

Testing the Difference of the ROC Curves in Bi-Pareto Model

Kanchan Jain* , Bhavna Kaushal and Suresh K. Sharma

Department of Statistics, Panjab University, Chandigarh, India.
Email: jaink14@gmail.com

Abstract. Receiver Operating Characteristic (ROC) curves are used for evaluation of classifier's performance. They have applications in clinical diagnostics in medicine, computational linguistics, machine learning and data mining. In this paper, we propose a test for the equality of two ROC curves for Bi-Pareto ROC model. The asymptotic distribution of the test statistic is determined through simulations and the size of the test is computed. A real data set has been used to demonstrate the testing procedure.

Keywords: Sensitivity, Specificity, Receiver Operating Characteristic curve (ROC), Area under ROC curve, Bi-Pareto.

1 Introduction

Receiver operating characteristic (ROC) curves have become a standard tool for evaluation of the discriminatory power of medical diagnostic tests and are commonly used in assessing the predictive ability of binary regression models. They are diagnostic tools that help in determining the accuracy of a test conducted on a person to know whether a particular disease is present or not. In a typical setting, with a binary indicator and a set of predictors or marker values, the motive is to see how well the marker values predict the binary indicator. The principal idea is to dichotomize the marker at various thresholds and compute the resulting sensitivity and specificity. Sensitivity or True positive rate (TPR) of a test is defined as the probability of a positive test result when the disease is present and specificity or True Negative Rate (TNR) is the probability of a negative test result when disease is absent. False Positive Rate (FPR) is given by (1-specificity). ROC curve is obtained by plotting the sensitivity versus (1-specificity) or TPR versus FPR.

In credit rating models in finance, sensitivity is termed as 'Hit Rate' (HR) whereas (1-specificity) is known as 'False Alarm Rate' (FAR). If the rating score of the debtor is lower than a cut-off value C , he is considered a defaulter. Otherwise, he is a non-defaulter on loan. Hence

$$\text{HR}(C) = \frac{\text{Number of defaulters classified correctly}}{\text{Total number of defaulters}}$$

and

$$\text{FAR}(C) = \frac{\text{Number of non-defaulters classified incorrectly}}{\text{Total number of non-defaulters}}$$

ROC curve plots HR versus FAR ([1]). For detailed discussion on ROC curves, one can refer to [2].

Let F and G be the cumulative distribution functions (cdf) of two populations N and P . Then the ROC curve has the form

$$\text{ROC} = G\{F^{-1}(x)\}.$$

The area under ROC curve (AUROC) is a widely used summary index ([3], [4], [5] and [6]). It is the average TPR taken uniformly over all FPRs on (0, 1) and is written as

$$\text{AUROC} = \int_0^1 G\{F^{-1}(x)\} dx \quad (1)$$

For credit rating models,

$$\text{AUROC} = \int_0^1 \text{HR}[\text{FAR}] d(\text{FAR}).$$

For the purpose of classifier comparison, [7] suggested a test for equality of ROC curves in the case of Binormal model. But there are situations where binormal model cannot be used. [8] gave test for equality of ROC curves in the case of Biexponential model. Bi-Pareto model was introduced by [9] and we propose a test for testing the equality of ROC curves for Bi-Pareto model.

For a Bi-Pareto model, the populations non diseased (N) and diseased (P) are assumed to follow Pareto(α_1, λ_1) and Pareto(α_2, λ_2) for $\alpha_1, \alpha_2, \lambda_1, \lambda_2 > 0$ respectively. The Bi-Pareto ROC curve has the form [9]

$$\text{ROC}_{\text{BP}} = G[F^{-1}(x)] = 1 - \left[\frac{\lambda_2 (1-x)^{1/\alpha_1}}{\lambda_1} \right]^{\alpha_2}. \quad (2)$$

The corresponding area under ROC is given by

$$\text{AUROC} = \left[1 - \left(\frac{\lambda_2}{\lambda_1} \right)^{\alpha_2} \right] \frac{\alpha_1}{\alpha_1 + \alpha_2}.$$

Section 2 consists of the proposed test statistic for testing the equality of two ROC curves for two classifiers A and B. In Section 3, the probability density function (pdf) and cumulative distribution function (cdf) of the proposed test statistic have been derived. Section 4 consists of simulation results for different sample sizes for different parametric setups. In Section 5, we apply the procedure to a real life data set.

