iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73819

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Estimation of Area Under BI-Generalized Exponential Distribution ROC Curve Using Confidence Intervals

Dr. Ch. Prasuna[1], Dr. R.V.S.S. Nagabhushana Rao[2]
[1]*Assistant Professor (c), Department of Statistics, Vikrama Simhapuri University, Nellore*
[2]*Assistant Professor (c), Department of Statistics, Vikrama Simhapuri University, Nellore*

***Abstract:** Many researchers have provided the mathematical formulation of the curve by assuming some specific distribution. Conventionally, much work has been carried out by assuming normal distribution. Both of the populations (groups), the bi-Generalized Exponential Distribution model is known to be a good model. In this paper we derive a new method of determining the AUC by utilizing the confidence intervals for Scale and location of the two groups and taking weighted average of possible AUC values with different patterns of weights. We also study the behavior of the new estimator by simulation.*
***Keywords:** ROC, AUC, Bi- Generalized Exponential Distribution, Confidence Intervals.*

I. INTRODUCTION

The Receiving Operating Characteristic (ROC) curve model is a useful graphical and statistical modeling tool for evaluation performance in any two population (group) classification task, example in medicine two groups of subjects (Diseased, Healthy). Even though it originated during Second World War, (Green and Swets, 1966, Lloyd (1998)), many researchers highlighted its significance in Medicine, Experimental Psychology, Finance, Banking, data mining etc., in later years. In biomedical applications there has been an increased use of ROC curve models for assessing the effectiveness of continuous diagnostic marker values (e.g. tumor volume) in distinguishing between diseased and healthy individuals. A person is assessed as diseased (positive) or healthy (negative) depending on whether the corresponding marker value is greater than or less than equal to a given threshold value.

A. The Generalized Exponential Distribution

Recently, Mudholkar, Srivastava&Freimer (1995) proposed a three-parameter (one scale and two shapes) distribution, the exponentiated Weibull distribution (see also Mudholkar & Srivastava, 1993). Both papers analyse certain datasets and show that the exponentiated Weibull, which has three parameters, has a better fit than the two-parameter (taking location parameter to be zero) Weibull or one-parameter exponential, which are special cases of the exponentiated Weibull. More recently, Gupta & Kundu (1997) considered a special case (exponentiated exponential) of the exponentiated Weibull model assuming the location parameter to be zero, and compared its performances with the two-parameter gamma family and the two-parameter Weibull family, mainly through data analysis and computer simulations.

The three-parameter gamma and three-parameter Weibull are the most popular distributions for analysing lifetime data. Both distributions have been studied in the literature, and both have applications in fields other than lifetime distributions; see e.g. Alexander (1962), Jackson (1969), van Kinken (1961) and Masuyama & Kuroiwa (1952). The three parameters, in both distributions, represent location, scale and shape, and because of them both distributions have quite a bit of flexibility for analysing skewed data.

The three-parameter GE has the following density function

$f(x;\alpha=1,\lambda,\mu)=1/\lambda\, e^{-(x-\mu)/\lambda}$; $(x>\mu;\ \lambda>0;\ \alpha=1)$. (1)

The three parameter GE has the following cumulative distribution function

$F(x;\alpha=1,\lambda,\mu)=(1-e^{-(x-\mu)/\lambda})$; $(x>\mu;\ \lambda>0;\ \alpha=1)$. (2)

Here ($\alpha=1$) is a shape parameter, λ is a scale parameter and μ is a location parameter. We denote the GE distribution with shape parameter $\alpha=1$; scale parameter λ and location parameter μ as GE($\alpha=1,\lambda,\mu$).

In next section, a newer version of ROC curve and Area under the Curve are derived. Here, the test scores of healthy (H) and diseased (D) populations follow Generalized Exponential Distribution respectively.

B. Bi-Generalized-Exponential-Distribution Model

A Bi-Generalized-Exponential-Distribution model assumes that, decision values follow two independent Generalized -Exponential-Distributions: one for diseased (negative)and another one is for healthy (positive). Let X and Y are from healthy and diseased populations with underlying GED functions *g(x), f(y)* respectively.

