BAYESIAN SEMI-PARAMETRIC AND NON-PARAMETRIC ESTIMATION OF RECEIVER OPERATING CHARACTERISTIC (ROC) SURFACE

BY

KIPRONO BEN KOECH

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS IN THE SCHOOL OF SCIENCE, UNIVERSITY OF ELDORET, KENYA

NOVEMBER, 2015

DECLARATION

Declaration by the Candidate

This thesis is my original work and it has not been submitted for any academic award any institution; and shall not be reproduced in part or full, or in any format without prior written permission from the author and/or University of Eldoret

Kiprono Ben Koech Signature:…… Date:……

SC/PGM/067/11

DECLARATION BY SUPERVISORS

This thesis has been submitted with our approval as University Supervisors.

Dr. Argwings Otieno Signature:……..…… Date: ……

University of Eldoret

Dr. Victor K. Kimeli Signature:…… Date:……

University of Eldoret.

DEDICATION

I dedicate this thesis to my parents William and Grace Soi, my sister Rodah, my brothers Japheth and Daudi and the science fraternity.

ABSTRACT

Receiver operating characteristic curve analysis is widely used in biomedical research to assess the performance of diagnostic tests. Estimation of receiver operating characteristic curves based on parametric approach has been widely used over years. However, this is limited by the fact that distribution of almost all diseases in epidemiology cannot be established quite easily. Bayesian methods are robust as it allows computability and the distributions based on this are flexible. Therefore, inference based on parametric distributions can be either misleading or insufficient. There is need for generalization of the receiver operating characteristic curve (since, the analysis largely assumes that test results are dichotomous) to allow tests to have more than two outcomes. The receiver operating characteristic curve was generalized to constitute a surface, which uses volume under the surface (VUS) to measure the accuracy of a diagnostic test. Dirichlet process mixtures of normals and Mixtures of Finite Polya Trees, which are robust models that can handle nonstandard features in data in modelling the diagnostic data, were used to model the test outcomes. The models proved to address difficulties in modelling continuous diagnostic data with skewness, multimodality, or other nonstandard features. Semiparametric and Nonparametric models for receiver operating characteristic surface estimation were fitted using Markov Chain Monte Carlo with simple Metropolis Hastings steps. The mixing parameters, means and variances were updated with random-walk type proposals centred at some definite values. The Semi-parametric and Nonparametric, parametric approaches were considered for estimating the receiver operating characteristic surface'svolume under the surface (VUS). Simulation results indicate that even when the parametric assumption holds, these models give accurate results as the volume under the surface (VUS) for both methods were greater than 1/6, the value of a "useless test" . Graphically, the semiparametric receiver operating characteristic surface has the appealing feature of being continuous and smooth, thus allowing for useful interpretation of the diagnostic performance at all thresholds. Similarly, the non-parametric methods lead substantially to the same conclusions. In summary, to overcome the strict assumptions of parametric models, Bayesian semi-parametric model involving Dirichlet process mixtures of normals as well as non-parametric model that involve mixtures of finite Polya trees can be applied for Receiver Operating Characteristics surface estimation as they both have desirable performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS/ ACRONYMS

ADC:	AIDS Dementia Complex
AIDS:	Acquired Immunodeficiency Syndrome
AUC:	Area Under Curve
CI:	Credibility Intervals/ Confidence Intervals
DPM:	Dirichlet Process Mixtures
EFF:	Efficiency of a Test
FPR:	False Positive Rate
FPR:	False Positive Rate
HIV:	Human-Immunodeficiency Virus
HUM:	Hypervolume Under ROC Manifold
i.i.d:	Independent and Identically Distributed
IAE:	Intergrated Absolute Error
KDE:	Kernel Density Estimate
MCMC:	Markov Chain Monte Carlo
MFPT:	Mixtures of Finite Polya Trees
NPV:	Negative Predictive Values
PAUC:	Partial Area Under Curve
PPV:	Positive Predictive Values
ROC:	Receiver Operating Characteristic
ROCS:	Receiver Operating Characteristic surface
TPR:	True Positive Rate
TTR:	True Transition Rate
VUS:	Volume Under ROC Surface

ACKNOWLEDGEMENTS

Foremost, I thank God for His favours, mercies and blessings throughout my studies. Secondly, would like to deferentially pass my sincere gratitude to my supervisors, Dr. Otieno, and Dr. Kimeli, due to their resolute guidance in the development and implementation of the research methodology. I highly acknowledge their advice and academic materials toward writing this thesis.

I also feel greatly appreciative to my parents and siblings for their relentless encouragements. Thank you very much and God Bless you

Furthermore, I acknowledge the invaluable support and discussions on the development and implementation of the methodologies from my course and research mate, Mr. King'ang'i Morris. Not forgetting to thank all my classmates, colleagues and friends who kept own advising and encouraging me to continue with the work.

Finally yet importantly, I would like to thank those I have not mentioned for their prayers and efforts are recognized and greatly appreciated only paper and time are limiting.

God bless you all.

CHAPTER ONE

INTRODUCTION

1.1 General background

Several fields such as computer science, meteorology, biochemistry and medical studies use statistical methodologies in classification and formulation of predictive inferences (Zhou et al, 2002). For instance in medical studies, the selection of the statistical methodology applied to the prediction and classification as well as the diagnosis of the status of a subject is of utmost importance. In particular, the accurate and timely diagnosis of a patient's condition is crucial to the ultimate treatment of the diseased condition. Detecting these conditions and evaluating the prediction of patients with disease can be achieved by analysing the clinical and laboratory data (Hanley & McNeil, 1982). Inaccurate diagnoses in many real-world biomedical settings carry emotionally stressful and financial consequences.

Results of classification in a diagnostic test can indicate the presence or absence of the specific disease-related material or can yield an entire array of non-binary results. In the case of non-binary ordinal (subjective) or continuous scales, a threshold value can be used for classification such that with results above or below such threshold classified as positive or negative for disease, as appropriate. For example, in cancer patients in which the progression of the disease is relatively fast, determining the stage of the disease is crucial to applying the appropriate treatment, and earlier detection of the stage of the disease can vastly increase survivability of the patient via the appropriate medical prognosis (Pepe, 2003).

Complexity in the classification of a diagnostic test varies depending on technological and procedural outlook. From a procedural perspective, the test may only involve

one-step, which results in one of only two outcomes, positive or negative, or it may involve a vast sequence of procedures that may result in one of an entire spectrum of possible classifications.

Diagnostic tests in the technological point of view may be a classic bacterial culture test, or it can be a complex application employing the latest in genetic sequencing technologies (Pepe et al., 2001). One of the main criteria that should be considered before implementing a diagnostic test is the accuracy of the test. An accurate test is one that correctly classifies its test population according to the disease or non-disease condition. Inaccurate tests cause those with actual disease to be misclassified as non-diseased, also known as a "false negative". Conversely, they cause those with no actual disease to be misclassified as diseased, also known as a "false positive". False negative errors leave diseased subjects untreated. False positive errors open subjects to being subjected to unnecessary procedures and emotional stress. Both false negatives and false positives may also create disillusionment and distrust within the general subjects towards the medical and diagnostic testing community as a whole, potentially making data collection more difficult, biased and costly (Zhou et al, 2002).

In practice, such errors must be kept to a minimum. As such, the accuracy of a diagnostic test is of utmost importance and must be thoroughly assessed and understood before it can be used. Thorough evaluation of the population of interest, the test itself and outcomes of the test is required to effectively assess a diagnostic test (Kraemer, 1992).

1.2 Diagnostic test accuracy measures

The accuracy is a test's ability to identify a condition correctly when the disease or condition is truly present and to exclude the disease or condition when it is actually

absent. The accuracy of a test is always measured by comparing the test results to the true condition status. Binary test results, that is, the true condition status is either "the condition is present" or "the condition is absent" are assumed.

1.2.1 Sensitivity and Specificity

Sensitivity and specificity are two basic measures of diagnostic accuracy. The contingency table, Table 1 can be used to illustrate the two definitions. Firstly, denote the true disease or condition status by the indicator variable *D,* where

$$D = \begin{cases} 1 & \text{with disease or condition}; \\ 0 & \text{without disease or condition}; \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Further, denote the result of the diagnostic test by the indicator variable *T.* Test results indicating the disease or condition's presence are called *positive,* denoted as $T = 1$, whereas those indicating the disease or condition's absence are called *negative,* denoted as $T = 0$, where

$$T = \begin{cases} 1 & \text{postive test results}; \\ 0 & \text{negative test results} \end{cases} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Table 1 below illustrates a basic count table specifying the different numbers under different categories

Table 1: Basic Count Table

	True disease or condition status		
Test results	Present(D=1)	Absent(D=0)	Total
Positive (T=1)	n_{11}	n_{12}	$n_{1.}$
Negative(T=0)	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	N

In table 1, above the total numbers with and without the condition are $n_{.1}$and $n_{.2}$ respectively. The total numbers with the condition whose test result is positive and negative are $n_{11}$and n_{21}respectively. The total numbers without the

condition whose test result is positive and negative are n_{12} and n_{22}, respectively.

The total number in the study is N, where

$$N = n_{11} + n_{12} + n_{21} + n_{22}$$

Sensitivity (Se) is the test's ability to detect the condition when the condition is present. Sensitivity is the probability that the test result is positive (T = 1), given the presence of the disease or condition (D = 1), written as:

$$Se = P(T = 1|D = 1) \quad \text{……………………...……………..}(3)$$

$$Se = n_{11}/n_{\cdot 1}$$

Specificity (Sp) is the test's ability to exclude the condition without the condition. It is the probability that the test result is negative (T = 0), given the absence of the disease or condition (D= 0), written as

$$Sp = P(T = 0|D = 0) \quad \text{…………………….……………..…}(4)$$

$$Sp = n_{22}/n_{\cdot 2}$$

The data can also be summarize by probabilities. The consequences associated with the test results are also considered. The test can have two types of errors. One is false positive errors and the other is false negative errors. The true positive rate (TPR) and false positive rate (FPR) are defined respectively as:

True Positive Rate = $TPR = P(T = 1|D = 1)$……………………………..…(5)

False Positive Rate = $FPR = P(T = 1|D = 0)$……………………………..…(6)

False negative fraction (FNF) is 1-TPF. True negative fraction (TNF) is 1-FPF. Under which, sensitivity is known as the TPR and specificity is known as TNR.

Under various applications, the terminology for TPR and FPR is often different. In biomedical research, the 'sensitivity' (TPR) and 'specificity' (1-FPR) are often descriptors of test performance (Cai et al., 2003).

1.2.2 Predictive values

The accuracy of a diagnostic test can also be quantified by how well the test results predict the true condition status. As such, another important measure of a diagnostic test *is*predictive value. The predictive values depend on the prevalence of the condition, such as in a disease condition. The predictive values are:

Positive Predictive Value = PPV = $P(D = 1|T = 1)$,……………..…….(5)

Negative Predictive Value = NPV = $P(D = 0|T = 0)$……………………(6)

A perfect test is one that predicts the condition perfectly. That is, PPV=1 and NPV=1. Contrarily, a useless test is one with no information about the true condition status. As such, a test, which does not reflect the true condition status, very well will result in a low PPV. The predictive values can tell us how likely the condition is given the test result. The values are affected by the prevalence of the condition. Low prevalence of the condition may be a reason for a low PPV. In research studies, both the classification probability (TPF and FPF) and the predictive values are important and there is a direct relationship between the two (Pepe et al., 2001). Suppose the prevalence is $\rho = P(D = 1)$. A result can be directly ascertained from the Bayes' theorem:

$$PPV = \frac{\rho TPR}{\rho TPR + (1-\rho)FPR} \quad \text{……} \quad (7)$$

$$NPV = \frac{(1-\rho)(1-FPF)}{(1-\rho)(1-FPR) + \rho(1-TPR)} \quad \text{……} \quad (8)$$

1.3 Receiver Operating Characteristics Curve Analysis

1.2.1 Receiver Operating Characteristics curves

Receiver operating characteristic (ROC) curves are useful for assessing the accuracy of medical diagnostic tests (Pepe, 2003). Green and Swets (1966) developed signal detection theory in psychophysics, which appeared to be a potential method for medical diagnostic testing.

Lusted (1971) pointed out that this method could be adopted for medical decision making and stated that the method could overcome limitations of a single sensitivity and specificity pairs. Since then, this method has been the most valuable and popular tool for describing and comparing diagnostic tests, particularly in medicine. Making predictions has become an essential part in disease diagnosis.

Analysis of Receiver Operating Curve examines the relationship between sensitivity and specificity of a binary diagnostic test. Sensitivity or true positive rate measures the proportion of positives correctly classified; specificity or true negative rate measures the proportion of negatives correctly classified. Normally, the true positive rate TPR is plotted against the false positive rate FPR, which is one minus true negative rate.

Below is a simple illustration of ROC graph.

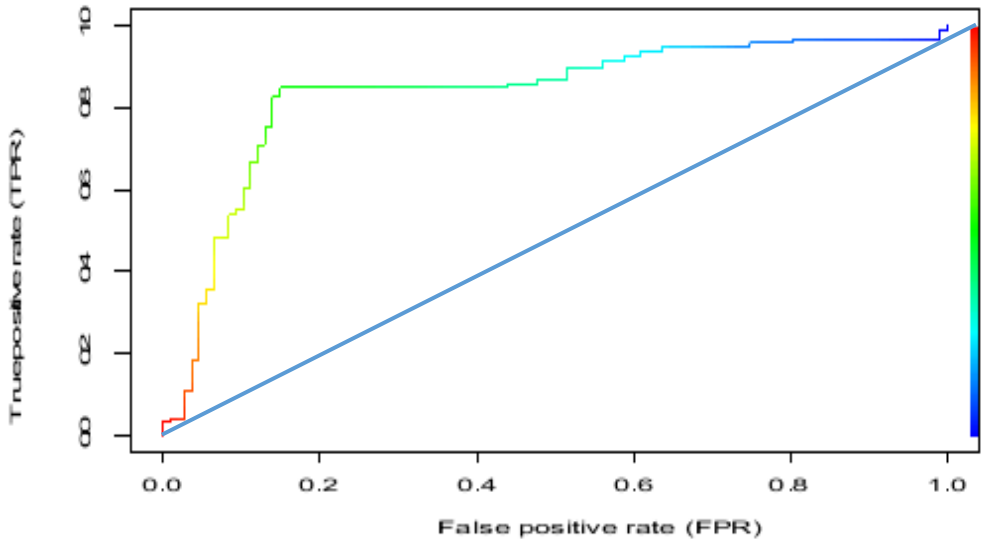

Figure 1: ROC graph (Source: Researcher)

The graph in Figure 1 above shows TPR rate plotted on the Y-axis against FPR is plotted on the X-axis. The ROC graph depicts relative trade-offs between true positives and false positives.The ROC plot has many advantages compared to other measures of accuracy (Zweig & Campbell, 1993). An ROC curve can visually represent the data's accuracy. The scales of the ROC curve plot are two basic measures of accuracy, which can be easily read from the plot.

The ROC curve includes all the possible decision thresholds so that there is no requirement to select a particular decision threshold. Because sensitivity and specificity are independent of prevalence, the ROC curve is independent of prevalence as well (Dwyer 1997).

The ROC curve is also independent of the scale of the test results. That is, the ROC curve does not vary to any monotonic (e.g., linear, logarithmic) transformations of the test results, which is a useful property (Campbell, 1994). Another advantage of the

ROC curve is that it can provide a direct and visual comparison of two or more tests on a single set of scales. It is possible to compare different tests at all decision thresholds by constructing the ROC curves.

1.3.2 Measuring the accuracy of a diagnostic test

Some summary indices associated with the ROC curve are often used to summarize the accuracy of a diagnostic test and provide important information about the ROC curve. When the ROC curve is not feasible to plot, such summary measures can also provide important information about the ROC curve. Area under the ROC curve (AUC) and partial area under the ROC curve (PAUC) are two important summary indices, which are particularly useful in certain situations.

1.3.3 Area under the ROC curve (AUC)

ROC curve is a useful measure to summarize the accuracy of a diagnostic test. Another valuable measure associated with the ROC curve is the area under the ROC curve (AUC). The area under the ROC curve takes values between 0 and 1.

