Journal of Basic and Applied Engineering Research
p-ISSN: 2350-0077; e-ISSN: 2350-0255; Volume 4, Issue 2; January-March, 2017, pp. 176-179

http://www.krishisanskriti.org/Publication.html

Estimation of ROC Curve and AUC for a Right-Truncated Sample from an Exponential Distribution

R. Amala[1] and G. Kiruthika[2]

[1]Department of Statistics S.D.N.B Vaishnav College for Women Chennai, India
[2]Department of Statistics Madras Christian College Chennai, India
E-mail: [1]amalar.statistics@gmail.com, [2]kiruthikagurunathan@ymail.com

Abstract—*Assessment of biomarker forms the basis for any clinical diagnosis. Evaluation of its diagnostic value enables us to discriminate the subjects who are susceptible to any particular disorder. This can be carried out with the help of Receiver Operating Characteristic (ROC) curve. In many situations, the biomarker value for some of the subjects cannot be measured due to some technical problem. Discarding those information as such and application of classical ROC procedure leads to biased estimation of accuracy. This study deals with modeling a parametric ROC curve for the right-truncated sample that follow exponential distribution. The ROC model, Area under the right-truncated Exponential Curve (AUC), the asymptotic variance and confidence interval for estimated AUC have been discussed.*

Keywords*: ROC, AUC, right-truncated Exponential distribution (RTE), approximate MLE.*

1. INTRODUCTION

Let Z be a right truncated exponential random variable i.e. $Z \sim Exp(b,\lambda)$ where λ is the scale parameter and b is a known constant above which the values are truncated. The PDF of Z is given by

$$f(z,\lambda,b)=\begin{cases}\dfrac{\lambda e^{-\frac{z}{\lambda}}}{1-e^{-\frac{b}{\lambda}}}; & 0<z\le b,\ \lambda>0\\ 0 & else\end{cases} \qquad (1.1)$$

The mean and variance of Z are given by

$$E(Z)=\lambda-b\left(e^{\frac{b}{\lambda}}-1\right)^{-1} \qquad (1.2)$$

and

$$V(Z)=\lambda^2-b^2e^{\frac{b}{\lambda}}\left(e^{\frac{b}{\lambda}}-1\right)^{-2} \qquad (1.3)$$

respectively. The cumulative distribution function is given by

$$F(Z)=\left(1-e^{\frac{-z}{\lambda}}\right)\left(1-e^{\frac{-b}{\lambda}}\right)^{-1} \qquad (1.4)$$

In practice, the exponential distribution has predominant application in the field of survival and reliability engineering to study the lifetime of manufactured items. In the last two decades, exponential distribution also finds its application in assessing the biomarker in clinical diagnosis. For example, Pundir and Amala [6] provided a methodology for evaluating the discriminatory power of continuous biomarker using exponential distribution (one parameter). In this paper, they have derived Bi-exponential ROC model which is able to evaluate the biomarker graphically and its summary measure AUC depicts the accuracy numerically. They have also measured the standard error of the estimate of AUC and its confidence interval under the assumption of asymptotic normality.

Many authors have investigated the application of exponential distribution in ROC analysis, in a diverse perspective, including Hussain [3], Pundir and Amala [6], where the former used a generalized exponential distribution and the latter used two parameter exponential distribution for evaluating the biomarkers that follows the under-taken distribution. Other works based on exponential distribution includes Betinec[1],Vardhan *et al.* [2] and Hussain[3] .

In many situations, the biomarker value for some of the subjects cannot be measured due to some technical problem. Discarding that information as such and application of classical ROC procedure leads to biased estimation of accuracy. This study deals with modelling a parametric ROC

curve for the right-truncated sample when the sample data follow exponential distribution.

In Section 2, Right -Truncated Bi-Exponential (RTE_2) ROC model and the expression of AUC of RTE_2 are developed. In Section 3, the maximum likelihood estimation of parameters of AUC of RTE_2 ROC curve is discussed. In Section 4, the RTE_2 ROC model is being applied to simulated data set and the conclusions are drawn.

