DIGITALCOMMONS
— @WAYNESTATE —

Journal of Modern Applied Statistical Methods

Volume 13 | Issue 1 Article 20

5-1-2014

Evaluation of Area under the Constant Shape Bi-Weibull ROC Curve

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

R Amala
Pondicherry University, Puducherry, India, amalar.statistics@gmail.com

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation

Pundir, Sudesh and Amala, R (2014) "Evaluation of Area under the Constant Shape Bi-Weibull ROC Curve," *Journal of Modern Applied Statistical Methods*: Vol. 13 : Iss. 1 , Article 20.
DOI: 10.22237/jmasm/1398917940
Available at: http://digitalcommons.wayne.edu/jmasm/vol13/iss1/20

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Journal of Modern Applied Statistical Methods
May 2014, Vol. 13, No. 1, 305-328.

Copyright © 2014 JMASM, Inc.
ISSN 1538 − 9472

Evaluation of Area under the Constant Shape Bi-Weibull ROC Curve

Sudesh Pundir
Pondicherry University
Puducherry, India

R. Amala
Pondicherry University
Puducherry, India

The Receiver Operating Characteristic (ROC) curve generated based on assuming a constant shape Bi-Weibull distribution is studied. In the context of ROC curve analysis, it is assumed that biomarker values from controls and cases follow some specific distribution and the accuracy is evaluated by using the ROC model developed from that specified 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, area under the ROC curve (AUC), asymptotic and bootstrap confidence intervals for the AUC are derived. Theoretical results are validated by simulation studies.

Keywords: Constant shape Bi-Weibull ROC model, area under the ROC curve, asymptotic variance of accuracy, confidence interval, parametric bootstrap variance

Notations and Terminologies

X	Random variable representing controls	t	Cut-off point of classification, $t \in x \cup y$
Y	Random variable representing cases	$I(\theta)$	Fisher Information matrix
m	Number of controls	$y(x)$	ROC model
n	Number of cases	MLE	Maximum Likelihood Estimate
$f(x)$	Probability Density Function (PDF) of X	$x(t)$	False Positive Rate (FPR) at cut-off t
$g(y)$	PDF of Y	$y(t)$	True Positive Rate (TPR) at cut-off t
$F(x)$	Distribution function of X	TPR	Probability that cases are correctly identified (Sensitivity)
$G(y)$	Distribution function of Y	FPR	Probability that controls are wrongly identified as cases (1-Specificity)
AUC	Population Area under the ROC curve	α_0, α_1	Shape parameters of X and Y, respectively
$A\widehat{U}C$	Observed Area Under the ROC curve	β_0, β_1	Scale parameters of X and Y, respectively

Dr. Sudesh Pundir is an Assistant Professor in the Department of Statistics. Email her at sudeshpundir19@gmail.com. R. Amala is a Doctoral Student in Applied Statistics. Email her at amalar.statistics@gmail.com.

Introduction

A Receiver Operating Characteristic (ROC) curve provides quick access to the quality of classification in many medical diagnoses. In ROC curve analysis, the accuracy has been analyzed in terms of a model relating the parameters of cases and controls called as the ROC model. ROC model can be defined as the TPR obtained as a function of FPR which takes the form

$$y(x) = 1 - G\left(F^{-1}\left(1 - x(t)\right)\right); 0 \le x(t) \le 1 \tag{1}$$

where $x(t)$ and $y(t)$ are defined as follows:

$$\left.\begin{aligned} x(t) = P(X > t) = \int_t^\infty f(x)dx = 1 - \int_0^t f(x)dx = 1 - F(t) \\ y(t) = P(Y > t) = \int_t^\infty g(y)dy = 1 - \int_0^t g(y)dy = 1 - G(t) \end{aligned}\right\} \tag{2}$$

Graphically, a ROC curve is a graph of TPR versus FPR for all possible threshold values. 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, area under the ROC curve is most commonly adopted index. In this article, the inference about the area under the ROC curve 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 (Zhou, Obuchowski & McClish, 2002), Bi-Logistic ROC model (Oglive & Creelman, 1968), Bi-Lomax ROC model (Campbell & Ratnaparkhi, 1993), Bi-Gamma ROC model (Dorfman et al., 1996), Bi-Exponential ROC model (Betinec, 2008), Generalized Bi-Exponential ROC model (Hussain, 2011), Bi-Rayleigh ROC model and its comparison with Bi-Normal model (Pundir & Amala, 2012), comparison of Bi-Rayleigh ROC model with Bi-Normal and Bi-Gamma ROC models (Pundir & Amala, 2012) and a review of all parametric ROC models in case of continuous data (Pundir & Amala, 2014), Normal-Exponential (Pundir & Amala, 2014).

A constant shape Bi-Weibull ROC model is proposed for non-normal data. 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. Let $X \sim W(\alpha_0, \beta_0)$ and $Y \sim W(\alpha_1, \beta_1)$, then the ROC model developed from two parameter Bi-Weibull distribution is given by

$$y(x) = Exp\left\{-\frac{\left(-\beta_0 \ln(x(t))\right)^{\frac{\alpha_1}{\alpha_0}}}{\beta_1}\right\}, \; \alpha_1, \alpha_0 > 0, \; \beta_1, \beta_0 > 0 \tag{3}$$

One major disadvantage of assuming two parameter Weibull distribution to the biomarker is that the accuracy cannot be expressed in closed form. By substituting the MLE's $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_0$ and $\hat{\beta}_1$, 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 are needed. Here, the standard error of accuracy is studied by different methods viz. Monte Carlo, asymptotic MLE, parametric bootstrap and non-parametric 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.

Constant Shape Bi-Weibull ROC Model

The constant shape Bi-Weibull ROC model assumes that the biomarker values from controls and cases follow two parameter Weibull distribution with same shape parameter and different scale parameters.

The PDF of controls and cases take the form

$$f(x) = \frac{\alpha}{\beta_0} x^{\alpha-1} \exp\left\{-\frac{x}{\beta_0}^{\alpha}\right\}, x > 0, \alpha, \beta_0 > 0 \tag{4}$$

and

$$g(y) = \frac{\alpha}{\beta_1} y^{\alpha-1} \exp\left\{-\frac{y}{\beta_1}^{\alpha}\right\}, y > 0, \alpha, \beta_1 > 0 \tag{5}$$

respectively.

