Probability of Misclassification Under Nelly Distribution Using Optimal Rule

Nnabude Chinelo Ijeoma[1]
S. I. Onyeagu (Professor)[2]
Dr. C. H. Nwankwo[3]
Department of Statistics, Faculty of Physical Sciences, Nnamdi Azikiwe University,
Awka, Anambra State, Nigeria

Abstract:- Errors of misclassification and their probabilities are studied for classification problems associated with univariate Nelly distribution. The effect of applying the linear discriminant function (LDF) based on normality to Nelly populations are assessed by comparing probability (Optimum) based on the linear discriminant function (LDF) with those based on the likelihood ratio rule (LR) for the Nelly. Both theoretical and empirical results are presented.

Keywords:- *Errors of Misclassification, Nelly Distribution, Linear Discriminant Function, Likelihood Ratio Rate, Error Rate*.

I. INTRODUCTION

The linear discriminant function (LDF) when used to categorize an observation that belongs to one of two normal population, has numerous advantageous qualities in classification issues. The parent populations' multivariate normal distributions are assured in the majority of the study. In the univariate case (denoted by $N(\mu,\sigma^2)$), the classification problem has been studied by John [6] and Sedransk and Okamoto [10]. Numerous kinds of non-normality have t heir effects studied. The linear discriminant function's resilience when the underlining distribution is a part of Johnson's system was investigated by Lachenbruch et. al (1975). Further exploring this issue, Ching'anda [4], Ching'anda and Subrohmaniam [5] developed distributions based on sizable samples for both conditional and unconditional probabilities of misclassification. Similar studies have been done for the inverse Gaussian distribution by Amoh and Kocherlakota [2] and Gamma distribution by Mahmoud and Moustafa [9].

In this paper, we will consider the probability of misclassification using optimal rule, when we have two classes and sampling from Nelly distributions. The probability of misclassification distributions is examined. The parameters X_1 and X_2 are known coming from Nelly distributions.

$$f_i(X) = \frac{(2\lambda_i)^{\propto} X^{\propto-1} e^{-2\lambda_i x}}{\Gamma(\propto)} \quad i = 1,2 \qquad (1)$$

- *The Robustness of the Linear Discriminant Function will be Examined in two ways*

- Supposing that in classifying an observation X from (1), the linear discriminant function (LDF) derived under the assumption of normality, how are the optimum (based on all parameters being known) probability of misclassification affected?
- The optimum probability of misclassification based on the likelihood ratio rule will be compared with those obtained from linear discriminant function.

II. THE CLASSIFICATION RULES

A classification rule or classifier is a process by which the elements of the population set are each expected to belong to one of the classes given a population whose members each belong to one of a number of different sets or classes. Every component of the population is assigned to the class which it actually belongs in a flawless categorization. If certain flaws are present in the classification, statistical analysis must be used to examine the classification. The general solution to the classification rule is to minimize the total probability of misclassification. (Anderson, 1958)

Suppose that $f_i(x)$ is the density function of X if it comes from the population

Π_i $(i = 1,2)$ and we assign X to Π_1 if X is in some region R_1 and to Π_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 Π_1. $P_1 + P_2 = 1$. The total probability of misclassification is

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

ISSN No:-2456-2165

$$E = P_1\left[1 - \int_{R_1} f_1(x)\,dx + P_2\int_{R_1} f_2(x)\,dx\right]$$

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

$$E = P_1 + \int_{R_1}[P_2 f_2(x) - P_1 f_1(x)]\,dx \qquad (2)$$

Where E is minimized (Neyman Pearson Lemma) if R_1 included 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

$$R_1 : \frac{f_1}{f_2} \geq \frac{P_2}{P_1}$$

$$R_2 : \frac{f_1}{f_2} < \frac{P_2}{P_1}$$

In what follows, assume $P_1 = P_2 = \frac{1}{2}$, it is well known that if $P_1 = P_2$ and $f_i(x)$ is univariate normal; the classification rule given above is equivalent to Fisher's Linear Discriminant function. (Lachenbruch, 1975).

