

ABSTRACT

GU, JIEZHUN. Nonparametric and Semiparametric Inference About ROC Curves. (Under the direction of Dr. Subhashis Ghosal).

Receiver Operating Characteristic (ROC) curve is defined as the true positive fraction versus the false positive fraction obtained by varying a decision threshold criterion. It has been widely used in medical science for its ability to measure the accuracy of diagnostic or prognostic tests. Many parametric and nonparametric estimation methods have been proposed for estimating the ROC curve and its functionals.

Although a nonparametric approach allows a lot of flexibility in a model, sometimes it is desirable to incorporate specific parametric features into the model, especially for better interpretability. A binormal model for ROC analysis is the simplest and most popular one among semiparametric models. This model has both parametric and nonparametric components and provides a natural balance between flexibility and interpretability. Under the binormality assumption, the measurements are normally distributed after some unknown strictly increasing transformation. Assuming binormality, we propose a conceptually simple and computationally accessible Bayesian estimation method using a partial likelihood (BPL) based on ranks. This likelihood is defined to be the probability of the transformed observations keeping the same ranks and labels as the originals given the binormal parameters. However, directly computing the maximum partial likelihood estimates involves calculating multiple integrals based on the multivariate normal distribution. We propose the Bayesian approach to calculate the posterior mean of the parameters using an augmentation technique. We find that our proposed method gives more accurate estimates and confidence intervals than other alternatives.

In this thesis, we then propose a fully nonparametric Bayesian bootstrap (BB) estimation method for the ROC curve and its functionals. The BB method performs

well in terms of robustness, simplicity and smoothness without choosing a bandwidth. In order to study the asymptotic behavior of the BB version of the ROC process, we prove a strong approximation result for the quantile processes based on the BB resampling distributions. Combined with other strong approximation results, we obtain the strong approximation for the BB version of the ROC process in terms of two independent Kiefer processes. The results imply asymptotically accurate coverage probabilities for the Bayesian bootstrap confidence bands, and accurate frequentist coverage probabilities of the Bayesian bootstrap confidence intervals for the area under the curve functional of the ROC. We also show that the BB method is useful in testing the assumption of binormality. We estimate the binormality transformation function using the BB distribution based on non-diseased samples. Then, the parameters for normality can be estimated by the transformed diseased samples. The proposed test statistic is defined as the supremum of absolute differences between the resulting binormal ROC curve and the BB version's estimates. We reject the binormality assumption when the statistic is larger than the threshold value, which can be obtained by a resampling scheme.

All of the methods proposed above are based on the gold standard for the truth, which means that the true disease status of each patient is known by the most accurate available diagnostic test. However, in complicated medical practice, due to ethical considerations, and because the gold standard test could be invasive and expensive, patients are given the gold standard test only when there is some indication of the presence of the disease in screening tests. Moreover, the gold standard test for some diseases may not be available for a particular time point. When only one continuous test variable is available without the gold standard, deriving the ROC curve estimation is not straightforward. By observing that the binormality assumption allows us to identify the underlying components of non-diseased and diseased populations up to some strictly increasing transformation H ,

we will be able to estimate the ROC curve without the gold standard. In order to estimate the parameters in the mixture of two normals, we propose a direct approach which models the transformation using the Bernstein polynomials. A slightly indirect approach is to treat H as latent and construct estimates using a rank based likelihood as before. These two methods are implemented using MCMC techniques. From simulation results and real data analysis, we conclude that the latent model works quite well without the gold standard.

Nonparametric and Semiparametric Inference About ROC Curves

by

Jiezhun Gu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

STATISTICS

Raleigh, North Carolina
2007

APPROVED BY:

Dr. Subhashis Ghosal
Chair of Advisory Committee

Dr. Bibhuti Bhattacharyya

Dr. Montserrat Fuentes

Dr. Daowen Zhang

Dedication

To my parents, my husband Jiwei Chen and my son Zitianyuan Chen.

Biography

Jiezhun Gu was born on December 2, 1967 in Shanghai, China. She got her Bachelor of Science degree in Probability and Statistics from Peking University, Beijing, China in July 1990. After that, she worked sequentially as a software engineer, a statistician and an auditor in Shanghai, China. To seek a graduate education, she returned to school with 11 years' working experience. She got her Master degree in Biomathematics at North Carolina State University (NCSU) in August 2005 and she has been to pursue a Ph.D. degree in Statistics at NCSU since then. Her research focuses on the methodology of Nonparametric Bayesian and its applications, including Receiver operating characteristic (ROC) curve, clinical trial, and among others.

Acknowledgements

This work has been supervised by Dr. Subhashis Ghosal. It is a great pleasure to be his student working under his outstanding education, suggestion and communication to me. I am grateful for his excellent advice and financial supports funded by NSF. Without his help, I would not be able to finish my Ph.D. degree at this stage. Also, I am grateful for his example not only in the strict academic attitude, but also the outlook of the world. In the mean time, I would like to thank all the faculty members and students in our Department of Statistics, especially Department head Dr. Sastry Pantula, my Ph.D. committee members, Dr. Sujit Ghosh, Dr. Hao (Helen) Zhang, Dr. Wenbin Lu, Dr. Dennis Boos, Dr. James F. Selgrade, Dr. Charlie Smith, Dr. William (Bill) Swallow, Dr. Leonard (Len) Stefanski, Dr. Pam Arroway and Terry Byron, Adrian Blue and Sasha Miao for their great supports.

Contents

List of Figures	vii
List of Tables	viii
Glossary	x
1 Introduction	1
1.1 Overview	1
1.2 Proposed methods and analyses	8
1.2.1 A Bayesian partial likelihood based on ranks	8
1.2.2 A Nonparametric Bayesian bootstrap (BB) method and an application to testing binormality	8
1.2.3 Strong approximation for the BB version ROC process	9
1.2.4 Estimation methods without the gold standard for the truth	10
1.3 Organization	10
2 Bayesian estimation of ROC curve based on ranks	12
2.1 ROC curve setting in case control study	12
2.2 Binormal model and existing methods	14
2.3 Bayesian rank-based partial likelihood estimation method	18
2.3.1 Notation of truncated normal distribution	18
2.3.2 A partial likelihood and the corresponding Bayesian procedure	18
2.4 Consistency of the posterior	19
2.5 Simulation studies	21
2.6 Real data analysis	26
3 Nonparametric estimation of ROC curve	29
3.1 A new nonparametric method based on the Bayesian bootstrap (BB)	30
3.2 Application to testing binormality	34
3.3 Simulation studies	35

3.4	Real data analyses	39
4	Strong approximations for ROC process	42
4.1	Notation	43
4.2	Strong approximations for the BB quantile processes	45
4.3	Limit theorem for BB estimator of ROC curve and its functionals	48
4.3.1	Functional limit theorems for ROC curve	48
4.3.2	Implication of the functional limit theorems for ROC curves	56
4.3.3	Limit theorems for AUC	57
4.4	Simulation studies	59
4.4.1	Comparison of the frequentist and the BB resampled variability	59
4.4.2	Comparison of the empirical and BB's estimates of ROC	59
4.5	Proof of the Lemmas in the sense of <i>a.s.</i>	60
4.5.1	Proof of the Lemma 4.1	60
4.5.2	Proof of the Lemma 4.2	63
5	ROC analysis without the gold standard	67
5.1	Binormal model assumption without the gold standard	70
5.2	Proposed models	71
5.2.1	Bernstein polynomial model	71
5.2.2	Latent model	73
5.3	Simulation studies	75
5.4	Real data analysis	78
6	Discussion and Future work	83
	Bibliography	85
	Appendices	95
A	Theorems used for strong approximation theories	95
B	Theorems used for the proofs	98
B.1	Theorems used for proof the consistency of BPL estimator	98
C	Programs used for simulation studies	99
C.1	Matlab code for the BPL method with the gold standard	99
C.2	Matlab code for the BB method	102
C.3	WinBUGS code for the Bernstein polynomial model	103
C.4	WinBUGS code for the Latent model	104

List of Figures

2.1	Trace plots of intercept (a) and slope (b) for biomarker CA 125.	28
2.2	Density plots of intercept (a) and slope (b) and ROC estimate (c) by BPL method for biomarker CA 125.	28
3.1	Comparison of empirical and the BB's estimates of ROC curve.	31
3.2	Comparison of IAE from different settings.	36
3.3	Boxplots of IAE using BB, BN-G, BN-T and SP methods.	38
3.4	Pointwise 90% credible band of ROC curve using biomarker CA 19-9.	40
4.1	Comparison of density plots obtained by empirical $\sqrt{N}(\hat{A} - A)$ based on 6000 simulated data sets, the BB's $(\sqrt{N}(\hat{A}^\# - \hat{A}) \text{data})$ based on one simulated data set and corresponding 6000 resamples, and the asymptotic normal distribution $N(0, \sigma^2(0, 1))$ where σ can be calculated by Corollary 4.3.	60
4.2	Comparison CDF of empirical (solid curve) and the BB's mid 50% percentile (dotted curves) of $R(0.5)$	61
5.1	CA 19-9: Comparison of ROC curves using the BB and BPL methods (with the gold standard, respectively), the Bernstein polynomial (abbreviated as Poly) and the latent methods (without the gold standard) and BPLS with surrogate labels obtained by biomarker CA 125.	80
5.2	CA 125: Comparison of ROC curves using the BB and BPL methods (with the gold standard, respectively), the Bernstein polynomial (abbreviated as Poly) and the latent methods (without the gold standard) and BPLS with surrogate labels obtained by biomarker CA 19-9.	81

List of Tables

1.1	Diagnostic test outcomes (binary data) classified by the gold standard for the truth.	3
1.2	Diagnostic test outcomes (five-category data) classified by the gold standard for the truth.	3
1.3	Diagnostic test outcomes (binary data) using the threshold value category I.	4
2.1	Estimates of binormal ROC's parameters a and b using methods BPL, MLE, PMLE, GLM and LABROC where we abbreviate it as LAB inside the table. Our BPL estimate is based on 100 simulated data sets, 95000 Gibbs samples with 100000 iterations and burn-in at 5000. The other estimates are based on 1000 simulated data sets.	23
2.2	Sampling properties of BPL and PMLE. Our BPL estimate is based on 100 simulated data sets, 95000 Gibbs samples with 100000 iterations and burn-in at 5000. The PMLE estimate is based on 1000 simulated data sets.	24
2.3	Coverage probabilities of AUC and corresponding average lengths of the 90% CI shown beneath in the parentheses obtained by BPL, BN-G and BN-T methods. Our BPL estimate is based on 100 simulated data sets and corresponding 100000 iterations burn-in at 5000, BN-G and BN-T's estimates are based on 1000 simulated data sets and corresponding 1000 resamples. Simulated data sets are generated by lognormal, location-scale exponential distributions (abbreviated as A, B , respectively) with different combinations of the parameters (A: \mathbf{X} and \mathbf{Y} dataset are generated from the lognormal with corresponding normal parameters (u_x, σ_x) and (u_y, σ_y) , respectively; B: \mathbf{X} and \mathbf{Y} are generated from the exponential distribution with rate=0.5 and the location and scale parameters (u_x, σ_x) and (u_y, σ_y) , respectively). The grid points on $[0,1]$ are chosen with equal interval length 0.05.	25
2.4	Real data analysis CA-125: comparison of estimates (standard errors) of binormal parameters obtained by our BPL and other semiparametric methods.	27

