

Errors of Misclassification Associated with Edgeworth Series Distribution (ESD)

Awogbemi Clement Adeyeye*, Onyeagu Sidney Iheanyi

Statistics Department, Nnamdi Azikiwe University, Awka, Nigeria

Email address:

awogbemiadeyeye@yahoo.com (A. C. Adeyeye)

*Corresponding author

To cite this article:

Awogbemi Clement Adeyeye, Onyeagu Sidney Iheanyi. Errors of Misclassifications Associated with Edgeworth Series Distribution (ESD). *American Journal of Theoretical and Applied Statistics*. Vol. 8, No. 6, 2019, pp. 203-213. doi: 10.11648/j.ajtas.20190806.12

Received: March 2, 2019; **Accepted:** March 30, 2019; **Published:** November 8, 2019

Abstract: This study investigates the errors of misclassification associated with Edgeworth Series Distribution (ESD) with a view to assessing the effects of sampling from non-normality. The effects of applying a normal classificatory rule when it is actually a persistent non-normal distribution were examined. These were achieved by comparing the errors of misclassification for ESD with ND using small sample sizes at every level of skewness factor. The simulation procedure for the experiment of the study was implemented using numerical inverse interpolation method in R program to generate a uniformly distributed random variable N. A configuration size of 1000 was obtained for the two training samples drawn at every level of skewness factor (λ_3), in the range (0.00625, 0.4). This was repeated for different small sample sizes by comparing errors of misclassification of ESD with ND. The simulation results showed that the optimum probabilities of misclassification by ESD: (E_{12E}) decreases and (E_{12N}) increases, as the skewness factor (λ_3) increases. The optimum total probability of misclassification is stable as λ_3 also increases. The probability of misclassification $E_{12E} \geq E_{12N}$ and $E_{21E} \geq E_{21N}$ at every level of λ_3 . Thus, the total probabilities of misclassification are not greatly affected by the skewness factor. This asserts that the normal classification procedure is robust against departure from normality.

Keywords: Errors of Misclassification, Edgeworth Series Distribution, Skewness Factor, Classificatory Rule, Optimum Probability of Misclassification

1. Introduction

1.1. Background to the Study

The study of discrimination and classification problems with a view to assessing the effects of departure from the usual assumptions of normality cannot be overemphasized. In discrimination, we are concerned with the existence of two or more groups and a sample of observations from each of the groups. We are therefore required to design a rule based on measurements from these observations to the correct population when we do not know from which of the two populations it emanates [1, 22].

Classification is concerned with prediction or allocation of observations into groups in which a sample of observations is also given. The problem is to classify the observations into groups which are as distinct as possible [16].

Classification problem occurs when a researcher makes a

number of measurements on observations and wishes to classify the observations into one of several groups on the basis of the measurements. The observations cannot be identified with a group directly without recourse to the measurements. Fisher [8], illustrating this concept, classified iris flower from unknown group (specie) to any of the three known species (Iris Setosa red, Iris Versicolour green, and Iris Virginica black) on the basis of their attributes (Sepal length in cm, Sepal width in cm, Petal length in cm and Petal width in cm).

The general procedure for classifying an observation, X with p observed characters (X_1, \dots, X_p) consists of determining a function of (X_1, \dots, X_p) approximately, and assigning X to one of two populations depending on the value of the observations [2]. Since the observation vector is

random and the parameters for determining this function are often unknown, the procedure could result into two types of errors defined by errors of misclassification. Errors of misclassification occur when there is selection of criteria that is not suitable for classification [10].

The observation X may be classified as belonging to population π_1 when it actually comes from population two π_2 or vice versa. These errors are of serious concern in the choice of the procedure and as such, one is required to reduce the errors or more appropriately their probabilities are made as small as possible.

Let $f_1(x)$ and $f_2(x)$ be the probability density functions associated with X for population π_1 and population π_2 respectively. If the prior probabilities for populations π_1 and π_2 are P_1 and P_2 respectively with the regions of classifying observations into π_i into $R_i (i=1,2)$, then the probabilities of correctly or incorrectly classifying

observations are:

$$\begin{aligned} & \Pr(\text{object is correctly classified into } \pi_1) \\ &= \Pr(X \in R_1 | \pi_1) = \Pr(1|1) = P_1 \int_{R_1} f_1(x) dx \end{aligned} \quad (1)$$

$$\begin{aligned} & \Pr(\text{object is misclassified into } \pi_1) \\ &= \Pr(X \in R_1 | \pi_2) = \Pr(1|2) = P_2 \int_{R_1} f_2(x) dx \end{aligned} \quad (2)$$

$$\begin{aligned} & \Pr(\text{object is correctly classified into } \pi_2) \\ &= \Pr(X \in R_2 | \pi_2) = \Pr(2|2) = P_2 \int_{R_2} f_2(x) dx \end{aligned} \quad (3)$$

$$\begin{aligned} & \Pr(\text{object is misclassified into } \pi_2) \\ &= \Pr(X \in R_2 | \pi_1) = \Pr(2|1) = P_1 \int_{R_2} f_1(x) dx \end{aligned} \quad (4)$$

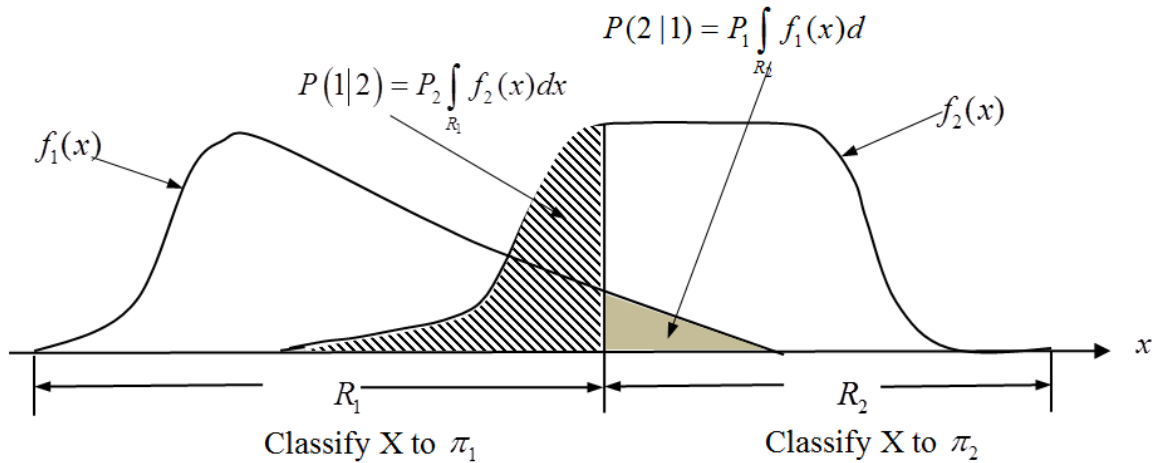


Figure 1. Probabilities of Misclassification.

