Errors of Misclassification Associated with Gamma Distribution

M. A. W. MAHMOUD
Faculty of Science, Department of Mathematics
Al-Azhar University, Nasr City, Cairo, Egypt

H. M. MOUSTAFA
Faculty of Science, Department of Mathematics
Assiut University, Assiut, Egypt

(Received May 1993; revised and accepted June 1994)

Abstract—Errors of misclassification and their probabilities are studied for classification problems associated with two classes of univariate gamma distribution. The effects of applying the normal classificatory rule to gamma populations are also studied and assessed by comparing probabilities (optimum and conditional) based on the linear discriminant function (LDF) for normality with those based on the likelihood ratio rule (LR) for gamma populations. To determine the form in the one-variable case, formulas for the exact distribution and density functions of the actual error rates are presented. Both theoretical and empirical results are considered.

Keywords—Errors of misclassification, Gamma distribution, Linear discriminant function, Likelihood ratio rule, Error rate.

1. INTRODUCTION

In classification problems, the linear discriminant function (LDF) has many desirable properties when used in classifying an observation which belongs to one of two normal populations. Most of the work assumes the parent populations have multivariate normal distributions. In the univariate case (denoted by $N(\mu, \sigma^2)$), the classification problem has been studied by John [1] and Sedransk and Okamoto [2].

The effects of several types of non-normality have been studied. Lachenbruch et al. [3] examined the robustness of the LDF when the underlying distribution belongs to Johnson’s system. Ching’anda [4] and Ching’anda and Subrohmaniam [5] further considered this problem and derived the distributions (based on large samples) for both conditional and unconditional probabilities of misclassification. Similar studies have been done for the inverse Gaussian distributions by Amoh and Kocherlakota [6].

In this paper, we will consider the errors of misclassification and their probabilities, when we have two classes, and sampling from gamma distributions. The distributions of probabilities of misclassification are studied. An observation $X$ may be originate and the parameters $\lambda_1, \lambda_2$ may be known or unknown; the same shape parameter $\theta$ is assumed to be known, from one of two gamma populations

$$f_i(x) = \frac{\lambda_i^\theta}{\Gamma(\theta)} x^{\theta-1} \exp(-\lambda_i x), \quad i = 1, 2.$$  \hspace{1cm} (1.1)

Both authors would like to thank A. H. El-Mawaziny for his useful comments. The authors would also like to thank the reviewer for his or her valuable remarks and comments.
We make a reparameterization for the density function in (1.1) to become in the form

\[ f_i(x, \theta, \mu_i) = \frac{(\theta/\mu_i)^\theta}{\Gamma(\theta)} x^{\theta-1} \exp \left( -\frac{\theta x}{\mu_i} \right), \quad \mu_i, \theta > 0, \]  

(1.2)

with \( E(x) = \mu_i \) and \( V(x) = \mu_i^2 / \theta \), where \( \lambda_i = \theta/\mu_i \) (i = 1, 2).

We will examine the robustness of the LDF in two ways:

(i) Supposing that in classifying an observation \( X \) from (1.2), we use the LDF derived under the assumption of normality, how are the optimum (based on all parameters being known) and conditional probabilities of misclassification affected?

(ii) Optimum and conditional probabilities of misclassification based on the likelihood ratio rule will be compared with those obtained.

2. CLASSIFICATION RULES

The general solution to the classification rule is to minimize the total probability of misclassification [7]. Suppose that \( f_i(x) \) is the density function of \( X \) if it comes from the population \( \Pi_i \) (i = 1, 2) and we assign \( X \) to \( \Pi_1 \) if \( X \) is in some region \( R_1 \) and to \( \Pi_2 \) if \( X \) is in some region \( R_2 \). We assume \( R_1 \cap R_2 = \emptyset, R_1 \cup R_2 = R \). Let \( P_i(i = 1, 2) \) be the proportion (Bayes assumption) of population \( \Pi_i \), \( P_1 + P_2 = 1 \). The total probability of misclassification is

\[ E = P_1 \int_{R_2} f_1(x) \, dx + P_2 \int_{R_1} f_2(x) \, dx \]

\[ = P_1 + \int_{R_1} \left[ P_2 f_2(x) - P_1 f_1(x) \right] \, dx, \]  

(2.1)

where \( E \) is minimized (Neyman-Pearson Lemma) if \( R_1 \) includes the points \( X \) such that \( [P_2 f_2 - P_1 f_1] < 0 \) and excludes the points for which \( [P_2 f_2 - P_1 f_1] > 0 \). Thus, the classification rule is

\[ R_1: \frac{f_1}{f_2} \geq \frac{P_2}{P_1}, \]

\[ R_2: \frac{f_1}{f_2} < \frac{P_2}{P_1}. \]  

(2.2)

In what follows, we will assume \( P_1 = P_2 = 1/2 \). It is well known that if \( P_1 = P_2 \) and \( f_1(x) \) is univariate normal, this classification rule given in (2.2) is equivalent to Fisher's linear discriminant function [8].