The probabilities, Sensitivity and 1−Specificity for constant shape Bi-Weibull distribution can be given as follows:

$$Sensitivity = \int_t^{\infty} \frac{\alpha}{\beta_1} y^{\alpha-1} \exp\left\{-\frac{y}{\beta_1}^{\alpha}\right\} dy = \exp\left\{\frac{-t^{\alpha}}{\beta_1}\right\} \tag{6}$$

$$1 - Specificity = \int_t^{\infty} \frac{\alpha}{\beta_0} x^{\alpha-1} \exp\left\{-\frac{x}{\beta_0}^{\alpha}\right\} dx = \exp\left\{\frac{-t^{\alpha}}{\beta_0}\right\} \tag{7}$$

Hence, the ROC model is given by

$$y(x) = x(t)^{\left(\frac{\beta_0}{\beta_1}\right)}; 0 \le x(t) \le 1. \tag{8}$$

The ROC curve can be estimated by substituting the MLE of parameters in equation (8) and plotted by taking $x(t)$ in equation (7) on x-axis and $y(x)$ in equation (8) on y-axis. Also, one can plot the ROC curve by taking 1−Specificity on X axis and Sensitivity on Y axis. The area under the ROC curve is obtained by integrating the joint density function of X and Y and it has the following form.

$$A = P(X < Y) = \int_0^\infty \int_0^y \frac{\alpha^2}{\beta_0 \beta_1} (xy)^{\alpha-1} Exp\left[-\left(\frac{y^\alpha}{\beta_1} + \frac{x^\alpha}{\beta_0} \right) \right] dxdy$$
$$= \frac{\beta_1}{\beta_1 + \beta_0} \tag{9}$$

The MLEs of β_0 and β_1 can be used again to estimate the AUC. And the performance of the estimator $A\widehat{U}C$ can be assessed through variance estimate.

Maximum Likelihood Estimate of Parameters

The MLE of two parameter Weibull distribution has been discussed by (Kundu & Gupta, 2006) in the context of Reliability estimation. Let $X_1, X_2, \ldots.. X_m$ be a random sample of size m from $W(\alpha, \beta_0)$ and $Y_1, Y_2, \ldots.. Y_n$ be a random sample of size n from $W(\alpha, \beta_1)$. The likelihood function of the selected sample is given by

$$L(x_i, y_j / \theta) = \prod_{i=1}^{m} f_X(x_i / \alpha, \beta_0) \prod_{j=1}^{n} f_Y(y_j / \alpha, \beta_1) \tag{10}$$

where $\theta = (\alpha, \beta_0, \beta_1)'$

The log-likelihood function is

$$LnL = (m+n) \ln \alpha + (\alpha - 1) \left[\sum_{j=1}^{n} \ln y_j + \sum_{i=1}^{m} \ln x_i \right] - n \ln \beta_1$$
$$- m \ln \beta_0 - \frac{1}{\beta_1} \sum_{j=1}^{n} y_j^\alpha - \frac{1}{\beta_0} \sum_{i=1}^{m} x_i^\alpha \tag{11}$$

Differentiating (11) with respect to α results in

$$\frac{\partial LnL}{\partial \alpha} = \frac{(m+n)}{\alpha} + \left[\sum_{j=1}^{n} \ln y_j + \sum_{i=1}^{m} \ln x_i \right]$$
$$- \frac{1}{\beta_1} \sum_{j=1}^{n} y_j^\alpha \ln y_j - \frac{1}{\beta_0} \sum_{i=1}^{m} x_i^\alpha \ln x_i \tag{12}$$

By differentiating the log-likelihood function with respect to β_0, β_1 and equating to zero, we get the estimates. The MLE's of β_1 and β_0 are determined as,

$$\hat{\beta}_1(\alpha) = \frac{\sum_{j=1}^{n} y_j^{\alpha}}{n} \quad \text{and} \quad \hat{\beta}_0(\alpha) = \frac{\sum_{i=1}^{m} x_i^{\alpha}}{m} \tag{13}$$

Substituting $\hat{\beta}_0$ and $\hat{\beta}_1$ in equation (12) and equating it to 0, results in a non-linear equation:

$$h(\hat{\alpha}) = \frac{m + n + \sum_{i=1}^{n} y_j^{\alpha} + \sum_{i=1}^{m} x_i^{\alpha}}{\dfrac{m\sum_{i=1}^{m} x_i^{\alpha} \ln x_i}{\sum_{i=1}^{m} x_i^{\alpha}} + \dfrac{n\sum_{i=1}^{n} y_i^{\alpha} \ln y_j}{\sum_{i=1}^{n} y_j^{\alpha}}} \tag{14}$$

Hence, $\hat{\alpha}$ can be determined as a solution of non-linear equation (14). By substituting equation (13) and (14) in equation (9), an estimate of AUC ($\widehat{AUC}$) will result.

Asymptotic Distribution of area under constant shape Bi-Weibull ROC Model

To evaluate the significance of the statistic AUC, its variance and standard error must be computed. The following theorem evaluates the variance of the estimate, $\widehat{AUC}$.

Theorem 1

The area under the constant shape Bi-Weibull ROC curve will converge in distribution to a Normal random variable with mean zero and variance

$$\tau = \frac{\beta_0^2 \beta_1^2}{\left(\beta_0+\beta_1\right)^4}\left[\frac{\left(m+n\right)}{mn}+\frac{\left[ln\left(\frac{\beta_0}{\beta_1}\right)\right]^2}{\left(m+n\right)\left(1+\Gamma_2''-\Gamma_2'^2\right)}\right] \text{ for large } N \text{, where } N=m+n.$$

Proof: Let $L\left(\theta / x,y\right)$; $\theta=\left(\alpha,\beta_0,\beta_1\right)'$ be the likelihood function of the sample observations from *X* and *Y* which is given by

$$\begin{aligned} ln\,L(\theta / x,y) &= (m+n)\,ln\,\alpha - m\,ln\,\beta_0 \\ &\quad -n\,ln\,\beta_1 + (\alpha-1)\left[\sum_{i=1}^{m} ln\,x_i + \sum_{j=1}^{n} ln\,y_j\right] \\ &\quad -\frac{1}{\beta_0}\sum_{i=1}^{m} x_i^{\alpha} - \frac{1}{\beta_1}\sum_{j=1}^{n} y_j^{\alpha} \end{aligned} \tag{15}$$

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

$$\Rightarrow \sqrt{N}(\hat{\theta}-\theta) \rightarrow N(0, I^{-1}(\theta)). \tag{16}$$

