International Journal of Statistics and Probability; Vol. 10, No. 2; March 2021
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Estimation of Receiver Operating Characteristic Surface Using Mixtures of Finite Polya Trees (MFPT)

Ben Kiprono Koech

Correspondence: School of Science, Alupe University College, P O Box 9591-30100, Eldoret, Kenya. E-mail: benykip@gmail.com

Received: October 21, 2020 Accepted: January 21, 2021 Online Published: January 25, 2021

doi:10.5539/ijsp.v10n2p18 URL: https://doi.org/10.5539/ijsp.v10n2p18

Abstract

Generalisation of Receiver operating characteristic (ROC) curve has become increasingly useful in evaluating the performance of diagnostic tests that have more than binary outcomes. While parametric approaches have been widely used over the years, the limitations associated with parametric assumptions often make it difficult to modelling the volume under surface for data that do not meet criteria under parametric distributions. As such, estimation of ROC surface using nonparametric approaches have been proposed to obtained insights on available data. One of the common approaches to non-parametric estimation is the use of Bayesian models where assumptions about priors can be made then posterior distributions obtained which can then be used to model the data. This study uses Polya tree priors where mixtures of Polya trees approach was used to model simulated three-way ROC data. The results of VUS estimation which is considered a suitable inference in evaluating performance of a diagnostic test, indicated that the mixtures of Polya trees model fitted well the ROC surface data. Further, the model performed relatively well compared to parametric and semiparametric models under similar assumptions.

Keywords: non-parametric estimation, mixtures of finite polya trees, receiver operating characteristics, volume under surface

1. Introduction

Generalisation of Receiver operating characteristic (ROC) curve in assessing how diagnostic tests perform can be a challenging task (Koech, 2018) especially since the tests more than two outcomes leading to multiple true class rates and false class rates. Non- parametric approaches can be used in modelling the volume under surface which can be used in measuring the accuracy of a diagnostic test. Ferguson (1983), West (1990) and Escobar & West (1995) are some of the notable scholarly works that have underpinned the use of non-parametric methods in in estimating the accuracy of diagnostic test. They have suggested that when observations on some random variable follow a distribution which is assumed to be is a random sample function of a non-parametric scholastic process (Dirichlet process), then the random measure's conditional distribution can be calculated. The assumption in this case are derived from Bayesian viewpoint.

Studies, such as Choi et al. (2006) have proposed Bayesian parametric multivariate ROC methodology as alternative to Gaussian estimation of diagnostic tests owing to the limitations of normality assumptions in cases especially rare diseases that have not been well studied. This argument follows some authors such as Hall and Zhou (2003) who faulted the normality assumptions in estimation of diagnostic tests as they developed multivariate distribution-free methods for estimation in biomedical studies. Studies by Erkanli et al. (2006) examined the use of non-parametric methods for binary outcome tests based on truncated Dirichlet process mixture model while Branscum et al. (2006) estimated the accuracy of diagnostic test when true infection status of a disease is unknown based on the mixture of finite Polya trees model. Similarly, Hanson et al (2008) estimated several ROC curves using the mixtures of Polya trees (MFPT) in estimation of multivariate serologic data.

Lavine (1992, 1994) and Hanson (2006) are considered as notable pioneers of computation based on finite Polya tree priors. On the other hand, Branscum et al. (2008) and in Hanson et al. (2008). According to Lavine (1992), Polya trees form a class of distributions for measure-valued random element (random probability measure). This can be considered as an intermediate between Dirichlet processes and tailfree processes (Tail-free processes can be defined as stochastic processes that can possess trajectories on the space relating to specified 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. Hanson (2006) also noted that Polya tree priors have the advantage in that some sampling situations

results in posterior mixtures of Dirichlet processes, can only lead to only one posterior Polya tree.

According to Hanson and Johnson (2002) MFPT can be considered as a more robust method for data driven models in in statistical modelling which includes evaluating the diagnostic testing. Regression models such as generalized linear models (Hanson, 2006) can be implemented under MFPT. The flexibility offered by the Polya tree priors allows for modelling of continuous distributions (Ferguson, 1974) which overcomes some of the notable limitations associated with parametric frequentist approaches (Heckerling, 2001). Li and Zhou (2009) estimation of ROC surface using both nonparametric and semiparametric approaches indicated that both approaches were robust in estimating the accuracy of three-way diagnostic outcomes. However, they noted that accuracy improved under their nonparametric approach and pointed out that the limitation of normality assumption for semiparametric approach as rare diseases and epidemics often do not follow Gaussian distributions. This view was corroborated by Inácio et al. (2012) who suggested that normality assumption under parametric models often fail to capture multimodality, nonstandard features, or skewness within data. As such, they proposed Bayesian nonparametric approach based on MFPT to estimate receiver operating characteristic surface.

Some of the merits of MFPT in estimating ROC surface as Inácio et al. (2012) suggested include the ability to broaden class of models as it may apply to large number of diseases, any population and varied diagnostic measures. More so, MFPT accommodates various forms of data where parametric distributions can be incorporated in computing larger non-parametric models. The ability of generalization under MFPT makes it robust and can allow for deeper insights into data which are becoming increasingly available in the age of data analytics and increase in computational power among available data analytics software and hardware solutions. Nevertheless, it would be reasonable to note that computation of MFPT and other non-parametric models can be difficult (Hanson, 2006) while implementation of the models may yet not be available within the statistical softwares available in markets. This poses significant challenge when working with such models. However, the possibility of computation of MFPT for ROC data motivated the current study. The aim of the study was to estimate receiver operating characteristic surface using MFPT based on simulated data.

2. Methods

The nonparametric model developed is based on finite Polya tree priors (Freedman, 1963; Fabius, 1964; Ferguson, 1973) for the distributions F_1, F_2 and F_3 corresponding to the three classes of outcomes in a three-way diagnostic test. To obtain the nonparametric estimator of ROC surface, all the distribution functions in (F_1, F_2 and F_3) shall be replaced with their empirical counterparts. The estimator shall be constructed as:

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

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

The general non-parametric model is

$$Y_{1i} \sim F_1,\ Y_{2j} \sim F_2 \text{ , and } Y_{3k} \sim F_3$$
$$F_i|c_i, \theta_i \sim \mathrm{FPT}_j(F_{\theta_i}, c_i),$$
$$\theta_i \sim p(\,\mathrm{d}\theta_i)$$

Random F_i is centred at $F_{\theta_i} = N(\mu_i, \sigma_i)$ 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 Gamma $(a_{\sigma_i}, b_{\sigma_i})$ which are independent gamma priors. The parameters are all with fixed hyperparameters. In this study, the finite Polya trees levels are represented by J_i. Similarly, the weight parameter c_i was also fixed. The likelihood function is proportional to

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

For equation (2), the f_i is a density function corresponding to distribution functions F_i. Furthermore, the $F_i(y|\chi_i, \theta_i)$ represents the cumulative distribution function while Markov Chain Monte Carlo (MCMC) methods can be used to approximate the joint posterior distribution for the 3 classes. The computation of the ROC surface plot can be computed given the likelihood function (Equation 2). MCMC with simple Metropolis–Hastings steps was used to fit the MFPT model for ROC surface where updating of the mixing parameters μ_i and σ_i followed random walk process.

All the analysis was implemented using R Statistical Computing program. Data was simulated for the three classes obtained from different populations drawn using normal random sample generators. For the distributions F_1, F_2 and F_3 the corresponding random variables were $Y_{1i} \sim N(1,1.5), Y_{2j} \sim N(2,1.5)$ and $Y_{3k} \sim N(3,1.5)$, where sample sizes

n_1 and n_2, and $n_3$were set to 16, 8 and 16 respectively. The random variables Y_{1i} represent the first group of test outcomes (non-diseased or healthy subjects) while Y_{2j} represent the second group of test outcomes (intermediate or transition subjects) who are at high risk of developing a condition. Further, the random variables Y_{3k} represent the last group of test outcomes (diseased or with condition subjects) such as in the case of Alzheimer's disease where the progression of the disease increases over time. An individual tested for Alzheimer's disease may be diagnosed as negative then falls in Y_{1i} or high risk Y_{2j} or Y_{3k} when a patient demonstrates apparent characteristics and symptoms of the disease.

It was also assumed that the Polya Tree is centred 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). It was further assumed that α, the precision parameter of the Polya Tree prior, $\alpha = 1$, and was considered as random. It was also assumed that the precision parameter followed gamma distribution, with hyperparameters for prior distribution being $\Gamma(a_0, b_0)$. The posterior distribution of the baseline as well as the precision parameters were implemented using Metropolis-Hastings steps as indicated earlier. The number of grid points for which the evaluation of the density estimate was considered, was set to ratio proportionate to the simulated sample size, that is, 2:1:2. That is, it was assumed 100, 50 and 100 for n_1, n_2 and n_3 respectively. This constituted the simulated case of a three-way diagnostic test. 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.

3. Analysis of Data and Results

3.1 Posterior Inference of Parameters for Y_{1i} Y_{2j}, and Y_{3k}

The study sought to examine the properties of the fitted Non-parametric distribution (that is under MFPT). More specifically, the posterior inference of parameters under non-parametric Bayesian density estimation was considered. The results for the fitted distribution for the test outcomes $Y_{1i,}$ 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. Likewise, the test outcomes Y_{2j} were fitted whereby it had 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. Finally, the test outcomes Y_{3k} where its posterior Inference of Parameters were $\mu_{Y_{3k}} = 3.22$ and $\sigma_{Y_{3k}} = 1.41$ while the acceptance rate for the Metropolis Step = 0.70 to 0.77. The results can be summarized in table 1 below.

Table 1. Posterior Inference of Parameters for $\boldsymbol{Y_{1i}}$ $\boldsymbol{Y_{2j}}$, and $\boldsymbol{Y_{3k}}$

$\boldsymbol{Y_{1i}}$		
$\mu_{Y_{1i}}$	$\sigma_{Y_{1i}}$	α
1.22	1.97	1.00
Acceptance Rate for Metropolis Step = [0.73 0.78]		
$\boldsymbol{Y_{2j}}$		
$\mu_{Y_{2j}}$	$\sigma_{Y_{2j}}$	α
2.18	1.70	1.00
Acceptance Rate for Metropolis Step = [0.78 0.85]		
$\boldsymbol{Y_{3k}}$		
$\mu_{Y_{3k}}$	$\sigma_{Y_{3k}}$	α
3.22	1.41	1.00
Acceptance Rate for Metropolis Step = [0.70 0.77]		

Further, to analyse 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} 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 1 gives a summary of the parameter plots.

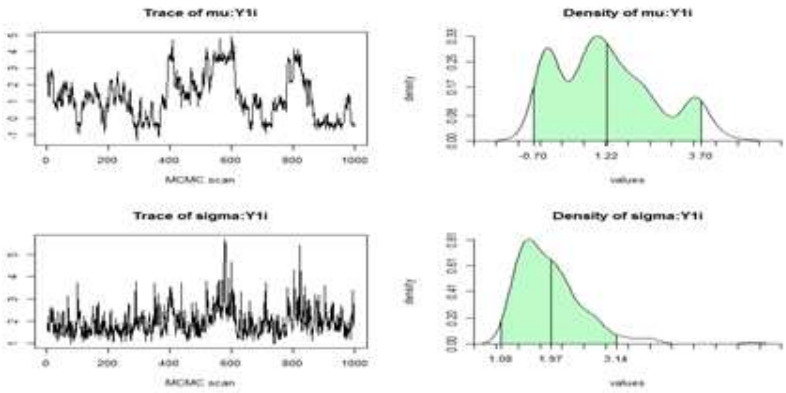

Figure 1. Posterior parameters for $\boldsymbol{Y_{1i}}$ under Non-parametric Bayesian density estimation

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. Similarly for Y_{2j} 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 2 gives a summary of the parameter plots.

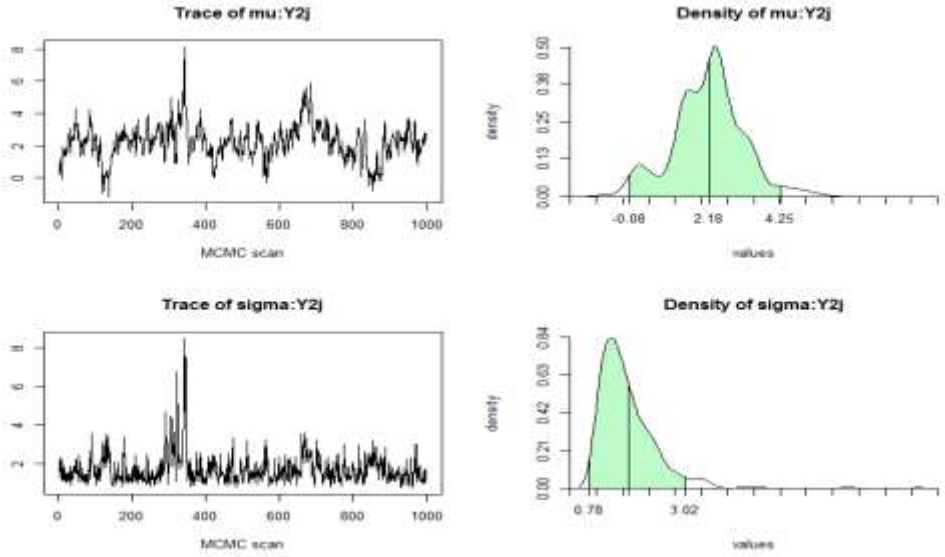

Figure 2. Posterior parameters for $\boldsymbol{Y_{2j}}$under Non-parametric Bayesian density estimation

It was concluded from the time series trace plots that the parameter estimates for Y_{2j} 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.

As well, the parameter plots for the fitted posterior parameters $\mu_{Y_{3k}}$ and $\sigma_{Y_{3k}}$ for Y_{3k} were computed. Similar to previous test outcomes parameter plots; the plots for MCMC scans and fitted histogram line for the parameter values were computed. Figure 3 gives a summary of the parameter plots.

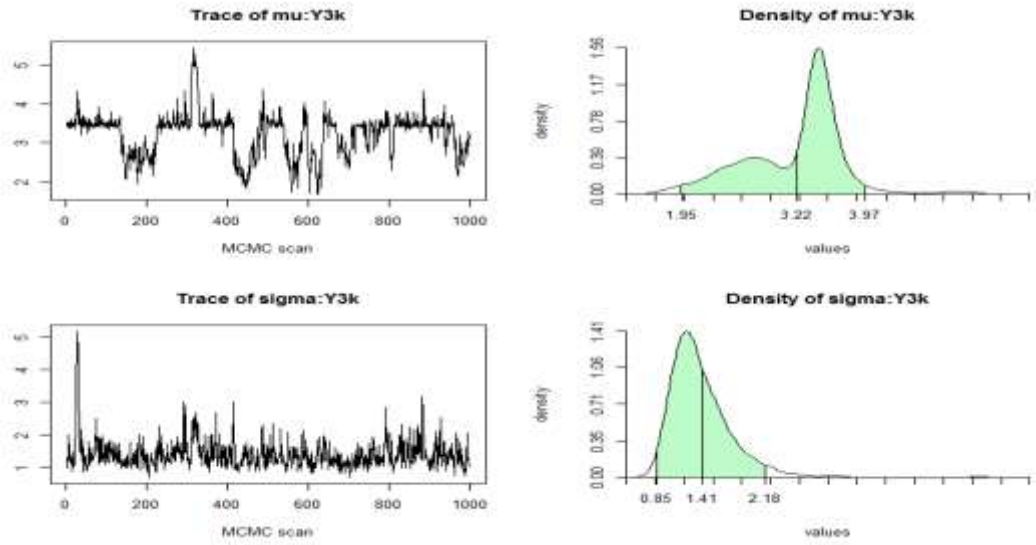

Figure 3. Posterior parameters for $\boldsymbol{Y_{3k}}$ under Non-parametric Bayesian density estimation

It was evident that the MCMC chains were convincingly stationary at the true parameter estimates for posterior parameters $\mu_{Y_{3k}}$ and $\sigma_{Y_{3k}}$for Y_{3k}. Further, the plots of the parameter estimates; mean and standard deviation show definite smoothness. Overall, the posterior parameters for the three simulated test outcomes indicate that posterior estimates; means and standard deviations converged or are stationary after 1000 iterations.

3.2 Data Plot for Y_{1i} Y_{2j}, and Y_{3k}

Data plots for the three classes were produced to observe the distributions of the three test outcomes by plotting the data using histograms. Histogram of the data points and curve fits (MFPT distribution and kernel density estimate curve fits) were obtained for Y_{1i} Y_{2j}, and Y_{3k}. The black continuous line represents the MFPT distribution curve fits for the posterior estimates while the red continuous line represents the kernel density estimate curve fits for the posterior estimates. The histogram of the data plot for Y_{1i} depicts that it follows some distribution, apparent in data peaks. The kernel 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, $\mu_{Y_{1i}} = 1$. Figure 4 shows summary of the result.

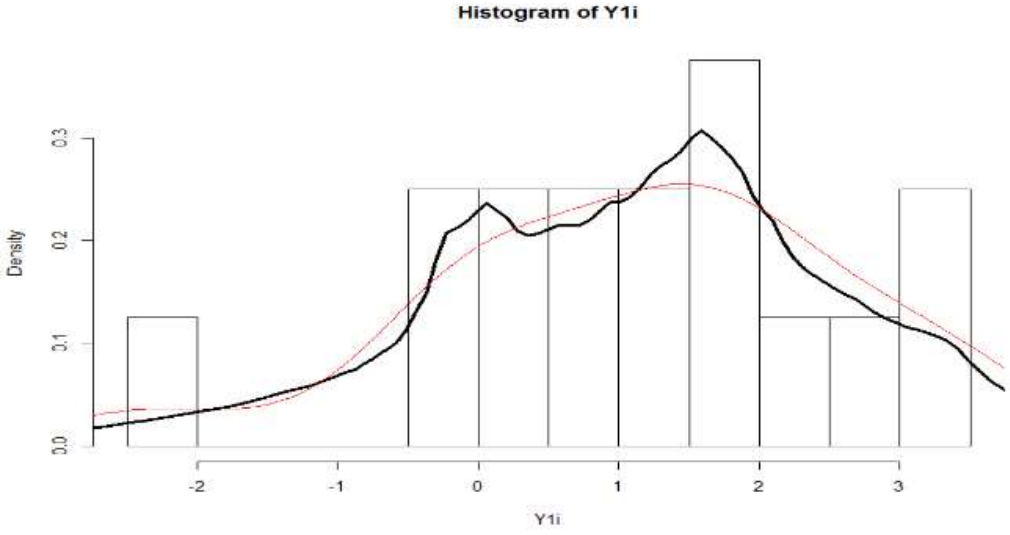

Figure 4. Distribution of $\boldsymbol{Y_{1i}}$

As well, a histogram was used to plot the data for Y_{2i}**.** The curve fits for MFPT and kernel density for the posterior distribution of the test outcomes were computed. It was also found from the plot of Y_{2j} indicates that the data assumes some distribution, as there were peaks in the histogram. The kernel density smooth curve fit shows that the MFPT fits the data considerably, as the peak for the $\mu_{Y_{2j}}$lies at the true parameter value, that is, $\mu_{Y_{2j}} = 2$. Figure 5 shows the

data, MFPT and kernel density plots.

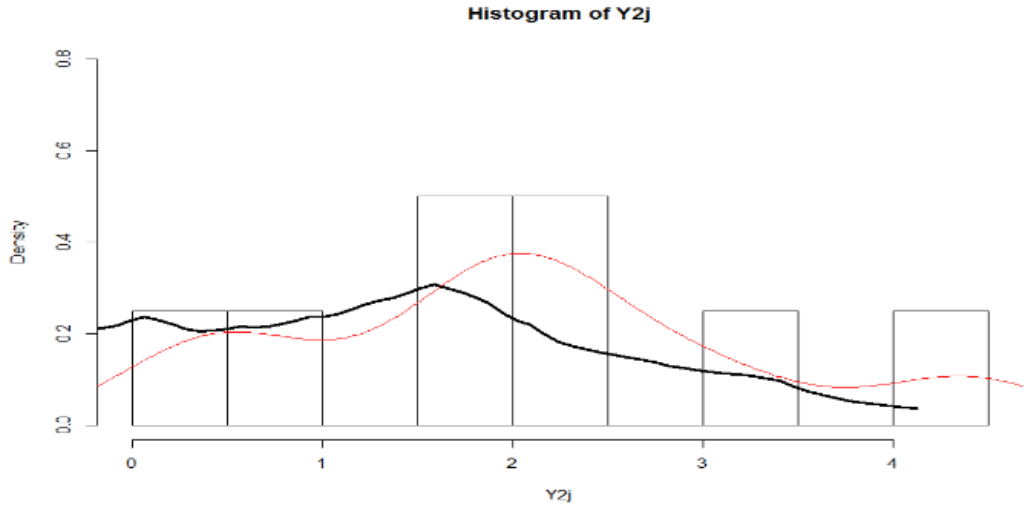

Figure 5. Distribution of $\boldsymbol{Y_{2j}}$

Finally, a histogram was plotted for Y_{3k} data**.** On the same histogram, curve fits for MFPT and kernel density for the fitted distribution of the test outcomes were computed. The data plot of Y_{3k} shows that the data assumes some distribution, evident in the existence of peaks. The kernel 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. Figure 6 shows the MFPT and kernel density plots for Y_{3k}.

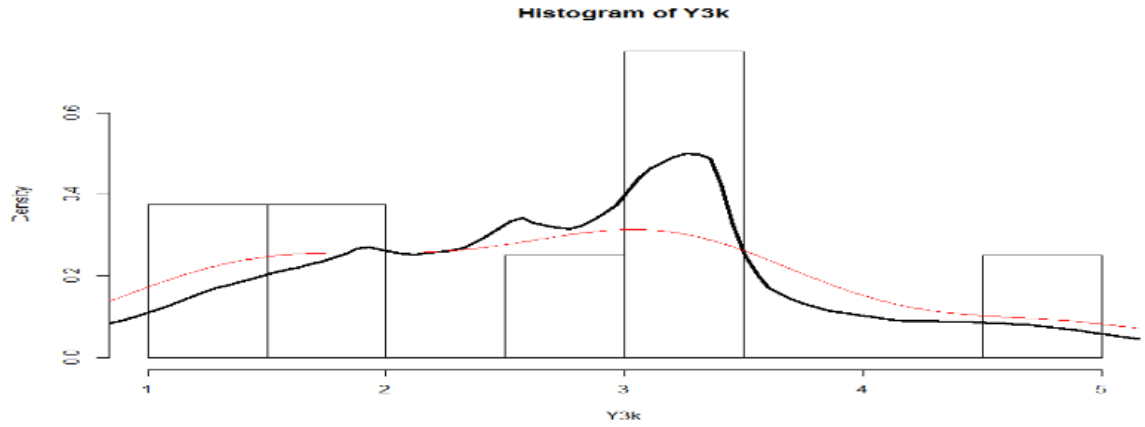

Figure 6. Distribution of $\boldsymbol{Y_{3k}}$

Overall, it was evident that for Y_{1i} Y_{2j}, and $Y_{3k}$the posterior distributions of MFPT indicate that the data fits the distribution well. Therefore, 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 could be drawn.

3.3 ROC Surface Estimation

Random samples were obtained for the three test outcomes. They represent simulated test outcomes groups of non-diseased or healthy subjects, intermediate or transition subjects and diseased or with condition subjects. Table 2 below provides a summary of the sample sizes and ordered means for the test outcomes drawn from the non-parametric distribution.

Table 2. Raw data for the simulated diagnostic test

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

From the data in table 2, ROC surface plot was obtained for the three-test outcome. The plot represents a plot for true positive fraction (TPF), against false positive fraction (FPF) and intermediate fraction (IDF). As such, it depicts *trade-offs* between the predictive measures as used in the three-way classification of test outcomes. Figure 7 shows a graphical illustration of ROC surface of the simulated diagnostic test using MFPT model.

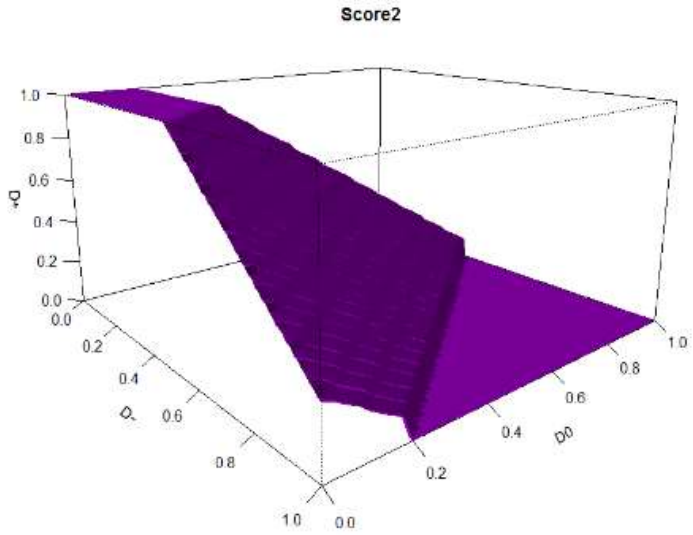

Figure 7. ROC surface plot under MFPT estimation

It is evident that the model performs well though the ROC surface is not very smooth. It is desired that the cut of points chosen would minimize FPF and maximize TPF. However, the examination of ROC surface plot indicates that there are *trade-offs*, that is, FPF increases with increase in TPF while a decrease in IDF causes TPF as well as TPF . This highlights the challenge of trade-offs between sensitivity (probability of correctly identifying positive status or subjects with a condition) and specificity (probability of correctly identifying negative status or subjects without a condition). From the inference about the ROC surface plot, it may not provide conclusive evidence on the efficiency of MFPT model hence volume under the surface (VUS) can be used for further inference.

Therefore, estimation of VUS was considered to obtain the overall performance of the test under MFPT model. To observe the relationships in VUS, a boxplot and scatter plot were used to summarize the data graphically for the non-parametric VUS. The optimal cut-points are labelled with dashed lines in the scatter plot while the estimated summary measure and corresponding CI were provided in the legend. As well, the box plot suggested that the test outcomes were are ordinal in nature evident in the varying positions of the box and whiskers. Figure 8 shows a boxplot and scatter plot and, with observations from non-diseased or healthy subjects, intermediate or transition subjects and diseased or with condition subjects coloured in green, blue and red, respectively.

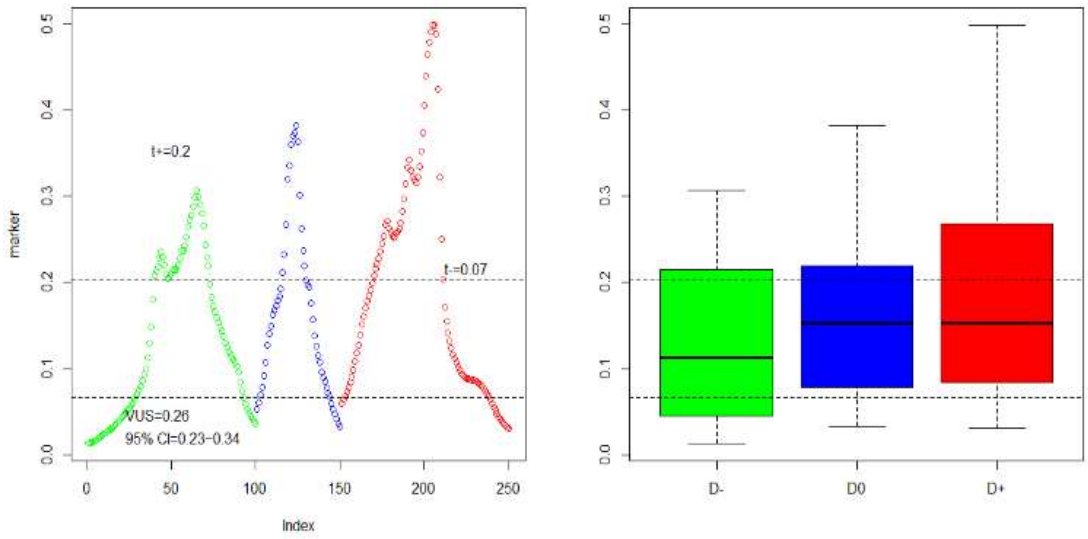

Figure 8. Boxplot and Scatter plot of VUS for the three groups

From the simulated diagnostic test representing a group of diagnostic test outcomes, VUS for the marker or group of test outcomes was computed. The computed VUS was found to be 0.26 with 95% confidence interval [0.23, 0.34]. Furthermore, the optimal cut-points off points were calculated for the samples whereby the best lower cut-point was estimated to be 0.0658 while 0.204 was the upper cut-point. The upper cut off point corresponds to IDF while lower cut-point corresponds to TPF and FPF where the selection of the optimal cut points depicts the optimal trade-off among IDF, TPF and FPF quantities. As such the three quantities, IDF, TPF and FPF were found to be 0.56, 0.52 and 0.36 respectively. The computation also suggested that the sample size of 114 was ideal for the sample size=114. That is, within a 5% margin of error, a sample size of 114 or more would be desirable for each test outcome group to estimate the VUS of the marker.

4. Discussion and Conclusions

The simulation study sought to examine whether the non-parametric estimator was efficient in modelling ROC and provide reasonable inference on performance of a three-way diagnostic test. Luo & Xiong (2012) suggestion of definite class ratio 2:1:2 for simulation of true-negative, intermediate, or suspicious and true-positive samples was considered. More specifically, the posterior parameters for MFPT estimation depicted desirable properties of stationarity at true values while the convergence of the MCMC chains were achieved. As such, there were high acceptance rate for the metropolis-hasting steps. Nevertheless, the MFPT posterior distributions depicted unsmooth fits though there was relative contiguity and symmetry to the kernel density fits. Further, the MFPT estimator shows the sampling characteristics influenced the diagnostic test performance for the three groups of diagnostic test outcomes. That is, the uninformative independent priors for test outcomes, $Y_{1i} \sim N(1, 1.5)$, $Y_{2i} \sim N(3, 1.5)$ and $Y_{3k} \sim N(3, 1.5)$ for the non-parametric model used. Comparing the results with Carvalho et al. (2013) and Koech (2018) semiparametric approaches, the non-parametric model appears to perform competitively. Interestingly, the MFPT model under same conditions as In ́acio (2012) appears to be competitive and accurate. Notably, the VUS obtained was 0.26 with 95% confidence interval [0.23, 0.34] which is higher compared to a 'useless test' (1/3!) corresponding to three groups of a three-way diagnostic tests = 0.167. Consequently, MFPT estimator appears to have a realistic discriminative power. Nevertheless, the means of the groups were set to be ordinal in nature to depict a progression of a condition (in bioassays) for which a diagnostic test would discriminate between the three groups. It was noted that the assumptions would be invalid especially if three-way diagnostic outcomes are not ordinal in nature or if samples are relatively small or if the test outcomes overlap. However, Jokiel-Rokita & Pulit (2013) suggested that small sample sizes are common clinical practise especially for rare diseases for which diagnostic tests are not well developed hence the current study may be used for inference.

Overall, the estimation of VUS under MFPT appears to be a promising area of discussing given that it produced desirable results even under assumptions of relatively small sample sizes, data with nonstandard features (multimodality and skewness). It is proposed that the model can be fit on data generated from other distributions such as exponential and gamma distributions as mixtures of distributions as long as posterior distributions can be obtained and formula for

ROC surface can be integrated or approximated. Also, hypervolume under surface can be estimated to deal with multiclass diagnostic test outcomes and tools be developed to implement ROC surface methodology for multiclass outcomes.

Acknowledgments

I acknowledge the partial funding by National Commission for Science, Technology and Innovation (NACOSTI) under the grant number NACOSTI/RCD/ST&1/5th call MSC/132. The funding went a long way in facilitating the computational aspects required for the simulation process and subsequently completing the study.

References

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. https://doi.org/10.1002/sim.3250

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. https://doi.org/10.1198/108571106X110883

Erkanli, A., Sung, M., Costello, E. J., & Angold, A. (2006). Bayesian semi-parametric ROC analysis. *Statistics in Medicine*, *25*, 3905-3928. https://doi.org/10.1002/sim.2496

Escobar, M. D., & West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*, *90*, 577-588. https://doi.org/10.1080/01621459.1995.10476550

Fabius, J. (1964). Asymptotic behavior of Bayes' estimates. *Annals of the Institute of Statistical Mathematics*, *35*, 846–856. https://doi.org/10.1214/aoms/1177703584

Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, (1), 209-230. https://doi.org/10.1214/aos/1176342360

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics*, *2*, 615-629. https://doi.org/10.1214/aos/1176342360

Ferguson, T. S. (1983). *Bayesian Density Estimation by Mixtures of Normal Distributions, in Recent Advances in Statistic*s. H. Rizvi and J. Rustagi, New York: Academic Press.

Freedman, D. A. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. *The Annals of Statistics*, *34*, 1194–1216. https://doi.org/10.1214/aoms/1177703856

Hall, P., & Zhou, X. H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Annals of Statistics*, *31*, 201-224. https://doi.org/10.1214/aos/1046294462

Hanson, T. E. (2006). Inference for mixtures of finite Polya tree models. *Journal of the American Statistical Association*, *101*(476), 1548-1565. https://doi.org/10.1198/016214506000000384

Hanson, T. E., Branscum, A. J., & Gardner, I. A. (2008). Multivariate mixtures of Polya trees for modeling ROC data. *Statistical Modelling*, *8*(1), 81-96. https://doi.org/10.1177/1471082X0700800106

Hanson, T., & Johnson, W. O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association*, *97*(460), 1020-1033. https://doi.org/10.1198/016214502388618843

Heckerling, P. S. (2001). Parametric three-way receiver operating characteristic surface analysis using Math-ematica. *Medical Decision Making*, *20*, 409-417. https://doi.org/10.1177/02729890122062703

Inacio de Carvalho, V., Jara, A., Hanson, T. E., & de Carvalho, M. (2013). Bayesian nonparametric ROC regression modeling. *Bayesian Analysis*, *3*, 623–646. https://doi.org/10.1214/13-BA825

Inácio, V., Turkman, A. A., Nakas, C. T., & Alonzo, T. A. (2012). Nonparametric Bayesian Estimation of the Three-Way Receiver Operating Characteristic Surface. *Biometrical Journal*, *53*(6), 1011–1024. https://doi.org/10.1002/bimj.201100070

Jokiel-Rokita, A., & Pulit, M. (2013). Nonparametric estimation of the ROC curve based on smoothed empirical distribution functions, *Statistics and Computing*, *23*, 703–712. https://doi.org/10.1007/s11222-012-9340-x

Koech, B. K. (2018). Semiparametric Estimation of Receiver Operating Characteristic Surface. *American Journal of Applied Mathematics and Statistics*, *6*(6), 218-223.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *The annals of statistics*, *20*(3), 1222-1235. https://doi.org/10.1214/aos/1176348767

Lavine, M. (1994). More aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, *22*(3),

1161-1176. https://doi.org/10.1214/aos/1176325623

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. https://doi.org/10.1016/j.jspi.2009.05.043

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. https://doi.org/10.18637/jss.v051.i03

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