2. RIGHT TRUNCATED BI-EXPONENTIAL (RTE_2) ROC MODEL AND ITS AUC

Let $X_i \sim RTE(\lambda_x)$; i = 1,2,3,…,m and $Y_j \sim RTE(\lambda_y)$; j = 1,2,…,n be independently distributed continuous biomarkers. The PDF of *X* and *Y* are given by

$$f_X(x) = \frac{1}{\lambda_x}\frac{\exp\left(\frac{-x}{\lambda_x}\right)}{1-\exp\left(\frac{-b_x}{\lambda_x}\right)} \;;\quad x \le b_x, \lambda_x > 0 \qquad (2.1)$$

and

$$f_Y(y) = \frac{1}{\lambda_y}\frac{\exp\left(\frac{-y}{\lambda_y}\right)}{1-\exp\left(\frac{-b_y}{\lambda_y}\right)} \;;\; y \le b_y, \lambda_y > 0 \qquad (2.2)$$

respectively, where λ_x and λ_y are the parameters of right truncated exponential distribution of normal and abnormal population respectively, b_x and b_y are known constants of *X* and *Y*. The sensitivity and 1-specificity, at the threshold *t* are obtained as

$$\overline{G}_Y(t) = \int_t^{\infty}\frac{\left(\frac{1}{\lambda_y}\right)\exp\left(\frac{-y}{\lambda_y}\right)}{1-\exp\left(\frac{-b_y}{\lambda_y}\right)}$$

$$= \frac{e^{\frac{-t}{\lambda_y}}}{1-\exp\left(\frac{-b_y}{\lambda_y}\right)}; y \le b_y, \lambda_y > 0 \qquad (2.3)$$

$$\overline{F}_X(t) = \int_t^{\infty}\left(\frac{1}{\lambda_x}\right)\exp\left(\frac{-x}{\lambda_x}\right)$$

$$= \frac{e^{\frac{-t}{\lambda_x}}}{1-\exp\left(\frac{-b_x}{\lambda_x}\right)} \;;\; x \le b_x, \lambda_x > 0 \qquad (2.4)$$

On plotting 1- specificity on the x-axis and sensitivity on the y-axis for various values of t, one can obtain the explicit function of RTE_2 ROC model as follows

$$RTE_2(t) = \frac{\left[F(b_x)\overline{F}_x(t)\right]^{\frac{\lambda_x}{\lambda_y}}}{F(b_y)} \;;\; 0 \le \overline{F}_x(t) \le 1 \qquad (2.5)$$

where $F(b_x) = 1-\exp\left(\frac{-b_x}{\lambda_x}\right)$ and

$$F(b_y) = 1-\exp\left(\frac{-b_y}{\lambda_y}\right) \qquad (2.6)$$

On plotting RTE_2 (t) on the y-axis and on x-axis for various values of t, we can also get the RTE_2 ROC curve.
The area under the RTE_2 ROC curve is obtained as follows

$$AUC = P(Y > X)$$

$$= \int_0^{b_x}\int_x^{b_y}\frac{1}{\lambda_x}\frac{1}{\lambda_y}\frac{\exp\left[-\left(\frac{x}{\lambda_x}+\frac{y}{\lambda_y}\right)\right]}{F(b_x)F(b_y)}dydx$$

$$AUC = \frac{-e^{-b_y/\lambda_y}}{F(b_y)} + \frac{\left(1-\exp\left[-b_x\left(\frac{1}{\lambda_x}+\frac{1}{\lambda_y}\right)\right]\right)}{F(b_x)F(b_y)}\frac{\lambda_y}{\lambda_x+\lambda_y} \qquad (2.7)$$

3. MAXIMUM LIKELIHOOD ESTIMATION OF AUC OF RIGHT-TRUNCATED EXPONENTIAL ROC CURVE

Let *X* and *Y* be the random variables representing the test result or biomarker value of normal and abnormal individual of size *m* and *n* respectively. Assuming that *X* and *Y* are continuous and independent random variables from exponential distribution.

Let $X_{(1)}, X_{(2)}, ..., X_{(r)}, X_{(r+1)}, X_{(r+2)}, ..., X_{(m)}$ be the ordered set of the complete observation from normal exponential population. Let us assume that the last (m-r) ordered statistics are truncated. Similarly let

$Y_{(1)}, Y_{(2)},..., Y_{(s)}, Y_{(s+1)}, Y_{(s+2)},..., Y_{(n)}$ be the ordered set of observations from abnormal exponential population and (*n-s*) observations are truncated.

