Research Article

ESTIMATION OF AUC FOR CONSTANT SHAPE BI-WEIBULL FAILURE TIME DISTRIBUTION

Leo Alexander, T* and Lavanya, A

Department of Statistics, Loyola College, Chennai–34, India

ARTICLE INFO	ABSTRACT
	The Receiver Operating Characteristic (ROC) curve is commonly used for evaluating the discriminatory ability of a biomarker. The conventional ways expressing the true accuracy of the test is by using its summary measure Area Under the ROC Curve (AUC) and intrinsic measures Sensitivity and Specificity. We propose a Bayesian approach for the estimation of the AUC under the Constant Shape Bi-Weibull Distribution using Extension of Jeffreys' Prior Information with three Loss Functions. Theoretical results are validated by simulation studies. Simulations indicated that estimate of AUC values was good even for relatively small sample sizes (n=25). Bayes estimation with General Entropy loss function provides the highest AUC values when the loss parameter is -1.6, according to the extension of Jeffreys prior value is 0.4 or 1.4. When AUC≤0.6, which indicated a marked overlap between the outcomes in diseased and non-diseased populations. An illustrative example is also provided to explain the concepts.

1. INTRODUCTION

Receiver operating characteristic (ROC) curves are widely used for the evaluation of continuous or ordinal diagnostic tests and biomarkers [6, 7, 22]. Measurements for a diagnostic test may be subject to an analytic limit of detection leading to immeasurable or unreported test results. Ignoring the scores that are beyond the limit of detection of a test leads to a biased assessment of its discriminatory ability, as reflected by indices such as the associated Area Under the ROC Curve (AUC). The ROC Curve is embedded by the two intrinsic measures Sensitivity (Sn) or True Positive Rate (TPR) and 1-Specificity (Sp) or False Positive Rate (FPR) along with its accuracy measure AUC. The corresponding AUC provides a global summary statistic indicating the overall discriminatory ability of a test independent of any cut-off point and may be used to compare the performance of different tests for detection of the condition of interest. The AUC is defined as the total Area under the ROC curve represented by "sensitivity" versus "1-specificity" corresponding to all possible cut-off points. The AUC can be interpreted as the average sensitivity, across all possible False-Positive Fractions [28]. An alternative interpretation is the proportion of the time that test scores for individuals with the condition of interest will exceed (or be less than) those of individuals without the condition. The term ROC analysis was coined during II world war to analyze the radar signals [21]. The application of ROC Curve technique was promoted in diversified fields such as experimental psychology [8], industrial quality control [5] and military monitoring [25]. [8] Was first to use the Gaussian model for estimating the ROC Curve. [3] Gave the maximum likelihood estimates for the ROC Curve parameters considering yes or no type responses and rating data. The importance of ROC Curve in medicine was due to [17] analyze the radiographic images. [19] Proposed a methodology to describe the scores of rating type by embedding the approach of [3]. [10] Explained the importance and robustness of Binormal ROC Curve. Recently [23] explained the Bayesianestimation of the receiver operating characteristic curve for a diagnostic test with a limit of detection in the absence of a gold standard.

We also developed Functional Relationship between Brier Score and Area Under the Constant Shape Bi-Weibull ROC Curve [16], Confidence Intervals Estimation for ROC Curve, AUC and Brier Score under the Constant Shape Bi-Weibull Distribution [13],Asymmetric and Symmetric Properties of Constant Shape Bi-Weibull ROC Curve Described by Kullback-Leibler Divergences [14], and Bayesian Estimation of Parameters under the Constant Shape Bi-Weibull Distribution Using Extension of Jeffreys' Prior Information with Three Loss Functions [15].

Corresponding author: ***Leo Alexander, T***
Department of Statistics, Loyola College, Chennai–34, India

The main purpose of this paper is to compare the traditional Maximum Likelihood Estimation of the AUC of the Constant Shape Bi-Weibull distribution with its Bayesian counterpart using Extension of Jeffreys' Prior Information obtained from Lindley's approximation procedure based on three Loss Functions.

In this paper, the Bayesian Estimation of AUC under the Constant Shape Bi-Weibull Distribution is studied by Using Extension of Jeffreys' Prior Information with Three Loss Functions. This paper is organized as follows: In Section 2, estimation of AUC under MLE, Jeffreys' Prior Information and Extension of Jeffreys' Prior Information with Three Loss functions is discussed. Section 3, provides simulation study for proposed theory. Section 4 deals with validation of the proposed theory based on real time data. Conclusions are given in Section 5.

2. A Constant Shape Bi-Weibull Roc Model And Its Auc

In medical science, a diagnostic test result called a biomarker [11, 2] is an indicator for disease status of patients. The accuracy of a medical diagnostic test is typically evaluated by sensitivity and specificity. Receiver Operating Characteristic (ROC) curve is a graphical representation of the relationship between sensitivity and specificity. Hence the main issue in assessing the accuracy of a diagnostic test is to estimate the ROC curve.