3.1	Coverage probabilities of AUC and corresponding average lengths of the 90% CI shown beneath in the parentheses using BB, BN-G, BN-T and SP methods.	38
3.2	Coverage probabilities and corresponding average lengths of 95% (shown beneath in the parentheses) CI for AUC obtained by BB, EL,MW, LT, PB and PTB methods. Our simulation results are based on 10000 simulated data sets and corresponding 1000 resamples. The grid points on [0,1] are chosen with equal interval length 0.005, $(m, n) = (50, 50)$	40
4.1	Notation for empirical and quantile processes	45
4.2	Values of $\mathbb{F}_n^{\#-1}(y)$ and $\tilde{\mathbb{U}}_{n-1}(y)$	47
5.1	Estimates of binormal ROC's parameters a and b without the gold standard. Simulations are based on the Bernstein polynomial and the latent model abbreviated as Poly and Lat in the table, respectively. All the estimates are based on 100 simulated data sets ($X \sim N(0,1)$, $Y \sim N(5^{1/2}\Phi^{-1}(AUC), 2^2)$, total sample size is equal to 50), and corresponding 3000 Gibbs samples (iteration 6000 and burn-in at 3000) for the latent model and 95000 Gibbs samples (iteration 100000 and burn-in at 5000) for both the Bernstein polynomial model and the BPLS method.	77
5.2	Coverage probabilities of AUC, related mean lengths of the 90% CI and IAE. The same simulation setting in Table 5.1 is used in this table. The grid points on [0,1] are chosen with equal interval length 0.01.	78

Glossary

$X \sim F$ –diagnostic variable for non-diseased population; $Y \sim G$ –diagnostic variable for diseased population;

(m, n) – training sample sizes from X and Y , respectively;

$\Phi(\cdot)$ and ϕ – cumulative distribution and density function of the standard normal distribution, respectively;

ϕ_{μ, σ_m} – the density of $\text{Normal}(\mu, \sigma_m^2)$ (hereafter denoted as $N(\mu, \sigma_m^2)$);

$\bar{\Phi}(t)$ –survival function of standard normal variable evaluated at t ;

$F_m(x)$ –empirical CDF function given X_1, \dots, X_m ;

$\bar{F}(x)$ and $\bar{G}(x)$ – survival functions of X and Y , respectively. That is, $\bar{F}(x) = P(X > x)$ and $\bar{G}(x) = P(Y > x)$;

$*$ – convolution operation symbol, for example $\Phi_{\sigma_m} * F_m(x) = \frac{1}{m} \sum_{i=1}^m \Phi\left(\frac{x-X_i}{\sigma_m}\right)$;

$H(\cdot)$ – strictly increasing transformation on the diagnostic variables;

c_t –threshold value, that is, positive diagnostic result is determined if the diagnostic test value is greater than c_t ;

$(X > c_t)$ –false positive event; $(Y > c_t)$ – true positive event;

FPF= $P(X > c_t)$ – false positive fraction, so-called one minus specificity;

TPF= $P(Y > c_t)$ – true positive fraction, so-called sensitivity;

PPV= $P(D=1|T=1)$ – positive predictive value, where $D=1$ and $T=1$ stand for that the patient is from the diseased group and the test result is positive, respectively;

NPV= $P(D=0|T=0)$ – negative predictive value, where $D=0$ and $T=0$ stand for that the patient is from the non-diseased group and the test result is negative, respectively;

$x \vee y = \max(x, y)$; $x \wedge y = \min(x, y)$;

i.i.d.–independently and identically distributed;

$\mathcal{D} \in (0, 1)$ — a prespecified set of FPF of interest.

Chapter 1

Introduction

1.1 Overview

Receiver operating characteristic (ROC) curve is rooted in statistical decision theory [Wald, 1950]. It is explicitly formulated in the context of electronic signal detection theory developed by Peterson, Birdsall and Fox (1954) at the University of Michigan, where the primary goal is to measure and model the discrimination process of signal detection. In medical contexts, Lusted (1960) first introduces the ROC curve in Roentgen diagnosis. The ROC curve methodology has become an indispensable tool in medical science for the past 30 years, especially in medical imaging, to quantify the accuracy of medical diagnostic tests [see the reviews by Swet and Pickett, 1982 and Hanley, 1989]. Its application has been widely used in many other disciplines such as psychology, quality control, economics and weather forecasting. In this thesis, we use the terminology in medical diagnoses to discuss the ROC curve. We assume that the diagnostic variables are continuous. Such an assumption seems more reasonable with the rapid growth of measuring capabilities of sophisticated diagnostic tools, especially in radiology reading studies. By convention,

larger diagnostic test values imply a greater chance to be diagnosed as positive.

The ROC curve is defined as follows. The diagnostic variables $X \sim F$ for the non-diseased group and $Y \sim G$ for the diseased group are well defined, because the true disease status of each patient is known beforehand. This is called the case with the gold standard for the truth. Positive or negative diagnostic results can be determined by comparing the diagnostic test values with the decision threshold value c_t ; i.e., if the diagnostic test value is greater than (less than or equal to) c_t , this diagnostic test will be determined as positive (negative). Two of the four combinations of true disease status and diagnostic results are called false positive (FP) and true positive (TP) status and are directly used in ROC formulation. The ROC curve can be plotted by pairs $(P(X > c_t), P(Y > c_t))$, $c_t \in \mathbb{R}$; that is, a plot of true positive fraction (TPF) versus false positive fraction (FPF) at different decision threshold values. In the terminology of medical diagnosis, TPF and FPF are also called sensitivity and one minus specificity, respectively. In statistical testing hypotheses terminology, ROC curve is a plot of type I error on the x -axis and the corresponding power on the y -axis by varying the decision threshold values. Thus ROC curve contains all possible decision operating points.

It is interesting to quote Swets (1973): “Psychologists have sought for more than a century to devise measurement procedures that minimize the extent of bias, and, indeed, one procedure more than a century old is largely successful in this respect. This procedure is to present two alternatives at a time, with the assurance that one is A and the other is B, and to ask which is which. ” ROC can play this role.

The construction of ROC curve can be illustrated by the following examples:

If the diagnostic test variable is binary, the diagnostic test results can be shown in a 2×2 contingency table (see Table 1.1). Hence, we can construct one ROC point as $(\frac{m_D}{m}, \frac{n_D}{n})$.

Table 1.1: Diagnostic test outcomes (binary data) classified by the gold standard for the truth.

Truth	Number of negative	Number of positive	Total
non-diseased	$m_{\bar{D}}$	m_D	m
Diseased	$n_{\bar{D}}$	n_D	n
Total	$m_{\bar{D}} + n_{\bar{D}}$	$m_D + n_D$	N

Table 1.2: Diagnostic test outcomes (five-category data) classified by the gold standard for the truth.

Truth	I	II	III	IV	V	Total
non-diseased	m_1	m_2	m_3	m_4	m_5	m
Diseased	n_1	n_2	n_3	n_4	n_5	n
Total	$m_1 + n_1$	$m_2 + n_2$	$m_3 + n_3$	$m_4 + n_4$	$m_5 + n_5$	$N = m + n$

If the diagnostic test variable is ordinal, for example, say it has five categories of disease severity denoted as I, II, III, IV and V meaning very likely normal, probably normal, possibly abnormal, probably abnormal and very likely abnormal, respectively, the diagnostic test results can be listed in a 2×5 contingency table; (see Table 1.2, where each entry shows the number of patients within that category.)

From Table 1.2, based on one decision threshold value, a 2×2 contingency table similar to Table 1.1 can be constructed. For example, by choosing the decision threshold value as category I, we can get the diagnostic test outcomes in Table 1.3. Therefore, we can obtain all ROC points as $(\frac{m_5}{m}, \frac{n_5}{n})$, $(\frac{m_4+m_5}{m}, \frac{n_4+n_5}{n})$, $(\frac{m_3+m_4+m_5}{m}, \frac{n_3+n_4+n_5}{n})$, $(\frac{m_2+m_3+m_4+m_5}{m}, \frac{n_2+n_3+n_4+n_5}{n})$ for decision threshold values IV, III, II and I, respectively.

When the diagnostic test variables are continuous, we can form an infinite number of ROC points which can be used to construct a continuous ROC curve.

The utility of the ROC curve lies in its ability to evaluate the accuracy of different

Table 1.3: Diagnostic test outcomes (binary data) using the threshold value category I.

Truth	Number of negative	Number of positive	Total
non-diseased	m_1	$m - m_1$	m
Diseased	n_1	$n - n_1$	n
Total	$m_1 + n_1$	$N - m_1 - n_1$	N

diagnostic modalities or readers within one curve measurement and in turn to provide meaningful threshold values of different diagnostic modalities or readers using the same critical ROC point based on some criterion (such as fixed TPF, fixed FPF, or proportion of FPF over TPF). A particular benefit of the method is that it gives a performance assessment technique for the technology that is independent of practice-of-medicine issues. It is also invariant under strictly increasing transformations of the diagnostic variables, emphasizing the relationships between the distributions of the diagnostic variables rather than the distributions themselves. Moreover, the most commonly used summary index of the ROC curve, called the area under the curve (AUC), can be interpreted as the probability of Y greater than X [Bamber, 1975] for continuous tests, which can be easily understandable to a wider audience. Other ROC functionals of interest such as the partial AUC (pAUC) also have similar intuitive explanations [cf. Zhou *et al.* 2002]. These features make the ROC analysis extremely popular in diagnostics research.

The main complexity of ROC analysis arises from the diversity of the diagnostic variable itself (e.g., it may be binary, ordinal or continuous), the inclusion covariate information, the diagnostic measurements without the gold standard for the truth, and correlated and clustered data sets. An ROC analysis involves many statistical aspects, such as design of ROC, curve fitting, indices of ROC accuracy, testing hypotheses, etc. In this thesis, the main focus is on ROC curve fitting with continuous diagnostic variables

in standard case-control studies (CCS) with or without the gold standard for the truth.

The literature for ROC analysis is extensive. In statistical decision theory, it is shown that the likelihood ratio can be used to quantify different decision rules. This forms the foundation of signal detection theory. Green and Swets (1966) point out three important issues from the perspective of signal detection tasks. First, any slope of the ROC curve is equal to the corresponding likelihood-ratio evaluated at that pair of (FPF, TPF). Second, a successive decision rule implies the underlying assumption that the larger values of the likelihood ratio indicate that there is greater chance to be able to detect the signal for the observers. Hence, the slope is a strictly decreasing function of the false positive fractions. Third, any point on the ROC curve is related to an accessible decision rule.

Comprehensive reviews on ROC methodology are given by Swets and Pickett (1982), Hanley (1989), Begg (1991), Shapiro (1999), Zhou *et al.* (2002) and Pepe (2003). Design based studies include prospective and retrospective designs, corresponding biases, and sample size considerations (see examples by Begg (1991), Pepe (2003) and Obuchowski (1998)).

There are three approaches for estimating ROC curves. First, pure parametric approaches [cf. Hanley, 1989] are based on the assumption of parametric distributions of the diagnostic variables.