2.1. Linear Discriminant Function for the Univariate Normal Distribution (known \( \mu_1 \neq \mu_2 \), and the Same Variance; \( \sigma^2 \))

If we assume that the distribution of \( X \) in \( \Pi_i \) is univariate normal, then it is well known (if \( \sigma^2 \) is the same) that

\[ R_1 = \{ X; x > A \text{ if } \mu_1 > \mu_2 \text{ or } X \leq A \text{ if } \mu_1 \leq \mu_2 \}, \]

where

\[ A = \frac{1}{2} (\mu_1 + \mu_2). \]  

(2.3)

2.2. Likelihood Ratio Rule (LR)

Now assume that the distributions of \( X \) in \( \Pi_i \) is given by (1.2); that is, \( X \sim \mathcal{G}(\theta, \theta/\mu_i) \). Then the classification rule is

\[ R_1 = \{ X; x \geq B \text{ if } \mu_1 \geq \mu_2 \text{ or } X < B \text{ if } \mu_1 < \mu_2 \}, \quad R_2 = R_1^c, \]  

(2.4)
where
\[ B = \frac{\mu_1 \mu_2}{\mu_1 - \mu_2} \ln \left( \frac{\mu_1}{\mu_2} \right). \] (2.5)

When the parameters \( \mu_1 \) and \( \mu_2 \) are not known, we obtain the classificatory rule by minimizing
\[ E = P_1 \int_{R_1} f_1(x) \, dx + P_2 \int_{R_2} f_2(x) \, dx, \] (2.6)
where the \( R_1 \) and \( R_2 \) are the new classificatory regions based on the samples. \( R_1 \) and \( R_2 \) are given in (2.4), (2.5), \( \mu_1 \) and \( \mu_2 \) replaced by their maximum likelihood estimators.

Using classification procedures, two kinds of errors are possible:
(i) we may classify an individual from \( \Pi_1 \) as belonging to \( \Pi_2 \), or
(ii) we may classify an individual from \( \Pi_2 \) as belonging to \( \Pi_1 \).

We shall study these probabilities.

3. OPTIMUM PROBABILITIES OF MISCLASSIFICATION

Let \( E_{12}(\mu_1, \mu_2) \) be the optimum probability of assigning an individual from \( \Pi_1 \) to \( \Pi_2 \), and \( E_{21}(\mu_1, \mu_2) \) be the optimum probability of assigning an individual from \( \Pi_2 \) to \( \Pi_1 \). Using the LDF classification rule given by (2.3), we have
\[ E_{12} = P\{X < A \mid X \in \Pi_1; \mu_1, \mu_2\}, \quad \text{if } \mu_1 > \mu_2. \] (3.1)

\( E_{21}(\mu_1, \mu_2) \) is similarly defined.

The cumulative distribution function of gamma distribution with parameters \( \mu_i \) and \( \theta \) (\( \mu_i = \mu_i/\lambda_i \)) is given by
\[ P(\leq x) = F(x; \theta, \mu) = \frac{\gamma(\theta, \mu x)}{\Gamma(\theta)} = P(\theta, \mu x), \] (3.2)
where
\[ \gamma(\theta, \mu x) = \sum_{n=0}^{\infty} \frac{(-1)^n (\mu x)^{\theta+n}}{n! (\theta+n)} \] (see [9]). Then using (3.2) we have
\[ E_{12} = F(A; \theta, \mu_1), \quad \mu_1 > \mu_2, \quad A = \frac{\mu_1 + \mu_2}{2}. \] (3.3)

The total optimum of probability of misclassification using LDF is
\[ E = \frac{(E_{12} + E_{21})}{2}, \quad \text{for } P_1 = P_2 = 0.5. \] (3.4)

Let \( E^*_{12}(\mu_1, \mu_2) \) be the probability of assigning an individual from \( \Pi_1 \) to \( \Pi_2 \) and denote by \( E^*_{21} \) the probability of assigning an individual from \( \Pi_2 \) to \( \Pi_1 \), when we are using the LR rule given in (2.5), which is defined by
\[ E^*_{12} = P\{X < B \mid X \in \Pi_1\} = F(B; \theta, \mu_1), \quad \text{if } \mu_1 > \mu_2, \]
\[ B = \frac{\mu_1 \mu_2}{\mu_1 - \mu_2} \ln \left( \frac{\mu_1}{\mu_2} \right). \]

