Evaluation of Diagnostic Accuracy and its Standard Error using Constant Shape Weibull Mixture ROC Curve

Sudesh Pundir
Department of Statistics, Pondicherry University, Puducherry, India
sudeshpundir19@gmail.com

Azharuddin
Department of Statistics, Pondicherry University, Puducherry, India
ziaazhar533@gmail.com

Abstract

Receiver Operating Characteristic (ROC) Curve is a widely used classification technique in Medical Diagnosis which classifies the healthy and diseased individuals on the basis of optimal cut off value of the biomarker. In this article, we have proposed Constant Shape Weibull Mixture ROC (CSWMROC) model. The properties of CSWMROC Curve are discussed and expressions for AUC, its variance and confidence interval are derived. The estimates of AUC of CSWMROC curve are obtained using Method of Moments (MOM). Numerical example is considered to support the proposed theory.

Keywords: CSWMROC Curve, AUC, Optimal cut-off value, Method of Moments, Monte Carlo simulation.

1. Introduction

Weibull Mixture distribution is very useful in medical diagnosis because it attains many shapes for different values of shape and scale parameters which helps in modeling different types of data. Here, we keep constant shape parameter to obtain the proper CSWMROC Curve so that it never crosses the chance diagonal otherwise it will become worthless.

Only a limited literature is available on the mixture of distributions. Some books on finite mixture distributions are written by Everitt and Hand [1981], Titterington *et al.* [1985] and McLachlan and Peel [2000]. Some authors like Newcomb [1886] studied the finite mixture distributions for outlier and Pearson [1894] estimated the parameters of the two component normal-mixture distribution by using the method of moments.

Other than the above mentioned monographs, some other works are attempted on Weibull mixture distribution. Kao [1959] derived the estimates of parameters of weibull mixture distribution using method of moments. Bucar *et al.* [2003] studied the finite weibull mixture distribution in Reliability theory. Arfa [2008] compared the sestimate of parameters of two component weibull mixture distribution by MOM and graphical method of estimation. Dwidayati *et al.* [2013] discussed the cure rate model in breast cancer patients through weibull mixture distribution. Dewan and Nandi [2009] estimated the parameters of the bivariate weibull distribution under random censoring using EM algorithm. Erisoglu and Erisoglu [2014] studied and compared the estimates of the weibull mixture distribution in case of heterogeneous data using EM algorithm, L-moment method and MLE method. They compared the bias, mean absolute error, total

mean error and time completion of the algorithm using different method of estimation by simulation studies. Pundir and Amala [2014] proposed and discussed the characteristics of the constant shape weibull ROC Curve.

ROC Curve is a graph between False positive rate (x(t)) and True positive rate (y(t)) for cut off value t. Till date, there are many authors like Green and Swets in [1966], Egan [1975], Zhou *et al.* [2002] and Krzanowski and Hand [2002] who discussed the ROC Curve for univariate distributions in case of continuous data. They gave the idea on theory of estimation on ROC Curve, AUC of ROC Curve and also used Statistical Inference on ROC Curve.

In practice, medical data is heterogeneous or it may consist of sub populations. Generally, we ignore this fact and apply the existing ROC models without checking for the heterogeneity which gives us the misleading results. Hence, there is a need to introduce mixture ROC models which will give exact accuracy of the diagnostic test with less standard error.

Only few authors discussed the mixture ROC Curve. The first article on the mixture ROC Curve is given by Dass and Kim [2011] where they discussed the Multivariate Bi-normal Mixture ROC Curve. Gonen [2013] also studied the ROC Curve and AUC using Bi-normal mixture distribution. It was found that if the heterogeneity is found in the data then Bi-normal mixture ROC Curve gives better smoothness as compare to bi-normal ROC Curve. Pundir and Azharuddin [2014] studied the Exponential Mixture ROC Curve and compared the estimates of AUC of Exponential Mixture ROC Curve using Method of Moments and MLE. Pundir and Azharuddin [2016] studied the Normal Mixture ROC Curve along with its properties and found the maximum likelihood estimates of parameters of AUC and confidence interval of AUC of Normal Mixture ROC Curve.

A mixture distribution can be applied if a population contains two or more sub-populations or in the presence of heterogeneity. A random variable X is said to follow a mixture distribution if it has the probability density function as

$$f(x)=\sum_{i=1}^{k} p_i\, f_i(x/\theta_i), \quad p_i>0,\ i=1,2,...,k \quad and \quad \sum_{i=1}^{k} p_i=1 \tag{1.1}$$

where p_i is the weight of the i^{th} component of mixture distribution.

Idenifiability is a necessary assumption for the estimation of mixture distributions. Without checking of identifiability in mixture distributions, one can not estimate the parameters. There are many authors who gave the idea on identifiability on mixture distribution. Teicher (1961, 1963) studied the identifiability of finite mixture distribution. Yakowitz and Spragins (1968) discussed the exponential families of mixture distribution are identifiable. Atienza N *et al.* (2006) discussed the new condition for identifiability on finite mixture distributions. They discussed the identifiability on Log Normal, Gamma and Weibull mixture distribution. In this paper, we are taking Constant Shape Weibull Mixture distribution where Weibull Mixture distribution is a member of exponential family, hence it is also identifiable.