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

$$I(\theta) = -\begin{bmatrix} E\left(\frac{\partial^2 \ln L}{\partial \alpha^2}\right) & E\left(\frac{\partial^2 \ln L}{\partial \alpha \partial \beta_0}\right) & E\left(\frac{\partial^2 \ln L}{\partial \alpha \partial \beta_1}\right) \\ E\left(\frac{\partial^2 \ln L}{\partial \beta_0 \partial \alpha}\right) & E\left(\frac{\partial^2 \ln L}{\partial \beta_0^2}\right) & E\left(\frac{\partial^2 \ln L}{\partial \beta_0 \partial \beta_1}\right) \\ E\left(\frac{\partial^2 \ln L}{\partial \beta_1 \partial \alpha}\right) & E\left(\frac{\partial^2 \ln L}{\partial \beta_1 \partial \beta_0}\right) & E\left(\frac{\partial^2 \ln L}{\partial \beta_1^2}\right) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \quad (17)$$

where

$$a_{11} = \frac{1}{\alpha^2}\left[(m+n)\left[1+\Gamma_2''\right] + 2(n\ln\beta_1 + m\ln\beta_0)\Gamma_2' + n(\ln\beta_1)^2 + m(\ln\beta_0)^2\right],$$

$$a_{22} = \frac{m}{\beta_0^2},\ a_{33} = \frac{n}{\beta_1^2},$$

$$a_{23} = \mathrm{a}_{32} = 0,\ a_{12} = a_{21} = \frac{-m}{\alpha\beta_0}\left(\Gamma_2' + \ln\beta_0\right),$$

$$a_{13} = a_{31} = \frac{-n}{\alpha\beta_1}\left(\Gamma_2' + \ln\beta_1\right),$$

$$V\left(\hat{\beta}_0\right) = \frac{\beta_0^2\left[n(m+n)(1+\Gamma_2'') + 2mn\log(\beta_0)\Gamma_2' + mn\left[\log(\beta_0)\right]^2 - n^2(\Gamma_2')^2\right]}{mn(m+n)\left(1-\Gamma_2''-(\Gamma_2')^2\right)},$$

$$V\left(\hat{\beta}_1\right) = \frac{\beta_1^2\left[n(m+n)(1+\Gamma_2'') + 2mn\log(\beta_1)\Gamma_2' + mn\left[\log(\beta_1)\right]^2 - m^2(\Gamma_2')^2\right]}{mn(m+n)\left(1-\Gamma_2''-(\Gamma_2')^2\right)}.$$

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

$$I^{-1}(\theta) = \frac{1}{a_{11}a_{22}a_{33} - a_{12}^2 a_{33} - a_{22}a_{13}^2}\begin{bmatrix} a_{22}a_{33} & -a_{21}a_{33} & -a_{22}a_{31} \\ -a_{12}a_{33} & a_{11}a_{33} - a_{13}^2 & a_{12}a_{31} \\ -a_{22}a_{13} & a_{21}a_{13} & a_{11}a_{22} - a_{12}^2 \end{bmatrix}. \quad (18)$$

$$= \begin{bmatrix} V(\hat{\alpha}) & Cov(\hat{\alpha},\hat{\beta}_0) & Cov(\hat{\alpha},\hat{\beta}_1) \\ Cov(\hat{\beta}_0,\hat{\alpha}) & V(\hat{\beta}_0) & Cov(\hat{\beta}_0,\hat{\beta}_1) \\ Cov(\hat{\beta}_1,\hat{\alpha}) & Cov(\hat{\beta}_1,\hat{\beta}_0) & V(\hat{\beta}_1) \end{bmatrix} \tag{19}$$

where

$$V(\hat{\alpha}) = \frac{\alpha^2}{(m+n)(1-\Gamma_2'' - (\Gamma_2')^2)},\ Cov(\hat{\alpha},\hat{\beta}_0) = \frac{\alpha\beta_0(\Gamma_2' + \ln\beta_0)}{(m+n)[1-\Gamma_2'' - (\Gamma_2')^2]},$$

$$Cov(\hat{\alpha},\hat{\beta}_1) = \frac{\alpha\beta_1(\Gamma_2' + \ln\beta_1)}{(m+n)[1-\Gamma_2'' - (\Gamma_2')^2]},$$

and $C(\hat{\beta}_0,\hat{\beta}_1) = \dfrac{\beta_0\beta_1(\Gamma_2' + \ln\beta_0)(\Gamma_2' + \ln\beta_1)}{(m+n)[1-\Gamma_2'' - (\Gamma_2')^2]}$

Because the area under the ROC curve is a function of parameters $\theta = (\alpha, \beta_0, \beta_1)'$, the Delta method will be adopted for finding the approximate variance. $V(\widehat{AUC})$ can be defined as:

$$\begin{aligned} V(AUC) = {} & \left(\frac{\partial AUC}{\partial\beta_1}\right)^2 V(\hat{\beta}_1) + \left(\frac{\partial AUC}{\partial\beta_0}\right)^2 V(\hat{\beta}_0) \\ & +2\left(\frac{\partial AUC}{\partial\beta_0}\right)\left(\frac{\partial AUC}{\partial\beta_1}\right) Cov(\hat{\beta}_0,\hat{\beta}_1). \end{aligned} \tag{20}$$

$$\tau = V(AUC) = \frac{\beta_0^2\beta_1^2}{(\beta_0+\beta_1)^4}\left[\frac{m+n}{mn} + \frac{\left[ln\left(\frac{\beta_0}{\beta_1}\right)\right]^2}{(m+n)\left(1+\Gamma_2'' - \Gamma_2'^2\right)}\right]. \tag{21}$$

where $V(\hat{\beta}_1)$, $V(\hat{\beta}_0)$ and $Cov(\hat{\beta}_0,\hat{\beta}_1)$ are taken from the matrix $I^{-1}(\theta)$. The estimate of variance is obtained by substituting the estimates of the parameters β_0, β_1. Hence, the estimate of accuracy follows that

$$\frac{\sqrt{N}(A\hat{U}C - AUC)}{\sqrt{\tau}} \to N(0,1). \tag{22}$$

where τ is obtained in equation (20) and it is proven that $A\hat{U}C \sim N\left(0,\tau\right)$, $\Gamma'_n = -\left(n-1\right)!\left[\frac{1}{n} + \gamma - \sum_{k=1}^{n}\frac{1}{k}\right]$ where γ is Euler-Mascheroni constant approximately equal to 0.5772. Note: $A\hat{U}C$ is an Unbiased Estimator of AUC (See Appendix D for the proof).