- *Linear Discriminant Function for the Univariate Normal Distribution*
  (known $\mu_1 \neq \mu_2$, and the same Variance σ^2)

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

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

If θ is the mean of the observation X and

$$H_0 : \theta = \mu_1 \text{ vs}$$

$$H_a : \theta \neq \mu_2$$

Then the likelihood when $\mu_1 < \mu_2$

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

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

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

$$= \left[X - \frac{1}{2}(\mu_1 + \mu_2)\right]\left(\frac{\mu_1 - \mu_2}{\sigma}\right) \qquad (3)$$

The above equation is the Anderson discriminant function, when the distributions in the populations are univariate normal with the same variance but different means. H_o is rejected $if\ L < K$, where K is a constant from equation 3.

The classification rule specifies as

Classify $X \in \pi_1\ if\ X < \frac{1}{2}(\mu_1 + \mu_2)$ and

Classify $X \in \pi_2\ if\ X \geq \frac{1}{2}(\mu_1 + \mu_2)$

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

Classify $X \in \pi_2 \; if \; X < \frac{1}{2}(\mu_1 + \mu_2)$ and

Classify $X \in \pi_1 \; if \; X \geq \frac{1}{2}(\mu_1 + \mu_2)$

➢ *Derivation of Classification Rule for Nelly Distribution*

We assume that the distributions of X in π_i is given by Nelly distribution. Then the classification rule for Nelly distribution is

$$f_i(X) = \frac{(2\lambda_i)^{\propto} X^{\propto-1} e^{-2\lambda_i x}}{\Gamma(\propto)}$$

$$L = \frac{f_1(X)}{f_2(X)} = \frac{\frac{(2\lambda_1)^{\propto} X^{\propto-1} e^{-2\lambda_1 x}}{\Gamma(\propto)}}{\frac{(2\lambda_2)^{\propto} X^{\propto-1} e^{-2\lambda_2 x}}{\Gamma(\propto)}} \quad (4)$$

$$= \left(\frac{\lambda_1}{\lambda_2}\right)^{\propto} e^{2(\lambda_2 - \lambda_1)x} \quad (5)$$

Since $In\frac{f_1}{f_2} \leq Ink$

$$2(\lambda_2 - \lambda_1)x \leq In\, k - \propto In\left(\frac{\lambda_1}{\lambda_2}\right)$$

$$X \leq \frac{1}{2(\lambda_2 - \lambda_1)}\left[In\, k - \propto In\left(\frac{\lambda_1}{\lambda_2}\right)\right]$$

$$\leq \frac{In\, k}{2(\lambda_2 - \lambda_1)} - \frac{\alpha}{2(\lambda_2 - \lambda_1)} In\left(\frac{\lambda_1}{\lambda_2}\right)$$

The classification rule would be

$$R_1 = \{X\,; x \geq B \;\; if \; \mu_1 \geq \mu_2\}$$

Where $B \; is = \frac{\alpha}{2(\lambda_2 - \lambda_1)} In\left(\frac{\lambda_1}{\lambda_2}\right)$

The optimum probabilities of misclassification of Nelly distribution using Linear discriminant function (LDF) and Likelihood ratio (LR)

For LDF we have

$E_{12} = P\,\{X < A \,/\, X \in \pi_1\,;\, \mu_1, \mu_2\} \;\; if \; \mu_1 > \mu_2$ and $E_{21}(\mu_1, \mu_2)$ similarly defined.

The cumulative distribution function of Nelly distribution with parameters μ_1 and $\alpha\left(\mu_i = \frac{\alpha}{2\lambda}\right)$ is given by

$$F(x, \propto, \mu) = \frac{\Upsilon(\propto, \mu x)}{\Gamma(\propto)} = P(\propto, \mu x) \quad (6)$$

Where $\Upsilon(\propto, \mu x) = \sum_{n=0}^{\infty} \frac{(-1)^n (\mu x)^{\propto+n}}{n!(\propto+n)}$

Then using equation 1 we have

$$E_{12} = F(A\!:\, \propto, \mu_1), \mu_1 > \mu_2\!: A = \frac{(\mu_1 + \mu_2)}{2}$$

Therefore, the total optimum probability of misclassification using LDF is given as