The total probability of misclassification is
\[ E^* = \frac{(E^*_{12} + E^*_{21})}{2}, \quad \text{for } P_1 = P_2 = 0.5. \]

The probabilities of misclassification, based on the LDF and LR rules, for various combinations of the parameters \( \lambda_1, \lambda_2 \) and \( \theta \) are given in Table 1. The values of the parameters chosen are \( \lambda_1 = 1.0, \lambda_2 = 2.0, 3.0, 4.0 \) and \( \theta = 2.0, 3.0, 4.0, 5.0 \). Gamma distribution is a useful model for data that arise from a positively skewed long tailed distribution; it approaches normality as \( \theta \) increases.

Table 1 gives a comparison between the LDF and LR classification rules when all the parameters are known. The results are obtained by evaluating the formulas given above.
Table 1. Comparison of the probabilities of misclassification based on the LDF and LR.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>LDF</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E_{12}$</td>
<td>$E_{21}$</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>$\lambda_2$</td>
<td>$\theta$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

4. CONDITIONAL PROBABILITIES OF MISCLASSIFICATION

When the population means $E_1X$ and $E_2X$ are unknown, their maximum likelihood estimates are used for formulating the classification rules based on the LDF or LR. Mahmoud and Moustafa [10] described procedures for finding maximum likelihood estimates of the parameters of mixture of two gamma distributions. They considered the estimation of discriminant function on the basis of a small sample size. It is difficult to obtain the distribution of the conditional errors of misclassification explicitly; this is due to the complex nature of the expression.

To investigate the robustness of the LDF when the distributions of $\Pi_1$ and $\Pi_2$ are gamma distributions and $\mu_1, \mu_2$ are unknown, we use GGAMR routine from IMSL for generating gamma random numbers. Using this procedure, two independent training samples $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ are generated from gamma $G(\theta, \mu_1)$ and $G(\theta, \mu_2)$, respectively. These are used to construct the classification rules (2.3) and (2.4) with the unknown parameters replaced by their estimates. Two further independent samples of sizes $N_1$ and $N_2$, $Y_{11}, Y_{12}, \ldots, Y_{1N_1}$ and $Y_{21}, Y_{22}, \ldots, Y_{2N_2}$ are generated from gamma $G(\theta, \mu_1)$ and $G(\theta, \mu_2)$, respectively.

Using the rules constructed from the training samples, these latter samples are reclassified into $\Pi_1$ or $\Pi_2$. To estimate the probabilities of misclassification, we define

$$e_{12} = \frac{1}{N_1} \sum_{j=1}^{N_1} \delta_j, \quad \delta_j = 1$$

if $Y_{1j} \in \Pi_1$ and is classified to $\Pi_2$. Similarly, we define

$$e_{21} = \frac{1}{N_2} \sum_{j=1}^{N_2} \gamma_j, \quad \gamma_j = 1$$

if $Y_{2j} \in \Pi_2$ and is classified to $\Pi_1$.

Tables 2 and 3 show the probabilities of conditional errors of misclassification obtained by simulation with $\lambda_1$ and $\lambda_2$ replaced by their estimates based on training samples of sizes $n_1 = n_2 = 20$ and 50.

The conditional errors of misclassification are similar in magnitude and nature to those in Table 1. In Table 4, the means and variances based on formulas (5.7), (5.8), (5.9) and (5.10) have been tabulated for $n_1 = n_2 = 10, 30, 50$ and 500.
Table 2. Comparison of the conditional probabilities of misclassification when $X_1$ and $X_2$ are unknown. Table entries represent average over 50 samples for simulation size 100. The estimates of $\lambda_1$ and $\lambda_2$ are based on samples of size 20.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>LDF</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$ $\lambda_2$ $\theta$ $c_{12}$ $c_{21}$ $\text{Var}(c_{12})$ $\text{Var}(c_{21})$ $c_{12}^{<em>}$ $c_{21}^{</em>}$ $\text{Var}(c_{12}^{<em>})$ $\text{Var}(c_{21}^{</em>})$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 2</td>
<td>.5440</td>
<td>.1234</td>
</tr>
<tr>
<td>3</td>
<td>.4838</td>
<td>.1086</td>
</tr>
<tr>
<td>4</td>
<td>.9744</td>
<td>.0808</td>
</tr>
<tr>
<td>5</td>
<td>.4400</td>
<td>.0716</td>
</tr>
<tr>
<td>3 2</td>
<td>.4829</td>
<td>.0454</td>
</tr>
<tr>
<td>3</td>
<td>.4110</td>
<td>.0314</td>
</tr>
<tr>
<td>4</td>
<td>.3872</td>
<td>.0152</td>
</tr>
<tr>
<td>5</td>
<td>.3486</td>
<td>.0100</td>
</tr>
<tr>
<td>4 2</td>
<td>.4499</td>
<td>.0164</td>
</tr>
<tr>
<td>3</td>
<td>.3718</td>
<td>.0060</td>
</tr>
<tr>
<td>4</td>
<td>.3452</td>
<td>.0028</td>
</tr>
<tr>
<td>5</td>
<td>.3020</td>
<td>.0010</td>
</tr>
</tbody>
</table>