Suppose that there are two groups of study subjects: diseased and no diseased. Let S be a continuous biomarker. Assume that ROC analysis based on the True Positive Probability (TPP),$P(S|D)$, and False Positive Probability (FPP),$P(S|H)$, in fundamental detection problems with only two events and two responses [26,8,4].

According to Signal Detection Theory(SDT), we assume that there are two probability distributions of the random variables X and *Y*, one associated with the signal event s and other with the non-signal event n[12]; these probability (or density) distributions of a given observation *x and y* are conditional upon the occurrence of *s* and *n* [4].

In the medical context, the signal event corresponds to the diseased group, and the nonsignal event to the no diseased group [9]. If the cutoff value is *c*, corresponding to a particular likelihood ratio, the TPP and FPP are given by the following expressions [4]:

Let x*, y* be the test scores observed from two populations with (diseased individuals) and without (no diseased individuals) condition respectively which follow Constant Shape Bi-Weibull distributions. The density functions of Constant Shape Bi-Weibull distributions are as follows,

$$f(x|D) = \frac{\beta}{\sigma_D} t^{\beta-1} e^{-\left[\frac{t^\beta}{\sigma_D}\right]}. \tag{1}$$

and

$$g(y|H) = \frac{\beta}{\sigma_H} t^{\beta-1} e^{-\left[\frac{t^\beta}{\sigma_H}\right]}. \tag{2}$$

The probabilistic definitions of the measures of ROC Curve are as follows:

$$\text{Sensitivity}(s_1) = P(S|D) = \int_t^\infty f(x|D)dx,$$

$$1 \quad \text{Specificity}(s_2) = P(S|H) = \int_t^\infty f(x|H)dx.$$

In this context, the (1-Specificity) and Sensitivity can be defined using equations (1) and (2) and are given in equations (3) and (4) respectively,

$$P(S|H) = y(t) = e^{-\left[\frac{t^\beta}{\sigma_H}\right]}. \tag{3}$$

and

$$P(S|D) = x(t) = e^{-\left[\frac{t^\beta}{\sigma_D}\right]}. \tag{4}$$

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

$$ROC(t) = x(t) = y(t)^b,$$

where$t = [\sigma_H \log y(t)]^{\frac{1}{\beta}}$is the threshold and$b = \frac{\sigma_H}{\sigma_D}$.

The accuracy of a diagnostic test can be explained using the Area under the Curve (AUC) of a ROC Curve.AUC describes the ability of the test to discriminate between diseased and no diseased 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,

$$AUC = \int_0^1 y(t)^b dx(t).$$

The closed form of AUC is as follows

$$AUC = \frac{1}{1+b}. \tag{5}$$

2.1 Maximum Likelihood Estimator of AUC

The MLE of two parameters Weibull distribution has been discussed [27] in the context of Reliability estimation. Let X1, X2, .. .…. Xm be a random sample of size m from $W(\beta, \sigma_D)$ and Y1, Y2, .. .…. Yn be a random sample of size n from $W(\beta, \sigma_H)$ .The likelihood function of the selected sample is given by

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

where$\theta = (\beta, \sigma_D, \sigma_H)'$

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

The log-likelihood function is

$$lnL = (m+n)ln\beta + (\beta - 1)\left[\sum_{i=1}^m lnx_i + \sum_{j=1}^n lny_j\right] - nln\sigma_H - mln\sigma_D - \frac{1}{\sigma_H}\sum_{j=1}^n y_j^\beta - \frac{1}{\sigma_D}\sum_{i=1}^m x_i^\beta, \tag{6}$$

Differentiating (5) with respect toβ, we get

$$\frac{\partial lnL}{\partial \beta} = \frac{(m+n)}{\beta} + \left[\sum_{i=1}^m lnx_i + \sum_{j=1}^n lny_j\right] - \frac{1}{\sigma_H}\sum_{j=1}^n y_j^\beta lny_j - \frac{1}{\sigma_D}\sum_{i=1}^m x_i^\beta lnx_i. \tag{7}$$

By differentiating the equation (5) with respect toσ_D, σ_H and equating to zero, we get the estimates. The MLE's of σ_D and σ_H are determined as,

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

$$\widehat{\sigma_D} = \frac{\sum_{i=1}^m x_i^\beta}{m}. \tag{8}$$

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

$$\widehat{\sigma_H} = \frac{\sum_{j=1}^n y_j^\beta}{n}. \tag{9}$$

Substituting equation (8) and (9) in equation (7) and equating it to zero, we get a non-linear equation:

$$(\hat{\beta}) = \frac{m + n + \sum_{j=1}^{n} y_j{}^{\beta} + \sum_{i=1}^{m} x_i{}^{\beta}}{\frac{n \sum_{j=1}^{n} y_j{}^{\beta} ln y_j}{\sum_{j=1}^{n} y_j{}^{\beta}} + \frac{m \sum_{i=1}^{m} x_i{}^{\beta} ln x_i}{\sum_{i=1}^{m} x_i{}^{\beta}}} \ . \tag{10}$$

