International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Confidence Intervals Estimation for ROC Curve, AUC and Brier Score under the Constant Shape Bi-Weibull Distribution

A. Lavanya[1], T. Leo Alexander[2]

[1]Research Scholar, Department of Statistics, Loyola College, Chennai–34, India

[2]Associate Professor, Department of Statistics, Loyola College, Chennai–34, India

Abstract: ***The accuracy of diagnostic test is typically evaluated by sensitivity and specificity. Receiver Operating Characteristic (ROC) curve analysis is one of the most familiar techniques and it will provide accuracy for the extent of correct classification of a test and it is a graphical representation of the relationship between sensitivity and specificity. The conventional way of expressing the true accuracy of test is by using its summary measures Area Under the Curve (AUC) and Brier Score ($\overline{B}$). Hence the main issue in assessing the accuracy of a diagnostic test is to estimate the ROC curve and its AUC and Brier Score. The ROC curve generated based on assuming a Constant Shape Bi-Weibull distribution. This article assumes that the biomarker values from the two groups follow Weibull distributions with equal shape parameter and different scale parameters. The ROC model, AUC, MLE, asymptotic, bootstrap confidence intervals for the AUC, asymptotic confidence intervals for the ROC curve and Brier Score are derived. However, the accuracy of a test is to be explained by involving the scale and shape parameters. Theoretical results are validated by simulation studies. An illustrative example is also provided to explain the concepts.***

Keywords: AUC, Brier Score, Constant Shape Bi-Weibull ROC model, Confidence Interval, MLE, Parametric bootstrap variance

1. Introduction

In medical science, a diagnostic test result is called a biomarker [1, 7] is an indicator for disease status of patients. The accuracy of a medical diagnostic test is typically evaluated by sensitivity and specificity. ROC curve is a graphical representation of the relationship between sensitivity and specificity. The AUC is an overall performance measure for the biomarker. Brier Score ($\overline{B}$) is shown as another summary measure in the context of ROC Curve to make the probabilistic judgements as well as to identify the extent of classification. Hence the main issue in assessing the accuracy of a diagnostic test is to estimate the ROC curve and its AUC and Brier Score. The ROC curve can be plotted by three approaches viz. parametric, non-parametric and semi-parametric. This article considers the parametric way of plotting the ROC curve. After the ROC curve is generated the intrinsic accuracy provided by the biomarker must be interpreted. To summarize the information contained in a ROC curve, many indices have been used. Among them, AUC curve is most commonly adopted index. In recent past work we developed Functional Relationship between Brier Score and Area Under the Constant Shape Bi-Weibull ROC Curve [10]. In this article, the inference about the AUC is of primary interest. The problem of assessing the accuracy of diagnosis/Biomarker has been studied by several authors by assuming various distributions to the biomarker values. They are Bi-Normal ROC model [17], Bi-Logistic ROC model [12], Bi-Lomax ROC model [3], Bi-Gamma ROC model [5],Bi-Exponential ROC model [2], Generalized Bi-Exponential ROC model [8], Bi-Rayleigh ROC model and its comparison with Bi-Normal model [9], comparison of Bi-Rayleigh ROC model with Bi-Normal and Bi-Gamma ROC models [13] and are view of all parametric ROC models in case of continuous data [14], Normal-Exponential [15]. Two parameter Weibull distribution is a most widely used life distribution in various fields viz. Survival analysis, Reliability engineering and recently in ROC curve analysis. One major disadvantage of assuming two parameter Weibull distributions to the biomarker is that the accuracy cannot be expressed in closed form. By substituting the MLE‟s, the accuracy can be evaluated numerically using Monte Carlo integration or any other numerical procedure. In the absence of closed form expression, the statistical inference on the accuracy measure will not be possible. To overcome this problem and to obtain a closed form expression, equal shape parameter and different scale parameters are assumed. Moreover, the original accuracy of the diagnosis is not affected by taking equal shape parameter. The ROC model developed from this assumption is called the constant shape Bi-Weibull ROC model. Research interest may lie in comparing the effectiveness of two separate diagnostic tests or the efficiency of biomarkers in predicting the disease. The comparison can be accomplished either by AUC or sensitivity of the test. In order to compare the AUC and to construct the confidence interval, the Standard Error (SE) of AUC is needed. Also the asymptotic confidence intervals for the ROC curve and Brier Score are derived. Here, the standard error of accuracy is studied by different methods viz. asymptotic MLE, parametric bootstrap methods. For parametric, the delta method will yield variance and SE with the help of asymptotic expressions for the variance and co-variances of the parameters.

2. A Constant Shape Bi-Weibull ROC Model and its AUC and Brier Score

Let x, *y* be the test scores observed from two populations with (abnormal individuals) and without (normal individuals) condition respectively which follow Constant Shape Bi-

Weibull distributions. The density functions of Constant Shape Bi-Weibull distributions are as follows,

$$f(x|n) = \frac{\beta}{\sigma_n} x^{\beta-1} e^{-\left[\frac{x^\beta}{\sigma_n}\right]} \quad (1)$$

and $$f(y|s) = \frac{\beta}{\sigma_s} y^{\beta-1} e^{-\left[\frac{y^\beta}{\sigma_s}\right]} \quad (2)$$

Let S be a continuous biomarker. The probabilistic definitions of the measures of ROC Curve are as follows:

$$\text{Sensitivity}(s_n) = P(S|s) = \int_t^\infty f(x|s)dx \quad ,$$

$$1 - \text{Specificity}(s_p) = P(S|n) = \int_t^\infty f(x|n)dx \quad .$$

In this context, the (1-Specificity) or False Positive Rate (FPR) and Sensitivity or True Positive Rate (TPR) can be defined using equations (1) and (2) and are given in equations (3) and (4) respectively,

$$P(S|n) = x(t) = e^{-\left[\frac{t^\beta}{\sigma_n}\right]} \quad (3)$$

and $$P(S|s) = y(t) = e^{-\left[\frac{t^\beta}{\sigma_s}\right]} \quad (4)$$

The ROC Curve is defined as a function of (1-Specificity) with scale parameters of distributions and is given as,

$$ROC(t) = y(t) = x(t)^{\frac{\sigma_n}{\sigma_s}} \quad (5)$$

where $t = -[\sigma_n log x(t)]^{\frac{1}{\beta}}$ is the threshold.