A class N of mixture is said to be *identifiable* if and only if for all $f(x)\varepsilon N$ and the equality of two representations

$$\sum_{i=1}^{n} p_i f_i(x;\theta_i) = \sum_{j=1}^{n'} \hat{p}_j f_j\left(x;\hat{\theta}_j\right) \tag{1.2}$$

holds where, n=n' and for all i there exit some j such that $p_i = \hat{p}_j$ and $\theta_i = \hat{\theta}_j$. A random variable X is said to follow the two component weibull mixture distribution with probability density function

$$f(x) = p\frac{\alpha_1}{\beta_1}x^{\alpha_1-1}\exp\left(-\frac{x^{\alpha_1}}{\beta_1}\right) + (1-p)\frac{\alpha_2}{\beta_2}x^{\alpha_2-1}\exp\left(-\frac{x^{\alpha_2}}{\beta_2}\right), \quad 0\langle x\langle\infty,\ 0\langle\alpha_i\langle\infty,\ 0\langle\beta_i\langle\infty,\ i=1,2. \tag{1.3}$$

The cumulative distribution function of the two component weibull mixture distribution is given as

$$F(x) = p\left(1-\exp\left(-\frac{x^{\alpha_1}}{\beta_1}\right)\right) + (1-p)\left(1-\exp\left(-\frac{x^{\alpha_2}}{\beta_2}\right)\right) \quad 0\langle x\langle\infty,\ 0\langle\alpha_i\langle\infty,\ 0\langle\beta_i\langle\infty,\ i=1,2. \tag{1.4}$$

where α_i and β_i are the shape and scale parameters of the weibull mixture distribution. In this paper, the shape parameter $\alpha_1 = \alpha_2 = \alpha$ is constant

The paper is organized as follows. In section 2, we have studied the CSWMROC model and its properties. The AUC and optimal cut-off value of biomarker using CSWMROC model are also derived. The moment estimates of AUC of CSWMROC Curve are also obtained in section 3. In section 4, the variance of AUC of CSWMROC model and confidence Interval (CI) are derived using delta method. In section 5, AUC, variance of AUC, Standard Error (SE) of AUC, Mean Square Error (MSE) of AUC, confidence interval and testing of AUC are done by using simulation studies. In the last section, conclusion is given.

2. *Constant Shape Weibull Mixture ROC model*

Let X be a random variable from healthy controls which follows Constant Shape Weibull Mixture Distribution with parameters α, β_{10} and β_{20} and Y be another random variable from disease cases which follows Constant shape Weibull Mixture Distribution with parameters α, β_{11} and β_{21}. The CSWMROC model is defined as

$$y(t) = p\left(x(t)\right)^{\frac{\beta_{10}}{\beta_{11}}} + (1-p)\left(x(t)\right)^{\frac{\beta_{20}}{\beta_{21}}}, \quad 0\langle p\langle 1;\ 0\langle t\langle\infty;\ 0\langle\beta_{ij}\langle\infty; i=0,1;\ j=1,2. \tag{2.1}$$

where,

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

Assumptions:

(1) The mean of disease cases should be greater than the mean of healthy cases for CSWMROC curve.

(2) The Shape parameter (α) should be fixed to obtain the proper CSWMROC curve.

(3) $\beta_{11} \rangle \beta_{10}$, $\beta_{21} \rangle \beta_{20}$, $\beta_{20} \rangle \beta_{10}$ and $\beta_{21} \rangle \beta_{11}$.

Properties:

(a) The CSWMROC curve remains unaltered if the test scores undergo a strictly increasing transformation.

(b) CSWMROC curve is monotonically increasing.

Proof: A function is said to be monotonically increasing function if the first derivative of the function is greater than zero. From (2.1), we have

$$\frac{dy(t)}{dx(t)} = p\frac{\beta_{10}}{\beta_{11}}\left\{y(t)\right\}^{\frac{\beta_{10}}{\beta_{11}}-1} + (1-p)\frac{\beta_{20}}{\beta_{21}}\left\{x(t)\right\}^{\frac{\beta_{20}}{\beta_{21}}-1} > 0, \quad p>0,\ \beta_{i0}>0,\ \beta_{i1}>0,\ i=1,2. \tag{2.2}$$

(c) CSWMROC Curve is a Concave.

Proof: A function is said to be concave if its second derivative is less than zero. From (2.2), we have

$$\frac{d^2y(t)}{dx^2(t)} = p\left(\frac{\beta_{10}}{\beta_{11}}\right)\left(\frac{\beta_{10}}{\beta_{11}}-1\right)\left\{y(t)\right\}^{\frac{\beta_{10}}{\beta_{11}}-2} + (1-p)\left(\frac{\beta_{20}}{\beta_{21}}\right)\left(\frac{\beta_{20}}{\beta_{21}}-1\right)\left\{x(t)\right\}^{\frac{\beta_{20}}{\beta_{21}}-2} \langle 0, \qquad p>0,\ \beta_{i0}>0,\ \beta_{i1}>0,\ i=1,2. \tag{2.3}$$

(d) CSWMROC Curve is TPR asymmetric.