In constructing a classification procedure, it is needful to minimize on the average, the bad effects of misclassification since a good classification procedure results to few misclassifications [18].

The Linear Discriminant Function (LDF) is a statistical procedure constructed as $(\mu_1 - \mu_2)^T \Sigma^{-1} \left[X - \frac{1}{2}(\mu_1 + \mu_2) \right]$. It assigns p dimensional observation vector X into one of the two populations π_i ($i = 1, 2$) and it is employed as an assignment rule when:

- The density functions of observations from populations π_1 and π_2 are multivariate normal: $\pi_i \sim N_p(\mu_i, \Sigma), i=1,2$
- The variance-covariance matrix (Σ_1) in population π_1 is the same as Σ_2 in population π_2 ;
- The prior probabilities of observations coming from

populations π_1 and π_2 are known;

- The parameters of the density functions in (a) are known.

Suppose the assumptions specified above are satisfied, then the Linear Discriminant Functions (LDF) provides optimal assignment rule in that it cannot be improved upon and the errors of misclassification are minimized. However, when some or all the assumptions are violated, it would be of interest to researchers to determine the effects of the violation on the procedures using LDF. If the parameters in (a) above are estimated from the samples, two problems may arise from the estimation stage. We may have missing values in the data and the initial sample may not be properly assigned due to inaccuracy in the initial assignment.

1.2. Statement of the Problem

For an experimenter who does not recognize an observation to be non-normal, he proceeds to use the normal

regions for classification. The question that emanates is: “how does this failure to transform to normality, prior to classification, affect the probability of misclassification”? This problem was investigated by comparing the errors of misclassification associated with Johnson’s system distributions in the appropriate transformable non-normal case with that of normal distribution [6]. Errors of misclassification associated with Gamma were also examined by [15]. Considerable work has been done by researchers in connection with errors of misclassification when the underlying distribution is transformable non-normal distribution, but the errors of misclassification associated with persistent non-normal distribution remain unresolved [12].

For any classification rule, the associated error rates are often used as criteria for evaluation of classification performance. These error rates are easily calculated when population parameters are known. However, when these parameters are unknown and must be estimated from the samples, the exact overall expected rate for the Fisher’s Linear Discriminant Function becomes virtually intractable. There is also a loss of information which affects the estimation of the probabilities of misclassification, in that it may be underestimated or overestimated. In order to rectify this problem, we derive the asymptotic distribution for the expected probability of misclassification of the distribution under consideration [17].

The aim of this study is to investigate errors of misclassification associated with Edgeworth Series Distribution (ESD). The research work seeks to achieve the following objectives:

- i. To examine the effect of applying the normal classificatory rule when the distribution is ESD by comparing the errors of misclassification using the Normal Distribution (ND) and ESD classification rules.
- ii. To use simulated data to validate the established results of the study.

2. Literature Review

The problem of estimating probabilities of misclassification has received remarkable attention in the literature ever since [8] introduced the Linear Discriminant Function. An extensive bibliography on this subject has been published by [21] since the probabilities of misclassification provide a way of evaluating the performance of the classification procedure.

Several investigations have also been conducted on the effects of non-normality on classification rules:

The robustness of the LDF and QDF with respect to certain types of non-normality was studied by [4], considering Johnson’s system of distributions which are transformable to normality. They considered three members of the family: log normal transformation $Y_i = \log(x_i)$, $0 < x_i < \infty$, Logit normal transformation $Y_i = \log\left(\frac{x_i}{1-x_i}\right)$, $0 < x_i < 1$ and Inverse hyperbolic sine normal

transformation $Y_i = \sin^{-1}(x_i)$, $-\infty < x_i < \infty$. Sampling studies were conducted in order to examine the behavior of the errors of misclassification, and they found out that the total error of misclassification is greatly increased as individual errors are distorted for all transformations in the case of the LDF. Approximate minimax rules were investigated because of the distortion in the errors and are found to reduce the errors of misclassification greatly.

The effect of non-normality on the QDF was investigated by [13]. They assumed that the data were transformable to normality. They derived random samples from non-normal distributions in order to study the effect of non-normality on QDF. Their results indicated that the actual error rates were considerably larger than the optimal rates in the case of zero mean difference.

The robustness of the Linear Discriminant Function (LDF) to non-normality using three Johnson system’s of distributions was examined by [6]. Though their work was restricted to transformable normality only, they opined that further work be carried out on a distribution that is non-normal with robustness on small sample sizes.

The effect of applying normal classificatory rule with focus on non-normality was also examined by [12]. He also obtained the asymptotic distribution of errors of misclassification in the non-normal case using Johnson’s system of distribution. The distribution function $G_i(z)$ and the expected value of the conditional distribution $E(e_{12})$ were evaluated for various values of given parameters theoretically using Johnson system of distributions. It was observed that the presence of outliers in one sample does not affect the behavior of the error rates in general.

Errors of misclassification for classification problems with two classes of univariate-gamma distribution were studied by [15]. The gamma density functions given as

$$f_i(x) = \frac{\lambda_i^\theta}{\Gamma(\theta)} x^{\theta-1} \exp(-\lambda_i x), i = 1, 2 \quad (5)$$

were reparameterized to the form

$$f_i(x, \theta, \mu_i) = \frac{\left(\frac{\theta}{\mu_i}\right)^\theta}{\Gamma(\theta)} x^{\theta-1} \exp\left(-\frac{\theta x}{\mu_i}\right), \mu_i, \theta > 0, \quad (6)$$

The effects of applying the normal classificatory rule to non-normal transformable gamma distribution were studied and assessed by comparing probabilities of misclassification (optimum and conditional). This was based on the Linear Discriminant Function (LDF) for normality and the likelihood ratio rule (LR) for gamma populations for various combination of λ_1, λ_2 and θ . They concluded that for small values of θ , n_1 and n_2 , the distribution functions do not become large and fast enough. This indicates that with a high probability, the errors of misclassification are likely to be large.

