ISSN 0971-0388

Aligarh Journal of Statistics
Vol. 40 (2020), 1-24

Evaluation of Biomarker in Case of Heterogeneous Data Through Rayleigh Mixture ROC Curve

Sudesh Pundir[1] and Azhar Uddin*
[Received on July, 2019. Accepted on July, 2020]

ABSTRACT

The Rayleigh Mixture ROC Curve can be applied when the lifetime data is heterogeneous or population consists of many subpopulations. We have proposed Rayleigh Mixture Receiver Operating Characteristic (ROC) model and also discussed its properties. Area under the Rayleigh Mixture ROC Curve (AUC) and its Variance are also found. Estimates of the parameters of Rayleigh Mixture ROC Curve are derived by using the Maximum Likelihood Method (EM) algorithm and Method of Moments. The proposed model is validated by using the simulation studies and real life example of head trauma data. It is found that the estimates using EM algorithm are closer to parameters as compared to the estimates using Method of Moments for large sample sizes.

1. Introduction

Rayleigh Mixture distribution has many real life applications in lifetime testing of an object. The lifetime of an object depends upon its age. The Rayleigh Mixture distribution is mainly used in the field of Reliability theory, Survival analysis, Probability theory and Operations research. This distribution has valuable attention in the field of medical diagnosis also, such as in the study of Magnetic Resonance Imaging (MRI).

The mixture distribution was introduced in 19^{th} century. Newcomb (1886) discussed the mixture distribution for outliers. Pearson (1894) also discussed the mixture distribution. He derived the estimates of two components normal mixture using method of moments. The first comprehensive study on finite mixture distribution is given by Everitt and Hand (1981). In this book, they discussed the

Corresponding author:* Azhar Uddin, Department of Statistics, Pondicherry University, Puducherry, India-605014. E-mail : ziaazhar533@gmail.com.
[1]Author: Sudesh Pundir, Department of Statistics, Pondicherry University, Puducherry, India-605014.

various estimation methods on finite mixture distribution along with the different types of mixture distribution. Titterington *et al.* (1985) also gave a nice discussion on finite mixture distribution with its properties in their text book. They also discussed different applications of the finite mixture distribution in their monograph. McLachlan and Peel (2000) discussed the different estimation methods of finite mixture distribution. They also mentioned the various applications on finite mixture models.

Identifiability is an important assumption for the estimation in finite mixture distribution. This lay out the unique characterization of the distribution. The estimation procedure of mixture distribution is not well defined without the identifiability. Yakowitz and Spragins (1968) discussed the exponential families of mixture distribution are identifiable. In this article, we are taking two component Rayleigh Mixture distribution. The Rayleigh Mixture distribution is a member of exponential family, so it is also identifiable.

Karim *et al.* (2011) discussed the Rayleigh Mixture distribution by taking the weight as sampling distributions like chi-square, *t* and *F* distribution. In this paper, they studied the moments, characteristic functions and shape characteristic of mixture distributions. Abushal *et al.* (2012) studied the estimation of Rayleigh Mixture distribution on the basis of generalized order statistics. In the article, they discussed the Bayesian estimates and MLE estimates both (point and interval) of the Rayleigh mixture distribution by using MCMC (Markov chain Monte Carlo) method. Dass and Kim (2012) discussed the ROC Curve using multivariate normal mixture. They used Bayesian and semi parametric inference for estimation. Gonen (2013) discussed the mixture ROC Curve. He proposed the AUC and partial AUC of normal mixture ROC Curve and also studied the comparison of bi-normal, two components normal and three components normal mixture ROC Curve. He discussed the estimation method of mixture ROC Curve using EM algorithm. Sindhu *et al.* (2014) studied the Bayes estimates of two components Rayleigh mixture distribution under doubly censoring. In this paper, they discussed the Bayes estimates of the Rayleigh mixture distribution under censoring. Ali (2014) derived the Bayesian estimates of mixture of the inverse Rayleigh distribution using informative and non-informative priors because of skewness and applicable in many fields. He did the simulation studies for censored data to compute the posterior distribution. Pundir and Azharuddin (2014) derived the Exponential Mixture ROC (EMROC) Curve. They have also studied the properties and AUC EMROC Curve. They have also discussed the comparative studies using the estimates of MLE via EM algorithm and Method of moments. Aslam *et al.* (2015) discussed the three component mixture of

Rayleigh Distributions. In this paper, they studied properties and estimation of Rayleigh mixture distribution in Bayesian perspective and also discussed Bayesian estimates of Rayleigh mixture distribution by using different prior and loss function. The sample sizes are large then the Bayes estimates are found to be more accurate as compared to the small sample size. Cheam and Mc Nicholas (2015) used the Gaussian mixture distribution for the heterogeneous data and check the diagnostic performance by the ROC Curve. He also discussed the EM algorithm, LABROC and Monte Carlo method to the comparison of Bi-normal Curve. Karim (2016) defined the scale of normal mixture distribution and also discussed properties of the inverted chi-square, Gamma, Exponential and Rayleigh distribution.

Pundir and Azharuddin (2016) discussed the Normal mixture ROC (NMROC) Curve. They have studied the properties and AUC and of NMROC Curve. Confidence interval of AUC of NMROC Curve is also derived. Simulation studies are also discussed. Pundir and Azharuddin (2016) derived the constant shape Weibull mixture ROC (CSWMROC) model. They have also discussed the properties, AUC and optimal cut-off of value of CSWMROC. Estimates of AUC of CSWMROC Curve are derived by using the method of moments. Monte Carlo simulation studies are also discussed. When the heterogeneity is found in the population and subpopulation then the CSWMROC Curve is compared to Bi constant shape Weibull ROC Curve. Pundir and Azharuddin (2016) further discussed a comparative study on NMROC Curve by using the estimates of MLE via EM algorithm and Method of Moments. They have also checked the accuracy of NMROC Curve using CK-BB data. The accuracy is found by MLE via EM algorithm and Method of Moments, which is approximately equal for the small dataset. They have also observed if the data have heterogeneous then NMROC Curve gives better accuracy as compared to the Bi-normal ROC Curve. Azharuddin and Pundir (2017) also further studied CSWMROC Curve. They have derived the optimal cut-off value and the estimates of CSWMROC Curve. Simulation studies are also discussed by using the estimates of MLE via EM algorithm and Method of Moments.

The whole paper is divided in 6 Sections. In Section 2, we proposed the Rayleigh Mixture ROC Curve. The properties of RMROC Curve are also studied. The AUC of RMROC Curve and optimal cut-off value are also obtained. In Section 3, we derived the estimates of RMROC Curve using Method of Moments (MOM) and Maximum Likelihood Method (MLE) via Expectation Maximization (EM) algorithm. In Section 4, the variance of Area under the Curve of RMROC Curve is also derived by delta method. Confidence Interval of AUC is also constructed. The test of significance of single AUC and two AUC's are also

proposed. In Section 5 simulation studies is also done. We compared the estimates of RMROC Curve using MOM and MLE via EM algorithm. The accuracy, standard error and confidence interval are also discussed. The real life example is discussed in Section 5. In the Section 6, the conclusion is given.