Proof: Let f(x) be comparison distribution and g(x) be reference distribution, then KL(f, g) and KL(g, f) are given as

$$KL(f,g) = \begin{bmatrix} \ln\beta_{10} + \ln\beta_{20} + \ln\left\{p\beta_{21}\exp\left(-\frac{t^\alpha}{\beta_{11}}\right) + (1-p)\beta_{11}\exp\left(-\frac{t^\alpha}{\beta_{21}}\right)\right\} - \ln\beta_{11} - \ln\beta_{21} \\ -\ln\left\{p\beta_{20}\exp\left(-\frac{t^\alpha}{\beta_{10}}\right) + (1-p)\beta_{10}\exp\left(-\frac{t^\alpha}{\beta_{20}}\right)\right\} \end{bmatrix} \left[p\frac{1}{\beta_{11}}\exp\left(-\frac{t}{\beta_{11}}\right) + (1-p)\frac{1}{\beta_{21}}\exp\left(-\frac{t}{\beta_{21}}\right)\right] \tag{2.4}$$

$$KL(f,g) = \begin{bmatrix} \ln\beta_{11} + \ln\beta_{21} + \ln\left\{p\beta_{20}\exp\left(-\frac{t^\alpha}{\beta_{10}}\right) + (1-p)\beta_{10}\exp\left(-\frac{t^\alpha}{\beta_{20}}\right)\right\} - \ln\beta_{10} - \ln\beta_{20} \\ -\ln\left\{p\beta_{21}\exp\left(-\frac{t^\alpha}{\beta_{11}}\right) + (1-p)\beta_{11}\exp\left(-\frac{t^\alpha}{\beta_{21}}\right)\right\} \end{bmatrix} \left[p\frac{1}{\beta_{10}}\exp\left(-\frac{t}{\beta_{10}}\right) + (1-p)\frac{1}{\beta_{20}}\exp\left(-\frac{t}{\beta_{20}}\right)\right]. \tag{2.5}$$

From (2.4) and (2.5), we can see that KL(g, f) > KL(f, g) i.e. the CSWMROC Curve is TPR asymmetric.

(e) The slope of the CSWMROC Curve at the cut off value t is given as

$$slope(t) = \frac{\beta_{10}\beta_{20}\left[p\beta_{21}\exp\left(-\frac{t^\alpha}{\beta_{11}}\right) + (1-p)\beta_{11}\exp\left(-\frac{t^\alpha}{\beta_{21}}\right)\right]}{\beta_{11}\beta_{21}\left[p\beta_{20}\exp\left(-\frac{t^\alpha}{\beta_{10}}\right) + (1-p)\beta_{10}\exp\left(-\frac{t^\alpha}{\beta_{20}}\right)\right]}. \tag{2.6}$$

The AUC of CSWMROC Curve is defined as

$$AUC = p\frac{\beta_{11}}{\beta_{10}+\beta_{11}} + (1-p)\frac{\beta_{21}}{\beta_{20}+\beta_{21}} = p(AUC_1) + (1-p)(AUC_2). \tag{2.7}$$

Optimal cut-off value

In medical diagnosis, the optimal cut-off value (t) tells us about the patient's situation whether his status of disease. The optimal cut-off value is defined by the Fluss *et al*. (2005) in the Youden index which is obtained by taking the maximum difference between the CDF of healthy and disease cases. The optimal threshold value or cut-off value of biomarker using CSWMROC curve is obtained as

$$t = p\max_t\left\{\left(\frac{\beta_{11}\beta_{10}}{\beta_{11}-\beta_{10}}\right)\ln\left(\frac{\beta_{11}}{\beta_{10}}\right)\right\} + (1-p)\max_t\left\{\left(\frac{\beta_{21}\beta_{20}}{\beta_{21}-\beta_{20}}\right)\ln\left(\frac{\beta_{21}}{\beta_{20}}\right)\right\}. \tag{2.8}$$

3. *Estimates of parameters of AUC of CSWMROC Curve using Method of Moments*

It is very old and easy method for estimating the parameters. The r^{th} sample moment of a mixture distribution is defined as

$$m_r^{'} = p\int_x x^r f_1(x)dx + (1-p)\int_x x^r f_2(x)dx \tag{3.1}$$

where, $f_1(x)$ and $f_2(x)$ are the densities of two sub-populations of mixture distribution. The r^{th} sample moment of Constant Shape Weibull Mixture distribution is obtained as

$$m_r^{'} = p\frac{\alpha}{\beta_1}\int_0^\infty x^{r+\alpha-1}\exp\left(-\frac{x^\alpha}{\beta_1}\right)dx + (1-p)\frac{\alpha}{\beta_2}\int_0^\infty x^{r+\alpha-1}\exp\left(-\frac{x^\alpha}{\beta_2}\right)dx. \tag{3.2}$$