Confidence Interval for *AÛC*

Asymptotic Confidence Interval

The asymptotic $100(1-\alpha)\%$ confidence interval for accuracy is given by

$$\left[A\hat{U}C - Z_{\alpha/2} SE(A\hat{U}C),\ A\hat{U}C + Z_{\alpha/2} SE(A\hat{U}C)\right]. \tag{23}$$

where SE($A\hat{U}C$) can be obtained from equation (21), α is the level of significance and $Z_{\frac{\alpha}{2}}$ is the critical value. For example, $Z_{\frac{\alpha}{2}}$ for a 5% level of significance is 1.96.

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 $W(\alpha_0, \beta_0)$ and $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from $W(\alpha_1, \beta_1)$. By using equation (13) and (14), the ML estimates of the parameters α, β_0, β_1 are estimated.

Step 2: By using the estimated parameters $\hat{\alpha}, \hat{\beta}_0$ *and* $\hat{\beta}_1$, 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. $\hat{\alpha}_b, \hat{\beta}_{b0}$ *and* $\hat{\beta}_{b1}$ are obtained. Using these bootstrap estimates the accuracy ($A\widehat{U}C_b$) is obtained.

Step 3: Step 2 is repeated 10,000 times. The mean of all 10,000 estimates of $\hat{\alpha}_b$'s, $\hat{\beta}_{b0}$'s *and* $\hat{\beta}_{b1}$'s are called the bootstrap estimates of parameters α, β_0 *and* β_1 respectively and mean of all $A\widehat{U}C_b$'s is called the estimated bootstrap accuracy. The standard deviation of all estimates $A\widehat{U}C_b$ is called the standard error of $A\widehat{U}C_b$.

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

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

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

Simulation Studies

Thus, the accuracy, standard error of $A\widehat{U}C$ and 95 % confidence interval for $A\widehat{U}C$ have been computed through four different techniques via Monte Carlo method, asymptotic MLE method, parametric bootstrap and non-parametric method.

Monte Carlo Method

The model in equation (3) does not possess a closed form, so Monte Carlo integration of equation (3) is necessary. A Monte Carlo simulation was performed to inspect the accuracy obtained by Monte Carlo integration. The Monte Carlo estimate of AUC, SE ($A\widehat{U}C$) and 95% confidence interval for $A\widehat{U}C$ is presented in Table 1. The R codes for the Monte Carlo simulation is provided in Appendix A.

Table 1. Accuracy, standard error and Confidence interval of $\widehat{AUC}$ based on Constant Shape Bi-Weibull ROC model through Monte Carlo Simulation

SL. No.	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\widehat{AUC}$	V($\widehat{AUC}$)	95% Confidence Interval	Band Width
1	3.0	2.0	9	45	0.9188	0.051969	[0.816923, 1]	0.1831
2	3.0	2.0	9	30	0.8835	0.072986	[0.740444, 1]	0.2596
3	2.5	1.5	9	12	0.7590	0.124569	[0.514798, 1]	0.4852
4	3.5	2.5	9	10	0.6727	0.180434	[0.319085, 1]	0.6809

Asymptotic MLE Method

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 column 2, 3 and 4 of Table 2. The corresponding accuracy, SE, 95% confidence interval and the band width are shown in 5, 6, 7, 8 columns of Table 1. As the accuracy increases, the SE tend to decrease, simultaneously, the coverage area of the confidence band are tends to decrease as accuracy increases. Because the asymptotic distribution is independent of α, α may be kept constant or it may vary. From the sample α is estimated using iterative procedure from equation (14) and using α, the other two parameters using were found using equation (13). Hence, the ML estimate of AUC is obtained. The 95% asymptotic confidence interval and the confidence width are also calculated.

Table 2. Accuracy, standard error and Confidence interval of $\hat{A}$ based on Constant Shape Bi-Weibull ROC model through Asymptotic MLE method

Sl. No.	$\hat{\alpha}$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{A}$	$V(\hat{A})$	95% Confidence Interval	Band Width
1	2.530	12136	245.4000	0.980	0.00913	[0.9623, 0.99133]	0.02903
2	3.140	66.123	6.8201	0.907	0.02924	[0.8491, 0.96380]	0.11460
3	1.520	249.980	43.7500	0.850	0.03960	[0.7735, 0.92860]	0.15510
4	1.430	167.430	47.3900	0.778	0.04950	[0.6824, 0.87638]	0.19398
5	1.085	36.290	18.7200	0.660	0.05990	[0.5425, 0.77700]	0.23450

Table 3 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 $\hat{\beta}_0 = 5$ and for different values of $\hat{\beta}_1$ viz. 8, 12, 20, 100. In Tables 3 and 4, first row represents the AUC, second row gives the SE, third row gives the lower confidence limit and the fourth row represents the upper confidence limit. It is observed that, as the sample size increases the variance decreases and the coverage area of confidence interval is narrow.

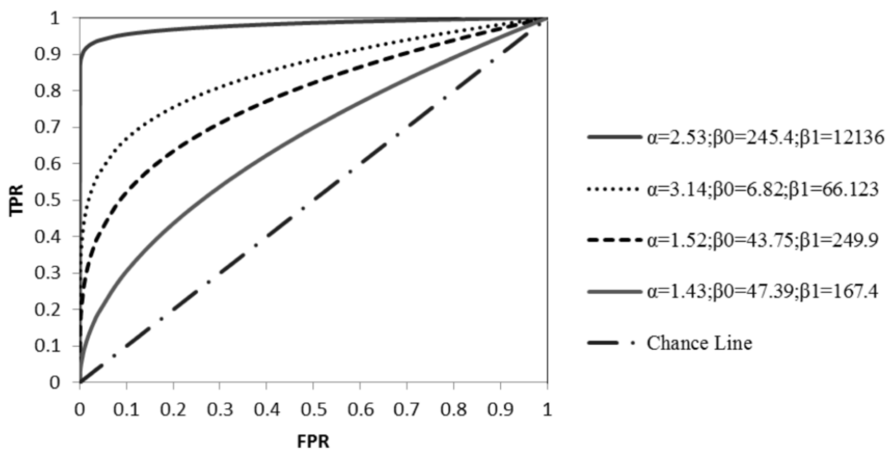

Figure 1. Constant Shape Bi-Weibull ROC model plotted for different AUC

Table 4 shows simulated independent samples of m controls and n cases ($m = n$ = 40, 50, 80, 100) to inspect the behavior of asymptotic MLE and confidence interval over different sample sizes by fixing $\hat{\beta}_1 = 45$ and for different values of $\hat{\beta}_0$ viz. 3, 8, 10, 20. It is observed that, as the sample size increases the variance decreases and the coverage area of confidence interval is narrow.