In this study, we examine the persistent non-transformable, non-normal distribution by investigating the effects of applying the normal classificatory rule when the distribution is ESD using empirical approach. We also develop expected probability of misclassification for ESD and its asymptotic distribution.

3. Method

The effects of non-normality in a two population discriminatory problem on the errors of misclassification are examined when the Anderson's statistic (W) defined by means of Edgeworth Series Distribution (ESD) is used for classifying an observation as emanating from population π_1 or π_2 . The effects would be studied for varying values of skewness factor based on the boundary of unimodal region for Edgeworth Series Distribution.

Optimum probabilities of misclassification for ESD are computed from known parameters and subsequently, the apparent probabilities of misclassification in respect of ESD for known and estimated parameters are generated.

3.1. Edgeworth Series Distribution (ESD)

Edgeworth Series Distribution (ESD) constitutes an expansion which is a series that approximates a probability distribution in terms of its cumulants and the Hermite polynomials. It relates the probability density function to that of a standard normal distribution [19].

The use of ESD is expedient because approximations to distribution of sample statistics of higher order than $n^{-\frac{1}{2}}$ is of concern interest in asymptotic theory of statistics. An important tool that evaluates the refinements is provided for by ESD. Its expansions take cognizance of a method of using information about a higher order moment to increase approximations accuracy [17].

Let $F(x)$ be the distribution to be approximated, $\{k_n\}$ its cumulants, γ_k the cumulants of a standard normal distribution function and D the differential operator with respect to x . Also, let Φ and ϕ be the standard normal distribution and standard normal density function respectively. Then

$$F(x) = \exp \sum_{n=1}^{\infty} (k_n - \gamma_n) \frac{(-D)^n}{n!} \Phi(x) \quad (7)$$

This is identical with the expansions in Hermite orthogonal function for a probability density function

$$P(x) = \sum_{n=0}^{\infty} c_n H_n(x) \phi(x) \quad (8)$$

where $H_{(n)}$ are Hermite polynomials and

$$C_k = \frac{1}{k!} \int_{-\infty}^{\infty} P(x) H_{(n)}(x) dx \quad (9)$$

By considering the standardized sum of n independent and identically distributed random variables, Edgeworth Series is obtained by collecting terms in equation (10) according to the power of n [11].

Let X_1, X_2, \dots, X_n be independent and identically distributed random variables with mean $\theta_0 = \mu$ and finite variance σ^2 .

If $\hat{\theta}_n$ is constructed from a sample of size n and $n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0)$ is asymptotically and normally distributed, then Edgeworth Series expansions are developed as approximations to distribution of estimates $\hat{\theta}_n$ of unknown quantities θ_0 . Thus the distribution functions of $n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0)$ is expanded as a power series in $n^{-\frac{1}{2}}$ so that

$$P\left(\frac{n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0)}{\sigma} \leq x\right) = \Phi(x) + n^{-\frac{1}{2}} P_1(x) \phi(x) + \dots + n^{-\frac{j}{2}} P_j(x) \phi(x) + \dots \quad (10)$$

where $\phi(x) = \sigma^{-1} (2\pi)^{-\frac{1}{2}} e^{-\frac{x^2}{2\sigma^2}}$ is the standard normal density function and

$\Phi(x) = \int_{-\infty}^x \phi(x) du$ is the standard normal distribution function

Equation (10) is the Edgeworth Series expansion. The functions P_j are polynomials with coefficients depending on cumulants of $\hat{\theta}_n - \theta_0$. In particular, P_j is a polynomial of degree at $3j-1$.

Suppose X_{ij} , $i=1, 2$, $j=1, 2, \dots, n_i$ denote two independent random samples from populations π_i , $i=1, 2$ respectively. The observations X_{ij} emanate from the common distribution defined by the density function

$$f_i(x) = \left(1 - \frac{\lambda_3}{6} D^3\right) \phi\left(\frac{x - \mu_i}{\sigma}\right), \quad -\infty < x < \infty \quad i=1, 2 \quad (11)$$

The parameter λ_3, μ_i ($i=1, 2$) satisfies the conditions: $-\infty < \lambda_3 < \infty$, $-\infty < \mu_i < \infty$ and, where D denotes the differential operator $\frac{d}{dx}$,

$$\phi\left(\frac{x - \mu_i}{\sigma}\right) = \frac{(2\pi)^{-\frac{1}{2}}}{\sigma} \exp\left[-\frac{(x - \mu_i)^2}{2\sigma^2}\right] \quad (12)$$

and λ_3 is the skewness factor [4].

3.2. Proposed Method of Estimating Probabilities of Misclassification

Let $X_{ij}, i=1,2; j=1,2,\dots,n_i$ be independent samples of sizes n_1, n_2 from populations π_1, π_2 . To estimate the apparent probabilities of misclassification, we define

$$E_{12E} = \sum_{j=1}^{n_1} \frac{\gamma_j}{n_1} \quad (13)$$

where $\gamma_j = 1$ if X_{1j} is classified as belonging to π_2 and $\gamma_j = 0$ if X_{1j} is classified as belonging to $\pi_1, j=1,2,\dots,n_1$.

The sample X_{11}, \dots, X_{1n_1} is taken from π_1 and each observation is classified in accordance with the rules in equations (35) and (36).

Similarly,

$$E_{21E} = \sum_{j=1}^{n_2} \frac{\delta_j}{n_2} \quad (14)$$

where $\delta_j = 1$ if X_{2j} is classified as belonging to π_1 and $\delta_j = 0$ if X_{2j} is classified as belonging to $\pi_2, j=1,2,\dots,n_2$

The sample X_{21}, \dots, X_{2n_2} is taken from π_2 and each observation is classified in accordance with the rules in equations (35) and (36).

The notation E_{12E} and E_{21E} represent the apparent probabilities of misclassification when observations from populations π_1 and π_2 are misclassified respectively by ESD rule.

For the purpose of comparison, the classification rule in equation (37) is successively applied to X_{1j}, X_{2j} . The proportion misclassified is estimated by the same procedure. Thus,

$$E_{12N} = \sum_{j=1}^{n_1} \frac{\gamma_j}{n_1} \quad (15)$$

And

$$E_{21N} = \sum_{j=1}^{n_2} \frac{\sigma_j}{n_2} \quad (16)$$

represent the two errors of misclassification.

