✚IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Standard Error of Area Under One Parameter Bi-exponential ROC Curve

Sudesh Pundir [*], R Amala
[*] Department of Statistics, Pondicherry University, India
sudeshpundir19@gmail.com

Abstract

In this paper, properties of Bi-Exponential Receiver Operating Characteristic (ROC) curve and Area Under the ROC curve (AUC) of Bi-Exponential ROC model are discussed. The evaluation of AUC and its inference is a crucial part of ROC curve analysis. We proposed the estimation of asymptotic and exact variances for AUC of Bi-Exponential ROC curve. The Confidence Intervals for AUC using asymptotic and exact variance methods are also proposed. The proposed methods are validated by extensive simulation study as well as real life example.

Keywords: Area Under the Bi-Exponential ROC Curve, Bi-Exponential ROC model, Kullback-Leibler divergence, and TPR asymmetry. (10pt Times New Roman, Justified).

Introduction

In Medical diagnosis, a subject is categorized into either *healthy* or *diseased* based on some clinical measurement based on the selected cut-off *t*. If the clinical measurement is 'greater than or equal to' *t*, then the subject is labeled as diseased and if the measurement is 'less than' *t*, then the subject is labeled as healthy subjects. The clinical measurements are often called as test results or test scores or Biomarker.

Let us denote the test results of diseased subject by the random variable *Y* with Probability Density Function (PDF), $g_Y(y)$ and Cumulative Distribution Function (CDF), $G_Y(y)$. Similarly, let us denote the test results of healthy subject by the random variable *X* with PDF, $f_X(x)$ and CDF, $F_X(x)$. Assume that *X* and *Y* are independent and continuous.

Sensitivity of the diagnostic test is defined as $\bar{G}_Y(t) = P(Y > t)$, which is the probability of correctly categorizing a diseased subject when a cut-off is given. Similarly, the *Specificity* of the diagnostic test is defined as $F_X(t) = P(X \leq t)$, which is the probability of correctly categorizing a healthy subject.

Receiver Operating Characteristic (ROC) curve is defined as a plot of "*True Positive Rate*" (TPR), $\bar{G}_Y(t)$ on the vertical axis versus the "*False Positive Rate*" (FPR), $\bar{F}_X(t)$ on the horizontal axis for different values of t, where $-\infty < t < \infty$. The mathematical model representing the ROC curve takes the form [8]

$$ROC[t] = \bar{G}_Y \circ \bar{F}_X^{-1}(t); 0 \leq \bar{F}_X(t) \leq 1 \qquad (1)$$

For an appropriate diagnostic test, the ROC curve should lie very close to upper left corner of the unit square. A typical ROC curve must satisfy the following properties.

1. ROC curve is invariant with respect to monotone increasing transformation of the test scores.
2. The test values of *X* are smaller than *Y*.
3. ROC curve is monotonically increasing function i.e. $\dfrac{dROC(t)}{d\bar{F}_X(t)} > 0$.
4. ROC curve is said to be concave, if $\dfrac{d^2 ROC(t)}{d\bar{F}_X^2(t)} < 0$ and convex, if $\dfrac{d^2 ROC(t)}{d\bar{F}_X^2(t)} > 0$.
5. The slope of ROC curve at any operating point is equal to the ratio of PDF of diseased to PDF of healthy at a particular cut-off point [8] which is given by

$$slope = \frac{g(t)}{f(t)} \qquad (2)$$

6. Let $KL(f, g)$ denote the Kullback – Leibler (K-L) divergence [6] between the distributions of healthy and diseased group with f(x) as the

comparison distribution and g(y) as the reference distribution. Then

$$KL(f,g) = \int_D f(x) \ln\left(\frac{f(x)}{g(y)}\right) dx \qquad (3)$$

where $D \in x \cap y;\ 0 < x < \infty;$

$0 < y < \infty$ since the range of x and y are same, we can represent x and y by x itself.

Similarly, let $KL(g,f)$ denote the K-L divergence between the distribution of diseased and healthy population with g(x) as the comparison distribution and f(x) as the reference distribution, then

$$KL(g,f) = \int_D g(x) \ln\left(\frac{g(x)}{f(x)}\right) dx \qquad (4)$$

where D is the common range of f and g. It is to be noted that $KL(f,g)$ and $KL(g,f)$ are positive and $KL(f,g) = KL(g,f) = 0$, if and only if $f(x) = g(x)$. These two measures tell us about the asymmetry of ROC curve about the negative diagonal. If $KL(f,g) < KL(g,f)$, then the ROC curve is said to be TPR asymmetric and if $KL(f,g) > KL(g,f)$, then the curve is said to be True Negative Rate (TNR) asymmetric.