The accuracy of a diagnostic test can be explained using the Area Under the Curve (AUC) of an ROC Curve.AUC describes the ability of the test to discriminate between abnormal and normal populations. A natural measure of the performance of the classifier producing the curve is AUC. This will range from 0.5 for a random classifier to 1 for a perfect classifier. The AUC is defined as,

$$AUC = \int_0^1 x(t)^{\frac{\sigma_n}{\sigma_s}} dx(t) \quad (6)$$

The closed form of AUC is as follows

$$AUC = \frac{\sigma_s}{\sigma_s + \sigma_n} \quad (7)$$

It is a well known evaluation measure for probabilistic classifiers and is proposed by Brier in 1950. In literature, procedures for calibrating classifiers have been proposed in different contexts such as Weather Prediction Tasks [4], Game Theory [6].

Further, Brier Score was considered in the context of signal Detection Theory by assuming that the calibration in the observers probability estimate is perfect and provided the theoretical relationship between Brier Score and Area Under the Binormal ROC Curve [11]. So Functional relationship between Brier Score and Area Under the ROC Curve was discussed in the Context of Skewed Distributions [16].

Now consider a set of M signal detection tasks with αM signal events and (1-α)M non-signal events (0≤α≤1) and αM, (1-α)M and α denote a priori probability of signal events.

Let $y_i = 0$ if the event is non-signal
$y_i = 1$ if the event is signal.

Let p_i denote the observers (or subject‟s) probability estimate that the i^{th} event will be the signal one, where the subscript i indicates the individual event [6] and the Brier Score ($\bar{B}$) can be defined as [7] using the expression

$$\bar{B} = \frac{1}{M}\sum_{i=1}^{M}(y_i - p_i)^2 \quad ,$$

where p_i is a function of x_i and is defined using Bayes theorem as follows,

$$p_i = p(x_i) = \frac{\alpha f(x_i|s)}{\alpha f(x_i|s) + (1-\alpha) f(x_i|n)}.$$

Now, we consider the expected value of $(y_i - p_i)^2$ in the expression $\bar{B}$, when the calibration in the observers probability estimate is perfect. In this case, p_i in the expression $\bar{B}$ is obtained by expression p_i from the Bayes theorem. Therefore, expected value of $(y_i - p_i)^2$is given as

$$E[(y_i - p_i)^2] = \int_{-\infty}^{\infty}(1-p(x))^2 \alpha f(x|s)dx + \int_{-\infty}^{\infty} p(x)^2(1-\alpha) f(x|n)dx$$

On simplification, we get,

$$E[(y_i - p_i)^2] = \int_{-\infty}^{\infty} \frac{\alpha(1-\alpha) f(x|n) f(x|s)}{(1-\alpha) f(x|n) + \alpha f(x|s)} dx.$$

Further, we assume that the convergence in probability of $\bar{B}$ given by the law of large numbers as M tends to infinity is the Brier score. Make the assumption that the calibration in the observer‟s probability estimate is perfect. That is, Expected Brier Score is equal to $\bar{B}$.

Therefore, [11] considered the Brier Score in the context of Signal Detection Theory by assuming the calibration in the observers probability estimate is perfect and provided the theoretical relationship between Brier Score and AUC. Therefore, Brier Score can be defined in the context of ROC Curve analysis as,

$$\bar{B} = \int_{-\infty}^{\infty} \frac{\alpha(1-\alpha) f(x|s)}{(1-\alpha) + \alpha \frac{f(x|s)}{f(x|n)}} dx \ . \quad \text{(with slope ROC Curve)}$$

or

$$\bar{B} = \int_{-\infty}^{\infty} \frac{\alpha(1-\alpha) f(x|n)}{\alpha + (1-\alpha) \frac{f(x|n)}{f(x|s)}} dx \ . \quad \text{(with inverse slope)}$$

Now we can describe $\bar{B}$ and AUC as the function of $\mathbf{b}=\frac{\sigma_n}{\sigma_s}$ and $\boldsymbol{\alpha}$ as

$$\bar{B} = \int_{-\infty}^{\infty} \frac{\alpha(1-\alpha)}{(1-\alpha) + \alpha \frac{dP(x|s)}{dP(x|n)}} dP(S|s) \ . \text{(slope)}$$

or

$$\bar{B} = \int_{-\infty}^{\infty} \frac{\alpha(1-\alpha)}{\alpha + (1-\alpha) \frac{dP(x|n)}{dP(x|s)}} dP(S|n). \text{(inverse slope)}$$

Therefore, the Brier Score can be defined as the function of likelihood ratio under a Bayesian set up. Now the expression for $\bar{B}$ as a function of **b** and **α** in the context of ROC Curve is follows,

$$\bar{B} = \int_0^1 \frac{\alpha(1-\alpha)}{\alpha + (1-\alpha)\left[\frac{1}{b} exp\left\{\left[\frac{t^\beta}{\sigma_s}\right] - \left[\frac{t^\beta}{\sigma_n}\right]\right\}\right]} dt \quad .$$

Now by substituting the threshold „t" value in the above equation and on further simplification, the Brier Score ($\bar{B}$) expression which is a function of **b** and **α** reduces to

$$\bar{B} = \int_0^1 \frac{\alpha(1-\alpha)}{\alpha + (1-\alpha)\left[\frac{1}{b} exp\left\{(1-b)\left[\left[logx(t)\right]^{\frac{1}{\beta}}\right]^\beta\right\}\right]} dx(t) \quad . \qquad (8)$$

The MLEs of σ_n and σ_s can be used again to estimate the AUC. The performance of the estimator $\widehat{AUC}$ can be accessed through variance estimate.

2.1 Maximum Likelihood Estimator of AUC

The MLE of two parameter Weibull distribution has been discussed [17] in the context of Reliability estimation. Let $X_1, X_2, \ldots X_m$ be a random sample of size m from $W(\beta, \sigma_n)$ and $Y_1, Y_2, \ldots Y_n$ be a random sample of size n from $W(\beta, \sigma_s)$ .The likelihood function of the selected sample is given by

$$L(x_i, y_j \mid \theta) = \prod_{i=1}^m f_X(x_i \mid \beta, \sigma_n) \prod_{j=1}^n f_Y(y_j \mid \beta, \sigma_s) \; .$$

where $\theta = (\beta, \sigma_n, \sigma_s)'$

$$L = \prod_{i=1}^m \frac{\beta}{\sigma_n} x_i^{\beta-1} e^{-\left[\frac{x_i^\beta}{\sigma_n}\right]} \prod_{j=1}^n \frac{\beta}{\sigma_s} y_j^{\beta-1} e^{-\left[\frac{y_j^\beta}{\sigma_s}\right]}$$

The log-likelihood function is

$$lnL = (m+n)ln\beta + (\beta-1)\left[\sum_{i=1}^m lnx_i + \sum_{j=1}^n lny_j\right] - nln\sigma_s - mln\sigma_n - \frac{1}{\sigma_s}\sum_{j=1}^n y_j^\beta - \frac{1}{\sigma_n}\sum_{i=1}^m x_i^\beta \; . \qquad (9)$$