3.3. Classification Rules for Normal Distribution

Let the probability density function of X in π_i ($i=1,2$) be

$$f_i(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma}\right)^2\right], -\infty < x < \infty, i=1,2 \quad (17)$$

If θ is the mean of the observation X and $H_0: \theta = \mu_1$ vs $H_a: \theta = \mu_2$, then the likelihood when $\mu_1 < \mu_2$

$$L = \frac{f_1(x)}{f_2(x)} = \exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2 + \frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2\right] \quad (18)$$

$$L^1 = -\frac{1}{2}\left(\frac{x-\mu_1}{\sigma}\right)^2 + \frac{1}{2}\left(\frac{x-\mu_2}{\sigma}\right)^2$$

$$= -\frac{1}{2\sigma^2} [2x - (\mu_1 + \mu_2)] (\mu_2 - \mu_1) \quad (19)$$

$$= \left[x - \frac{1}{2}(\mu_1 + \mu_2)\right] \left[\frac{\mu_1 - \mu_2}{\sigma}\right]$$

Equation (19) is the Anderson's discriminant function (W) when the distributions in the two populations are univariate normal with the same variance but different means [20]. We reject H_0 if $L < K$, where K is a constant.

From Equation (19) and the decision rule made, the classification rule specifies as follows:

$$\begin{aligned} &\text{Classify } X \in \pi_1 \text{ if } W > 0 \\ &\text{Classify } X \in \pi_2 \text{ if } W \leq 0 \end{aligned} \quad (20)$$

Equation (20) reduces to

$$\begin{aligned} &\text{Classify } X \in \pi_1 \text{ if } X < \frac{1}{2}(\mu_1 + \mu_2) \\ &\text{Classify } X \in \pi_2 \text{ if } X \geq \frac{1}{2}(\mu_1 + \mu_2) \end{aligned} \quad (21)$$

Similarly, when $\mu_1 > \mu_2$ the classification rule becomes

$$\begin{aligned} &\text{Classify } X \in \pi_2 \text{ if } X < \frac{1}{2}(\mu_1 + \mu_2) \\ &\text{Classify } X \in \pi_1 \text{ if } X \geq \frac{1}{2}(\mu_1 + \mu_2) \end{aligned} \quad (22)$$

When the parameters μ_1, μ_2 are unknown, and estimated by \bar{X}_1, \bar{X}_2 from the sample sizes of n_1 and n_2 respectively, the classification rule becomes:

$$\begin{aligned} &\text{Classify } X \in \pi_1 \text{ if } X < \frac{\bar{X}_1 + \bar{X}_2}{2}, \bar{X}_1 < \bar{X}_2 \\ &\text{and} \end{aligned} \quad (23)$$

$$\text{Classify } X \in \pi_1 \text{ if } X > \frac{\bar{X}_1 + \bar{X}_2}{2}, \bar{X}_1 \geq \bar{X}_2$$

$$\begin{aligned} \text{Similarly, } X \in \pi_2 \text{ if } X \geq \frac{(\overline{X}_1 + \overline{X}_2)}{2}, \overline{X}_1 < \overline{X}_2 \\ \text{and classify } X \in \pi_2 \text{ if } X \leq \frac{\overline{X}_1 + \overline{X}_2}{2}, \overline{X}_1 \geq \overline{X}_2 \end{aligned} \quad (24)$$

3.4. Classification Rule for Edgeworth Series Distribution (ESD)

Let the pdf of X in π_i be

$$f_i(x) = \left[1 - \frac{\lambda_3}{6} D^3 \right] \phi \left(\frac{x - \mu_i}{\sigma} \right), -\infty < x < \infty, i = 1, 2 \quad (25)$$

When $\mu_1 < \mu_2$, the likelihood ratio becomes

$$L = \frac{f_1(x)}{f_2(x)} = \frac{\left[1 - \left\{ \frac{\lambda_3}{2\sigma^3} \right\} \left(\frac{x - \mu_1}{\sigma} \right) + \left\{ \frac{\lambda_3}{6\sigma^3} \right\} \left(\frac{x - \mu_1}{\sigma} \right)^3 \right] \phi \left(\frac{x - \mu_1}{\sigma} \right)}{\left[1 - \left\{ \frac{\lambda_3}{2\sigma^3} \right\} \left(\frac{x - \mu_2}{\sigma} \right) + \left\{ \frac{\lambda_3}{6\sigma^3} \right\} \left(\frac{x - \mu_2}{\sigma} \right)^3 \right] \phi \left(\frac{x - \mu_2}{\sigma} \right)} \quad (26)$$

Equation (26) becomes

$$L = \frac{P \exp \left[-\frac{1}{2} \left(\frac{x - \mu_1}{\sigma} \right)^2 \right]}{Q \exp \left[-\frac{1}{2} \left(\frac{x - \mu_2}{\sigma} \right)^2 \right]} \quad (27)$$

$$\text{where } P = \left[1 - \frac{\lambda_3}{2\sigma^3} \left(\frac{x - \mu_1}{\sigma} \right) + \frac{\lambda_3}{6\sigma^3} \left(\frac{x - \mu_1}{\sigma} \right)^3 \right] \quad (28)$$

and

$$Q = \left[1 - \frac{\lambda_3}{2\sigma^3} \left(\frac{x - \mu_2}{\sigma} \right) + \frac{\lambda_3}{6\sigma^3} \left(\frac{x - \mu_2}{\sigma} \right)^3 \right] \quad (29)$$

We reject H_0 by if

$$\ln L < K \quad (30)$$

Taking $K=1$ reduces equation (30) to $\ln L < 0$

Thus, we reject $X \in \pi_1$ if

$$\ln P - \frac{1}{2} \left(\frac{x - \mu_1}{\sigma} \right)^2 - \ln Q + \frac{1}{2} \left(\frac{x - \mu_2}{\sigma} \right)^2 < 0 \quad (31)$$

This reduces to

$$\ln \left(\frac{P}{Q} \right) + \left[x - \left(\frac{\mu_1 + \mu_2}{2} \right) \right] \left(\frac{\mu_1 - \mu_2}{\sigma^2} \right) < 0$$

From equation (30), the classification rule takes the form:

For $\mu_1 < \mu_2$,

$$\text{classify } X \in \pi_1 \text{ if } \ln \left(\frac{P}{Q} \right) + W > 0 \quad (32)$$

and

$$\text{classify } X \in \pi_2 \text{ if } \ln \left(\frac{P}{Q} \right) + W \leq 0$$

For $\mu_1 > \mu_2$, (33)

$$\text{classify } X \in \pi_1 \text{ if } \ln \left(\frac{P}{Q} \right) - W > 0$$

and

$$\text{classify } X \in \pi_2 \text{ if } \ln \left(\frac{P}{Q} \right) - W \leq 0 \quad (34)$$

When the parameters μ_1, μ_2 are unknown, they are estimated by $\overline{X}_1, \overline{X}_2$ respectively and plugged in equation (34) before classification begins.