Second, semiparametric methods for ROC analysis have particularly been popular since the presence of nonparametric components makes these models considerably flexible, yet they can incorporate specific parametric features. Under the semiparametric framework, there are several different approaches. The simplest one is the binormal [Green and Swets, 1966] model, which assumes the binormality of the diagnostic test variables after some strictly increasing transformation. The feasibility of the binormality assumption are well discussed by Hanley (1988, 1989, 1996), Swets (1986a) and Metz *et*

al. (1998). The intercept and slope in the binormal model can be estimated by several methods, such as those given by Hsieh and Turnbull (1996), Metz *et al.* (1998), Zou and Hall (2000), Pepe (2000) and (2003), Cai and Moskowitz (2004), among others. However, see Goddard and Hinberg (1990) for some examples where binormality fails. Similarly, bi-gamma [Dorfman *et al.*, 1997] and bi-beta [Zou *et al.*, 2004] models have also been considered. Li *et al.* (1999) propose a model where F and G are nonparametrically and parametrically specified, respectively. Qin and Zhang (2003) model the functional form of the likelihood ratio to estimate the parameters. A normal mixture model is studied by Hall and Zhou (2003).

Third, completely nonparametric methods, such as kernel estimates of F and G , are discussed by Zou *et al.* (1997), Lloyd (1998), Zhou and Harezlak (2002), among others. The empirical estimate of ROC is studied by Hsieh and Turnbull (1996), along with its asymptotic property. Li *et al.* (1996) obtain the weak convergence theory for the ROC estimator under censoring by plugging in the product-limit estimators.

Summary indices of ROC curves include positive predictive value (PPV) and negative predictive value (NPV) used to correct the verification bias, AUC, Partial AUC (pAUC), Youden's Index (defined as the maximum value of the sum of the sensitivity and specificity [cf. Hsieh, 1991]), Kolmogorov-Smirnov measure of distance between two curves [Campbell, 1994], and the integrated absolute error (IAE) [Moise *et al.*, 1988] (see review by Shapiro (1999)).

Two methods deserve particular attention. First, DeLong *et al.* (1988) use a U-statistic to compare the AUC for correlated data, which can be substituted by jackknife method [Hanley and Hajian-Tilaki, 1997]. Second, Qin and Zhou (2006) propose an empirical likelihood inference for AUC which is more efficient compared to other non-parametric methods.

Curve estimation methods incorporating CCS covariate effects can be classified by directly modeling the distributions of the diagnostic variables or directly modelling the ROC functional form; (see work done by Tosteson and Begg (1988), Toledano and Gatsonis (1995), Le (1997), Pepe (1998), Ishwaran and Gatsonis (2000), Cai and Moskowitz (2004).)

The definition of the ROC curve implies that its underlying assumption is the gold standard for the truth, which means that the true disease status of each patient is known by the most accurate (gold standard) diagnostic test. However, in clinical practice, the gold standard test is not always available for various reasons. Verification bias occurs in the situation when we estimate the sensitivity and specificity of the diagnostic test based on those patients with verified disease status. Imperfect gold standard test, which is the best available reference test, is used when the gold standard is not available. Most of the existing statistical methods deal with the binary diagnostic variables and assume the conditional independence of the diagnostic tests. A few methods deal with the ordinal or continuous diagnostic tests to obtain the ROC curve estimate; (see work done by Begg and Greenes (1983), Zhou (1993, 1994), Reilly and Pepe (1995), Clayton *et al.* (1998), Gray *et al.* (1984), Hunink *et al.* (1993), Rodenberg and Zhou (2000), Alonzo (2000) for correcting the verification bias and for the reference gold standard tests; see the reviews by Walter and Irwig (1988), Hui and Zhou (1998); also see work done by Gart and Buck (1966), Mantel (1951), Staquet *et al.* (1981), Vacek (1985), Joseph *et al.* (1995), Pepe (2003).) Hall and Zhou (2003) model the continuous diagnostic variables under the conditional independence assumption without the gold standard as a mixture of two distributions. Zhou *et al.* (2005) propose a nonparametric maximum likelihood method to estimate the ROC curve when the diagnostic tests are ordinal and there are more than two tests. When only one continuous test variable is available without the

gold standard and without covariate information, little discussion about estimating the ROC curve can be found in the literature.

1.2 Proposed methods and analyses

1.2.1 A Bayesian partial likelihood based on ranks

We propose a Bayesian curve estimation method under binormality using a partial likelihood based on ranks. The motivation for this new method lies in the preference to treat the transformation as latent and to construct a Bayes estimator maximum invariant based on the ranks. A similar situation arises in the Gaussian copula model [Hoff, 2006]. We also establish the posterior consistency conditional on the ranks of the samples in an almost everywhere (a.e.) sense by applying Doob's Theorem [cf. Ghosal and Van der Vaart, 2008].

1.2.2 A Nonparametric Bayesian bootstrap (BB) method and an application to testing binormality

Motivated by the empirical counterpart of the ROC curve, we propose completely nonparametric modeling for estimating and building credible intervals for the ROC curves that is based on the Bayesian bootstrap [Rubin, 1981] resampling distributions. The BB method leads to smooth estimates without requiring the choice of a bandwidth. Our simulations show that these bands and intervals have approximate frequentist validity even for very small sample sizes. In comparison with other existing curve estimation methods, our BB method performs well in terms of accuracy, robustness, simplicity and smoothness. We also propose a procedure to test the binormality assumption as an

application of our BB method.

1.2.3 Strong approximation for the BB version ROC process

In order to study the asymptotic behavior of the BB version of the ROC process, we prove the strong approximation for the quantile processes based on the BB resampling distributions. Combining these results with some earlier results from strong approximation theory, we develop strong approximations for the empirical, bootstrap and BB version of the ROC process, its AUC and other functionals.

The existing strong approximation theory primarily includes these findings:

- Komlós, Major and Tusnády (1975) show that the uniform empirical process can be strongly approximated by a sequence of Brownian bridges obtained from a single Kiefer process;
- Csörgő and Révész (1978) prove that under suitable conditions a quantile process can be strongly approximated by a Kiefer process;
- Lo (1987) studies the strong approximation theory for the cumulative distribution function (c.d.f.) of bootstrap and BB processes.
- Hsieh and Turnbull (1996) examine the strong approximation for the empirical ROC estimate on a fixed subinterval of $[0, 1]$.

Interestingly, it will be seen that the forms of these Gaussian approximations are identical, and therefore the distribution of the ROC function, conditioned on the samples, is identical to the Gaussian approximation of the empirical ROC estimate. Also, the result implies that for functionals like AUC, the empirical estimator is asymptotically

normal and asymptotically equivalent to the BB estimator, and that the corresponding confidence intervals have asymptotic frequentist validity.

1.2.4 Estimation methods without the gold standard for the truth

The underlying assumption of the ROC curve is the existence of a gold standard for the truth, which means that the true disease status of each patient is known by some other independent accurate diagnostic test. However, in complicated medical practice, a gold standard is not always available. In this thesis, we model the diagnostic variable S without the gold standard transformed by H as a mixture of two normal distributions, where one normal component stands exactly for the non-diseased population and the other for the diseased population. The ROC curve can be defined in the usual way and an explicit expression is available because of the normality of the components. In order to estimate the parameters in the resulting mixture normal model, direct (Bernstein polynomial model) and indirect (Latent model) approaches are proposed. Monte Carlo Markov Chain sampling techniques, especially coupled with WinBUGS software, make the computation feasible.

1.3 Organization

This thesis is arranged as follows: Bayesian rank-based partial likelihood estimation method is introduced in Chapter 2. A new nonparametric method based on the Bayesian bootstrap (BB) and its application to testing binormality are presented in Chapter 3. The asymptotic behavior of the BB estimator is examined in Chapter 4. ROC curve

estimation methods without the gold standard for the truth are explored in Chapter 5. Finally, a discussion and future work are included in Chapter 6.

Chapter 2

Bayesian estimation of ROC curve based on ranks

2.1 ROC curve setting in case control study

Let $X \sim F$ and $Y \sim G$ be two independent continuous variables; for instance, two diagnostic variables coming from two populations, one without disease and one with disease, respectively. By varying the decision threshold value c_t (if $X > c_t$ or $Y > c_t$, false or true positive event occurs) and plotting the true positive fraction (sensitivity) versus the false positive fraction (one minus specificity), the ROC curve is obtained. It is given by $\{(P(X > c_t), P(Y > c_t))\} = \{(t, R(t))\}$, where $c_t \in \mathbb{R}$, and $t = P(X > c_t)$ is called the false positive fraction. Mathematically, we can write the functional form of ROC curve [Pepe, 2003, page 106] as follows:

$$R(t) = P(Y > c_t) = P(Y > \bar{F}^{-1}(t)) = \bar{G}(\bar{F}^{-1}(t)) = P(\bar{F}(Y) \leq t), \quad (2.1)$$

where $\bar{F}(u) = P(X > u)$ and $\bar{G}(u) = P(Y > u)$ are survival functions of X and Y , respectively. A commonly used index to compare the accuracy of the modalities is the area under the curve (AUC) A and its estimate \hat{A} , defined as

$$A = \int_0^1 R(t)dt \quad \text{and} \quad \hat{A} = \int_0^1 \hat{R}(t)dt, \quad (2.2)$$

where $\hat{R}(t)$ is some estimate of $R(t)$. The AUC can be interpreted as $P(Y > X)$, because $A = \int_0^1 R(t)dt = \int_0^1 \bar{G}(\bar{F}^{-1}(t))dt = \int_{-\infty}^{-\infty} \bar{G}(y)dF(y) = \int_{-\infty}^{\infty} P(Y > y)f(y)dy = P(Y > X)$. This interpretation makes the ROC analysis extremely popular in diagnostics research.

The accuracy of estimation for the entire ROC curve can be measured by the integrated absolute error (IAE) [Moise *et al.*, 1988]: $IAE = \int_0^1 |\hat{R}(t) - R(t)|dt$. Clearly $|\hat{A} - A| \leq IAE$. To construct a uniform credible band for ROC, it is advantageous to map the domain to the real line via a transformation ψ , such as the logistic transformation $\psi(x) = \log(x/(1-x))$, $x \in (0, 1)$. The maximum possible estimation error in the ψ -scale is $\epsilon(\psi, R, \hat{R}) = \sup\{|\psi(\hat{R}(t)) - \psi(R(t))| : t \in (0, 1)\}$. The width of a uniform $100(1-\alpha)\%$ credible band for the transformed ROC, denoted by $d_\alpha(\psi, R)$, is given by:

$$d_\alpha = d_\alpha(\psi, R) = 100(1-\alpha)\% \text{ percentile of the distribution of } \epsilon(\psi, R, \hat{R}). \quad (2.3)$$

Thus the uniform $100(1-\alpha)\%$ credible band for the ROC curve can be constructed by

$$\psi^{-1}(\psi(\hat{R}(t)) - d_\alpha) \leq R(t) \leq \psi^{-1}(\psi(\hat{R}(t)) + d_\alpha). \quad (2.4)$$

The transformation-retransformation automatically ensures that the credible band lies within the unit square. In practice, d_α has to be estimated, usually by some resampling technique. If the uniform credible band for ROC on $t \in (0, 1)$ is too wide, other alternatives such as pointwise $100(1-\alpha)\%$ credible band or uniform $100(1-\alpha)\%$ (or less) credible band restricted on a small subinterval of interest should be considered instead.