Ehtesham Hussain (2011) has shown that the bi-one parameter generalized exponential ROC is given by

$$\text{ROC}(t) = 1-(1-t)^{\frac{\lambda_D}{\lambda_H}} = 1-(1-t)^{\beta} \quad 0\le t\le 1 \tag{3}$$

Where $\beta = \frac{\lambda D}{\lambda H}$ $\quad 0\le t\le 1$

With density function for the healthy group

$$f(x;\alpha_H,\lambda_H,\mu_H) = \frac{1}{\lambda_H} e^{-\left(\frac{X-\mu_H}{\lambda_H}\right)} \quad \alpha_H = 1 \;\&\; \lambda_H, \mu_H > 0,\ x > 0 \tag{4}$$

otherwise for the diseased group

$$f(x;\alpha_D,\lambda_D,\mu_D) = \frac{1}{\lambda_D} e^{-\left(\frac{X-\mu_D}{\lambda_D}\right)} \quad \alpha_D = 1 \;\&\; \lambda_D, \mu_D > 0,\ x > 0 \tag{5}$$

The false positive rate with threshold t is given by

$X(t) = P(S>t|\ H)$

$$= 1-\left(1-e^{-\frac{t-\mu_H}{\lambda_H}}\right)$$

$$= e^{-\frac{t-\mu_H}{\lambda_H}}$$

$t=\mu_H-\lambda_H \log(x(t))$

and the ROC curve is given as

$Y(x) = P(S>t|\ D)$

$$= 1-\left(1-e^{-\frac{t-\mu_D}{\lambda_D}}\right)$$

$$= e^{-\frac{t-\mu_D}{\lambda_D}}$$

$$= \left(e^{-(\mu_H-\lambda_H \log(x(t))-\mu_D)/\lambda_D}\right)$$

$$= \left(e^{\left(\frac{\mu_D-\mu_H}{\lambda_D}\right)} t^{\frac{\lambda_H}{\lambda_D}}\right) \tag{6}$$

The three parameter generalized exponential AUC is given by

$$AUC = \frac{\lambda_D}{\lambda_D+\lambda_H} e^{\left(\frac{\mu_D-\mu_H}{\lambda_D}\right)} \tag{7}$$

C. Bi- Generalized Exponential AUC with confidence intervals for Locations and Scales

Assuming that n1 and n2 are the sample sizes of the H and D groups respectively and 100 (1-α)% confidence intervals of location(μ) is obtained from χ^2- distribution as follows.

Group	Lower	Middle	Upper
H	$L_1 = \hat{\mu}_H - \left[\frac{n_1\hat{\lambda}_H F_{(2,2n_1-2,1-\alpha/2)}}{n_1(n_1-1)}\right]$	$\hat{\mu}_H$	$U_1 = \hat{\mu}_H - \left[\frac{n_1\hat{\lambda}_H F_{(2,2n_1-2,\alpha/2)}}{n_1(n_1-1)}\right]$
D	$L_2 = \hat{\mu}_D - \left[\frac{n_2\hat{\lambda}_D F_{(2,2n_2-2,1-\alpha/2)}}{n_2(n_2-1)}\right]$	$\hat{\mu}_D$	$U_2 = \hat{\mu}_D - \left[\frac{n_2\hat{\lambda}_D F_{(2,2n_2-2,\alpha/2)}}{n_2(n_2-1)}\right]$

Table-1: Extreme estimates of the locations

Since the true location in the worst case reaches the limits of the confidence interval, there will be 9 possible locations (used (8)) as given below.

$S_1 = (L_2 - L_1)/\lambda_D$, $S_2 = (L_2 - \hat{\mu}_H)/\lambda_D$, $S_3 = (L_2 - U_1)/\lambda_D$, $S_4 = (U_2 - L_1)/\lambda_D$, $S_5 = (U2 - \hat{\mu}_H)/\lambda_D$, $S_6 = (U_2 - U_1)/\lambda_D$, $S_7 = (\hat{\mu}_D - L_1)/\lambda_D$, $\Delta_8 = (\hat{\mu}_D - \hat{\mu}_H)/\lambda_D$ and $\Delta_9 = (\hat{\mu}_D - U_1)/\lambda_D$ (8)

Similarly the 100(1-α)% confidence intervals of Scales (λ) is obtained from χ^2- distribution as follows

.