Differentiating (9) with respect to β, we get

$$\frac{\partial lnL}{\partial \beta} = \frac{(m+n)}{\beta} + \left[\sum_{i=1}^m lnx_i + \sum_{j=1}^n lny_j\right] - \frac{1}{\sigma_s}\sum_{j=1}^n y_j^\beta lny_j - \frac{1}{\sigma_n}\sum_{i=1}^m x_i^\beta lnx_i \quad . \qquad (10)$$

By differentiating the equation (9) with respect to σ_n , σ_s and equating to zero, we get the estimates. The MLE"s of σ_n and σ_s are determined as,

$$\frac{\partial lnL}{\partial \sigma_n} = -\frac{m}{\sigma_n} + \frac{\sum_{i=1}^m x_i^\beta}{\sigma_n^2}$$

$$\widehat{\sigma_n} = \frac{\sum_{i=1}^m x_i^\beta}{m} \quad . \qquad (11)$$

$$\frac{\partial lnL}{\partial \sigma_s} = -\frac{n}{\sigma_s} + \frac{\sum_{j=1}^n y_j^\beta}{\sigma_s^2}$$

$$\widehat{\sigma_s} = \frac{\sum_{j=1}^n y_j^\beta}{n} \quad . \qquad (12)$$

Substituting equation (11) and (12) in equation (10) and equating it to zero, we get a non-linear equation:

$$h(\beta) = \frac{m + n + \sum_{j=1}^n y_j^\beta + \sum_{i=1}^m x_i^\beta}{\frac{n\sum_{j=1}^n y_j^\beta lny_j}{\sum_{j=1}^n y_j^\beta} + \frac{m\sum_{i=1}^m x_i^\beta lnx_i}{\sum_{i=1}^m x_i^\beta}} \; . \qquad (13)$$

Hence, $\hat{\beta}$ can be determined as a solution of Non-linear equation (13). By substituting equations (11), (12) and (13) in equation (7), will get an MLE estimate of AUC.

2.2 Asymptotic Property of Area Under the Constant Shape Bi-Weibull ROC Model

To evaluate the significance of the statistic AUC, its variance and standard error must be computed. Let $L(\theta \mid x, y)$; $\theta = (\beta, \sigma_n, \sigma_s)'$ be the likelihood function of the sample observations from X and Y which is given by

$$lnL(\theta \mid x, y) = (m+n)ln\beta - nln\sigma_s - mln\sigma_n + (\beta-1)\left[\sum_{i=1}^m lnx_i + \sum_{j=1}^n lny_j\right] - \frac{1}{\sigma_s}\sum_{j=1}^n y_j^\beta - \frac{1}{\sigma_n}\sum_{i=1}^m x_i^\beta \; . \qquad (14)$$

Asymptotic normality of MLE states that a consistent solution of the likelihood equation is asymptotically normally distributed about the true value θ, that is $\hat{\theta} \sim N(\theta, I^{-1}(\theta))$.

$$=> \sqrt{N}(\hat{\theta} - \theta) \sim N(0, I^{-1}(\theta)) \quad . \qquad (15)$$

where $I(\theta)$ is the Fisher Information matrix which is given by

$$I(\theta) = -\begin{bmatrix} E\left(\frac{\partial^2 lnL}{\partial \beta^2}\right) & E\left(\frac{\partial^2 lnL}{\partial \beta\, \partial \sigma_n}\right) & E\left(\frac{\partial^2 lnL}{\partial \beta\, \partial \sigma_s}\right) \\ E\left(\frac{\partial^2 lnL}{\partial \sigma_n\, \partial \beta}\right) & E\left(\frac{\partial^2 lnL}{\partial \sigma_n^2}\right) & E\left(\frac{\partial^2 lnL}{\partial \sigma_n\, \partial \sigma_s}\right) \\ E\left(\frac{\partial^2 lnL}{\partial \sigma_s\, \partial \beta}\right) & E\left(\frac{\partial^2 lnL}{\partial \sigma_s\, \partial \sigma_n}\right) & E\left(\frac{\partial^2 lnL}{\partial \sigma_s^2}\right) \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad . \qquad (16)$$

where $a_{11} = \frac{1}{\beta^2}[(m+n)[1+\Gamma_2''] + 2(n\ln\sigma_s + m\ln\sigma_n)\Gamma_2' + n(\ln\sigma_s)^2 + m(\ln\sigma_n)^2]$

$a_{22} = \frac{m}{{\sigma_n}^2}, a_{33} = \frac{n}{{\sigma_s}^2}, a_{23} = a_{32} = 0,$

$a_{12} = a_{21} = -\frac{m}{\beta\sigma_n}(\Gamma_2' + \ln\sigma_n),$

$a_{13} = a_{31} = -\frac{n}{\beta\sigma_s}(\Gamma_2' + \ln\sigma_s).$

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

$$I^{-1}(\theta) = \frac{1}{a_{11}a_{22}a_{33} - {a_{12}}^2 a_{33} - a_{22}{a_{13}}^2} \times \begin{bmatrix} a_{22}a_{33} & -a_{21}a_{33} & -a_{22}a_{31} \\ -a_{12}a_{33} & a_{11}a_{33} - {a_{13}}^2 & a_{12}a_{31} \\ -a_{22}a_{13} & a_{21}a_{13} & a_{11}a_{22} - {a_{12}}^2 \end{bmatrix}. \quad (17)$$

$$= \begin{bmatrix} V(\hat\beta) & Cov(\hat\beta,\hat\sigma_n) & Cov(\hat\beta,\hat\sigma_s) \\ Cov(\hat\sigma_n,\hat\beta) & V(\hat\sigma_n) & Cov(\hat\sigma_n,\hat\sigma_s) \\ Cov(\hat\sigma_s,\hat\beta) & Cov(\hat\sigma_s,\hat\sigma_n) & V(\hat\sigma_s) \end{bmatrix} \quad (18)$$

where

$$V(\hat\sigma_n) = \frac{{\sigma_n}^2}{mn(m+n)(1-\Gamma_2''-(\Gamma_2')^2)}[n(m+n)[1+\Gamma_2''] + 2mn\log(\sigma_n)\Gamma_2' + mn(\log\sigma_n)^2 - n^2(\Gamma_2')^2],$$

$$V(\hat\sigma_s) = \frac{{\sigma_s}^2}{mn(m+n)(1-\Gamma_2''-(\Gamma_2')^2)}[n(m+n)[1+\Gamma_2''] + 2mn\log(\sigma_s)\Gamma_2' + mn(\log\sigma_s)^2 - m^2(\Gamma_2')^2],$$

$$V(\hat\beta) = \frac{\beta^2}{(m+n)(1-\Gamma_2''-(\Gamma_2')^2)},$$

$$Cov(\hat\beta,\hat\sigma_n) = \frac{\beta\sigma_n(\Gamma_2' + \ln\sigma_n)}{(m+n)(1-\Gamma_2''-(\Gamma_2')^2)},$$

$$Cov(\hat\beta,\hat\sigma_s) = \frac{\beta\sigma_s(\Gamma_2' + \ln\sigma_s)}{(m+n)(1-\Gamma_2''-(\Gamma_2')^2)},$$

$$Cov(\hat\sigma_n,\hat\sigma_s) = \frac{\sigma_n\sigma_s(\Gamma_2' + \ln\sigma_n)(\Gamma_2' + \ln\sigma_s)}{(m+n)(1-\Gamma_2''-(\Gamma_2')^2)},$$