A perfect diagnostic test is one with an area under the ROC curve of 1 and consists of two line segments: (0,0)-(0,1) and (0,1)-(1,1). In contrast, a test with an area of 0 is perfectly inaccurate. However, perfect diagnostic tests are rare. The area under the ROC curve can be interpreted as the average of sensitivity for all possible values of specificity. It can also be interpreted as the average value of specificity for all possible values of sensitivity.

The area under the ROC curve is a widely used summary measure for comparing ROC curves which is defined as in Bamber (1975)

$$AUC = \int_0^1 sensitivity\ d(1 - specificity) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (9)$$

1.3.4 Partial area under the ROC curve

Another summary measure associated with the ROC curve is the partial area under the ROC curve (PAUC). There is particular interest in the area under a portion of the ROC curve. The partial area under the ROC curve is the area between two sensitivities, which is defined as

$$PAUC(t_0) = \int_0^{t_0} ROC(t)dt \ldots\ldots\ldots\ldots \quad (10)$$

where $t_0 \epsilon$ (0,1). Its values range from $t_0^2/2$ for a completely uninformative test to t_0 for a perfect test.

Dwyer (1997) interpreted the partial area under the ROC curve as the probability that a randomly chosen subject without the condition will be classified correctly from a randomly chosen subject with the condition who tested negative in a diagnostic test. The partial area of test performance is appealing for some special cases and is well established in many clinical tests.

Generally, ROC curve is the main tool used in ROC analysis. It can be used to address a range of problems, including:

1. Determining a decision threshold that minimizes error rate or misclassification cost under given class and cost distributions.

2. Identifying regions where one classifier outperforms another.

3. Identifying regions where a classifier performs worse than chance.

4. Obtaining calibrated estimates of the class posterior.

In most of the recent diagnostic studies, the evaluation of diagnostic accuracy in

ordered three-class problems have suggested the use of receiver operating characteristic surfaces as a direct generalization of the ROC curve.

ROC surfaces have been proposed for the evaluation of diagnostic accuracy in ordered three-class classification problems as a direct generalization of the ROC curve (Nakas & Yiannoutsos, 2004).

1.4 Receiver Operating Characteristic Surface

Let $\{X_{1i}|\ i$=1,....,$n_1\}$ and $\{X_{2j}|\ j$=1,....,n_2 } be two test outcomes for " Healthy" and "Diseased" populations, respectively. Then the estimated AUC can be expressed as:

$$\widehat{AUC} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} P(X_{1i} \leq X_{2j}) \ldots\ldots \quad (13)$$

This scheme can be generalized to 3-class case as follows: Let the three classes be "Healthy (H)", "Transition (T)" and "Diseased (D)". Further, let the associated test outcomes be X_1 , X_2 and X_3 for "Positive (diagnosed D)", "Transition (diagnosed T)" and "Negative (diagnosed H)" respectively. C_1 and C_2 are used as the cut off points where ($C_1 \leq C_2$) , to obtain a similar table as 2-class case (table 1). This can be summarized in table 2 below.

Table 2: Classification of diagnostic test outcomes into three groups

		True disease or condition status			
		Healthy	Transition	Diseased	**TOTAL**
Test results	Positive	n_{11}	n_{12}	n_{13}	$n_{1\cdot}$
	Transition	n_{21}	n_{22}	n_{23}	$n_{2\cdot}$
	Negative	n_{31}	n_{32}	n_{33}	$n_{3\cdot}$
TOTAL		$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	N

It follows then that

$$\textit{True Positive Rate} = P\ (diagnosed\ H \mid H) = P\ (X_1 \leq C_1)$$

$$\textit{True Transition Rate} = P\ (diagnosed\ T \mid T) = P\ (C_1 \leq X_2 \leq C_2)$$

$$\textit{True Negative Rate} = P\ (diagnosed\ D \mid D) = P\ (X_3 \geq C_2)$$

If the various misclassifications are ignored, that is. no specificities are concerned, then theVUS can be expressed as, $VUS = P\ (X_1 \leq X_2 \leq X_3)$(Nakas & Yiannoutsos 2004).Consider a diagnostic test in which subjects can be classified into three different ordered categories. A continuous diagnostic test for all subjects can be performed and their test results can be recorded.

Let $\mathbf{Y_1} = (Y_{11}, Y_{12}, \ldots, Y_{1n_1})^T$denote test results on n_1 subjects from Class 1; $\mathbf{Y_2} = (Y_{21}, Y_{22}, \ldots, Y_{2n_2})^T$ denote test results on $n_2$subjects from Class 2; and

$\mathbf{Y_3} = (Y_{31}, Y_{32}, \ldots, Y_{3n_3})^T$denote test results on n_3 subjects from Class 3.

The test results for the individuals from the three classes are modelled according to the distribution F_1, $F_2$and F_3; continuous probability distributions on $\mathbb{R}^n$.

The test results Y_{1i}($i = 1,\ldots,n_1$) are independent and identically distributed. with distribution F_1; the test results Y_{2j} $(j = 1,\ldots,n_2)$ are independent and identically distributed. with distributions F_2; and the test results Y_{3k} $(k = 1,\ldots,n_3)$ are independent and identically distributed with distributions F_3. The test resultsY_{1i}, Y_{2j}andY_{3k} are mutually exclusive or independent to each other.

A decision rule, based on specifying two ordered decision thresholds$C_1 < C_2$, which classifies each subject into one of the three classes can be defined as follows:

IF $T_i < C_1$,, THEN assign subject to Class 1.

ELSE IF $C_1 < T_i < C_2$ assign subject to Class 2.

ELSE assign to Class 3.

The procedure is repeated for $i = 1, \ldots, n_1 + n_2 + n_3$to make diagnostic decisions for all subjects (Nakas & Yiannoutsos, 2004). The probabilities of correct classification into three classes for the pair of thresholds (C_1, C_2) can be computed.

$p_1 = \mathbb{P}(Y_{1i} \leq C_1) = F_1(C_1)$ Correct classification of Class 1,

$p_2 = \mathbb{P}(C_1 < Y_{2j} \leq C_2) = F_2(C_2) - F_2(C_1)$ Correct classification of Class 2,

$p_3 = \mathbb{P}(Y_{3k} > C_2) = 1 - F_3(C_2)$ Correct classification of Class 3.

A three-dimensional space forming the ROC surface is obtained when the probabilities (p_1, p_2, p_3) are plotted for all cutoff points (C_1, C_2), with$C_1 < C_2$, in the support of the diagnostic marker measurements.

By writing p_2-the correct classification probability for the intermediary or transition class**,** as a function of $p_1$and p_3,the equation for defining an ROC surface for the test can be written as:

$$ROCS(p_1, p_3) = \begin{cases} F_2(F_3^{-1}(1 - p_1)) - F_2(F_1^{-1}(p_3)), if\ F_1^{-1}(p_3) \leq F_3^{-1}(1 - p_1), \\ 0 \qquad \text{otherwise.} \end{cases}$$

…… *(14)*

From the definition above, for both arguments the map $(p_1, p_3) \mapsto$ ROCS (p_1, p_3) is monotone non-increasing. ROC curve for binary test is a representation of the relationship between the correct classification probabilities for two classes (sensitivity and specificity). Similarly, ROC surface represents the three-way trade-off among the correct classification probabilities for the three classes. According to Scurfield (1996), two other equivalent forms of defining ROC surface is the use of $p_1$orp_3 as functions of the two other elements. The current form of expression is chosen for algebraical

simplicity to presenting the results. However, the statistical inferences remain invariant for the three choices. Below is an illustration of ROC surface.

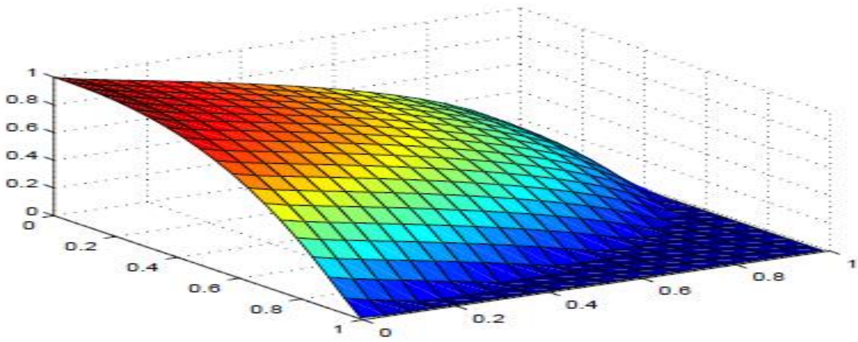

Figure 2: Receiver Operating Characteristics Surface (Nakas & Yiannoutsos, 2004)

According to Gu et al. (2008) the accuracy of the estimation of the entire ROC surface can be measured by the integrated absolute error (IAE).

$$\mathrm{IAE} = \int_0^1 \int_0^1 |\,\widehat{\mathrm{ROCS}}\,(p_1,\ p_3) - \mathrm{ROCS}(p_1,\ p_3)|\mathrm{d}p_1\mathrm{d}p_3 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(15)}$$

Where $\widehat{\mathrm{ROCS}}(p_1,\ p_3)$ is an estimate of $\mathrm{ROCS}(p_1,\ p_3)$

In summary, the ROC surface and VUS are two measures which are extensions of the two-class ROC curve and AUC.

1.4.1 Volume under the ROC surface

Volume under the ROC surface (VUS) be considered as a measure of accuracy of a diagnostic test in a single summary value. This can be defined as:

$$VUS = \int_0^1 \int_0^1 ROCS\ (p_1, p_3)\ dp_1 dp_3 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(16)}$$

Mossman (1999) outlines the generalization of the two dimensional area under the ROC curve by representing VUS mathematically to an equivalent probability defined as $P(Y_{1i} < Y_{2j} < Y_{3k})forY_{1i},\ Y_{2j}$ and Y_{3k} randomly selected from the three classes,

respectively.

It is noted that when the distributions F_1, $F_2$and F_3 are identical or completely overlap, hence a useless test then the volume under surface is equivalent to 1/6 and the value 1 when the three populations are perfectly discriminated in the anticipated ordering (Mossman, 1999). In two-category classification, rejecting the null hypothesis that AUC is equal to 1/2 would imply that the test is able to differentiate between the two classes with a probability higher than that of a random guess (Pepe, 2003).

For a three-category classification, it is required that the test to have at least some ability to differentiate three categories instead of only two categories. If the null hypothesis that VUS is equal to 1/6 is rejected, it can only be argued that the test is not the one that completely guesses the three classes. In fact, the test with a VUS greater than 1/6 might be able to differentially pick out one class but completely guess the other two classes (Mossman, 1999).

In that case, the test is still useless for a three-category classification and cannot be recommended for use. For any three-category classifier, it has several pairwise AUCs. Tests with any of these pairwise AUC values being too close to 1/2 should be screened out. The lower bound of VUS in three-category ROC analysis should be jointly considered with the lower bound of AUC in pairwise two-category ROC analysis (Obuchowski, 2005).

1.5 Problem statement

Receiver operating characteristic curve analysis is widely used in biomedical research to assess the performance of diagnostic tests. However, cases whereby test outcomes are more than two requires generalisation of the receiver operating characteristic

curve to constitute a surface, which uses volume under the surface to measure the accuracy of a diagnostic test. Constructing three-dimensional ROC surfaces to evaluate accuracy of continuous diagnostic tests over years have been widely based on parametric approaches.

There is vast literature on pureparametric approaches to estimation of three-dimensional ROC surfaces. However, in medical practice, the distribution of most diseases especially in epidemiology cannot be established quite easily. Therefore, strict assumptions of parametric distributions for the diseases can result in misleading or insufficient inference.

Parametric models are often not sufficiently flexible to capture skewness, multimodality, or other nonstandard features of the data. Therefore, there is necessity of implementation of alternative models such as Bayesian methods which have proved to be robust as it allows computability and the distributions based on this are flexible. In the past little had been discussed on Bayesian semiparametric and non-parametric methods for modelling diagnostic tests due to computational issues associated with the intractable distributions. However, with the advancement of computational tools, modelling of diagnostic data can be achieved through the Bayesian methods.

In particular, this research finds the need to adopt the semi parametric based on DPM and non-parametric estimation based on MFPT of three-dimensional ROC surfaces to overcome strict assumptions of parametric modelling of diagnostic data.

1.6 Justification

The volume under the three-class ROC surface has extensive applications in various areas since it provides a global measure of differences between or among populations.

Specifically, in biomedical diagnosis, it is imperative to have desirable levels of performance for diagnostic markers. Cancer studies for instance, has been an area under focus as the disease prevalence is notably on the rise. Therefore, study will be worthwhile in providing insight to the performance of the diagnostic tests and subsequent inferences.

1.7 Objectives

1.7.1 General objective

The broad objective of the study was to estimate the ROC surface based on Bayesian semi parametric and non-parametric methods.

1.7.2 Specific objectives

1. To estimate the ROC surface under semi parametric estimation procedures.
2. To estimate the ROC surface under non-parametric estimation procedures.
3. To compare the estimated ROC surfaces under both semi parametric estimation and non-parametric estimation procedures.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The measure of accuracy of a test introduced is often based upon decision thresholds, which may be difficult to detect. Lusted (1971) illustrated a way in which it could overcome the limitation of a single sensitivity and specificity pair, which he first applied to psychophysics. Lusted argued that the method could overcome the limitation by considering all of the decision thresholds. By applying the receiver operating characteristic (ROC) curve, the accuracy of a diagnostic test can be described without the limitations of decision thresholds. Lusted stated that ROC curves offer an ideal means of examining the performance of the diagnostic tests. Subsequently, the ROC curve has been the most valuable and most widely used tool to describe and compare diagnostic tests in various disciplines of medicine.

2.2 Receiver Operating Characteristics Curve

ROC curve, is a plot of the sensitivity of a diagnostic test versus the false-positive rate. ROC curves were originally developed for electronic signal-detection theory (Peterson et al, 1954). ROC curves and ROC analysis have subsequently formed the basis of statistical decision theory, having been applied to various medical and nonmedical studies, including studies of human perception (Drury & Fox, 1975) and military monitoring (Swets, 1977).

In medical diagnostic testing, the main interest is measuring the observer's abilities for interpreting test results rather than the criteria used for such decisions. As such, Lusted (1971) discussed how in medical diagnostics, a distinction must be made between the observer's cognitive and sensory abilities to interpret the test results for

detecting the condition and the observer's criteria used in deciding whether a condition is present or absent.

Swets and Pickett (1982) discussed how ROC curves display all possible cut points and thus can estimate the frequency of various outcomes at each cut point. Furthermore, ROC curves can apply previously generated probabilities of the condition, as well as calculations of the costs and benefits of correct and incorrect decisions, to determine the optimum cut point. They were also the first to study the analysis of multi reader studies in which several observers interpret the test results of the same sample of patients. They identified several sources of variability, as well as correlations in multi reader studies and then created a methodology for estimating and comparing the test accuracy for such studies.

The first to use the Gaussian model for estimating the ROC curve were Green and Swets (1966). They assumed the numerical value of a sensory event (defined as X) affects the observer's confidence about whether the condition is present or absent. They also assumed a cut point (defined as t) such that if $X < t$ and $X > t$, then the observer will choose the hypothesis that the condition is absent and present, respectively. Additionally, they assumed the Gaussian distribution of t under each hypothesis.

Furthermore, Dorfman and Alf (1968, 1969) proposed maximum-likelihood estimates for the parameters of a binormal ROC curve, and provided methodologies for obtaining the variance-covariance matrix and the corresponding confidence intervals.

The most widely used summary measure for the test accuracy of ROC analysis is the area under the ROC curve (AUC). Hanley and McNeil (1982) provided a relatively simple methodology to estimate AUC without having to assume the distribution of the

test results. Interestingly, they noted that AUC is equivalent to the Wilcoxon 2-sample test statistic. They developed a method for calculating sample size for studies that apply the ROC curve area. Several other nonparametric methodologies have subsequently been developed for estimating and comparing ROC curves.

McClish (1989) stated that AUC was a global measure of a test's accuracy. He provided parametric methods for estimating and comparing the partial area under the ROC curve. These parametric methods are based upon a binormal model and parallel the MLEs of the area under the total ROC curve. Many statistical methods were developed shortly after these investigations for the estimation of the ROC analysis for two-way classification.

