Statistics and Applications {ISSN 2454-7395 (online)}
Volume 21, No. 1, 2023 (New Series), pp 41–49

Estimation of AUC of Mixture ROC Curve in the Presence of Measurement Errors

G. Siva and R. Vishnu Vardhan
Department of Statistics, Ramanujan School of Mathematical Sciences, Pondicherry University, Puducherry, India

Received: 17 December 2021; Revised: 15 March 2022; Accepted: 18 March 2022

Abstract

Receiver Operating Characteristic (ROC) curve is one of the widely used classification tool and its applications can be seen in diversified fields of science and engineering. In this work, we made an attempt to examine the influence of measurement errors on the AUC of a mixture ROC curve. A bias corrected estimator is proposed and derived. The proposed work is supported with real dataset and simulation studies and results show that the proposed bias corrected estimator helps in correcting the AUC with minimum bias and minimum mean square error.

Key words: Mixture ROC curve; Area under the curve; Measurement errors.

1. Introduction

Over the years, classification problems have gained a lot of attention in terms of theoretical development and practical applications in various disciplines. To handle such problems, one of the classification tool is the Receiver Operating Characteristic (ROC) curve, originated during World War II for analyzing the radar images. In diagnostic medicine, ROC curve is widely used for evaluating the test's performance and also useful in comparing diagnostic tests by means of Area under the Curve (AUC) and Sensitivities. It is a unit square graphical plot between false positive rate (1-specificity) and true positive rate (sensitivity) at various threshold values. The AUC of an ROC curve plays an important role in assessing the performance of a diagnostic test(s) and also measures the ability of a biomarker to distinguish between two groups.

Measurement error (ME) problems are among the oldest in the history of statistics and can be of great practical and economic importance. It is the difference between a measured quantity and its true value. In diagnostic medicine, markers are subject to substantial measurement errors which may be attributed to instruments used in the laboratory, knowledge of the technicians, biological variability, temporal changes in subjects, *etc.* Shear *et al.* (1987) has taken the measurements of systolic and diastolic blood pressure on children's, which were used as forecasters of future hypertension. Carracio *et al.* (1995) has done a study on the children to predict the presence or absence of bacterial menengitis using cerebrospinal

Corresponding Author: R. Vishnu Vardhan
Email: vrstatsguru@gmail.com

fluid. Since the outcome of the test in identifying the menengitis is attributed to either laboratory equipment or technician, which may lead to the phenomenon of observing errors in the measured quantities. With the above two examples, it can be understood that in most of the data collections which purely depend upon the laboratory equipments, technicians etc. there are high chances of having errors in the measurements. For more examples on ME, readers can look into Begg and Greene (1983), Begg and McNeil (1988), Berbaum *et al.* (1989), Buonaccorsi (2010) and Fuller (2009).

In ROC analysis, the most popular one is the Bi-normal ROC model, where the two populations assumed to follow normal distribution. The estimation of AUC and its measures have been addressed by several authors and a few to mention are Hanley and McNeil (1982), Faraggi and Reiser (2002), Zhou *et al.* (2009), Vishnu Vardhan and Sarma (2010). However, when the data is exposed to measurement errors the estimation of AUC will be a problem of interest. Because as the measurements are deviated from their true value, it leads to produce spurious AUC. Hence, the AUC has to be corrected by means of an estimator. The seminal work on providing an estimator to correct the AUC was addressed by Coffin and Sukhatme (1996). They showed that in the presence of measurement errors, the AUC will be biased downwards and also came out with a bias corrected estimator that corrects the AUC. In similar lines, Faraggi (2000) and Reiser (2000) have worked on estimating the confidence intervals for the AUC in the presence of measurement error. Tosteson *et al.* (2005) studied the effect of measurement errors on AUC of an ROC curve by expressing the magnitude of the measurement error as a ratio of two variances; graphical and simulated environment were presented to show the effect of ME.

The above methodologies works well only when the knowledge on class labels is known. Even though the class labels are known, in most of the practical situations we may get observed with bi-modal or multi-model patterns within each known population. In such scenarios the existing binormal structure and correction of AUC in the presence of measurement error may not feasible to execute.

In this work, we proposed a Mixture ROC model which takes into the account of a possible mean differences between populations. Let us assume that two sub components are identified in diseased population and defined as D_1 and D_2. Now, we take into the account of the following possible mean differences such as $\mu_{D_1} - \mu_H$ and $\mu_{D_2} - \mu_{D_1}$ ($\mu_{D_2} \geq \mu_{D_1} \geq \mu_H$) and the same is shown in Figure 1. In Section 3, the same scenario is illustrated using OGTT dataset.

In section 2, we present the methodology of mixture ROC curves and its correction in measurement error. In section 3, a real dataset is considered to assess the performance of the proposed methodology and in section 4, monte carlo simulations are performed to compare the MSE of estimated and bias corrected estimator of the true AUC values. This has helped to examine how the bias and MSE of the estimators are influenced by measurement errors at different sample sizes.

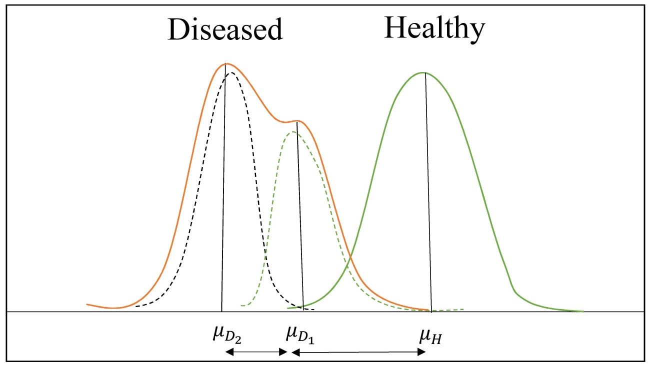

Figure 1: Hypothetically overlapping density curves of Healthy and Diseased populations

2. Methodology

2.1. Mixture receiver operating characteristic curve with measurement error

Let us consider the data where the class labels of subjects are known. In most of the cases, we directly start with developing a classifier rule. But there are chances of having several subgroups in each of the known populations. For instance, consider the oral glucose tolerance test (OGTT) data, where the subjects disease status is defined. However, on investigating the diseased population, it resulted with a bi-modal pattern. This indicates that there are two sub populations with in the diseased population (Figure 2).

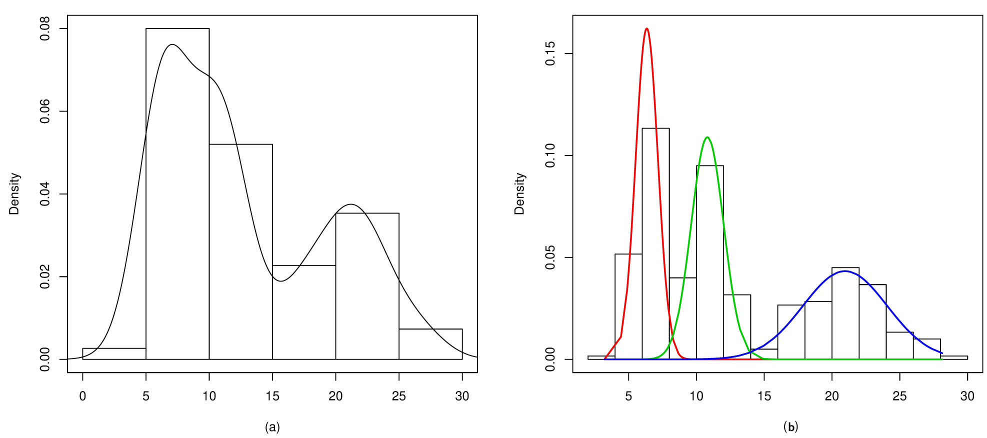

Figure 2: (a) The overall density plot of OGTT, (b) Plot after identifying the components in the OGTT data set.

Let us consider a binary classified data (Healthy, H and Diseased, D) where the D population consists of two sub populations within it. The identification of two sub populations (D_1 and D_2) will be done through EM algorithm. Let $\mu_H, \mu_{D_1}, \mu_{D_2}$ and $\sigma^2_H, \sigma^2_{D_1}, \sigma^2_{D_2}$ are the means and variances of three populations, respectively.

The expressions for the False Positive Rate (1-specificity) and True Positive Rate (sensitivity) in the mixture form is defined as

$$FPR = x(c) = \lambda_1\, x(c_1) + \lambda_2\, x(c_2) \tag{1}$$

$$TPR = y(c) = \lambda_1\, y(c_1) + \lambda_2\, y(c_2) \tag{2}$$

here, λ_1 and λ_2 are the mixing proportions; c_1 and c_2 are threshold values for the pairs (D_1, H) and (D_2, D_1).

By definition, we write

$$x(c_1) = \Phi\left(\frac{\mu_H - c_1}{\sigma_H}\right) \quad ; \quad x(c_2) = \Phi\left(\frac{\mu_{D_1} - c_2}{\sigma_{D_1}}\right) \tag{3}$$

$$y(c_1) = \Phi\left(\frac{\mu_{D_1} - c_1}{\sigma_{D_1}}\right) \quad ; \quad y(c_2) = \Phi\left(\frac{\mu_{D_2} - c_2}{\sigma_{D_2}}\right) \tag{4}$$

The expressions for c_1 and c_2 will take the following form

$$c_1 = \mu_H - \sigma_H \Phi^{-1}[x(c_1)] \quad ; \quad c_2 = \mu_{D_1} - \sigma_{D_2} \Phi^{-1}[x(c_2)] \tag{5}$$

where Φ^{-1} is the inverse cumulative distribution function of normal. The mixture ROC expression is derived by substituting (5) in (2) and is given in (6)

$$ROC = \lambda_1\left[\Phi\left(\frac{\mu_{D_1} - \mu_H}{\sigma_{D_1}} + \frac{\sigma_H}{\sigma_{D_1}}\Phi^{-1}[x(c_1)]\right)\right] + \lambda_2\left[\Phi\left(\frac{\mu_{D_2} - \mu_{D_1}}{\sigma_{D_2}} + \frac{\sigma_{D_1}}{\sigma_{D_2}}\Phi^{-1}[x(c_2)]\right)\right] \tag{6}$$

In general, if the diseased component has 'p' sub populations then (6) can be rewritten as

$$ROC(c) = \sum_{i=1}^{p} \lambda_i \left[\Phi\left(A_i + B_i\; \Phi^{-1}[FPR]\right)\right] \tag{7}$$

where $\sum_{i=1}^{p} \lambda_i = 1$; $A_i = \dfrac{\mu_i - \mu_{i-1}}{\sigma_i}$; $B_i = \dfrac{\sigma_{i-1}}{\sigma_i}$

2.2. Corrected bias approximation

Let us define $X_1, X_2, \ldots, X_m \overset{iid}{\sim} N(\mu_H, \sigma_H^2)$, $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(\mu_{D_1}, \sigma_{D_1}^2)$ and $Z_1, Z_2, \ldots, Z_k \overset{iid}{\sim} N(\mu_{D_2}, \sigma_{D_2}^2)$, then the AUC expression for mixture ROC curve is given as

$$mAUC = \theta = \lambda_1 \Phi\left(\frac{\mu_{D_1} - \mu_H}{\sqrt{\sigma_{D_1}^2 + \sigma_H^2}}\right) + \lambda_2 \Phi\left(\frac{\mu_{D_2} - \mu_{D_1}}{\sqrt{\sigma_{D_2}^2 + \sigma_{D_1}^2}}\right)$$

If the observations in H, D_1 and D_2 are observed with measurement errors then we define

$$x_i = X_i + u_i, \quad i = 1, 2, \ldots, m; \;\; u_i \sim iid\; N(0, \sigma_u^2)$$

$$y_i = Y_i + v_j, \quad j = 1, 2, \ldots, n; \;\; v_i \sim iid\; N(0, \sigma_v^2)$$

$$z_k = Z_k + \gamma_k, \quad k = 1, 2, \ldots, l; \;\; \gamma_k \sim iid\ N(0, \sigma_\gamma^2)$$

we assume u_i, v_j, z_k, X_i, Y_j and Z_k are all independent. The natural estimator of θ is

$$m\hat{AUC} = \hat{\theta} = \lambda_1\, \hat{\theta}_1 + \lambda_2\, \hat{\theta}_2$$

where $\hat{\theta}_1 = \Phi\left(\frac{\hat{\mu}_{D_1} - \hat{\mu}_H}{\sqrt{s_{D_1}^2 + s_H^2}}\right)$, $\hat{\theta}_2 = \Phi\left(\frac{\hat{\mu}_{D_2} - \hat{\mu}_{D_1}}{\sqrt{s_{D_2}^2 + s_{D_1}^2}}\right)$

here s_H^2, $s_{D_1}^2$ and $s_{D_2}^2$ are the sample variances. Using Taylor series expansion, it can be shown that $E(\hat{\theta}) = \theta + O(1)$. Since, the observations are measured with errors, the resulting area estimates i.e., AUC's will be biased downward. By adopting the methodology of Coffin and Sukhathme (1996), the expressions for $\hat{\theta_1}$ and $\hat{\theta_2}$ are

$$\begin{aligned} E(\hat{\theta}_1) &\approx P(Y > X + \delta_1) = \int\int [1 - G_Y(s+t)] f_X(s) f_{\delta_1}(t) dt ds \\ &\approx \theta_1 - \frac{1}{2} Var(\delta_1) \int g_Y^T(s) f_X(s) ds \\ E(\hat{\theta}_2) &\approx P(Z > Y + \delta_2) = \int\int [1 - G_Z(s+t)] f_Y(s) f_{\delta_2}(t) dt ds \\ &\approx \theta_2 - \frac{1}{2} Var(\delta_2) \int g_Z^T(s) f_Y(s) ds \end{aligned}$$

where $\delta_1 = u - v \sim N(0, \sigma_u^2 + \sigma_v^2)$ and $\delta_2 = v - \gamma \sim N(0, \sigma_v^2 + \sigma_\gamma^2)$, here $G_Y(.)$, $G_Z(.)$ are distribution functions of Y, Z and $f_{\delta_1}(.), f_{\delta_2}(.)$ are density functions of δ_1, δ_2. Thus, the approximate bias in using $\hat{\theta}_1$ and $\hat{\theta}_2$ to estimate θ will be

$$\begin{aligned} -B_1 &= -\frac{1}{2} Var(\delta_1) \int g_Y^T(s) f_X(s) ds \\ &= -\frac{\frac{1}{2}(\sigma_u^2 + \sigma_v^2)}{\sqrt{2\pi}\tau_{XY}^2} \left(\frac{\mu_{D_1} - \mu_H}{\tau_{XY}}\right) exp\left\{-\frac{1}{2}\left(\frac{\mu_{D_1} - \mu_H}{\tau_{XY}}\right)^2\right\} \\ -B_2 &= -\frac{1}{2} Var(\delta_2) \int g_Z^T(s) f_Y(s) ds \\ &= -\frac{\frac{1}{2}(\sigma_v^2 + \sigma_\gamma^2)}{\sqrt{2\pi}\tau_{YZ}^2} \left(\frac{\mu_{D_2} - \mu_{D_1}}{\tau_{YZ}}\right) exp\left\{-\frac{1}{2}\left(\frac{\mu_{D_2} - \mu_{D_1}}{\tau_{YZ}}\right)^2\right\} \end{aligned}$$

where $\tau_{XY} = \sqrt{\sigma_H^2 + \sigma_{D_1}^2}, \tau_{YZ} = \sqrt{\sigma_{D_1}^2 + \sigma_{D_2}^2}$, then the bias corrected estimator for θ in the mixture form is defined as

$$mAUC_{corr} = \theta^* = \lambda_1\, \theta_1^* + \lambda_2\, \theta_2^* \tag{8}$$

where $\theta_1^* = \hat{\theta}_1 + \hat{B}_1$ and $\theta_2^* = \hat{\theta}_2 + \hat{B}_2$. Using the unbiased estimates $\hat{\sigma}_u^2$, $\hat{\sigma}_v^2$ and $\hat{\sigma}_\gamma^2$, the estimated value of B_1 and B_2 will be

$$\hat{B}_1 = \frac{(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)}{2\sqrt{2\pi}(s_H^2 + s_{D_1}^2 - \hat{\sigma}_u^2 - \hat{\sigma}_v^2)} \left(\frac{\hat{\mu}_{D_1} - \hat{\mu}_H}{\sqrt{s_H^2 + s_{D_1}^2 - \hat{\sigma}_u^2 - \hat{\sigma}_v^2}}\right) exp\left\{-\frac{1}{2}\left(\frac{\hat{\mu}_{D_1} - \hat{\mu}_H}{s_H^2 + s_{D_1}^2 - \hat{\sigma}_u^2 - \hat{\sigma}_v^2}\right)^2\right\}$$

$$\hat{B}_2 = \frac{(\hat{\sigma}_v^2 + \hat{\sigma}_\gamma^2)}{2\sqrt{2\pi}(s_{D_1}^2 + s_{D_2}^2 - \hat{\sigma}_v^2 - \hat{\sigma}_\gamma^2)} \left(\frac{\hat{\mu}_{D_2} - \hat{\mu}_{D_1}}{\sqrt{s_{D_1}^2 + s_{D_2}^2 - \hat{\sigma}_v^2 - \hat{\sigma}_\gamma^2}}\right) exp\left\{-\frac{1}{2}\left(\frac{\hat{\mu}_{D_2} - \hat{\mu}_{D_1}}{s_{D_1}^2 + s_{D_2}^2 - \hat{\sigma}_v^2 - \hat{\sigma}_\gamma^2}\right)^2\right\}$$

The confidence intervals (CI) for corrected AUC measures are obtained using

$$\widehat{mAUC}_{corr} \pm Z_{\left(1-\frac{\alpha}{2}\right)}\ S.E(\widehat{mAUC}_{corr})$$

3. Real data set

The OGTT dataset (Lasko *et al.*, 2005) consists of 21 samples of Healthy and a mixture of Diseased individuals. In order to show the measurement error in the data, random error observations are generated $N(0, 1.2)$ and added to the original samples. This is done to mimic the situation where the actual data is affected with ME.

Along with the accuracy measures, it's bias and MSE's are obtained and presented in table (1). From the results, it is shown that by adding error observations to the original data, the accuracy measure is affected and biased downwards (i.e., from $\theta = 0.94626$ to $\hat{\theta} = 0.91641$). In such situation, the proposed bias corrected estimator helps in achieving the true accuracy and which has minimum bias and minimum MSE when compared with the estimated accuracy. The ROC curves are drawn for the original dataset (True ROC)

Table 1: Bias and MSE of estimated and corrected estimator of AUC of OGTT dataset

	$\hat{\theta}$ (True AUC)	$\hat{\theta}_{ME}$ (Uncorrected AUC)	Bias	MSE	$\hat{\theta}^*$ (Corrected AUC)	Bias	MSE
Mixture ROC	0.94626	0.91641	-0.02985	0.00089	**0.94061**	-0.00565	**0.00003**

and after adding error observations to the data (ROC with ME). From Figure 3, it is clearly seen that errors in measurement will affect the shape of the ROC curve and it is downwards than the true ROC curve.

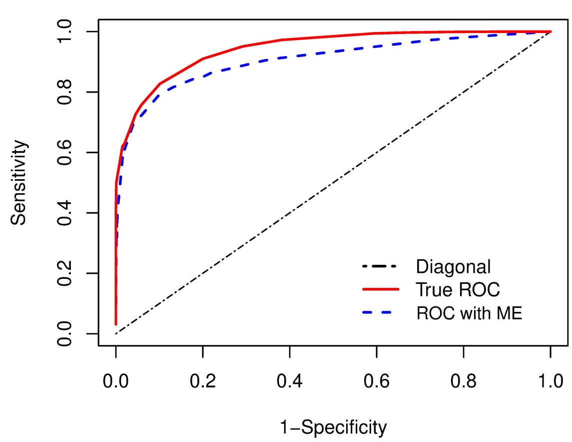

Figure 3: True and contaminated ROC (with ME) curves for OGTT dataset

4. Simulation studies

Monte Carlo simulations are carried out to illustrate the behavior of the proposed bias corrected estimator in the mixture ROC forms when the observations are measured with error.

In Table 2, two sets of means and variances are considered along with the initial values for mixing proportions. Set A and set B has unequal and equal variances, respectively. To show the influence of measurement errors in the data, the error component, $\epsilon \sim N(0, 1.9)$ is added to set A and B & AUC's are estimated (before and after correction). In each population, random samples of size $n = \{25, 50, 100, 200\}$ were generated using the parameter values listed in Table 2.

Table 2: Considered parameters for simulation studies

Sets	λ_1	λ_2	μ_H	μ_{D_1}	μ_{D_2}	σ_H	σ_{D_1}	σ_{D_2}
A	0.5	0.5	29.3	32.5	35.2	1.0	1.5	2.0
B	0.5	0.5	29.3	32.5	35.2	1.5	1.5	1.5

The estimated and bias corrected AUC values along with its bias and mean square errors at various sample sizes are presented in Table 3.

Table 3: The Bias, MSE of the estimated and bias-corrected estimator of AUC

Sets	$\hat{\theta}$	n	$\hat{\theta}_{ME}$ (CI_L,CI_U)	Bias	MSE	$\hat{\theta}^*$ (CI_L,CI_U)	Bias	MSE
A	0.91099	25	0.83777 (0.82228,0.85326)	-0.07322	0.01855	0.94898 (0.92649,0.97146)	0.03799	0.00144
		50	0.85403 (0.83667,0.87139)	-0.05696	0.01356	0.93654 (0.91489,0.95820)	0.02555	0.00065
		100	0.86416 (0.84844,0.87988)	-0.04683	0.01212	0.92650 (0.90554,0.94746)	0.01551	0.00024
		200	0.86922 (0.84916,0.88929)	-0.04177	0.01101	0.91561 (0.88365,0.94756)	0.00461	0.00002
B	0.91673	25	0.83488 (0.81679,0.85298)	-0.08149	0.01508	0.93858 (0.91483,0.96234)	0.02221	0.00049
		50	0.86306 (0.84866,0.87746)	-0.03150	0.00917	0.93614 (0.90054,0.97174)	0.01977	0.00039
		100	0.87897 (0.86447,0.89347)	-0.05331	0.00819	0.92754 (0.90604,0.94904)	0.01117	0.00012
		200	0.88487 (0.86705,0.90268)	-0.03740	0.00682	0.90702 (0.87457,0.93947)	-0.00935	0.00009

From the results, it is understood that the area estimates ($\hat{\theta}_{ME}$) are biased downward at each sample size. Using the proposed mixture of bias corrected approximation, it is observed that the bias corrected estimator of AUC's ($\hat{\theta}^*$) are closer to the true AUC's ($\hat{\theta}$) values and has minimum MSE when compared with the estimated AUC's ($\hat{\theta}$). Using the proposed methodology of bias corrected approximation in mixture ROC, we can obtain the reliable estimates of AUC's in the presence of measurement errors.

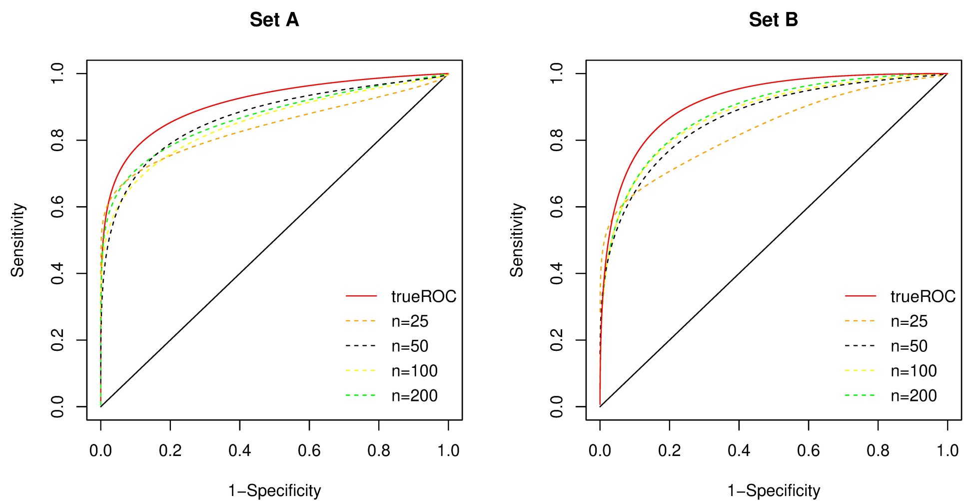

Figure 4: The true and estimated ROC curves at various sample sizes

The graphical representation of the true mixture ROC curve and the estimated mixture ROC curves (errors in the data) at various sample sizes is presented in Figure 4. From this graphical ROC plots also it is understood that, the resulting area estimates are downward in the presence of measurement errors.

5. Summary

In this paper, we made an attempt to address the problem of measurement errors in estimating the AUC of mixture normal ROC model. A bias corrected approximation has been defined in the mixture form. The methodology is supported by a OGTT dataset and monte carlo simulation studies. Results indicates that the proposed bias corrected estimator provides the corrected AUC's and it will be closer to the true AUC values with minimum bias and minimum MSE.

References

Begg, C. B. and Greenes, R. A. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics*, **39**, 207–215.

Begg, C. B. and McNeil, B. J. (1988). Assessment of radiologic tests: control of bias and other design considerations. *Radiology*, **167**, 565–569.

Berbaum, K. S., Dorfman, D. D. and Franken Jr, E. A. (1989). Measuring observer performance by ROC analysis: indications and complications. *Investigative Radiology*, **24**, 228–233.

Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications.* Chapman & Hall/CRC Press.

Carraccio, C., Blotney, K. and Franken, E. A. (1989). Cerebrospinal fluid analysis in systematically children without central nervous system disease. *Pediatrics*, **96**, 48–51.

Cheam, A. S., and McNicholas, P. D. (2016). Modelling receiver operating characteristic curves using Gaussian mixtures. *Computational Statistics & Data Analysis*, **93**, 192-208.

Coffin, M. and Sukhatme, S. (1997). Receiver operating characteristic studies and measurement errors. *Biometrics*, **53**, 823–837.

Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.

Gonen, M. (2013). Mixtures of receiver operating characteristic curves. *Academic Radiology*, **20**, 831–837.

Faraggi, D. (2000). The effect of random measurement error on receiver operating characteristic (ROC) curves. *Statistics in Medicine*, **19**, 61–70.

Fuller, W. A. (2009). *Measurement Error Models.* John Wiley & Sons.

McClish, D. K. (1992). Combining and Comparing Area Estimates across Studies or Strata. *Medical Decision Making*, **12**, 274–279.

Thomas A Lasko et al. (2005). The use of receiver operating characteristics curves in biomedical informatics. *Journal of Biomedical Informatics*, **38**, 404–415.

Perkins, N. J., Schisterman, E. F. and Vexler, A. (2009). Generalized ROC curve inference for a biomarker subject to a limit of detection and measurement error. *Statistics in Medicine*, **28**, 1841–1860.

Reiser, B. (2000). Measuring the effectiveness of diagnostic markers in the presence of measurement error through the use of ROC curves. *Statistics in Medicine*, **19**, 2115–2129.

Shear et al. (1987). Designation of children with high blood pressure—considerations on percentile cut points and subsequent high blood pressure: the Bogalusa Heart Study. *American Journal of Epidemiology*, **125**, 73–84.

Tosteson et al. (2005). Measurement error and confidence intervals for ROC curves. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, **47**, 409–416.

Vardhan, R. V., and Sarma, K. V. S. (2010). Estimation of the area under the ROC curve using confidence intervals of mean. *ANU Journal of Physical Sciences*, **2**, 29-39.

Zhou, X. H., McClish, D. K., and Obuchowski, N. A. (2009). *Statistical Methods in Diagnostic Medicine.* John Wiley & Sons.