Because the area under the ROC curve is a function of parameters $\theta = (\beta, \sigma_n, \sigma_s)'$, the Delta method will be adopted for finding the approximate variance $V(\widehat{AUC})$, Can be defined as:

$$V(AUC) = \left(\frac{\partial AUC}{\partial\sigma_s}\right)^2 V(\hat\sigma_s) + \left(\frac{\partial AUC}{\partial\sigma_n}\right)^2 V(\hat\sigma_n) + 2\left(\frac{\partial AUC}{\partial\sigma_n}\right)\left(\frac{\partial AUC}{\partial\sigma_s}\right)Cov(\hat\sigma_n,\hat\sigma_s) \quad . \quad (19)$$

$$\tau = V(AUC) = \frac{{\sigma_n}^2{\sigma_s}^2}{(\sigma_n+\sigma_s)^4}\left[\frac{m+n}{mn} + \frac{\left[\ln\left(\frac{\sigma_n}{\sigma_s}\right)\right]^2}{(m+n)(1-\Gamma_2''-(\Gamma_2')^2)}\right]. \quad (20)$$

Where $V(\hat\sigma_s)$, $V(\hat\sigma_n)$, and $Cov(\hat\sigma_n,\hat\sigma_s)$ are taken from the matrix $I^{-1}(\theta)$. The estimate of variance is obtained by substituting the estimates of the parameters σ_n, σ_s. Hence, the estimate of accuracy follows that

$$\frac{\sqrt{N}(\widehat{AUC} - AUC)}{\sqrt{\tau}} \rightarrow N(0,1) \quad . \quad (21)$$

where τ is obtained in equation (19) and it is proven that

$$\widehat{AUC} \sim N(0,\tau), \qquad \Gamma_n' = -(n-1)!\left[\frac{1}{n} + \gamma - \sum_{k=1}^{n}\frac{1}{k}\right],$$

where γ is Euler-Mascheroni constant approximately equal to 0.5772. We note that $\widehat{AUC}$ is an Unbiased Estimator of AUC.

2.3 Confidence Interval for AUC

a) Asymptotic Confidence Interval

The asymptotic 100(1−α)% confidence interval for accuracy is given by

$$\left[\widehat{AUC} - Z_{\alpha/2}SE(\widehat{AUC}), \widehat{AUC} + Z_{\alpha/2}SE(\widehat{AUC})\right] \quad . \quad (22)$$

where $SE(\widehat{AUC})$ can be obtained from equation (20), α is the level of significance and $Z_{\alpha/2}$ is the critical value. For example, $Z_{\alpha/2}$ for a 5% level of significance is 1.96.

b) Bootstrap Confidence Interval

The parametric bootstrap is a resampling technique which can be used to find the variance of any estimator. The idea of bootstrap is to create or resample an artificial dataset from an empirical distribution with same sample size and structure as the original for large number of times. Once the dataset is created, the parameters of interest are to be estimated for each data set. The bootstrap variance of parameter is nothing but the variance of all estimated parameters. Parametric bootstrap is very similar to the non-parametric bootstrap method. In non-parametric bootstrap the sample is simulated from empirical distribution but in parametric bootstrap it is simulated from specified parametric distribution.
The following are the steps involved in finding the parametric bootstrap estimate:

Step 1: Let $X_1, X_2, \ldots, X_m$ be a random sample of size m from $W(\beta,\sigma_n)$ and $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from $W(\beta,\sigma_s)$. By using equations (11), (12) and (13), the ML estimates of the parameters β, σ_n and σ_s are estimated.

Step 2: By using the estimated parameters $\hat\beta, \hat\sigma_n, \hat\sigma_s$, the random observations X_b of size m and Y_b of size n (Bootstrap samples) are generated. From X_b and Y_b, the bootstrap estimates viz. $\hat\beta_b$, $\hat\sigma_{bn}$ and $\hat\sigma_{bs}$ are obtained. Using these bootstrap estimates the accuracy ($\widehat{AUC}_b$) is obtained.

Step 3: Step 2 is repeated 500 times. The mean of all 500 estimates of $\hat\beta$'s, $\hat\sigma_n'$s and $\hat\sigma_s$'s are called the bootstrap estimates of parameters β, σ_n and σ_s respectively and mean of all $\widehat{AUC}_b$'s is called the estimated bootstrap accuracy. The standard deviation of all estimates $\widehat{AUC}_b$ is called the standard error of $\widehat{AUC}_b$.

Step 4: The 100(1−α)% confidence interval for AUC_b is obtained as follows:

$$\left[\widehat{AUC}_b - Z_{\alpha/2}SE(\widehat{AUC}_b), \widehat{AUC}_b + Z_{\alpha/2}SE(\widehat{AUC}_b)\right] \quad .$$