2. Rayleigh Mixture ROC Curve

Let X be a random variable from healthy controls which follows Rayleigh Mixture distribution with parameters σ_{10}, σ_{20} and Y be a random variable from disease cases which follows Rayleigh Mixture distribution with parameters σ_{11}, σ_{21}. The PDF and CDF of X and Y are given as

$$f(x) = p\frac{x}{\sigma_{10}^2}\exp\left(-\frac{x^2}{\sigma_{10}^2}\right) + (1-p)\frac{x}{\sigma_{20}^2}\exp\left(-\frac{x^2}{\sigma_{20}^2}\right), \quad x \geq 0, \ \ \sigma_{d0} > 0, \ \ d = 1,2$$

$$g(y) = p\frac{y}{\sigma_{11}^2}\exp\left(-\frac{y^2}{\sigma_{11}^2}\right) + (1-p)\frac{y}{\sigma_{21}^2}\exp\left(-\frac{y^2}{\sigma_{21}^2}\right), \quad y \geq 0, \ \ \sigma_{d1} > 0, \ \ d = 1,2$$

$$F(x) = p\left(1-\exp\left(-\frac{x^2}{\sigma_{10}^2}\right)\right) + (1-p)\left(1-\exp\left(-\frac{x^2}{\sigma_{20}^2}\right)\right), \quad x \geq 0, \ \ \sigma_{d0} > 0, \ \ d = 1,2$$

$$G(y) = p\left(1-\exp\left(-\frac{y^2}{\sigma_{11}^2}\right)\right) + (1-p)\left(1-\exp\left(-\frac{y^2}{\sigma_{21}^2}\right)\right), \quad y \geq 0, \ \ \sigma_{d1} > 0, \ \ d = 1,2$$

where, the means and variances of random variables X and Y are

$$\text{Mean}(X) = p_0\sigma_{10}\sqrt{\frac{\pi}{2}} + (1-p_0)\sigma_{20}\sqrt{\frac{\pi}{2}} \approx 1.253p_0\sigma_{10} + 1.253(1-p_0)\sigma_{20}$$

$$\text{Mean}(Y) = p_1\sigma_{11}\sqrt{\frac{\pi}{2}} + (1-p_1)\sigma_{21}\sqrt{\frac{\pi}{2}} \approx 1.253p_1\sigma_{11} + 1.253(1-p_1)\sigma_{21}$$

$$\text{Variance}(X) = p_0\frac{4-\pi}{2}\sigma_{10}^2 + (1-p_0)\frac{4-\pi}{2}\sigma_{20}^2 \approx 0.429p_0\sigma_{10}^2 + 0.429(1-p_0)\sigma_{20}^2$$

$$\text{Variance}(Y) = p_1\frac{4-\pi}{2}\sigma_{11}^2 + (1-p_1)\frac{4-\pi}{2}\sigma_{21}^2 \approx 0.429p_1\sigma_{11}^2 + 0.429(1-p_1)\sigma_{21}^2$$

The Rayleigh Mixture ROC Model is defined as

$$y(t) = p\{x(t)\}^{\frac{\sigma_{10}^2}{\sigma_{11}^2}} + (1-p)\{x(t)\}^{\frac{\sigma_{20}^2}{\sigma_{21}^2}} \qquad (2.1\text{-}2.9)$$

where

$$x(t) = p\exp\left(-\frac{t^2}{2\sigma_{10}^2}\right) + (1-p)\exp\left(-\frac{t^2}{2\sigma_{20}^2}\right).$$

The first subscript, 1 shows 1st subpopulation and 2 shows 2nd subpopulation in complete population. The second subscript 1 denotes that the observation is coming from disease cases and 0 denotes that the observation is coming from healthy cases.

Assumptions of ROC Curve

i) The mean of disease cases should be greater than the mean of healthy cases for RMROC Curve.

ii) $\sigma_{11} > \sigma_{10}$ and $\sigma_{21} > \sigma_{20}$

iii) $\sigma_{10} > \sigma_{20}$ and $\sigma_{11} > \sigma_{21}$

Properties of ROC Curve

i) RMROC Curve remains unaltered if the test scores undergo a strictly increasing transformation.

ii) RMROC Curve is monotonically increasing function.

Proof: A function is said to be monotonically increasing function if first derivative of the function is greater than zero. The first derivative of (2.9) with respect to x(t) is given as

$$\frac{dy(t)}{dx(t)} = p\frac{\sigma_{10}^2}{\sigma_{11}^2}\left\{x(t)\right\}^{\frac{\sigma_{10}^2}{\sigma_{11}^2}-1} + (1-p)\frac{\sigma_{20}^2}{\sigma_{21}^2}\left\{x(t)\right\}^{\frac{\sigma_{20}^2}{\sigma_{21}^2}-1} > 0$$

iii) RMROC Curve is a concave.

Proof: A function is said to be concave if its second derivative is less than zero. From (2.9), we have

$$\frac{d^2y(t)}{d\left\{x(t)\right\}^2} = p\left(\frac{\sigma_{10}^2}{\sigma_{11}^2}\right)\left(\frac{\sigma_{10}^2}{\sigma_{11}^2}-1\right)\left\{x(t)\right\}^{\frac{\sigma_{10}^2}{\sigma_{11}^2}-2} + (1-p)\left(\frac{\sigma_{20}^2}{\sigma_{21}^2}\right)\left(\frac{\sigma_{20}^2}{\sigma_{21}^2}-1\right)\left\{x(t)\right\}^{\frac{\sigma_{20}^2}{\sigma_{21}^2}-2} < 0$$

iv) The slope of the RMROC curve at the cut off value '*t*' is obtained by the ratio of probability density function of disease cases to the probability density function of healthy controls. It is given as

$$slope = \frac{\sigma_{10}^2\sigma_{20}^2\left[\, pt\sigma_{21}^2 \exp\left(-\frac{t^2}{2\sigma_{11}^2}\right) + (1-p)t\sigma_{11}^2 \exp\left(-\frac{t^2}{2\sigma_{21}^2}\right)\right]}{\sigma_{11}^2\sigma_{21}^2\left[\, pt\sigma_{20}^2 \exp\left(-\frac{t^2}{2\sigma_{10}^2}\right) + (1-p)t\sigma_{10}^2 \exp\left(-\frac{t^2}{2\sigma_{20}^2}\right)\right]}.$$

The slope of the ROC Curve at a threshold value *t* gives the Likelihood Ratio, which tells us how much ROC Curve is nearer to FPR and TPR.

v) The RMROC Curve is TNR asymmetric.

Proof: Kullback-Leibler divergence describes the symmetric property of ROC Curve (Ref. Hughes and Bhattacharya (2013)). RMROC Curve is TNR asymmetric if the Kullback-Leibler divergence KL(p,q) is greater than KL(q,p) and which are given as follows

$$KL(p,q) = \left[2\ln\sigma_{11} + 2\ln\sigma_{21} - 2\ln\sigma_{10} - 2\ln\sigma_{20} + \ln\left\{\left[p\sigma_{20}^2 t\exp\left(-\frac{t^2}{2\sigma_{10}^2}\right) + (1-p)\sigma_{10}^2 t\exp\left(-\frac{t^2}{2\sigma_{20}^2}\right)\right\}\right] - \ln\left\{pt\sigma_{21}^2\exp\left(-\frac{t^2}{2\sigma_{11}^2}\right) + (1-p)t\sigma_{11}^2\exp\left(-\frac{t^2}{2\sigma_{21}^2}\right)\right\}\right] \left[p\frac{t}{\sigma_{10}^2}\exp\left(-\frac{t^2}{2\sigma_{10}^2}\right) + (1-p)\frac{t}{\sigma_{20}^2}\exp\left(-\frac{t^2}{2\sigma_{20}^2}\right)\right].$$

$$KL(q,p) = \left[2\ln\sigma_{10} + 2\ln\sigma_{20} - 2\ln\sigma_{11} - 2\ln\sigma_{21} + \ln\left\{\left[p\sigma_{21}^2 t\exp\left(-\frac{t^2}{2\sigma_{11}^2}\right) + (1-p)\sigma_{11}^2 t\exp\left(-\frac{t^2}{2\sigma_{21}^2}\right)\right\}\right. - \left.\ln\left\{pt\sigma_{20}^2\exp\left(-\frac{t^2}{2\sigma_{10}^2}\right) + (1-p)t\sigma_{10}^2\exp\left(-\frac{t^2}{2\sigma_{20}^2}\right)\right\}\right] \left[p\frac{t}{\sigma_{11}^2}\exp\left(-\frac{t^2}{2\sigma_{11}^2}\right) + (1-p)\frac{t}{\sigma_{21}^2}\exp\left(-\frac{t^2}{2\sigma_{21}^2}\right)\right].$$

(2.10-2.11)

From (2.10) and (2.11), we can find that the KL(p,q) > KL(q,p) i.e., the RMROC Curve is TNR asymmetric.