2.3 Three-Way ROC

Many real-world classification problems involve more than just two categories and the extension of the two-way ROC analysis is needed. Scurfield (1996) first mapped the mathematical definition of a proper ROC measure for more than two categories. These diagnoses have an ordered gradation of illness from notdiseased to seriously illFor example, cognitive function declines from normal function to mild impairment, tosevere impairment or dementia. Another example would be the stage of cancer progression at the time of detection, from localized cancer through distant metastases alreadypresent.

There is need for statistical methods are needed for the assessment of diagnostic accuracy whenthe true disease status has an ordinal scale. This would necessitate the dichotomization of the binary gold standard so that the existing methods for binary classification can beapplied subsequently.

In the recent past, ROC methodology was then extended to multiple-class diagnostic problems by introducing a three-dimensional ROC surface. Mossman (1999) introduced the concept of three-class ROC analysis into medical decision making. Nakas and Yiannoutsos (2004) were the first to consider the estimation of the volume under the ROC surface for ordered three-class problems by using U-statistic theory. They also considered the volume under the ROC surface(VUS), and its relation to the probability of correctly ordered observationsfrom the three groups. They studied a continuous diagnostic test of neuropsychological battery to detect the presence of HIV-related cognitive dysfunction (AIDS dementia complex (ADC)), which generates three test outcomes: unimpaired, ADC stage 0.5, and ADC stage 1 to 3.

The multiple test outcomes would render the conventional ROC analysis unable to assess the accuracy of the test. In the three-class diagnostic problem studied by Nakas and Yiannoutsos (2004), a diagnostic marker resulting in continuous measurements is used for the discrimination of three classes of patients from three distinct possibly overlapping distributions of marker measurements. Here, a usual assumption is that subjects from class 3 tend to have higher measurements than subjects from class 2 and the latter tend to have higher measurements than subjects from class 1.

The three-group ROC surface generalizes the popularttwo-group ROC curve, which in recent years has attracted much theoreticalattention and has been widely applied for analysis of accuracy of diagnostictests. For these diagnostic tasks, one potential solution is todichotomize the gold standardat one or two cut-off points so that the existing methods for binary classification canbe applied. However, Obuchowski (2005) has shown that creating this artificial binarygold standard can induce a bias in the estimation of the test's accuracy.

2.4 ROC analysis

The difficulty in generalizing the ROC curve to more than two disease classes results fromthe fact that a decision rule for a K-group classification will produce K true class ratesand K(K-1) false class rates. Li and Fine (2008) further proposed the estimation of the volume under the ROC surface (VUS) and the hypervolume under the ROC manifold (HUM). They also provided the estimation of the multiple-class ROC measures and applied the multiple-class ROC analysis as a model of selection criterion in microarray studies.

Li and Zhou (2009) considered non-parametric and semi parametric estimation of the ROC surfaces by approximating the asymptotic ROC surfaces with multivariate Brownian bridge processes.

In medical research, it is also important to evaluate the various factors that can influence the medical performance. Great interest has been shown in developing methods for combining biomarkers. Statistical regression analysis has recently been studied to make inferences about such factors and biomarkers.

Han (1987) originally developed the maximum rank correlation estimator (MRC), which was considered as a generalized regression model of nonparametric analysis. It has recently been applied to assess classifications because of its close relationship to the ROC curve. Optimization algorithms that maximize the area under the ROC curve have also recently been proposed. Pepe (2003) developed optimal prognostic scores by applying binary regressions. The optimal linear combination is attained from several available diagnostic biomarkers from which it is seeked to maximize the area

under the ROC curve among all the possible linear combinations in the binary data analysis.

Enrique et al. (2004) suggested how to obtain the confidence interval for the generalized ROC criterion, conditional on given covariate values and derived some inferences under the normal distribution assumption. Theory of the consistency of the optimal confidence interval is based upon the argument which comes from Sherman (1993), relying on a general method for establishing the limiting distribution of a maximization estimator.

Models for uncertain data distributions based on mixtures of standard components, such as normal mixtures, underlie mainstream approaches to density estimation, including kernel techniques, nonparametric maximum likelihood, and Bayesian approaches using mixtures of Dirichlet processes (Ferguson 1974). The latter provide theoretical bases for more traditional Semiparametric and non- parametric methods, such as kernel techniques, and hence a modeling framework within which the various practical problems of local versus global smoothing, smoothing parameter estimation, and the assessment of uncertainty about density estimates may be addressed.

Semiparametric and non- parametric approaches, a formal model allows these problems to be addressed directly via inference about the relevant model parameters. These issues are discussed using data distributions derived as normal mixtures in the framework of mixtures of Dirichlet processes, essentially the framework of Ferguson (1983). West (1990) discussed these models in a special case of the framework by developing approximations to predictive distributions based on a clustering algorithm motivated by the model structure and draws obvious connections with kernel approaches.

For single test designs, Erkanli et al. (2006) used truncated Dirichlet process mixture models and Branscum et al. (2006) developed mixture of finite Polya trees models for Bayesian nonparametric ROC data analysis when true infection status is unknown. Bayesian parametric multivariate ROC methodology was developed recently by Choi et al. (2006). Hall and Zhou (2003) developed a multivariate distribution-free frequentist approach.

The utility of Polya trees (PT) and mixtures of Polya trees (MPT) priors for providing robust, flexible inferences in statistical modelling has been demonstrated within several prominent domains of statistics including linear regression (Hansonand Johnson, 2002), generalized linear models (Hanson, 2006). Most applications to date have involved modelling univariate data because, unlike Dirichlet process priors, the initial development of PT priors focused on continuous distributions supported on the real line (Ferguson, 1974). Heckerling (2001), proposed a simple parametric frequentist approach under the assumption that test results all follow normal distributions.

Li and Zhou (2009) develop a frequentist nonparametric and semiparametric approach. The nonparametric approach is based on the empirical counterparts of the distribution functions of the test results in each group, whereas the semiparametric approach attempts to generalize a parametric (normal) functional form of the ROC surface. However, this latter approach, as pointed out by the authors, relies heavily on the normality assumption.

Parametric models are often not sufficiently flexible to capture skewness, multimodality, or other nonstandard features of the data. Inácio et al. (2011) proposed a Bayesian nonparametric approach that uses a mixture of finite Polya trees (MFPT)

model to estimate the ROC surface. Bayesian nonparametric models allow for broadening the class of models under consideration, and hence for a widely applicable approach that can be used for practically any population and for a large number of diseases and diagnostic measures. Particularly, an important feature of Polya tree priors, besides that they can accommodate most forms of data, is that they can include a parametric distribution in the larger non-parametric family. This generalization has the potential to make the inference robust to departures from an assumed parametric distribution while still having good performance if the actual distribution is the parametric one.

The research seeked to complement the modelling of diagnostic data of the test measurement using a three-sample density ratio model. The advantage of applying density ratio models into the ROC surface analysis is that it not only allows estimation of the ROC surface semi parametrically but also enables implementation of the method easy as the usual procedures in many statistical software packages can be employed.

CHAPTER THREE

SEMI-PARAMETRIC AND NONPARAMETRIC ESTIMATION

3.1 Introduction

This chapter defines the methods to be used in the estimation of the parameters under semi parametric and non-parametric estimations. It also outlines the conditions and assumptions therein.

3.2 Semi-parametric Estimation using DPM of normals model

Suppose that the normal means and variances, πcome from some prior distribution $G(\cdot)$ on $\Re \times \Re^{+}$. If$G(\cdot)$ is uncertain and modeled as a Dirichlet process, then the data come from a Dirichlet mixture of normals (Escobar 1995; Ferguson 1983; West 1990).

In particular, it is supposed that $G \sim DP(\alpha G_0)$, a Dirichlet process defined by α, a positive scalar, and$G_0(\cdot)$, a specified bivariate distribution function over $\Re \times \Re^{+}$.$G_0(\cdot)$ is the prior expectation of$G(\cdot)$, so that $E\{G(\pi)\} = G_0(\pi)$ for all $\pi \in \Re \times \Re^{+}$, and α is a precision parameter, determining the concentration of the prior for $G(\cdot)$ about$G_0(\cdot)$ .Write parameters of concern as$\pi = \{\pi_1, \ldots, \pi_n\}$.

A key feature of the model structure, and of its analysis, relates to the discreteness of$G(\cdot)$under the Dirichlet process assumption (Ferguson, 1973.) Briefly, in any sample π of size n from $G(\cdot)$there is positive probability of coincident values. Thus, given π, a sample of size n from$G(\cdot)$, the subsequent estimates represents a new, distinct probability values. In practical density estimation, suitable values ofαwill typically be small relative to the initial prior$G_0(\cdot)$. Dirichlet process mixture models are based on Dirichlet process priors for the primary parameters.

Such a model assumes that the prior distribution function itself is uncertain drawn from a Dirichlet process in standard notation such as in Antoniak (1974). Hsieh and Turnbull (1996), considered similar estimation methodology for binary ROC curves. For this research, the methodology was extended to accommodate three group test outcomes as a generalization of the binary ROC curves.

A generic function that performs ROC surface analysis based on Dirichlet process mixture of normals models for density estimation was considered (Escobar & West, 1995).A diagnostic test in which subjects can be classified into three different ordered categories was assumed.

Considering $\mathbf{Y_1} = (Y_{11}, Y_{12}, \dots, Y_{1n_1})^T$test results on n_1 subjects from Class 1; $\mathbf{Y_2} = (Y_{21}, Y_{22}, \dots, Y_{2n_2})^T$ test results on n_2 subjects from Class 2; and $\mathbf{Y_3} = (Y_{31}, Y_{32}, \dots, Y_{3n_3})^T$test results on n_3 subjects from Class 3.

The test results for the individuals from the three classes are modeled according to the distribution $G_{Y_{1i}}$, $G_{Y_{2j}}$and $G_{Y_{3k}}$; continuous probability distributions on $\mathbb{R}^n$. Functional form of Dirichlet Process Mixtures of Normals for $Y_{1i}$is given by:

$$Y_{1i}|G_{Y_{1i}} \sim \int G_{1i}\left(Y_{1i}\middle|\mu_{Y_{1i}}, \sigma_{Y_{1i}}\right) dG_{Y_{1i}}\left(\mu_{Y_{1i}}, \sigma_{Y_{1i}}\right)$$

$$G_{Y_{1i}} \sim DP(\alpha_{Y_{1i}}, G_{Y_{10}})$$

This can be substituted for Y_{2j}andY_{3k} as well. The distribution for Y_{1i}(distribution for the non-diseased or healthy group) can be expressed as follows:

$$Y_{1i}|\ \mu_{Y_{1i}}, \sigma_{Y_{1i}} \sim N(\mu_{Y_{1i}}, \sigma_{Y_{1i}}), i = 1, \dots, n_1$$

$$(\mu_{Y_{1i}}, \sigma_{Y_{1i}})\ |\ G_{Y_{1i}} \sim G_{Y_{1i}}$$

$G_{Y_{1i}} | \alpha_{Y_{1i}} G_{Y10} \sim DP(\alpha_{Y_{1i}} G_{Y_{10}})$………………....………………………………………. *(17)*

where, the baseline distribution is the conjugate normal-inverted-Wishart,

$$G_{Y_{10}} = N(\mu_{Y_{1i}}/m_1, (1/k_1)\, \sigma_{Y_{1i}})\, IW\, (\sigma_{Y_{1i}}/\, \nu_1, \psi_1)$$

Similarly for Y_{2j} (distribution for the transition or suspicious group)

$Y_{2j} |\, \mu_{Y_{2j}}, \sigma_{Y_{2j}} \sim N(\mu_{Y_{2j}}, \sigma_{Y_{2j}}), j = 1, \dots, n_2$

$(\mu_{Y_{2j}}, \sigma_{Y_{2j}}) \mid G_{Y_{2j}} \sim G_{Y_{2j}}$

$G_{Y_{2j}} | \alpha_{Y_{2j}} G_{Y_{20}} \sim DP(\alpha_{Y_{2j}} G_{Y_{20}})$……………………………………………..……… *(18)*

It was assumed the baseline distribution is the conjugate normal-inverted-Wishart,

$$G_{Y_{20}} = N(\mu_{Y_{2j}}/m_2, (1/k_2)\, \sigma_{Y_{2j}})\, IW\, (\sigma_{Y_{2j}}/\, \nu_2, \psi_2)$$

Finally, for Y_{3k} (distribution for the diseased group)

$Y_{3k} |\, \mu_{Y_{3k}}, \sigma_{Y_{3k}} \sim N(\mu_{Y_{3k}}, \sigma_{Y_{3k}}), k = 1, \dots, n_3$

$(\mu_{Y_{3k}}, \sigma_{Y_{3k}}) \mid G_{Y_{3k}} \sim G_{Y_{3k}}$

$G_{Y_{3k}} | \alpha_{Y_{3k}} G_{Y_{30}} \sim DP(\alpha_{Y_{3k}} G_{Y_{30}})$…………………………………………..……… *(19)*

It was also assumed that the baseline distribution is the conjugate normal-inverted-Wishart,

$$G_{Y_{30}} = N(\mu_{Y_{3k}}/m_3, (1/k_3)\, \sigma_{Y_{3k}})\, IW\, (\sigma_{Y_{3k}}/\, \nu_3, \psi_3)$$

To let part of the baseline distribution fixed at a particular value, the corresponding hyperparameters of the prior distributions were set to null in the hyperprior specification of the model.

Although the baseline distribution, $G_{Y_{10}}$, $G_{Y_{20}}$and $G_{Y_{30}}$, are conjugate priors in the model specifications, the algorithms with auxiliary parameters described in MacEachern and Muller (1998) and Neal (2000) are adopted.

The ROC surface using a Monte Carlo approximation to the posterior means $\mathrm{E}(G_{Y_{1i}}|Y_{1i})$, $\mathrm{E}(G_{Y_{2j}}|Y_{2j})$, and $\mathrm{E}(G_{Y_{3k}}|Y_{3k})$, which is based on MCMC samples from posterior predictive distribution for a future observation.Y_{1i}, Y_{2j}andY_{3k} are the vectors containing the diagnostic marker measurements in the non-diseased, transition or suspicious group and diseased subjects, respectively. The optimal cut-off point is based on the efficiency test, EFF = TP + TTR+TN, and is built on Cohen's kappa as defined in Kraemer (1992).

3.3 Bayesian nonparametric method based on Mixtures of Finite Polya trees

The nonparametric model developed involves finite Polya tree priors for the distributions F_1, $F_2$and F_3. Freedman (1963), Fabius (1964), and Ferguson (1974) are the most notable pioneers of Polya tree priors. However, Lavine (1992, 1994), became the natural starting point for understanding their potential use in modeling data while Hanson (2006) considered some computational details. Recent applications of these priors in ROC curve analysis can be found in Branscum et al. (2008) and in Hanson et al. (2008).

According to Lavine (1992), polya trees form a class of distributions for random probability measure which is an intermediate between Dirichlet processes and tailfree

processes (Tail-free processes are stochastic processes that can be defined to have trajectories on the space of probability distributions). Polya trees offers better approach to Dirichlet processes since they can allow construction of probability 1 to a set of continuous or absolutely continuous probability measures, whereas their advantage over more general tailfree processes is their much greater tractability.

Another advantage is that in some sampling situations results in posterior mixtures of Dirichlet processes, can only lead to just a single posterior Polya tree.

3.3.1 Properties of polya trees

Let $E = \{0,1\}, E^0 = \emptyset, E^m$be the $m-$fold product $E \times E \times \cdots \times E$, $E^* = \bigcup_0^\infty E^m$ and E^N be the set of infinite sequences of elements of E. Further, $\varepsilon \epsilon E^*$ has Beta distribution with parameters $\alpha_{\varepsilon 0}$ and $\alpha_{\varepsilon 1}$

1. Dirichlet processes are special cases of Polya trees. A Polya tree is a Dirichlet process if, for every, $\varepsilon \epsilon E^*, \alpha_\varepsilon = \alpha_{\varepsilon 0} + \alpha_{\varepsilon 1}$

2. Some Polya trees assign probability 1 to the set of continuous distributions.

Polya trees can accommodate most forms of data. A finite Polya tree prior with *J* levels on a random probability measure *F* augments a standard family of cumulative distributions $\{F_\theta : \theta \in \Theta\}$ with 2^J-1 additional parameters X that stochastically adjust the density $f_\theta(.)$ to place additional mass in areas where data are seen more often than expected under $f_\theta(.)$. The idea of centering the Polya tree at a parametric family is that if the data really follow the parametric family, the Polya tree should be more efficient than other nonparametric priors.