The shape parameter α is constant for both sub populations and β_1 and β_2 are the scale parameters of Weibull Mixture distribution. On putting r=1, 2, 3, 4 in (3.2), we get

$$m_1^{'} = p\beta_{10}^{1/\alpha}\Gamma\left(1+\frac{1}{\alpha}\right) + (1-p)\beta_{20}^{1/\alpha}\Gamma\left(1+\frac{1}{\alpha}\right) \tag{3.3}$$

$$m_2^{'} = p\beta_{10}^{2/\alpha}\Gamma\left(1+\frac{2}{\alpha}\right) + (1-p)\beta_{20}^{2/\alpha}\Gamma\left(1+\frac{2}{\alpha}\right) \tag{3.4}$$

$$m_3^{'} = p\beta_{10}^{3/\alpha}\Gamma\left(1+\frac{3}{\alpha}\right) + (1-p)\beta_{20}^{3/\alpha}\Gamma\left(1+\frac{3}{\alpha}\right) \tag{3.5}$$

$$m_4^{'} = p\beta_{10}^{4/\alpha}\Gamma\left(1+\frac{4}{\alpha}\right) + (1-p)\beta_{20}^{4/\alpha}\Gamma\left(1+\frac{4}{\alpha}\right) \tag{3.6}$$

On solving (3.3)-(3.6), one can obtain $\hat{p}$, $\hat{\alpha}$, $\hat{\beta}_1$, and $\hat{\beta}_2$ by using the Newton Raphson method in MATHEMATICA software.

4. *Variance of AUC of CSWMROC Curve using delta method*

The approximate variance of AUC of CSWMROC Curve by Delta method gives the approximate variance as

$$V\left(\hat{AUC}\right) = pV\left(\hat{AUC}_1\right) + (1-p)V\left(\hat{AUC}_2\right) \tag{4.1}$$

where

$$AUC_1 = \frac{\beta_{11}}{\beta_{10}+\beta_{11}} \qquad \text{and} \qquad AUC_2 = \frac{\beta_{21}}{\beta_{20}+\beta_{21}}. \tag{4.2}$$

Using delta method, we have

$$\left.\begin{aligned} V\left(\hat{AUC}_1\right) &= \left(\frac{\partial AUC_1}{\partial \beta_{11}}\right)^2 V\left(\hat{\beta}_{11}\right) + \left(\frac{\partial AUC_1}{\partial \beta_{10}}\right)^2 V\left(\hat{\beta}_{10}\right) + 2\left(\frac{\partial AUC_1}{\partial \beta_{11}}\right)\left(\frac{\partial AUC_1}{\partial \beta_{10}}\right)\operatorname{cov}\left(\hat{\beta}_{11}, \hat{\beta}_{10}\right) \\ V\left(\hat{AUC}_2\right) &= \left(\frac{\partial AUC_2}{\partial \beta_{21}}\right)^2 V\left(\hat{\beta}_{21}\right) + \left(\frac{\partial AUC_2}{\partial \beta_{20}}\right)^2 V\left(\hat{\beta}_{20}\right) + 2\left(\frac{\partial AUC_2}{\partial \beta_{21}}\right)\left(\frac{\partial AUC_2}{\partial \beta_{20}}\right)\operatorname{cov}\left(\hat{\beta}_{21}, \hat{\beta}_{20}\right) \end{aligned}\right\} \tag{4.3}$$

On substituting (4.3) in (4.1), we get

$$\begin{aligned} V\left(\hat{AUC}\right) = p&\left[\left(\frac{\partial AUC_1}{\partial \beta_{11}}\right)^2 V\left(\hat{\beta}_{11}\right) + \left(\frac{\partial AUC_1}{\partial \beta_{10}}\right)^2 V\left(\hat{\beta}_{10}\right) + 2\left(\frac{\partial AUC_1}{\partial \beta_{11}}\right)\left(\frac{\partial AUC_1}{\partial \beta_{10}}\right)\operatorname{cov}\left(\hat{\beta}_{11}, \hat{\beta}_{10}\right)\right] + \\ (1-p)&\left[\left(\frac{\partial AUC_2}{\partial \beta_{21}}\right)^2 V\left(\hat{\beta}_{21}\right) + \left(\frac{\partial AUC_2}{\partial \beta_{20}}\right)^2 V\left(\hat{\beta}_{20}\right) + 2\left(\frac{\partial AUC_2}{\partial \beta_{21}}\right)\left(\frac{\partial AUC_2}{\partial \beta_{20}}\right)\operatorname{cov}\left(\hat{\beta}_{21}, \hat{\beta}_{20}\right)\right]. \end{aligned} \tag{4.4}$$

Differentiating (4.2) with respect to $\beta_{11}, \beta_{10}, \beta_{21}$ and β_{20}, we get

$$\frac{\partial AUC_1}{\partial \beta_{11}} = \frac{\beta_{10}}{(\beta_{11}+\beta_{10})^2}, \frac{\partial AUC_1}{\partial \beta_{10}} = -\frac{\beta_{11}}{(\beta_{11}+\beta_{10})^2}, \frac{\partial AUC_2}{\partial \beta_{21}} = \frac{\beta_{20}}{(\beta_{21}+\beta_{20})^2}, \frac{\partial AUC_2}{\partial \beta_{20}} = -\frac{\beta_{21}}{(\beta_{21}+\beta_{20})^2}$$