Table 3. Comparison of the conditional probabilities of misclassification when $X_1$ and $X_2$ are unknown. Table entries represent average over 50 samples for simulation size 100. The estimates of $\lambda_1$ and $\lambda_2$ are based on samples of size 50.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>LDF</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1$ $\lambda_2$ $\theta$ $c_{12}$ $c_{21}$ $\text{Var}(c_{12})$ $\text{Var}(c_{21})$ $c_{12}^{<em>}$ $c_{21}^{</em>}$ $\text{Var}(c_{12}^{<em>})$ $\text{Var}(c_{21}^{</em>})$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 2 2</td>
<td>.4784</td>
<td>.1550</td>
</tr>
<tr>
<td>3</td>
<td>.4520</td>
<td>.1266</td>
</tr>
<tr>
<td>4</td>
<td>.3982</td>
<td>.1130</td>
</tr>
<tr>
<td>5</td>
<td>.3926</td>
<td>.0920</td>
</tr>
<tr>
<td>3 2</td>
<td>.4172</td>
<td>.0656</td>
</tr>
<tr>
<td>3</td>
<td>.3786</td>
<td>.0354</td>
</tr>
<tr>
<td>4</td>
<td>.3166</td>
<td>.0284</td>
</tr>
<tr>
<td>5</td>
<td>.3010</td>
<td>.0174</td>
</tr>
<tr>
<td>4 2</td>
<td>.3856</td>
<td>.0254</td>
</tr>
<tr>
<td>3</td>
<td>.3404</td>
<td>.0128</td>
</tr>
<tr>
<td>4</td>
<td>.2786</td>
<td>.0060</td>
</tr>
<tr>
<td>5</td>
<td>.2600</td>
<td>.0028</td>
</tr>
</tbody>
</table>

5. ASYMPTOTIC DISTRIBUTION OF THE CONDITIONAL PROBABILITIES OF MISCLASSIFICATION

When the sample sizes are large, we can obtain the asymptotic distribution of the error of misclassification by using Taylor's expansion. The approximated mean and variance of the error of misclassification and associated probabilities can be evaluated.

Following [6], we can write the asymptotic distribution as a univariate normal distribution with different parameters depending on whether $X_1 > X_2$ or $X_1 < X_2$. We use the second approach similar to the one used in [5]. Probability distribution of the error of misclassification is given as the sum of two bivariate normal distributions. Because of symmetry, we shall only consider $e_{12}(\bar{X}_1, \bar{X}_2)$ and $e_{12}^{*}(\bar{X}_1, \bar{X}_2)$.

5.1. Asymptotic Expansion of the Conditional Error Probabilities

Neglecting terms containing partial derivatives of order higher than the second, we can expand $e_{12}(\bar{X}_1, \bar{X}_2)$ in a Taylor series form around $(\mu_1, \mu_2)$ as follows:
Table 4. Asymptotic means and variances of $e_{12}$ and $e^*_1$.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$n_1 = n_2 = 10$</th>
<th>$n_1 = n_2 = 30$</th>
<th>$n_1 = n_2 = 50$</th>
<th>$n_1 = n_2 = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_1$</td>
<td>$\lambda_2$</td>
<td>$\theta$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_1$</td>
<td>$\hat{\theta}_2$</td>
<td>$E(e_{12})$</td>
<td>$\text{Var}(e_{12})$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>.4386</td>
<td>.0068</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>.3825</td>
<td>.0067</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>.3233</td>
<td>.0061</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>.2801</td>
<td>.0054</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>.4414</td>
<td>.0013</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>.3903</td>
<td>.0013</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>.3527</td>
<td>.0012</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>.3226</td>
<td>.0011</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>.3544</td>
<td>.0013</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>.3233</td>
<td>.0012</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>.2788</td>
<td>.0010</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>.2440</td>
<td>.0009</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>.3549</td>
<td>.0013</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>.2896</td>
<td>.0011</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>.2428</td>
<td>.0009</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>.2068</td>
<td>.0008</td>
</tr>
</tbody>
</table>