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

2.4 Confidence Intervals for ROC Curve

The $100(1-\alpha)\%$ confidence intervals for the ROC curve are estimated using delta method. This confidence interval for the ROC Curve represents the range at each point of False Positive Rate (FPR) and its corresponding True Positive Rate (TPR). Therefore, the $100(1-\alpha)\%$ confidence intervals for FPR and TPR are as follows:

$$\left[\widehat{\mathrm{FPR}} - Z_{\alpha/2}\mathrm{SE}(\widehat{\mathrm{FPR}}), \widehat{\mathrm{FPR}} + Z_{\alpha/2}\mathrm{SE}(\widehat{\mathrm{FPR}})\right] .$$

$$\left[\widehat{\mathrm{TPR}} - Z_{\alpha/2}\mathrm{SE}(\widehat{\mathrm{TPR}}), \widehat{\mathrm{TPR}} + Z_{\alpha/2}\mathrm{SE}(\widehat{\mathrm{TPR}})\right] .$$

where $\widehat{\mathrm{FPR}}$ and $\widehat{\mathrm{TPR}}$ are the estimated FPR and TPR, respectively, and their variances are

$$V(\widehat{\mathrm{FPR}}) = \left(\frac{\partial \mathrm{FPR}}{\partial \sigma_n}\right)^2 V(\hat{\sigma}_n) + \left(\frac{\partial \mathrm{FPR}}{\partial \beta}\right)^2 V(\hat{\beta}) .$$

$$V(\widehat{\mathrm{TPR}}) = \left(\frac{\partial \mathrm{TPR}}{\partial \sigma_s}\right)^2 V(\hat{\sigma}_s) + \left(\frac{\partial \mathrm{TPR}}{\partial \beta}\right)^2 V(\hat{\beta}) .$$

where $\left(\frac{\partial \mathrm{FPR}}{\partial \sigma_n}\right)^2 = \left(\frac{t^\beta}{\sigma_n{}^2} e^{-\left[\frac{t^\beta}{\sigma_n}\right]}\right)^2$, $\left(\frac{\partial \mathrm{TPR}}{\partial \sigma_s}\right)^2 = \left(\frac{t^\beta}{\sigma_s{}^2} e^{-\left[\frac{t^\beta}{\sigma_s}\right]}\right)^2$,

$$\left(\frac{\partial \mathrm{TPR}}{\partial \beta}\right)^2 = \left(-\frac{t^\beta \log t}{\sigma_s} e^{-\left[\frac{t^\beta}{\sigma_s}\right]}\right)^2,$$

$$\left(\frac{\partial \mathrm{FPR}}{\partial \beta}\right)^2 = \left(-\frac{t^\beta \log t}{\sigma_n} e^{-\left[\frac{t^\beta}{\sigma_n}\right]}\right)^2,$$

$$V(\hat{\sigma}_n) = \frac{\sigma_n{}^2}{mn(m+n)(1-\Gamma_2'' - (\Gamma_2')^2)}[n(m+n)[1+\Gamma_2''] + 2mn\log(\sigma_n)\Gamma_2' + mn(\log\sigma_n)^2 - n^2(\Gamma_2')^2],$$

$$V(\hat{\sigma}_s) = \frac{\sigma_s{}^2}{mn(m+n)(1-\Gamma_2'' - (\Gamma_2')^2)}[n(m+n)[1+\Gamma_2''] + 2mn\log(\sigma_s)\Gamma_2' + mn(\log\sigma_s)^2 - m^2(\Gamma_2')^2],$$

$$V(\hat{\beta}) = \frac{\beta^2}{(m+n)(1-\Gamma_2'' - (\Gamma_2')^2)} .$$

Further, the confidence intervals for FPR and TPR can be obtained using the following expressions.

$$\left[\widehat{\mathrm{FPR}} - Z_{\alpha/2}\sqrt{\left(\frac{t^\beta}{\sigma_n{}^2} e^{-\left[\frac{t^\beta}{\sigma_n}\right]}\right)^2 V(\hat{\sigma}_n) + \left(-\frac{t^\beta \log t}{\sigma_n} e^{-\left[\frac{t^\beta}{\sigma_n}\right]}\right)^2 V(\hat{\beta})}\ , \right.$$
$$\left. \widehat{\mathrm{FPR}} + Z_{\alpha/2}\sqrt{\left(\frac{t^\beta}{\sigma_n{}^2} e^{-\left[\frac{t^\beta}{\sigma_n}\right]}\right)^2 V(\hat{\sigma}_n) + \left(-\frac{t^\beta \log t}{\sigma_n} e^{-\left[\frac{t^\beta}{\sigma_n}\right]}\right)^2 V(\hat{\beta})}\right] .$$

$$\left[\widehat{\mathrm{TPR}} - Z_{\alpha/2}\sqrt{\left(\frac{t^\beta}{\sigma_s{}^2} e^{-\left[\frac{t^\beta}{\sigma_s}\right]}\right)^2 V(\hat{\sigma}_s) + \left(-\frac{t^\beta \log t}{\sigma_s} e^{-\left[\frac{t^\beta}{\sigma_s}\right]}\right)^2 V(\hat{\beta})}\ , \right.$$
$$\left. \widehat{\mathrm{TPR}} + Z_{\alpha/2}\sqrt{\left(\frac{t^\beta}{\sigma_s{}^2} e^{-\left[\frac{t^\beta}{\sigma_s}\right]}\right)^2 V(\hat{\sigma}_s) + \left(-\frac{t^\beta \log t}{\sigma_s} e^{-\left[\frac{t^\beta}{\sigma_s}\right]}\right)^2 V(\hat{\beta})}\right] .$$

These confidence interval lines show the variability of the proposed ROC curve at each and every point on the ROC curve. In the next section, the results are carried out using simulation study to explain the proposed methodology. Further, the confidence intervals are evaluated for the summary measure AUC and the intrinsic measures FPR and TPR.

2.5 Confidence Intervals for Brier Score

The Confidence Intervals for Brier Score are as follows:

$$[\hat{B} \pm Z_{\alpha/2}\ \mathrm{SE}(\hat{B})]$$

Where $\hat{B}$ is the estimated B, and their variance is

$$V(\hat{B}) = E(\hat{B})^2 - [E(\hat{B})]^2.$$

Where

$$E(\hat{B})^2 = (1-\alpha)\left[0.333 + \frac{\hat{\sigma}_s}{\hat{\sigma}_n}\right]$$

$$E(\hat{B}) = (1-\alpha)\left[0.5 + \frac{\hat{\sigma}_s}{\hat{\sigma}_n}\right]$$

3. Simulation Study

The accuracy, confidence intervals for AUC, ROC curve and Brier Score has been computed through different techniques via asymptotic MLE method, parametric bootstrap.

3.1 Asymptotic MLE Method

Simulation studies are conducted with different combinations of scale and shape parameters of both normal and abnormal populations. At every parameter combination and sample size, the AUC and its confidence intervals are obtained. The main purpose of conducting simulations is to show how the AUC of ROC curve possesses different values as the scale and shape parameters of the normal and abnormal distributions change. The variations in the parameter values of both populations are used to explain the overlapping area in terms of AUC; this mean that the higher the AUC, the lesser the overlapping area and vice versa. Further, to demonstrate the behavior of AUC, the entire simulation work is carried out with three different experiments. In the first experiment, the results are reported in Table 3.1. Numerical experiments were carried out to inspect how the MLE‟s of AUC and their asymptotic results work for simulated data sets. Four different samples of size (m, n) = (30, 30) with different parametric values were considered as mentioned in columns 1, 2 and 3 Table 3.1. The corresponding accuracy, 95% confidence interval are shown in 4, 5 columns of Table 3.1.