Then the joint PDF of $X_{(1)}, X_{(2)},..., X_{(m)}$ and $Y_{(1)}, Y_{(2)},..., Y_{(n)}$ is given by

$$f_{(i,j)}(x,y)=\frac{k}{\lambda_x\lambda_y}\left[F(b_x)\right]^{m-r+1}\left[1-\frac{1}{F(b_x)}e^{\frac{-x}{\lambda_x}}\right]^{r-1}$$
$$\exp\left[-(m-r+1)\frac{x}{\lambda_x}\right]\left[F(b_y)\right]^{n-s+1}\left[1-\frac{1}{F(b_y)}e^{\frac{-y}{\lambda_y}}\right]^{s-1}$$
$$\exp\left[-(n-s+1)\frac{y}{\lambda_y}\right]$$

$$i=1,\ 2,\ ...,r\ ;\quad 0<x<b_x$$
$$j=1,\ 2,...,\ s\ ;\quad 0<y<b_y \qquad (3.1)$$

where $k=\frac{m!n!}{(r-1)!(s-1)!(m-r)!(n-s)!}$

$F(b_x)$ and $F(b_y)$ are defined in equation (2.6).

where b_x and b_y are the point of truncation of X and Y respectively, and $F(b_x)$ and $F(b_y)$ are the CDF of X and Y at b_x and b_y .

In this case the log-likelihood function of equation (3.1) is given by

$$\ln L=-m\ln\lambda_x-\frac{1}{\lambda_x}\sum_{i=1}^{m}x_i-m\ln F(b_x)-n\ln\lambda_y-\frac{1}{\lambda_y}\sum_{j=1}^{n}y_j-n\ln F(b_y) \qquad (3.2)$$

Differentiating (2.7) with respect to λ_x and λ_y we get,

$$\frac{\partial\ln L}{\partial\lambda_x}=\frac{-m}{\lambda_x}+\frac{1}{\lambda_x^{\,2}}\sum_i(x_i)+\frac{mb_x}{\lambda_x^{\,2}}\frac{e^{\frac{-b_x}{\lambda_x}}}{1-e^{\frac{-b_x}{\lambda_x}}} \qquad (3.3)$$
$$=\frac{-m}{\lambda_x}+\frac{1}{\lambda_x^{\,2}}\sum_i(x_i)+\frac{mb_x}{\lambda_x^{\,2}}\frac{f(z)}{F(Z)}$$

where $z=\frac{b_x}{\lambda_x}$, $f(z)$ and $F(z)$ are the PDF and CDF of standard exponential variable given by $f(z)=e^{-z}$ and $F(z)=1-e^{-z}$

Since, it is not possible to express the parameter as an explicit function of *x* while maximizing the log-likelihood, we use approximate MLE by using Taylors, series approximation, [for more details of approximate MLE of parameters refer Kang and Cho [4] and Wu *et al.*[5]

$$\hat{\lambda}_x=\frac{B_x-\sqrt{B_x^{\,2}-m^2\beta_x b_x^{\,2}}}{2m} \qquad (3.4)$$

where $B_x=\sum_{i=1}^{n}x_i+mb_x\alpha_x$

$$\beta_x=\frac{f\left(\frac{b_x}{2}\right)F\left(\frac{b_x}{2}\right)+f^2\left(\frac{b_x}{2}\right)}{\left[F^2\left(\frac{b_x}{2}\right)\right]}\ \&$$

$$\alpha_x=\frac{f\left(\frac{b_x}{2}\right)}{F\left(\frac{b_x}{2}\right)}+\frac{f\left(\frac{b_x}{2}\right)F\left(\frac{b_x}{2}\right)+f^2\left(\frac{b_x}{2}\right)}{\left[F\left(\frac{b_x}{2}\right)\right]^2}\left(\frac{b_x}{2}\right)$$