ISSN No:-2456-2165

$$E = \frac{(E_{12} + E_{21})}{2} \quad for\ P_1 = P_2 = 0.5 \qquad (7)$$

- *For Likelihood Ratio*

$$E_{12}^{*} = P\{X < B \mid X \in \pi_1\} = F(B, \propto, \mu), \qquad if\ \mu_1 > \mu_2$$

Recall B = $\frac{\alpha}{2(\lambda_2 - \lambda_1)} In \left(\frac{\lambda_1}{\lambda_2}\right)$

Therefore, the total probability of misclassification using LR is given as

$$E^{*} = \frac{(E_{12}^{*} + E_{21}^{*})}{2} \qquad for\ P_1 = P_2 = 0.5 \qquad (8)$$

Therefore, the probabilities of misclassification based on the linear discriminant function (LDF) and likelihood ratio (LR) rules for various combination of the parameter $\lambda_1, \lambda_2\ and\ \propto$ the values chosen are $\lambda_1 = 1.0,\ \lambda_2 = 2.0, 3.0, 4.0\ and\ \propto = 2.0, 3.0, 4.0, 5.0, \ldots, 15.0$

III. RESULTS

Table 1 Comparison of the optimum probability of misclassification based on the LDF and LR when λ_1=1 and λ_2=2

Parameters			**Linear Discriminant Function (LDF)**			**Likelihood Ratio (LR)**		
λ_1	λ_2	α	E_{12}	E_{21}	E	E_{12}^{*}	E_{21}^{*}	E^{*}
1	2	2	0.1734	0.0549	0.1142	0.1398	0.0431	0.0914
		3	0.4789	0.1078	0.2934	0.3674	0.0771	0.2223
		4	0.1156	0.0075	0.0095	0.0830	0.0556	0.0693
		5	0.0115	0.0087	0.0101	0.1708	0.1314	0.1511
		6	0.0072	0.0052	0.0062	0.1981	0.1474	0.1727
		7	0.0059	0.0055	0.0057	0.3880	0.3605	0.3743
		8	0.0053	0.0037	0.0045	0.5117	0.3650	0.4382
		9	0.0010	0.0007	0.0009	0.2141	0.1572	0.1856
		10	0.0005	0.0006	0.0005	0.3322	0.3431	0.3376
		11	0.0004	0.0003	0.0004	0.4437	0.3603	0.4020
		12	0.00001	0.00004	0.00001	0.1039	0.0282	0.0661
		13	0.00001	0.000002	0.000006	0.1171	0.0303	0.0737
		14	0.000005	0.000001	0.000003	0.1210	0.0279	0.0745
		15	0.0000003	0.0000006	0.0000001	0.1113	0.0212	0.0663

Table 1 gives a comparison between the linear discriminant function of Nelly distribution and likelihood ratio of Nelly distribution when all the parameters are known; that is (λ_1=1, λ_2=2 and α =2 to 15)

Table 2 Comparison of the Optimum Probability of Misclassification based on the LDF and LR when λ_1=1 and λ_2=3

Parameters			**Linear Discriminant Function (LDF)**			**Likelihood Ratio (LR)**		
λ_1	λ_2	α	E_{12}	E_{21}	E	E_{12}^{*}	E_{21}^{*}	E^{*}
1	3	2	0.1443	0.0213	0.0828	0.0370	0.0047	0.0209
		3	0.3823	0.0288	0.2055	0.0626	0.0032	0.0329
		4	0.0093	0.0005	0.0049	0.2030	0.0047	0.1041
		5	0.0029	0.0010	0.0019	0.0228	0.0084	0.0156
		6	0.0022	0.0011	0.0017	0.0484	0.0256	0.0370
		7	0.0018	0.0009	0.0014	0.0759	0.0412	0.0585
		8	0.0009	0.0006	0.0007	0.1000	0.0699	0.0850
		9	0.0008	0.0005	0.0007	0.1702	0.1134	0.1418
		10	0.0007	0.0006	0.0007	0.3588	0.3036	0.3312
		11	0.0004	0.0003	0.0004	0.3955	0.3245	0.3599
		12	0.0003	0.0002	0.0003	0.5845	0.4462	0.5154
		13	0.0002	0.0001	0.0002	0.6956	0.5181	0.6069
		14	0.000001	0.000001	0.000001	0.9051	0.6947	0.7999
		15	0.000001	0.0000006	0.000001	0.000004	0.000006	0.000003