Hence, $\hat{\beta}$ can be determined as a solution of Non-linear equation (10). By substituting equations (8), (9) and (10) in equation (5), will get an MLE estimate of AUC and is given as

$$\widehat{AUC}_{ML} = \frac{1}{1 + \hat{b}_{ML}} \ , w \ ere \hat{b}_{ML} = \frac{\hat{\sigma}_{HML}}{\hat{\sigma}_{DML}}.$$

2.2 Bayesian Estimation of AUC under Constant Shape Bi-Weibull Distribution

Bayesian Estimation approach has received a lot of attention in recent times for analyzing Failure Time data, which has mostly been proposed as an alternative to that of the traditional methods.

Bayesian Estimation approach makes use of once prior knowledge about the parameters as well as the available data. When once prior knowledge about the parameter is not available, it is possible to make use of the non-informative prior in Bayesian analysis.

Since we have no knowledge of the parameters, we seek to use the Extension of Jeffreys' Prior Information, where Jeffreys' Prior is the square root of the determinant of the Fisher information. According to [1], the extension of Jeffreys' prior is by taking $u(\theta) \propto [I(\theta)]^c$, $c\epsilon R^+$, so that

$$u(\theta) \propto \left[\frac{1}{\theta}\right]^{2c}$$

Thus,

$$u(\sigma, \beta) \propto \left[\frac{1}{\beta \sigma_D \sigma_H}\right]^{2c}.$$

Let X1, X2, .. …. Xm be a random sample of size m from $W(\beta, \sigma_D)$ and Y1, Y2, .. …. Yn be a random sample of size n from $W(\beta, \sigma_H)$ .The likelihood function of the selected sample is given by

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

where $\theta = (\beta, \sigma_D, \sigma_H)'$

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

With Bayes theorem, the joint posterior distribution of the parameters θ is

$$\pi(\theta|t) \propto L(t| \theta)u(\theta)$$

$$= \frac{k}{(\beta \sigma_D \sigma_H)^{2c}} \prod_{i=1}^{m} \frac{\beta}{\sigma_D} x_i{}^{\beta-1} e^{-\left[\frac{x_i{}^{\beta}}{\sigma_D}\right]} \prod_{j=1}^{n} \frac{\beta}{\sigma_H} y_j{}^{\beta-1} e^{-\left[\frac{y_j{}^{\beta}}{\sigma_H}\right]},$$

where k is the normalizing constant that makes π a proper pdf.

Remark 2.1

Here we consider two Asymmetric Loss Functions namely Linear Exponential (LINEX) Loss Function and General Entropy Loss Function. Also the Symmetric Loss Function namely Squared Error Loss Function considered in order to estimate AUC values.

2.2.1 Linear Exponential Loss Function (LINEX)

The LINEX Loss Function is under the assumption that the minimal loss occurs at $\hat{\theta} = \theta$ and is expressed as

$$L(\hat{\theta}\quad\theta) \propto \exp\left(a(\hat{\theta}\quad\theta)\right)\quad a(\hat{\theta}\quad\theta)\quad 1$$

where $\hat{\theta}$ is an estimation of θ and $a \neq 0$. The sign and magnitude of the shape parameter 'a' represents the direction and degree of symmetry, respectively. There is overestimation if **a**>0 and underestimation if **a**<0 but when **a** 0 , the LINEX Loss Function is approximately the Squared Error Loss Function. The posterior expectation of the LINEX Loss Function, according to [20], is

$$E_\theta L(\hat{\theta}\quad\theta) \propto exp(a\hat{\theta})E_\theta(\exp(\quad a\theta))\quad a(\hat{\theta}\quad E_\theta(\theta))\quad 1. \tag{11}$$

The Bayes Estimator of **θ**, represented by $\hat{\theta}_{BL}$ under LINEX Loss Function, is the value of $\hat{\theta}$ which minimizes equation (11) and is given as

$$\hat{\theta}_{BL} = \quad\frac{1}{a} \ln E_\theta(exp(\quad a\theta)).$$

Provided $E_\theta(exp(\quad a\theta)$ exists and is finite. The Bayes Estimator $\hat{u}_{BL}$ of a function

$$u = u(exp(\quad a\sigma_D), exp(\quad a\sigma_H), exp(\quad a\beta)) \text{ is given as}$$

$$\hat{u}_{BL} = E(\exp(\quad a\sigma_D), exp(\quad a\sigma_H), exp(\quad a\beta)|t)$$

$$= \frac{\iint u[exp(\quad a\sigma_D), exp(\quad a\sigma_H), exp(\quad a\beta)]\pi\ (\boldsymbol{\theta})d\boldsymbol{\theta}}{\iint \pi\ (\sigma_D, \sigma_H, \beta)d\sigma_D d\sigma_H d\beta} . \tag{12}$$