For simplicity, we introduce the following notations:

$$e_{12}(\bar{X}_1, \bar{X}_2) = e_{12}(\bar{X}_1, \bar{X}_2)$$

(5.1)

Under the LDF classification rule, and

$$e_1 = e_{12}(\bar{X}_1, \bar{X}_2) = F(\alpha, \theta, \mu_1) \quad \text{with} \quad \alpha = \frac{\bar{X}_1 + \bar{X}_2}{2}$$

(5.2)

Under the LR classification rule. Then the asymptotic expansion of $e_1$ and $e_2$ become

$$e_1 - K_1 + K_2 [(\bar{X}_1 - \mu_1) + (\bar{X}_2 - \mu_2)] + \frac{1}{2} K_3 \left[ (\bar{X}_1 - \mu_1) + (\bar{X}_2 - \mu_2) \right]^2 + O\left(\frac{1}{n}\right),$$

(5.3)
\[ e_2 = K_1^* + K_2^* [ (\bar{x}_1 - \mu_1) + C (\bar{x}_2 - \mu_2) ] + \frac{1}{2} K_3^* [ (\bar{x}_1 - \mu_1)^2 + K_4^* [ (\bar{x}_1 - \mu_1) (\bar{x}_2 - \mu_2) ] + \frac{1}{2} K_2^* [ (\bar{x}_2 - \mu_2)^2 + O \left( \frac{1}{n} \right), \tag{5.5} \]

where

\[ K_1 = e_1 (\bar{x}_1, \bar{x}_2) |_{(\mu_1, \mu_2)}, \]

\[ K_2 = \frac{1}{2} f (\alpha, \theta, \mu_1) |_{(\mu_1, \mu_2)}, \quad \text{where } f \text{ in (1.2)}, \]

\[ K_3 = \left( \frac{1}{2} \right)^2 \left( \frac{\theta - 1}{\alpha} - \frac{\theta}{\mu_1} \right) f (\alpha, \theta, \mu_1) |_{(\mu_1, \mu_2)}, \]

\[ K_4^* = e_2 (\bar{x}_1, \bar{x}_2) |_{(\mu_1, \mu_2)}, \]

\[ K_5^* = \left( \frac{\bar{x}_2}{(\bar{x}_1 - \bar{x}_2)^2} - \frac{\bar{x}_2^2}{(\bar{x}_1 - \bar{x}_2)^2} \ln \left( \frac{\bar{x}_1}{\bar{x}_2} \right) \right) f (\beta, \theta, \mu_1) |_{(\mu_1, \mu_2)}, \]

\[ K_6^* = \left( \frac{2 \bar{x}_2}{(\bar{x}_1 - \bar{x}_2)^3} \ln \left( \frac{\bar{x}_1}{\bar{x}_2} \right) - \frac{\bar{x}_2^2}{(\bar{x}_1 - \bar{x}_2)^2} - \frac{\bar{x}_2}{(\bar{x}_1 - \bar{x}_2)^2} \ln \left( \frac{\bar{x}_1}{\bar{x}_2} \right) \right)^2 \]

\[ K_7^* = \left( \frac{\bar{x}_1}{(\bar{x}_1 - \bar{x}_2)^2} - \frac{2 \bar{x}_1 \bar{x}_2}{(\bar{x}_1 - \bar{x}_2)^3} \ln \left( \frac{\bar{x}_1}{\bar{x}_2} \right) - \frac{\bar{x}_1^2}{(\bar{x}_1 - \bar{x}_2)^2} \ln \left( \frac{\bar{x}_1}{\bar{x}_2} \right) \right) \]

\[ C = 1 + (\mu_1 / \mu_2) \ln (\mu_1 / \mu_2) - \mu_1 / \mu_2 \]

since \( M_X(t) = |M_X(t/n)|^n, M_X(t) = E(e^{tX}), \) we have

\[ E(\bar{X}) = \mu, \]

\[ E(\bar{X}^2) = \mu^2 \left( 1 + \frac{1}{n \theta} \right), \]

\[ E(\bar{X}^3) = \mu^3 \left( 1 + \frac{3}{n \theta} + \frac{2}{n^2 \theta^2} \right), \]

\[ E(\bar{X}^4) = \mu^4 \left( 1 + \frac{6}{n \theta} + \frac{11}{n^2 \theta^2} \right) + \mu^4 \left( \frac{1}{\beta} + \frac{3}{\theta^2} + \frac{8}{\theta^3} \right). \]