To determine the variance of $\hat{\beta}_{11}, \hat{\beta}_{10}, \hat{\beta}_{21}$ and $\hat{\beta}_{20}$, we use the Fisher information matrix. The likelihood function is given as

$$L = \prod_{i=1}^{m_{10}} f_{10}(x) \prod_{j=1}^{n_{11}} g_{11}(y).$$

The Fisher information matrix is given as

$$I(\theta_1) = \begin{bmatrix} E\left(\frac{\partial^2 \ln L}{\partial \alpha^2}\right) & E\left(\frac{\partial^2 \ln L}{\partial \alpha \partial \beta_{11}}\right) & E\left(\frac{\partial^2 \ln L}{\partial \alpha \partial \beta_{10}}\right) \\ E\left(\frac{\partial^2 \ln L}{\partial \beta_{11} \partial \alpha}\right) & E\left(\frac{\partial^2 \ln L}{\partial \beta_{11}^2}\right) & E\left(\frac{\partial^2 \ln L}{\partial \beta_{11} \partial \beta_{10}}\right) \\ E\left(\frac{\partial^2 \ln L}{\partial \beta_{10} \partial \alpha}\right) & E\left(\frac{\partial^2 \ln L}{\partial \beta_{10} \partial \beta_{11}}\right) & E\left(\frac{\partial^2 \ln L}{\partial \beta_{10}^2}\right) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{4.5}$$

where,

$$\theta_1 = (\alpha, \beta_{10}, \beta_{11})$$

$$\left.\begin{aligned} a_{11} &= \left[(m_{10}+n_{11})(1+\Gamma_2^{'}) + 2(n_{11}\ln\beta_{11} + m_{10}\ln\beta_{10})\Gamma_2^{'} + n_{11}(\ln\beta_{11})^2 + m_{10}(\ln\beta_{10})^2\right] \\ a_{22} &= \frac{n_{11}}{\beta_{11}^2}, a_{33} = \frac{m_{10}}{\beta_{10}^2}, a_{12} = a_{21} = -\frac{n_{11}}{\alpha\beta_{11}}\left(\Gamma_2^{'} + \ln\beta_{11}\right), a_{23} = a_{32} = 0 \\ a_{13} &= a_{31} = -\frac{m_{10}}{\alpha\beta_{10}}\left(\Gamma_2^{'} + \ln\beta_{10}\right) \end{aligned}\right\} \quad (4.6)$$

and

$$\Gamma_n^{'} = -(n-1)!\left[\frac{1}{n} + \gamma - \sum_{k=1}^{n}\frac{1}{k}\right],$$

γ : Euler-Mascheroni constant approximately equal to 0.5772

m_{10}, m_{20}: sample sizes of healthy controls

n_{11}, n_{21} : sample sizes of disease cases.

On substituting(4.6) in (4.5), the inverse Fisher Information matrix is given as

$$I^{-1}(\theta_1) = \frac{1}{a_{11}a_{22}a_{33} - a_{12}^2 a_{33} - a_{22}a_{13}^2}\begin{bmatrix} a_{22}a_{33} & -a_{21}a_{33} & -a_{22}a_{31} \\ -a_{12}a_{33} & a_{11}a_{33} - a_{13}^2 & a_{12}a_{31} \\ -a_{22}a_{13} & a_{21}a_{13} & a_{11}a_{22} - a_{12}^2 \end{bmatrix}$$

$$= \begin{bmatrix} V(\hat{\alpha}) & COV(\hat{\alpha}, \hat{\beta}_{11}) & COV(\hat{\alpha}, \hat{\beta}_{10}) \\ COV(\hat{\beta}_{11}, \hat{\alpha}) & V(\hat{\beta}_{11}) & COV(\hat{\beta}_{11}, \hat{\beta}_{10}) \\ COV(\hat{\beta}_{10}, \hat{\alpha}) & COV(\hat{\beta}_{10}, \hat{\beta}_{11}) & V(\hat{\beta}_{10}) \end{bmatrix} \quad (4.7)$$

where

$$\left.\begin{aligned} V(\hat{\alpha}) &= \frac{\alpha^2}{(m_{10}+n_{11})\left(1-\Gamma_2^{''} - \left(\Gamma_2^{'2}\right)\right)}, \operatorname{cov}(\hat{\alpha}, \hat{\beta}_{10}) = \frac{\alpha\beta_{10}\left(\Gamma_2^{'} + \ln\beta_{10}\right)}{(m_{10}+n_{11})}, \\ V(\hat{\beta}_{11}) &= \frac{\beta_{11}^2\left[n_{11}(n_{11}+m_{10})(1+\Gamma_2^{''}) + 2n_{11}m_{10}\log(\beta_{11})\Gamma_2^{'} + m_{10}n_{11}\left[\log(\beta_{11})^2 - m_{10}^2\left(\Gamma_2^{'}\right)^2\right]\right]}{n_{11}m_{10}(n_{11}+m_{10})\left(1-\Gamma_2^{''} - \left(\Gamma_2^{'}\right)^2\right)}, \\ V(\hat{\beta}_{10}) &= \frac{\beta_{10}^2\left[n_{11}(n_{11}+m_{10})(1+\Gamma_2^{''}) + 2n_{11}m_{10}\log(\beta_{10})\Gamma_2^{'} + m_{10}n_{11}\left[\log(\beta_{10})^2 - n_{11}^2\left(\Gamma_2^{'}\right)^2\right]\right]}{n_{11}m_{10}(n_{11}+m_{10})\left(1-\Gamma_2^{''} - \left(\Gamma_2^{'}\right)^2\right)}, \\ \operatorname{cov}(\hat{\alpha}, \hat{\beta}_{11}) &= \frac{\alpha\beta_{11}\left(\Gamma_2^{'} + \ln\beta_{11}\right)}{(m_{10}+n_{11})\left(1-\Gamma_2^{''} - \left(\Gamma_2^{'}\right)^2\right)}, \operatorname{cov}(\hat{\beta}_{11}, \hat{\beta}_{10}) = \frac{\beta_{10}\beta_{11}\left(\Gamma_2^{'} + \ln\beta_{10}\right)\left(\Gamma_2^{'} + \ln\beta_{11}\right)}{(m_{10}+n_{11})\left(1-\Gamma_2^{''} - \left(\Gamma_2^{'}\right)^2\right)} \\ \operatorname{cov}(\hat{\alpha}, \hat{\beta}_{10}) &= \frac{\alpha\beta_{10}\left(\Gamma_2^{'} + \ln\beta_{10}\right)}{(m_{10}+n_{11})\left(1-\Gamma_2^{''} - \left(\Gamma_2^{'}\right)^2\right)}. \end{aligned}\right\} \quad (4.8)$$