The optimal cut-off value gives maximum classification accuracy between the healthy individual and disease cases in medical diagnosis. The optimal cut-off value of RMROC Curve is given as

$$t = p\sqrt{2\frac{\sigma_{10}^2\sigma_{11}^2}{\left(\sigma_{11}^2-\sigma_{10}^2\right)}\ln\left(\frac{\sigma_{11}^2}{\sigma_{10}^2}\right)} + (1-p)\sqrt{2\frac{\sigma_{20}^2\sigma_{21}^2}{\left(\sigma_{21}^2-\sigma_{20}^2\right)}\ln\left(\frac{\sigma_{21}^2}{\sigma_{20}^2}\right)}.$$

The AUC of RMROC Curve is defined as

$$AUC = pAUC_1 + (1-p)AUC_2 .$$
$$= p\frac{\sigma_{11}^2}{\sigma_{11}^2+\sigma_{10}^2} + (1-p)\frac{\sigma_{21}^2}{\sigma_{20}^2+\sigma_{21}^2}. \qquad (2.12)$$

3. Estimates of the parameters of Rayleigh Mixture ROC Curve

a) Method of Moments

The r^{th} theoretical moment about origin of Rayleigh Mixture distribution is defined as

$$\mu'_{r0} = p\int_0^\infty x^r \frac{x}{\sigma_{10}^2}\exp\left(-\frac{x^2}{2\sigma_{10}^2}\right)dx + (1-p)\int_0^\infty x^r \frac{x}{\sigma_{20}^2}\exp\left(-\frac{x^2}{2\sigma_{20}^2}\right)dx$$

Equating r^{th} theoretical moment about origin to the corresponding r^{th} sample moment, we get

$$m'_{r0} = p\sigma_{10}^r 2^{\frac{r}{2}}\Gamma\left(\frac{r}{2}+1\right) + (1-p)\sigma_{20}^r 2^{\frac{r}{2}}\Gamma\left(\frac{r}{2}+1\right).$$

Substituting *r*=1,2,3 in (5.2), we get first three sample moments of healthy controls as follows

$$m'_{10} = p\sigma_{10}2^{\frac{1}{2}}\Gamma\left(\frac{3}{2}\right) + (1-p)\sigma_{20}2^{\frac{1}{2}}\Gamma\left(\frac{3}{2}\right)$$
$$m'_{20} = 2p\sigma_{10}^2 + 2(1-p)\sigma_{20}^2$$
$$m'_{30} = p\sigma_{10}^3 2^{\frac{3}{2}}\Gamma\left(\frac{5}{2}\right) + (1-p)\sigma_{20}^3 2^{\frac{3}{2}}\Gamma\left(\frac{5}{2}\right) \qquad (3.1\text{-}3.5)$$

On solving (3.3), (3.4) and (3.5) in MATHEMATICA software, one can get the estimates of p, σ_{10} and σ_{20} for healthy cases and similarly one can get the estimates for disease cases.

b) Maximum Likelihood estimates using EM algorithm

The EM algorithm is used when the likelihood is not found in closed form and it is discussed by Dempster *et al.* (1977) to estimate the two component normal mixture distribution by using EM algorithm. We use an iterative technique for estimating the parameters known as Expectation Maximization (EM) algorithm. Mostly, mixture distributions are not found in closed form.

The likelihood function of Rayleigh Mixture density function from (2.1) is defined as

$$L = f\left(x_1, x_2, ..., x_{m|} \mid p_0, \sigma_{10}^2, \sigma_{20}^2\right) = \prod_{i=1}^{m}\left[p_0 \frac{x_i}{\sigma_{10}^2}\exp\left\{-\frac{x_i^2}{2\sigma_{10}^2}\right\} + (1-p_0)\frac{x_i}{\sigma_{20}^2}\exp\left\{-\frac{x_i^2}{2\sigma_{20}^2}\right\}\right].$$

Taking log on both sides, we get

$$\ln L = \sum_{i=1}^{m}\ln\left[p_0 \frac{x_i}{\sigma_{10}^2}\exp\left\{-\frac{x_i^2}{2\sigma_{10}^2}\right\} + (1-p_0)\frac{x_i}{\sigma_{20}^2}\exp\left\{-\frac{x_i^2}{2\sigma_{20}^2}\right\}\right]. \quad (3.6\text{-}3.7)$$

Differentiating (3.7) with respect to p_0, σ_{10}^2 and σ_{20}^2, we get

$$\frac{\partial \ln L}{\partial p_0} = \sum_{i=1}^{m}\left[\frac{\frac{x_i}{\sigma_{10}^2}\exp\left(-\frac{x_i}{2\sigma_{10}^2}\right) - \frac{x_i}{\sigma_{20}^2}\exp\left(-\frac{x_i}{2\sigma_{20}^2}\right)}{p_0\frac{x_i}{\sigma_{10}^2}\exp\left(-\frac{x_i}{2\sigma_{10}^2}\right) + (1-p_0)\frac{x_i}{\sigma_{20}^2}\exp\left(-\frac{x_i}{2\sigma_{20}^2}\right)}\right] = \sum_{i=1}^{n}\left[\frac{\Delta_i(x)}{p} - \frac{\{1-\Delta_i(x)\}}{(1-p)}\right]$$

where

$$\Delta_i(x) = \frac{pf_1(x)}{pf_1(x)+(1-p)f_2(x)} \text{ and } 1-\Delta_i(x) = \frac{(1-p)f_2(x)}{pf_1(x)+(1-p)f_2(x)}.$$

$$\frac{\partial \ln L}{\partial \sigma_{10}^2} = \sum_{i=1}^{m}\left[\frac{\left(\frac{x_i-2\sigma_{10}^2}{2\sigma_{10}^4}\right)p_0\frac{x_i}{\sigma_{10}^2}\exp\left(-\frac{x_i^2}{2\sigma_{10}^2}\right)}{p_0\frac{x_i}{\sigma_{10}^2}\exp\left(-\frac{x_i^2}{2\sigma_{10}^2}\right) + (1-p_0)\frac{x_i}{\sigma_{20}^2}\exp\left(-\frac{x_i^2}{2\sigma_{20}^2}\right)}\right] = \sum_{i=1}^{m}\left[\Delta_i(x)\left(\frac{x_i-2\sigma_{10}^2}{2\sigma_{10}^4}\right)\right]$$

$$\frac{\partial \ln L}{\partial \sigma_{20}^2} = \sum_{i=1}^{m}\left[\frac{(1-p_0)\left(\frac{x_i-2\sigma_{20}^2}{2\sigma_{20}^4}\right)\frac{x_i}{\sigma_{20}^2}\exp\left(-\frac{x_i^2}{2\sigma_{20}^2}\right)}{p_0\frac{x_i}{\sigma_{10}^2}\exp\left(-\frac{x_i^2}{2\sigma_{10}^2}\right) + (1-p_0)\frac{x_i}{\sigma_{20}^2}\exp\left(-\frac{x_i^2}{2\sigma_{20}^2}\right)}\right] = \sum_{i=1}^{m}\left[(1-\Delta_i(x))\left(\frac{x_i-2\sigma_{20}^2}{2\sigma_{20}^2}\right)\right]$$

(3.8-3.10)

After solving (3.8), (3.9) and (3.10), we get, $\hat{p}_0$, $\hat{\sigma}_{10}^2$, and $\hat{\sigma}_{20}^2$.
For healthy cases, the estimates are given as

$$\hat{p}_0 = \frac{\sum_{i=1}^{m}\Delta_i(x)}{m},\ \hat{\sigma}_{10}^2 = \frac{\sum_{i=1}^{m}[x_i\Delta_i(x)]}{2\sum_{i=1}^{m}\Delta_i(x)} \text{ and } \hat{\sigma}_{20}^2 = \frac{\sum_{i=1}^{m}x_i[\{1-\Delta_i(x)\}]}{2\sum_{i=1}^{m}\{1-\Delta_i(x)\}}.$$

Similarly, for disease cases the estimates are given as

$$\hat{p}_1 = \frac{\sum_{j=1}^{n}\Delta_i(y)}{n},\ \hat{\sigma}_{11}^2 = \frac{\sum_{j=1}^{n}\left[y_j\Delta_j(y)\right]}{2\sum_{j=1}^{n}\Delta_j(y)} \text{ and } \hat{\sigma}_{21}^2 = \frac{\sum_{j=1}^{n} y_j\left[\left\{1-\Delta_j(y)\right\}\right]}{2\sum_{j=1}^{n}\left\{1-\Delta_j(y)\right\}}. \quad (3.11\text{-}3.16)$$

Substituting above estimates of parameters in (2.12), we can get the estimate of AUC.