Table 3. Accuracy, Variance and 95% confidence Interval for AUC when $\hat{\beta}_0$ = 5 for different sample size

Sample Size	***m* = *n* = 5**	***m* = *n* = 10**	***m* = *n* = 40**	***m* = *n* = 50**	***m* = *n* = 80**	***m* = *n* = 100**
$\hat{\beta}_1 = 8$	0.6154 0.0232 0.3171 0.9137	0.6137 0.0116 0.4045 0.8263	0.6137 0.0029 0.5099 0.7192	0.6137 0.0023 0.5210 0.7080	0.6137 0.0012 0.5408 0.6883	0.6137 0.0012 0.5487 0.6804
$\hat{\beta}_1 = 12$	0.7059 0.0193 0.4339 0.9778	0.7057 0.0096 0.5136 0.8982	0.7057 0.0024 0.6097 0.8020	0.7057 0.0019 0.6199 0.7918	0.7057 0.0012 0.6379 0.7738	0.70570 0.00096 0.64510 0.76610
$\hat{\beta}_1 = 20$	0.8000 0.0132 0.5745 1.0000	0.8000 0.0066 0.6406 0.9594	0.8000 0.0017 0.7203 0.8797	0.8000 0.0013 0.7287 0.8730	0.80000 0.00083 0.74360 0.85640	0.8000 0.0007 0.7496 0.8504
$\hat{\beta}_1 = 100$	0.9524 0.0019 0.8659 1.0000	0.9500 0.0028 0.7961 1.0000	0.95000 0.00024 0.92180 0.98060	0.9500 0.0001 0.9250 0.9773	0.95000 0.00012 0.93080 0.97160	0.95000 0.00009 0.93310 0.96930

Table 4: Accuracy, SE and 95% confidence Interval for AUC when $\hat{\beta}_1 = 45$ for different sample size

Sample Size	***m* = *n* = 5**	***m* = *n* = 10**	***m* = *n* = 40**	***m* = *n* = 50**	***m* = *n* = 80**	***m* = *n* = 100**
$\hat{\beta}_0 = 3$	0.9375 0.0029 0.8318 1.0000	0.9375 0.0015 0.8628 1.0000	0.93750 0.00036 0.90010 0.97490	0.93750 0.00029 0.90410 0.97090	0.93750 0.00018 0.91110 0.96390	0.93750 0.00015 0.91390 0.96110
$\hat{\beta}_0 = 8$	0.8490 0.0096 0.6575 1.0000	0.8491 0.0048 0.7136 0.9845	0.849100 0.001194 0.781300 0.916782	0.84910 0.00096 0.78850 0.90960	0.8491 0.0006 0.8012 0.8969	0.84910 0.00047 0.80620 0.89190
$\hat{\beta}_0 = 10$	0.8182 0.0119 0.6044 1.0000	0.8182 0.0059 0.6670 0.9694	0.8182 0.0015 0.7426 0.8938	0.8182 0.0012 0.7506 0.8858	0.81820 0.00074 0.76470 0.87160	0.818200 0.000595 0.770400 0.865900
$\hat{\beta}_0 = 20$	0.6923 0.0200 0.4154 0.9693	0.6923 0.0100 0.8881 0.4965	0.9500 0.0025 0.5944 0.7902	0.6923 0.0019 0.6048 0.7799	0.6923 0.0012 0.6231 0.7615	0.692300 0.000998 0.630400 0.754200

Estimation of Bootstrap Variance

For parametric bootstrapping, the data was generated from a uniform distribution using (*m, n*) as specified in Table 5. Then by inverse transformation method, it is converted into Weibull variate with the values of $\alpha_0, \alpha_1, \beta_0$ and β_1. Using Step 1 results in the estimates as $\hat{\alpha}$, $\hat{\beta}_0$, $\hat{\beta}_1$. By using Step 2, the estimate of bootstrap sample is obtained as $\hat{\alpha}_b$, $\hat{\beta}_{b0}$ and $\hat{\beta}_{b1}$. From these 10,000 estimates of parameters, one can find an estimate of AUC by using the equation (10). By averaging these 10,000 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 5 shows the bootstrap area under the curve, SE and confidence interval for $\widehat{AUC}_b$.

Table 5. Accuracy, standard error and Confidence interval of $\widehat{AUC}$ based on Constant Shape Bi-Weibull ROC model through Bootstrap Simulation

(*m*, *n*)	$\hat{\alpha}_b$	$\hat{\beta}_{b0}$	$\hat{\beta}_{b1}$	$\widehat{AUC}_b$	$SE(\widehat{AUC}_b)$	**95% Confidence Interval**	**Band Width**
(10, 10)	2.6709	7.5414	201.8100	0.9249	0.04700	[0.8328, 1.0000]	0.1672
(20, 20)	2.5274	6.3739	103.9060	0.9240	0.03366	[0.8581, 0.9899]	0.1318
(30, 30)	2.4662	5.8770	84.2817	0.9220	0.02680	[0.8695, 0.9745]	0.1050
(50, 50)	2.4340	5.6493	74.3820	0.9222	0.02110	[0.8581, 0.9636]	0.1055
(100,100)	2.3948	5.4283	66.5260	0.9211	0.01450	[0.8926, 0.9496]	0.0570

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.

Sensitivity and Specificity

To generate a Weibull random variate with parametric values $\alpha_0 = 3; \alpha_1 = 2; \beta_0 = 9$ and $\beta_1 = 45$. The data is

X = {0.76261, 0.803019, 0.863084, 0.905439, 1.146029, 1.338408, 1.366008, 1.39672, 1.415312, 1.432053, 1.592267, 1.608494, 1.673259, 1.710255, 1.81614, 1.899346, 1.903763, 1.991144, 2.011153, 2.024541, 2.05607, 2.31567, 2.36017, 2.376429, 2.516461, 2.660371, 2.663695, 2.669402, 2.73371, 3.092265}

Y = {1.183838, 1.472276, 1.849655, 3.121439, 3.298009, 3.478297, 3.512602, 3.853157, 4.751021, 5.094757, 5.143248, 5.263026, 5.682114, 5.824499, 6.555983, 6.71353, 6.747835, 7.373468, 7.736402, 7.743548, 8.111, 8.393854, 9.171785, 9.313726, 9.789551, 10.28716, 10.63431, 11.08168, 12.01407, 12.10905}

Using equations (12) and (13), ML estimates are found to be $\alpha = 2.705, \beta_0 = 6.539$ and $\beta_1 = 245.0269$. Using equations (6) and (7) the sensitivity and specificity of the test were also calculated: the sensitivity of the test is 94% and specificity is 89%. To the data generated above all the four methods were applied and compared (see Table 6). The non-parametric estimates are obtained by the method of Hanley and McNeil (1982), and the R codes are given in Appendix F.