From (12), it can be observed that ratio of integrals which cannot be solved analytically and for that we employ Lindley's approximation procedure to estimate the parameters. Lindley considered an approximation for the ratio of integrals for evaluating the posterior expectation of an arbitrary function $\hat{u}(\boldsymbol{\theta})$ as

$$E[u(\theta)|x] = \frac{\int u(\theta)v(\theta)[L(\theta)]d\theta}{\int v(\theta)[L(\theta)]d\theta}.$$

According to [24], Lindley's expansion can be approximated asymptotically by

$$\hat{\theta} = u + \frac{1}{2}[u_{200}\delta_{200} + u_{020}\delta_{020} + u_{002}\delta_{002}] + u_{100}\rho_{100}\delta_{200} + u_{010}\rho_{010}\delta_{020} + u_{001}\rho_{001}\delta_{002}$$

$$+ \frac{1}{2}[L_{300}u_{100}\delta^2_{200} + L_{030}u_{010}\delta^2_{020} + L_{003}u_{002}\delta^2_{002}], \tag{13}$$

where L is the log-likelihood function in equation (6),

$$u(\sigma_D) = exp(\quad a\sigma_D), u_{100} = \frac{\partial u}{\partial\sigma_D} = \quad a\,exp(\quad a\sigma_D),$$

$$u_{200} = \frac{\partial^2 u}{\partial^2\sigma_D} = \quad a^2\,exp(\quad a\sigma_D),$$

$$u_{010} = u_{020} = u_{001} = u_{002} = 0,$$

$$u(\sigma_H) = exp(\quad a\sigma_H), u_{010} = \frac{\partial u}{\partial\sigma_H} = \quad a\,exp(\quad a\sigma_H),$$

$$u_{020} = \frac{\partial^2 u}{\partial^2\sigma_H} = \quad a^2\,exp(\quad a\sigma_H),$$

$$u_{100} = u_{001} = u_{200} = u_{002} = 0$$

$$u(\beta) = exp(\quad a\beta), u_{001} = \frac{\partial u}{\partial\beta} = \quad a\,exp(\quad a\beta),$$

$$u_{002} = \frac{\partial^2 u}{\partial^2\beta} = \quad a^2\,exp(\quad a\beta),$$

$$u_{100} = u_{010} = u_{200} = u_{020} = 0,$$

$$\rho(\sigma_D, \sigma_H, \beta) = \quad ln(\sigma_D{}^{2c}) \quad ln(\sigma_H{}^{2c}) \quad ln(\beta^{2c}),$$

$$\rho_{100} = \frac{\partial \rho}{\partial \sigma_D} = \quad \frac{1}{\sigma_D{}^{2c}},$$

$$\rho_{010} = \frac{\partial \rho}{\partial \sigma_H} = \quad \frac{1}{\sigma_H{}^{2c}},$$

$$\rho_{001} = \frac{\partial \rho}{\partial \beta} = \quad \frac{1}{\beta^{2c}},$$

$$\delta_{200} = (\ L_{200})^{-1}, \delta_{020} = (\ L_{020})^{-1}, \delta_{002} = (\ L_{002})^{-1},$$

$$L_{002} = \frac{\partial^2 L}{\partial^2 \beta} = \quad \left(\frac{m}{\beta^2}\right) \quad \frac{1}{\sigma_D} \sum_{i=1}^{m} x_i{}^\beta \, (lnx_i\ )^2 \quad \left(\frac{n}{\beta^2}\right) \quad \frac{1}{\sigma_H} \sum_{j=1}^{n} y_i{}^\beta \, (lny_j)^2,$$

$$L_{003} = \frac{\partial^3 L}{\partial^3 \beta} = 2\left(\frac{m}{\beta^3}\right) \quad \frac{1}{\sigma_D} \sum_{i=1}^{m} x_i{}^\beta \, (lnx_i\ )^3 + 2\left(\frac{n}{\beta^3}\right) \quad \frac{1}{\sigma_H} \sum_{j=1}^{n} y_i{}^\beta \, (lny_j\ )^3,$$

$$L_{200} = \frac{\partial^2 L}{\partial^2 \sigma_D} = \frac{m}{\sigma_D{}^2} \quad 2\frac{\sum_{i=1}^{m} x_i{}^\beta}{\sigma_D{}^3},$$

$$L_{020} = \frac{\partial^2 L}{\partial^2 \sigma_H} = \frac{n}{\sigma_H{}^2} \quad 2\frac{\sum_{j=1}^{n} y_j{}^\beta}{\sigma_H{}^3},$$

$$L_{300} = \frac{\partial^3 L}{\partial^3 \sigma_D} = \quad 2\frac{m}{\sigma_D{}^3} + 6\frac{\sum_{i=1}^{m} x_i{}^\beta}{\sigma_D{}^4},$$

$$and L_{030} = \frac{\partial^3 L}{\partial^3 \sigma_H} = \quad 2\frac{n}{\sigma_H{}^3} + 6\frac{\sum_{j=1}^{n} y_j{}^\beta}{\sigma_H{}^4}.$$