A finite Polya tree for a distribution *F* is constructed by dividing the sample space

into finer-and-finer disjoint sets using successive binary partitioning. Denote the series of nested partitions by $\boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_J$ where, for $\boldsymbol{1 \leq j \leq J}$, $\boldsymbol{\Pi}_{j+1}$ a refinement of the partition Π_j in that each set in Π_j is the union of two sets in Π_{j+1}. At level $j \leq J$ the tree, the sample space is partitioned into 2^j sets each with a corresponding branch probability. The products of the branch probabilities that lead to sets at level *j* provide the marginal probabilities of those sets. The sets that partition the sample space at the *j*-th level of the tree are denoted by $B_\theta(j,k), k = 1,\ldots,2^j$, and the standard parametrization defines $B_\theta(j,k) = (F_\theta^{-1}((k-1)/2^j), F_\theta^{-1}(k)/2^j))$, for a parametric distribution F_θ. The corresponding random branch probabilities, denoted by $X_{j,k}$, are modeled according to the independent gamma distributions, namely, $X_{j,k} \sim \Gamma(cj^2, cj^2)$ with $X_{j,k+1} = 1 - X_{j,k}, j = 1,\ldots,J, k = 1,\ldots,2^j - 1.$

The weight parameter $c > 0$ determines how concentrated F is around F_θ. Larger values of c lead to inferences that approach those for a parametric analysis, whereas small values allow less concentration and thus lead to a more nonparametric analysis. In practice, c often set to a fixed value. For instance, setting $c = 1$ allows for a great deal of prior flexibility. For this work, c will be fixed, although a prior distribution could be placed on it for general MFPT models as discussed by Hanson (2006). The notation $F \sim \mathrm{FPT}_J(F_\theta, c)$ is used to denote that *F* has a finite Polya tree prior with *J* levels, centered at F_θ with weight parameter c. Finite Polya tree priors treat θ as a constant, while mixtures of finite Polya trees model θ with a prior distribution, $\theta \sim p(\mathrm{d}\theta)$. In general, mixtures of Polya trees are useful when a standard parametric Bayesian analysis is suspect because the family of sampling densities is not known exactly. The standard Bayesian analysis requires specification of a family of sampling

densities and prior densities for some parameter.

3.3.2 Inference for VUS based on mixtures of finite Polya trees (MFPT) priors

To obtain the nonparametric estimator of ROC surface, all the distribution functions in (14) shall be replaced with their empirical counterparts. The estimator shall be constructed as:

$$ROCS(p_1,\ p_3) =$$

$$\begin{cases} \widehat{F_2}(\widehat{F_3^{-1}}\ (1-p_1)) - \widehat{G_2}(\widehat{F_1^{-1}}(p_3)), if\ \widehat{F_1^{-1}}(p3)\ \leq\ \widehat{F_3^{-1}}(1-p_1), \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{otherwise.} \end{cases}$$

…… *(20)*

Where$\widehat{F_1}$, $\widehat{F_2}$and $\widehat{F_3}$are the nonparametric distribution functions for test results from the three classes, respectively. The model to be used is a specified hierarchical model involving the specification of independent mixture of finite polya tree priors forG_i, (*i=1, 2, 3*) conditional on hyperparameters.

The general non-parametric model is

$Y_{1i} \sim F_1$, $Y_{2j} \sim F_2$, and $Y_{3k} \sim F_3$

$F_i | c_i, \theta_i \sim \text{FPT}_j(F_{\theta_i}, c_i)$,

$\theta_i \sim p(\text{d}\theta_i)$

Random $F_i$is centered at F_{θ_i}= $N(\mu_i, \sigma_i)$ where$\theta_i = N(\mu_i, \sigma_i)$. Let $\chi_i = \{X_{i,j,k}\}$ denote the set of branch probabilities for F_i. The mixing parameters μ_i have independent normal priors$N(a_{\mu_i}, b_{\mu_i})$whereas σ_i have independent gamma priors Gamma $(a_{\sigma_i}, b_{\sigma_i})$, all with fixed hyperparameters. The levels of the finite Polya trees are set

equal to $J_i$and determine the level of detail that is accommodated by the model. Similarly, the weight parameter c_i was also fixed. The likelihood function is proportional to

$$\prod_{i=1}^{n_1} f_1(y_{1i}|\chi_1,\theta_1) * \prod_{j=1}^{n_2} f_2(y_{2j}|\chi_2,\theta_2) * \prod_{k=1}^{n_3} f_3(y_{3k}|\chi_3,\theta_3) \ldots\ldots\ldots\ldots\ldots\ldots \textit{(21)}$$

With $f_i$being the density corresponding toF_i. The cumulative distribution function is $F_i(y|\chi_i,\theta_i)$. The joint posterior distribution is approximated via Markov Chain Monte Carlo (MCMC) methods. Given the likelihood function, the expression of the ROC surface in formula (14) can be computed. The mixture of finite Polya trees model was fitted using MCMC with simple Metropolis–Hastings steps. The mixing parameters μ_i and σ_i are updated with random-walk-type proposals centered at the previous values.

CHAPTER FOUR

RESULTS AND ANALYSIS

4.1 Introduction

This chapter gives a presentation of results in the form of tabulations graphs. Discussion of the results was also made. All graphical and numerical computations were conducted in R version 3.1.0 (R Development Core Team, 2014)

4.2 Semiparametric Bayesian density estimation and ROC surface estimation.

The simulation study to assess the performance of the Non-parametric estimation method covers test scores for the three classes generated from three different normal populations: Y_{1i}~N(1,1.5), Y_{2j}~N(2,1.5) and Y_{3k} ~N(3,1.5), where sample sizes n_1 and n_2, and $n_3$were set to 100, 50 and 100. The means are ordered to ensure that the observations are monotonically increasing while the sample sizes were considered under conditions for diagnostic tests in clinical practice (Jokiel-Rokita& Pulit, 2013).

To fit the Semiparametric Bayesian density estimation, Dirichlet Process Mixtures (DPM) of normals model for density estimation was used. The prior parameters are defined as; α=1 gives the value of the precision parameter, m_i replicated for the values *i=1,2* and *3* gives the mean of the normal part of the baseline distribution. In addition, $\psi_i^{-1}$of a solution of scale matrix gives the hyperparameters of the inverted Wishart prior distribution for the scale matrix,ψ_i, of the inverted Wishart part of the baseline distribution.

Further, ν_i represents the hyperparameters of the inverted Wishart part of the baseline distribution. $\tau_1 = 1$and $\tau_2 = 100$ gives the hyperparameters for the gamma prior distribution of the scale parameter k_i of the normal part of the baseline distribution.

The k_i and ψ_i parameters replicated for the values 1,2 and 3 fitted in the posterior distribution represents the scale parameter of the normal part of the baseline distribution and the scale matrix of the inverted-Wishart part of the baseline distribution respectively.

4.3 Properties of the fitted semiparametric distribution

4.3.1 Posterior Inference of Parameters under Semiparametric estimation

The fitted density estimation using DPM of normals model for the test outcomes $Y_{1i,}$ representing the non-diseased subjects to obtain Posterior Inference of Parameters. The k_1, the scale parameter of the normal part of the baseline distribution and solution of the matrix $\psi_1 - Y_{1i}$of the inverted-Wishart part of the baseline distribution was computed. The results are summarized in the Table 3 below.

Table 3: Posterior Inference of Parameters for Y_{1i} (test outcomes for diseased)

k_1	$\psi_1 - Y_{1i}$
0.03252	0.76822

Dirichlet Process Mixtures model of normals for the test outcomes$Y_{2j,}$ representing the transition or suspicious group was also fitted. The Posterior Inference of Parameters for the scale parameter of the normal part of the baseline distribution $k_2$and solution of the matrix $\psi_2 - Y_{2j}$ of the inverted-Wishart part of the baseline distribution was computed. The results are summarized in table 4 below.

Table 4: Posterior Inference of Parameters for Y_{2j} (test outcomes for diseased)

k_2	$\psi_2 - Y_{2j}$
0.02477	0.65813

Lastly, the density estimates for the test outcomes Y_{3k}representing the diseased subjects using Dirichlet Process Mixtures DPM of normals model were fitted. The Posterior Inference for the scale parameter of the normal part of the baseline distribution parameters $k_3$and solution of the matrix $\psi_3 - Y_{3k}$ of the inverted-Wishart part of the baseline distribution were obtained. The results are summarized in table 5 below.

Table 5: Posterior Inference of Parameters for Y_{3k} (test outcomes for diseased)

k_3	$\psi_3 - Y_{3k}$
0.02197	0.73261

4.3.2 Posterior parameters for Y_{1i} under semiparametric estimation

The posterior parameters $k_1$and $\psi_1 - Y_{1i}$summarized by time series MCMC scans and fitted line for the parameter values for Y_{1i} (test outcomes for the non-diseased group) were also computed. Figure 3 below gives a summary of the parameter plots.

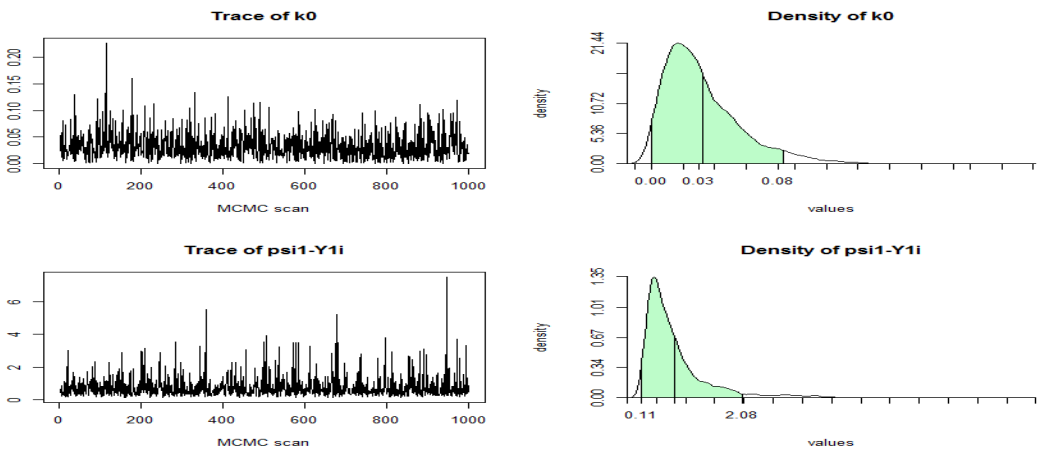

Figure 3: Posterior parameters for Y_{1i} under semiparametric estimation

It was evident that the chains for the posterior parameters; the scale parameter of the normal part of the baseline distribution $k_1$and solution of the matrix $\psi_1 - Y_{1i}$of the inverted-Wishart part of the baseline distribution depicts stationarity at the true parameter values. The line plots of the two posterior parameters for the 1000

iterations of the Metropolis Hastings steps sampler produced smooth plots.

4.3.3 Posterior parameters for Y_{2j} under semiparametric estimation

The posterior parameters plot for Y_{2j}(test outcomes for the transition or suspicious group) were also computed. MCMC scans and fitted line for the parameter values for k_2 and $\psi_2 - Y_{2j}$were summarized in figure 4 below.

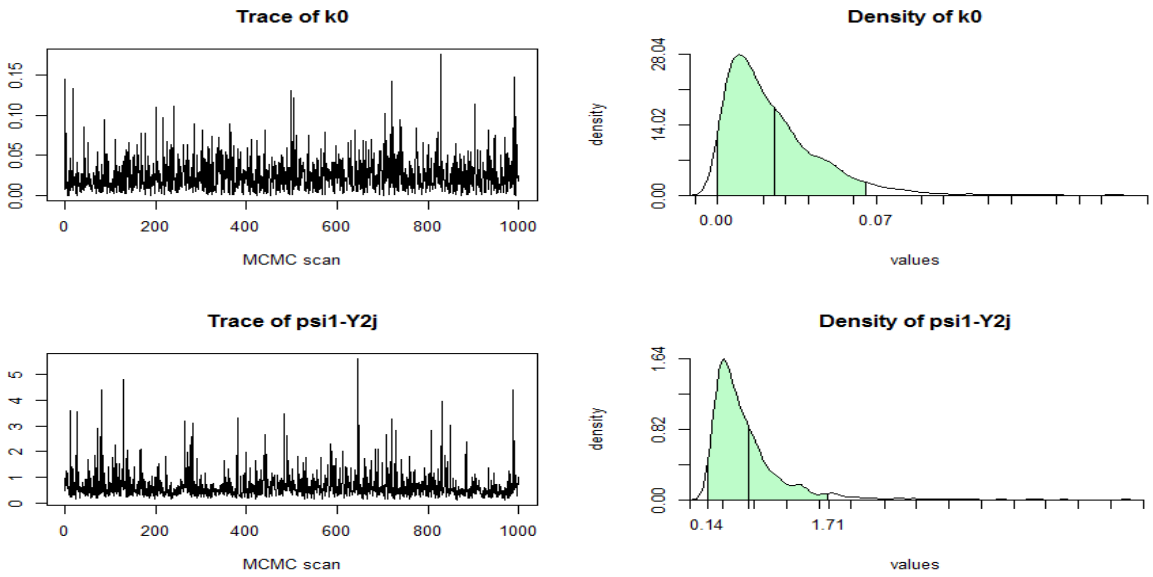

Figure 4: Posterior parameters for Y_{2j} under semiparametric estimation

It was evident that the MCMC iterates for the posterior parameters converges to the true parameter values for the 1000 iterations of the Metropolis Hastings steps sampler. The sampler produced smooth plots as well. In particular, it was found out that the scale parameter of the normal part of the baseline distribution $k_2$and solution of the matrix and $\psi_2 - Y_{2j}$of the inverted-Wishart part of the baseline distribution portrays some stationary state near the true parameter values.

4.3.4 Posterior parameters for Y_{3k} under semiparametric estimation

The posterior parameters for Y_{3k} (test outcomes for the diseased group) summarized series MCMC scans and fitted line was computed. The plots for the posterior parameters $k_3$and $\psi_3 - Y_{3k}$were represented in figure 5 below.