Area under the ROC curve is the frequently used measure for quantifying the performance of the diagnostic test. It is defined as the probability that in a randomly selected pair of healthy and diseased subject, the test result of diseased subject is higher than the healthy subject. Mathematically, it is defined as

$$AUC = \int_0^1 \bar{G}_Y(t) d\bar{F}_X(t). \qquad (5)$$

The evaluation of AUC and its inference is a crucial part of ROC curve analysis. The estimation of ROC curve is done in three ways like Non-Parametric, Parametric and Semi-Parametric. In this paper, we have studied the parametric approach of plotting ROC curve.

In Parametric approach a specific distribution is assumed to X and Y with different parametric values. The conventional Bi-Normal ROC model assumes that X and Y or any monotone transformation of X and Y follow normal distribution with different parameters where the parameters with subscript '0' represents healthy parameters and with subscript '1' represents diseased parameter.

The other distributional models that are considered in literature are Bi-Exponential ROC model [2], Bi-Gamma ROC model [4], Bi-Lomax ROC model [3], Generalized Bi-Exponential ROC model [7], Bi-Lognormal ROC model [1], Bi-Rayleigh ROC model [9], [10], Bi-Weibull ROC model [11] and a review of all parametric ROC models in case of continuous data [12].

This paper is organized as follows: In Section 2, Bi-Exponential ROC model, its properties and ML estimation of parameters are discussed. Section 3, provides estimation of AUC, asymptotic distribution of estimated AUC and confidence interval for $\hat{AUC}$. In Section 4, exact distribution of $\hat{AUC}$ and confidence interval for $\hat{AUC}$ are discussed. The proposed theory is validated by simulation studies and real life example in Section 5. Section 6 contains the concluding remarks.

Materials and methods
Bi-Exponential ROC Model

Let Z be a random variable that follows the one parametric Exponential distribution with inverse scale parameter λ is denoted by Exp (λ). It possess the PDF as

$$f_Z(z,\lambda) = \lambda e^{-\lambda z};\ z > 0, \lambda > 0. \qquad (6)$$

The CDF of random variable Z is given by

$$F_Z(z) = P(Z \le z) = 1 - e^{-\lambda z},\ \lambda > 0 \qquad (7)$$

Bi-Exponential ROC model which is given by [2]

$$eROC(t) = \bar{F}_X(t)^{\frac{\lambda_1}{\lambda_0}};\ 0 \le \bar{F}_X(t) \le 1, \lambda_0 > \lambda_1 \qquad (8)$$

where λ_0 and λ_1 are the parameter of healthy and diseased group respectively, "eROC (t) " represent the Bi-Exponential ROC model.

Now, we will discuss some of the properties of eROC curve

Properties
1. eROC curve is monotonically increasing function.

Proof: A function is said to be a monotone increasing function, if the first derivative of the function is positive. Since, the first derivative of eROC curve with respect to $\bar{F}_X(t)$ is positive.

i.e. $\dfrac{deROC(t)}{d\bar{F}_X(t)} = \dfrac{\lambda_1}{\lambda_0}\left[\bar{F}_X(t)\right]^{\left(\frac{\lambda_1}{\lambda_0}-1\right)} > 0$
$\qquad$ (9)

Hence, eROC curve is monotonically increasing function.

2. eROC curve is concave and never lies below the chance line.

Proof: From equation (9), the second derivative of eROC(t) is given by

$$\frac{d^2 eROC(t)}{d\bar{F}_X^2(t)} = \frac{\lambda_1}{\lambda_0}\left(\frac{\lambda_1}{\lambda_0}-1\right)\left[\bar{F}_X(t)\right]^{\left(\frac{\lambda_1}{\lambda_0}-2\right)} < 0$$
$\qquad$ (10)

The ratio $\dfrac{\lambda_1}{\lambda_0}$ will always be less than zero since

we assumed that $\lambda_0 > \lambda_1$ and hence the term

$\left(\dfrac{\lambda_1}{\lambda_0}-1\right) < 0$ and $\left[\bar{F}_X(t)\right]^{\left(\frac{\lambda_1}{\lambda_0}-2\right)} > 0$ since

$0 \le \bar{F}_X(t) \le 1$. On the whole, we will get $\dfrac{d^2 eROC(t)}{d\bar{F}_X^2(t)} < 0$. Hence, eROC curve is concave in nature. Now, let us prove that it never lies below the chance line.

ROC curve is said to be proper ROC curve if it never crosses the chance line or if the decision variable is a strictly increasing function of the likelihood ratio. Consider any two points *a* and *b* (say) where 0<*a, b*<1 on eROC curve. Since we have proved that the eROC curve is concave, the line segment connecting the point *a* and *b* never lies above the curve. If we take the extreme point i.e. *a = 0* and *b = 1,* it becomes the chance line, so the chance line also never lies above the curve. Hence we have proved that the eROC curve never crosses the chance line.