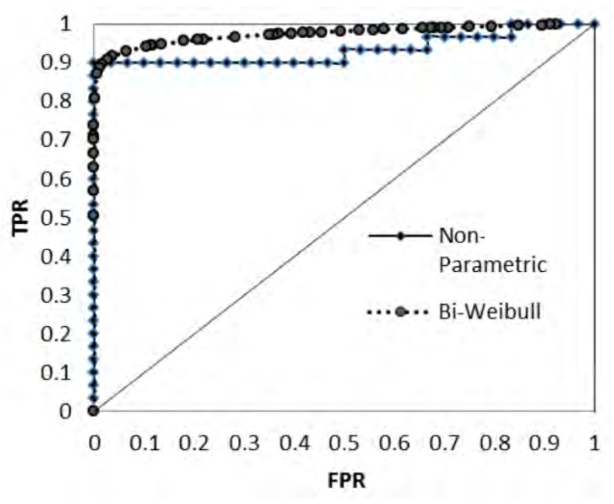

Figure 2. Constant Shape Bi-Weibull ROC curve plotted for simulated data

Conclusion

This article considered a ROC model developed from two parameter Weibull distributions for evaluating the accuracy of biomarkers in predicting disease status. 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 assumed. It should be noted that, the accuracy remains unchanged by this assumption. 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 $A\hat{U}C$, $SE(A\hat{U}C)$ and 95% confidence interval were found. The behavior of asymptotic SE and confidence interval is studied through simulation. The parametric AUC is higher than the AUC obtained by other methods including Monte Carlo, non-parametric and parametric bootstrap.

References

Amala, R., & Pundir, S. (2012). Statistical inference on AUC from a bi-lognormal ROC model for continuous data. *International Journal of Engineering Science and Innovative Technology*, *1*(2): 283-295.

Betinec, M. (2008). Testing the difference of the ROC curves in biexponential model. *Tatra Mountains Mathematical Publications*, *39*: 215-223.

Campbell, G., & Ratnaparkhi, M. V. (1993). An application of Lomax distributions in receiver operating characteristic (ROC) curve analysis. *Communication in Statistics*, *22*(6): 1681-1687.

Dorfman, D. D., et al. (1996). Proper receiver operating characteristics analysis: The bi-Gamma model. *Academic Radiology*, *4*: 138-149.

Hussain, E. (2011). The ROC curve model from generalized-exponential distribution. *Pakistan Journal of Statistics and Operations Research*, *7*(2): 323-330.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1): 29-36.

Kundu, D., & Gupta, R. D. (2006). Estimation of P[Y<X] for Weibull Distribution. *IEEE Transactions on Reliability*, *55*(2): 270-280.

Oglive, J. C., & Creelman, C. D. (1968). Maximum-likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology*, *5*: 377-391.

Pundir, S., & Amala, R. (2012a). A study on the Bi-Rayleigh ROC model. *Bonfring International Journal of Data Mining*, *2*(2): 42-47.

Pundir, S., & Amala, R. (2012b). A study on the comparison of bi-Rayleigh ROC model with bi-Gamma ROC model. In *Application of Reliability Theory and Survival Analysis*, 196-209. Coimbatore, India: Bonfring Publications.

Pundir, S., & Amala, R. (2013). Parametric receiver operating characteristic modeling for continuous data: A glance. *Model Assisted Statistics and Application*, *9*(2): 121-135. doi: 10.3233/MASA -130284.

Pundir, S. & Amala, R. (2014). Statistical Inference on AUC in Normal-Exponential ROC model. *Aligarh Journal of Statistics, 34*: 67-84.

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

Appendix A. R Code for Evaluation of AUC and Estimation of Standard Error Using Monte Carlo Simulation

```
m<-100; a0<-2.9753; a1=2.30387;
b0<-10295.0304;b1<-20646.898;x<-runif(m)
auc<-mean(exp(-(1/b1)* ( (-b0*log(x))^(a1/a0) ) ))
print(auc)
v.auc<-var(exp(-(1/b1)* ( (-b0*log(x))^(a1/a0) ) ))
print(v.auc); print(sqrt(v.auc))
```

Appendix B. Evaluation of AUC

$$\int_0^{\infty}\int_0^{y} \frac{\alpha}{\beta 1} * y^{\alpha-1} * Exp\left[\frac{-y^{\alpha}}{\beta 1}\right] * \frac{\alpha}{\beta 0} * x^{\alpha-1} * Exp\left[\frac{-x^{\alpha}}{\beta 0}\right] dxdy$$

$$\text{Conditional Expression}\left[\frac{\beta 1}{\beta 0+\beta 1}, \text{Re}\left[\frac{1}{\beta 0}+\frac{1}{\beta 1}\right] > 0 \,\&\&\, \text{Re}[\beta 1] > 0 \,\&\&\, \text{Re}[\alpha] > 0\right]$$

Appendix C. Evaluation of Asymptotic Distribution of AUC

i) $E\left[X^{\alpha}\right] = \int_0^{\infty} x^{\alpha} * \frac{\alpha}{\beta 0} * x^{\alpha-1} * Exp\left[\frac{-x^{\alpha}}{\beta 0}\right] dx$

Conditional Expression $[\beta 0, \text{Re}[\alpha] > 0 \,\&\&\, Re[\beta 0] > 0]$

ii) $E\left[X^{\alpha}\left(LogX\right)^2\right] \int_0^{\infty} x^{\alpha}\left(Log[x]\right)^\wedge 2 * \frac{\alpha}{\beta 0} * x^{\alpha-1} * Exp\left[\frac{-x^{\alpha}}{\beta 0}\right] dx$

$$\text{Conditional Expression}\left[\frac{1}{6\alpha^2}\beta 0\begin{pmatrix} 6(-2+\text{EulerGamma}) \\ \text{EulerGamma}+\pi^2+6\text{Log}[\beta 0] \\ (2-2\text{EulerGamma}+\text{Log}[\beta 0]) \end{pmatrix},\right.$$