Bayesian Estimation of AUC using LINEX Loss Function is given as

$$\widehat{AUC}_{BL} = \frac{1}{1 + \hat{b}_{BL}} , w \ ere \hat{b}_{BL} = \frac{\hat{\sigma}_{HBL}}{\hat{\sigma}_{DBL}}$$

2.2.2 General Entropy Loss Function

Another useful Asymmetric Loss Function is the General Entropy (GE) Loss which is a generalization of the Entropy Loss and is given as

$$L(\hat{\theta} \quad \theta) \propto \left(\frac{\hat{\theta}}{\theta}\right)^k \quad kln\left(\frac{\hat{\theta}}{\theta}\right) \quad 1 \ .$$

The Bayes Estimator $\hat{\boldsymbol{\theta}}_{BG}$ of $\boldsymbol{\theta}$under the General Entropy Loss is

$$\hat{\theta}_{BG} = [\ E_\theta(\theta^{-k})]^{-\frac{1}{k}} \quad ,$$

provided $E_\theta(\theta^{-k})$exists and is finite.

The Bayes Estimator for this Loss Function is

$$\hat{u}_{BG} = E\{u[\sigma_D{}^{-k}, \sigma_H{}^{-k}, \beta^{-k}]|t\}$$

$$= \frac{\iint u[\sigma_D{}^{-k}, \sigma_H{}^{-k}, \beta^{-k}]\pi\ (\sigma_D, \sigma_H, \beta)d\sigma_D d\sigma_H d\beta}{\iint \pi\ (\sigma_D, \sigma_H, \beta)d\sigma_D d\sigma_H d\beta} .$$

Applying the same Lindley approach here as in (13) with $u100$, $u200$, $u010$, $u020$ and $u001$, $u002$ are the first and second derivatives for σ_D, σ_H and β, respectively, and are given as

$$u = [\sigma_D^{-k}], u_{100} = \frac{\partial u}{\partial \sigma_D} = k[\sigma_D^{-k-1}],$$

$$u_{200} = \frac{\partial^2 u}{\partial^2 \sigma_D} = (k^2 \quad k)\sigma_D^{-k-2},$$

$$u_{010} = u_{020} = u_{001} = u_{002} = 0,$$

$$u = [\sigma_H^{-k}], u_{010} = \frac{\partial u}{\partial \sigma_H} = k[\sigma_H^{-k-1}],$$

$$u_{020} = \frac{\partial^2 u}{\partial^2 \sigma_H} = (k^2 \quad k)\sigma_H^{-k-2},$$

$$u_{100} = u_{200} = u_{001} = u_{002} = 0,$$

$$u = [\beta^{-k}], u_2 = \frac{\partial u}{\partial \beta} = k[\beta^{-k-1}],$$

$$u_{22} = \frac{\partial^2 u}{\partial^2 \beta} = (k^2 \quad k)\beta^{-k-2},$$

$$and \ u_{100} = u_{010} = u_{200} = u_{020} = 0.$$

Bayesian Estimation of AUC using the General Entropy Loss is given as

$$\widehat{AUC}_{BG} = \frac{1}{1 + \hat{b}_{BG}}, w \ ere\hat{b}_{BG} = \frac{\hat{\sigma}_{HBG}}{\hat{\sigma}_{DBG}}.$$

2.2.3 Symmetric Loss Function

The Squared Error Loss is given by

$$L(\hat{\theta} \quad \theta) \propto (\hat{\theta} \quad \theta)^2.$$

This Loss Function is symmetric in nature, that is, it gives equal weightage to both over and under estimation.

In real life, we encounter many situations where overestimation may be more serious than underestimation or vice versa.

The Bayes Estimator $\hat{u}_{BS}$ of a function $u = u(\sigma_D, \sigma_H, \beta)$ of the unknown parameters under Square Error Loss Function (SELF) is the posterior mean, where

$$\hat{u}_{BS} = E\{u[\sigma_D, \sigma_H, \beta]|t\} = \frac{\iint u[\sigma, \beta] \pi (\sigma_D, \sigma_H, \beta)d\sigma_D d\sigma_H d\beta}{\iint \pi (\sigma_D, \sigma_H, \beta)d\sigma_D d\sigma_H d\beta}.$$