In the process of comparing errors of misclassification using ESD and ND classification rules, and data generated from the ESD, the effect of applying normal classification rule(likelihood ratio) when the distribution is ESD would be investigated by empirical method. Thus, the classification rule for ESD is given in the form:

For $\mu_1 < \mu_2$,

$$\text{classify } X \in \pi_1 \text{ if } \frac{P \exp \left[-\frac{1}{2} \left(\frac{x - \mu_1}{\sigma} \right)^2 \right]}{Q \exp \left[-\frac{1}{2} \left(\frac{x - \mu_2}{\sigma} \right)^2 \right]} < 1 \quad (35)$$

and

$$\text{classify } X \in \pi_2 \text{ if } \frac{P \exp \left[-\frac{1}{2} \left(\frac{x - \mu_1}{\sigma} \right)^2 \right]}{Q \exp \left[-\frac{1}{2} \left(\frac{x - \mu_2}{\sigma} \right)^2 \right]} \geq 1 \quad (36)$$

where P and Q remain as earlier defined in equations (35) and (36)

The normal classificatory rule for $\mu_1 < \mu_2$ is

$$\text{classify } X \in \pi_1 \text{ if } X < \left(\frac{\mu_1 + \mu_2}{2} \right)$$

and

$$\text{classify } X \in \pi_2 \text{ if otherwise} \quad (37)$$

3.5. Comparison of Errors of Misclassification

We estimate the errors of misclassification with focus on the small sample sizes. This is based on the fact that the asymptotic expansion of the errors does not indicate the behaviour of the error for small sample sizes [5, 9].

Estimation of the optimum probability of misclassification in the ESD when the skewness factor is in the range (0.00625, 0.4) is considered.

The apparent error rate for the Normal Distribution and ESD classification rules are examined using simulated data from ESD. The classification rules for the two distributions are also derived using likelihood criterion. The form of the estimators and the choice of values for skewness factor are also presented. The errors of misclassification are subsequently compared using the likelihood ratio rules for the Normal Distribution and ESD.

3.6. Choice of Skewness Factor Values

The choice of the values for skewness factor (λ_3) is anchored on the boundary of the positive unimodal regions for ESD where its probability density function is only valid. Thus, the skewness factor is chosen to lie within the range (0.00625, 0.4) as suggested by [3, 7].

3.7. Optimum Probability of Misclassification of ESD

When all the parameters of the distributions in the populations are known, the probability of misclassification is optimal in the sense that we cannot improve upon it. When an observation from π_1 is misclassified, the optimum probability of misclassification is given by

$$\begin{aligned}\alpha_1[R, f] &= \Pr\left[X \geq \frac{(\mu_1 + \mu_2)}{2}\right] \\ &= \int_{\delta}^{\infty} \left[1 - \frac{\lambda_3}{6} D^3\right] \phi\left(\frac{x - \mu_1}{\sigma}\right) dx \\ &= \int_{\delta}^{\infty} \left[1 + \frac{\lambda_3}{6\sigma^3} H_3\left(\frac{x - \mu_1}{\sigma}\right)\right] \phi\left(\frac{x - \mu_1}{\sigma}\right) dx \\ &= \int_{\delta}^{\infty} \phi\left(\frac{x - \mu_1}{\sigma}\right) dx + \frac{\lambda_3}{6\sigma^3} \int_{\delta}^{\infty} H_3\left(\frac{x - \mu_1}{\sigma}\right) \phi\left(\frac{x - \mu_1}{\sigma}\right) dx\end{aligned}\quad (38)$$

where $\delta = \left(\frac{\mu_1 + \mu_2}{2}\right)$ and $H_n(x)$ is Chebyshev's-Hermite polynomial of degree r and defined by the identity:

$$H_n(x)\phi(x) = (-D)^n \phi(x) \quad (39)$$

See [11].

If $\phi(x)$ denotes the standard normal density function, then we define the Hermite polynomial $H_n(x)$ for any integer n by

$$\frac{(-1)^n}{\sqrt{2\pi}} \frac{d^n}{dx^n} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{t^2}{2}} dt = (-1)^n \frac{d^n}{dx^n} \phi(x) = H_n(x)\phi(x) \quad (40)$$

and setting $z = \frac{x - \mu_1}{\sigma}$ as in equation (38), we have

$$\begin{aligned}\alpha_1[R, f] &= \int_{\frac{\delta - \mu_1}{\sigma}}^{\infty} \phi(z) dz + \frac{\lambda_3}{6\sigma^2} \int_{\frac{\delta - \mu_1}{\sigma}}^{\infty} H_3(z) \phi(z) dz \\ &= 1 - \phi\left(\frac{\delta - \mu_1}{\sigma}\right) + \frac{\lambda_3}{6\sigma^2} H_2\left(\frac{\delta - \mu_1}{\sigma}\right) \phi\left(\frac{\delta - \mu_1}{\sigma}\right) \\ &= 1 - \phi\left(\frac{\mu_2 - \mu_1}{2\sigma}\right) + \frac{\lambda_3}{6\sigma^2} \left[\left(\frac{\mu_2 - \mu_1}{2\sigma}\right) - 1\right] \phi\left(\frac{\mu_2 - \mu_1}{2\sigma}\right)\end{aligned}\quad (41)$$

If an observation from π_2 is misclassified, the optimum probability of misclassification is given by

$$\begin{aligned}\alpha_2[R, f] &= \Pr\left[X < \frac{(\mu_1 + \mu_2)}{2}\right] \\ &= \int_{-\infty}^{\delta} \left[1 - \frac{\lambda_3}{6} D^3\right] \phi\left(\frac{x - \mu_2}{\sigma}\right) dx \\ &= \int_{-\infty}^{\delta} \phi\left(\frac{x - \mu_2}{\sigma}\right) dx + \frac{\lambda_3}{6\sigma^3} \int_{-\infty}^{\delta} H_3\left(\frac{x - \mu_2}{\sigma}\right) \phi\left(\frac{x - \mu_2}{\sigma}\right) dx\end{aligned}\quad (42)$$

where $\delta = \frac{(\mu_1 + \mu_2)}{2}$.