4. Variance of AUC of RMROC Curve

The approximate variance of AUC of RMROC Curve using delta method is given as

$$V(AUC) = pV(AUC_1) + (1-p)V(AUC_2)$$

$$= p\left[\left(\frac{\partial AUC_1}{\partial\sigma_{11}^2}\right)^2 V(\hat{\sigma}_{11}^2) + \left(\frac{\partial AUC_1}{\partial\sigma_{10}^2}\right)^2 V(\hat{\sigma}_{10}^2) + 2\left(\frac{\partial AUC_1}{\partial\sigma_{11}^2}\right)\left(\frac{\partial AUC_1}{\partial\sigma_{10}^2}\right)\mathrm{cov}(\hat{\sigma}_{11}^2,\hat{\sigma}_{10}^2)\right] +$$

$$(1-p)\left[\left(\frac{\partial AUC_2}{\partial\sigma_{21}^2}\right)^2 V(\hat{\sigma}_{21}^2) + \left(\frac{\partial AUC_2}{\partial\sigma_{20}^2}\right)^2 V(\hat{\sigma}_{20}^2) + 2\left(\frac{\partial AUC_2}{\partial\sigma_{21}^2}\right)\left(\frac{\partial AUC_2}{\partial\sigma_{20}^2}\right)\mathrm{cov}(\hat{\sigma}_{21}^2,\hat{\sigma}_{20}^2)\right] \quad (4.1)$$

where

$$\frac{\partial AUC_1}{\partial\sigma_{11}^2} = \frac{\sigma_{10}^2}{\left(\sigma_{11}^2+\sigma_{10}^2\right)^2},\ \frac{\partial AUC_1}{\partial\sigma_{11}^2} = -\frac{\sigma_{11}^2}{\left(\sigma_{11}^2+\sigma_{10}^2\right)^2} \quad (4.2)$$

$$\frac{\partial AUC_1}{\partial\sigma_{21}^2} = \frac{\sigma_{20}^2}{\left(\sigma_{21}^2+\sigma_{20}^2\right)^2},\ \frac{\partial AUC_1}{\partial\sigma_{20}^2} = -\frac{\sigma_{21}^2}{\left(\sigma_{21}^2+\sigma_{20}^2\right)^2} \quad (4.3)$$

Let $x_{1_{10}}, x_{2_{10}}, x_{3_{10}}, \ldots x_{m_{10}}$ and $y_{1_{11}}, y_{2_{11}}, y_{3_{11}}, \ldots y_{n_{11}}$ are the random sample of sizes m_{10} and n_{11} from $f_1(x_i)$ and $g_1(y_j)$. The likelihood function is given as follows

$$L = \prod_{i=1}^{m_{10}} f_1(x_i)\prod_{j=1}^{n_{11}} g_1(y_j) = \prod_{i=1}^{m_{10}}\left(\frac{x_i}{\sigma_{10}^2}\exp\left(-\frac{x_i^2}{2\sigma_{10}^2}\right)\right)\prod_{j=1}^{n_{11}}\left(\frac{y_j}{\sigma_{11}^2}\exp\left(-\frac{y_j^2}{2\sigma_{11}^2}\right)\right)$$

Taking log on both sides, we get

$$\ln L = -m_{10}\ln\sigma_{10}^2 + \sum_{i=1}^{m_{10}}\ln x_i - \frac{\sum_{i=1}^{m_{10}} x_i^2}{2\sigma_{10}^2} - n_{11}\ln\sigma_{11}^2 + \sum_{j=1}^{n_{10}}\ln y_j - \frac{\sum_{j=1}^{n_{11}} y_j^2}{2\sigma_{11}^2}$$

For finding $V(\hat{\sigma}_{11}^2), V(\hat{\sigma}_{10}^2), V(\hat{\sigma}_{21}^2) \text{ and } V(\hat{\sigma}_{20}^2)$ and $Cov(\hat{\sigma}_{11}^2,\hat{\sigma}_{10}^2), Cov(\hat{\sigma}_{21}^2,\hat{\sigma}_{20}^2)$, we will use the Fisher Information matrix which is given as

$$I(\theta) = -\begin{pmatrix} E\left(\frac{\partial^2 \ln L}{\partial \sigma_{10}^4}\right) & E\left(\frac{\partial^2 \ln L}{\partial \sigma_{10}^2 \partial \sigma_{11}^2}\right) \\ E\left(\frac{\partial^2 \ln L}{\partial \sigma_{11}^2 \partial \sigma_{10}^2}\right) & E\left(\frac{\partial^2 \ln L}{\partial \sigma_{11}^4}\right) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

where, $a_{11} = \frac{m_{10}}{\sigma_{10}^4}$, $a_{22} = \frac{n_{11}}{\sigma_{11}^4}$ and $a_{12} = a_{21} = 0$.

The $I^{-1}(\theta)$ is given as

$$I^{-1}(\theta) = \frac{1}{a_{11}a_{22}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} = \begin{pmatrix} V\left(\sigma_{10}^2\right) & Cov\left(\sigma_{10}^2, \sigma_{11}^2\right) \\ Cov\left(\sigma_{11}^2, \sigma_{10}^2\right) & V\left(\sigma_{11}^2\right) \end{pmatrix}$$

where

$$V\left(\hat{\sigma}_{11}^2\right) = \frac{\sigma_{11}^4}{n_{11}}, V\left(\hat{\sigma}_{10}^2\right) = \frac{\sigma_{10}^4}{m_{10}}, \text{ and } Cov\left(\sigma_{10}^2, \sigma_{11}^2\right) = Cov\left(\sigma_{11}^2, \sigma_{10}^2\right) = 0. \tag{4.4}$$

Similarly, we can obtain

$$V\left(\hat{\sigma}_{21}^2\right) = \frac{\sigma_{21}^4}{n_{21}}, V\left(\hat{\sigma}_{20}^2\right) = \frac{\sigma_{20}^4}{m_{20}}, \text{and } Cov\left(\sigma_{20}^2, \sigma_{21}^2\right) = Cov\left(\sigma_{21}^2, \sigma_{20}^2\right) = 0. \tag{4.5}$$

Substituting (4.2), (4.3), (4.4) and (4.5) in (4.1), the variance of estimated AUC is

$$V(AUC) = p\left[\left\{\frac{\sigma_{10}^2}{\left(\sigma_{10}^2 + \sigma_{11}^2\right)^2}\right\}^2 \frac{\sigma_{11}^4}{n_{11}} + \left\{-\frac{\sigma_{11}^2}{\left(\sigma_{11}^2 + \sigma_{10}^2\right)^2}\right\}^2 \frac{\sigma_{10}^4}{m_{10}}\right] + (1-p)\left[\left\{\frac{\sigma_{20}^2}{\left(\sigma_{20}^2 + \sigma_{21}^2\right)^2}\right\}^2 \frac{\sigma_{21}^4}{n_{21}} + \left\{-\frac{\sigma_{21}^2}{\left(\sigma_{21}^2 + \sigma_{20}^2\right)^2}\right\}^2 \frac{\sigma_{20}^4}{m_{20}}\right]$$

$$= p\left[\frac{\sigma_{11}^4 \sigma_{10}^4}{\left(\sigma_{11}^4 + \sigma_{10}^4\right)^4}\left(\frac{m_{10} + n_{11}}{m_{10} n_{11}}\right)\right] + (1-p)\left[\frac{\sigma_{21}^4 \sigma_{20}^4}{\left(\sigma_{21}^4 + \sigma_{20}^4\right)^4}\left(\frac{m_{20} + n_{21}}{m_{20} n_{21}}\right)\right]$$

given as (4.6)

where, m_{10}, m_{20} are the samples of healthy controls of Rayleigh mixture distribution and n_{11}, n_{21} are the samples of disease cases of Rayleigh mixture distribution.