Similarly, one can derive $\hat{\lambda}_y$ as follows

$$\hat{\lambda}_y=\frac{B_y-\sqrt{B_y^{\,2}-n^2\beta_y b_y^{\,2}}}{2n} \qquad (3.5)$$

where $B_y=\sum_{j=1}^{n}y_j+nb_y\alpha_y$

$$\beta_y=\frac{f\left(\frac{b_y}{2}\right)F\left(\frac{b_y}{2}\right)+f^2\left(\frac{b_y}{2}\right)}{\left[F^2\left(\frac{b_y}{2}\right)\right]}\ \text{and}$$

$$\alpha_y=\frac{f\left(\frac{b_y}{2}\right)}{F\left(\frac{b_y}{2}\right)}+\frac{f\left(\frac{b_y}{2}\right)F\left(\frac{b_y}{2}\right)+f^2\left(\frac{b_y}{2}\right)}{F^2\left(\frac{b_y}{2}\right)}\left(\frac{b_y}{2}\right)$$

4. SIMULATION STUDIES

In this section, we have applied the proposed RTE_2 ROC model to a simulated data from right-truncated exponential distribution. the data have been simulated from exponential distribution for different values of parameters as mentioned in the first column of Table 4.1, then the samples are truncated at (r, s) = (2, 2) and with the truncated samples the parameters and AUC are estimated using equation (3.4), (3.5) and (2.7). The following table presents the details of simulation using assumed parametric value, estimated parametric of RTE_2 distribution. The data are also checked for goodness of fit test of right truncated exponential distribution using Kolmogrov-Smirnov statistic and the results are also presented in the Table 4.1

Table 4.1: Assumed parameters, estimated values of the parameters of abnormal and normal population, K-S statistic and AUC of RTE_2 ROC Curve.

(λ_x, λ_y)	$(\hat{\lambda}_x, \hat{\lambda}_y)$	**K-S test statistic**	**AUC**
(4, 8)	(6.25, 9.04)	(0.1577, 0.1353)	0.6499
(4.5,9)	(6.78, 9.7)	(0.1357,0.1419)	0.6486
(4.2,12)	(5.427, 17.5)	(0.1752, 0.2852)	0.7500

In Table 4.1, from the table it as observed as we increase the difference between mean of X and Y and the difference between the parameter of two populations, the value of AUC tends to increase which in turn maximizing the discriminatory capability of the biomarker.

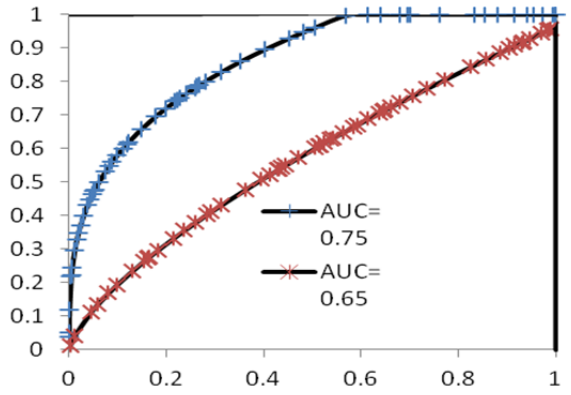

Fig. 1: RTE ROC curve with two different AUC measures 0.75 and 0.65

5. CONCLUSION

In this paper, we have proposed a ROC model that adjusts for right-truncated observations from exponential distribution and evaluates the accuracy of biomarker in classifying the subjects with abnormality. The RTE_2 ROC model, its accuracy measure AUC has been estimated. The RTE_2 ROC curves for different sets of simulated data sets are analyzed for their accuracies.

REFERENCES

[1] Betinec, M., "Testing the difference of the ROC Curves in Biexponential model", *Tatra Mountains Mathematical Publications*, 39, 2008, pp. 215-223.

[2] Vardhan, R.V. Pundir, S. and Sameera, G., "Estimation of Area under the ROC curve Using Exponential and Weibull Distributions", *Bonfring International Journal of Data Mining* 2(2), 2012, pp. 52-56.

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

[4] Suk-Bok Kang and Young-Suk Cho, "Approximate MLE of exponential distribution for truncated samples", *Journal of Information and Optimization Sciences*, 19(3), 1998, pp. 712-749.

[5] Jong-Wuu Wu , Wen-Liang Hung and Ching-Yi Chen, "Approximate MLE of the scale parameter of the truncated Rayleigh distribution under the first failure-censored data", *Journal of Information and Optimization Sciences*, 25:2, 2004, pp. 221-235.

[6] Pundir, S. and Amala, R., "Standard error of Area under the Bi-Exponential ROC curve", *International Journal of Engineering Sciences and Research Technology,* 3(8), 2014, pp.712-721.