Using the result in equation (38) and setting $z = \frac{x - \mu_2}{\sigma}$, we have

$$\begin{aligned}\alpha_2[R, f] &= \Pr\int_{-\infty}^{\frac{(\delta - \mu_2)}{\sigma}} \phi(z) dz + \frac{\lambda_3}{6\sigma^2} \int_{-\infty}^{\frac{(\delta - \mu_2)}{\sigma}} H_3(z) \phi(z) dz \\ &= \phi\left[\frac{\delta - \mu_2}{\sigma}\right] - \frac{\lambda_3}{6\sigma^2} H_2\left[\frac{\delta - \mu_2}{\sigma}\right] \phi\left[\frac{\delta - \mu_2}{\sigma}\right] \\ &= \phi\left[\frac{\mu_1 - \mu_2}{2\sigma}\right] - \frac{\lambda_3}{6\sigma^2} \left[\left(\frac{\mu_1 - \mu_2}{2\sigma}\right)^2 - 1\right] \phi\left(\frac{\mu_1 - \mu_2}{2\sigma}\right)\end{aligned}\quad (43)$$

4. Results

The optimum probabilities of misclassification are generated from equations (41) and (42). We assume that the known parameters: $\mu_1 = 0$, $\mu_2 = 1$ and $\sigma^2 = 1$. This implies that equations (41) and (42) are now functions of the skewness factor (λ_3) in the range (0.00625, 0.4). The probabilities of misclassification (E_{12E} and E_{21E}) are also computed and the results displayed in Table 1.

The apparent probabilities of misclassification are also examined when μ_1 and μ_2 are known and when the parameters are estimated from the samples.

Two independent samples of configuration size of 1000 each are at each value of the skewness factor (λ_3) from populations π_1 and π_2 . Their distributions are of ESD with respective parameters $\mu_1 = 0$, $\sigma^2 = 1$, $\mu_2 = 1$ and.

Using the ESD and Normal Distribution (ND) classification rules, the proportions misclassified in populations π_1 and π_2 are obtained and repeated for small samples ($n = 5, 10, 15, 20$ and 25). The probabilities of

misclassification obtained from the *samples* are averaged and the results displayed in Tables 2-6.

The notation E_{12E} is the apparent probability of misclassification when an observation from π_1 is misclassified by ESD and E_{21E} is the apparent probability of misclassification when an observation from π_2 is misclassified by ESD. Similarly, E_{12N} and E_{21N} are apparent probabilities of

misclassification when observation from populations π_1 and π_2 are misclassified respectively by ND classification rule.

The simulation experiments have been implemented using R programs and all the simulation results are obtained and displayed along with the total probabilities of misclassification in Tables 1-6.

Table 1. Optimum Probabilities of Misclassification at Different Values of Skewness for ESD.

Skewness Factor (λ_3)	Optimum Probability of Misclassification		
	E_{12E}	E_{21E}	Total
0.00625	0.3082	0.3088	0.6170
0.0125	0.3079	0.3091	0.6170
0.025	0.3074	0.3096	0.6170
0.05	0.3063	0.3107	0.6170
0.10	0.3041	0.3129	0.6170
0.15	0.3019	0.3151	0.6170
0.20	0.2997	0.3173	0.6170
0.25	0.2975	0.3195	0.6170
0.30	0.2953	0.3217	0.6170
0.35	0.2931	0.3239	0.6170
0.40	0.2909	0.3261	0.6170

The results in Table 1. show that E_{12E} decreases as the skewness factor λ_3 increases and E_{21E} increases as λ_3 increases. The total probability of misclassification is also stable (constant) as λ_3 increases.

Table 2. Comparison of Errors of Misclassification for Means unknown and Estimated by Average Values over 5 Samples.

Skewness Factor (λ_3)	ESD			ND		
	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.140	0.400	0.540	0.140	0.400	0.540
0.0125	0.220	0.410	0.630	0.220	0.410	0.630
0.025	0.225	0.465	0.690	0.220	0.475	0.695
0.05	0.210	0.395	0.605	0.205	0.400	0.605
0.10	0.205	0.475	0.680	0.175	0.495	0.670
0.15	0.260	0.285	0.545	0.230	0.320	0.550
0.20	0.305	0.365	0.670	0.295	0.395	0.690
0.25	0.455	0.185	0.640	0.420	0.230	0.650
0.30	0.195	0.465	0.660	0.115	0.545	0.660
0.35	0.225	0.465	0.660	0.125	0.520	0.645
0.40	0.440	0.180	0.610	0.360	0.250	0.610

From Table 3, E_{12E} is either equal to or greater than E_{12N} and E_{21E} is either equal to or less than or greater than E_{21N} at every level of skewness factor. The equality of the probability also occurs when the skewness factor λ_3 is very small.

Table 3. Comparison of Errors of Misclassification for Means unknown and Estimated by Average Values over 10 Samples.

Skewness Factor (λ_3)	ESD			ND		
	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.252	0.249	0.501	0.252	0.315	0.567
0.0125	0.236	0.236	0.472	0.236	0.236	0.472
0.025	0.266	0.219	0.485	0.231	0.295	0.526
0.05	0.224	0.282	0.506	0.216	0.314	0.530
0.10	0.290	0.278	0.568	0.208	0.336	0.544
0.15	0.387	0.203	0.590	0.215	0.220	0.435
0.20	0.277	0.320	0.597	0.270	0.337	0.607
0.25	0.255	0.245	0.500	0.230	0.292	0.522
0.30	0.248	0.334	0.582	0.182	0.394	0.576
0.35	0.216	0.339	0.555	0.175	0.354	0.529
0.40	0.253	0.209	0.462	0.170	0.196	0.366

From Table 3, E_{12E} is either equal to or greater than E_{21N} and E_{21E} is also equal to or less than E_{21N} at every level of skewness λ_3 .