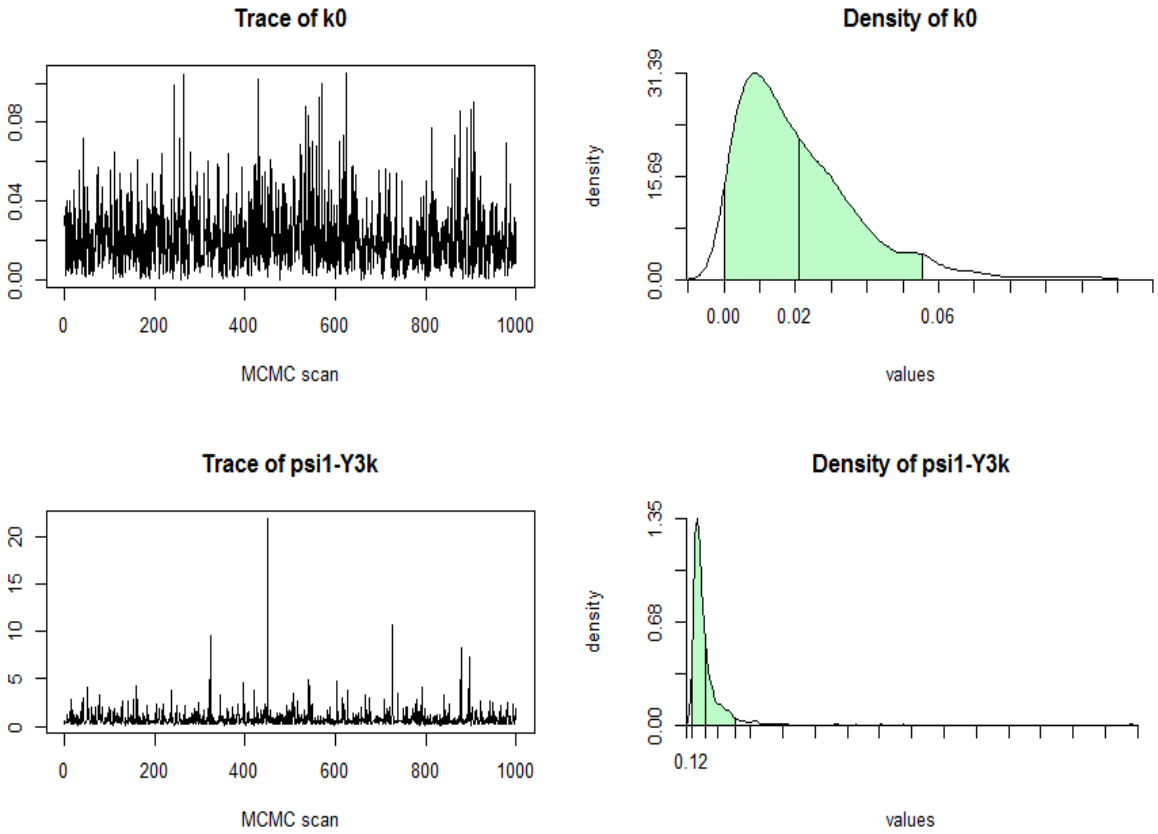

Figure 5: Posterior parameters for Y_{3k} under semiparametric estimation

For the posterior parameters for Y_{3k} (test outcomes for the diseased group) it was also apparent that the 1000 MCMC iterates for the posterior parameters converges to the true parameter values. It was also evident that the Metropolis Hastings steps sampler produced smooth plots. In general, it was evident that the posterior parameters $k_3$and $\psi_3 - Y_{3k}$were stationary as they approach the true parameter values.

4.3.5 Data plots under semiparametric estimation

It was evident from these plots of posterior parameters; $k_i$and $\psi_i - Y_{ij}$that the process appear to be stationary. All plots suggested that convergence was achieved after 1000 iterations of the Metropolis Hastings steps Sampler. Inference thereof is that each plot seemed to confirm that the parameter posterior parameters converged to stationarity after 1000 or so iterations. Further, the fitted distribution was analysed by

comparing it with the data plots and kernel density estimate plots. It was shown that the posterior distribution curve- DPM of normals model curve fit and kernel density estimate curve fit for the data. The kernel density estimate curve fits for the posterior estimates represented by faint line while dark line represented the posterior distribution curve fit.

4.3.6 Distribution of Y_{1i} (test outcomes for the nondiseased group) under Semiparametric estimation

The data plot for Y_{1i} (test outcomes for the non-diseased group)was represented using a histogram. The posterior distribution of the test outcomes for the non-diseased group and curve fits for DPM of normals model and kernel density estimate were also fitted on the same data plot. Figure 6 below shows the data plot, DPM of normals model curve fit and kernel density estimate curve fit.

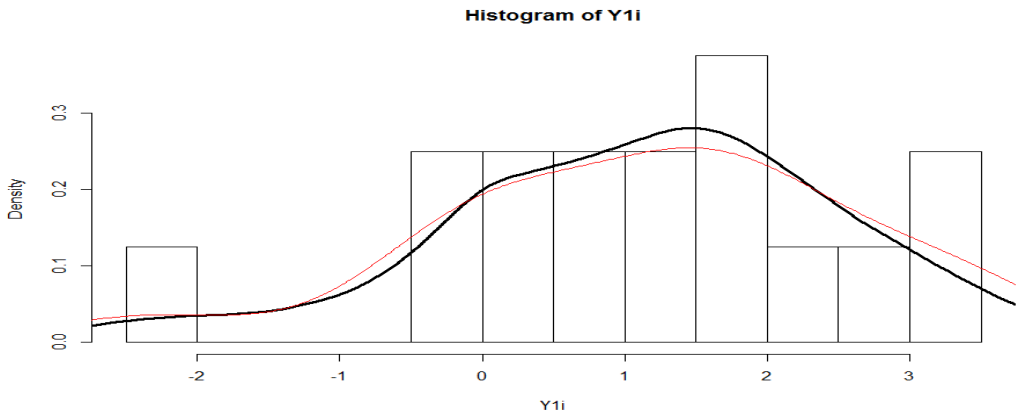

Figure 6: Distribution of Y_{1i} (test outcomes for the nondiseased group) under Semiparametric estimation

Figure 6 above shows that plot forY_{1i} (test outcomes for the nondiseased group) portrays that the data assumes some distribution, evident in the presence of peaks. TheDPM of normals model posterior distribution fit for the test outcomes is a near symmetric curve fit indicating that the posterior distribution fits the data well. Thekernel density smooth curve fit further confirmed that that the DPM of normals

model fits the data convincingly.

4.3.7 Distribution of Y_{2j}(test outcomes for the transition or suspicious group) under Semiparametric estimation

The plots for Y_{2j} (test outcomes for the transition or suspicious group) were computedusing a histogram. The posterior distribution of the test outcomes for the transition or suspicious group and curve fits for DPM of normals model and kernel density estimate were also fitted on the same data plot. DPM of normals model curve fit and kernel density estimate curve fit are shown in figure 7 below.

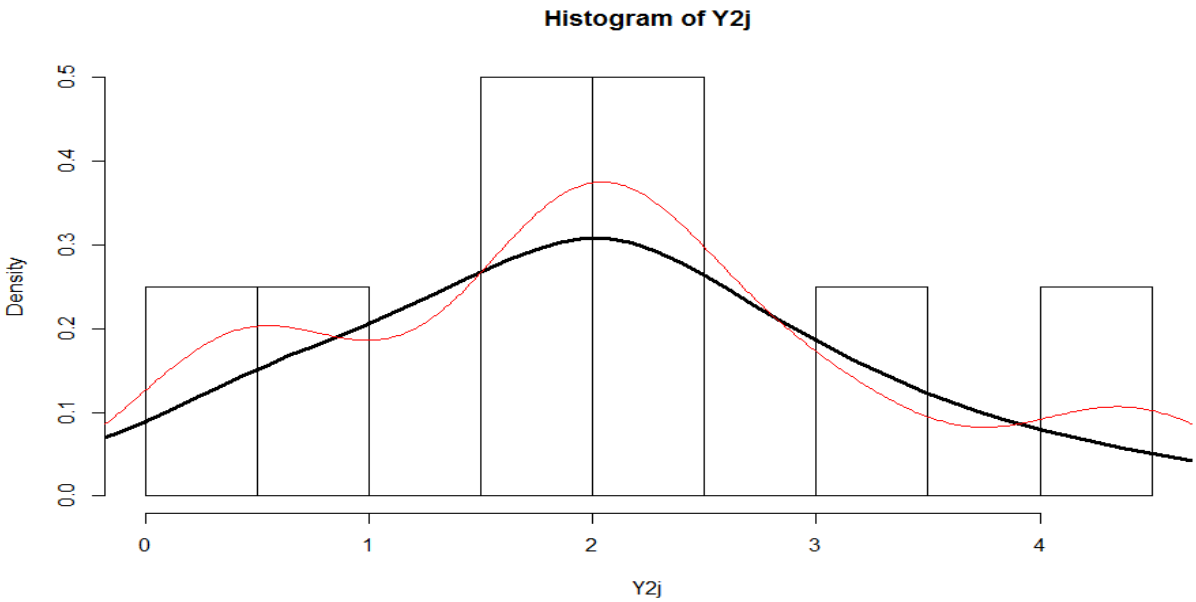

Figure 7: Distribution of Y_{2j} (test outcomes for the transition or suspicious group) under Semiparametric estimation

The presence of peaks in data plot forY_{2j} (test outcomes for the transition or suspicious group) portrays that the test outcomes assumes some distribution. TheDPM of normals model posterior distribution fit for these test outcomes was a near perfect symmetric curve or normal distribution fit indicating that the posterior distribution fits the data well. Thekernel density smooth curve fit further confirmed that that the DPM of normals model fits the data convincingly, especially a true parameter values.

4.3.8 Distribution of Y_{3k}(test outcomes for the diseased group) under semiparametric estimation

The data for Y_{3k}(test outcomes for the diseasedgroup) was also plottedusing ahistogram. The posterior distribution of the test outcomes for the diseasedgroup and curve fits for DPM of normals model and kernel density estimate were also fitted on the same data plot. DPM of normals model curve fit and kernel density estimate curve fit are shown in figure 8 below.

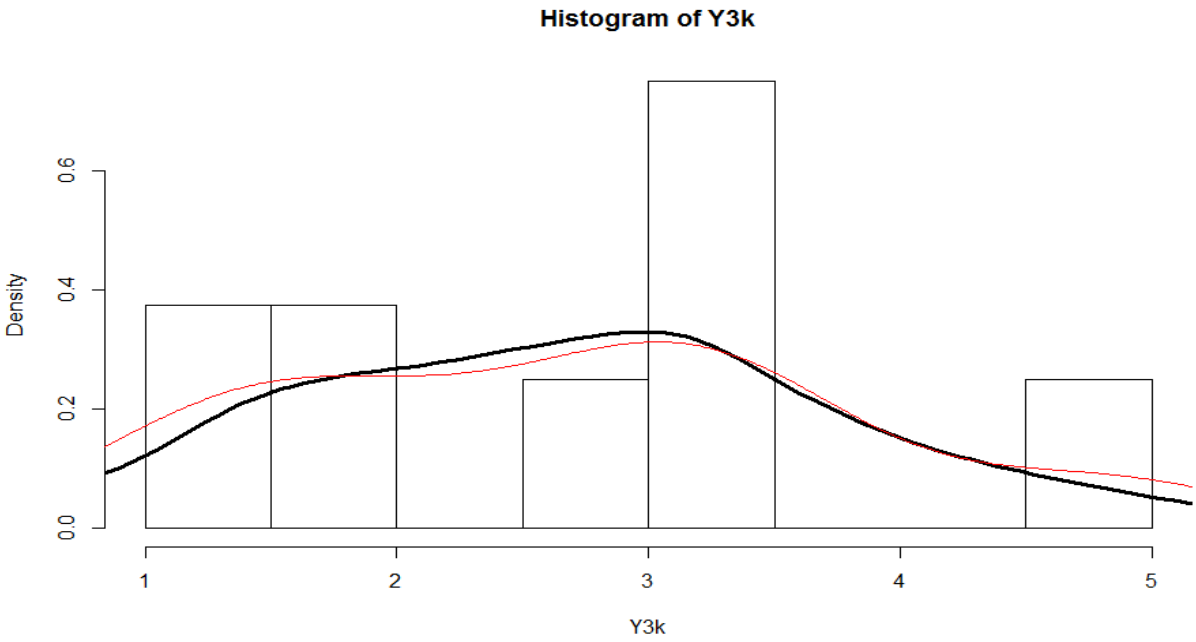

Figure 8: Distribution of Y_{3k} (test outcomes for the diseased group) under Semiparametric estimation

The histogram or data plot representing posterior distribution ofY_{3k}(test outcomes for the diseased group) reveals that the test outcomes follow some distribution. The fit TheDPM of normals model posterior distribution for these test outcomes lacked smooth fit though it exhibited peak at true parameter values. Thekernel density smooth curve fit and the DPM of normals model curve fit were found to be adjacent

indicating that the DPM of normals model fits the data well.

4.4 Semiparametric Estimation of ROC surface

Having seen that the properties semiparametric bayesian density estimators; the posterior parameters were desirable, samples were drawn to estimate the ROC surface.

Random samples for the three test outcomes using the DPM of Normals procedures. These data represents test outcomes of a simulated diagnostic test that classifies disease or condition into three ordered groups namely D- (non diseased), D0 (transition or suspicious) and D+ (diseased). It was assumed that the test outcomes are ordinal and that the simulated diagnostic test classifies the groups without overlap. Table 6 below provides a summary of the test outcomes drawn from the distribution.

Table 6: Simulated diagnostic test Raw Data Summary for Semiparametric model

Simulated diagnostic test Raw Data Summary for Semiparametric model

	n=Sample size	μ= means
D- (non diseased)	100	0.1313193
Do (transition/suspicious)	50	0.1606993
D+ (diseased)	100	0.1867411

4.4.1 ROC surface plot

From the simulated diagnostic test classification of test outcomes, a three way ROC surface corresponding to the three test outcomes was plotted. The Semiparametric model defined in to estimate the ROC surface plot were used. Figure 9 below represents three dimensional surface plot depicting trade-offs between the predictive measures for classification of the three test outcomes.

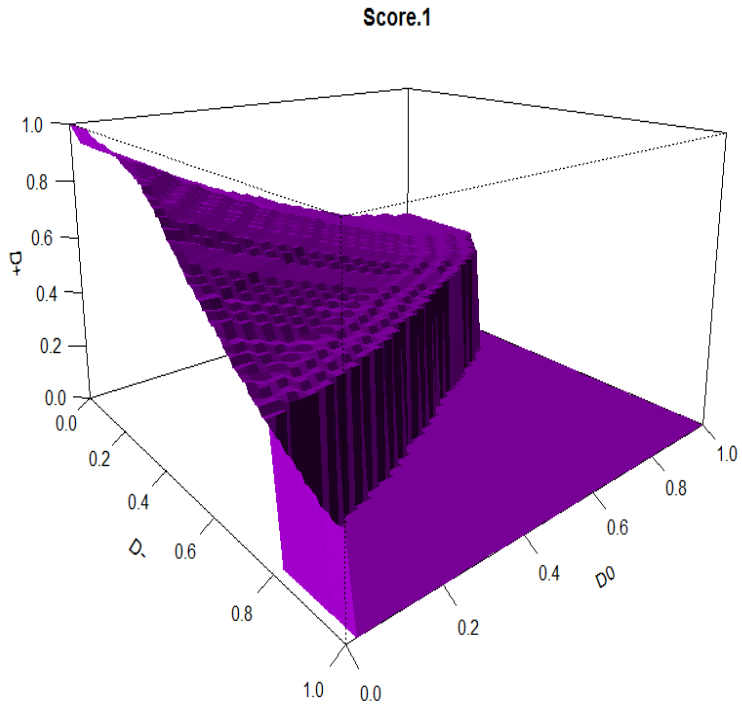

Figure 9: Three-dimensional ROC surface plot depicting tradeoffs between the predictive measures for classification of the three test outcomes under semiparametric model

Figure 9 above represents a plot to estimate the ROC surface under the semiparametric model assumptions. It was evident that the semiparametric model gives a near smooth ROC surface indicating that the ROC surface has good coverage thus the simulated diagnostic test performs well in classifying the test outcomes.

4.4.2 Semiparametric estimation of volume under ROC surface (VUS)

For the simulated diagnostic test that classified the test outcomes into the three groups to estimate the volume under ROC surface, volume under surface group of test outcomes under the Semiparametric assumptions based on DPM of normals model was computed. It was found out that the volume under the surface VUS= 0.3411. The 95% confidence interval was also computed where the lower confidence interval was found to be 0.317 while the upper confidence interval was found to be 0.4598.

Additionally, optimal cut-points off points were computed whereby the best lower cut-point was found to be 0.0935 while the upper cut-point was found to be 0.2307. The group correct classification probabilities were found to be; specificity= 0.48, true transition rate=0.74 and sensitivity=0.42. It was also derived that the estimate of the sample size=115 for the predefined precision. As such, to better estimate the diagnostic accuracy of the marker or group of test outcomes, minimum sample size of 115 will be desired for each group in order to estimate the VUS of the marker within a 5% margin of error.

4.4.3 Scatter plot and a boxplot for the semi-parametric VUS

To provide a summary of the semiparametric volume under ROC surface, a general scatter plot and a boxplot were plotted. The graphical summary of the data for D- , D0 and D+ colored in green, blue and red, respectively and the estimated summary measure for the confidence interval (CI) is provided in the legend while the optimal cut-points are labeled for the Semiparametric VUS. Figure 10 below gives the summary.