Group	Lower	Middle	Upper
H	$L_1 = \frac{2n_1\hat{\lambda}_H}{\chi^2_{(2n_1-2,1-\alpha/2)}}$	$\hat{\lambda}_H$	$\frac{2n_1\hat{\lambda}_H}{\chi^2_{(2n_1-2,\alpha/2)}}$
D	$\frac{2n_2\hat{\lambda}_D}{\chi^2_{(2n_2-2,1-\alpha/2)}}$	$\hat{\lambda}_D$	$\frac{2n_2\hat{\lambda}_D}{\chi^2_{(2n_2-2,\alpha/2)}}$

Table-2: Extreme estimates of the scales

Then the pooled Scales is obtained by considering 9 different combinations as below

$V_1 = L_2/(L_2+L_1)$, $V_2 = L_2/(L_2+U_1)$, $V_3 = L_2/(L2+\hat{\lambda}_H)$, $V_4 = U_2/(U_2+L_1)$, $V_5 = U_2/(U_2+L_2)$, $V_6 = U_2/(U_2+\hat{\lambda}_H)$, $V_7 = \hat{\lambda}_D/(\hat{\lambda}_D+L1)$, $V_8 = \hat{\lambda}_D/(\hat{\lambda}_D+U1)$ and $V_9 = \hat{\mu}_D/(\hat{\mu}_D+\hat{\lambda}_H)$ (9)

When both locations and scales are replaced with their interval estimates, there will be 9 x 9 = 81 possible combinations for evaluating the AUC given in (11). Let A_{ij} denote the AUC with S_i and V_j for i , j = 1,2,.. 9.

Then the matrix A = $[A_{ij}]_{9\times 9}$, where $A_{ij} = V_i e^{S_j}$ produces 81 mutually distinct estimates of the AUC that arise due to the use of interval estimates. We call this the *joint AUC* matrix.

We get $\hat{AUC} = \frac{\lambda_D}{\lambda_D+\lambda_H} e^{\left(\frac{\mu_D-\mu_H}{\lambda_D}\right)}$

$$SE(A\hat{U}C) = \sqrt{\frac{A(1-A)+(n_1-1)(Q_1-A)^2+(n_2-1)(Q_2-A)^2}{n_1 n_2}} \quad (10)$$

Where A = AUC , $Q_1 = A/(2-A)$ and $Q_2 = 2A2/(1+A)$

At each combination of λ and μ then SE is calculated as

$$SE(\hat{A}_{ij}) = \sqrt{\frac{A_{ij}(1-A_{ij})+(n_1-1)(Q_1-A_{ij})^2+(n_2-1)(Q_2-A_{ij})^2}{n_1 n_2}} \quad (11)$$

There will be 81 combinations now the pooled estimate of AUC & the corresponding AUC is found by using two different weighing schemes as listed below.

1) Simple Average Method

In this method, the weights are taken as $W_{ij} = 1/81 \ \forall$ i, j = 1, 2, ….., 9 and the estimate is given as

$$AUC_{AVG} = \sum_{i=1}^{9} \sum_{j=1}^{9} W_{ij} A_{ij} \quad (12)$$

2) Fixed Weights Method (FW-Method)