Table 3.1: Accuracy, standard error and Confidence interval of AUC based on Constant Shape Bi-Weibull ROC model through Asymptotic MLE method

$\hat{\beta}$	$\hat{\sigma}_n$	$\hat{\sigma}_s$	$\widehat{AUC}$	95% Confidence Interval
3.0	0.6333	2.8000	0.8155	[0.5391,1.0919]
2.5	1.8418	6.2552	0.7723	[0.5365,1.0084]
1.52	1.3849	3.8046	0.7331	[0.5271,0.9391]
1.0	1.1256	2.5665	0.7125	[0.5145,0.9154]

From table 3.1, it is observed that, the accuracy increases, simultaneously, the confidence interval increases. Because the asymptotic distribution is independent of β, β may be kept constant or it may vary.

The sample β is estimated using iterative procedure from equation (13) and using β, the other two parameters using were found using equations (11) and (12). Hence, the ML estimate of AUC is obtained. The 95% asymptotic confidence interval is also calculated.

If β takes higher values as 3 and 2.5, the AUC is observed to have a better value indicating high level of accuracy, thus, reflecting the scenario that as the discrepancy between shape parameters of both normal and abnormal population"s increases, AUC attains a larger value indicating a better extent of correct classification with minimum percentage of overlapping area.

In second experiment, the scale parameter of normal population is varied by fixing the other parameters as constant and, Table 3.2 shows simulated independent samples of m controls and n cases (m= n = 5, 10, 40, 50, 80, 100) to assess the behavior of asymptotic MLE"s and confidence interval over different sample sizes by fixing $\hat{\sigma}_n$=5 and for different values of $\hat{\sigma}_s$ viz. 8, 12, 20, 100.

Table 3.2: Accuracy and Confidence interval of AUC when $\hat{\sigma}_n$=5 for different sample size

Sample size	(m, n) (5, 5)	(m, n) (10,10)	(m, n) (40,40)	(m, n) (50,50)	(m, n) (80,80)	(m, n) 100,100
$\hat{\sigma}_s$=8	0.6154 0.2570 0.9738	0.6154 0.3620 0.8688	0.6154 0.4887 0.7421	0.6154 0.5021 0.7287	0.6154 0.5258 0.7050	0.6154 0.5352 0.6955
$\hat{\sigma}_s$=12	0.7059 0.2442 1.1676	0.7059 0.3794 1.0324	0.7059 0.5426 0.8691	0.7059 0.5598 0.8519	0.7059 0.5905 0.8213	0.7059 0.6257 0.7860
$\hat{\sigma}_s$=20	0.8000 0.1614 1.4386	0.8000 0.3484 1.2516	0.8000 0.5742 1.0258	0.8000 0.5981 1.0019	0.8000 0.6404 0.9596	0.8000 0.6572 0.9428
$\hat{\sigma}_s$=30	0.8571 0.0580 1.6562	0.8571 0.2921 1.4222	0.8571 0.5746 1.1396	0.8571 0.6044 1.1098	0.8571 0.6574 1.0569	0.8571 0.6785 1.0358

From Table 3.2, it is observed that the scale parameter of normal population ($\hat{\sigma}_s$) is varied by keeping all the other parameters as constant. Moderate levels of discrepancy in the shape values and scale parameters influence the accuracy of the classification. As $\hat{\sigma}_s$ attains a larger value, the AUC of ROC curve tends to have better values of accuracy.

So this reveals that along with discrepancy in shape parameters of both populations, scale parameter also tends to explain better variability in the data giving rise to talk about the exact performance of the test considered. In Tables 3.2, first row represents the $\widehat{AUC}$, second and third rows gives the confidence limits. It is observed that, as the sample size increases the confidence interval is increases.

To demonstrate the proposed methodology with the help of graphical visualization, Figures 3.1(a) and 3.1 (b) are drawn for the experiment 2. That is the scale parameter of normal population ($\hat{\sigma}_s$) is varied by keeping all the other parameters as constant.

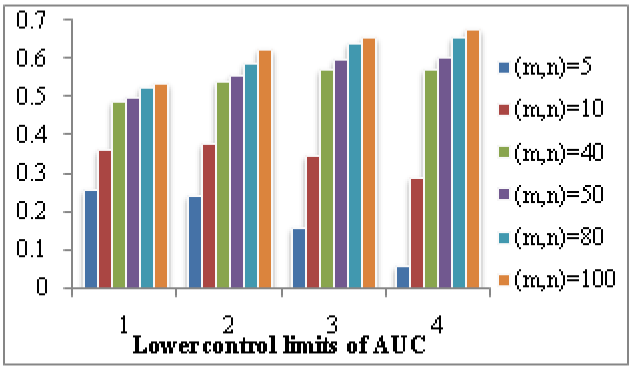

Figure 3.1(a): Effect on Lower Control limits for AUC values by varied different scale parameter in normal populations

From Figure 3.1(a), it is visualized that the experiment 2 and it is observed that, as the sample size increases the Lower Control limits are increases.

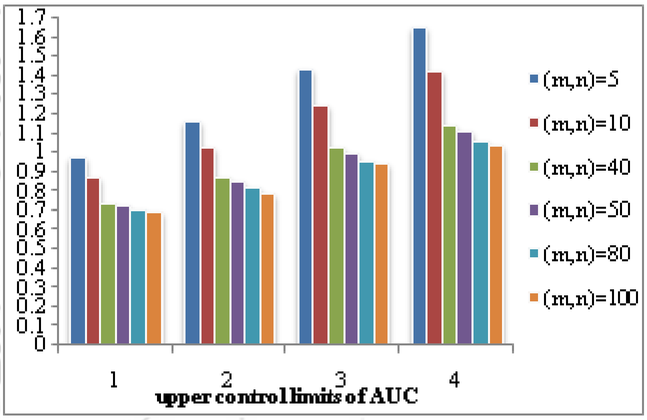

Figure 3.1(b): Effect on Upper Control limits for AUC values by varied different scale parameter in normal populations

From Figure 3.1(b), it is visualized that the experiment 2 and it is observed that, Upper Control limits increases over different sample sizes by fixing $\hat{\sigma}_n$=5 and for different values of $\hat{\sigma}_s$ viz. 8, 12, 20, 100. In the third experiment, the scale parameter of abnormal population is varied by fixing the other parameters as constant and, Table 3.3 shows simulated independent samples of m controls and n cases (m= n = 5, 10, 40, 50, 80, 100) to inspect the behavior of asymptotic MLE and confidence interval over different sample sizes by fixing $\hat{\sigma}_s$= 45 and for different values of $\hat{\sigma}_n$ viz. 3, 8, 10, 20.