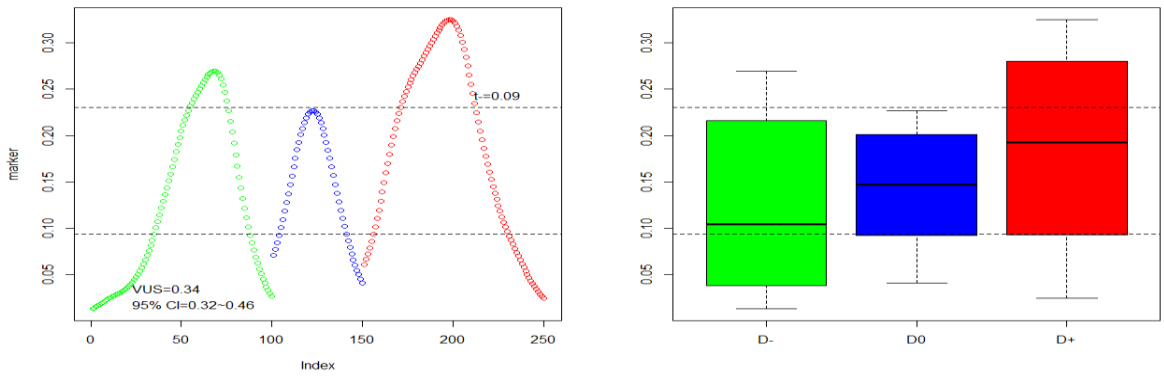

Figure 10: Scatter plot and a boxplot of the marker under semiparametric model

From figure 10 above, the dashed lines show the lower and uppercut points for the Semiparametric VUS. The value of the VUS 0.34 is also shown. The box plots portrays that the test outcomes show some ordering and that the test outcomes are

ordinal or values are monotonically increasing.

4.5 Non-parametric Bayesian density estimation and ROC surface estimation

The simulation study to assess the performance of the Non-parametric estimation method covers cases of normal distribution where test scores for the three classes were generated from three different normal populations: $Y_{1i} \sim N(1,1.5), Y_{2j} \sim N(2,1.5)$ and $Y_{3k} \sim N(3,1.5)$, where sample sizes n_1 andn_2, and $n_3$were set to 16, 8 and 16 respectively. It was also assumed that the Polya Tree is centered on normal distribution, $PT \sim N(0,1)$ distribution, by taking each *J=4* levels of the partitions where$n_i \approx 2^J$ (Hanson & Johnson, 2002). Similar to the semiparametric case, the means are ordered to ensure that the observations are monotonically increasing.

It was assumed that α, the precision parameter of the Polya Tree prior, $\alpha = 1$, and was considered as random, having a gamma distribution, $\Gamma(a_0, b_0)$: hyperparameters for prior distribution of the precision parameter. In the computational implementation of the model, Metropolis-Hastings steps were used to sample the posterior distribution of the baseline and precision parameters. The number of grid points where the density estimate was evaluated was set to ratio proportionate to the simulated sample size, that is, 2:1:2, it was assumed assumed 100, 50 and 100 for $n_1, n_2$and n_3; the simulated case.

The MCMC parameters included nburn = 1000 gives the number of burn-in scans, nskip= 50 gives the thinning interval, nsave= 1000 giving the total number of scans to be saved. Further, ndisplay= 100 giving the number of saved scans to be displayed on screen, tune1=0.15, tune2=1.1, and tune3=1.1 giving the positive Metropolis tuning parameter for the baseline mean, variance, and precision parameter, respectively.

4.6 Properties of the fitted Non-parametric distribution

4.6.1 Posterior Inference of Parameters under Non-parametric Bayesian estimation

The fitted distribution (Bayesian Density Estimation Using MFPT) for the test outcomes$Y_{1i,}$ representing the non-diseased subjects has the Posterior Inference of Parameters $\mu_{Y_{1i}} = 1.173$ and$\sigma_{Y_{1i}} = 1.786$. The acceptance rate for the Metropolis Step = 0.7350577 to 0.7764615. The results are summarized in table 7 below.

Table 7: Posterior Inference of Parameters for Y_{1i} (test outcomes for diseased)

$\mu_{Y_{1i}}$	$\sigma_{Y_{1i}}$	α
1.22	1.97	1.00

Acceptance Rate for Metropolis Step = [0.73 0.78]

Likewise, the test outcomes $Y_{2j}$for the transition or suspicious state condition were fitted using Bayesian Density Estimation Using MFPT whereby the posterior Inference of Parameters $\mu_{Y_{2j}} = 2.18$ and $\sigma_{Y_{2j}} = 1.70$ while the acceptance rate for the Metropolis Step = 0.78 0.85. Table 8 below provides a summary of the results.

Table 8: Posterior Inference of Parameters for Y_{2j} (test outcomes for transition or suspicious)

$\mu_{Y_{2j}}$	$\sigma_{Y_{2j}}$	α
2.18	1.70	1.00

Acceptance Rate for Metropolis Step = [0.78 0.85]

Finally, the test outcomes Y_{3k}representing the diseased or ‘with condition’ group were fitted under the Bayesian Density Estimation Using MFPT. The posterior Inference of Parameters $\mu_{Y_{3k}} = 3.22$ and $\sigma_{Y_{3k}} = 1.41$ while the acceptance rate for the Metropolis Step = 0.70 to 0.77. The results were summarized using table 9 below.

Table 9: Posterior Inference of Parameters for Y_{3k} (test outcomes for diseased or ‘with

condition')

$\mu_{Y_{3k}}$	$\sigma_{Y_{3k}}$	α
3.22	1.41	1.00

Acceptance Rate for Metropolis Step = [0.70 0.77]

4.6.2 Posterior parameters for Y_{1i} under Non-parametric Bayesian density estimation

Further, to analyze the properties of the posterior parameters, the parameters plots for the fitted distribution for each of the three simulated test outcomes were computed. The plot posterior parameters for Y_{1i} (test outcomes for the nondiseased group) were summarized by time series MCMC scans and fitted histogram line for the parameter values for mean and standard deviation namely $\mu_{Y_{1i}}$ and $\sigma_{Y_{1i}}$. Figure 11 below gives a summary of the parameter plots.

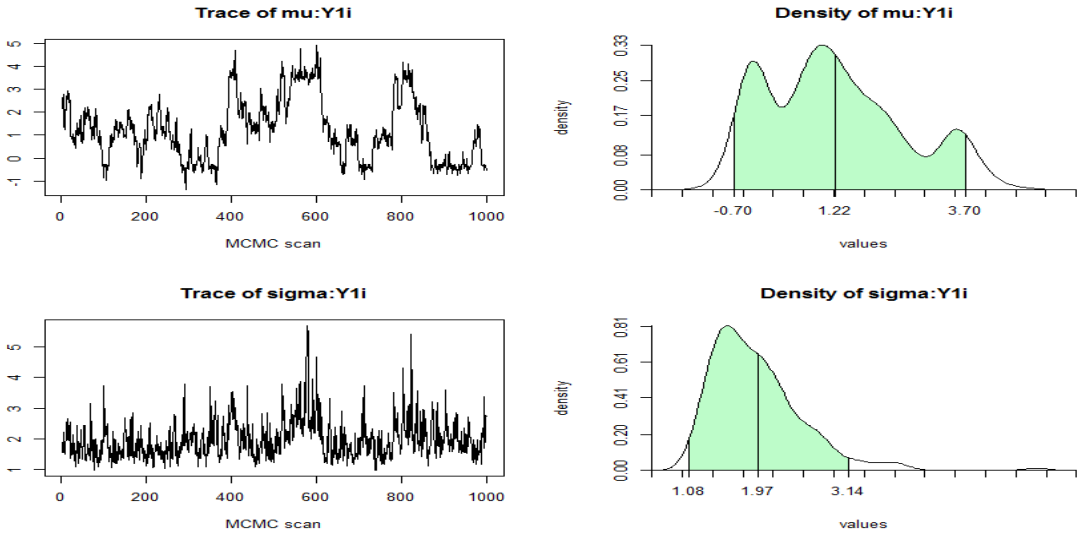

Figure 11: Posterior parameters for Y_{1i} under Non-parametric Bayesian density estimation.

It was evident that the chains for both parameters were desirably stationary at the true parameter values. The plots of the mean of the parameters and standard deviation for

the 1000 iterations of the sampler produced near- smooth plots.

4.6.3 Posterior parameters for Y_{2j} under Non-parametric Bayesian density estimation

Similarly for Y_{2j} (test outcomes for the transition or suspicious group), the fitted posterior parameters $\mu_{Y_{2j}}$ and $\sigma_{Y_{2j}}$ were plotted. A summary of the plots is represented by MCMC scans and fitted histogram line for the parameter values. Figure 12 below gives a summary of the parameter plots.

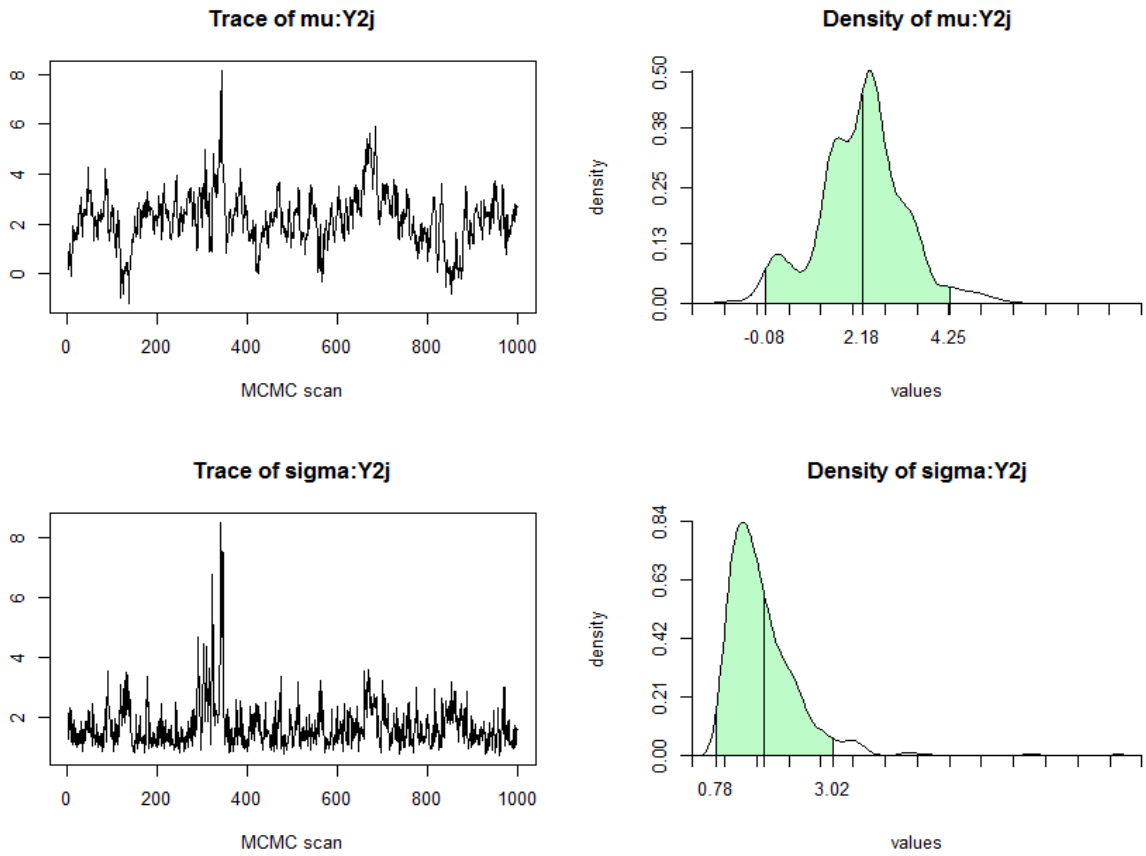

Figure 12: Posterior parameters for Y_{2j}under Non-parametric Bayesian density estimation

It was concluded from the time series trace plots that the parameter estimates for Y_{2j}(test outcomes for the transition or suspicious group), were all reasonably

convergent to the true parameter values. The parameter curve plots also depicted desirable degree of smoothness and predicted the true parameter values.

4.6.4 Posterior parameters for Y_{3k}under Non-parametric Bayesian density estimation

The parameter plots for the fitted posterior parameters $\mu_{Y_{3k}}$ and $\sigma_{Y_{3k}}$for Y_{3k}(test outcomes for the diseased or 'with condition') were also computed. Similar to previous test outcomes parameter plots; the plots for MCMC scans and fitted histogram line for the parameter values were computed. Figure 13 below gives a summary of the parameter plots.

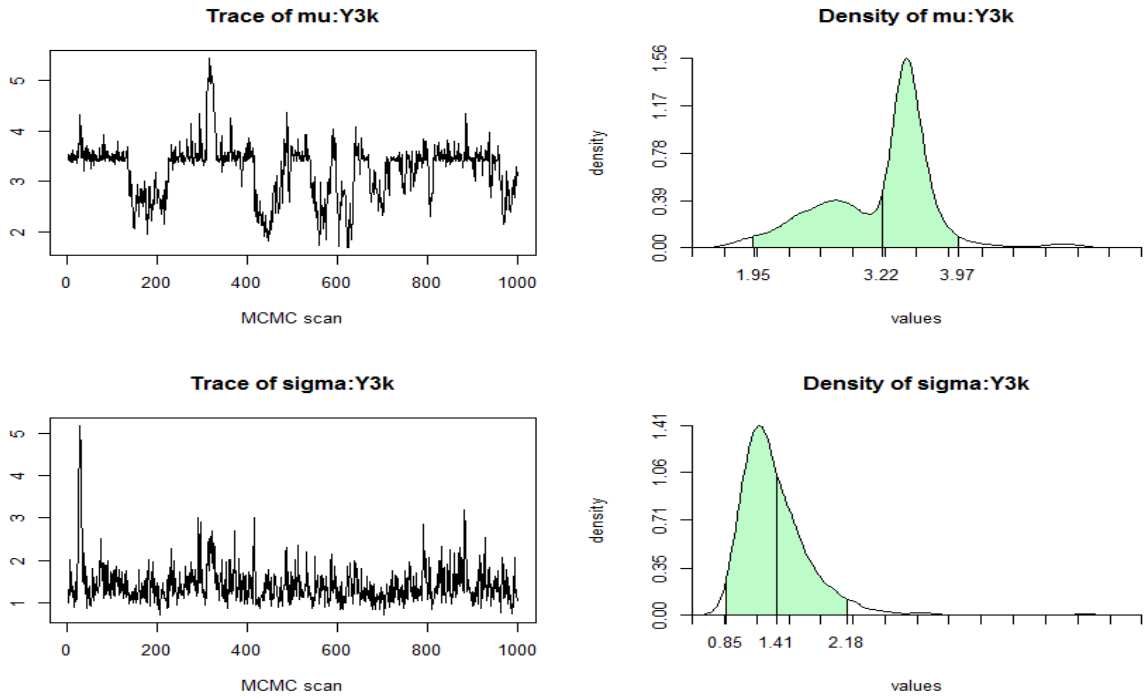

Figure 13: Posterior parameters for Y_{3k} under Non-parametric Bayesian density estimation

It was evident that the MCMC chains were convincingly stationary at the true parameter estimates for posterior parameters $\mu_{Y_{3k}}$ and $\sigma_{Y_{3k}}$forY_{3k}(test outcomes for the diseased or 'with condition'). Further, theplots of the parameter estimates; mean

and standard deviation show definite smoothness.

4.6.5 Data plots under Non-parametric Bayesian density estimation

The posterior parameters for the three simulated test outcomes indicate that posterior estimates; means and standard deviations converged or are stationary after 1000 iterations, then the distributions of the three test outcomes by plotting the data using histograms were analysed. The MFPT distribution curve fits for the posterior estimates represented by the dark line and the kernel density estimate curve fits for the posterior estimates represented by faint line for comparison purposes were computed.

4.6.6 Distribution of Y_{1i} (test outcomes for the non-diseased group) under Non-parametric Bayesian density estimation

The data for Y_{1i}(test outcomes for the non-diseased group)were plottedusing a histogram. The posterior distribution of the test outcomes for the non-diseased group and curve fits for MFPT and kernel density were computed. Figure 14 below shows the data, MFPT and kernel density plots.