Let p_j denote the probability that A_j occurs in a given context. Then { $p_1, p_2 \ldots\ldots p_9$}, $p_j \geq 0$, $\sum_{j=1}^{9} p_j = 1$ is a probability distribution. The following is one method of assigning probabilities

$$W_{ij} = \frac{\hat{p}}{m} \quad \text{for (i,j)}$$

$$= \frac{1-\hat{p}}{N-m} \quad \text{for(i , j)}$$

Where m= no of combinations & N=81

A_{ij} = AUC

The estimator of AUC is then given by

$$AUC_{FW} = \sum_{j=1}^{9} A_{ij} W_{ij} \quad (13)$$

D. Numerical illustration

We have conducted trials in which n_1, n_2, μ_H, μ_D, λ_H, λ_D are given such that the ROC curve will have a target AUC say 0.8108. The 81 combinations of the joint AUC matrix with elements $A_{ij} = V_i e^{S_j}$ for i = j = 1,2,…,9 are evaluated using Excel functions. Taking the input parameters as n1 = 100, n_2 = 100, μ_H = 0.06, μ_D= 0.10 λ_H =0.4, λ_D =1.5 we get the joint AUC matrix as shown in table-1.

Trail -1: Target AUC = 0.8108, n_1 = = 100, n_2 = 100, μ_H = 0.06, μ_D= 0.10 λ_H =0.4, λ_D =1.5									
	S1	S2	S3	S4	S5	S6	S7	S8	S9
V1	0.8108	0.8128	0.8118	0.8033	0.8053	0.8043	0.8070	0.8090	0.8080
V2	0.8424	0.8445	0.8434	0.8347	0.8367	0.8357	0.8385	0.8406	0.8395
V3	0.8263	0.8283	0.8273	0.8187	0.8207	0.8197	0.8225	0.8245	0.8235
V4	0.7754	0.7773	0.7763	0.7682	0.7701	0.7692	0.7718	0.7737	0.7727
V5	0.8108	0.8128	0.8118	0.8033	0.8053	0.8043	0.8070	0.8090	0.8080
V6	0.7927	0.7946	0.7936	0.7854	0.7873	0.7864	0.7890	0.7910	0.7900
V7	0.7944	0.7964	0.7954	0.7871	0.7890	0.7881	0.7907	0.7927	0.7917
V8	0.8278	0.8299	0.8288	0.8202	0.8222	0.8212	0.8240	0.8260	0.8250
V9	0.8108	0.8128	0.8118	0.8033	0.8053	0.8043	0.8070	0.8090	0.8080

Table -3: Simulated AUC values with interval estimates for both location and scale parameters.

The new estimates (mean ± standard error) derived by the three methods are obtained as follows. a) Simple Average Method: 0.834396 ± 0.161143

b) Fixed Weights Method: 0.872413 ± 0.161422

Keeping other parameters fixed and changing the sample sizes gives the following results.

Sample Size	Method of averaging	Joint AUC method
10	Simple Average	0.6744 ± 0.0788
	Fixed Weights	0.7393 ±0.071668
20	Simple Average	0.7426 ± 0.0463
	Fixed Weights	0.7750 ± 0.043086

50	Simple Average	0.7836 ± 0.0250
	Fixed Weights	0.7965± 0.02407
110	Simple Average	0.7985 ± 0.0157
	Fixed Weights	0.8043 ± 0.015415
200	Simple Average	0.8004 ± 0.0143
	Fixed Weights	0.8053 ± 0.014082

Table-4: Estimated AUC ± Standard Error by different methods for different sample sizes.

The pattern of AUC with the two methods against the sample size is shown in Figure-1. It can be seen that both methods produce the same result when the sample size is large.