On putting (4.8) in (4.3), $V\left(A\hat{U}C_1\right)$ is given as

$$V\left(A\hat{U}C_1\right)=\left[\frac{\beta_{11}^2\beta_{10}^2}{\left(\beta_{11}+\beta_{10}\right)^4}\left\{\left(\frac{\left(n_{11}+m_{10}\right)}{n_{11}m_{10}}\right)+\frac{\left[\ln\left(\frac{\beta_{10}}{\beta_{11}}\right)\right]^2}{\left(n_{11}+m_{10}\right)\left(1+\Gamma_2^{''}-\Gamma_2^{'2}\right)}\right\}\right]. \tag{4.9}$$

Similarly, the $V\left(A\hat{U}C_2\right)$ is given as

$$V\left(A\hat{U}C_2\right)=\left[\frac{\beta_{21}^2\beta_{20}^2}{\left(\beta_{21}+\beta_{20}\right)^4}\left\{\left(\frac{\left(n_{21}+m_{20}\right)}{n_{21}m_{20}}\right)+\frac{\left[\ln\left(\frac{\beta_{20}}{\beta_{21}}\right)\right]^2}{\left(n_{21}+m_{20}\right)\left(1+\Gamma_2^{''}-\Gamma_2^{'2}\right)}\right\}\right]. \tag{4.10}$$

On substituting (4.9) and (4.10) in (4.1), we get

$$\begin{aligned}V\left(A\hat{U}C\right)= p&\left[\frac{\beta_{10}^2\beta_{11}^2}{\left(\beta_{10}+\beta_{11}\right)^4}\left(\frac{\left(m_{11}+n_{10}\right)}{m_{11}n_{10}}+\frac{\left(\ln\left(\frac{\beta_{10}}{\beta_{11}}\right)\right)^2}{\left(m_{11}+n_{10}\right)\left(1+\Gamma_2^{''}-\Gamma_2^{'2}\right)}\right)\right]+\\ &\left(1-p\right)\left[\frac{\beta_{20}^2\beta_{21}^2}{\left(\beta_{20}+\beta_{21}\right)^4}\left(\frac{\left(m_{21}+n_{20}\right)}{m_{21}n_{20}}+\frac{\left(\ln\left(\frac{\beta_{20}}{\beta_{21}}\right)\right)^2}{\left(m_{21}+n_{20}\right)\left(1+\Gamma_2^{''}-\Gamma_2^{'2}\right)}\right)\right].\end{aligned} \tag{4.11}$$

Using V(AUC), one can easily find confidence interval, MSE and test of significance for AUC.

(i) The 100(1-α)% confidence interval of AUC is given as

$$AUC\pm SE\left(A\hat{U}C\right)Z_{\alpha/2} \tag{4.12}$$

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

(ii) The Mean Square Error (MSE) is used to identify the quality of an estimator. It is defined as

$$MSE\left(A\hat{U}C\right)=Variance\left(A\hat{U}C\right)+Bias\left(A\hat{U}C\right)^2 \tag{4.13}$$

where

$$Bias\left(\hat{AUC}\right) = E\left(\hat{AUC}\right) - AUC.$$

(iii) Consider the problem of testing of AUC of CSWMROC Curve as

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

The test statistic is given as

$$Z = \frac{\sqrt{N}\left(\hat{AUC} - AUC\right)}{\sqrt{V\left(\hat{AUC}\right)}} \sim N(0,1) \tag{4.14}$$

where N = m + n, m is the sample size of healthy controls and n is the sample size of disease cases.

5. Simulation Studies

The random numbers are generated from Weibull mixture distribution with fixed values of shape parameter and scale parameters of healthy controls and disease cases for the sample sizes N=10, 20, 30, 100, 200 and 300. The sample sizes are equal for healthy controls and disease cases. The value of weight of healthy controls and disease case are also taken as equal i.e., p=0.7. The values of shape parameter is same for healthy controls and disease cases $\alpha = \alpha_{10} = \alpha_{20} = \alpha_{11} = \alpha_{21} = 2$. The values of scale parameters of healthy controls and disease cases are $\beta_{10} = 1,\ \beta_{20} = 2$ and $\beta_{11} = 10,\ \beta_{21} = 9$.