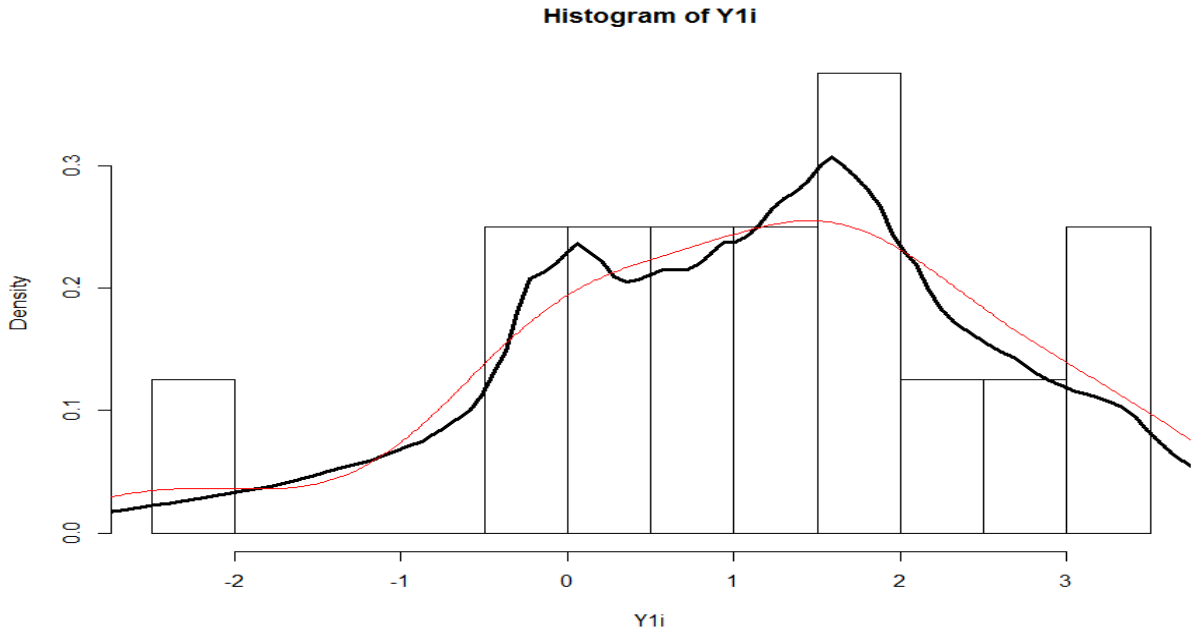

Figure 14: Distribution of Y_{1i} (test outcomes for the non-diseased group) under Non-parametric Bayesian density estimation

The histogram of the plot forY_{1i} (test outcomes for the non-diseased group) data depicts that the data follows some distribution, evident in the existence of peaks in the data. Theposterior distributions of the fit using MFPT indicate that the data fits the distribution well. Thekernel density smooth curve fit shows that the MFPT model fits the data convincingly, as the peak for the $\mu_{Y_{1i}}$lies at the true parameter value, that is, $\mu_{Y_{1i}} = 1$.

4.6.7 Distribution of Y_{2j}(test outcomes for the transition or suspicious group) under Non-parametric Bayesian density estimation

A histogram was used to plot the data for Y_{2j}(test outcomes for the transition or suspicious group)**.**The curve fits for MFPT and kernel density for the posterior distribution of the test outcomes for the transition or suspicious group were computed. Figure 15 below shows the data, MFPT and kernel density plots.

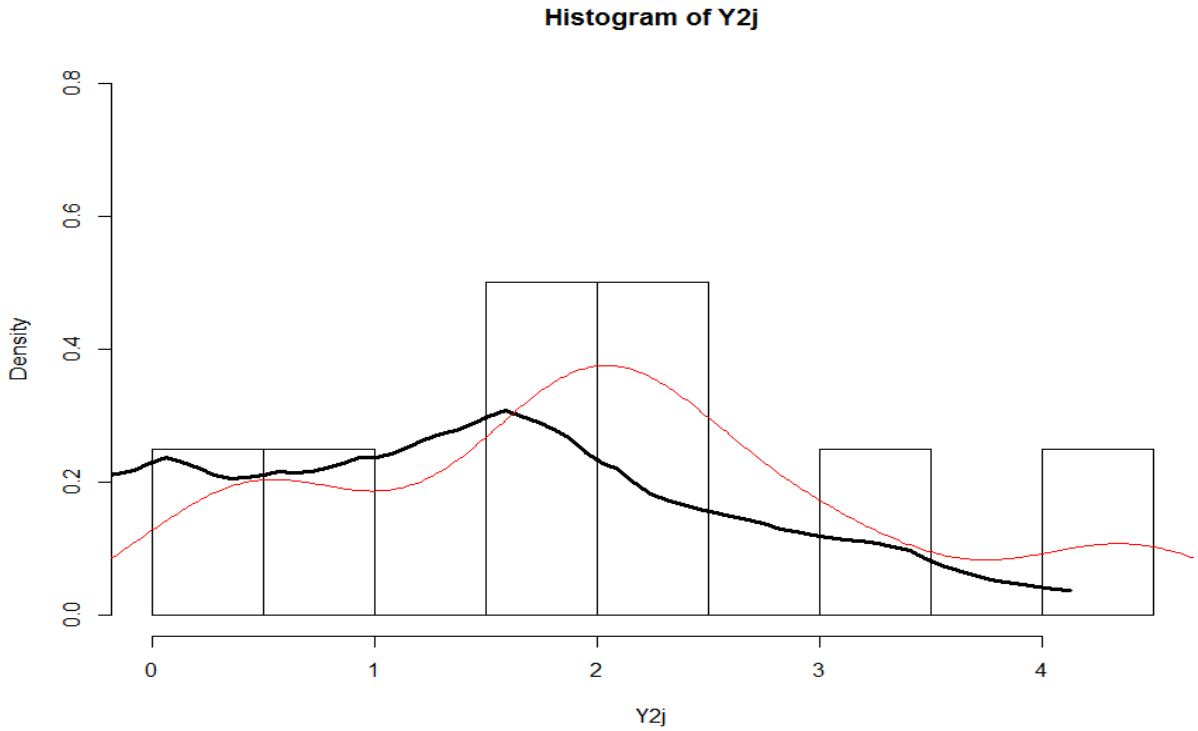

Figure 15: Distribution of Y_{2j} (test outcomes for the transition or suspicious group) under Non-parametric Bayesian density estimation

It was also found from the plot ofY_{2j} (test outcomes for the transition or suspicious group) data that the data assumes some distribution, evident in the existence of peaks in the histogram. It was also evident that the posterior distributions of the fit using MFPT indicate that the data fits the distribution well. Thekernel density smooth curve fit shows that the MFPT fits the data convincing, as the peak for the $\mu_{Y_{2j}}$lies at the true parameter value, that is, $\mu_{Y_{2j}} = 2$.

4.6.8 Distribution of Y_{3k}(test outcomes for the diseased group) under Non-parametric Bayesian density estimation

Finally, a histogram was plotted for Y_{3k}(test outcomes for the diseased group) data**.**On the same histogram, curve fits for MFPT and kernel density for the fitted distribution of the test outcomes were computed. Figure 16 below shows the data, MFPT and kernel density plots.

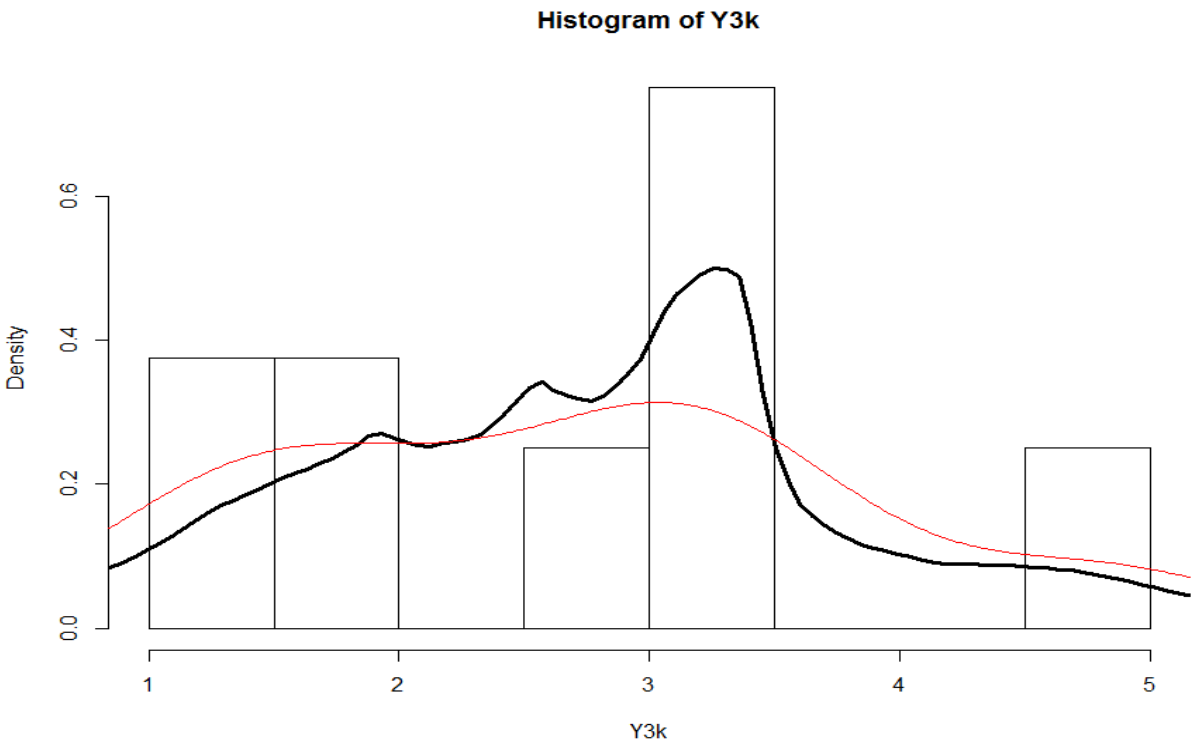

Figure 16: Distribution of Y_{3k} (test outcomes for the diseased group) under Non-parametric Bayesian density estimation

From the data plot of Y_{3k}(test outcomes for the diseased group), it was also found that the data assumes some distribution, evident in the existence of peaks. It is also evident that the posterior distributions of the fit using MFPT indicate that the data fits the distribution well. Thekernel density smooth curve fit shows that the MFPT fits the data well, as the peak lies at the true parameter value, that is, $\mu_{Y_{3k}}$=3.

4.7 Nonparametric ROC surface estimation

Since the properties of the parameters from the fitted non-parametric Bayesian Density Estimation Using MFPT distribution were desirable, random samples for inference for the ROC surface were drawn. Random samples for the three test outcomes using the Bayesian Density Estimation Using MFPT procedures were obtained. The samples represent a simulated diagnostic test that classifies outcomes into three groups. Table 10 below provides a summary of the test outcomes drawn from the distribution.

Table 10: Simulated diagnostic test Raw Data Summary under non-parametric estimation

	n=Sample size	μ= means
D- (non diseased)	100	0.54
Do (transition/suspicious)	50	2.34
D+ (diseased)	100	3.04

4.7.1 ROC surface plot

From the drawn data, a three way ROC surface corresponding to the three test outcomes was plotted. The model was used to estimate the ROC surface plot. Figure 17 below represents three-dimensional surface plot depicting trade-offs between the predictive measures for classification of the three test outcomes.

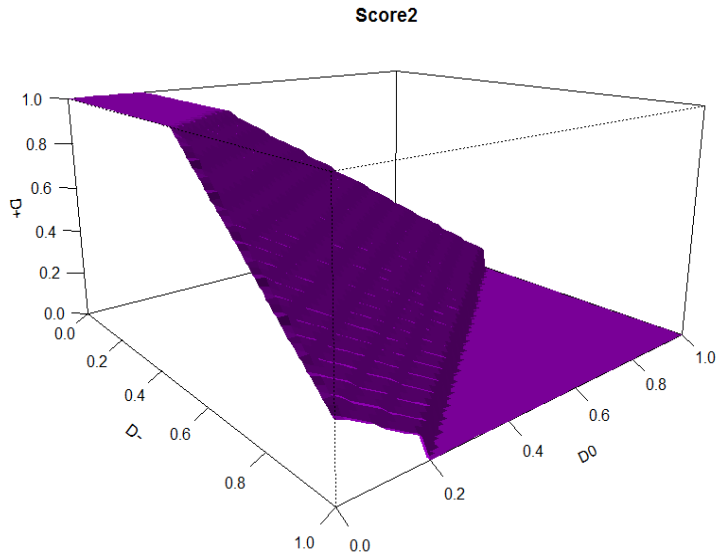

Figure 17: Three-dimensional surface plot-depicting tradeoffs between the predictive measures for classification of the three test outcomes under nonparametric estimation

Figure 17 above represents a plot of ROC surface for the nonparametric approach that uses a mixture of finite Polya trees (MFPT) model. It is evident that the model performs well though the ROC surface is not very smooth.

4.7.2 Nonparametric estimation of volume under ROC surface (VUS)

From the simulated diagnostic test representing a group of diagnostic test outcomes, volume under surface for the marker or group of test outcomes under the nonparametric methods was computed. It was found out that the volume under the surface VUS= 0.26. The 95% confidence interval was also computed. The lower confidence interval was found to be 0.23 while the upper confidence interval was found to be 0.34. Further, the optimal cut-points off points were derived. The best lower cut-point was found to be 0.0658 while the upper cut-point was found to be 0.204. The group correct classification probabilities were found to be; specificity= 0.36, true transition rate=0.56 and sensitivity=0.52. The estimate of the sample size=114 for the predefined precision was also derived. In other words, for the accuracy of the diagnostic test or group of test outcomes, a sample size of 114 will be

desirable for each group in order to estimate the VUS of the marker within a 5% margin of error.

4.7.3 Scatter plot and a boxplot for the non-parametric VUS

A generic Scatter plot and a boxplot to summarize the data graphically for the non-parametric VUS were also computed. Figure 18 below shows a scatter plot and a boxplot, with observations from D- , D0 and D+ colored in green, blue and red, respectively. The estimated summary measure along with the CI is provided in the legend while the optimal cut-points are labeled.

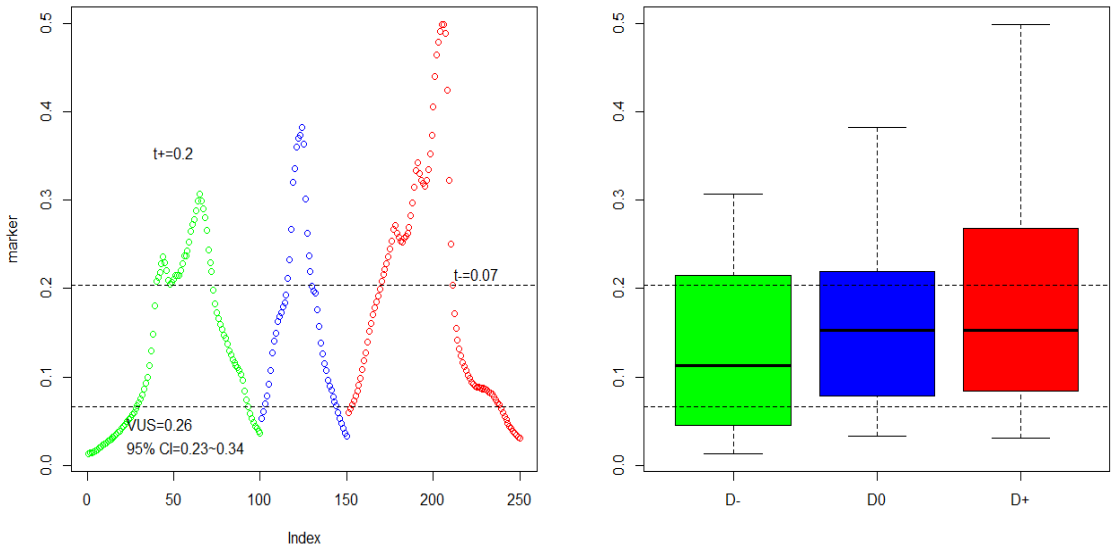

Figure 18: Scatter plot and boxplot of the marker under nonparametric estimation

From figure 18 above, it can be shown that the optimal cut off points for the marker represented by the dashed lines can be computed for the VUS. The box plots shows that the test outcomes show some ordering and that the test outcomes are ordinal in nature.