Now, we can find standard error, confidence interval and develop test of significance of AUC of RMROC curve using variance of estimated AUC.

i) For testing the AUC of RMROC Curve,

$$H_0 : AUC = AUC_0 \text{ vs } H_1 : AUC \neq AUC_0$$

The test statistic is given as,

$$Z = \frac{\sqrt{N}\left(\hat{AUC} - AUC\right)}{\sqrt{V\left(\hat{AUC}\right)}} \sim N(0,1)$$

where N = m + n

ii) The 100(1-α)% confidence interval of AUC of RMROC Curve is given as

$$\hat{AUC} \pm SE\left(\hat{AUC}\right)Z_{\alpha/2}$$

where α is the level of significance and $Z_{\alpha/2}$ is the critical value of the confidence interval.

iii) Consider the problem of testing for AUCs of mixture ROC Curve and univariate ROC Curve. The Hypotheses are given as follows:

$$H_0: AUC_{mixture} = AUC_{univariate}$$
$$H_1: AUC_{mixture} > AUC_{univariate}$$

The test statistic is,

$$Z = \frac{\sqrt{N}\left(\hat{AUC}_{mixture} - \hat{AUC}_{univariate}\right)}{\sqrt{V\left(\hat{AUC}_{mixture}\right) + V\left(\hat{AUC}_{univariate}\right)}} \sim N(0,1). \qquad (4.7)$$

5. Numerical Studies

a) Simulation Studies

In this section, we will discuss the simulation studies for checking the behavior of RMROC Curve. The random samples are generated from two component Rayleigh Mixture distribution. Let us fix the weights for healthy controls and disease cases as p = $p_0 = p_1 = 0.5$. The parameters of healthy controls are $\sigma_{10} = 3$, $\sigma_{20} = 2$ and the parameters of disease cases are $\sigma_{11} = (8, 11, 14, 18, 20)$ and $\sigma_{21} = (6, 9, 12, 15, 17)$.

The sample sizes are taken as $m = n = (10, 20, 30, 40, 50, 100, 200, 300)$. The level of significance (α) is fixed as 0.05. The studies are done by using the MATHEMATICA Software and the commands are given in Appendix (a).

Tables 5.1 and 5.2 show the estimates of parameters of RMROC Curve by MLE (EM) algorithm and MOM for different sample sizes. The biases of parameters are also shown in bracket.

Table 5.1: Estimates and bias of parameters of RMROC Curve using MLE (EM) algorithm.

N	$\hat{p}_0$	$\hat{\sigma}_{10}$	$\hat{\sigma}_{20}$	$\hat{p}_1$	$\hat{\sigma}_{11}$	$\hat{\sigma}_{21}$
10	0.551 (0.5)	2.839 (-0.161)	1.435 (-0565)	0.34 (-0.16)	9.872 (1.872)	6.012 (0.012)
20	0.320 (-0.18)	3.312 (0.371)	1.677 (-0.323)	0..32 (-0.18)	10.942 (-0.058)	8.139 (-0.861)
30	0.453 (-0.05)	2.740 (-0.26)	1.871 (-0.129)	0.739 (0.23)	14.171 (0.171)	10.274 (-1.726)
40	0.562 (0.06)	2.940 (-0.06)	1.532 (-0.468)	0.496 (-0.01)	18.340 (0.34)	12.790 (-2.21)
50	0.640 (0.14)	2.959 (-0.041)	1.560 (-0.44)	0.51 (0.01)	22.219 (2.919)	16.103 (-0.897)
100	0.718 (0.21)	2.823 (-0.177)	1.688 (-0.312)	0.839 (0.33)	10.555 (2.555)	6.520 (0.52)
200	0.035 (0.03)	4.286 (1.286)	2.297 (0.297)	0.565 (0.065)	10.855 (-0.145)	7.779 (-1.221)
300	0.827 (0.32)	2.718 (-0.282)	1.374 (-0.626)	0.491 (-0.009)	15.193 (1.193)	10.753 (-1.247)

Table 5.2: Estimates and bias of parameters of RMROC Curve using MOM.

N	$\hat{p}_0$	$\hat{\sigma}_{10}$	$\hat{\sigma}_{20}$	$\hat{p}_1$	$\hat{\sigma}_{11}$	$\hat{\sigma}_{21}$
10	0.806 (0.31)	3.643 (0.643)	1.348 (-0.652)	0.83 (0.33)	6.446 (-1.554)	3.618 (-2.382)
20	0.800 (0.30)	2.908 (-0.092)	1.186 (-0.814)	0.747 (0.24)	11.721 (0.721)	4.393 (-4.607)
30	0.184 (-0.516)	3.832 (0.832)	2.091 (0.091)	0.11 (-0.59)	16.705 (2.705)	12.559 (0.559)

40	0.30 (-0.2)	2.686 (-0.81)	2.025 (0.025)	0.42 (-0.08)	16.786 (-1.214)	11.782 (-3.218)
50	0.36 (-0.14)	2.858 (-0.142)	1.899 (-0.142)	0.52 (0.02)	17.428 (-2.572)	16.616 (-0.384)
100	0.875 (0.37)	2.680 (-0.320)	1.373 (-0.627)	0.778 (0.278)	10.667 (2.667)	7.293 (1.293)
200	0.006 (-0.49)	5.872 (2.872)	2.350 (0.35)	0.256 (-0.244)	11.781 (0.781)	8.781 (-0.219)
300	0.930 (0.43)	2.621 (-0.379)	0.834 (-1.166)	0.719 (0.219)	14.472 (0.472)	13.821 (1.821)

From Tables 5.1 and 5.2, it is observed that as the sample size increases, the estimates become closer to the parameters of RMROC Curve. The bias of parameter also decreases with increase in sample size. On comparing both estimates, it is found that as the sample size increases, the estimates by MLE (EM) algorithm are more efficient than the estimates using MOM.

For testing the AUC of RMROC Curve for given different values of parameters, the hypotheses are given below:

$$H_0 : AUC = 0.88 \ \ vs \ \ H_1 : AUC \neq 0.88$$

$$H_0 : AUC = 0.94 \ \ vs \ \ H_1 : AUC \neq 0.94$$

$$H_0 : AUC = 0.96 \ \ vs \ \ H_1 : AUC \neq 0.96$$

$$H_0 : AUC = 0.97 \ \ vs \ \ H_1 : AUC \neq 0.97$$

$$H_0 : AUC = 0.98 \ \ vs \ \ H_1 : AUC \neq 0.98 \qquad (5.1\text{-}5.5)$$

Using (4.7), one can calculate the values of Z-statistic which are given in Tables 5.3 and 5.4.

Table 5.3 shows that the estimate of AUC, variance, SE, MSE and CI of AUC by MLE (EM) algorithm using estimates of parameters from Table 5.1. The estimates of AUC, variance, SE, MSE of AUC by MOM are also shown in Table 5.4 using estimates of parameters from Table 5.2. The bias of AUC is also shown in bracket.

Table 5.3: $\hat{AUC}, V(\hat{AUC}), SE(\hat{AUC}), MSE(\hat{AUC})$, CI(AUC) and Z-values of RMROC Curve using MLE (EM) algorithm.

N	$\hat{AUC}$	$V(\hat{AUC})$	$SE(\hat{AUC})$	$MSE(\hat{AUC})$	**CI(AUC)**	**Z-values**
10	0.938 (0.055)	0.000681	0.0261	0.0037	[0.887, 0.989]	1.902
20	0.945 (0.004)	0.000292	0.0171	0.0003	[0.911, 0.978]	0.183
30	0.965 (0.001)	0.000076	0.0087	0.0000	[0.947, 0.982]	0.053
40	0.980 (0.003)	0.000019	0.0044	0.0000	[0.971, 0.989]	0.510
50	0.987 (0.005)	0.000006	0.0026	0.0000	[0.981, 0.992]	1.857
100	0.934 (0.054)	0.000075	0.0086	0.0029	[0.917, 0.951]	88.601
200	0.916 (-0.024)	0.000059	0.0076	0.0006	[0.901, 0.931]	-60.301
300	0.971 (0.011)	0.000052	0.0022	0.0001	[0.967, 0.976]	123.525

Table 5.4: $\hat{AUC}, V(\hat{AUC}), SE(\hat{AUC}), MSE(\hat{AUC})$, CI(AUC) and Z-values of RMROC Curve using MOM.