Taking the expectation of the expansion of \( e_1 \) and \( e_2 \) (in (5.4) and (5.5)), we have

\[ E(e_1) = K_1 + \frac{K_3}{2 \theta} \left( \frac{\mu_1^2}{n_1} + \frac{\mu_2^2}{n_2} \right) + O \left( \frac{1}{n} \right), \tag{5.7} \]

\[ E(e_2) = K_4^* + \frac{1}{2 \theta} \left( K_5^* \frac{\mu_1^2}{n_1} + K_5^* \frac{\mu_2^2}{n_2} \right) + O \left( \frac{1}{n} \right). \tag{5.8} \]
Also the variances of $e_1$ and $e_2$ are approximately given by

$$V(e_i) = \frac{K^2_2}{\theta} \left( \frac{\mu_1^2}{n_1} + \frac{\mu_2^2}{n_2} \right) + \frac{K^2_4}{4} \left\{ \left( \frac{1}{\theta} + \frac{3}{\theta^2} + \frac{8}{\theta^3} \right) \left( \frac{\mu_1^4}{n_1^2} + \frac{\mu_2^4}{n_2^2} \right) \right\},$$

$$V(e_2) = \frac{K^2_2}{\theta} \left( \frac{\mu_1^2 + C^2 \mu_2^2}{n_2} \right) + \frac{K^2_4}{4} \frac{K^2_1 \mu_1^2 \mu_2^2}{n_1 n_2} + \frac{1}{2 \theta^2} \left( K^2_3 \frac{\mu_1^4}{n_1^2} + K^2_5 \frac{\mu_2^4}{n_2^2} \right)$$

$$+ \frac{1}{4} \left( \frac{1}{\theta} + \frac{3}{\theta^2} + \frac{8}{\theta^3} \right) \left( K^2_3 \frac{\mu_1^2}{n_1^2} + K^2_5 \frac{\mu_2^2}{n_2^2} \right)$$

$$+ \frac{4}{\theta^2} K^2_2 \left( K^2_3 \frac{\mu_1^2}{n_1^2} + K^2_5 \frac{\mu_2^2}{n_2^2} \right).$$

### 5.2. Asymptotic Distribution Using Univariate Approach

As $n_1$ and $n_2$ become large, the distribution of

$$\frac{e_i - E(e_i)}{\sqrt{\text{Var}(e_i)}}, \quad i = 1, 2,$$

is approximately normal with mean zero and variance unity. Ignoring terms of order $n^{-2}$ in the expansion of (5.9), (5.10), we obtain for $\bar{X}_1 > \bar{X}_2$

$$F_1(Z) = P(e_1 \leq Z) = \Phi \left[ \frac{1}{K_2 S} \left( Z - K_1 - \frac{K_3 S^2}{2} \right) \right],$$

where $\Phi$ is the standard normal cdf with $S^2 = (1/\theta) (\mu_1^2/n_1 + \mu_2^2/n_2)$ and

$$F'_1(Z) = P(e_2 \leq Z) = \Phi \left[ \frac{1}{K_2 S} \left( Z - K'_1 - S'^2 \right) \right].$$

with

$$S'^2 = \frac{1}{\theta} \left( \frac{\mu_1^2}{n_1} + C^2 \frac{\mu_2^2}{n_2} \right), \quad S^2 = \frac{1}{2 \theta} \left( K^2_3 \frac{\mu_1^2}{n_1} + K^2_5 \frac{\mu_2^2}{n_2} \right).$$

When $\bar{X}_1 \leq \bar{X}_2$, the above analysis holds true with $K_1, K_2$ and $K_3$ replaced by $K'_1, K'_2$ and $K'_3$ for $e_2$, and $K'_1 = 1 - K_1, K'_2 = -K_2$ and $K'_3 = -K_3$, and the same line for $e_2$.

### 5.3. Asymptotic Conditional Distribution Using Bivariate Approach

Using the approach of [6], we can obtain asymptotic distribution of $e_{12}(\bar{X}_1, \bar{X}_2)$ as a sum of two bivariate normal distribution functions. Neglecting terms containing partial derivatives of order higher than the second, we can expand $e_1 = e_{12}(\bar{X}_1, \bar{X}_2)$ in a Taylor series form around $(\mu_1, \mu_2)$. For $\bar{X}_1 > \bar{X}_2$,

$$e_1 = K_1 + K_2 \left[ (\bar{X}_1 - \mu_1) + (\bar{X}_2 - \mu_2) \right] + O \left( \frac{1}{n} \right),$$