3. The slope of the eROC curve at the threshold t is given by

$$slope = \frac{\lambda_1}{\lambda_0} e^{\{t(\lambda_0 - \lambda_1)\}}$$
$\qquad$ (11)

4. It is invariant with respect to monotone increasing transformation of the test scores.

5. eROC curve is TPR asymmetric [6].

Proof:

The K-L divergence between two exponential distributions has been studied by [6]. The K-L divergence between the distribution of diseased and healthy group with f(x) as the comparison distribution and g(x) as the reference distribution has been given as

$$KL(f,g) = \frac{\lambda_1}{\lambda_0} - 1 - \ln\left(\frac{\lambda_1}{\lambda_0}\right)$$
$\qquad$ (12)

Similarly, the K-L divergence between the distribution of healthy and diseased group with g(x) as the comparison distribution and f(x) as the reference distribution has been given as

$$KL(g,f) = \frac{\lambda_0}{\lambda_1} - 1 - \ln\left(\frac{\lambda_0}{\lambda_1}\right)$$
$\qquad$ (13)

It was found that $KL(g,f) > KL(f,g).$ These two divergence measures would be zero, if the healthy and diseased group is identical. Hence, we have proved that, the eROC curve is TPR asymmetric.

Asymptotic Distribution of Bi-Exponential $A\hat{U}C$ and Confidence Interval for $A\hat{U}C$

The area under the eROC curve is computed as

$$AUC = P(Y > X) = \frac{\lambda_0}{\lambda_0 + \lambda_1}.$$
$\qquad$ (14)

where λ_0 and λ_1 are the parameters of healthy and diseased group respectively. In order to compute the estimated AUC, we need to find the MLE's of λ_0 and λ_1. In the following section, we will find the MLEs of λ_0 and λ_1.

By substituting the ML estimates $\hat{\lambda}_0$ and $\hat{\lambda}_1$ we get the ML estimate of AUC i.e.

$$A\hat{U}C = \frac{m\sum_{j=1}^{n} y_j}{n\sum_{i=1}^{m} x_i + m\sum_{j=1}^{n} y_j}$$
$\qquad$ (15)

Now, we will derive the asymptotic distribution and confidence interval of AUC. To evaluate the significance of the statistic $A\hat{U}C$, we need to compute its variance and standard error. The following theorem evaluates the variance of the estimate, $A\hat{U}C$.

Theorem 3.1: The area under the eROC curve will converge in distribution to a Normal random variable with mean zero and variance $\dfrac{(m+n)}{mn}\dfrac{\lambda_0^2\lambda_1^2}{(\lambda_1+\lambda_0)^4}$ for large $N(=m+n)$.

Proof:

Let $L(\theta / x, y); \theta = (\lambda_0, \lambda_1)'$ be the likelihood function of the sample observations from X and Y which is given by

$$\ln L = m \ln \lambda_0 - \lambda_0 \sum_{i=1}^{m} x_i + n \ln \lambda_1 - \lambda_1 \sum_{j=1}^{n} y_j \qquad (16)$$

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

$$\Rightarrow \sqrt{N}(\hat{\theta} - \theta) \rightarrow N(0, I^{-1}(\theta)). \qquad (17)$$

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

$$I(\theta) = -\begin{bmatrix} E\left(\dfrac{\partial^2 \ln L}{\partial \lambda_0^2}\right) & E\left(\dfrac{\partial^2 \ln L}{\partial \lambda_0 \partial \lambda_1}\right) \\ E\left(\dfrac{\partial^2 \ln L}{\partial \lambda_1 \partial \lambda_0}\right) & E\left(\dfrac{\partial^2 \ln L}{\partial \lambda_1^2}\right) \end{bmatrix} \qquad (18)$$

where

$$a_{11} = \frac{m}{\lambda_0^2}, \; a_{22} = \frac{n}{\lambda_1^2}, \qquad a_{12} = a_{21} = 0.$$

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

$$I^{-1}(\theta) = \begin{bmatrix} V(\hat{\lambda}_0) & Cov(\hat{\lambda}_0, \hat{\lambda}_1) \\ Cov(\hat{\lambda}_1, \hat{\lambda}_0) & V(\hat{\lambda}_1) \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{1}{a_{11}} & 0 \\ 0 & \dfrac{1}{a_{22}} \end{bmatrix}.$$

$$(19)$$

where

$$V(\hat{\lambda}_0) = \frac{\lambda_0^2}{m}, \; Cov(\hat{\lambda}_0, \hat{\lambda}_1) = 0$$

$$Cov(\hat{\lambda}_1, \hat{\lambda}_0) = 0, V(\hat{\lambda}_1) = \frac{\lambda_1^2}{n}.$$

$$(20)$$

Since area under the ROC curve is a function of parameters $\theta = (\lambda_0, \lambda_1)'$, we will adopt the Delta method for finding the approximate variance. $V(A\hat{U}C)$ can be defined as follows:

$$V(A\hat{U}C) = \left(\frac{\partial AUC}{\partial \lambda_0}\right)^2 V(\hat{\lambda}_0) + \left(\frac{\partial AUC}{\partial \lambda_1}\right)^2 V(\hat{\lambda}_1) +$$

$$2\left(\frac{\partial AUC}{\partial \lambda_0}\right)\left(\frac{\partial AUC}{\partial \lambda_1}\right) Cov(\hat{\lambda}_0, \hat{\lambda}_1).$$

$$= \frac{\lambda_0^2 \lambda_1^2}{(\lambda_0 + \lambda_1)^4}\left(\frac{m+n}{mn}\right) \qquad (21)$$

where $V(\hat{\lambda}_0)$, $V(\hat{\lambda}_1)$ and $Cov(\hat{\lambda}_0, \hat{\lambda}_1)$ and $Cov(\hat{\lambda}_1, \hat{\lambda}_0)$ are taken from the equation (20). The estimate of variance is obtained by substituting the estimates of the parameters λ_0, λ_1.

The standard error of $A\hat{U}C$ can be obtained by taking square root of $V(A\hat{U}C)$ in equation (21). The $100(1-\alpha)$ % confidence interval is obtained by

$$[A\hat{U}C \pm SE(A\hat{U}C)Z_{\frac{\alpha}{2}}] \qquad (22)$$

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

It is observed that the asymptotic confidence interval do not perform well for small sample sizes. Hence, we propose the exact distribution of estimated AUC and confidence interval for small sample size in the following section.

Exact Confidence interval for $A\hat{U}C$

From Section 3, the ML estimate of AUC was found to be

$$A\hat{U}C = \frac{\hat{\lambda}_0}{\hat{\lambda}_0 + \hat{\lambda}_0} = \frac{(m/U)}{(m/U) + (n/V)} \qquad (23)$$

where

$$\hat{\lambda}_0 = \frac{m}{\sum_{i=1}^{m} x_i} = \frac{m}{U} \text{ and } \hat{\lambda}_0 = \frac{n}{\sum_{j=1}^{n} y_j} = \frac{n}{V}.$$

$$(24)$$

It is well known that, if $X \sim Exp(\lambda_0)$ then U follows gamma distribution with shape parameter *m* and scale parameter λ_0 and similarly *V* follows gamma distribution with shape parameter *n* and scale parameter λ_1. Hence, $2\lambda_0 V$ and $2\lambda_1 U$ are independently distributed as χ^2 with 2m and 2n degrees of freedom.

Lemma 1: Let δ_1 and δ_2 be independent χ^2 random variables with m and n degrees of freedom respectively, the random variable [13]

$$F = \frac{\delta_1 / m}{\delta_2 / n} \qquad (25)$$

is said to have an F- distribution with (m, n) degrees of freedom. From the Lemma, it follows that,

$$F = \frac{(2\lambda_1 U / 2m)}{(2\lambda_0 V / 2n)} = \frac{(\lambda_1 U / m)}{(\lambda_0 V / n)} \sim F(2m, 2n) \qquad (26)$$

The 100(1-α) % confidence interval for AUC can be obtained from F distribution through $P(F_l < F < F_u)$ where F_l and F_u satisfies the condition

where F_l and F_u satisfies the condition
$P(F_l < F < F_u) = 1 - \alpha.$

After a simple manipulation, the 100 (1-α)% confidence interval is found to be

$$\left[\frac{A\hat{U}C\, F_{(2m,\, 2n,\, \alpha/2)}}{A\hat{U}C\, F_{(2m,\, 2n,\, \alpha/2)} + (1 - A\hat{U}C)},\; \frac{A\hat{U}C\, F_{(2m,\, 2n,\, 1-\alpha/2)}}{A\hat{U}C\, F_{(2m,\, 2n,\, 1-\alpha/2)} + (1 - A\hat{U}C)} \right]$$

$$(27)$$

where $F_{2m,\, 2n,\alpha/2}$ and $F_{2m,\, 2n,1-\alpha/2}$ are the lower and upper $\alpha/2^{th}$ percentile points of F distribution with 2m and 2n degrees of freedom.

Results and discussion
Simulation Studies

In this section, we provide the results of estimation of asymptotic, exact variance and confidence interval for AUC using simulated datasets.

Asymptotic Variance Method

In this section, we did simulation studies to observe how the asymptotic variance of AUC behaves using simulated data sets. We have considered four different samples of size (m, n) = (30, 30) with different parametric values for λ_0 = {0.0829, 0.1456, 0.2035, 0.4167} and λ_1= {0.0466, 0.0423, 0.0267, 0.0311}. The estimated parameters, $A\hat{U}C$, $SE(A\hat{U}C)$, 95% Confidence Interval for $A\hat{U}C$ are shown in Table 1.