2 Proposed Test Procedure

Let there be two classifiers A and B with ROC^A and ROC^B as the corresponding Receiver Operating Characteristic curves. We wish to test the hypothesis

$$H_0 : \text{ROC}^A = \text{ROC}^B$$

versus

$$H_1 : \text{ROC}^A > \text{ROC}^B \text{ or } H_1' : \text{ROC}^A < \text{ROC}^B.$$

This is equivalent to testing the null hypothesis

$$H_0 : \text{AUROC}^A = \text{AUROC}^B$$

versus

$$H_1 : \text{AUROC}^A > \text{AUROC}^B \text{ or } H_1' : \text{AUROC}^A < \text{AUROC}^B.$$

For classifier A, the Bi-Pareto model has the underlying distribution as Pareto(α_1^A, λ_1^A) and Pareto(α_2^A, λ_2^A). Similarly for classifier B, Pareto(α_1^B, λ_1^B) and Pareto(α_2^B, λ_2^B) give rise to a Bi-Pareto model. For $I \in \{A, B\}$, we write

$$\delta_I = \frac{\lambda_2^I}{\lambda_1^I} \text{ and } \tau_I = (\delta_I)^{\alpha^I}.$$

It is assumed that $\alpha_1^A = \alpha_2^A = \alpha^A$ and $\alpha_1^B = \alpha_2^B = \alpha^B$ but $\lambda_1^A \neq \lambda_2^A$ and $\lambda_1^B \neq \lambda_2^B$.

For the sake of simplicity we shall write $\alpha^A = \beta_1, \alpha^B = \beta_2$.

Using (2), $(\text{ROC})^A = (\text{ROC})^B$ if $\left(\frac{\lambda_2^A}{\lambda_1^A}\right)^{\alpha^A} = \left(\frac{\lambda_2^B}{\lambda_1^B}\right)^{\alpha^B}$, that is $\tau_A = \tau_B$.

Hence H_0 and H_1 can be equivalently written as

$$H_0 : \tau_A = \tau_B$$

versus

$$H_1 : \tau_A > \tau_B \quad \text{or} \quad H_1' : \tau_A < \tau_B$$

For testing H_0 versus H_1' , we propose the test statistic $T = \frac{\hat{\tau}_A / \tau_A}{\hat{\tau}_B / \tau_B}$ where $\hat{\tau}_A = (\hat{\delta}_A)^{\hat{\beta}_1}$ and

$$\hat{\tau}_B = (\hat{\delta}_B)^{\hat{\beta}_2}.$$

Under H_0 , the test statistic $T = \frac{\hat{\tau}_A}{\hat{\tau}_B}$.

Since it is difficult to find the asymptotic distribution of T , we decide on it through simulations in Section 4. This is done by deriving the expressions for the probability density function (pdf) and the cumulative distribution function (cdf) of T and then using the probability integral transformation.

3 Distribution of the Proposed Test Statistic

For the two classifiers A and B, we let

- n and p : number of sample observations corresponding to A from populations N and P ;
- m and q : number of sample observations corresponding to B from populations N and P .

For classifier A, using the maximum likelihood estimators (MLEs) of λ_1 and λ_2 , we can write [10]

$$\hat{\delta}_A = \frac{\min_{1 \leq i \leq p} Y_i}{\min_{1 \leq i \leq n} X_i} = \frac{z_1}{z_2}$$

Since X_i follows Pareto distribution $(\alpha_1^A, \lambda_1^A)$, hence the survival function of $Z_1 = \min_{1 \leq i \leq n} X_i$ is

$$\bar{F}_{z_1}^A(x) = 1 - F_{z_1}^A(x) = \left(\frac{\lambda_1^A}{x}\right)^{n\alpha^A} = \left(\frac{\gamma_1}{x}\right)^{n\beta_1}$$

for $x > 0$, where $\lambda_1^A = \gamma_1$, and we have $\bar{F}_{z_1}^A = \left(\frac{\gamma_1}{x}\right)^{n\beta_1}$.