N	$\hat{AUC}$	$V(\hat{AUC})$	$SE(\hat{AUC})$	$MSE(\hat{AUC})$	**CI(AUC)**	**Z-values**
10	0.778 (-0.105)	0.0059	0.0773	0.0169	[0.626, 0.929]	-1.423
20	0.939 (-0.002)	0.0003	0.0180	0.0003	[0.904, 0.974]	-0.111
30	0.961 (-0.003)	0.00009	0.0099	0.0000	[0.942, 0.980]	-0.33

40	0.972 (-0.005)	0.00003	0.0059	0.0000	[0.961, 0.984]	-0.847
50	0.980 (-0.002)	0.00001	0.0040	0.0000	[0.972, 0.988]	-0.5
100	0.943 (0.063)	0.000057	0.0075	0.0040	[0.928, 0.958]	119.088
200	0.932 (-0.008)	0.000040	0.0063	0.0001	[0.919, 0.944]	-24.080
300	0.970 (0.01)	0.000005	0.0024	0.0001	[0.965, 0.974]	103.14

From Tables 5.3 and 5.4, we observe that the estimates of AUC are closer to the true AUC. The variance, SE and MSE of AUC decreases with increase in sample size. It is also observed that maximum likelihood estimates are more efficient as compared to MOM. The MLE via EM algorithm gives more accuracy and less bias than MOM. We found that all values of Z-statistic lie between -1.96 and 1.96. Hence, all the null hypotheses are accepted.

The RMROC Curves for different sample sizes are shown in Fig. 5.1.

The Figure 5.1 depict that RMROC Curve gives maximum accuracy with increase in difference in scale parameters of disease cases.

b) Real Life Example

Zhou *et al.* (2002) has published head trauma data. The biomarker is CK-BB on the basis of which one can decide about the status of disease of a person. It consists of 57 subjects out of which 19 individuals are with good outcome (healthy) and 38 individuals are with bad outcome (disease). The histogram of CK-BB data for bad outcome and good outcome are shown below:

The above Fig. 5.2 histograms are asymmetric and right skewed. Tables 5.5 and 5.6, show the goodness of fit tests for poor outcome (disease cases) and good outcome (healthy controls). In these Tables, the test Statistics and p-values are given.

Table 5.5: Goodness of fit tests of Rayleigh Mixture distribution for poor outcome (disease individual) of head trauma data.

Test	**Statistics**	***p*-Value**
Anderson-Darling	0.9477	0.3846
Cramer-Von Mises	0.1516	0.3846
Kuiper	0.1974	0.4359
Pearson Chi Square	8.8947	0.3512
Watson U Square	0.0795	0.4061
Kolmogorov Smirnov	0.1390	0.4163

Table 5.6: Goodness of fit tests of Rayleigh Mixture distribution for good outcome (healthy individual) of head trauma data.

Test	**Statistics**	***p*-Value**
Anderson-Darling	0.8602	0.4368
Cramer-Von Mises	0.1659	0.3444
Kuiper	0.2670	0.4848
Watson U Square	0.0893	0.3364
Kolmogorov Smirnov	0.1931	0.4239

From Tables 5.5 and 5.6, we observe that all p-values are more than α, hence Rayleigh Mixture distribution fits well to the head trauma data. It is also obvious from PP-Plot in Fig. 5.3.

In Table 5.7, the estimates of AUC, using MLE (EM) algorithm and MOM for the head trauma data are given.

Table 5.7: Estimates of parameters of RMROC Curve for head trauma data.

Methods	$\hat{p}_0$	$\hat{\sigma}_{10}$	$\hat{\sigma}_{20}$	$\hat{p}_1$	$\hat{\sigma}_{11}$	$\hat{\sigma}_{21}$
MLE (EM)	0.73	120.88	21.87	0.48	566.46	160.86
MOM	0.76	118.75	11.87	0.46	580.39	158.98

Table 5.7 show that the estimates of parameters of RMROC Curve using MLE (EM) algorithm and MOM for head trauma data. The estimates are obtained by using the MATHEMATICA Software. The program is given in Appendix (a).

Using Table 5.7, the Table 5.8 shows that the estimates of AUC, variance, standard error and confidence interval of AUC, Sensitivity, and Specificity using MLE (EM) algorithm and MOM for the head trauma data are given.

Table 5.8: Estimates of AUC, Variance, SE, Sensitivity, Specificity, cut-off value (*t*) and Confidence Interval of AUC of RMROC Curve for head trauma data.

Methods	$\hat{AUC}$	$V(\hat{AUC})$	$SE(\hat{AUC})$	**Sen.**	**Spe.**	**(*t*)**	$CI(\hat{AUC})$
MLE (EM)	0.96	0.00007	0.008	0.945	0.855	165.89	[0.98, 0.95]
MOM	0.97	0.00005	0.007	0.963	0.817	160.761	[0.99, 0.96]

From Table 5.8, we observe that the estimates by MLE (EM) algorithm of AUC are almost same as estimates by MOM.

Now, we want to compare Bi-Rayleigh ROC Curve to RMROC Curve. The AUC of Bi-Rayleigh ROC Curve is found as 93% with standard error of 0.018. The Confidence Interval of AUC of Bi-Rayleigh ROC Curve is [0.894, 0.965]. The Sensitivity and Specificity of CK-BB by Bi-Rayleigh ROC Curve are 88% and 80% respectively at threshold 169.5 U/L (Ref. Pundir and Amala (2012)). The AUC of RMROC Curve is found as 96% with standard error 0.008. The confidence interval of AUC is [0.95, 0.98]. The Sensitivity and Specificity of RMROC Curve are 94.5% and 85.5% at the optimal threshold value is 165.89 U/L.

For testing the significance of AUC of Rayleigh Mixture $(AUC_{RM})$and AUC of Bi-Rayleigh (AUC_{BR}) ROC Curves, the hypotheses are $H_0 : AUC_{RM} = AUC_{BR}$ vs $H_1 : AUC_{RM} \rangle AUC_{BR}$. Using (4.15), we got the value of Z-statistic as 1.677, which is more than 1.645. Therefore, it is found that AUC of RMROC Curve is more than AUC of Bi-Rayleigh ROC Curve. Hence, RMROC Curve gives better accuracy than Bi-Rayleigh ROC Curve when the heterogeneity is found in data.

RMROC Curve and Bi-Rayleigh ROC Curve for head trauma data are shown in Fig. 5.4. It shows that the RMROC Curve is closer to the proper ROC Curve for

head trauma data and gives maximum accuracy as compared to Bi-Rayleigh ROC curve.

6. Conclusion

In real life situations, when a diseased population has a finite number of categories such that the test scores belongs to one of these categories then finite mixture model is the most appropriate model for such type of application. Similarly, if a healthy population also consists of finite number of categories and the test scores belongs to one of these categories then we can apply mixture model. Hence, mixture models are the best models for the population consisting of many categories due to which heterogeneity and complexity comes in the population. ROC Curve is used to measure the power or ability of the biomarker for diagnosis of a disease. Mixture ROC models are helpful for datasets where population has higher variability than univariate ROC models. In this paper, we have found the role of Mixture distribution in the context of ROC Curve which improves the accuracy of the diagnostic tests. The purpose of this research work is to measure the ability of a biomarker by using the proposed Mixture ROC models classifying the individuals as healthy or disease with high accuracy.