Table 3.3: Accuracy and Confidence interval of AUC when $\hat{\sigma}_s$=45 for different sample size

Sample size	(m, n) (5, 5)	(m, n) (10,10)	(m, n) (40,40)	(m, n) (50,50)	(m, n) (80,80)	(m, n) (100,100)
$\hat{\sigma}_n$=5	0.9000 0.0685 1.8685	0.9000 0.2151 1.5849	0.9000 0.5576 1.2424	0.9000 0.6125 1.0565	0.9000 0.6579 1.1421	0.9000 0.6834 1.1166
$\hat{\sigma}_n$=8	0.8491 0.0763 1.6019	0.8491 0.3026 1.3955	0.8491 0.5758 1.222	0.8491 0.6047 1.0934	0.8491 0.6559 1.0423	0.8491 0.6763 1.0219
$\hat{\sigma}_n$=10	0.8182 0.1343 1.5021	0.8182 0.3346 1.3018	0.8182 0.5764 1.0599	0.8182 0.6019 1.0344	0.8182 0.6472 0.9892	0.8182 0.6653 0.9711
$\hat{\sigma}_n$=20	0.6923 0.2498 1.1348	0.6923 0.3794 1.0052	0.6923 0.5359 0.8488	0.6923 0.5524 0.8322	0.6923 0.5817 0.8029	0.6923 0.5934 0.7913

From Table 3.3, it is observed that the scale parameter of abnormal population ($\hat{\sigma}_n$) is varied by keeping all the other parameters as constant. Moderate levels of discrepancy in the shape values and scale parameters influence the accuracy of the classification. As $\hat{\sigma}_n$ attains a larger value, the AUC of ROC curve tends to have better values of accuracy.

So this reveals that along with discrepancy in shape parameters of both populations, scale parameter also tends to explain better variability in the data giving rise to talk about the exact performance of the test considered. In Tables 3.3, first row represents the $\widehat{AUC}$, second and third rows gives the confidence limits. It is observed that, as the sample size increases the confidence interval is increases.

To demonstrate the proposed methodology with the help of graphical visualization, Figures 3.1(c) and 3.1 (c) are drawn for the experiment 3. That is the scale parameter of abnormal population ($\hat{\sigma}_n$) is varied by keeping all the other parameters as constant.

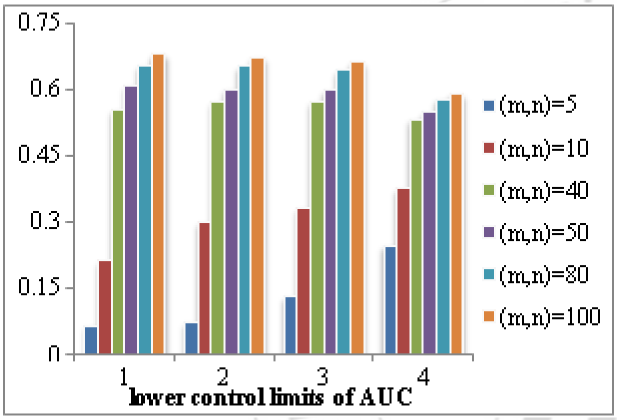

Figure 3.1(c): Effect on lower control limits for AUC values by varied different scale parameter in abnormal populations

From Figure 3.1(c), it is visualized that the experiment 3 and it is observed that, as the sample size increases the Lower control limits are increases.

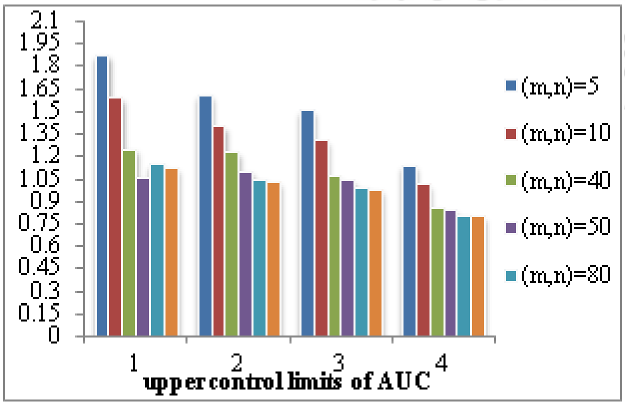

Figure 3.1 (d): Effect on Upper Control limits for AUC values by varied different scale parameter in abnormal populations

From Figure 3.1(d), it is visualized that the experiment 3 and it is observed that, the Upper Control limits increases over different sample sizes by fixing $\hat{\sigma}_s$=5 and for different values of $\hat{\sigma}_n$ viz. 5, 8, 10, 20.

3.2 Estimation of Bootstrap Variance

For parametric bootstrapping, the data was generated from a uniform distribution using (m, n) as specified in Table 3.4. Then by inverse transformation method, it is converted into Weibull variate with the values of β, σ_n and σ_s. Using Step 1 results in the estimates as $\hat{\beta}$, $\hat{\sigma}_n$ and $\hat{\sigma}_s$. By using Step 2, the estimate of bootstrap sample is obtained as $\hat{\beta}_b$, $\hat{\sigma}_{bn}$ and $\hat{\sigma}_{bs}$. From these 500 estimates of parameters, one can find an estimate of AUC by using the equation (8). By averaging these 500 numbers of estimates of AUC, one can estimate the bootstrap estimate AUC. Standard error of $\widehat{AUC}_b$ is nothing but the standard deviation of the **b** number $\widehat{AUC}_b$'s. By Step 4, the 95% confidence interval for bootstrap AUC is obtained as usual. Table 3.4 shows the bootstrap AUC, SE and confidence interval for $\widehat{AUC}_b$.

Table 3.4: Accuracy, SE, and Confidence interval of AUC through Bootstrap Simulation

Sample size	$\hat{\beta}$	$\hat{\sigma}_n$	$\hat{\sigma}_s$	$\widehat{AUC}$	95% Confidence Interval
(10,10)	2.6709	7.5414	201.8100	0.9249	[0.8328,1.0000]
(40,40)	2.5274	6.3739	103.9060	0.9240	[0.8581,0.9899]
(50,50)	2.4662	5.8770	84.2817	0.9220	[0.8695,0.9745]
(80,80)	2.4340	5.6493	74.3820	0.9222	[0.8581,0.9636]
(100,100)	2.3948	5.4283	66.5260	0.9211	[0.8926,0.9496]

Comparing asymptotic and bootstrap variance, both perform at the same level. The asymptotic variance does not perform well for small samples such as (5, 5) and (10, 10) where the bound for accuracy has reached below 0.5 which is not regarded as a good estimate. Hence, the asymptotic variance holds for large samples only.