2.2 Binormal model and existing methods

The binormal model is one of the most popular models in ROC study. The binormal model assumes normality after some strictly increasing transformation $H(\cdot)$ of the diagnostic variables. Specifically, the binormal model assumes that $H(X) \sim \text{Normal}(\mu_{\bar{D}}, \sigma_{\bar{D}}^2)$ and $H(Y) \sim \text{Normal}(\mu_D, \sigma_D^2)$. Then, the binormal ROC curve is given by

$$R(t) = \Phi(a + b\Phi^{-1}(t)), \quad (2.5)$$

$$a = (\mu_D - \mu_{\bar{D}})/\sigma_D, \quad (2.6)$$

$$b = \sigma_{\bar{D}}/\sigma_D.$$

The corresponding AUC has a closed form expression given by $A = \Phi(a/\sqrt{1+b^2})$, due to $H(Y) - H(X) \sim \text{Normal}(\mu_D - \mu_{\bar{D}}, \sigma_D^2 + \sigma_{\bar{D}}^2)$, $A = P(Y > X) = P(H(Y) > H(X))$ and the definitions of a and b (see (2.6)).

Some interesting implications of the binormality assumption are explored by several authors, which will be helpful in understanding what is meant by binormality assumption.

1. Green and Swets (1966), Metz *et al.* (1998) display the binormal model of ROC visually in the following way:

$$TPF(c_t) = \Phi(a + b\Phi^{-1}(FPF(c_t))), \quad c_t \in \mathbb{R}. \quad (2.7)$$

From (2.7), we can see that the binormality assumption also implies a perfect linear relationship between TPF and FPF on “normal-deviate axes”; that is,

$$\Phi^{-1}(TPF(c_t)) = a + b\Phi^{-1}(FPF(c_t)), \quad c_t \in \mathbb{R}.$$

2. Lloyd (1998) considers a special case to interpret the meaning of the parameters a and b in the binormal model. If $F = N(0, \sigma^2)$, $G = N(\delta, \sigma^2)$, then $a = \delta/\sigma$ and

$b = 1$. Notice that the intercept parameter a is also equal to the square root of the Mahalanobis distance between F and G . Hence, we can see that the parameters a and b quantify the standardized separation and the ratio of standard deviations of the two random variables.

3. Cai and Moskowitz (2004) point out that binormality implies that the ratio of the density function for the diseased group over that for the non-diseased group evaluated at y is equal to the ratio of the normal density with mean μ_D and standard deviation σ_D over that of the standard normal evaluated at $-H(y)$; that is,

$$\frac{f_D(y)}{f_{\bar{D}}(y)} = \frac{b\phi(a - bH(y))}{\phi(-H(y))}. \quad (2.8)$$

To estimate the parameters a and b , the following methods have been considered in the literature.

1. **LABROC**: Metz *et al.* (1998) propose a maximum likelihood (ML) algorithm LABROC4 based on the key observation that ordering the given combined samples may classify the data into several categories using the sequence of truth-state runs. The likelihood can be shown as

$$L(k, l | a, b, t) = \prod_{i=1}^I (P_{i|\bar{D}})^{k_i} \prod_{j=1}^I (P_{i|D})^{l_j},$$

where k_i is the number of observations from the non-diseased group dropped in i^{th} category, l_j is the number of observations from the diseased group dropped in j^{th} category, $P_{i|\bar{D}}$ is the probability of one observation from the non-diseased group dropped in i^{th} category, $P_{i|D}$ is the probability of one observation from the diseased group dropped in i^{th} category. Also $t = (t_0, t_1, \dots, t_I)$, where $[t_{i-1}, t_i)$ is the range of i^{th} category, $t_0 = -\infty, t_I = +\infty$, and t_i is the latent fixed boundary

value generated by truth-state runs. However, the accuracy of the estimates will depend on the initial guess for the parameter values. Moreover, increasing sample size may also inflate the number of parameters, which can be solved by LABROC5 [Metz *et al.*, 1998] algorithm, but will be an ad hoc solution. Compared to our rank-based partial likelihood estimation method of binormal model proposed later, the similarity is that we do Gibbs sampling based on the same truth-state runs, which is invariant up to strictly increasing transformation. The difference is that we treat the latent boundary values as nuisance parameters by defining a constraint set based on the observed combined ranks.

2. **Box-Cox Transformation method (BN-T):** Zou and Hall (2000) use the Box-Cox transformation to estimate the strictly increasing transformation $H(\cdot)$. The estimating procedure includes: (1) Transforming the observed values of the diagnostic variables to positive quantities by changing location and exponentiation; (2) Applying the Box-Cox transformation ($x(\lambda) = (x^\lambda - 1)/\lambda$ if $\lambda > 0$, and $\log x$ if $\lambda = 0$) to obtain transformed observations $X'_i(\lambda)$ and $Y'_j(\lambda)$, where λ is the optimum power transformation parameter in the Box-Cox transformation. (3) Assuming $X'_i(\lambda) \sim \text{i.i.d. } N(\alpha, \beta^2)$ and $Y'_j(\lambda) \sim \text{i.i.d. } N(v, \tau^2)$, get MLEs of λ , α , β , v and τ ; (4) The ROC curve estimated by Box-Cox transformation method is given by $\hat{R}(t) = \Phi(\frac{\hat{v}-\hat{\alpha}}{\hat{\tau}} + \frac{\hat{\beta}}{\hat{\tau}}\Phi^{-1}(t))$. In the same paper, they also directly construct a rank-based likelihood $L(\mu, \sigma | \text{ranks of } Y_1, \dots, Y_n \text{ among the data})$ and obtain ML estimates of μ and σ approximated by using a Monte Carlo procedure.

3. **Generalized linear model (GLM) method (BN-G):** Using the relationship

$$E(1(Y > X) | \bar{F}(X) = t) = P(Y > X | X = \bar{F}^{-1}(t)) = \bar{G}(\bar{F}^{-1}(t)) = R(t),$$

Pepe (2000) suggests a GLM method for estimating the ROC curve. By conditioning on $t_i = \bar{F}(X_i)$, we have $E(U_{ji}) = R(t_i)$, where $U_{ji} = 1(Y_j > X_i)$. Thus, the following probit regression model

$$\Phi^{-1}(U_{ji}) = a + b \Phi^{-1}(\hat{t}_i), \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

can be used for estimating a and b , where $\hat{t}_i = \hat{F}_m(x_i)$, $i = 1, \dots, m$, are obtained by plugging in the empirical survival function of X . However, the U_{ji} 's are dependent and the \hat{t}_i 's are random; hence, the applicability of the GLM estimation method is not fully justified in this context. Metz et al. (1998), Qin and Zhou (2006) also point out this issue in their settings.

Alonzo and Pepe (2002) propose another GLM procedure abbreviated by GLM, defining the quantity $U_{jt} = 1(Y_j > \bar{F}^{-1}(t))$ instead of the BN-G method's U_{ji} , where $t \in \mathcal{D}$. This method is computationally simpler than the BN-G method.

4. Maximum profile likelihood and pseudo maximum likelihood estimator.

Based on observation of (2.8), Cai and Moskowitz (2004) propose a maximum profile likelihood estimation method. However, the citation of the result on profile maximum likelihood [Murphy and van der Vaart, 2000] is questionable, since the estimates of the nuisance parameters used to plug-in are based on an estimating equation, not obtained from maximizing the likelihood. In the same paper, they also propose a pseudo maximum likelihood estimator (PMLE) based on the imputation of the transformation function H based on the data from two groups of subjects. From this viewpoint, PMLE is more efficient than the estimator given by Alonzo and Pepe (2002).

2.3 Bayesian rank-based partial likelihood estimation method

2.3.1 Notation of truncated normal distribution

Two-sided-truncated normal with non-truncated normal mean μ , variance σ^2 and corresponding truncation region (e_1, e_2) is denoted by $\text{TN}(\mu, \sigma^2, (e_1, e_2))$; its density can be expressed as $f_{\text{TN}}(x) = \phi(x) / \int_{e_1}^{e_2} \phi(x) dx$, when $e_1 < x < e_2$ and $f_{\text{TN}}(x) = 0$, otherwise.

2.3.2 A partial likelihood and the corresponding Bayesian procedure

A Binormal model assumes that $X_1, \dots, X_m \sim \text{i.i.d. } F$, $Y_1, \dots, Y_n \sim \text{i.i.d. } G$, and \mathbf{X} and \mathbf{Y} are continuous and independent; without loss of generality, there exists a strictly increasing transformation H such that

$$Z_i = H(X_i) \sim \text{i.i.d. } \text{N}(0, 1), W_j = H(Y_j) \sim \text{i.i.d. } \text{N}(\mu, \sigma^2), \quad (2.9)$$

$\mathbf{S} = S_N = (\mathbf{X}, \mathbf{Y})$, $\mathbf{X} = (X_1, \dots, X_m)$, similar notations for $\mathbf{Y}, \mathbf{Z}, \mathbf{W}$,

combined ranks $\mathbf{R}_N = R(S_N) = (R_{1N}, \dots, R_{NN})$, $N = m + n$,

labels $\mathbf{L} = L(\mathbf{S}) = (L_{1N}, \dots, L_{NN})$, where $L_{iN} = 0$ (respectively, 1) if S_i from \mathbf{X} (respectively, \mathbf{Y}).

Under the binormality assumption, the ranks and labels of (\mathbf{Z}, \mathbf{W}) preserve those of \mathbf{S} . Therefore we can define a set \mathcal{D}_{obs} invariant under H as follows:

$$\begin{aligned} \mathcal{D}_{\text{obs}} &= \{(\mathbf{z}, \mathbf{w}) \in \mathbb{R}^{m+n} : R(\mathbf{z}, \mathbf{w}) = R(\mathbf{S}), L(\mathbf{z}, \mathbf{w}) = L(\mathbf{S})\}, \\ &= \{(\mathbf{z}, \mathbf{w}) \in \mathbb{R}^{m+n} : \underline{Z}_k < z_k < \overline{Z}_k, \underline{W}_l < w_l < \overline{W}_l, L(z, w) = L(\mathbf{S})\}, \end{aligned} \quad (2.10)$$

where $\mathbf{z} = (z_1, \dots, z_m)$, $\mathbf{w} = (w_1, \dots, w_n)$, $\underline{Z}_k = \max_i\{Z_i : X_i < X_k\} \vee \max_j\{W_j : Y_j < X_k\}$, $\overline{Z}_k = \min_i\{Z_i : X_k < X_i\} \wedge \min_j\{W_j : X_k < Y_j\}$, $\underline{W}_l = \max_i\{Z_i : X_i < Y_l\} \vee \max_j\{W_j : Y_j < Y_l\}$, $\overline{W}_l = \min_i\{Z_i : Y_l < X_i\} \wedge \min_j\{W_j : Y_l < Y_j\}$, for all $k = 1, \dots, m$, $l = 1, \dots, n$. Hence, a rank-based partial likelihood can be defined based on this invariant set (2.10) as

$$\begin{aligned} & \Pr((\mathbf{z}, \mathbf{w}) \in \mathbb{R}^{m+n} : R(\mathbf{z}, \mathbf{w}) = R(\mathbf{S}), L(\mathbf{z}, \mathbf{w}) = L(\mathbf{S}) | \mu, \sigma) \\ &= \Pr((\mathbf{z}, \mathbf{w}) \in \mathcal{D}_{\text{obs}} | \mu, \sigma). \end{aligned} \tag{2.11}$$

However, evaluation of (2.11) involves high dimensional integrations, and hence direct computation of the maximum partial likelihood estimates will be extremely challenging. An alternative is to adopt the Bayesian approach. The Bayes partial likelihood estimator (BPL) is defined as the posterior mean of (μ, σ) , where the posterior distribution is given by

$$\pi(\mu, \sigma | R(\mathbf{S}), L(\mathbf{S})). \tag{2.12}$$

We shall use a data augmentation technique and implement Gibbs sampling to compute the posterior in (2.12). More specifically, letting (\mathbf{Z}, \mathbf{W}) be the augmentation variables, we shall describe a procedure to sample from the conditional distribution of $((\mu, \sigma) | \mathbf{Z}, \mathbf{W}, R(\mathbf{S}), L(\mathbf{S}))$ and $((\mathbf{Z}, \mathbf{W}) \in \mathcal{D}_{\text{obs}} | \mu, \sigma)$; see Section 2.5 for details.