Table 2 gives a comparison between the linear discriminant function of Nelly distribution and likelihood ratio of Nelly distribution when all the parameters are known; that is (λ_1=1, λ_2=3 and α =2 to 15)

Table 3 Comparison of the Optimum Probability of Misclassification based on the LDF and LR when λ_1=1 and λ_2=4

parameters			Linear discriminant function (LDF)			Likelihood ratio (LR)		
λ_1	λ_2	α	E_{12}	E_{21}	E	E_{12}^*	E_{21}^*	E^*
1	4	2	0.0161	0.0071	0.0116	0.0237	0.0105	0.0171
		3	0.3360	0.0111	0.1736	0.0106	0.0002	0.0054
		4	0.0034	0.0010	0.0022	0.0123	0.0039	0.0081
		5	0.0053	0.0013	0.0033	0.0262	0.0065	0.0164
		6	0.0040	0.0008	0.0024	0.0326	0.0071	0.0199
		7	0.0013	0.0003	0.0008	0.0260	0.0069	0.0165
		8	0.0007	0.0002	0.0005	0.0383	0.0136	0.0259
		9	0.0006	0.0002	0.0004	0.0711	0.0268	0.0489
		10	0.0006	0.0004	0.0005	0.2196	0.1469	0.1833
		11	0.0004	0.0001	0.0003	0.1660	0.0664	0.1162
		12	0.0003	0.0002	0.0002	0.3967	0.2600	0.3283
		13	0.0002	0.0001	0.0002	0.6532	0.5239	0.5886
		14	0.000001	0.000006	0.000008	0.6103	0.4085	0.5094
		15	0.0000002	0.0000005	0.0000001	0.0000003	0.0000008	0.0000006

Table 3 gives a comparison between the linear discriminant function of Nelly distribution and likelihood ratio of Nelly distribution when all the parameters are known; that is (λ_1=1, λ_2=4 and α =2 to 15)

IV. CONCLUSION

Based on the result of the analysis of data, it can be concluded that the likelihood ratio of the Nelly distribution performs better than its linear discriminant function

REFERENCES

[1]. Abramowitz, M. and Stegun, I. A. (1970) Editors, *Handbook of* Mathematical Functions, Dover, New York, .

[2]. Amoh, R.K. and Kocherlakota, K. (1986). *Errors of misclassification associated with the inverse Gaussian distribution.* Communications in Statistics, theory and methods, 15(2), pp.589-612. doi:10.1080/03610928608829139

[3]. Anderson, T. W. (1958). *An introduction to multivariate statistical analysis.* 1st edition. John Wiley and sons, inc., New York.

[4]. Ching'anda E. F. and Subrahmaniam, K. (1976). *Misclassification errors and their distribution: Robustness*, M.Sc thesis, University of Manitoba

[5]. Ching'anda E. F. and Subrahmaniam, K. (1979). *Robustness of the linear discriminant function to non-normality:* Johnson's system, J. statist. Planning and inference 3, 69 – 77

[6]. John, S. (1961). *Errors in discrimination.* The annals of mathematical statistics vol. 32, No 4, pp. 1125 – 1144. https://www.jstor.org/stable/2237911

[7]. Lachenbruch, P. A., Sneeringer C. and Reve L. T. (1973). *Robustness of the linear and Quadratic discriminant function to certain types of non-normality, Commun. Statist.* 1, 39 – 57 (1973)

[8]. Lachenbruch, P. A. (1975). *Discriminant analysis.* 1st edition Hafner press, New York

[9]. Mahmoud and Moustafa (1995). Estimation of a discriminant function from a mixture of two gamma distributions when sample size is small, *Mathl. Comput. Modelling* 18 (5), 87-95 (1993).

[10]. Sedransk, N and Okamoto, M. (1971). *Estimation of the probability of misclassification for a linear discriminant function in the univariate normal case.* Annals of the institute of statistical mathematics 23, 419 - 43