4. Illustration

The real data set is about the ICU scoring system was extracted from [18]; SAPS III is a system for predicting mortality (dead or alive) status of a patient in ICU. SAPS III has been designed to provide a real-life predicted mortality for a patient by following a well defined procedure, based on a mathematical model that needs calibration. This data consists of a total of 200 respondents of which 40 are alive and 160 are dead. From this data set it is observed that the SAPS III scores for dead patients follow abnormal population where as the scores for patients who are alive follow normal population.

Table 4.1: Results for SAPS III using GHROC curve methodology

$\hat{\beta}$	$\hat{\sigma}_n$	$\hat{\sigma}_s$	$\widehat{AUC}$	95% Confidence Interval
2.5234	31635.43	42155.17	0.5712	[0.4819, 0.6607]

The results for the prognosis of disease are reported in Table 4.1. It is observed that the accuracy of the test is 57.12% indicating that the SAPS III score is able to identify the status of mortality about 57.12% and the variance is 0.002080202. Further, the confidence interval of AUC is [0.4819, 0.6607] and the proposed ROC curve for SAPS III uniformly lies above the chance line to explain the mortality rate.

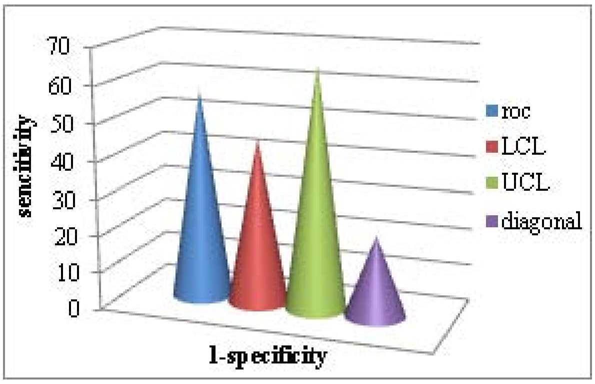

Figure 4.1: ROC curve for SAPS III with its confidence intervals.

5. Conclusion

The present work is carried out to establish a ROC model developed from two parameter Weibull distributions for evaluating the accuracy of biomarkers in predicting disease status. In advance it did not yield a closed form expression for Area Under the ROC curve. For this reason, equal shape parameter and different scale parameter were considered. Moreover, the original accuracy of the diagnosis is not affected by taking equal shape parameter. Hence, estimation of Area Under the Constant Shape Bi-Weibull ROC curve is a main objective for this study. The Maximum Likelihood technique is adopted for estimating the parameters. The technique yielded an asymptotically unbiased estimate of the accuracy. The asymptotic distribution of $\hat{AUC}$ and 95% confidence interval were found. The behavior of asymptotic confidence interval is studied through simulation. The bootstrap AUC is higher than the AUC obtained by other methods.

References

[1] C.B. Begg, "Advances in statistical methodology for diagnostic medicine in the 1980‟s," Statistics in Medicine, 10(12), 1887–1895, 1991.
[2] M. Betinec, "Testing the difference of the ROC curves in biexponential model," Tatra Mountains Mathematical Publications, 39, 215-223, 2008.
[3] G. Campbell, M.V. Ratnaparkhi, "An application of Lomax distributions in receiver operating characteristic (ROC) curve analysis," Communication in Statistics, 22(6), 1681-1687, 1993.
[4] M. DeGroot, S Fineberg, "The comparison and evaluation of forecasters," The statistician, 32, 12-22. 1983.
[5] D.D. Dorfman et al., "Proper receiver operating characteristics analysis: The bi-Gamma model," Academic Radiology, 4, 138-149, 1996.
[6] D. Foster, R.V. Vohra, "Asymptotic calibration," Biometrika, 379-390, 1998.
[7] J.A. Hanley, "Receiver operating characteristic (ROC) methodology: the state of the art," Critical Reviews in Diagnostic Imaging, 29(3), 307–335, 1989.
[8] E. Hussain, "The ROC curve model from generalized-exponential distribution," Pakistan Journal of Statistics and Operations Research, 7(2), 323-330, 2011.
[9] D. Kundu, R.D. Gupta, "Estimation of P[Y<X] for Weibull Distribution," IEEE Transactions on Reliability, 55(2), 270-280, 2006.
[10] T. Leo Alexander, A. Lavanya, "Functional Relationship between Brier Score and Area Under the Constant Shape Bi-Weibull Roc Curve," International Journal of Recent Scientific Research, 7(7), 12330-12336, 2016.
[11] Mitsuru Ikeda, Takeo Ishigaki, Kazunobu Yamauchi, "Relationship between brier score and area under the binormal roc curve," Computer methods and programs in Biomedicine , 67, 187-194, 2002.
[12] J.C. Oglive, C.D. Creelman, "Maximum-likelihood estimation of receiver operating characteristic curve parameters," Journal of Mathematical Psychology, 5, 377-391, 1968.
[13] S. Pundir, R. Amala, "A study on the Bi-Rayleigh ROC model," Bonfring International Journal of Data Mining, 2(2), 42-47, 1968.
[14] S. Pundir, R. Amala, "A study on the comparison of bi-Rayleigh ROC model with bi-Gamma ROC model," In Application of Reliability Theory and Survival Analysis, Coimbatore, India: Bonfring Publications, 196-209, 2012.
[15] S. Pundir, R. Amala, "Parametric receiver operating characteristic modeling for continuous data: A glance," Model Assisted Statistics and Application, 9(2), 121-135, 2013, doi: 10.3233/MASA -130284.
[16] R. Vishnu Vardhan, A. Lavanya, and T. Ganesh, "Functional relationship between Brier Score and Area Under the ROC Curve in the Context of Skewed Distributions," IJSIMR ,3, 935-939, 2015.
[17] X.H. Zhou, N.A. Obuchowski and D.K. McClish, Statistical Methods in Diagnostic Medicine, New York, NY: John Wiley & Sons, Inc, 2002.
[18] http://www.umass.edu/statdata/statdata/data/icu.dat

Author Profile

Dr. T. Leo Alexander currently serving as an Associate Professor in the Department of Statistics, Loyola College, Chennai, India. For the past three decades. Teaching both Under Graduate, Post Graduate courses and guiding research students for their M.Phil and Ph.D. Also served as Controller of Examinations as well as Registrar i/c at University of Madras.

Allada Lavanya- Pursuing PhD in the field of Statistical Inference and particularly in the area of Survival Analysis and its applications. Currently doing research under the supervision of Dr. T. Leo Alexander, Associate Professor, Department of Statistics, Loyola College, Chennai, India