(5.14)

and for $\bar{X}_1 \leq \bar{X}_2$,

$$e_1 = K'_1 + K'_2 \left[ (\bar{X}_1 - \mu_1) + (\bar{X}_2 - \mu_2) \right] + O \left( \frac{1}{n} \right).$$

Denoting the distribution function of $e_1$ by $G$, we obtain

$$G(Z) = P(e_1 \leq Z) = P(e_1 \leq Z, \bar{X}_1 > \bar{X}_2) + P(e_1 \leq Z, \bar{X}_1 \leq \bar{X}_2).$$

(5.15)
Using the expansions for $e_1$ in (5.14) and (5.15), we have

\[
G(Z) = P \left\{ \left( \bar{X}_1 + \bar{X}_2 \right) - (\mu_1 + \mu_2) < \frac{Z - K_1}{K_2}, \bar{X}_1 - \bar{X}_2 > 0 \right\} \\
+ P \left\{ \left( \bar{X}_1 + \bar{X}_2 \right) - (\mu_1 + \mu_2) < \frac{Z - K_1'}{K_2'}, \bar{X}_1 - \bar{X}_2 < 0 \right\}.
\]

This probability can be written in the form

\[
G(Z) = P \left\{ \left( \bar{X}_1 + \bar{X}_2 \right) - (\mu_1 + \mu_2) < \frac{Z - K_1}{K_2}, (\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1) < (\mu_1 - \mu_2) \right\} \\
+ P \left\{ \left( \bar{X}_1 + \bar{X}_2 \right) - (\mu_1 + \mu_2) < \frac{Z - K_1'}{K_2'}, (\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2) \leq - (\mu_1 - \mu_2) \right\}.
\]

Let $U = \bar{X}_1 + \bar{X}_2$ and $V = \bar{X}_1 - \bar{X}_2$. For large $n_1$ and $n_2$, $U$ and $V$ have a joint BVN $(\mu_u, \sigma^2_u, \sigma^2_v, \rho)$, where

\[
\mu_u = \mu_1 + \mu_2, \quad \mu_v = \mu_1 - \mu_2, \quad \sigma^2_u = \sigma^2_v = \left( \frac{\nu^2_1 + \nu^2_2}{n_1} \right),
\]

\[
\sigma_{uv} = \left( \frac{\nu^2_1 - \nu^2_2}{n_1 - n_2} \right), \quad \rho = \frac{\sigma_{uv}}{\sigma_u \sigma_v}.
\]

(see [4, p. 62]).

Since $\bar{X}_1$ and $\bar{X}_2$ are, respectively, $\mathcal{N}(\mu_1, n_1\theta)$ and $\mathcal{N}(\mu_2, n_2\theta)$, then $\sigma^2_i = \mu^2_i/n_i\theta$ ($i = 1, 2$). From equations (5.16) and (5.17), we have

\[
G(z) = H(a_1, b_1, \rho) + H(a_2, b_2, \rho),
\]

where

\[
H(z_1, z_2, \rho) = P(Z_1 \leq z_1, Z_2 \leq z_2),
\]

and $Z_1, Z_2$ has a standard BVN distribution with correlation coefficient $\rho$. Hence, the distribution function of $e_1$ is given by (5.18) with

\[
a_1 = \frac{Z - K_1}{\sigma_u K_2}, \quad b_1 = \frac{\mu_v}{\sigma_v}, \quad \rho = \frac{\sigma_{uv}}{\sigma_u \sigma_v},
\]

\[
a_2 = \frac{Z - K_1'}{\sigma_u K_2'}, \quad b_2 = -\frac{\mu_v}{\sigma_v}.
\]

Note that the distribution function of $H(a_1, b_1, \rho)$ can be written in the form

\[
H(a_1, b_1, \rho) = \int_{-\infty}^{k(z)} \phi(z_1) dz_1 \int_{-\infty}^{b_1} \Phi(w) dw,
\]

where

\[
k(z) = \frac{Z - K_1}{\sigma_u K_2}, \quad w = \frac{Z_2 - \rho z_1}{\sqrt{1 - \rho^2}}, \quad \phi(z_1) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{z_1^2}{2} \right).
\]

Then the pdf of the distribution in equation (5.19) is obtained by differentiating $H(a_1, b_1, \rho)$ with respect to $z$, then

\[
h(a_1, b_1, \rho) = k'(z) \phi(k(z)) \Phi \left( \frac{b_1 - \rho a_1}{\sqrt{1 - \rho^2}} \right).
\]
Figure 1. Asymptotic distribution functions of $E_{12}$ univariate approach for $n_1 = n_2 = 10$. 
Figure 2. Asymptotic distribution functions of $E_{12}$ univariate approach for $n_1 = n_2 = 30$. 