Applying the same Lindley approach here as in (13) where $u100$, $u200$, $u010$, $u020$ and $u001$, $u002$ are the first and second derivatives for σ_D, σ_H and β, respectively, and are given as

$$u = \sigma_D, \qquad u_{100} = \frac{\partial u}{\partial \sigma_D} = 1,$$

$$u_{010} = u_{001} = u_{200} = u_{020} = u_{002} = 0,$$

$$u = \sigma_H, \qquad u_{010} = \frac{\partial u}{\partial \sigma_H} = 1,$$

$$u_{100} = u_{001} = u_{200} = u_{020} = u_{002} = 0,$$

$$u = \beta, u_{001} = \frac{\partial u}{\partial \beta} = 1,$$

$and\ u_{100} = u_{010} = u_{200} = u_{020} = u_{002} = 0.$

Bayesian Estimation of AUC using The Squared Error Loss is given as

$$\widehat{AUC}_{BS} = \frac{1}{1+\hat{b}_{BS}}\ ,w\ ere\hat{b}_{BS} = \frac{\hat{\sigma}_{HBS}}{\hat{\sigma}_{DBS}}.$$

3. Simulation Study

Simulation studies are conducted with different combinations of scale and shape parameters of both diseased and non-diseased populations. At every parameter combination and sample size, the AUC is 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 AUC has been computed through different methods via MLE and Bayesian Estimation using Extension of Jeffreys' Prior Information obtained from Lindley's approximation procedure with three Loss Functions.

In our Simulation study, we chose a sample size of n=25, 50, and 100 to represent small, medium, and large dataset. The assumed scale and shape parameters of both populations are $\sigma_D = \sigma_H = 0.5$ and 1.5, $\beta = 0.8$ and 1.2. The values of Jeffreys' Extension are $c = 0.4$ and 1.4. The values for the Loss parameters **(a, k)** are $a=k=\pm0.6$ and ±1.6.

In Table 1 we present the estimated values for AUCfor both the Maximum Likelihood Estimation and Bayesian Estimation using anextension of Jeffrey's prior information with the three loss functions.

Table 1 Estimated values for AUC

(m, n)	σ_D	σ_H	β	c	$\widehat{AUC}_{ML}$	$\widehat{AUC}_{BS}$	$\widehat{AUC}_{BL}$ a=k=0.6	$\widehat{AUC}_{BG}$ a=k=0.6	$\widehat{AUC}_{BL}$ a=k=-0.6	$\widehat{AUC}_{BG}$ a=k=-0.6	$\widehat{AUC}_{BL}$ a=k=1.6	$\widehat{AUC}_{BG}$ a=k=1.6	$\widehat{AUC}_{BL}$ a=k=-1.6	$\widehat{AUC}_{BG}$ a=k=-1.6
(25,25)	1.5	0.5	0.8	0.4	0.6518	0.7641	0.3251	0.3352	0.6588	0.6684	0.0961	0.1438	0.8353	0.8693
	1.5	0.5	1.2	1.4	0.6492	0.7629	0.3276	0.3360	0.6574	0.6675	0.1019	0.1446	0.8339	0.8681
	1.5	0.5	0.8	0.4	0.6620	0.7734	0.3271	0.3266	0.6620	0.6757	0.1095	0.1299	0.8459	0.8778
	1.5	0.5	1.2	1.4	0.7641	0.7777	0.3283	0.3199	0.6628	0.6798	0.1144	0.1193	0.8495	0.8800
(50,50)	1.5	0.5	0.8	0.4	0.7482	0.7494	0.3510	0.3414	0.6460	0.6586	0.1584	0.1478	0.8287	0.8523
	1.5	0.5	1.2	1.4	0.8011	0.7423	0.3535	0.3446	0.6437	0.6540	0.1626	0.1507	0.8251	0.8438
	1.5	0.5	0.8	0.4	0.7402	0.7585	0.3468	0.3350	0.6505	0.6652	0.1521	0.1387	0.8364	0.8619
	1.5	0.5	1.2	1.4	0.7529	0.7648	0.3402	0.3301	0.6551	0.6699	0.1387	0.1319	0.8414	0.8681
(100,100)	1.5	0.5	0.8	0.4	0.7063	0.7509	0.3511	0.3408	0.6467	0.6596	0.1597	0.1473	0.8308	0.8541
	1.5	0.5	1.2	1.4	0.8021	0.7466	0.3539	0.3425	0.6448	0.6569	0.1655	0.1485	0.8289	0.8489
	1.5	0.5	0.8	0.4	0.6719	0.7561	0.3478	0.3372	0.6498	0.6633	0.1540	0.1424	0.8356	0.8598
	1.5	0.5	1.2	1.4	0.8078	0.7556	0.3502	0.3360	0.6487	0.6634	0.1597	0.1393	0.8354	0.8583

ML: Maximum Likelihood, BS: Squared Error Loss function, BL: LINEX Loss function, BG: General Entropy Loss function

From Table 1 it is observed that Bayes estimator under LINEX and General Entropy Loss functions tends to underestimate the AUC values when loss parameters value are (0.6,1.6).