In this paper, we have proposed RMROC model. The assumptions and properties of RMROC Curve are derived. The RMROC Curve is monotonically increasing, concave and TNR asymmetric. We have obtained AUC and optimal cut-off value of RMROC Curve and also checked the identifiability condition of Rayleigh mixture distribution. The estimates of the parameters of RMROC by MLE via EM algorithm and MOM are derived. The approximate variance and 95% confidence interval of AUC of RMROC Curve are also obtained. Testing of AUC of RMROC and comparison of two AUCs is also proposed. We have done the simulation study and also validate the RMROC model by the head trauma data. The bias, variance, standard error and MSE of AUC decrease with increase in sample size. MLE via EM algorithm gives closer estimates as compared to MOM. Therefore, it is found that estimates of RMROC Curve by MLE via EM algorithm are more efficient than MOM. From Z-values, it is obvious that estimated AUC is closer to the actual AUC.

Rayleigh Mixture distribution fits well to the head trauma data. The accuracy of CK-BB using RMROC Curve is found to be 96% and 97% by MLE via EM algorithm and MOM with standard errors 0.008 and 0.007 at optimal cut-off values 165.89 U/L and 160.76U/L whereas the accuracy of CK-BB using Bi-

Rayleigh ROC Curve is 93% with standard error of AUC as 0.018 at optimal cut-off value 169.5 U/L. RMROC Curve gives maximum accuracy as compared to Bi-Rayleigh ROC Curve in case of head trauma data when the heterogeneity is found.

References

Abushal, T.A. and Al-Zaydi, A.M. (2012): Prediction Based on Generalized Order Statistics from a Mixture of Rayleigh Distributions Using MCMC Algorithm. *Open Journal of Statistics*, **2(3)**, 356-367.

Aslam, M., Tahir, M., Hussain Z. and Zahrani, B.A. (2015): A 3-Component Mixture of Rayleigh Distributions: properties and Estimation in Bayesian Framework. *PLOS ONE,* **10(5)**, 1-14, *e0126183.*doi: 10.1371/journal. pone. 0126183.

Azharuddin and Pundir , S. (2017): A Study of Weibull Mixture ROC Curve with Constant Shape Parameter, *IAPQR*, **41(2)**, 89-98.

Cheam, A.S.M and McNicholas, P.D. (2015): Modelling Receiver Operating Characteristic Curves using Gaussian Mixtures, *Computational Statistics & Data Analysis*, **93**, 192-208.

Dass, S.C. and Kim, S.W. (2012): A Semi-parametric approach to estimation of ROC curves for multivariate binormal mixtures, www.stt.msu.edu/Links/ Research_Memoranda/RM/RM_683.pdf.

Dempster, A.P., Laired, N.M. and Rubin, D.B. (1977): Maximum- Likelihood from incomplete data, via the EM algorithm, *Journal of Royal Statistics Society Series B (Methodological),* **39(1)**, 1-38.

Everitt, B. and Hand, D. J. (1981): *Finite Mixture Distributions*, Chapman and Hall, New York.

Gonen, M. (2013): Mixtures of Receiver Operating Characteristic Curves, *Academic Radiology,* **20(7)**, 831-837.

Karim, R., Hossain, P., Begum, S., and Hossain, F. (2011): Rayleigh Mixture Distribution, *Journal of Applied Mathematics, Article ID 238290,* 17 pages, doi: 10.1155/2011/238290.

Karim, R. (2016): A Class of scale Mixture Normal Distribution, *J BiomBiotat*, **7(2)**, 1-2.

McLachlan, G. and Peel, D. (2000): *Finite Mixture Models*, John Wiley and Sons, New York, 2000.

Newcomb, S. (1886): A generalized theory of the combination of observations so as to obtain the best result, *American Journal of Mathematics,* **8**, 343-346.

Pearson, K. (1894): Contribution to the mathematical theory of evolution, *Philosophical Transactions of the Royal Society A*, **185**, 71-110.

Ali, S. (2015): Mixture of the Inverse Rayleigh Distribution: Properties and estimation in a Bayesian framework, *Applied Mathematical Modelling*, **39**, 515-530.

Pundir, S. and R. Amala (2012): A Study on the Bi-Rayleigh ROC Model, *Bonfring International Journal of Data mining*, **2(2)**, 42-47.

Pundir, S. and Azharuddin (2014): A Study of Exponential Mixture ROC Model, Proceedings of 3rd International Conference on "*Innovative Approach in Applied Physical, Mathematical/Statistical, Chemical Sciences and Emerging Energy Technology for sustainable Development",* Jawaharlal Nehru University, New Delhi.

Pundir, S. and Azharuddin (2016): Detection of AUC and Confidence Interval using Normal Mixture ROC Curve, *IOSR Journal of Mathematics (IOSR-JM)*, **12(2)**, 77-86.

Pundir, S. and Azharuddin (2016): Evaluation of Diagnostic Accuracy and Standard Error using Constant Shape Weibull Mixture ROC Curve, *Pakistan Journal of Statistics and Operation Research*, **12(4)**, 739-751.

Pundir, S. and Azharuddin (2016): Estimation of Accuracy of Normal Mixture ROC Curve, *Journal of Indian Association for Productivity, Quality and Reliability*, **41(1)**, 61-77.

Sindhu, T. N., Feroze, N. and Aslam, M. (2014): Bayesian Estimation of the Parameters of Two component mixture of Rayleigh Distribution under Doubly Censoring, *Journal of Modern Applied Statistical Methods*, **13(2)**, 259-286.

Titterington, D.M., Smith A.F.M. and Makov, U.E. (1985): *Statistical Analysis of Finite Mixture Distributions*, John Wiley and Sons, New York.

Yakowitz, S.J. and Spragins, J.D. (1968): On the Identifiability of Finite Mixtures, *The Annals of Mathematical Statistics*, **39**, 209-214.

Zhou, X.H., Obuchowski, N.A. and McClish, D.K. (2002): *Statistical Methods in Diagnostic Medicine*, John Wiley and Sons, New York.

Fig. 2.1: Rayleigh Mixture distribution with parameters p=0.5, $\sigma_1 = 12$ and $\sigma_2 = 10$.

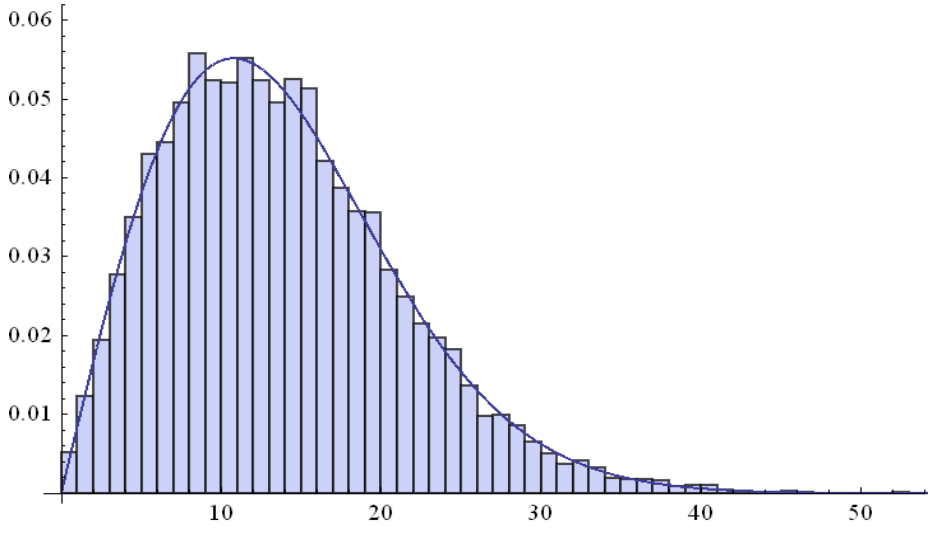


Fig. 1

Fig. 5.1: RMROC Curve for different sample sizes with different parameters.