2.4 Consistency of the posterior

Let (μ_0, σ_0) be the true value of (μ, σ) . It is important to know if the consistency of the posterior at (μ_0, σ_0) holds; i.e., whether or not the posterior distribution is concentrated around (μ_0, σ_0) [Ghosh and Ramamoorthi, 2003]. Assume that the disease prevalence is

asymptotically stable; i.e., when $N \rightarrow \infty$, $\frac{n}{N} \rightarrow \lambda$, where $0 < \lambda < 1$. Each observation X_i or Y_j is randomly sampled from the whole population, and with the gold standard for the truth it can be labeled into either the X or Y group. Also, assume the joint prior density $\pi(\mu, \sigma)$ is dominated by the Lebesgue measure (denoted as $[\nu]$).

Theorem 2.1 Suppose that $\mathbf{S} \sim (1 - \lambda)F + \lambda G$, and assume that binormality (2.9) holds conditional on the labels, and (μ, σ) has joint prior density $\pi(\mu, \sigma) > 0$ a.e. over $\mathbb{R} \times \mathbb{R}^+$ with respect to the Lebesgue measure $[\nu]$. Then for (μ_0, σ_0) a.e. $[\nu]$, and for any neighborhood \mathcal{U}_0 of (μ_0, σ_0) , we have that

$$\lim_{N \rightarrow \infty} \pi((\mu, \sigma) \in \mathcal{U}_0 | R(\mathbf{S}), L(\mathbf{S})) = 1 \quad a.s. \quad [P_{\mu_0, \sigma_0, H}^\infty], \quad (2.13)$$

where $P_{\mu_0, \sigma_0, H}^\infty$ denotes the joint distribution of all \mathbf{X} 's and \mathbf{Y} 's under the binormal model (2.9) with (μ_0, σ_0) as the true value of (μ, σ) .

Proof. Because $S_i \sim$ i.i.d. $(1 - \lambda)F + \lambda G$, then $U_i = ((1 - \lambda)F + \lambda G)(S_i)$ are i.i.d. Uniform(0,1). Hence, we have

$$U_k = \lim_{N \rightarrow \infty} \frac{R_{kN}}{N + 1} \text{ for all fixed } k, \quad (2.14)$$

where R_{kN} is the rank of the k^{th} element among the total N observations. This result is contained in the proof of Theorem *a* on page 157 of Hájek and Šidák (1967).

Notice unconditionally, $U_i = ((1 - \lambda)\Phi + \lambda\Phi_{\mu, \sigma})(N_i)$ are i.i.d. Uniform(0,1), where $N_i = H(S_i)$. Based on the sample size N , we denote a subsequence labeling the observations conditionally from F as (i_1, \dots, i_m) . Hence, conditionally on $L_{i_j N} = 0$, $j = 1, \dots, m$, $N_{i_j} \sim N(0, 1)$ and $U_{i_j} = ((1 - \lambda)\Phi + \lambda\Phi_{\mu, \sigma})(N_{i_j}) \sim$ i.i.d. $g_{\mu, \sigma}$ (say), which belongs to a regular parametric family indexed by parameters (μ, σ) . More specifically, Cramér type regularity conditions can be verified by applying the inverse function theorem. Thus

some consistent estimator for (μ, σ) can be easily obtained, such as the MLE, a Bayes estimator, or the method of moments estimator. Hence, there exists a series of functions h_N and h , such that $(\mu, \sigma) = \lim_{N \rightarrow \infty} h_N(U_{i_1}, \dots, U_{i_m}) = h(U_{i_1}, U_{i_2}, \dots)$, say.

Therefore, there exist functions $\tilde{h}, f_1, f_2, \dots$, conditional on observations from F ,

$$\begin{aligned} (\mu, \sigma) &= h(U_{i_1}, U_{i_2}, \dots) \\ &= h\left(\lim_{N \rightarrow \infty} f_1(\mathbf{R}_N), \lim_{N \rightarrow \infty} f_2(\mathbf{R}_N), \dots\right) \\ &= \tilde{h}(R(S_N) : N \geq 1) \quad a.s. \quad [P_{\mu_0, \sigma_0, H}^\infty], \end{aligned}$$

where $\mathbf{R}_N = R(S_N)$ is defined in (2.9). By Doob's Theorem [Ghosal and Van der Vaart, 2008], the consistency of posterior (2.13) at (μ_0, σ_0) holds a.e. $[\pi]$, and hence at (μ_0, σ_0) a.e. $[\nu]$. ■

Remark 2.1 *In most infinite dimensional applications of Doob's Theorem, the lack of existence of an analog of the Lebesgue measure implies that the exceptional set of points where the posterior may be inconsistent is null when measured with respect to the prior only. This severely dilutes the importance of the conclusion since what is null with respect to a given prior may be quite large when measured with respect to another prior. In the present case, existence of a dominating Lebesgue measure removes this arbitrariness.*

2.5 Simulation studies

We compare the BPL estimator of the binormal model for the ROC curve obtained by (2.12) with other methods, such as maximum profile likelihood estimator (MLE) and pseudo maximum likelihood estimator (PMLE) proposed by Cai and Moskowitz (2004), binary regression approach (GLM) by Alonzo and Pepe (2002), and LABROC by Metz *et*

al. (1998), through the same data generating scheme used in Cai and Moskowitz (2004). We choose the commonly used improper prior $\pi(\mu, \sigma^2) \propto 1/\sigma^2$ and define the same initial values of $\mu = 3$ and $\sigma = 2$ for each simulated data set. Gibbs sampling procedure from the posterior (2.12) is described as follows using the same denotation in (2.9):

1. Generate $(X_1, \dots, X_m) \sim \text{i.i.d. } N(0, 1)$, $(Y_1, \dots, Y_n) \sim \text{i.i.d. } N(\mu, \sigma^2)$, order the combined vector (\mathbf{X}, \mathbf{Y}) denoted as $OS(\mathbf{X}, \mathbf{Y})$ and record the corresponding labels as $L(OS(\mathbf{X}, \mathbf{Y}))$;
2. Generate the initial values (\mathbf{Z}, \mathbf{W}) satisfying $L(OS(\mathbf{Z}, \mathbf{W})) = L(OS(\mathbf{X}, \mathbf{Y}))$;
3. Start the iterations:

(1) Conditional on μ, σ , update $\mathbf{Q} = OS(\mathbf{Z}, \mathbf{W}) = (Q[1], \dots, Q[N])$ with constraint $((\mathbf{Z}, \mathbf{W}) \in \mathcal{D}_{\text{obs}}|\mu, \sigma)$, equivalently, satisfying $L(\mathbf{Q}) = L(OS(\mathbf{X}, \mathbf{Y}))$ through the following sequential truncated normal componentwise simulations ($i = 1, \dots, N$):

$$Q^{(new)}[i] \sim \begin{cases} \text{TN}(0, 1, (Q[i-1], Q[i+1])), & \text{when } L_{iN} = 0; \\ \text{TN}(\mu, \sigma^2, (Q[i-1], Q[i+1])), & \text{when } L_{iN} = 1, \end{cases}$$

where $Q[0] = -\infty$, $Q[N+1] = \infty$;

(2) Update \mathbf{Z} (or \mathbf{W}) values based on $Q^{(new)}$ and $L(OS(\mathbf{X}, \mathbf{Y}))$, i.e., every component of vector \mathbf{Z} belongs to the set

$$\{Q^{(new)}[i] : L_{iN} = 0, i = 1, \dots, N\}.$$

Similarly, every component of vector \mathbf{W} belongs to the set

$$\{Q^{(new)}[i] : L_{iN} = 1, i = 1, \dots, N\};$$

(3) Update μ and σ by their posterior distributions as

$$\begin{aligned}\sigma^2|\text{rest} &\sim \text{inverse gamma } ((n-1)/2, (n-1)s_w^2/2), \\ \mu|\text{rest} &\sim \text{Normal}(\bar{W}_n, \sigma^2/n),\end{aligned}$$

where $s_w^2 = \sum_{j=1}^n (W_j - \bar{W}_n)^2 / (n-1)$, $\bar{W}_n = \sum_{j=1}^n W_j / n$;

4. After burn-in, we can obtain the BPL estimate of the binormal parameters a and b by the corresponding posterior means for the given data \mathbf{X}, \mathbf{Y} . In the mean time, we can calculate $100(1-\alpha)\%$ credible intervals for \hat{a} and \hat{b} as $(q_{a,\alpha/2}, q_{a,1-\alpha/2})$ and $(q_{b,\alpha/2}, q_{b,1-\alpha/2})$, where $q_{a,\alpha/2}$ and $q_{a,1-\alpha/2}$ denote $\alpha/2$ and $1-\alpha/2$ quantiles of Gibbs samples of a , respectively; similar definition of $q_{b,\alpha/2}, q_{b,1-\alpha/2}$ for b .

Table 2.1: Estimates of binormal ROC's parameters a and b using methods BPL, MLE, PMLE, GLM and LABROC where we abbreviate it as LAB inside the table. Our BPL estimate is based on 100 simulated data sets, 95000 Gibbs samples with 100000 iterations and burn-in at 5000. The other estimates are based on 1000 simulated data sets.

m	n	Bias					Mean square error				
		BPL	MLE	PMLE	GLM	LAB	BPL	MLE	PMLE	GLM	LAB
50	a	.071	.079	.064	.034	.032	.051	.063	.064	.059	.059
50	b	-.004	.043	.062	-.009	.002	.020	.022	.022	.022	.022
200	a	-.007	.026	.018	.013	.002	.010	.125	.138	.126	.128
200	b	-.006	.014	.017	-.004	.001	.004	.048	.051	.052	.049

We calculate the bias and the MSE of the estimates using methods BPL, MLE, PMLE, GLM and LABROC. The true values of parameters are chosen to be $a = 1.245$ and $b = 0.667$, equivalently, $\mu = 1.868$, $\sigma = 1.5$. Simulation results in Table 2.1 also includes the information contained in Cai and Moskowitz (2004). For limited simulation results not shown in Table 2.1, when the sample size exceeds one thousand, mean square

errors obtained by BPL tend to be much smaller than those by MLE, PMLE, GLM and LABROC methods.

Table 2.2: Sampling properties of BPL and PMLE. Our BPL estimate is based on 100 simulated data sets, 95000 Gibbs samples with 100000 iterations and burn-in at 5000. The PMLE estimate is based on 1000 simulated data sets.

m	n	Sampling SE		Ave(\hat{SE})		Coverage(95%)	
		BPL	PMLE	BPL	PMLE	BPL	PMLE
50	a	.215	.245	.234	.248	.98	.974
50	b	.141	.135	.143	.193	.97	.968
200	a	.098	.116	.112	.113	.96	.945
200	b	.066	.070	.064	.073	.95	.949

In order to check whether or not the frequentist variability of the estimate can be asymptotically approximated by the sampling variability, we examine the sampling standard error (Sampling SE), the average of the estimated standard error estimator and MCMC standard error obtained by PMLE and BPL respectively, denoted as Ave(\hat{SE}) in Table 2.2, and coverage probabilities of the 95% confidence interval for the estimates a and b . Simulation results in Table 2.2 also contain the information in Cai and Moskowit (2004).