This gives the probability density function (pdf) of Z_1 as

$$f_{z_1}^A(x) = \frac{n\beta_1 (\gamma_1)^{n\beta_1}}{x^{n\beta_1+1}}, \quad x > 0. \tag{3}$$

Writing $\gamma_2 = \lambda_2^A$, the survival and probability density functions of Z_2 are

$$\bar{F}_{z_2}^B(x) = \left(\frac{\lambda_2^A}{x}\right)^{m\alpha^A} = \left(\frac{\gamma_2}{x}\right)^{m\beta_1} \quad \text{and} \quad f_{z_2}^B(x) = \frac{m\beta_1 (\gamma_2)^{m\beta_1}}{x^{m\beta_1+1}}. \tag{4}$$

Using (3) and (4), the pdf of $\hat{\delta}_A = \frac{\lambda_2^A}{\lambda_1^A}$ is given by

$$g_{\hat{\delta}_A}(u) = \int_{-\infty}^{\infty} f_{z_2}^A(uv) f_{z_1}^A(v) |v| dv$$

$$= \begin{cases} \frac{nm\beta_1 \left(\frac{\gamma_1}{\gamma_2}\right)^{n\beta_1}}{n+m} u^{n\beta_1-1}, & 0 < u < \frac{\gamma_2}{\gamma_1} \\ \frac{nm\beta_1 \left(\frac{\gamma_2}{\gamma_1}\right)^{m\beta_1}}{n+m} \frac{1}{u^{m\beta_1+1}}, & \frac{\gamma_2}{\gamma_1} < u < \infty \end{cases}. \quad (5)$$

On similar lines, writing $\theta_1 = \lambda_1^B$, $\theta_2 = \lambda_2^B$, we have

$$f_{z_1}^B(x) = \frac{p\beta_2(\theta_1)^{p\beta_2}}{x^{p\beta_2+1}}, \quad x > 0 \quad \text{and} \quad f_{z_2}^B(x) = \frac{q\beta_2(\theta_2)^{q\beta_2}}{x^{q\beta_2+1}}, \quad x > 0. \quad (6)$$

Hence the pdf of $\hat{\delta}_B = \frac{\lambda_2^B}{\lambda_1^B}$ is given by

$$g_{\hat{\delta}_B}(u) = \int_{-\infty}^{\infty} f_{z_2}^B(uv) f_{z_1}^B(v) |v| dv$$

$$= \begin{cases} \frac{pq\beta_2 \left(\frac{\theta_1}{\theta_2}\right)^{p\beta_2}}{p+q} u^{p\beta_2-1}, & 0 < u < \frac{\theta_2}{\theta_1} \\ \frac{pq\beta_2 \left(\frac{\theta_2}{\theta_1}\right)^{q\beta_2}}{p+q} \frac{1}{u^{q\beta_2+1}}, & \frac{\theta_2}{\theta_1} < u < \infty \end{cases} \quad (7)$$

Using (5) and (6), we can write pdfs of $\hat{\tau}_A$ and $\hat{\tau}_B$ as

$$h_{\hat{\tau}_A}(y) = \frac{y^{\frac{1}{\beta_1}}}{\beta_1} g_{\hat{\delta}_A}\left(y^{\frac{1}{\beta_1}}\right)$$

$$= \begin{cases} \frac{nm}{n+m} \left(\frac{\gamma_1}{\gamma_2}\right)^{n\beta_1} y^{n-1}, & 0 < y < \left(\frac{\gamma_2}{\gamma_1}\right)^{\beta_1} \\ \frac{nm}{n+m} \left(\frac{\gamma_2}{\gamma_1}\right)^{m\beta_1} \frac{1}{y^{m+1}}, & \left(\frac{\gamma_2}{\gamma_1}\right)^{\beta_1} < y < \infty \end{cases} \quad (8)$$

Under H_0 , the pdf of the proposed statistic $T = \frac{\hat{\tau}_A}{\hat{\tau}_B}$ can be written using (7) and (8) and is given by

$$f_T(t) = \begin{cases} c \left\{ \frac{t^{q-1}}{(m+q)d^q} + \frac{t^{q-1}}{(n-q)d^q} - \frac{t^{n-1}}{(n-q)d^n} + \frac{t^{n-1}}{(n+p)d^n} \right\}, & 0 < t < d; \\ c \left\{ \frac{1}{p-m} (d^m t^{-m-1} - d^p t^{-p-1}) + \frac{1}{n+p} d^p t^{-p-1} + \frac{1}{m+q} d^m t^{-m-1} \right\}, & d < t < \infty \end{cases} \quad (9)$$

where $c = \frac{nmpq}{(n+m)(p+q)}$ and $d = \left(\frac{\gamma_2}{\gamma_1}\right)^{\beta_1} \left(\frac{\theta_1}{\theta_2}\right)^{\beta_2}$.