Errors of Misclassification

$\lambda_1 = 1, \lambda_2 = 2, \theta = 3$

$\lambda_1 = 1, \lambda_2 = 4, \theta = 3$

$\lambda_1 = 1, \lambda_2 = 2, \theta = 5$

$\lambda_1 = 1, \lambda_2 = 4, \theta = 5$
Figure 3. Asymptotic distribution functions of $E_{12}$ univariate approach for $n_1 = n_2 = 50$. 

- $\lambda_1 = 1, \lambda_2 = 2, \theta = 3$
- $\lambda_1 = 1, \lambda_2 = 4, \theta = 3$
- $\lambda_1 = 1, \lambda_2 = 2, \theta = 5$
- $\lambda_1 = 1, \lambda_2 = 4, \theta = 5$
Figure 4. Asymptotic distribution functions of \( E_{12} \) bivariate approach for \( n_1 = n_2 = 10 \).
Hence, the pdf of $e_1$ for large $n_1, n_2$ is obtained in the form

$$g(z) = \frac{1}{\sigma_u K_2} \phi \left( \frac{z - K_1}{\sigma_u K_2} \right) \Phi \left( \frac{b_1 - \rho a_1}{\sqrt{1 - \rho^2}} \right) + \frac{1}{\sigma_u K_2} \phi \left( \frac{z - K_1}{\sigma_u K_2} \right) \Phi \left( \frac{b_2 - \rho a_2}{\sqrt{1 - \rho^2}} \right).$$

Similarly, we obtain the asymptotic distribution of $e_2$ as follows:

$$G^*(z) = H (a_1', b_1, \rho^*) + H (a_2', b_2, \rho^*),$$

where

$$a_1' = \frac{z - K_1^*}{\sigma_u K_2^*}, \quad b_1 = \frac{\mu_u - \mu_v}{\sigma_v}, \quad a_2' = \frac{z - K_1^*}{\sigma_u K_2^*}, \quad b_2 = \frac{\mu_v}{\sigma_v},$$

and

$$\mu_u = \mu_1 + C \mu_2, \quad \mu_v = \mu_1 - \mu_2, \quad \sigma^2_u = \left( \frac{\sigma_1^2}{n_1} + C^2 \frac{\sigma_2^2}{n_2} \right),$$

$$\sigma^2_{uv} = \left( \frac{\sigma_1^2}{n_1} - C \frac{\sigma_2^2}{n_2} \right), \quad \rho^* = \frac{\sigma^2_{uv}}{\sigma_u \sigma_v}.$$

Then the probability density function of $e_2^*$ is given by

$$g^*(z) = \frac{1}{\sigma_u K_2} \phi \left( \frac{z - K_1^*}{\sigma_u K_2} \right) \Phi \left( \frac{b_1 - \rho^* a_1}{\sqrt{1 - \rho^{*2}}} \right) + \frac{1}{\sigma_u K_2} \Phi \left( \frac{b_2 - \rho^* a_2}{\sqrt{1 - \rho^{*2}}} \right).$$

### 5.4. Numerical Comparisons of LDF and LR Using the Asymptotic Distributions

When $n_1$ and $n_2$ are large, the robustness of the LDF can be studied by comparing the asymptotic distributions of $e_1$ and $e_2$ using either the univariate or bivariate approach. Figures 1–3 give the asymptotic distribution function of $e_1$ and $e_2$, $F_1(z)$ and $F_2(z)$ based on (5.12) and (5.13) for $n_1 = n_2 = 10, 30$ and $50$ and for $z = 0.1, 0.2, \ldots, 0.6; S^2$ and $S^2$, $S_{*2}$ are calculated, respectively.

The distribution functions of $e_1$ and $e_2$ are based on the assumption that $X_1 > X_2$. For small $\theta$, there is not much difference between the values of $F_1(z)$ and $F_2(z)$; however, as $\theta = |\lambda_1 - \lambda_2|$ increases, $F_1^*(z)$ is found to be superior to $F_2^*(z)$. For small values of $\theta$ and $n_1, n_2$, the distribution functions do not become large fast enough, thus indicating that with a high probability, the errors of misclassification are likely to be large. As $\theta$ increases, we find that for small values of $z$, the values of $F_1(z)$ and $F_2(z)$ tend to unity, indicating that the errors tend to be small.

In Figure 4, the results using the bivariate approach have been illustrated for $n_1 = n_2 = 10$. We note that they are indeed parallel to those obtained using the univariate approach. The two techniques give similar results for all sample sizes and parameter values.

### REFERENCES