From the Table 1, it is observed that, as the accuracy increases and the standard error decreases simultaneously. The ROC curves for different parametric values are plotted in Figure 1. In table 2, we have simulated independent samples of *m* healthy and *n* diseased (*m = n* = 5, 10, 30, 60, 80) to assess the behavior of asymptotic variance and confidence interval by fixing λ_0 and by varying λ_1. Similarly, in Table 3, we have simulated independent samples of *m* healthy and *n* diseased of size (*m = n* = 5, 10, 30, 60, 80) by fixing λ_1 and varying λ_0. In each row, the first element represents the test's $A\hat{U}C$, second element (bolded) represents the $SE(A\hat{U}C)$, third and

fourth element represent the lower and upper confidence limit respectively. Figure 2 represents, the pattern of standard error with respect to the sample size by keeping λ_0 constant as in Table 2 and Figure 3 represents the pattern of standard error with respect to the sample size by keeping λ_1 constant as in Table 3 by using asymptotic variance method.

From Table 2 and 3, it is observed that $SE(A\hat{U}C)$, decreases with increase in sample size and increase in accuracy. The behavior is depicted in Figure 2 and 3. It is observed that the asymptotic MLE do not perform well for small sample size. If we observe the lower confidence limit, it is crossing the lowest accuracy 0.5 which is not being considered as a good AUC estimate. As we mentioned earlier, if the curve reaches upper left corner of unit square then AUC is close to one. The topmost curve corresponds to the estimates $(\hat{\lambda}_0, \hat{\lambda}_1)$ = (0.4001, 0.0305), the curve just below the topmost curve corresponds to (0.2134, 0.0238), the third curve from top corresponds to (0.1054, 0.0359) and the lowest curve corresponds to the estimates (0.0833, 0.0447).

Figure 1:
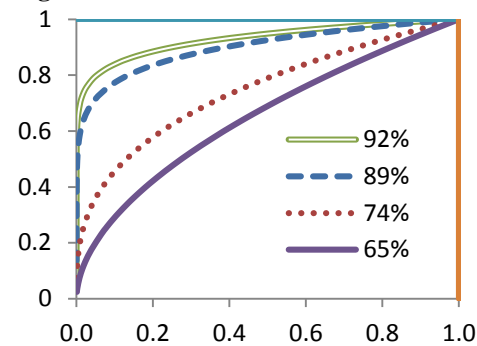
*Bi-Exponential ROC curve for different parametric values*

Exact Variance Method

In this section, we used the same parameters and estimated value to observe how the exact variance method behaves. The estimated parameters, $A\hat{U}C$, $SE(A\hat{U}C)$, 95% Confidence Interval for $A\hat{U}C$, are shown in Table 4.

In Table 5 and 6, the AUC and SE are analyzed by fixing one parametric value and varying the other and vice versa for different sample sizes and results are presented.

By comparing Table 2 with Table 5 and Table 3 with Table 6, we observe that there is a considerable difference in the two method adopted. The exact variance method possesses a shorter confidence
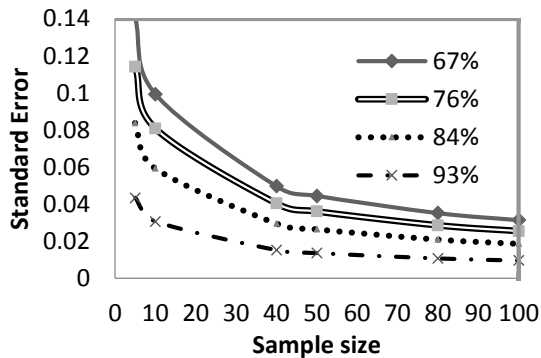
interval when compared with asymptotic variance method.

Figure 2:
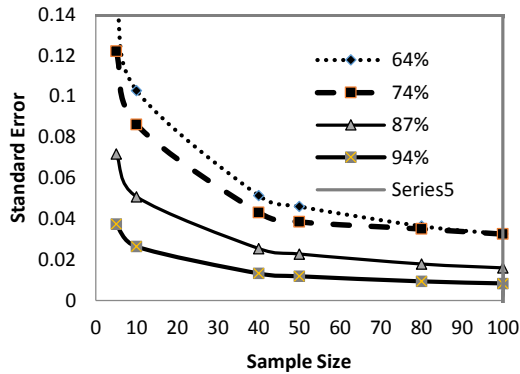
Standard error versus sample size

Figure 3:

Standard error versus sample size

Real Life Example

Biomarkers are extremely useful for detecting the disease at early stages. Prostate Specific Antigen (PSA) is a biomarker which is significant in detecting the prostate cancer. For the application of proposed model, we consider the biomarker called total serum PSA. The data has been obtained from a Phase 3 prostate cancer Case-Control study nested in the Beta Carotene [6]. The data used in this paper consisted randomly selected individuals who are affected by prostate cancer and 50 healthy individuals who were participated in a lung cancer prevention trial.