$\text{Re}[\alpha] > 0 \,\&\&\, Re[\beta 0] > 0]$

iii) $E\left[Y^{\alpha}\right] = \int_0^{\infty} y^{\alpha} * \frac{\alpha}{\beta 1} * y^{\alpha-1} * Exp\left[\frac{-y^{\alpha}}{\beta 1}\right] dy$

Conditional Expression$[\beta 1, \text{Re}[\alpha] > 0 \,\&\&\, Re[\beta 1] > 0]$

iv) $$E\left[Y^{\alpha}\left(Log\ Y\right)^{2}\right]=\int_{0}^{\infty}y^{\alpha}*\left(Log\left[y\right]\right)^{\wedge}2*\frac{\alpha}{\beta 1}*y^{\alpha-1}*Exp\left[\frac{-y^{\alpha}}{\beta 1}\right]dy$$

$$\text{Conditional Expression}[\frac{1}{6\alpha^{2}}\beta 1\begin{pmatrix}6\left(-2+\text{EulerGamma}\right)\\ \text{EulerGamma}+\pi^{2}+6\text{Log}\left[\beta 1\right]\\ \left(2-2\text{EulerGamma}+\text{Log}\left[\beta 1\right]\right)\end{pmatrix},$$

$$\text{Re}\left[\alpha\right]>0\&\&Re\left[\beta 1\right]>0]$$

v) $$E\left[X^{\alpha}Log\ X\right]=\int_{0}^{\infty}x^{\alpha}*Log\left[x\right]*\frac{\alpha}{\beta 0}*x^{\alpha-1}*Exp\left[\frac{-x^{\alpha}}{\beta 0}\right]dx$$

$$\text{Conditional Expression}[-\frac{\beta 0\left(-1+\text{EulerGamma}+\text{Log}\left[\frac{1}{\beta 0}\right]\right)}{\alpha},$$

$$\text{Re}\left[\alpha\right]>0\&\&Re\left[\beta 0\right]>0]$$

vi) $$E\left[Y^{\alpha}Log\ Y\right]=\int_{0}^{\infty}y^{\alpha}*Log\left[y\right]*\frac{\alpha}{\beta 1}*y^{\alpha-1}*Exp\left[\frac{-y^{\alpha}}{\beta 1}\right]dy$$

$$\text{ConditionalExpression}[-\frac{\beta 1\left(-1+\text{EulerGamma}+\text{Log}\left[\frac{1}{\beta 1}\right]\right)}{\alpha},$$

$$\text{Re}\left[\alpha\right]>0\&\&Re\left[\beta 1\right]>0]$$

vii) $$E\left[Y\log Y\right]\int_{0}^{\infty}y*Log\left[y\right]*\frac{\alpha}{\beta 1}*y^{\alpha-1}*Exp\left[\frac{-y^{\alpha}}{\beta 1}\right]dy$$

$$\text{ConditionalExpression}[-\frac{1}{\alpha^{2}}\beta 1^{\frac{1}{\alpha}}\text{Gamma}\left[\frac{1}{\alpha}\right]$$

$$\left(\text{EulerGamma}-\text{HarmonicNumber}\left[\frac{1}{\alpha}\right]+\text{Log}[\frac{1}{\beta 1}]\right),$$

$$\text{Re}\left[\alpha\right]>0\&\&Re\left[\beta 1\right]>0]$$

viii) $E\left[Y\log Y\right]\int_0^{\infty} y * Log\left[y\right] * \frac{\alpha}{\beta 1} * y^{\alpha-1} * Exp\left[\frac{-y^{\alpha}}{\beta 1}\right]dy$

$$\text{ConditionalExpression}[-\frac{1}{\alpha^2}\beta 1^{\frac{1}{\alpha}}\text{Gamma}\left[\frac{1}{\alpha}\right]$$

$$\left(\text{EulerGamma}-\text{HarmonicNumber}\left[\frac{1}{\alpha}\right]+\text{Log}[\frac{1}{\beta 1}]\right),$$

$$\text{Re}\left[\alpha\right]>0\,\&\&\,Re\left[\beta 1\right]>0]$$

ix) The first order differentiation of Γ_n is given by $\Gamma_n\psi(n)$ where $\psi(n)$ is called the digamma function. The value of Γ'_n at n is equal to is $1-\gamma$; where γ is the Euler-Mascheroni constant has the approximate value 0.5772. The second order differentiation of Γ_n can be representated as $\Gamma''_n = \int_0^{\infty} x^{n-1}e^{-x}\left(\log x\right)^2 dx$ has the value $-1+\left(1-\gamma\right)^2+\frac{\pi^2}{6}$.

In general the m^{th} derivative of Γ_n is obtained by

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

x) $\psi\left(n\right)$

xi) $-1+\left(1-\gamma\right)^2+\frac{\pi^2}{6}$.

Appendix D. Unbiasedness of Estimated AUC

An estimator T is said to be an unbiased estimator if it satisfies the condition $E(T) = \mu$. The estimated accuracy is

$$A\hat{U}C = \frac{\hat{\beta}_1}{\hat{\beta}_1+\hat{\beta}_0} = \frac{\frac{\sum_{j=1}^{n} y_j^{\alpha}}{n}}{\frac{\sum_{j=1}^{n} y_j^{\alpha}}{n}+\frac{\sum_{i=1}^{m} x_i^{\alpha}}{m}}$$

Taking the expectation results in

$$E(A\hat{U}C) = \frac{\sum_{j=1}^{n} E(y_j^{\alpha})}{\sum_{j=1}^{n} E(y_j^{\alpha}) + \sum_{i=1}^{m} E(x_i^{\alpha})} = \frac{\beta_1}{\beta_1 + \beta_0} = AUC$$

Hence $A\hat{U}C$ is an unbiased estimator of AUC.