Besides considering the performance of the estimators in the binormal model, we also compare the accuracy of the estimators of ROC functional AUC obtained by our BPL with BN-G [ROC GLM method by Pepe, 2000], BN-T Box-Cox [Zou and Hall, 2000]. Coverage probabilities of AUC and corresponding average lengths of the 90% CI (shown beneath in the parentheses) are displayed in Table 2.3.

Based on the limited simulation results shown in Table 2.1 and 2.2, we conclude that the BPL estimator of (a, b) has a considerably smaller MSE than the estimators given by MLE, PML, GLM, LABROC methods in this simulation setting. When the sample

Table 2.3: Coverage probabilities of AUC and corresponding average lengths of the 90% CI shown beneath in the parentheses obtained by BPL, BN-G and BN-T methods. Our BPL estimate is based on 100 simulated data sets and corresponding 100000 iterations burn-in at 5000, BN-G and BN-T's estimates are based on 1000 simulated data sets and corresponding 1000 resamples. Simulated data sets are generated by lognormal, location-scale exponential distributions (abbreviated as A, B , respectively) with different combinations of the parameters (A: \mathbf{X} and \mathbf{Y} dataset are generated from the lognormal with corresponding normal parameters (u_x, σ_x) and (u_y, σ_y) , respectively; B: \mathbf{X} and \mathbf{Y} are generated from the exponential distribution with rate=0.5 and the location and scale parameters (u_x, σ_x) and (u_y, σ_y) , respectively). The grid points on $[0,1]$ are chosen with equal interval length 0.05.

Data		$m = n = 15$			$m = n = 50$		
u_x, σ_x	u_y, σ_y	BPL	BN-G	BN-T	BPL	BN-G	BN-T
A		.97	.886	.866	0.89	.871	.882
0,1	1, 1	(.249)	(.254)	(.250)	(.141)	(.141)	(.143)
		.94	.859	.801	.86	.890	.873
0,1	3, 3	(.237)	(.233)	(.205)	(.132)	(.134)	(.122)
B		.93	.862	.860	.91	.888	.856
0,1	1, 1	(.258)	(.288)	(.285)	(.158)	(.159)	(.158)
		.92	.857	.910	.91	.772	.925
0,1	3, 3	(.134)	(.097)	(.076)	(.054)	(.062)	(.045)

size increases, although BPL estimator tends to have slightly larger bias, its mean square error is much smaller than the others in Table 2.1. Compared to PMLE, our BPL estimates have smaller sampling variation in most cases and similar coverage probabilities in Table 2.2. Also, because the simulation results of $\text{Ave}(\hat{SE})$, which shows the sampling variability and asymptotic variability using BPL and PMLE methods, respectively, are not far from those of Sampling SE, which reflects the frequentist variability using BPL and PMLE methods, we can use the sampling variability to estimate the frequentist variability. Moreover, compared to PMLE, our BPL estimator performs better when estimating b based on sample size $m = n = 50$. Compared to BN-G and BN-T, our BPL estimator tends to have larger average length of CI for AUC but with much higher coverage probability when the simulated data sets are from location-scale exponential distributions. When the data are from lognormal, which means the binormality assumption holds exactly, the BPL estimator performs better in terms of both accuracy and efficiency.

2.6 Real data analysis

We use the data set published by Wieand *et al.* (1989). This study is based on 51 patients in the control group diagnosed with pancreatitis and 90 patients in the case group diagnosed with pancreatic cancer by two biomarkers: a cancer antigen (CA 125) and a carbohydrate antigen (CA 19-9). The purpose is to decide which marker would better distinguish the case group from the control group. We compare our BPL estimator with MLE and PMLE by Cai and Moskowitz (2004), Zou and Hall (2000), GLM [Alonzo and Pepe, 2002] and LABROC [Metz, *et al.* 1998] using biomarker CA 125 for illustration. Our BPL estimate is based on 95000 Gibbs samples with 100000 iterations and burn-in

at 5000. We use the same priors as described before. Both of the initial values of a and b are chosen to be equal to 2. Table 2.4 also reflects the information contained in Cai and Moskowitz (2004). Convergence of MCMC is examined by trace plots of MCMC samples in Figure 2.1. The density plots of MCMC samples of a , b and our BPL estimate of ROC curve are shown in Figure 2.2, respectively. Compared to the other estimates, our BPL estimate tends to have slightly smaller estimated standard error.

Table 2.4: Real data analysis CA-125: comparison of estimates (standard errors) of binormal parameters obtained by our BPL and other semiparametric methods.

	BPL	MLE	PMLE	Zou and Hall	GLM (2002)	LABROC
a	.748(.188)	.76(.191)	.719(.198)	.727(.190)	.778(.197)	.720(.185)
b	1.024(.139)	1.065(.140)	1.020(.148)	1.007(.142)	1.017(.167)	1.002(.137)

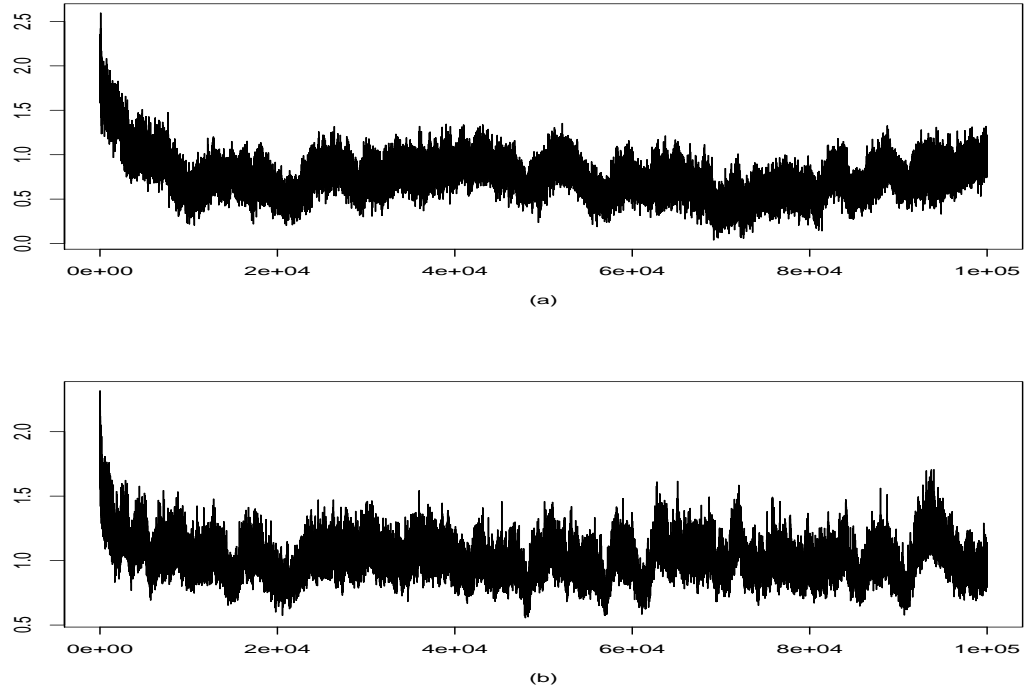


Figure 2.1: Trace plots of intercept (a) and slope (b) for biomarker CA 125.

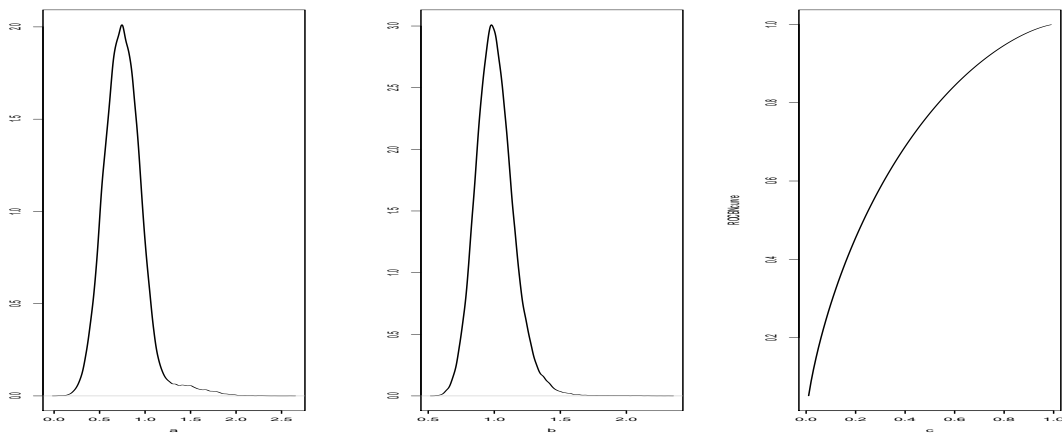


Figure 2.2: Density plots of intercept (a) and slope (b) and ROC estimate (c) by BPL method for biomarker CA 125.

Chapter 3

Nonparametric estimation of ROC curve

As we introduced before, many of the existing methods available in the literature to estimate the ROC curve are based on data generating schemes; here, we only list a semiparametric method used in simulation study.

Semiparametric method (SP): Semiparametric location-scale model [Pepe, 2003] assumes $X = \mu_{\bar{D}} + \sigma_{\bar{D}}\epsilon$ and $Y = \mu_D + \sigma_D\epsilon$ where μ 's and σ 's are the location and scale parameters, respectively, and ϵ has some unspecified survival function S_0 with mean 0 and variance 1. The functional form of the ROC is given by $R(t) = S_0(-a + bS_0^{-1}(t))$. The empirical survival function of the ϵ is then defined as

$$\hat{S}_0(y) = \frac{1}{m+n} \left\{ \sum_{i=1}^m 1 \left(\frac{X_i - \hat{\mu}_{\bar{D}}}{\hat{\sigma}_{\bar{D}}} > y \right) + \sum_{j=1}^n 1 \left(\frac{Y_j - \hat{\mu}_D}{\hat{\sigma}_D} > y \right) \right\},$$

where $\hat{\mu}_{\bar{D}}$, $\hat{\sigma}_{\bar{D}}$, $\hat{\mu}_D$, $\hat{\sigma}_D$ are the sample means and the sample standard errors of X 's and Y 's, respectively. By plugging the empirical survival function \hat{S}_0 and estimates of location scale parameters into the expression for $R(t)$, we get the semiparametric ROC

estimate $\hat{R}(t) = \hat{S}_0(-\hat{a} + \hat{b}\hat{S}_0^{-1}(t))$, where $\hat{a} = (\hat{\mu}_D - \hat{\mu}_{\bar{D}})/\hat{\sigma}_D$, $\hat{b} = \hat{\sigma}_{\bar{D}}/\hat{\sigma}_D$. This method is restricted to a location-scale model only.