The corresponding cdf of $T = \frac{\hat{\tau}_A}{\hat{\tau}_B}$ is obtained as

$$F_T(t) = \begin{cases} \left[\frac{c(m+n)}{q(m+q)(n-q)} - \frac{c(p+q)}{n(n+p)(n-q)} - \left[\frac{cd^m t^{-m}(p+q)}{m(p-m)(m+q)} + \frac{cd^p t^{-p}(m+n)}{p(m-p)(n+p)} \right] + \frac{c(p+q)}{m(m+q)(p-m)} + \frac{c(m+n)}{p(n+p)(m-p)} \right], & t > d \\ \frac{ct^q}{qd^q} \frac{m+n}{(m+q)(n-q)} - \frac{ct^n}{nd^n} \frac{p+q}{(n+p)(n-q)}, & 0 < t < d \\ 0, & t < 0 \end{cases} \quad (10)$$

Using (9),

$$E(T) = cd \left\{ \frac{\frac{(m+n)}{(m+q)(n-q)(q+1)} - \frac{(p+q)}{(n-q)(n+p)(n+1)}}{\frac{(p+q)}{(p-m)(m+q)(m-1)} - \frac{(m+n)}{(p-m)(n+p)(p-1)}} + \right\} \quad (11)$$

$$V(T) = cd^2 \left\{ \frac{\frac{(m+n)}{(m+q)(n-q)(q+2)} - \frac{(p+q)}{(n-q)(n+p)(n+2)} + \frac{(p+q)}{(p-m)(m+q)(m-2)} - \left[\frac{(m+n)}{(p-m)(n+p)(p-2)} - c \left[\frac{(m+n)}{(m+q)(n-q)(q+1)} - \frac{(p+q)}{(n-q)(n+p)(n+1)} \right] + \frac{(p+q)}{(p-m)(m+q)(m-1)} - \frac{(m+n)}{(p-m)(n+p)(p-1)} \right]^2}{\left[\frac{(p+q)}{(p-m)(m+q)(m-1)} - \frac{(m+n)}{(p-m)(n+p)(p-1)} \right]^2} \right\} \quad (12)$$

As it is difficult to identify the distribution of T analytically, we show through simulations that under H_0 , $Z = \frac{T - E(T)}{\sqrt{V(T)}}$ follows standard normal distribution. This is validated using Kolmogorov-Smirnov test.

In the next section, we carry out simulations using R software and calculate the value of Z .

4 Simulations Studies

For two classifiers A and B, simulations are carried out using different values of n, m, p and q . The corresponding p -values of KS test are given in Tables 1-3. For all the parametric combinations, it is observed that Z follows standard normal distribution.

From the p -values in Tables 1-3, it is observed that the test statistic Z follows standard normal distribution for all considered combinations. Table 4 shows the power of the proposed test.

Table 1. p values of KS test and size of the proposed test

$\gamma_1 = 3, \gamma_2 = 1.5, \theta_1 = 2.5, \theta_2 = 2, \alpha_1 = 2.5, \alpha_2 = 3$						
N	m	p	q	p-value	Size	
5	10	15	10	0.835	0.04	
15	20	15	10	0.6044	0.04	
40	30	45	50	0.9068	0.05	
70	80	65	75	0.8529	0.04	
75	80	90	85	0.9291	0.05	
100	110	90	105	0.1569	0.04	
150	100	175	125	0.2734	0.05	
200	150	175	225	0.8613	0.04	

250	200	175	275	0.7252	0.05
275	300	200	225	0.1519	0.04
325	350	250	275	0.8316	0.04
400	425	300	350	0.7002	0.04
450	400	325	375	0.604	0.05
500	450	350	400	0.6666	0.04

Table 2. p values of KS test and size of the proposed test

$$\gamma_1 = 2.5, \gamma_2 = 2, \theta_1 = 3, \theta_2 = 1.5, \alpha_1 = 2, \alpha_2 = 2.5$$

N	m	p	q	p-value	Size
25	30	35	20	0.1283	0.05
50	60	45	70	0.9252	0.04
100	110	90	105	0.7612	0.04
150	100	175	125	0.5269	0.05
200	150	175	225	0.8917	0.04
250	200	175	275	0.6879	0.04
275	300	200	225	0.7067	0.05
300	325	200	275	0.6466	0.04
325	350	250	275	0.6313	0.05
350	375	275	300	0.8375	0.04
400	425	300	350	0.1527	0.04

Table 3. p values of KS test and size of the proposed test

$$\gamma_1 = 3.5, \gamma_2 = 2.5, \theta_1 = 2, \theta_2 = 3, \alpha_1 = 2.5, \alpha_2 = 3$$

N	m	p	Q	p-value	Size
5	10	15	10	0.8325	0.05
15	20	15	10	0.6289	0.04
40	30	45	50	0.3886	0.05
50	60	45	70	0.4939	0.04
150	100	175	125	0.4945	0.05
200	150	175	225	0.6285	0.04
275	300	200	225	0.4649	0.05
300	325	200	275	0.2417	0.05
325	350	250	275	0.2159	0
350	375	275	300	0.6403	0

Table 4. Power of the proposed test

Sample(m, n, p, q)	Power
$\gamma_1 = 1, \gamma_2 = 1.5, \theta_1 = 2.5, \theta_2 = 2, \alpha_1 = 2.5, \alpha_2 = 3$	
(1, 2, 3, 4)	0.9463148
(2, 5, 3, 4)	0.9907312
(5, 6, 4, 5)	0.9951299
(5, 6, 4, 8)	0.9971126
(9, 7, 4, 6)	0.9983687
(8, 6, 5, 7)	0.9992284
(9, 10, 11, 8)	1
$\gamma_1 = 2.5, \gamma_2 = 2, \theta_1 = 3, \theta_2 = 1.5, \alpha_1 = 2, \alpha_2 = 2.5$	

(1, 2, 3, 4)	0.9640934
(2, 5, 3, 4)	0.990099
(5, 6, 4, 5)	0.9937304
(5, 6, 4, 8)	0.9970414
(9, 7, 4, 6)	0.9960254
(8, 6, 5, 7)	0.9960876
(9, 10, 11, 8)	1
$\gamma_1 = 2, \gamma_2 = 2.5, \theta_1 = 2, \theta_2 = 3, \alpha_1 = 2.5, \alpha_2 = 3$	
(1, 2, 3, 4)	0.9260143
(2, 5, 3, 4)	0.9326923
(5, 6, 4, 5)	0.9354839
(5, 6, 4, 8)	0.9444444
(9, 7, 4, 6)	0.9473684
(8, 6, 5, 7)	0.9565217
(9, 10, 11, 8)	1
$\gamma_1 = 1.5, \gamma_2 = 2.5, \theta_1 = 2, \theta_2 = 3.5, \alpha_1 = 3, \alpha_2 = 4$	
(1, 2, 3, 4)	0.5452482
(2, 5, 3, 4)	0.6774194
(5, 6, 4, 5)	0.7727273
(5, 6, 4, 8)	0.8571429
(9, 7, 4, 6)	0.9090909
(10, 6, 5, 7)	0.9230769
(9, 10, 11, 8)	1

It is observed from the above table that for certain combinations, the value of power is quite high and even close to 1 for the combination (9,10,11,8) .

5 Real Life Example

The data given below consist of survival times of 50 patients ([11]) with advanced acute myelogenous leukemia reported to the International Bone Marrow Transplant registry. Twenty eight of these patients had received an autologous (auto) bone marrow transplant in which, after high doses of chemotherapy, their own marrow was reinfused to replace their destroyed immune system. 22 patients had an allogeneic (allo) bone marrow transplant where marrow from an HLA (Histocompatibility Leukocyte Antigen) matched sibling was used to replenish their immune systems.

Table 5. Leukemia free-survival times (in months) for Autologous and Allogeneic Transplants

Allo transplant patients: 0.030, 0.493, 0.855, 1.184, 1.283, 1.480, 1.776, 2.138, 2.500, 2.763, 2.993, 3.224, 3.421, 4.178, 5.691, 6.941, 8.882, 8.882, 11.480, 11.513, 12.796, 20.066
Auto patients: 0.658, 0.822, 1.414, 2.500, 3.322, 3.816, 4.737, 4.934, 5.033, 5.757, 5.855, 5.987, 6.151, 6.217, 8.651, 8.717, 10.329, 11.480, 12.007, 12.237, 15.461, 15.757, 16.480, 16.711, 17.237, 18.092, 23.158, 56.086.

In the above data set, we take median as the threshold value. The values which are less than median are in the 1st group and the values greater than median are in the 2nd group for both the populations. Group 1 values correspond to classifier A and the values corresponding to classifier B are in group 2 for both the populations. In this case, $m = 11$, $n = 11$, $p = 14$, $q = 14$, $\hat{\gamma}_1 = 0.030$, $\hat{\gamma}_2 = 3.224$,

$\hat{\theta}_1 = 0.658$, $\hat{\theta}_2 = 8.651$, $\hat{\beta}_1 = 0.5463$, $\hat{\beta}_2 = 1.8477$, $\hat{\delta}_A = 107.467$, $\hat{\delta}_B = 13.147$, $\hat{\tau}_A = 12.873$,
 $\hat{\tau}_B = 116.756$.

Hence, $T = \frac{\hat{\tau}_A}{\hat{\tau}_B} = 0.110255$.

Using (13) and (14), $E(T) = 0.1115$, $V(T) = 0.0003288$ and hence $S.D(T) = 0.0181$.

This gives $Z = \frac{T - E(T)}{\sqrt{V(T)}} = -0.06878$.

As $|Z| < 1.96$ for $\alpha = 0.05$, we don't reject H_0 implying that the ROC curves for both classifiers A and B are the same.

Acknowledgement. The corresponding author is thankful to University Grants Commission, Government of India, for providing financial support for this work.

References

1. S. Satchell and W. Xia, Parametrical Models of the ROC Curve: Applications to Credit Rating Model Validation, *Quantitative Finance Research Centre Research*, Paper 181, University of Technology, Sydney, 2006.
2. W. J. Krzanowski and D. J. Hand, ROC Curves for Continuous Data, Taylor and Francis Group, New York, 2009.
3. D. Bamber, The Area above the Ordinal Dominance Graph and the Area below the Receiver Operating Characteristic Graph, *Journal of Mathematical Psychology*, 12, 1975, pp. 387-415.
4. J.A. Hanley and B.J. McNeil, The Meaning and Use of the Area Under ROC Curve, *Radiology*, 4, 1982, pp. 49-58.
5. D. M. Green and J.A. Swets, Signal Detection Theory and Psychophysics, Wiley, New York, 1966.
6. A.P. Bradley, The Use of the Area Under ROC Curve in the Evaluation of Machine Learning Algorithms, *Pattern Recognition*, 30, 1077, pp.1145-1159.
7. C. E. Metz and H. B. Kronman, Statistical Significance Tests for Binormal ROC Curves, *Journal of Mathematical Psychology*, 22, 1980, pp.218-243.
8. M. Betinec, Testing The Difference of the ROC Curves in Biexponential Model, *Tatra Mountains Mathematical Publications*, 39, 2008, pp.215-223.
9. B. Kaushal, K. Jain, and S. K. Sharma, Estimation of Area under Receiver Operating Characteristic Curve for Bi-Pareto and Bi-Two Parameter Exponential Models, *Open Journal of Statistics*, 4(1), 2014, pp.1-10.
10. N. L. Johnson, S. Kotz and A. W. Kemp, Univariate Discrete Distributions, (2nd ed.), John Wiley & Sons, New York, 1993.
11. J. P. Klein and M. L. Moeschberger, Survival Analysis Techniques for Censored and Truncated Data., Springer-Verlag, New York, 2003.