Table 5.1: ***Estimates of parameters of AUC of CSWMROC Curve by MOM for different sample sizes***

N	$\hat{\beta}_{10}$	$\hat{\beta}_{20}$	$\hat{\beta}_{11}$	$\hat{\beta}_{21}$
10	1.245	1.418	72.875	58.75
20	1.933	6.744	76.552	17.449
30	0.331	3.654	89.311	23.110
100	0.543	3.517	43.096	37.776
200	0.835	3.186	13.191	10.871
300	0.615	3.655	14.767	9.780

From above table, it is observed that with the increase in sample size, the estimates of CSWMROC model become closer to the parameters. Using the estimators in Table 5.1, one can see $\hat{AUC}$, $V\left(\hat{AUC}\right)$, $SE\left(\hat{AUC}\right)$, confidence interval and Z-values to test $AUC_0 = 0.88$ in Table 5.2.

Table 5.2: $\hat{AUC}$, $V(\hat{AUC})$, $SE(\hat{AUC})$, $MSE(\hat{AUC})$, ***95% Confidence Interval (CI) of*** $\hat{AUC}$ ***and Z-values***

N	$\hat{AUC}$	$V(\hat{AUC})$	$SE(\hat{AUC})$	$MSE(\hat{AUC})$	$CI(\hat{AUC})$	Z-values
10	0.981	0.0002	0.0152	0.0101	[0.951, 1.01]	31.939
20	0.899	0.0015	0.0387	0.0018	[0.823, 0.975]	3.102
30	0.956	0.0004	0.0206	0.0059	[0.916, 0.996]	29.434
100	0.965	0.0000	0.0087	0.0071	[0.948, 0.982]	120.208
200	0.890	0.0001	0.0126	0.0002	[0.865, 0.915]	20.00
300	0.890	0.0001	0.0103	0.0001	[0.870, 0.910]	24.494

It is observed from Table 5.2 that $\hat{AUC}$ become closer to the true value of AUC as the sample size increases but $V(\hat{AUC})$, $SE(\hat{AUC})$ and MSE(AUC) decreases with increase in the sample size because the variance of AUC and standard error of AUC depends on sample sizes. From Z values, one can see that all $\hat{AUC}$ values are greater than 0.88, so we reject the null hypothesis and concludes that AUC is not equal to 0.88.

Fig. 5.1 shows the CSWMROC curves for different sample sizes and fixed values of parameters mentioned above.

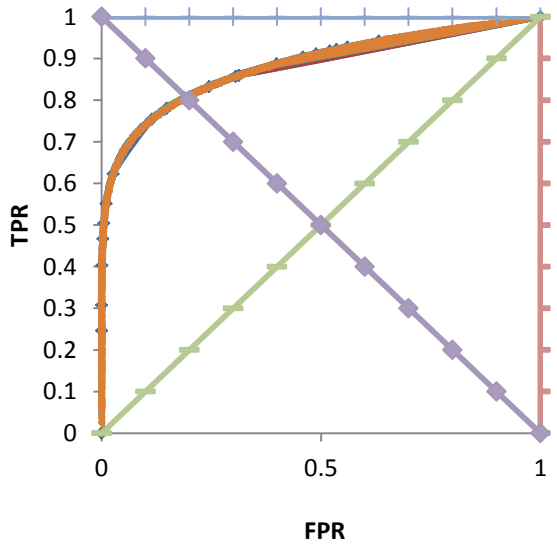

Fig. 5.1 CSWMROC Curve with fixed parameters and different sample sizes.

6. Conclusion

In this paper, we have proposed CSWMROC model and found that CSWMROC curve is monotonically increasing, concave in nature and TPR asymmetric. The Area under the CSWMROC Curve, its variance and the optimal cut-off value of biomarker using CSWMROC Curve are also derived. The estimates of parameters of AUC are obtained by MOM. The variance of AUC of ROC Curve is also derived. The MSE of AUC, confidence interval of AUC and test for AUC are also discussed. From simulation studies, it is concluded that the estimates of parameters of AUC of CSWMROC Curve using MOM become approximately closer to the population parameters for large sample size. It is concluded that when heterogeneity is found in the data and Weibull mixture distribution fits well to the data then one should use Weibull mixture ROC model instead of Bi-Weibull ROC model.

References

1. Arfa M. and Aslam M. (2008). "A comparative study to estimate the parameters of mixed-weibull distribution", Pak.J.Stat.Oper.Res., 4(1), 1-8.

2. Atienza N., Garcia-Heras J., and Munoz Pichardo J. m. (2006). "A new condition for Identifiability of finite mixture distributions". Metrika (2006) 63:215-221, Doi 10.1007/s00184-005-0013-z.

3. Bucar T., Nagode M., and Fajdiga M. (2003). "Reliability approximation using finite Weibull mixture distributions", Reliability Engineering and System Safety, 84, 241–251.

4. Dass S.C. and Kim S.W. (2011). "A Semi-parameteric Approach to Estimation of ROC Curves for Multivariate Binormal Mixtures", www.stt.msu.edu/Links/Research_Memoranda/RM/RM_683.pdf.