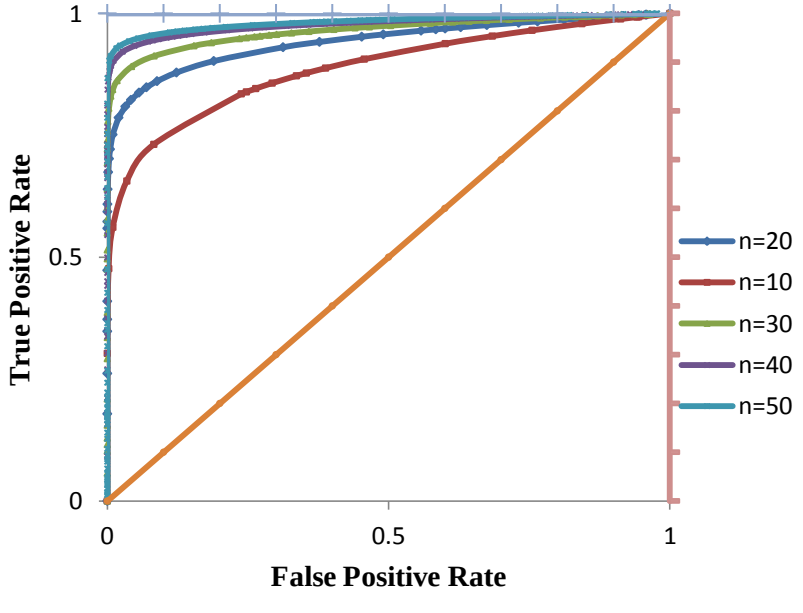


Fig. 5.2: Histogram of poor outcome (disease) and good outcome (healthy).

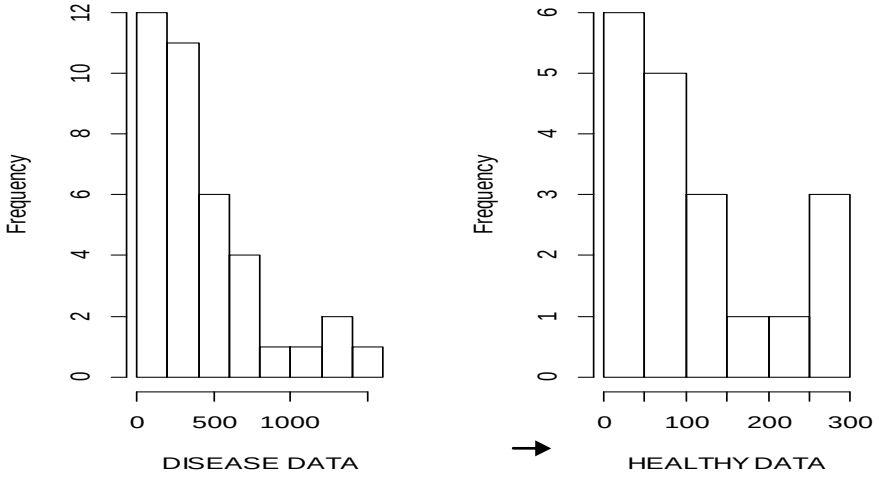


Fig. 2

Fig. 5.3: PP-Plot for poor outcome and good outcome of head trauma data.

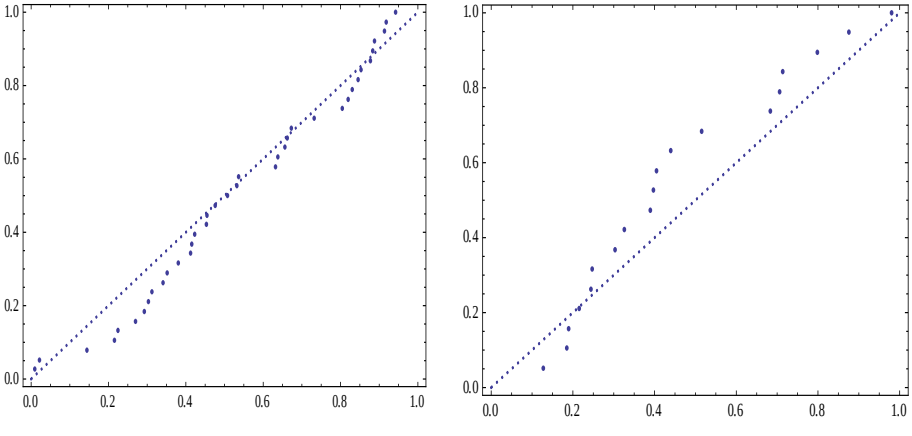
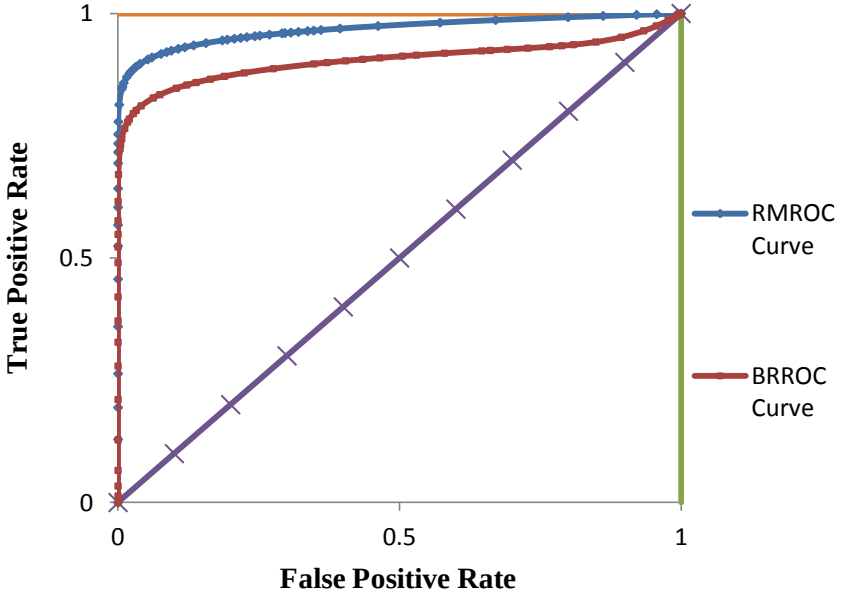


Fig. 3

Fig. 5.4: RMROC and BRROC Curves for head trauma data.



Appendix

Programs

(a) MATHEMATICA Command

```
Dist=MixtureDistribution[{p,1-
p},{RayleighDistribution[],RayleighDistribution[]}]
Data=RandomVariate[Dist,sample size]
FindDistributionParameters[Data,MixtureDistribution[{p,1p},{RayleighDistribut
ion[],RayleighDistribution[]}],ParameterEstimators->"MaximumLikelihood"]
FindDistributionParameters[Data,MixtureDistribution[{p,1-
p},{RayleighDistribution[],RayleighDistribution[]}],ParameterEstimators>"Meth
odOfMoments"]
Hist=Histogram[data,Automatic,"ProbabilityDensity"]
Plot=Plot[PDF[MixtureDistribution[{p,1-
p},{RayleighDistribution[],RayleighDistribution[]}],x],{x,range}]
Show[Hist,Plot]
DistributionFitTest[data,densityfunction,{"TestDataTable",All}]
```

ProbabilityPlot[data,densityfunction]
For MOM estimates, the following command is used
Solve

$$\left[\left\{p\sigma_1\sqrt{\frac{\pi}{2}}+(1-p)\sigma_2\sqrt{\frac{\pi}{2}}==m_1,\ 2p\sigma_1^2+2(1-p)\sigma_2^2==m_2,\ 3p\sigma_1^3\sqrt{\frac{\pi}{2}}+3(1-p)\sigma_2^3==m_3\right\},\{p,\sigma_1,\sigma_2\}\right]$$

(b) R-Codes

```
data1=c(140,1087,230,183,1256,700,16,800,253,740,126,153,283,90,303,193,76
,1370,543,913,509,576,671,80,490,156,356,350,323,1560,120,216,443,523,76,3
03,353,206)
data2=c(136,286,281,23,200,146,220,96,100,60,17,27,126,100,253,70,40,6,46)
par(mfrow=c(1,2))
hist(data1,main="")
hist(data2,main="")

hist(data1,xlim=c(5,1500),ylim=c(0,15),breaks=10,main="",xlab="number")
```