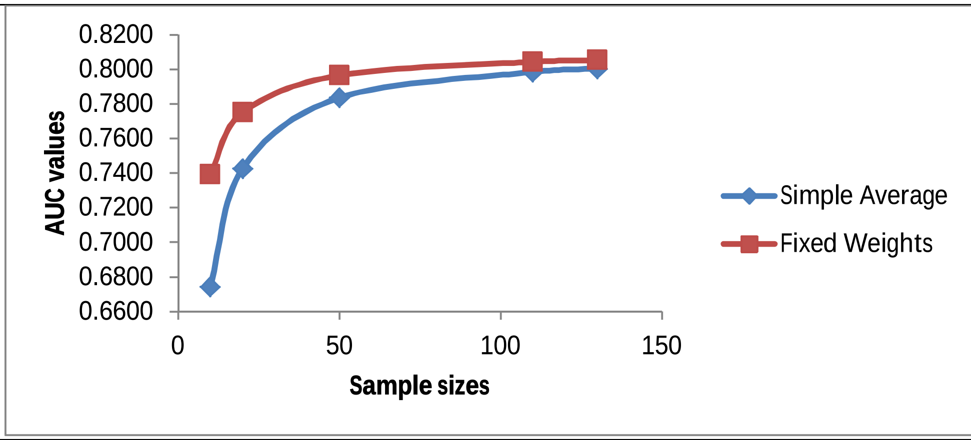


Figure-1: Estimated AUC as a function of the sample size

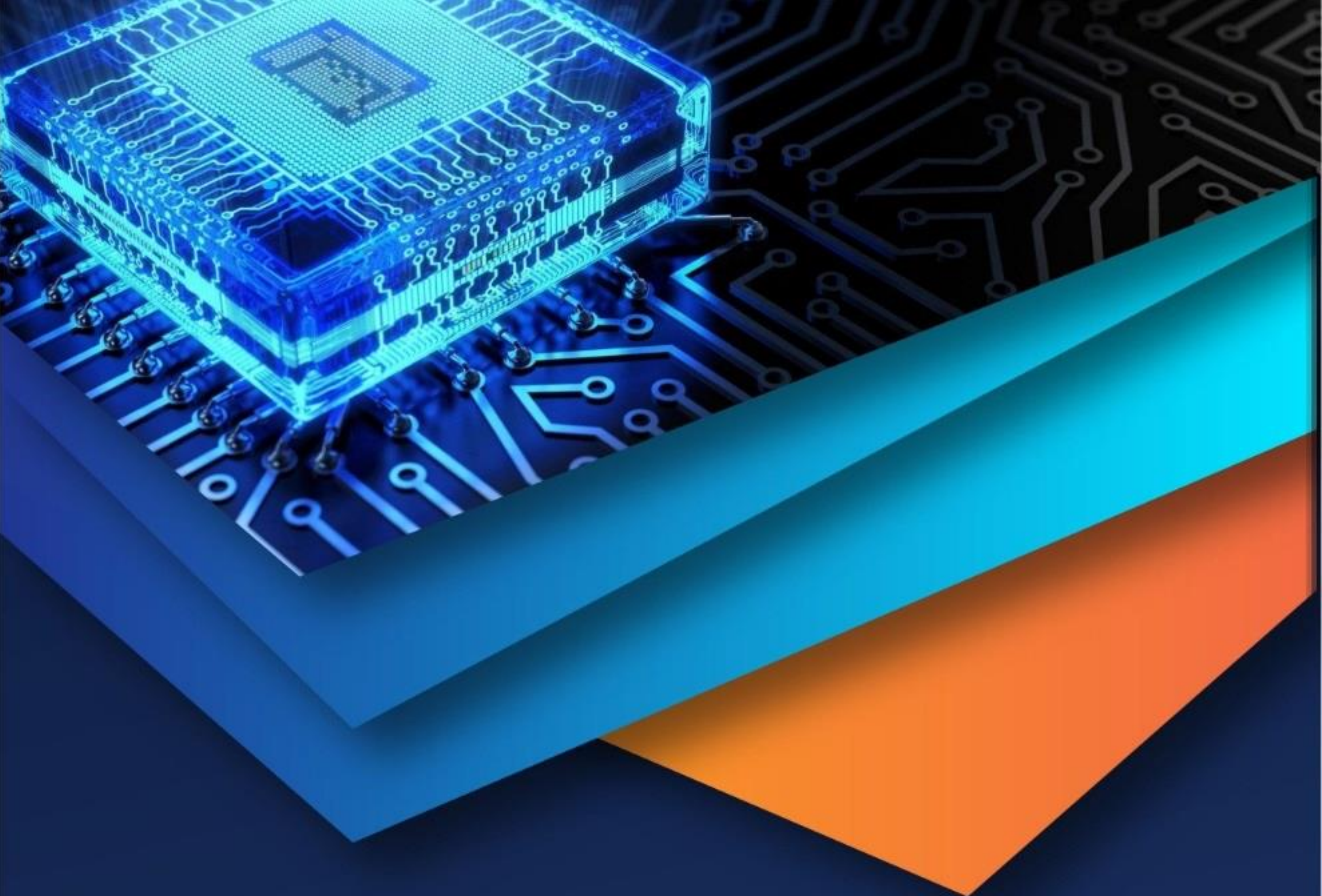
II. CONCLUSIONS

In this paper we have developed a new method of estimating the AUC of bi-generalized exponential distribution ROC model by using interval estimates .We have identified 9 different combinations that can be considered for evaluating the AUC using the Normal probability distribution. The new estimate is based on 81 possible AUC values and combining them as a weighted average with two different weighting schemes. It is shown that for large samples all the methods work equally well. Further the standard error of the estimate steeply decreases as the sample size increases.

REFERENCES

[1] Green, D.M and Swets, J.A (1966). *"Signal Detection theory and Psychophysics"*. Wiley, Newyork.
[2] ALEXANDER, G.N. (1962). The use of the gamma distribution in estimating the regulated output from the storage. *Trans. Civil Engineering, Institute of Engineers, Australia* **4**, 29–34.
[3] JACKSON, O.A.Y. (1969). Fitting a gamma or log-normal distribution to fibre-diameter measurements of wool tops. *Appl. Statist.* **18**, 70–75.
[4] VAN KINKEN, J. (1961). A method for inquiring whether the *0* distribution represents the frequency distribution of industrial accident costs. *Acturielle Studien* **3**, 83–92.
[5] MASUYAMA,M.&KUROIWA, Y. (1952). Table for the likelihood solutions of gamma distribution and its medical applications. *Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs.* **1**, 18–23.
[6] Ehtesham Hussain. (2011), " the ROC Curve Model from Generalized –Exponential distribution", Pak.j.stat.oper.res.Vol.VII No.2, pp323-330.