Table 4. Comparison of Errors of Misclassification for Means unknown and Estimated by Average Values over 15 Samples.

Skewness Factor (λ_3)	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.345	0.145	0.490	0.345	0.150	0.495
0.0125	0.310	0.310	0.620	0.310	0.310	0.620
0.025	0.405	0.280	0.685	0.400	0.285	0.685
0.05	0.230	0.390	0.620	0.225	0.395	0.620
0.10	0.375	0.305	0.680	0.350	0.315	0.665
0.15	0.405	0.180	0.585	0.360	0.225	0.585
0.20	0.355	0.325	0.680	0.320	0.355	0.675
0.25	0.295	0.340	0.635	0.235	0.395	0.630
0.30	0.320	0.350	0.670	0.230	0.385	0.615
0.35	0.260	0.345	0.605	0.200	0.430	0.630
0.40	0.315	0.375	0.690	0.145	0.415	0.560

From Table 4, E_{12E} is equal to or greater than E_{21N} and E_{21E} is either less or greater than or equal to E_{21N} at every level of skewness factor λ_3 .

Table 5. Comparison of Errors of Misclassification for Means unknown and Estimated by Average Values over 20 Samples.

Skewness Factor (λ_3)	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.220	0.206	0.426	0.220	0.206	0.426
0.0125	0.280	0.280	0.560	0.192	0.295	0.487
0.025	0.330	0.210	0.540	0.290	0.230	0.520
0.05	0.345	0.205	0.550	0.295	0.250	0.545
0.10	0.265	0.300	0.565	0.230	0.390	0.620
0.15	0.340	0.350	0.690	0.330	0.375	0.705
0.20	0.350	0.240	0.590	0.320	0.255	0.575
0.25	0.295	0.270	0.565	0.270	0.295	0.565
0.30	0.300	0.195	0.495	0.265	0.200	0.465
0.35	0.310	0.350	0.660	0.270	0.360	0.630
0.40	0.405	0.285	0.690	0.380	0.400	0.780

From Table 5, E_{12E} is either equal to or greater than E_{12N} and E_{21E} is either equal to or less than E_{21N} at every level of skewness λ_3 .

Table 6. Comparison of Errors of Misclassification for Means unknown and Estimated by Average Values over 25 Samples.

Skewness Factor (λ_3)	E_{12E}	E_{21E}	Total	E_{12N}	E_{21N}	Total
0.00625	0.270	0.220	0.490	0.270	0.220	0.490
0.0125	0.290	0.330	0.620	0.290	0.235	0.525
0.025	0.390	0.295	0.685	0.375	0.310	0.685
0.05	0.340	0.270	0.610	0.335	0.280	0.615
0.10	0.375	0.305	0.680	0.360	0.315	0.675
0.15	0.360	0.230	0.590	0.345	0.245	0.590
0.20	0.275	0.430	0.705	0.225	0.480	0.705
0.25	0.375	0.255	0.630	0.320	0.290	0.610
0.30	0.390	0.240	0.630	0.300	0.330	0.630
0.35	0.290	0.300	0.590	0.240	0.345	0.585
0.40	0.405	0.225	0.630	0.305	0.290	0.595

From Table 6, E_{12E} is equal to or greater than E_{12N} and E_{21E} is either equal to or greater than E_{21N} at every level of skewness λ_3 .

5. Discussion

From the simulation results, It is evident that the total probability of misclassification at every value of λ_3 is either under or overestimated when small samples are employed to estimate μ_1 and μ_2 . The differences between taking small sample sizes are also not apparent.

The fact that $E_{12E} \geq E_{12N}$ and $E_{21E} \leq E_{21N}$ at every level

of skewness factor needed some algebraic justification and that has been established in this study. The actual probabilities of misclassification E_{12E} and E_{21E} as well as their sum remain close to the corresponding errors induced by the normal rule when λ_3 is small. As λ_3 increases, E_{12E} tends to be larger than E_{12N} and E_{21E} is smaller than E_{21N} .

Also, from Tables 4-6, the total probabilities of misclassification for the ESD and ND classification rules indicate no major difference between them at each value of λ_3 . The behaviours of the individual probabilities of misclassification E_{12E} and E_{12N} at every level of λ_3 show

that for small sample sizes, $E_{12E} \geq E_{12N}$ and $E_{21E} \leq E_{21N}$. The observed equality occurs when λ_3 is very small with an increasing parity as λ_3 increases. The observed equality of probabilities requires some algebraic justification.

From Normal Distribution (ND) classification rule and ESD classification rule, we earlier specified that

$$P = \left[1 - \left(\frac{\lambda_3}{2\sigma^3} \right) \left(\frac{X - \mu}{\sigma} \right) + \left(\frac{\lambda_3}{6\sigma^3} \right) \left(\frac{X - \mu}{\sigma} \right)^3 \right] \quad (44)$$

$$Q = \left[1 - \left(\frac{\lambda_3}{2\sigma^3} \right) \left(\frac{X - \mu_2}{\sigma} \right) + \left(\frac{\lambda_3}{6\sigma^3} \right) \left(\frac{X - \mu_2}{\sigma} \right)^3 \right] \quad (45)$$

When $\mu_1 = 0$, $\mu_2 = 1$ and $\sigma = 1$, we have,

$$P = 1 - \frac{\lambda_3 X}{2} + \frac{\lambda_3 X^3}{6} \quad (46)$$

$$Q = 1 - \frac{\lambda_3}{2} (X - 1) + \frac{\lambda_3}{6} (X - 1)^3 \quad (47)$$

We consider the error of misclassifying an observation $X \in \pi_1$. By normal classification rule, we classify an observation $X \in \pi_2$ wrongly if we use the rule:

$$\text{Classify } X \in \pi_1 \text{ if } W \leq 0. \quad (48)$$

The corresponding wrong classification of an observation $X \in \pi_1$ using ESD classification rule is to use

$$\text{Classify } X \in \pi_1 \text{ if } \ln\left(\frac{P}{Q}\right) + W \leq 0. \quad (49)$$

From equation (49), $\ln\left(\frac{P}{Q}\right)$ is a function of random variable X and the skewness factor (λ_3) . Given that $\ln\left(\frac{P}{Q}\right) = 0$, the two procedures are equivalent.