First of all, let us evaluate the goodness of fit test for one parameter exponential distribution using Kolmogorov-Smirnov, Anderson-Darling and Chi-Square statistics. The statistic and p-value for all the three procedures is shown in Table 7.

For healthy marker values, the probabilities of the modeled K-S (0.1342), A-D (0.9977) and Chi-Square (6.2572) was greater than or equal to the level of significance of 20% (0.2000). Hence, there is no

evident that the null hypothesis "The healthy marker values fits the one parameter Exponential distribution" to be rejected. Similarly in diseased group, the p-values of the modeled statistics were greater than the level of significance of 10% (0.1000). Hence, we conclude that the diseased marker values fit the assumed Exponential distribution.

The estimated parameters are $\hat{\lambda}_0 = 2.3086$ and $\hat{\lambda}_1 = 0.6479$. Hence the AUC, standard errors using asymptotic MLE and exact variance methods are estimated as 0.7799, 0.034498 and 0.034477 respectively. The 95% asymptotic and exact confidence intervals for estimated AUC are constructed as [0.7123, 0.8476] and [0.7133, 0.8484] respectively. From the result, it is obvious that the asymptotic MLE method and exact variance method both performs in the similar way. The sensitivity and specificity of the marker are found to be 71% and 71% respectively at the threshold value of 0.534. From the sensitivity and specificity rates, we infer that an individual whose is having the "ratio of free to total PSA" marker value greater than 0.534 are 71% likely to be detected with the prostate cancer. Similarly, an individual having the marker value less than 0.534 are 71% likely to be not having the prostate cancer. The ROC curve plotted is plotted for the PSA marker and the asymmetric property of the ROC curve is also studied which is shown in Figure 6.

The line segment connecting (0,1) and (1,0) is called the negative chace line and it is obtained by plotting FPR on X-axis and 1-FPR on Y-axis. The dashed vertical line segment let us call it as S_1 corresponds to the co-ordinate [FPR=a (0.08, say), 0 ≤ TPR ≤ 1]. The dashed horizontal line (S_2) segment corresponds to the co-ordinate [0 ≤ FPR ≤ 1, TPR=1-a (0.95)]. Let A=[a,0.5], B=[0.5, 1-a] and C=[a*>a, 1-a*]. A ROC curve is said to be symmetric if it passes through the co-ordinate A, B and C. Any ROC curve is said be TPR asymmetric if it passess through S2 after the co-ordinate B and the one that passes though S2 before the co-ordinate B is called TNP asymmetric. Thus from Figure 6, it is proved that the Bi-Exponential ROC curve is TPR asymmetric graphically. From equation (3) and (4), the K-L divergence measures

$KL(g, f)(1.2927) > KL(f, g)(0.5514)$

Hence, the eROC curve is TPR asymmetry is proved numerically.

Conclusion

Some of the properties of the Bi-Exponential ROC model have been studied. It was found that Bi-Exponential ROC curve satisfies some common properties like monotonicity and invariance property. It also satisfies some specific properties like concavity and TPR asymmetric. We have developed two methods to estimate the variance of Bi-Exponential AUC namely asymptotic variance method and exact variance method. The 100(1-α) % confidence interval for AUC using both method have been computed. Through extensive simulation studies and by looking at the real life example, we found that exact variance methods perform slightly better than the asymptotic variance method. Each of the method has to be adopted according to the specific conditions i.e. asymptotic variance can be adopted if the sample size is large else one can go for exact variance.

Acknowledgements

References:

[1] 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, Vol. 1(2), 283-295.

[2] Betinec, M. (2008). Testing the difference of the ROC Curves in Biexponential model, Tatra Mountains Mathematical Publications, Vol. 39, 215-223.

[3] Campbell, G. and Ratnaparkhi, M.V. (1993). An application of Lomax distributions in receiver operating characteristic (roc) curve analysis, Communication in Statistics, Vol. 22(6), 1681–1687.

[4] Dorfman, D.D., Berbaum, K.S., Metz, C.E., Lenth, R.V., Hanley, J.A. and Dagga, H.A. (1996). Proper Receiver Operating Characteristics Analysis: The bigamma model, ademic Radiology, Vol. 4, 138-149.

[5] Etzioni, R. Pepe, M., Longton, G., Hu, C. and Goodman, G. (1999). Incorporating the time dimension in receiver operating characteristic curves: A case study of prostate cancer, Medical Decision Making, Vol. 19(3), 242-251.