Bayes estimation with General Entropy loss function provides the highest AUC values when the loss parameter is -1.6, according to the extension of Jeffreys prior value is 0.4 or 1.4. So that the Bayes estimators of AUC under General Entropy loss function is best estimation method for Constant Shape Bi-Weibull Distribution.

4. Illustration

The real data set namely Coronary Heart Disease [CHDAGE] Data extracted from [29].Data consists of 100 Observations, 3 variables. For this data, we have to find estimated AUC values. This data consists of patients who are Diseased and who are Non-diseased. We have to know the patients with Diseased and patients without Diseased and age is most influential variable for diagnose. Table 2 depicts the Estimated AUC values using CHDAGEData.

Table 2 Estimated AUC values using CHDAGE Data

(m, n)=(43,57)	σ_D=1.5, σ_H=0.5, c=0.4 β=0.8	β=1.2	σ_D=1.5, σ_H=0.5, c=1.4 β=0.8	β=1.2
MLE	0.5542	0.5792	0.5541	0.5793
BS	0.7500	0.7500	0.7502	0.7500
BL(a=k=0.6)	0.3542	0.3543	0.3540	0.3543
BL(a=k=-0.6)	0.6456	0.6456	0.6457	0.6456
BL(a=k=1.6)	0.1675	0.1679	0.1673	0.1678
BL(a=k=-1.6)	0.8318	0.8319	0.8320	0.8320
BG(a=k=0.6)	0.3410	0.3409	0.3408	0.3409
BG(a=k=-0.6)	0.6590	0.6590	0.6592	0.6591
BG(a=k=1.6)	0.1471	0.1470	0.1469	0.1470
BG(a=k=-1.6)	0.8530	0.8529	0.8532	0.8529

From Table 2, we observe that Bayes estimation with General Entropy loss function provides the highest AUC values when the loss parameter is -1.6, according to the extension of Jeffreys prior value is 0.4 or 1.4. So that the Bayes estimators of AUC under General Entropy loss function is best estimation method for Constant Shape Bi-Weibull Distribution using CHDAGE Data.

To demonstrate the proposed methodology with the help of graphical visualization, Figures 1 is drawn for comparing the Estimated AUC values under Constant Shape Bi-Weibull distribution using CHDAGE Data.
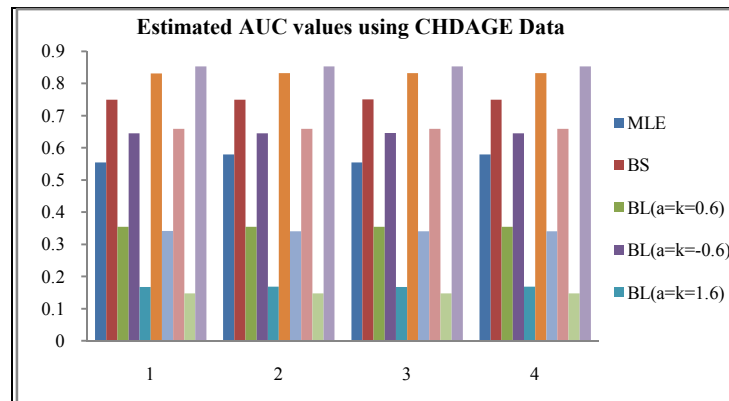

Figure 1 Effect on AUC values comparing different Estimation methods

From Figure 1, it is visualized that Bayes estimation with General Entropy loss function provides the highest AUC values when the loss parameter is -1.6.

5. CONCLUSION

The main objective of this paper is Bayesian estimation of AUC for the Constant Shape Bi-Weibull distribution, under three Loss functions, namely, the Linear Exponential (LINEX) Loss, General Entropy Loss, and Square Error Loss functions. Also,The Maximum Likelihood Estimation of AUC is discussed. Bayes estimators were obtained using Lindley approximation while MLE was obtained using Newton-Raphson method.

A Simulation study was conducted to examine and compare the performance of the estimators for different sample sizes with different values for the extension of Jeffreys' prior and the loss functions.

We also observe that Bayes estimation with General Entropy loss function provides the highest AUC values when the loss parameter is -1.6, according to the extension of Jeffreys prior value is 0.4 or 1.4.

So that the Bayes estimators of AUC under General Entropy loss function is best estimation method for Constant Shape Bi-Weibull Distribution.