Appendix E. R Code for Evaluation of Bootstrap AUC and Confidence Interval

```
k<-10000 ;al0<-3.9; al1<-2.84;be1<-38.56;be0<-11.31;m<-30;n<-30;
df1 <-data.frame(array(dim=c(n,k))); df0 <-data.frame(array(dim=c(m,k)))
dfw0<-data.frame(array(dim=c(m,k))); dfw1<-data.frame(array(dim=c(n,k)))
a<-array(dim=k); ave<-array(dim=k); b1<-array(dim=k); b0<-array(dim=k)
auc<-array(dim=k); SE<-array(dim=k); for(i in 1:k)
{
df1[i]<-runif(30); df0[i]<-runif(30);
dfw0[i]<-(-be0*log(1-df0[i]))^(1/al0);
dfw1[i]<-(-be1*log(1-df1[i]))^(1/al1);
loglik<-function(param)
{
a[i]<-param[1]; b0[i]<-param[2]; b1[i]<-param[3]
ll<-(m+n)*log(a[i])+(a[i]-1)*(sum(log(dfw1[i]))+sum(log(dfw0[i])))-n*log(b1[i])-
m*log(b0[i])-(sum(dfw1[i]^a[i])/b1[i])-(sum(dfw0[i]^a[i])/b0[i])
ll
}
M0<-maxNR(loglik,start=c(1,2,3))
a[i]<-M0$estimate[1]; b0[i]<-M0$estimate[2]; b1[i]<-M0$estimate[3]
auc[i]<-(b1[i]/(b1[i]+b0[i])); dt<-data.frame(a[i],b0[i],b1[i],auc[i])
}
print(dt); b.auc<-mean(auc); b.se.auc<-sd(auc);
cat("Bootstrap Accuracy=","\n",b.auc,"\n")
cat("Bootstrap Standard Error=","\n", b.se.auc)
lcl<-(b.se.auc-(1.96*b.se.auc)); ucl<-(b.se.auc+(1.96*b.se.auc))
```

Appendix F. R code for Sensitivity and Specificity Analysis

```
s<-sort(c(h,d)); n<-length(d); m<-length(h); X<-array(dim=m+n-1)
k<-m+n-1;
for(i in 1:k)
       {
X[i]<-(s[i]+s[i+1])/2;
```

```
        }
t<-c(s[1]-1,X,s[m+n]+1); print(t); a<-2.705492 ; b0<-6.538623 ;
b1<-245.026947; # Estimated by MLE from data
Sen<-exp((-t^a)/b1); Sp<-1-exp((-t^a)/b0); dt<-data.frame(t, Sen, Sp);
```

Appendix G. R code for Non-Parametric Method

```
NP.ROC<-function(h,d) # Creating a function named NP.ROC()
{
s<-sort(c(h,d)); n<-length(d); m0<-mean(h);m1<-mean(d);m<-length(h)
X<-array(dim=m+n-1);k<-m+n-1;
for(i in 1:k)
        {
X[i]<-(s[i]+s[i+1])/2;
        }
t<-c(s[1]-1,X,s[m+n]+1); print(t);
TPR<-array(dim=length(t)) # Defining empty array to save calculations
FPR<-array(dim=length(t)); TP<-array(dim=length(t)); TN<-array(dim=length(t));
FN<-array(dim=length(t)); FP<-array(dim=length(t)); AUC<-array(dim=length(t));
SP<-array(dim=length(t)); TNR<-array(dim=length(t)); SplusS<-
array(dim=length(t));
se<-array(dim=length(t));q1<-array(dim=length(t));q2<-array(dim=length(t));v<-
array(dim=length(t));
        for(i in 1:length(t))
        {
A<-d[d>=t[i]]# observations greater than or equal to t among diseased i.e. True
Positives
B<-d[d<t[i]] # observations less than t among diseased i.e. False Negatives
C<-h[h>=t[i]]# observations greater than or equal to t among healthy i.e. False
Positives
D<-h[h<t[i]] # observations less than t among healthy i.e. True Negatives
TP[i]<-length(A) # No. of TPs
FP[i]<-length(C) # No. of FPs
FN[i]<-length(B) # No. of FNs
TN[i]<-length(D) # No. of TNs
TPR[i]<-(TP[i]/n)
FPR[i]<-(FP[i]/m)
TNR[i]<-1-FPR[i] # or TN[i]/m
AUC[i]<-(TP[i]+TN[i])/(TP[i]+TN[i]+FN[i]+FP[i])
SplusS[i]<-TPR[i]+TNR[i] # TNR+TPR
q1[i]<-AUC[i]/(2-AUC[i]); q2[i]<-(2*AUC[i]^2)/(1+AUC[i])
v[i]<-(AUC[i]*(1-AUC[i])+(n-1)*(q1[i]-AUC[i]^2)+(m-1)*(q2[i]-AUC[i]^2))/(m*n)
se[i]<-sqrt(v[i]);
        }
library(utils)
write.csv(dt,"msanalysis.csv") # writing the data frame in CSV format for usage
        m<-length(h); n<-length(d); l<-m*n;
        sum=0;
                for(i in 1 : m)
                {
                s<-c(0);
```

```
                    for(j in 1 : n)
                    {
                           if(d[j]>h[i])
                           {
                           s[j]=1;
                           }
                                  else if (d[j]==h[i])
                                  {
                                  s[j]=0.5;
                                  }
                           else
                           s[j]=0
                    }
             output=data.frame(s)
             sum=sum+sum(output)
             }
value= sum/(m*n)
print(value)
dt<-data.frame (t, FPR, TPR, TP, TN, FP, FN, AUC, se)
print(dt); Q1<-value/(2-value); Q2<-(2*value^2)/(1+value)
V<-(value*(1-value)+(n-1)*(Q1-value^2)+(m-1)*(Q2-value^2))/(m*n)
SE<-sqrt(V); # Standard Error of AUC
lc<-value-(SE*1.96) # Lower Confidence Limit of AUC
uc<-value+(SE*1.96) # Upper Confidence Limit of AUC
if(uc>1){ # Sometimes if the standard error is high, the upper CI may go greater
that one in which case approximating it to one.
uc<-1.0
}
cat("---------------------------------------",
"\n", "Healthy Mean","\t",":","\t",m0,
"\n", "Diseased Mean","\t",":","\t",m1,
"\n", "AUC","\t","\t",":","\t", value,
"\n", "SE","\t","\t",":","\t", SE,
"\n", "CI","\t","\t",":","\t","[", lc,",","\t",uc,"]",
"\n","---------------------------------------","\n")
plot(TPR~FPR,type="b",main="",xlab="FPR",ylab="TPR",xlim=c(0,1),ylim=c(0,1))
abline(lm(c(0:1)~c(0,1)))
}
NP.ROC(h,d);
```