[6] Hughes, G. and Bhattacharya, B. (2013). Symmetry Properties of Bi-Normal and Bi-Gamma Receiver Operating Curves are Described by Kullback-Leibler Divergences, Entropy, Vol. 15, 1342-1356.

[7] Hussain, E. (2011). The ROC Curve Model from Generalized-Exponential Distribution, Pakistan Journal of Statistics and Operations Research, Vol. 7(2), 323-330.

[8] Krzanowski, W.J. and Hand, D.J. (2009). ROC curves for continuous data, Monographs on Statistics and Applied Probability, CRC Press, Taylor and Francis Group, NY.

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

[10] Pundir, S. and Amala, R. (2012: b). A study on the comparison of Bi-Rayleigh ROC model with Bi-Gamma ROC model, Edited volume, Application of Reliability Theory and Survival Analysis, Bonfring Publication 196-209.

[11] Pundir, S. and Amala, R. (2014). Evaluation of Area Under the Constant Shape Bi-Weibull ROC Curve, International Journal of Modern and Applied Statistical Methods, accepted for publication.

[12] Pundir, S. and Amala, R. (2014). Parametric Receiver Operating Characteristic Modeling for continuous data: A Glance, Model Assisted Statistics and Application, $9(2)$, $121\text{-}135$.

[13] Rohatgi, V.K. and Saleh, E.A.K. (2001). An introduction to probability and statistics, Second edition, John Wiley and Sons, New Delhi.

Table 1 $A\hat{U}C$, $SE(A\hat{U}C)$ and 95% confidence interval for $A\hat{U}C$ based on eROC through asymptotic variance method

S.N.	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$A\hat{U}C$	$SE(A\hat{U}C)$	95% Confidence Interval
1	0.0833	0.0447	0.6508	0.0587	[0.5358, 0.7658]
2	0.1054	0.0359	0.7458	0.0489	[0.6499, 0.8418]
3	0.2134	0.0238	0.8996	0.0233	[0.8540, 0.9453]
4	0.4001	0.0305	0.9292	0.0170	[0.8959, 0.9645]

Table 2 $A\hat{U}C$, $SE(A\hat{U}C)$, 95% confidence interval for $A\hat{U}C$ by fixing λ_0=0.0829 and varying λ_1

Sample Size $\hat{\lambda}_1$	(5,5)	(10,10)	(30,30)	(60,60)	(80,80)
0.0466	0.6402 **0.1457** [0.3546, 0.9257]	0.6402 **0.1030** [0.4382, 0.8421]	0.6402 **0.0595** [0.5236, 0.7567]	0.6402 **0.0421** [0.5577, 0.7226]	0.6402 **0.0364** [0.5688, 0.7115]
0.0293	0.7389 **0.1220** [0.4997 0.9780]	0.7389 **0.0863** [0.5697 0.9080]	0.7389 **0.0498** [0.6412 0.8365]	0.7389 **0.0352** [0.6698 0.8079]	0.7389 **0.0305** [0.6791 0.7987]
0.0125	0.8689 **0.0720** [0.7278 1.0000]	0.8689 **0.0509** [0.7692 0.9688]	0.8689 **0.0294** [0.8114 0.9266]	0.8689 **0.0208** [0.8282 0.9097]	0.8689 **0.0180** [0.8337 0.9043]
0.0056	0.9367 **0.0375** [0.8632 1.000]	0.9367 **0.0265** [0.8848 0.9887]	0.9367 **0.0153** [0.9067 0.9667]	0.9367 **0.0108** [0.9155 0.9579]	0.9367 **0.0094** [0.9184 0.9551]

Table 3 $A\hat{U}C$, $SE(A\hat{U}C)$, 95% confidence interval for $A\hat{U}C$ by fixing λ_1 = 0.0466 and varying λ_0

Sample Size $\hat{\lambda}_0$	(5,5)	(10,10)	(30,30)	(60,60)	(80,80)
0.0829	0.6402 **0.1457** [0.3546 0.9257]	0.6402 **0.1030** [0.4382 0.8421]	0.6402 **0.0595** [0.5236 0.7567]	0.6402 **0.0421** [0.5577 0.7226]	0.6402 **0.0364** [0.5688 0.7115]
0.1278	0.7328 **0.1238** [0.4901 0.9755]	0.7328 **0.0876** [0.5612 0.9044]	0.7328 **0.0506** [0.6337 0.8319]	0.7328 **0.0358** [0.6627 0.8029]	0.7328 **0.0310** [0.6721 0.7935]
0.2056	0.8152 **0.0953** [0.6285 1.0000]	0.8152 **0.0674** [0.6832 0.9473]	0.8152 **0.0389** [0.7390 0.8915]	0.8152 **0.0275** [0.7613 0.8691]	0.8152 **0.0238** [0.7685 0.8619]
0.5562	0.9227 **0.0451** [0.8343 1.0000]	0.9227 **0.0319** [0.8602 0.9852]	0.9227 **0.0184** [0.8866 0.9588]	0.9227 **0.0130** [0.8972 0.9482]	0.9227 **0.0113** [0.9006 0.9448]