CHAPTER FIVE

DISCUSSION

The simulation studies were conducted to assess the semiparametric and nonparametric estimators of ROC surface. In every simulation, true-negative, true transition or suspicious and true-positive samples for the two methods were generated. Samples for the semiparametric approach; the true-negative class, true transition or suspicious and the true-positive class follows normal distributions with standard deviation of 1.5 with varying mean values. Similarly for non-parametric case, the true-negative class, true transition or suspicious and the true-positive class follows normal distributions with standard deviation of 1.5 with varying mean values. The generated samples were set to same initial conditions so that comparison between the performances of two approaches can be made. The set of simulations, test outcomes both studies were assumed to have different sample sizes according to definite class ratio 2:1:2 as recommended by (Luo & Xiong, 2012).

As expected, the methods perform better when the sample size is larger. The estimators had some drawbacks, and it may suffer from large variability, particularly for small sample sizes. However, this is not a major problem as small samples are common-place in clinical practice (Jokiel-Rokita& Pulit, 2013).

The parameter plots suggested that the Metropolis Hastings steps sampler used achieved convergence and were valid for subsequent analysis. It was deduced from the trace plots for all parametersthat inference based on the parameters was robust. In particular, the semiparametric posterior parameters were all stationary at the true parameter values and that all the MCMC chains were all convergent. Similarly, for the non-parametric estimators, the posterior parameter values were all stationary at the true parameter values and the chains were convergent; the acceptance rate for all

metropolis-hasting steps were all desirable.

That is, the high acceptance rate does indicate that the algorithm is behaving satisfactorily since it correspond to the fact that the chain is moving too faster on the surface.

The plots for the posterior distributions for the three test outcomes under the Semiparametric case (DPM of normals) were all smooth and symmetric except for Y_{3k}which had asymmetric data plot; this can be attributed to the small sample size used. On the other hand, the non-parametric posterior distribution (MFPT) plot for the three test outcomes portrayed unsmooth fits though the fit of MFPT model depicted some symmetry and adjacency to the kernel density fits. It was noted that the Kernel Density Estimation for both Semiparametric and non-parametric posterior distributions tend toward the true density therefore it is convincing that a much narrower variation on the sample values can be obtained. The Semiparametric posterior distribution fit was found to have smooth and more adjacent fit to kernel density estimate fits compared to the nonparametric posterior distribution fits.

The Semiparametric estimator provides a good mathematical model under assumed population distributions for the test results in three classes. It was also noticed that the choice of normal distribution in semiparametric model may be replaced by other well-known parametric families such as Weibull or gamma, where appropriate. Nonparametric estimator shows the observed diagnostic performance of the test among three classes reflects the sampling characteristics. uninformative independent priors for test outcomes, that is, $Y_{1i} \sim N\ (1, 1.5)$, $Y_{2j} \sim N\ (3, 1.5)$ and $Y_{3k} \sim N\ (3, 1.5)$ for both Semiparametric and non-parametric models were used. The semiparametric model is based on Dirichlet Processes of Normals and allows the entire distribution in

each group to smoothly change as a function of the covariates.

The results compared with the semiparametric approach of Carvalho et al. (2013) shows that the model performs competitively. The Bayesian nonparametric approach developed as based on mixtures of finite Polya trees priors outperforms the estimators proposed by In´acio (2012). However, the developed methods are prone to overlap between test outcomes.

The ROC surface under the Semiparametric and nonparametric models were found to have good coverage thus the two surfaces was considered useful to examine the diagnostic accuracy of the test for the three classes at different threshold values. It was notable, however, that the Semiparametric ROC surface had a smoother surface than the Nonparametric ROC surface; this can be attributed to the fact that Semiparametric case was a result of definite distribution, that is, mixture of normal distribution has smooth parametric fit.

The computed semiparametric and nonparametric ROC surface had similar methods used by Nakas and Yiannoutsos (2004). The estimated VUS under Semiparametric model was 0.34 and a 95% credible interval was (0.317, 0.460). On the other hand, the estimated VUS under Nonparametric model was 0.26 and a 95% credible interval was (0.232, 0.204). These values was contrasted to the VUS of a useless test or uninformative level of VUS=1/6 =0.167, which led us to conclude that the assumed marker has a reasonable discriminative power under both model assumptions. From these results, it can be concluded that for this particular analysis all methods lead substantially to the same conclusions.

The Semiparametric and Nonparametric approaches considered for estimating the ROC surface and the VUS indicate that even when the semiparametric assumption

holds, nonparametric estimation for Polya tree priors gives accurate results. These priors are especially useful when modelling data with nonstandard features, such as skewness and multimodality, as shown in the plot of posterior parameters for the MFPT case. Even with small sample sizes, the DPM of normals model and MFPT approach behaves quite well. Moreover, it was experienced that these approaches fits normal data well and is robust enough to fit data generated from other distributions for example gamma and for exponential distributions as mixtures of such distributions.

It was also noted that full inference is available with the Bayesian approach employed since the pair of thresholds that should be used to make the diagnostic decision was obtained. Nakas et al. (2010) earlier suggested that once models have been fitted, a posterior distribution for the optimal thresholds can easily be obtained.

It is noted also that there is no "rule of thumb" when choosing whether to use a Semiparametric or nonparametric model; the choice of models depends on the data. Moreover, since the DPM and MFPT priors are centered at a parametric family, they can inherit the overall shape of the underlying density when the weight parameter is relatively large (Jara, 2007).

Although the Semiparametric model yielded the best results, the MFPT is model is competitive as well. The two methods perform equally well in the simulation studies. The computation for MFPT method was more intensive.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1: Conclusions

In summary, diagnostic tests for three ordinal groups are important in biomedical practice. A useful summary measure (VUS) which can be adopted to evaluate the discriminative ability of a diagnostic test when there are three ordinal groups was computed.The applicability of DPM and MFPT models in solving difficulty in the modeling of continuous diagnostic data with skewness, multimodality or other nonstandard features were discussed. These data-driven models provided robust inference for the ROC surface and for the volume under the ROC surface (VUS).

Further, the summary of ROC surface under the semiparametric and nonparametric cases indicated that the two methods perform equally well in the simulation studies. The produced ROC surface plots has the appealing feature of being continuous and smooth, thus allowing for useful interpretation of the diagnostic performance at all thresholds. Comparably, the nonparametric ROC surface plot is less smooth compared to the semiparametric ROC surface plot. The surface plots produced are useful to examine the diagnostic accuracy of the test for the three classes at different threshold values. The three axes correspond to the probabilities of correct classification into the three groups. Volume under surface for semiparametric model was higher than the volume under surface for nonparametric model indicating higher accuracy. It can be attributed to the fact that the semiparametric model is largely centred on the Gaussian distribution.

It is noteworthy that semiparametric and nonparametric modelling of diagnostic data does not mean that there are no parameters in the models that were assumed. It is evident that the two models are massively parametric as the posterior parameters for the models were fitted.

The terms were used to indicate that the models are free of restrictive, inappropriate constraints that are implied particular parametric models.

6.2: Recommendations

A generalization of ROC curve to K=3; the computed ROC surface was developed in this research. It can be recommended that a generalization for more than three classes (K > 3), to produce a ROC hypersurface be developed. The Hypervolume Under the ROC Manifold (HUM) can be used for inferences for the ROC manifold especially in genetics where gene classifications involve several categories.

Further, most of the researchers have dealt with generalizations of theoretical findings from the two-class case and the geometric properties of the ROC surface. There is need for inference for multiple-class classification such as lack of ordinality in the test outcomes, (Li and Fine, 2008) where tests classifications were nominal, calls for more theoretical developments for the robustification of the framework of ROC surface analysis.

It is also recommend that researchers who are interested ROC surface methodology use the Comprehensive R Archive Network repository and contribute to further developments of R-packages for the implementation of ROC surface analysis in order to facilitate more researchers to implement ROC surface methodology.

REFERENCES

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics*, 2, 1152–74.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*. 12, 387-415.

Branscum, A. J., Johnson, W. O., Hanson, T. E. &Gardner, I. A. (2008). Bayesian semiparametric ROC curve estimation and disease diagnosis. *Statistics in Medicine*. 27, 2474–2496.

Cai, T., Pepe, M. S., Lumley, T., Zheng, Y. & J Enney, N. S. (2003). The sensitivity and specificity of markers for event times. *UW Biostatistics Working Paper Series*.

Campbell, G. (1994). Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Statistics in Medicine*. 13, 499-508.

Choi, Y., Johnson, W.O., Gardner, I. A., & Collins, M.T. (2006). Bayesian inference for receiver operating characteristic curves in the absence of a gold standard. *Journal of Agricultural Biology and Environmental Statistics*. 11: 210–229.

Dorfman, D.D. &Alf, J.E. (1968). Maximum-likelihood estimation of parameters of signal-detection theory-a direct solution. *Psychometrika*. 33, 113-124.

Dorfman, D.D. &Alf, J.E. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals-rating method data. *Mathematical Psychometrics.* 6, 487-496.

Drury, C.G. &Fox, J.G. (1975). *Human reliability in quality control*. Halsted, New York.

Dwyer, A.J. (1997). In pursuit of a piece of the ROC. *Radiology*. 202, 621-625.

Enrique, F.S, David, F. and Benjamin, R. (2004). Adjusting the generalized ROC curve for covariates. *Statistics in Medicine*. (23) 3319-3331.

Erkanli, A.,Sung, M.,Costello, E.J., & Angold, A.(2006). Bayesian semi-parametric ROC analysis. *Statistics in Medicine*. 25, 3905-3928.

Escobar, M. D. (1990). Estimating Normal Means With a Dirichlet Process Prior *Technical Report* 512, Department of Statistics: Carnegie Mellon University.

Escobar, M.D. &West, M. (1995) Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, 90: 577-588.

Fabius, J. (1964). Asymptotic behavior of Bayes' estimates. *Annals of the Institute of Statistical Mathematics*. 35, 846–856.

Ferguson, T. S (1983). *Bayesian Density Estimation by Mixtures of Normal*

Distributions, inRecent Advances in Statistics**,** eds. H. Rizvi and J. Rustagi, New York: Academic Press.

Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*,(1), 209-230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics* 2, 615–629.

Freedman, D. A. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. . *The Annals of Statistics*.34, 1194–1216.

Green, D.M. &Swets, J.A. (1966). *Signal detection theory and psychophysics*. Wiley, New York.

Gu, J., Ghosal, S. &Roy, A. (2008). Bayesian bootstrap estimation of ROC curve. *Statistics in Medicine*. 27, 5407–5420.

Hall, P. & Zhou, X. H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Annals of Statistics*.31, 201–224.

Han, A.K. (1987). Non-parametric analysis of a generalized regression model:the maximum rank correlation estimator. *Journal of Econometrics* (35) 303-316.

Hanley, J.A. & McNeil, B.J. (1982). The meaning and use of the area under an ROC curve. *Radiology*. 143, 29-36.

Hanson, T. & Johnson, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*, 97, 1020–33

Hanson, T. E. (2006). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association.* 101, 1548–1565.

Hanson, T. E., Kottas, A. &Branscum, A. J. (2008). Modelling stochastic order in the analysis of receiver operating characteristic data: Bayesian non-parametric approaches. *Journal of the Royal Statistical Society*: 57, 207–225.

Heckerling, P. S. (2001). Parametric three-way receiver operating characteristic surface analysis using Math-ematica. *Medical Decision Making*. 20, 409–417.

Hsieh, F. & Turnbull, B.W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics*. (24), 25-40

Inacio de Carvalho, V., Jara, A., Hanson, T. E. & de Carvalho, M. (2013). Bayesian nonparametric ROC regression modeling, *Bayesian Analysis*, 3, 623–646

Inácio V, Turkman AA, Nakas CT, Alonzo TA. Nonparametric Bayesian Estimation of the Three-Way Receiver Operating Characteristic Surface. *Biometrical Journal*. 53(6):1011–1024.

Jara, A. (2007). Applied Bayesian non- and semi-parametric inference using dpackage. *Rnews*, 17–26.

Jokiel-Rokita, A. & Pulit, M. (2013). Nonparametric estimation of the ROC curve based on smoothed empirical distribution functions, *Statistics andComputing*, 23, 703–712.

Kottas, A., Muller, P., & Quintana, F.(2005). Nonparametric Bayesian modeling for multivariate ordinal data. *Journal of Computational and Graphical Statistics*,14, 610–25.

Kraemer, H. C. (1992). *Evaluating Medical Tests*. Sage Publications.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modeling. *The Annals of Statistics.* 20, 1222–1235.

Lavine, M. (1994). More aspects of Polya tree distributions for statistical modeling. *The Annals of Statistics*. 22, 1161–1176.

Li, J. & Fine, J.P. (2008). ROC analysis for multiple classes and multiple categories and its application in microarray study. *Biostatistics*.9, 566–576.

Li, J., Zhou, X.H. (2009). Nonparametric and Semiparametric Estimation of the Three Way Receiver Operating Characteristic Surface. *Journal of Statistical Planning and Inference*. 139, 4133–4142.

Luo, J. & Xiong, C. (2012). DiagTest3Grp: An R Package for Analyzing Diagnostic Tests with Three Ordinal Groups. *Journal of Statistical Software*. 51(3), 1–24.

Lusted, L.B. (1971). Signal detectability and medical decision making. *Science*. 171, 1217-1219.

MacEachern, S. N. &Muller, P. (1998). Estimating mixture of Dirichlet Process Models.*Journal of Computational and Graphical Statistics*, 7 (2): 223-338.

McClish, D.K. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making* 9, 190-195.

Mossman, D. (1999). Three-way ROCs. *Medical Decision Making.* 19, 78–89

Nakas, C. T. and Yiannoutsos, C. T. (2004). Ordered multiple-class ROC analysis with continuous measurements. *Statistics in Medicine.* 23, 3437–3449.

Nakas, C. T., Alonzo, T. A. &Yiannoutsos, C. T. (2010). Accuracy and cut-off point selection in three-class classification problems using a generalization of the Youden index. *Statistics in Medicine.* 29, 2946–2955

Neal, R. M. (2000). Markov Chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9: 249-265.

Obuchowski, N.A. (2005). Estimating and comparing diagnostic tests accuracy when the gold standard is not binary. *Academic Radiology.* 12, 1198-1204.

Pepe, M. S. (2003). T*he Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford: Oxford University Press.

Pepe, M. S., Etzioni, R., Feng, Z. D., Potter, J., Thompson, M., Thornqui St, M., Winget, M. & Yasui, Y. (2001). Phases of biomarker development for early detection of cancer. *J. Natl. Cancer*. 93, 1054-61.

Peterson, W.W., Birdsall, T.G. and Fox, W.C. (1954). The theory of signal detection theory. Transactions of the IRE Professional Group on Information Theory 1, 171-212.

R Development Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Scurfield, B. K. (1996). Multiple-event forced-choice tasks in the theory of signal detectability. *Journal of Mathematical Psychology*.40: 253-269.

Sherman, R.P (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica (*61), 123-13.

Swets, J.A. (1977). *Vigilance: Relationships among theory, physiological correlates and operational performance*. New York: Plenum.

Swets, J.A. &Pickett, R.M. (1982). *Evaluation of diagnostic systems: Methods form signal detection theory*. Academic Press, New York.

West, M. (1990). *Bayesian kernel density estimation*. Institute of Statistics and Decision Sciences, Duke University.

Xiong, C., van Belle, G., Miller, J. P., &Morris, J. C. (2006). Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Statistics in Medicine*. 25, 1251-1273.

Zhou, X. H., Obuchowski, N.A. & McClish, D.K (2002). *Statistical methods in diagnostic medicine*. New York: Wiley.

Zou, K.H., Hall, W.J. & Shapiro, D.E. (1997). Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistic in Medicine*. 16, 2143-2156.

Zweig, M.H., & Campbell, G (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry.* 39: 561.