References

1. Al-Kutubi, H.S., Ibrahim, N.A., 2009, Bayes estimator for exponential distribution with extension of Jeffery prior information, *Malaysian Journal of Mathematical Sciences***3(2)**, 297–313.
2. Betinec, M., 2008, Testing the difference of the ROC curves in biexponential model, *Tatra Mt. Math. Publ***39**, 215-223.
3. Dorfman., Alf., 1969, Maximum – Likelihood Estimation of parameters of signal detection theory and determination of confidence interval–rating method data, *Journal of Mathematical Psychology***6**, 487-496.
4. Dorfman, D.D., Berbaum, K.S., Metz, C.E., Lenth, R.V., Hanley, J.A., Dagga, H.A., 1996, Proper receiver operating characteristic analysis: the bigamma model, *Acad. Radiol***4**, 138-149.
5. Drury, C.G., Fox, J.G., 1975, Human reliability in quality control, *Halsted*, New York.
6. Gardner, I.A., Greiner, M., 2006, Receiver-operating characteristic curves and likelihood ratios: improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests, *Veterinary Clinical Pathology* **35(1)**, 8-17.
7. Greiner, M.D., Smith, R.D., 2000, Principles and practical application of the receiver operating characteristic analysis for diagnostic tests, *Preventive Veterinary Medicine***45(1–2)**, 23-41. DOI: 10.1016/S0167-5877(00)00115-X.
8. Green, D.M., Swets, J.A., 1966, Signal Detection Theory and Psychophysics, *John Wiley and Sons Inc*, New York, NY, USA.
9. Hanley, J.A., Mc Neil, B.J., 1982, A Meaning and use of the area under a Receiver Operating Characteristics (ROC) Curves, *Radiology***143**, 29-36.
10. Hanley, J.A., 1988, The Robustness of the Binormal Assumption used in fitting ROC curves, *Medical Decision Making***8**,197-203.
11. Hanley, J.A., *et al*., 1989, Receiver Operating Characteristic (ROC) methodology: the state of the art, *Critical Reviews in Diagnostic Imaging***29(3)**, 307-335.

12. Hughes, G., Bhattacharya, B., 2013, Symmetry Properties of Bi-Normal and Bi-Gamma Receiver Operating Characteristic Curves are described by Kullback-Leibler Divergences, *Entropy***15**, 1342-1356.
13. Lavanya, A., Leo Alexander, T., 2016, Confidence Intervals Estimation for ROC Curve, AUC and Brier Score under the Constant Shape Bi-Weibull Distribution, *International Journal of Science and Research* **5(8)**, 371-378.
14. Lavanya, A., Leo Alexander, T., 2016, Asymmetric and Symmetric Properties of Constant Shape Bi-Weibull ROC Curve Described by Kullback-Leibler Divergences, *International Journal of Applied Research* **2(8)**, 713-720.
15. Lavanya, A., Leo Alexander, T., 2016, Bayesian Estimation of Parameters under the Constant Shape Bi-Weibull Distribution Using Extension of Jeffreys' Prior Information with Three Loss Functions, *International Journal of Science and Research***5(9)**, 96-103.
16. Leo Alexander, T., Lavanya, A., 2016, 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.
17. Lusted, L.B., 1971, Signal detectability and medical decision making, *Science* 171,1217-1219.
18. Metz, C.E., 1989, Some practical issues of experimental design and data analysis in radiological ROC studies, *Invest Radiol***24**, 234-245.
19. Metz, C.E., Herman, B.E., Shen, J.H., 1998, Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously distributed data, *Statistics in Medicine***17(9)**, 1033-1053.
20. Pandey, B.N., Dwividi, N., Pulastya, B., 2011, Comparison between bayesian and maximum likelihood estimation of the scale parameter in Weibull distribution with known shape under linex loss function, Journal of Scientific Research**55**, 163–172.
21. Peterson, W.W., Birdsall, T.G., Fox, W.C., 1954, The theory of signal detectability, Transactions of the IRE Professional Group on Information Theory, *PGIT***2(4)**,171-212.
22. Pepe, M.S., 2003, The Statistical Evaluation of Medical Tests for Classification and Prediction, *Oxford University Press*: New York, NY.
23. Seyed Reza Jafarzadeha., Johnsonb, W.O., Uttsb, J.M., Gardnera, I.A., 2010, Bayesian estimation of the receiver operating characteristic curve for a diagnostic test with a limit of detection in the absence of a gold standard, *Statist. Med***29**, 2090—2106.
24. Sinha, S.K., 1986, Bayesian estimation of the reliability function and hazard rate of a Weibull failure time distribution, *Tranbajos De Estadistica***1(2)**, 47–56.
25. Swets, J.A., Egilance., 1977, Relationships among theory, Physiological correlates and operational performance, *Plenum*, New York.
26. Swets, J.A., Pickett, R.M., 1982, Evaluation of Diagnostic Systems: Methods from Signal Detection Theory, *Academic Press*, New York.
27. Zhou, S.H., Obuchowski, N.A., McClish DK., 2002, Statistical Methods in Diagnostic Medicine, New York, NY: John Wiley & Sons, Inc.
28. Zweig, M.H., Campbell, G., 1993, Receiver operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clinical Chemistry***39(4)**, 561--577.
29. https://www.umass.edu/statdata/statdata/data/chdage.txt

****

How to cite this article:

Leo Alexander, T and Lavanya, A.2016, Estimation of Auc for Constant Shape Bi-Weibull Failure Time Distribution. *Int J Recent Sci Res.* 7(10), pp. 13826-13836.