Table 4 $A\hat{U}C$, $SE(A\hat{U}C)$, 95% confidence interval for $A\hat{U}C$

based on Bi-Exponential ROC through exact variance method

S.N.	$\hat{\lambda}_0$	$\hat{\lambda}_1$	$A\hat{U}C$	95% Confidence Interval
1	0.0833	0.0447	0.6508	[0.5279, 0.7545]
2	0.1054	0.0359	0.7458	[0.6377, 0.8302]
3	0.2134	0.0238	0.8996	[0.8432, 0.9373]
4	0.4001	0.0305	0.9292	[0.8873, 0.9563]

Table 5 $A\hat{U}C$, $SE(A\hat{U}C)$, 95% confidence interval for $A\hat{U}C$ by fixing λ_0 and varying λ_1

Sample Size $\hat{\lambda}_1$	(5,5)	(10,10)	(40,40)	(50,50)	(80,80)
0.0466	0.6402 [0.3237 0.8686]	0.6402 [0.4192 0.8143]	0.6402 [0.5163 0.7478]	0.6402 [0.5539 0.7182]	0.6402 [0.5838 0.6929]
0.0293	0.7389 [0.4322 0.9132]	0.7389 [0.5345 0.8746]	0.7389 [0.6293 0.8251]	0.7389 [0.6638 0.8021]	0.7389 [0.6905 0.7821]
0.0125	0.8689 [0.6409 0.9822]	0.8689 [0.7291 0.9420]	0.8689 [0.7992 0.9170]	0.8689 [0.8223 0.9048]	0.8689 [0.8395 0.8938]
0.0056	0.9367 [0.7990 0.9820]	0.9367 [0.8573 0.9733]	0.9367 [0.8988 0.9611]	0.9367 [0.9118 0.9550]	0.9367 [0.9212 0.9494]

Table 6 AÛC , SE(AÛC) , 95% confidence interval for AÛC by fixing λ_1 = 0.0466 and varying λ_0

Sample Size $\hat{\lambda}_0$	(5,5)	(10,10)	(30,30)	(60,60)	(80,80)
0.0829	0.6402 [0.3237 0.8686]	0.6402 [0.4190 0.8143]	0.6402 [0.5063 0.7478]	0.6402 [0.5539 0.7182]	0.6402 [0.5838 0.6929]
0.1278	0.7328 [0.4246 0.9107]	0.7328 [0.5267 0.8711]	0.7328 [0.6220 0.8205]	0.7328 [0.6569 0.7971]	0.7328 [0.6838 0.7767]
0.2056	0.8152 [0.5428 0.9425]	0.8152 [0.6416 0.9158]	0.8152 [0.7258 0.8803]	0.8152 [0.7549 0.8634]	0.8152 [0.7767 0.8484]
0.5562	0.9227 [0.7654 0.9780]	0.9227 [0.8289 0.9671]	0.9227 [0.8775 0.9520]	0.9227 [0.8928 0.9448]	0.9227 [0.9039 0.9380]

Table 7 Results of Goodness of Fit test

	Test	Statistic	P-value	Rank	α %
Healthy	Kolmogorov-Smirnov	0.1342	0.3016	35	20, 10, 5, 2, 1
	χ^2	6.2572	0.2820	36	20, 10, 5, 2, 1
	Anderson- Darling	0.9977	-	29	20, 10, 5, 2, 1
Diseased	Kolmogorov-Smirnov	0.1422	0.2404	28	20, 10, 5, 2, 1
	χ^2	6.3019	0.2779	23	20, 10, 5, 2, 1
	Anderson- Darling	1.833	-	28	10, 5, 2, 1

Author Biblography

Sudesh Pundir
Dr. Sudesh Pundir is currently working as an Assistant Professor in the Department of Statistics, Pondicherry University, Puducherry. Her areas of research are Biostatistics, Applied Statistics, Reliability. She has a no. of published research papers in reputed journals. She has also participated in many conferences in India as well as abroad. She has organized one International Conference and acted as an organizing committee member in many conferences. She has presented many research papers and gave invited talks, special invited talks and acted as a resource person in various National and International Conferences/Workshops. She is a life member of ISPS and IISA.

Email: sudeshpundir19@gmail.com

R Amala
Amala. R is presently pursuing her Ph.D. in Statistics under the guidance of Dr. Sudesh Pundir, Department of Statistics, Pondicherry University, Puducherry. Thrust area of her research is Applied Statistics. She has presented seven research papers, She has participated in eight National and international conferences. She has published eight research paper in an international Journal.

Email: amalar.statistics@gmail.com