$$\begin{aligned} \frac{P}{Q} &= \frac{1 - \frac{\lambda_3}{2} X + \frac{\lambda_3}{6} X^3}{1 - \frac{\lambda_3}{2} (X - 1) + \frac{\lambda_3}{6} (X - 1)^3} \\ &= \frac{1 - \frac{\lambda_3}{2} X + \frac{\lambda_3}{6} X^3}{1 - \frac{\lambda_3}{2} X + \frac{\lambda_3}{2} + \frac{\lambda_3}{6} (X^3 - 3X^2 + 3X - 1)} \end{aligned} \quad (50)$$

But then

$$\begin{aligned} P - Q &= \frac{\lambda_3}{2} X^2 - \frac{\lambda_3}{2} + \frac{\lambda_3}{6} - \frac{\lambda_3}{2} \\ &= \frac{\lambda_3}{Q} [3X^2 - 3X - 2] \\ &\Rightarrow 3X^2 - 3X - 2 = 0. \end{aligned} \quad (51)$$

The quadratic function in equation (51) is equal to zero if the solutions of (51) are

$$X = +1.457 \text{ and } -0.457 \quad (52)$$

The parabola of equation (52) faces upwards and indicates that when $-0.457 < X < 1.457$, $P < Q$

$$\text{Since } P < Q, \ln\left(\frac{P}{Q}\right) = \ln P - \ln Q < 0. \quad (53)$$

With this, the cut-off point of the ESD classification rule is higher than the ND classification cut-off point. This results to E_{12E} being greater than E_{12N} .

Also, if an observation from π_1 is wrongly classified, the cut-off point of the ESD classification rule is lower than the ND classification rule cut-off point.

Hence, $E_{21E} < E_{21N}$

6. Conclusion

We have investigated the effect of sampling from persistent non-normal distribution by examining the normal classificatory rule when it is actually an Edgeworth Series Distribution (ESD). From the results obtained in this study, it is asserted that the normal procedure is sturdy against departures from normality. Thus, the skewness factor (λ_3) has a very little effect on the total probability of misclassification, which implies that it is not affected by the departures from normality. Nevertheless, the skewness factor indicates an increase or decrease in their errors of misclassification. Besides, the estimation of the errors when small sample sizes are used to estimate the means is an indication that the optimum probability of misclassification is underestimated or overestimated. This is anchored on the data generated and strictly restricted to this work.

References

- [1] Alvin, C. R. (2002). *Methods of Multivariate Analysis*, 2nd ed. John Wiley & Sons, New York, pp. 270.
- [2] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Methods*, New York: John Wiley, pp. 126-127.
- [3] Barton, D. E. and Dennis, K. E. (1952). The Condition under which Gram-Charlier and Edgeworth Curves are Positive Definite and Unimodal, *Biometrika*, 39, pp. 425- 427.
- [4] Bhuyan, C. K. (2010). *Probability Theory and Statistical Inference*, 1st ed., New Central Book Agency (Ltd), London, pp. 203.
- [5] Broffit, J. D. and Williams, J. S. (1973). Minimum Variance Estimators for Misclassification Probabilities in Discriminant Analysis, *Journal of Multivariate Analysis*, (3), pp. 311-327.
- [6] Chingada, E. F. and Kocherlakota, S. (1978). Robustness of the Linear Discriminant Function to nonnormality, *Journal of Statistical planning and Inference*, 3, pp. 74.

- [7] Draper, N, and Tierny, D. (1972). Region of Positive and Unimodal Series Expansion of the Edgeworth and Gramcharlier Approximations. *Biometrika*, 59 (2), pp. 463-465.
- [8] Fisher, F. A. (1936). The Use of Multiple Measurements in Taxonomic Problems, *Annals of Eugenics*, 7, pp. 179-188.
- [9] Geisser, S. (1967). Estimation Associated with Linear Discriminants. *Annals of Mathematical Statistics*, (38) pp. 807-817.
- [10] John, P. B (2010). Measurement Error, Model, Methods and Applications, CRC Press, Taylor and Francis, pp. 1-10.
- [11] Kendall, M. G. and Stuart, A. (1958). The Advanced Theory of Statistics, Vol. 3, Hafner: New York. pp. 155.
- [12] Kocherlakota, S., Kocherlakota, K. and Balakrishnan, N (1987). Expansions of Errors of Misclassification, *Advances in Multivariate Statistical Analysis*, pp. 191-211.
- [13] Lachenbruch, P. A., Clarke, B. B. & Lin, L. (1977). The Effect of Non-normality on Quadratic Discriminant Function, *MEDINFO 77 Shives/Wolf IFEP, North Holland Publishing Co.* pp. 101-104.
- [14] Lachenbruch, P. A., Sneeringer, C. &Revo, L. T. (1973). Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-normality. *Journal of Communication Statistics*, 1, pp. 39-57.
- [15] Mahmoud, M. A. and Moustafa, H. M. (1995). Errors of Misclassification Associated with Gamma Distribution, *Journal of Mathematical Computing Modeling*, Vol. 22, No. 3, pp. 105-119.
- [16] Ogum, G. E. O. (2002). Introduction to Methods of Multivariate Analysis, *Afri-Towers Ltd.*, Aba, Nigeria. pp. 107.
- [17] Peter, H. (1997). The Bootstrap and Edgeworth Expansions, *Springer Series in Statistics*, pp. 145.
- [18] Richard, A. J. and Dean, W. W. (1988). Applied Multivariate Statistical analysis, 3rd ed., Prentice Hall, Inc. New Jersey, pp. 497.
- [19] Ruby, C. W. (2010). A Bayesian Edgeworth Expansion by Stein's Identity, *Bayesian Analysis*, 5 (4), pp. 741-764.
- [20] Sedransk, N. and Okamoto, M. C. (1971). Estimation of the Probabilities of Misclassification for the Linear Discriminant Function in the Univariate Normal Case. *Annals of Statistics*, 23 (3), pp. 419-427.
- [21] Toussaint, G. T. (1974). Bibliography on Estimation of Misclassification. *IEEE Transactions on Information Theory*. Vol. 2 (4), pp. 472-479.
- [22] William, R. D. and Matthew, G. (1984). Multivariate Analysis, Methods and Applications. John Wiley & Sons, Inc. N. Y. pp. 360-361.