thus so is \( F^{-1}[E(\hat{P}_2 | \bar{x}_1, \bar{x}_2)] \). Hence (5) says that there is a function \( h = h(\bar{x}_1, \bar{x}_2) \) such that

\[
h(\bar{x}_1, \bar{x}_2) - \mu_2 = o_p(N^{-\frac{1}{2}})
\]

where for \( d > 0 \), \( h = F^{-1}[E(\hat{P}_2 | \bar{x}_1, \bar{x}_2)] + \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \), and for \( d < 0 \),

\( h = -F^{-1}[E(\hat{P}_2 | \bar{x}_1, \bar{x}_2)] + \frac{1}{2}(\bar{x}_1 + \bar{x}_2) \).
If we can show that no \( h(x_1, x_2) \) exists that satisfies (6), we have the desired contradiction. We know from Appendix IIIB that 
\[ \bar{x}_2 - \mu_2 = o_p(N^{-1/2} + e), \]
and since \( \bar{x}_2 \) is the "best" estimator of \( \mu_2 \) by various criteria, it is intuitively plausible that no \( h(x_1, x_2) \) satisfying (6) exists or perhaps that no such \( h \) from among some subclass of functions of \( x_1 \) and \( x_2 \) exists.

Thus from Pitman (1939, p. 401) we know that no invariant estimator \( h(s_2) \) exists, where \( s_2 \) indicates the original sample from \( \Pi_2 \), such that \( h(s_2) - \mu_2 \) is smaller, in the \( o_p(\cdot) \) sense, than \( \bar{x}_2 - \mu_2 \).

Presumably letting \( h \) depend also on \( x_1 \) (or \( s_1 \)) will not lead to a better estimator since \( s_1 \) is independent of \( \mu_2 \).

Furthermore if \( h(x_1, x_2) \) is the same function for all \( N \), then a well-known theorem (Wilks (1962), p. 260) establishes that \( h \) is asymptotically normal with mean \( h(\mu_1, \mu_2) \) and variance proportional to \( N^{-1} \), the constant depending on the partial derivatives of \( h \) at \( (\mu_1, \mu_2) \). Thus \( h(x_1, x_2) - \mu_2 \overset{p}{\to} 0 \) in probability unless \( h(\mu_1, \mu_2) \equiv \mu_2 \), in which case \( h(x_1, x_2) = \bar{x}_2 \). Therefore if \( h(x_1, x_2) \neq \bar{x}_2 \) then \( h - \mu_2 \) cannot even be \( o_p(1) \). On the other hand if \( h(x_1, x_2) = h_N(\bar{x}_1, \bar{x}_2) \), for example \( h = \frac{N}{N+1} \bar{x}_2 \), then this argument fails. (Note that the Pitman argument also does not apply to the particular \( h \), \( h = \frac{N}{N+1} \bar{x}_2 \), because this estimator is not invariant.)

The problem of finding the largest class of \( h \) functions for which (6) cannot be satisfied and the alternative problem of finding an estimator \( P_2 \) which does satisfy (3) are unsolved.

\[ ^\dagger \text{An estimator } A \text{ based on } (x_1, \ldots, x_N) \text{ is invariant if} \]
\[ A(x_1 + \lambda, \ldots, x_N + \lambda) \equiv A(x_1, \ldots, x_N) + \lambda. \]
BIBLIOGRAPHY

A. This section lists work referenced in the text of the paper.


Stoller, David S. (1954). Univariate two-population distribution-free


Normal Distribution Function and Related Functions. Applied
Mathematics Series No. 50.

Van Ryzin, J. (1965). Non-parametric Bayesian decision procedures
for (pattern) classification with stochastic learning. Transactions
of the Fourth Prague Conference on Information Theory, Statistical
Decision Functions and Random Processes.

Van Ryzin, J. (1966). A stochastic a posteriori updating algorithm
for pattern recognition. Technical Report No. 121, Dept. of
Statistics, Stanford University, Stanford, California.


B. Some classification references since 1962.

multivariate normal distributions with different covariance matrices.


Bhattacharya, P. K. and Das Gupta, S. (1964). Classification between
univariate exponential populations. Sankhya A 26 17-24.

Cacoullos, Theophilos (1961). Comparing distances between multi­

Cacoullos, Theophilos (1962a). Comparing distances between multi­

Cacoullos, Theophilos (1962b). On the distribution of bilinear forms

Cochran, W. G. (1964a). Comparison of two methods of handling covariates

function. Technometrics 6 179-190.


Lachenbruch, P. A. (1966). Discriminant analysis when the initial samples are misclassified. Technometrics 8 657-662.