Other two nonparametric methods are well known. DeLong *et al.* (1988) use a U-statistics approach to estimate $\theta = \text{AUC}$ by $\hat{\theta} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \gamma(X_i, Y_j)$, where $\gamma(x, y) = 1\{x < y\} + \frac{1}{2}1\{x = y\}$. They also provide consistent estimators of $\text{Var}(\hat{\theta})$ given by $V_y^2/(m(m-1)) + V_x^2/(n(n-1))$, where $V_x(X_i) = \frac{1}{n} \sum_{j=1}^n \gamma(X_i, Y_j) - \hat{\theta}$, $V_y(Y_j) = \frac{1}{m} \sum_{i=1}^m \gamma(X_i, Y_j) - \hat{\theta}$ and $V_x = \sum_{i=1}^m V_x^2(X_i)$, $V_y = \sum_{j=1}^n V_y^2(Y_j)$. This method can be used to estimate AUC only; it does not yield any estimate of the whole ROC curve. Lloyd (1998) proposes a semiparametric kernel estimator of ROC by using kernel estimators of the survival functions F and G , respectively. The choice of the bandwidth is discussed by Zou *et al.* (1997), among others, but the method is still ad hoc in general. Also, in limited simulations, we find this method is not reliable when the data is skewed on the boundary.

3.1 A new nonparametric method based on the Bayesian bootstrap (BB)

The motivation of our BB estimator is twofold as shown below. On the one hand, we can view this as a bandwidth-free smoothing of the empirical estimate. On the other hand, we argue that it is a non-informative limit of a Bayesian estimate based on the Dirichlet process prior.

1. Empirical ROC estimators [Hsieh and Turnbull, 1996] are easily obtained by plugging the empirical counterparts into the ROC functional form. In order to have continuous estimators of the ROC curve, the jumps in the empirical CDF can be

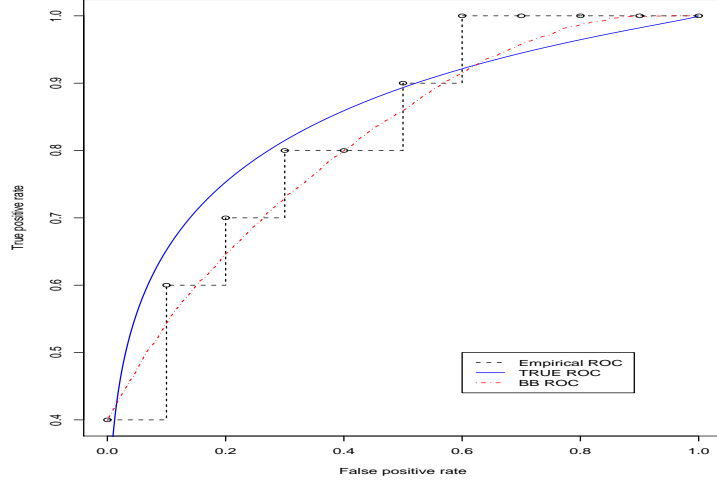


Figure 3.1: Comparison of empirical and the BB's estimates of ROC curve.

interpolated linearly. The bootstrap method [Efron, 1979] can be used to get the error of the curve estimate. However, the inherent discreteness of the estimate is partially due to a finite choice of the weights. The BB method proposed below assigns the Dirichlet distribution to the weights. By forming an ensemble of estimators and averaging, it provides a smoother version of the bootstrap. Figure 3.1 gives an illustration of these differences even when the sample size is small. Data is generated from $X_1, \dots, X_m \sim \text{i.i.d. } N(0, 1)$, $Y_1, \dots, Y_n \sim \text{i.i.d. } N(1.868, 1.5^2)$, where $m=n=10$, with corresponding 5000 resamples. The grid points on $[0,1]$ are chosen with equal interval length 0.001.

2. To implement a Bayesian analysis, a natural choice of priors on F and G is Dirichlet process, denoted as DP, with certain pairs of precision M and center measure ξ , say $F \sim \text{DP}(M_1, \xi_1)$, $G \sim \text{DP}(M_2, \xi_2)$. Conditional on data X_1, \dots, X_m and Y_1, \dots, Y_n , the posterior of $(F|\text{data})$ is $\text{DP}(M_1 + m, \frac{M_1 \xi_1 + m \mathbb{F}_m}{M_1 + m})$, $(R|F, G, \text{data})$ is $\text{DP}(M_2 + n, \frac{M_2 \xi_2 \circ \bar{F}^{-1} + n \mathbb{G}_n \circ \bar{F}^{-1}}{M_2 + n})$. Unfortunately, the above posterior can be obtained

only by the Sethuraman (1994) representation of the Dirichlet process and function inversion. The Sethuraman representation involves generating an infinite collection of random variables which is computationally intensive. We consider the non-informative limit of the Dirichlet processes by letting $M_1 \rightarrow 0$ and $M_2 \rightarrow 0$. This simplifies the procedure to an infinitely easier simulation problem. In fact, we do not even have to specify the center measures ξ_1 and ξ_2 . We need only to generate from the uniform distribution over the simplex, which can be done quite easily; see Remark 3.1. Our BB estimator is just based on this simplification; i.e., the posterior of $(F|\text{data})$ is $\text{DP}(m, \mathbb{F}_m)$, and $(R|F, G, \text{data})$ is $\text{DP}(n, \mathbb{G}_n \circ \bar{F}^{-1})$.

The BB estimator of the ROC curve and its associated summary measures can be computed as follows: Recall $R(t) = \Pr(\bar{F}(Y) \leq t)$, $t \in \mathcal{D} \subset [0, 1]$, where \mathcal{D} denotes a prespecified set of FPF of interest. If we can impute the variable $Z = \bar{F}(Y)$ by plugging in the survival distribution of F based on the BB resampling distribution, then the CDF of Z based on the BB resampling distribution is one realization of the ROC curve from the corresponding posterior distribution.

1. Step 1. (Imputing the placement variables based on the BB resampling distribution.) Let $Z_j = \bar{F}^\#(Y_j) = 1 - F^\#(Y_j)$, $F^\#(Y_j) = \sum_{i=1}^m p_i 1(X_i \leq u)$, $(p_1, \dots, p_m) \sim \text{Dirichlet}(m; 1, \dots, 1)$ independent of others, equivalently to generate $\bar{F}^\# \sim \text{DP}(m, \mathbb{F}_m)$ evaluated at Y_j 's as Z_j (Z_j is also called non-disease placement value [Pepe, 2003, page 105] evaluated at Y_j). The difference between our method and Pepe's is that we choose the survival function using the BB resampling distribution instead of the empirical one.
2. Step 2. (Generating one random realization of the ROC curve.) Generate one realization of $R_{m,n}^\#(t)$, i.e., CDF of Z_1, \dots, Z_n , where $R_{m,n}^\#(t) = \sum_{j=1}^n q_j 1(Z_j \leq t)$,

$(q_1, \dots, q_n) \sim \text{Dirichlet}(n; 1, \dots, 1)$ independent of others, equivalent to generate $R_{m,n}^\#(t) \sim \text{DP}(n, \mathbb{G}_n \circ \bar{F}^{-1})$ evaluated at $t \in \mathcal{D}$. Corresponding random realization of AUC is denoted as $A^\#$, plus some subscript as the index of the realization.

3. Step 3. (Averaging the ensemble of random ROC curves.) The BB estimate, denoted as $\hat{R}_{m,n}^{BB}(t)$, is obtained by averaging the random realizations of the ROC curves, i.e., $\hat{R}_{m,n}^{BB}(t) = \text{mean}(R_{m,n}^\#(t))$, $t \in \mathcal{D}$. Similarly, we obtain the BB estimate of AUC denoted as A^{BB} by plugging $\hat{R}_{m,n}^{BB}(t)$ into (2.2).

Because of the averaging over two levels of random variations, the BB estimate is much smoother than the empirical one. Note that to smooth it out, we do not need a kernel.

Remark 3.1 *A convenient method for generating $(p_1, \dots, p_m) \sim \text{Dirichlet}(m; 1, \dots, 1)$ is to generate $w_1, \dots, w_m \sim \text{i.i.d. exponential distribution with rate 1}$ and put $p_i = w_i / \sum_{j=1}^m w_j$, $i = 1, \dots, m$.*

In order to compute error estimates for the BB estimators of the ROC curve and associated indices, the above steps need to be repeated K times (where K is a reasonably large number). For example, the BB standard error of A^{BB} is given by

$$s = \sqrt{\frac{1}{K-1} \sum_{l=1}^K (A_l^\# - A^{BB})^2}. \quad (3.1)$$

Also a $100(1-\alpha)\%$ BB credible interval for A can be obtained from

$$\text{the percentiles of } \{A_l^\#, l = 1, \dots, N\} \text{ at level } \alpha. \quad (3.2)$$

To obtain a uniform credible band for R based on BB samples, we may estimate d_α by the $100(1-\alpha)\%$ percentile of the sample $\sup\{|\psi(R^{(l)}(t)) - \psi(\hat{R}(t))| : t \in (0, 1), l =$

$1, \dots, N\}$, where $R^{(l)}(t)$ and $\hat{R}(t)$ are l^{th} random realization and BB estimate of $R(t)$, respectively, and substitute \hat{d}_α in (2.4).

Similar ideas may be used to estimate the IAE.

3.2 Application to testing binormality

Because the binormal model is popularly used in practice, it is important to validate model assumptions before using it. Several methods are available, such as one based on the linearity property of TPF and FPF on the “normal-deviate axes”; the graphic method is mentioned by Swets (1986b); Cai and Moskowitz (2004) propose a residual plot using bootstrap sampling method; Dorfman and Alf (1968), Lin and Mudholkar (1980), and Bozdogan and Ramirez (1986), propose a goodness-of-fit test.

Our testing of binormality procedure is motivated as follows. In the binormal model, $H(X) \sim N(0,1)$ and $H(Y) \sim N(\mu, \sigma^2)$, by convention $\mu > 0$, and the continuous strictly transformation is easily identifiable as $H(x) = \Phi^{-1}(F(x))$, where Φ^{-1} and Φ denote the quantile function and CDF of standard normal distribution, respectively. By plugging in the kernel smoothed empirical estimate \tilde{F}_m of F , we can estimate $H(x)$ by $\hat{H}(x) = \Phi^{-1}(\tilde{F}_m(x))$, where $\tilde{F}_m = \Phi_{\sigma_m} * F_m$, “ $*$ ” stands for the convolution operation and σ_m is the bandwidth. See Zou *et al.* (1997), Lloyd (1998), Zhou and Harezlak (2002), among others for a discussion about the choice of the bandwidth. Now μ and σ can be estimated by the sample mean and sample standard error of $\hat{H}(Y_1), \dots, \hat{H}(Y_n)$, which are denoted by $\hat{\mu}$ and $\hat{\sigma}$, respectively. Under the null hypothesis that the binormal model is true, the ROC function is given by $R(t) = \Phi(a + b\Phi^{-1}(t))$, where $a = \mu/\sigma$ and $b = 1/\sigma$.