[7] MUDHOLKAR, G.S. & SRIVASTAVA, D.K.(1993). Exponentiated Weibull family for analyzing bathtub failure data. *IEEE Trans. Reliability* **42**, 299–302.
[8] GUPTA, R.D. & KUNDU, D. (1997). Exponentiated exponential family: an alternative to gamma and Weibull distribution. Technical report. Dept of Math., Stat. & Comp. Sci., University of New Brusnwick, Saint- John, NB, Canada.
[9] Mudholkar, G.S. Srivastava, D.K. &Freimer. M. (1995). The exponentiated Weibull family: a reanalysis of the bus motor failure data. *Technometrics* **37**, 436–445.
[10] Lloyd, C. J. (1998). Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. Journal of the American Statistical Association, 93, 1356-1364.
[11] Faraggi. D and Reiser.B (2002), Estimation of the Area under the ROC curve ,Statistics in Medicine; 21 :3093-3106
[12] Betinec. M (2008), "Testing the difference of the ROC Curves in Bi-exponential Model" Tatra Mountains Mathematical Publications,39,215-223.
[13] R.VishnuVardhan and, Sarma KVS (2010), " Estimation of the Area under the ROC Curve using confidence intervals of means " ANU Journal of Physical Sciences 2(1), 29-39.

[14] R.Vishnu Vardhan, SudeshPundir and G.Sameera (2012), " Estimating of Area Under the ROC Curve Using Exponential and Weibull distributions", Bonfring International Journal of Data Mining Vol.2,No.2,June.
[15] Prasuna. Ch and Sarma. K.V.S (2013), **"**Estimating the Area under the ROC curve and confidence intervals using Bi-exponential model", International Journal of Statistics and Analysis. ISSN 2248-9959 Volume 3, Number 3 (2013), pp. 323-331.
[16] Suresh babu. N and Sarma. K.V.S(2013)" On the Estimation of Area under Binormal ROC curve using Confidence IntervalsforMeans and Variances"



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY