Stochastic Modeling and Applications
Vol. 24 No. 2 (July-December, 2020)

Received: 15th September 2020 Revised: 19th October 2020 Accepted: 24th October 2020

ESTIMATING THE MEASURES OF MROC CURVE AND BOOSTING THE AUC IN THE PRESENCE OF MONOTONE MISSING DATA

DASHINA.P AND R.VISHNU VARDHAN*

Abstract. Receiver Operating Characteristic (ROC) curve is widely known and mostly applied classification tool for assessing the performance of diagnostic test(s) and also to provide better classification and allocation. In this paper, we present a situation of missing data with monotone pattern. This is dealt with some matrix decomposition approaches and the methodology is developed in Multivariate ROC curve framework. Using a real data set, the application of proposed work is illustrated.

1. Introduction

ROC Methodology got originated during world war II in analysing radar images. The theoretical development and its applications in the fields of psychophysical research, radiology, preventive medicine and many more dates back to early 1950's and 60's. It has been proved that the methodology of ROC has addressed many clinical issues, particularly in comparing several diagnostic tests. One of the objectives of ROC curve analysis is to classify subjects/individuals into one of the 'k' populations, in particular k=2, namely healthy (H) and diseased (D) categories. In order to have better correct classification, the cut-off which will be used for classification must be an optimal one. The overall accuracy (or) summary of any diagnostic test can be explained using Area Under the Curve(AUC), which lies between 0 and 1. Practically, any test's AUC should lie between 0.5 and 1. AUC = 0.5 indicates an imperfect test and AUC = 1.0 indicates a perfect test. ROC curve lies in a unit square plot with 1-specificity $(1 - S_p)$ on X-axis and sensitivity (S_n) on Y-axis.

The definition of ROC curve, sensitivity and specificity are usually given as,

$$ROC = 1 - G[F^{-1}(1 - x(t)] \tag{1.1}$$

where F and G are the distribution functions of H and D populations.
Sensitivity,

$$S_n = P\left(U > c \,|D\right) \tag{1.2}$$

2000 *Mathematics Subject Classification.* Primary 92B15; Secondary 62P10.
Key words and phrases. Monotone missing data, MROC curves, minimax principle, Cholesky decomposition.
* Corresponding Author.

and Specificity,

$$S_p = P\left(U \leq c \,|H\right) \tag{1.3}$$

where U is a test score, c is cut-off. If an individual's score $U > c$, it is categorized to population D , otherwise H.
As we have several co-ordinates of sensitivity and specificity, we will be having corresponding cutoffs. So, we need to identify an optimal threshold. In ROC analysis, optimal cutoff can be determined with the help of Youden's index (J) and is given as:

$$J = max[S_n + S_p - 1] \tag{1.4}$$

In the parametric form of ROC, the well known ROC form is the Binormal ROC curve, where the diseased(with condition) and healthy(without condition) underlie Normal distribution. In later years, the decision of classifying an individual into one of the two populations became essential with a set of variables rather than with a single variable. Hence it became a need to come out with a multivariate version of ROC curve, where the feature set is defined in the form of a linear combination. There are good number of articles on this multivariate extension of ROC corve, a few to mention are [6, 10, 11, 12, 13, 14]. We consider the multivariate ROC form proposed by [6], where the vector 'b' is obtained using *minimax* approach. In next section, the MROC curve methodology is given in detail and in subsequent section the concept of monotone missing data, estimation of μ and Σ , illustrations and discussions are presented.

Let us consider two multivariate normal random vectors X and Y with mean vectors μ_0, μ_1 and covariance matrices Σ_0 and Σ_1 respectively, i.e., $X \sim MVN(\mu_0, \Sigma_0)$ *and* $Y \sim MVN(\mu_1, \Sigma_1)$. Let x(c) denote the false positive rate (FPR) and y(c) denote the true positive rate (TPR) where "c" is the threshold value. The expressions for FPR, TPR and AUC derived by [6] are given below:

$$FPR = x(t) = \Phi\left(\frac{c - b^{'}\mu_0}{\sqrt{b'\Sigma_0 b}}\right) \tag{1.5}$$

where b ($\neq 0$) is a k x 1 vector and Φ denotes the cumulative distribution function of the normal distribution.

then the threshold can be expressed as,

$$t = b^{'}\mu_0 + \sqrt{b'\Sigma_0 b} \quad \Phi^{\text{-}1}(1 - x(t)) \tag{1.6}$$

then the true positive rate is obtained as,

$$TPR = y(t) = \Phi\left(\frac{b^{'}\mu_1 - c}{\sqrt{b'\Sigma_1 b}}\right) \tag{1.7}$$

and the MROC curve is given as,

$$TPR(fpr) = y(x) = \Phi\left(\frac{b'(\mu_1 - \mu_0) - \sqrt{b'\Sigma_0 b} \quad \Phi^{-1}(1 - x(t))}{\sqrt{b'\Sigma_1 b}}\right) \tag{1.8}$$

and the linear combination is defined as,

$$U = b^{'} X = b_1 x_1 + b_2 x_2 + ... + b_k x_k \tag{1.9}$$

where b is obtained using the minimax procedure as,

$$b = [t\,\Sigma_1 + (1 - t)\,\Sigma_0]^{-1}(\mu_1 - \mu_0) \tag{1.10}$$

Here t takes values between 0 and 1 and should be determined by trial and error method.

In practice, presence of missing or incomplete data is a very common problem. There are many types of missing data and how to tackle with those different types of missing in data is very crucial in the data analysis. In such situation, the parameter estimates may not be reliable and consistent. In any classification problem, if we come across such missing data pattern, then it is very important to address it first and then a classifer rule can be determined. Confining this discussion to the MROC methodology, we wish to focus on estimating the μ and Σ of two classes if such missing pattern is witnessed. Hence, we came out with a solution to estimate μ and Σ of H and D if data is of monotone missing pattern without any imputation. The nested structure of monotone data allows explicit derivation of maximum likelihood estimators and likelihood ratio test statistics for the mean μ and covariance matrix Σ. Therefore it has received a special attention in the literature [1]. The maximum likelihood estimates of parameters with various patterns of missing data with simple mathematical manipulation can be read from [3]. Further, in the same paper, matrix differentiation and matrix transformations were used to derive the maximum likelihood estimates of the means under rank constraint and of the covariances when the observations are missing. Closed forms were obtained for the maximum likelihood estimators of the mean vector and the covariance matrix of a multivariate normal model with a k-step monotone missing data pattern [2] using matrix derivatives as an extension of 2-step missing data pattern [3]. However, some simple test procedures for obtaining the maximum likelihood estimators of the mean vector and covariance matrix were given by [1]. Fewer studies are reported in literature where the situation of monotone pattern is dealt in both univariate and multivariate cases. There are good number of articles that provide realistic situations where one can notice the monotone missing pattern. A few to mention are [1, 15, 16, 17, 18]. None of the studies dealt with a scenario where monotone missing data is present in a classification problem. In this paper, we wish to propose the methodology of estimating the measures of MROC curves such as sensitivity, Specificity, AUC where, the data has the nature of monotone missing. Methodology is supported with a numerical illustration.

In this paper we adopted the method of [1] in obtaining the mean vector and covariance matrix in a classification scenario when there is a presence of monotone missing pattern, instead of using most common procedures like imputation in the

presence of missing data. To demonstrate the practical importance and application of the proposed methodology, we created a monotone missing pattern in both H and D populations. All these exercises are made on a real data set namely IUGRFDS [5]. In next section, we detail out what actually monotone missing data means and how μ and Σ will be estimated. Further, few matrix decomposition techniques are used to boost the AUC and other measures.

2. Monotone Missing Data

Let $X_{px1} \sim MVN(\mu, \Sigma)$, here μ and Σ are unknown; Σ is a positive definite matrix. Let X be partitioned as $(x_1', x_2', \ldots., x_k')'$ such that x_i is p_i x 1 vector, $i = 1, 2, .., k$ and $p_1 + p_2 + \ldots + p_k = p$. Partition the mean vector and covariance matrix accordingly. Consider a random sample of N_1 independent observations from the above distribution that has the following pattern.
$x_{11}, \ldots, x_1N_k, \ldots, x_1N_2, \ldots, x_1N1$
$x_{21}, \ldots, x_2N_k, \ldots, x_2N_2$
.....
$x_{k1}, \ldots, x_kN_k$
which is known as Monotone or Triangular pattern, that is, N_i observations are available on $p_1 + \ldots + p_i$ components, where i = 1, 2, ..., k.

Estimation of the parameters of Monotone Data

Maximum Likelihood Estimates given by [1] are:

$$\hat{\mu} = \begin{bmatrix} \hat{\mu_1} \\ \hat{\mu_2} \\ \hat{\mu_3} \end{bmatrix} \; and \quad \hat{\Sigma} = \begin{bmatrix} \hat{\Sigma_{11}} & \hat{\Sigma_{12}} & \hat{\Sigma_{13}} \\ \hat{\Sigma_{21}} & \hat{\Sigma_{22}} & \hat{\Sigma_{23}} \\ \hat{\Sigma_{31}} & \hat{\Sigma_{32}} & \hat{\Sigma_{33}} \end{bmatrix} \tag{2.1}$$

From the monotone pattern, for the first N_1 observations, we get $\overline{x}_{1,1}$ and $S^{(1)}$, and for N_2 observations, we get the sample mean vector

$$\overline{x}_2 = \begin{bmatrix} \overline{x}_{1,2} \\ \overline{x}_{2,2} \end{bmatrix} \tag{2.2}$$

and the sample covariance matrix

$$S^{(2)} = \begin{bmatrix} S_{11,2} & S_{12,2} \\ S_{21,2} & S_{22,2} \end{bmatrix} \tag{2.3}$$

and for N_3 observations, we get the sample mean vector

$$\overline{x}_3 = \begin{bmatrix} \overline{x}_{1,3} \\ \overline{x}_{2,3} \\ \overline{x}_{3,3} \end{bmatrix} \tag{2.4}$$

and the sample covariance matrix

$$S^{(3)} = \begin{bmatrix} S_{11,3} & S_{12,3} & S_{13,3} \\ S_{21,3} & S_{22,3} & S_{23,3} \\ S_{31,3} & S_{32,3} & S_{33,3} \end{bmatrix} \tag{2.5}$$

$$\hat{\mu_1} = \overline{x}_{1,1} \tag{2.6}$$

$$\hat{\mu_2} = \overline{x}_{2,2} - B_{21}(\overline{x}_{1,2} - \hat{\mu_1}) \tag{2.7}$$

$$\hat{\mu_3} = \overline{x}_{3,3} - [B_{31}(\overline{x}_{1,3} - \hat{\mu_1}) + B_{32}(\overline{x}_{2,3} - \hat{\mu_2})] \tag{2.8}$$

$$\hat{\Sigma_{11}} = \frac{S^{(1)}}{N_1} \tag{2.9}$$

$$\hat{\Sigma_{2.1}} = \frac{(S_{22,2} - B_{21}S_{12,2})}{N_2} \tag{2.10}$$

$$\hat{\Sigma_{22}} = \hat{\Sigma_{2.1}} + B_{21}\hat{\Sigma_{12}} \tag{2.11}$$

$$\hat{\Sigma_{33}} = \hat{\Sigma_{3.21}} + B_{31}\hat{\Sigma_{13}} + B_{32}\hat{\Sigma_{23}} \tag{2.12}$$

$$(\Sigma_{31}, \Sigma_{32}) = (B_{31}, B_{32}) \begin{bmatrix} \hat{\Sigma_{11}} & \hat{\Sigma_{12}} \\ \hat{\Sigma_{21}} & \hat{\Sigma_{22}} \end{bmatrix} \tag{2.13}$$

$$(B_{l1}, ..., B_{l\,\overline{l-1}}) = (S_{l1,l}, ..., S_{l\,\overline{l-1},l}) \begin{pmatrix} S_{11,l} & \cdots & S_{1\,\overline{l-1},l} \\ \vdots & \ddots & \vdots \\ S_{\overline{l-1}\,1,l} & \cdots & S_{\overline{l-1}\,\overline{l-1},l} \end{pmatrix} \tag{2.14}$$

where l = 1, 2, ..., k.
$S^{(1)}$, $S^{(2)}$ and $S^{(3)}$ are unbiased estimates of Σ at each N_i.

Further, an attempt is made to decompose the matrices obtained with N_2 and N_3 observations. The most popularly used matrix decomposition methods such as Cholesky's and Singular Value have been considered to decompose $S^{(2)}$ and $S^{(3)}$ matrices.

Cholesky decomposition or Cholesky factorization is a useful decomposition method for efficient numerical solutions, e.g., Monte Carlo simulations. Let A be a Hermitian positive-definite matrix and it's Cholesky's decomposition is A = [L][L]$^{\mathrm{T}}$. Here L and L$^{\mathrm{T}}$ denote the lower triangular matrix and conjugate transpose of L respectively.

SVD is a matrix decomposition that applies to any matrix, real, or complex. The Singular Value decomposition is a powerful tool for many matrix computations because it reveals a great deal about the structure of a matrix.
Singular value decomposition is a method of decomposing a matrix into three other matrices :

A= U S V$^{\mathrm{T}}$ where,
A is an m x n matrix
U is an m x n orthogonal matrix
that is; U$^{\mathrm{T}}$U = I

S is an n x n diagonal matrix
V is an n x n orthogonal matrix
that is; $V^T V=I$.

In next section, the above explained methodology is illustrated using a real data set.

3. Results and Discussion

For the numerical illustration, Intra Uterine Growth Restricted Fetal Doppler Study (IUGRFDS) dataset is collected from Sri Venkateswara Medical College, Tirupati by [6], which is a tertiary cum care hospital meets the above criterion where a procedure namely MCA is used to identify whether the blood flow from the womb of the mother to the baby is sufficient enough for its growth. The dataset consists of 82 samples in which 42 (n_0) are healthy and 40 (n_1) are diseased. Three indices are used to observe the growth of baby namely Pulsatility Index (PI), Resistivity Index (RI) and Systolic/Diastolic (S/D) ratio. Some observations were randomly removed from the from variables RI and S/D of MCA procedure, to obtain a dataset that exhibits monotone missing pattern. Now n_0 and n_1 of the monotone data are 34 and 32 respectively. Mean vector and Covariance matrix were computed using the methods described by [1].

In Table 1, the covariance matrices $S^{(2)}$ and $S^{(3)}$ are given and their decomposed forms using Cholesky's decomposition and Singular Value decomposition methods are shown in Table 2,3 and 4. Table 5 depicts the mean vectors and covariance matrices of H and D populations obtained using method given by [1], Cholesky's decomposition and Singular Value decomposition. In Table 6, the measures of MROC curve are reported. In Figure 1, the MROC curve generated through complete data, monotone missing data, Cholesky's decomposition and Singular Value decomposition are depicted. From the result of Table 5, it is understood that the adopted method is able to provide the estimates of $\hat{\mu}$ and $\hat{\Sigma}$ with small deviations in decimal values when compared with the $\hat{\mu}$ and $\hat{\Sigma}$ of complete data. The AUC values obtained through monotone missing data is almost similar to that of the complete data AUC. However, the experiment to decompose the $S^{(2)}$ and $S^{(3)}$ matrices in $\hat{\Sigma}$ of monotone missing data has shown some better results. The AUC value through Cholesky decomposition is much better than AUC of complete data, whereas, Singular Value decomposition attained similar AUC values as that of complete data. So, even in estimating $\hat{\mu}$ and $\hat{\Sigma}$ under monotone missing pattern, better estimates can be obtained when we further decompose the covariance matrix. Of the Cholesky decomposition and Singular Value decomposition, the first method is considered to be a better one in providing similar estimates as that of complete data and also boosts the AUC. Apart from this, the 1-specificity value observed using Cholesky decomposition is far better than any other method with reasonably good sensitivity. The $(1-S_p)$ obtained using Cholesky's decomposition is about 18% where as, other methods possessed higher false positive values. In terms of ROC curves also, the curve obtained by Cholesky decomposition has better concavity than that of monotone missing data, Singular value decomposition

and complete data.

TABLE 1. Sample covariance matrices of monotone data

Method	$\mathbf{S^{(2)}}$	$\mathbf{S^{(3)}}$
Healthy	$\begin{pmatrix} 1.597 & 0.795 \\ 0.795 & 6.457 \end{pmatrix}$	$\begin{pmatrix} 1.163 & 0.993 & 2.491 \\ 0.993 & 6.315 & 2.301 \\ 2.491 & 2.301 & 6.047 \end{pmatrix}$
Diseased	$\begin{pmatrix} 1.006 & -0.137 \\ -0.137 & 6.174 \end{pmatrix}$	$\begin{pmatrix} 0.854 & 0.031 & 2.409 \\ 0.031 & 5.870 & -0.454 \\ 2.409 & -0.454 & 9.361 \end{pmatrix}$

TABLE 2. Decomposed matrices of $S^{(2)}$ using CD and SVD methods

Methods	**Cholesky Decomposition** $\mathbf{LL^T}$		**Singular Value Decomposition** $\mathbf{USV^T}$		
Healthy	$\begin{pmatrix} 1.264 & \\ 0.629 & 2.462 \end{pmatrix}$	$\begin{pmatrix} 1.264 & 0.629 \\ & 2.462 \end{pmatrix}$	$\begin{pmatrix} 0.157 & -0.987 \\ 0.987 & 0.157 \end{pmatrix}$	$\begin{pmatrix} 6.58 & \\ & 1.469 \end{pmatrix}$	$\begin{pmatrix} 0.157 & 0.987 \\ -0.987 & 0.157 \end{pmatrix}$
Diseased	$\begin{pmatrix} 1.003 & \\ -0.136 & 2.481 \end{pmatrix}$	$\begin{pmatrix} 1.003 & -0.136 \\ & 2.481 \end{pmatrix}$	$\begin{pmatrix} -0.026 & -0.999 \\ 0.999 & -0.026 \end{pmatrix}$	$\begin{pmatrix} 6.17 & \\ & 1.002 \end{pmatrix}$	$\begin{pmatrix} -0.026 & 0.999 \\ -0.999 & -0.026 \end{pmatrix}$

TABLE 3. Decomposed matrices of $S^{(3)}$ using Cholesky Decomposition

Method	**Cholesky Decomposition** $\mathbf{A = LL^T}$
Healthy	$\begin{pmatrix} 1.163 & 0.993 & 2.491 \\ 0.993 & 6.315 & 2.301 \\ 2.491 & 2.301 & 6.047 \end{pmatrix} = \begin{pmatrix} 1.078 & & \\ 0.921 & 2.338 & \\ 2.311 & 0.074 & 0.837 \end{pmatrix} \begin{pmatrix} 1.078 & 0.921 & 2.311 \\ & 2.388 & 0.074 \\ & & 0.837 \end{pmatrix}$
Diseased	$\begin{pmatrix} 0.854 & 0.031 & 2.409 \\ 0.031 & 5.870 & -0.454 \\ 2.409 & -0.454 & 9.361 \end{pmatrix} = \begin{pmatrix} 0.924 & & \\ 0.034 & 2.423 & \\ 2.607 & -0.224 & 1.586 \end{pmatrix} \begin{pmatrix} 0.924 & 0.034 & 2.607 \\ & 2.423 & -0.224 \\ & & 1.586 \end{pmatrix}$

TABLE 4. Decomposed matrices of $S^{(3)}$ using Singular Value Decomposition

Method	**Singular Value Decomposition** $\mathbf{A = USV^T}$
Healthy	$\begin{pmatrix} 0.295 & 0.247 & 0.922 \\ 0.650 & -0.759 & -0.004 \\ 0.699 & 0.601 & -0.385 \end{pmatrix} \begin{pmatrix} 9.24 & & \\ & 4.16 & \\ & & 0.1 \end{pmatrix} \begin{pmatrix} 0.295 & 0.650 & 0.699 \\ 0.247 & -0.759 & 0.601 \\ 0.922 & -0.004 & -0.385 \end{pmatrix}$
Diseased	$\begin{pmatrix} 0.251 & 0.051 & 0.966 \\ -0.103 & 0.994 & -0.025 \\ 0.962 & 0.092 & -0.255 \end{pmatrix} \begin{pmatrix} 10.04 & & \\ & 5.83 & \\ & & 0.21 \end{pmatrix} \begin{pmatrix} 0.251 & -0.103 & 0.962 \\ 0.051 & 0.994 & 0.092 \\ 0.966 & -0.025 & -0.255 \end{pmatrix}$

TABLE 5. Mean vectors and Covariance matrices obtained from Different Methods

	Healthy Population				**Diseased Population**			
Methods	μ_0	Σ_0			μ_1	Σ_1		
CompleteData	1.034	0.055	0.013	0.116	1.148	0.034	−0.009	0.092
	1.162	0.013	0.172	0.036	1.323	−0.009	0.187	−0.037
	1.092	0.116	0.036	0.275	1.454	0.092	−0.037	0.321
MonotoneData	1.034	0.054	0.027	0.115	1.148	0.033	−0.005	0.094
	1.193	0.027	0.177	0.063	1.316	−0.005	0.177	−0.029
	1.099	0.115	0.063	0.266	1.438	0.094	−0.029	0.347
Cholesky Decomposition	1.034	0.054	0.027	0.093	1.148	0.033	−0.005	0.093
	1.193	0.027	0.192	−0.098	1.316	−0.005	0.071	−0.022
	1.046	0.093	−0.098	0.303	1.437	0.093	−0.022	0.313
Singular Value Decomposition	1.034	0.054	0.027	0.153	1.148	0.033	−0.005	0.121
	1.193	0.027	0.177	0.208	1.316	−0.005	0.176	0.067
	1.121	0.153	0.208	0.465	1.436	0.121	0.067	0.300

TABLE 6. Measures of MROC

Methods	**AUC Measure**	**cutoff c**	**sensitivity** S_n	**1-specificity 1-**S_p	**Youden's Index J**
Complete Data	0.71	1.546	0.65	0.35	0.30
Monotone Data	0.68	2.317	0.79	0.54	0.24
Cholesky Method	0.77	1.893	0.65	0.18	0.46
SVD Method	0.70	0.737	0.96	0.80	0.15

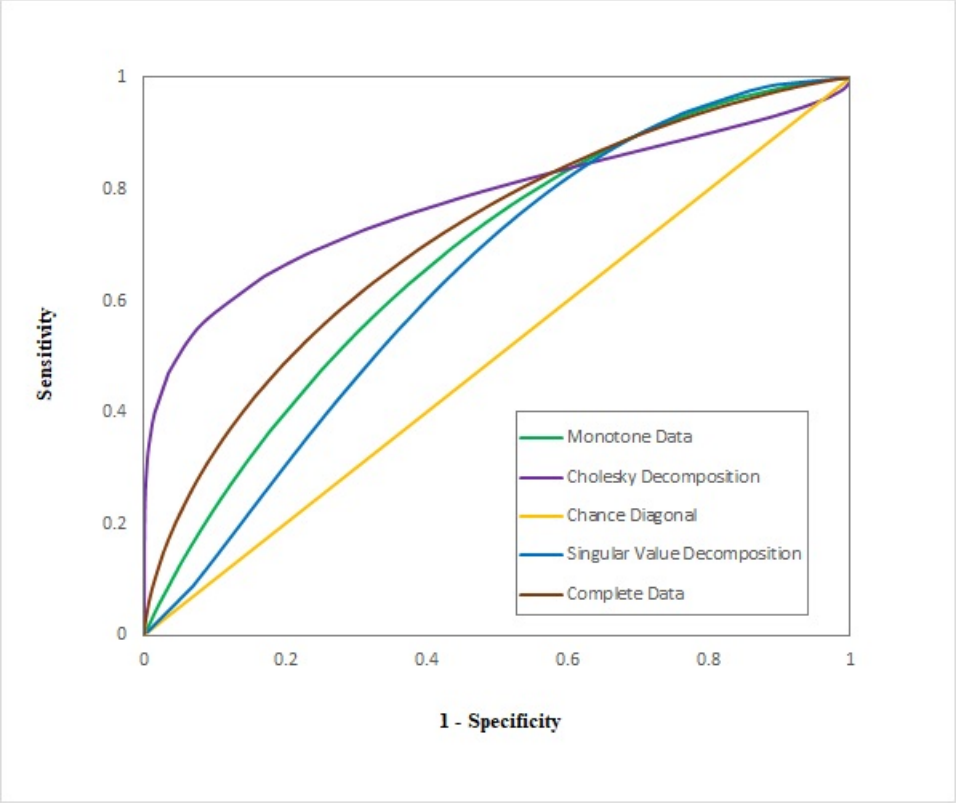

FIGURE 1. ROC Curves for all the methods

4. Summary

In this paper, an attempt is made in estimating the mean vectors and covariance matrices in the case where data is observed to have monotone missing pattern and considering this scenario in a classification framework. We adapted the method proposed by [1] and further changes are made in decomposing the covariance matrix using well known Cholesky's and Singular Value decomposition methods. Results depict that information in terms of $\hat{\mu}$ and $\hat{\Sigma}$ can be well retained and closer values can be attained as that of $\hat{\mu}$ and $\hat{\Sigma}$ of complete data. Further, it is shown that, on using Cholesky decomposition, the true information in terms of AUC can be explained in better manner. When compared with AUCs, the AUC obtained through Cholesky decomposition is much better than with low

1-specificity and reasonably good sensitivity. Hence, if we have observed missing cases in a data, using the methods discussed in the paper, the parameter estimates and measures of MROC curve can be estimated in a proper manner and can provide better results in terms of accuracy and $1 - S_p$.

References

1. Krishnamoorthy, K and Pannala, Maruthy K. : Some simple test procedures for normal mean vector with in complete data,*Annals of the Institute of Statistical Mathematics* **50**(1998) 531–542.
2. Jinadasa, KG and Tracy, DS.: Maximum likelihood estimation for multivariate normal distribution with monotone sample, *Communications in Statistics-Theory and Methods* **21** (1992) 41–50.
3. Anderson, Theodore Wilbur and Olkin, Ingram.: Maximum-likelihood estimation of the parameters of a multivariate normal distribution, *Linear algebra and its applications* **70** (1985) 147–171.
4. Anderson, Theodore W.: Maximum likelihood estimates for a multivariate normal distribution when some observations are missing, *Journal of the american Statistical Association* **52** (1957) 200–203.
5. Vardhan, R Vishnu and Sameera, G and Chandrasekharan, PA and Beere, Thulasi.: Inferential procedures for comparing the accuracy and intrinsic measures of multivariate receiver operating characteristic (MROC) Curve *International Journal of Statistics in Medical Research* **4** (2015) 87–93.
6. Sameera, G and Vardhan, R Vishnu and Sarma, KVS.: Binary classification using multivariate receiver operating characteristic curve for continuous data *Journal of biopharmaceutical statistics* **26** (2016) 421–431.
7. Gupta, Arjun K and Nagar, Daya K.: *Matrix variate distributions*, **104**, CRC Press, 2018.
8. Anderson, Theodore W and Bahadur, Raghu Raj and others.: Classification into two multivariate normal distributions with different covariance matrices, *The annals of mathematical statistics* **33** (1962) 420–431.
9. Lipschutz, Seymour and Lipson, Marc.: *Schaum's outline of theory and problems of linear algebra*, Erlangga, 2001.
10. Su, John Q and Liu, Jun S.: Linear combinations of multiple diagnostic markers, *Journal of the American Statistical Association* **88**(1993) 1350–1355.
11. Reiser, Benjamin and Faraggi, David.: Confidence intervals for the generalized ROC criterion, *Biometrics* (1997) 644–652.
12. Yuan, Zheng and Ghosh, Debashis.: Combining multiple biomarker models in logistic regression, *Biometrics* **64** (2008) 431–439.
13. Yin, Jingjing and Tian, Lili.: Optimal linear combinations of multiple diagnostic biomarkers based on Youden index, *Statistics in medicine* **33**, Wiley Online Library, 2014.
14. Liu, Aiyi and Schisterman, Enrique F and Zhu, Yan.: On linear combinations of biomarkers to improve diagnostic accuracy, *Statistics in medicine* **24** (2005) 37–47.
15. Bhargava, RP.: Some one-sample hypothesis testing problems when there is a monotone sample from a multivariate normal population, *Annals of the Institute of Statistical Mathematics* **27** (1975) 327–339.
16. Yagi, Ayaka and Seo, Takashi and Srivastava, Muni S.: Testing equality of mean vectors in a one-way MANOVA with monotone missing data, *Communications in Statistics-Theory and Methods* **47** (2018) 5534–5546.
17. Little Roderick, JA and Rubin Donald, B.: *Statistical analysis with missing data*, Hoboken, NJ: Wiley, 1987.

18. Cook, R Dennis and Weisberg, Sanford.: *An introduction to regression graphics*, John Wiley & Sons, 1994.

DASHINA.P: Department of Statistics,Pondicherry University,Pondicherry,India
Email address: dashinap3@gmail.com

R.VISHNU VARDHAN: Department of Statistics,Pondicherry University,Pondicherry,India
Email address: vrstatsguru@gmail.com