5. *Dewan I. and Nandi S. (2009). "An em algorithm for estimating the parameters of bivariate weibull distribution under random censoring",* www.isid.ac.in/~statmath/eprints.

6. Dwidayati N. *et al* (2013). "Estimation of the Parameters of a Mixture Weibull Model for Analyze Cure Rate", Applied Mathematical Sciences, 7(116), 5767 – 5778.

7. Egan J.P. (1975). "Signal Detection Theory and ROC Analysis*"*, Academic Press, New York.

8. Erisoglu U. and Erisoglu M. (2014). "L-Moments estimates for mixture of weibull distributions", Journal of data science 12, 69-85.

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

10. Fluss, R., Faraggi, D. and Reiser, B. (2005). "Estimation of the Youden Index and its associated cutoff point", Biometrical Journal, 47(4), 458-472.

11. Gonen M. (2013). "Mixtures of Receiver Operating Characteristic Curves. Memorial Sloan-Katter in Cancer", Doi: 10.1016/7 acra.2013.03.003.

12. Green D.M. and Swets J.A. (1966). "Signal Detection Theory and Psychophysics*"*, Wiley, New York.

13. Kao J.H.K. (1959). "A graphical estimation of mixed weibull parameters in life-testing of electron tubes", Technometrics 1, 389-407.

14. Krzanowski W.J. and Hand D.J. (2002). "ROC curves for continuous data, Monographs on Statistics and Applied Probability", CRC Press, Taylor and Francis Group, New York.

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

16. Newcomb S. (1886). "A generalized theory of the combination of observations so as to obtain the best result", Amer. J. Math, 8, 343-346.

17. Pearson K. (1894). "Contribution to the mathematical theory of evolution", Phil. Trans. Roy. Soc. *A*, 185, 71-110.

18. Pundir S. and Amala R. (2014). Evaluation of area under the Constant Shape Bi-Weibull ROC Curve, Journal of Modern Applied Statistical Methods, 13(1), 305-328.

19. Pundir S. and Azharuddin (2014). "A Study of Exponential Mixture ROC Model, 3rd International Conference on Innovative Approach in Applied Physical, Mathematical/Statistical, Chemical Sciences and Emerging Energy Technology for sustainable Development", 18-26.

20. Pundir S. and Azharuddin (2016). "Detection of AUC and Confidence Interval using Normal Mixture ROC Curve", IOSR Journal of Mathematics (IOSR-JM), e-ISSN: 2278-5728, p-ISSN: 2319-765X. 12(2), 77-86.

21. Teicher, H. (1961). "Identifiability of Mixtures". The Annals of Mathematical Statistics 32, 244-248.

22. Teicher, H. (1963). "Identifiability of finite Mixtures", The Annals of Mathematical Statistics 34, 1265-1269.

23. Titterington D.M. *et.al* (1985). "Statistical Analysis of Finite Mixture Distributions", John Wiley and Sons Ltd.

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

25. Zhou X.H. *et.al*. (2002). "Statistical Methods in Diagnostic Medicine*"*, Wiley, New York.

Programs

(a) R-Command

The random numbers are generated by using the following command

```
y<-p*runif(n)+(1-p)*runif(n)
x<-p*((log((1-y)^(-b1)))^(1/a))+(1-p)*((log((1-y)^(-b2)))^(1/a))
```

n: sample size, a: shape parameter, b1: scale parameter of 1st sub-population, b2: scale parameter of 2nd sub-population.

(b) MATHEMATICA Command

The moment estimators of AUC of CSWMROC Curve are obtained by using the following command

```
FindRoot[{p*b1^(1/\[Alpha])*Gamma[1 + 1/\[Alpha]] + (1 - p)*b2^(
   1/\[Alpha])*Gamma[1 + 1/\[Alpha]] == m1,
 p*b1^(2/\[Alpha])*Gamma[1 + 2/\[Alpha]] + (1 - p)*b2^(
   2/\[Alpha])*Gamma[1 + 2/\[Alpha]] == m2,
 p*b1^(3/\[Alpha])*Gamma[1 + 3/\[Alpha]] + (1 - p)*b2^(
   3/\[Alpha])*Gamma[1 + 3/\[Alpha]] == m3,
 p*b10^(4/\[Alpha])*Gamma[1 + 4/\[Alpha]] + (1 - p)*b2^(
   4/\[Alpha])*Gamma[1 + 4/\[Alpha]] == m4}, {p,p0
 }, {\[Alpha], [Alpha0]}, {b1, b10}, {b2, b20}]
```

(c) Euler-Mascheroni Constant

The first order differentiation of is given as $\Gamma_n \psi(n)$ where $\psi(n)$ called the digamma function. The value of $\Gamma_n^{'}$ at n is equal to 1-γ', where γ is Euler-Mascheroni constant and its approximate value is 0.5772. The second order differentiation at n is define as

$\Gamma_n^{''} = \int_0^\infty x^{n-1} e^{-x} (\log x)^2 \, dx$ has the value $-1+(1-\gamma)^2 + \frac{\pi^2}{6}$.

The general m[th] derivative of Γ_n is obtained by

$\Gamma_n^m = \int_0^\infty x^{n-1} e^{-x} (\log x)^m \, dx.$