Thus, we consider the test statistic $T = \sup_t |\hat{R}(t) - \Phi(\hat{a} + \hat{b}\Phi^{-1}(t))|$, where $\hat{R}(t)$ is empirical (or the BB) estimate of $R(t)$, $\hat{a} = \hat{\mu}/\hat{\sigma}$ and $\hat{b} = 1/\hat{\sigma}$. We reject the null

hypothesis of binormality for large values of T ; that is, at level α we reject if $T \geq c_\alpha$. However, it is hard to analytically compute or approximate c_α , so we employ a resampling technique. Here, the Bayesian bootstrap method is used. Because strong approximation ensures under fairly mild restrictions that, asymptotically, the sampling distribution of any test statistic $\gamma_{m,n}(F_m, G_n)$ is equal to the BB resampling distribution of $\gamma_{m,n}(F_m^\#, G_n^\#)$ conditional on the samples X_1, \dots, X_m and Y_1, \dots, Y_n (see definitions of $F^\#(u)$, p_i 's and q_j 's in Section 3.1, where $G^\#(u) = \sum_{j=1}^n q_j 1(Y_j \leq u)$). In this case, we may define $T^\# = \gamma_{m,n}(F_m^\#, G_n^\#) = \sup_t |R_{m,n}^\#(t) - \Phi(a^\# + b^\# \Phi^{-1}(t))|$, where $R_{m,n}^\#(t)$ is defined in Section 3.1, $a^\#$ and $b^\#$ are obtained through $\mu^\#$ and $\sigma^\#$, which are estimated by the sample mean and sample standard error of $H^\#(Y_1), \dots, H^\#(Y_n)$. Here we define $H^\#(x) = \Phi^{-1}(\tilde{F}_m^\#(x))$, where $\tilde{F}_m^\# = \Phi_{\sigma_m} * F_m^\#$. Then c_α can be estimated by $\hat{c}_\alpha = 100(1 - \alpha)\%$ percentile of the BB distribution of $T^\#$ conditional on the given samples. Thus, we reject the null hypothesis of binormality if $T \geq \hat{c}_\alpha$.

3.3 Simulation studies

Before doing our simulation study, we first examine the difference between two survival function estimates as BB and smoothed BB method. Smoothed Bayesian bootstrap survival function is defined as $\bar{F}_1(u) = \sum_{i=1}^m p_i \bar{\Phi}(\frac{u-x_i}{h_m})$; the Bayesian bootstrap survival function is defined as $\bar{F}_2(u) = \sum_{i=1}^m p_i 1(x_i > u)$, $(p_1, \dots, p_m) \sim \text{Dirichlet}(m; 1, \dots, 1)$, $h_m = m^{-\frac{1}{3}}$.

Generate two independent observations X and Y as $X_1, \dots, X_m \sim \text{i.i.d. } F = N(1, 1)$; $Y_1, \dots, Y_n \sim \text{i.i.d. } G = N(1.5, 1)$, $m = n = 50$, simulating 1000 data sets and 1000 corresponding resamples. Comparing the integrated absolute error by using smoothed BB, BB and parametric method [Pepe, 2003], our BB method is at least as good as the

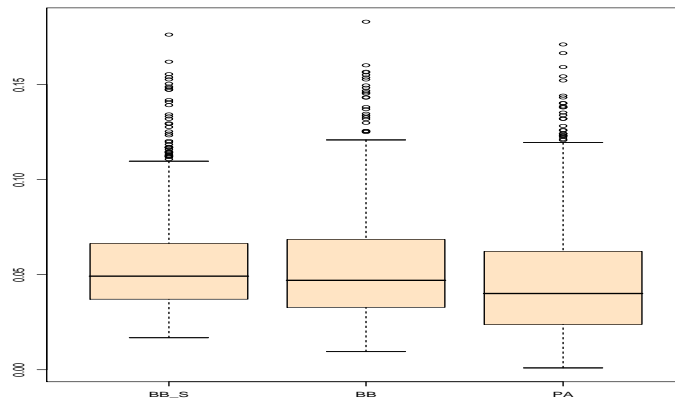


Figure 3.2: Comparison of IAE from different settings.

alternative method of smoothed BB. The output is shown in the Figure (3.2), with the first boxplot using smoothed BB, the second boxplot using BB method, and the third boxplot using the parametric method (true one). Thus, the extra computation due to kernel smoothing cannot be strongly recommended.

In order to check the performance of the BB estimator compared to some other existing alternatives, we conduct the simulations under various situations.

1. *Comparison with some semiparametric estimation methods based on binormality:*

We compare accuracy of the estimates of ROC curve and the AUC functional obtained by our BB method with BN-G [ROC GLM method by Pepe, 2000], BN-T Box-Cox [Zou and Hall, 2000] and SP [semiparametric location-scale models by Pepe, 2003, page 112]. Note that BN-G and BN-T assume a binormal model. For BN-G, BN-T and SP methods, the bootstrap is used to estimate standard errors or construct confidence intervals. Our simulation results are based on 1000 simulated data sets and corresponding 1000 resamples. Data are generated by log-normal, location-scale exponential, gamma and beta distributions (abbreviated as

A, B, C, D , respectively) with different combinations of the parameters (A: X and Y data sets are generated from the lognormal with corresponding normal parameters (u_x, σ_x) and (u_y, σ_y) , respectively; B: X 's and Y 's are generated from the exponential distribution with rate=0.5 and the location and scale parameters (u_x, σ_x) and (u_y, σ_y) , respectively; C: X 's and Y 's are generated from gamma distribution with mean and standard deviation (u_x, σ_x) and (u_y, σ_y) , respectively; D: X 's and Y 's are generated from beta distribution with mean and standard deviation (u_x, σ_x) and (u_y, σ_y) , respectively). We compare coverage probabilities and average lengths of 90% credible intervals for AUC based on different estimation methods in Table 3.1 for various parameter combinations, and also examine IAEs (see Figure 3.3) to evaluate the fitness of the curve estimates by different methods for certain parameter combinations labeled with asterisks in Table 3.1. In Figure 3.3, from (a) to (d), the boxplots of IAE are shown using the first data set of A,B,C,D, respectively in Table 3.1. The first index A (respectively, B) stands for $m=n=15$ (respectively, $m=n=50$) and second index 1, 2, 3, 4 stands for BB, BN-G, BN-T and SP methods, respectively. The grid points on $[0,1]$ are chosen with equal interval length 0.05 to calculate AUC.

From the simulation results shown in Table 3.1 and Figure 3.3, we can observe that the proposed BB method performs well in terms of accuracy and robustness. The method BN-G gives larger IAE in some data sets, along with less coverage of AUC with larger sample size in one of the location-scale exponential data sets. The method BN-T gives the lower coverage probability in some cases. In general, the coverage probability increases when the sample size increases from 15 to 50, while the mean lengths of 90% CI of AUC, IAE and their variations decrease significantly.

Table 3.1: Coverage probabilities of AUC and corresponding average lengths of the 90% CI shown beneath in the parentheses using BB, BN-G, BN-T and SP methods.

Data		$m = n = 15$				$m = n = 50$			
u_x, σ_x	u_y, σ_y	BB	BN-G	BN-T	SP	BB	BN-G	BN-T	SP
A		.899	.886	.866	.923	.893	.871	.882	.900
0,1	1, 1*	(.262)	(.254)	(.250)	(.276)	(.149)	(.141)	(.143)	(.176)
		.861	.859	.801	.820	.886	.890	.873	.150
0,1	3, 3	(.230)	(.233)	(.205)	(.275)	(.136)	(.134)	(.122)	(.162)
B		.875	.862	.860	.880	.90	.888	.856	.885
0,1	1, 1*	(.305)	(.288)	(.285)	(.304)	(.158)	(.159)	(.158)	(.167)
		.854	.857	.910	.930	.91	.772	.925	.919
0,1	3, 3	(.098)	(.097)	(.076)	(.110)	(.054)	(.062)	(.045)	(.065)
C		.886	.876	.840	.873	.886	.851	.860	.867
1,1	2, 1*	(.244)	(.234)	(.223)	(.244)	(.140)	(.132)	(.133)	(.146)
		.852	.861	.858	.882	.906	.837	.886	.847
1,1	5, 3	(.101)	(.105)	(.084)	(.108)	(.059)	(.061)	(.051)	(.057)
D		.891	.878	.867	.865	.898	.892	.774	.821
.15, .15	.2, .3*	(.321)	(.323)	(.336)	(.348)	(.182)	(.177)	(.212)	(.184)
		.896	.876	.874	.834	.886	.912	.796	.662
.15, .15	.5, .45	(.333)	(.332)	(.440)	(.341)	(.189)	(.187)	(.266)	(.211)

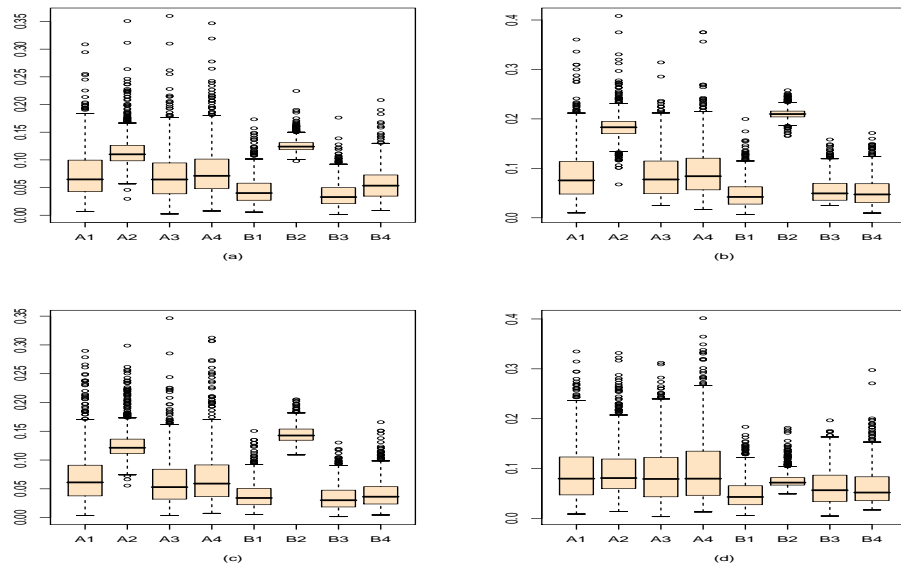


Figure 3.3: Boxplots of IAE using BB, BN-G, BN-T and SP methods.

2. Comparison with some nonparametric estimation methods:

There are several nonparametric estimation methods available to estimate the area under the curve. Qin and Zhou (2006) conduct extensive simulations to compare the accuracy and efficiency of the estimates using various methods, which are EL [Qin and Zhou, 2006], MW (Mann-Whitney two-sample rank statistics), LT [logistic transformation by Pepe, 2003, page 107], PB (standard percentile bootstrap) and PTB (percentile- t bootstrap). Two out of three simulated data are used which are under the same setting as proposed in Qin and Zhou (2006). They are distributed as normal with mean and standard deviation $(0, 1)$ and $(5^{1/2}\Phi^{-1}(AUC), 2)$ for F and G , respectively; exponential with rate=1 for F and rate= $(\frac{1}{AUC} - 1)$ for G , where the pdf of exponential distribution with rate= λ is defined by $f_{\lambda}(x) = \lambda e^{-\lambda x}$. We will compare the performance of our BB estimator with these estimators. From simulation results (see Table 3.2), we can see BB estimator performs well; especially, the BB intervals tend to be shorter.

3.4 Real data analyses

We shall illustrate the BB method to construct a credible band of curve estimate and a credible interval for the AUC estimate using the data set published by Wieand *et al.* (1989). This study is based on 51 patients in the control group diagnosed with pancreatitis and 90 patients in the case group diagnosed with pancreatic cancer by two biomarkers: a cancer antigen (CA 125) and a carbohydrate antigen (CA 19-9). For the purpose of illustration, we only choose biomarker CA 19-9. The BB estimates are based on 5000 resamples and grid points with even interval length 0.01 on $[0,1]$. We only consider a pointwise 90% credible band in this case (see Figure 3.4).

Table 3.2: Coverage probabilities and corresponding average lengths of 95% (shown beneath in the parentheses) CI for AUC obtained by BB, EL, MW, LT, PB and PTB methods. Our simulation results are based on 10000 simulated data sets and corresponding 1000 resamples. The grid points on $[0,1]$ are chosen with equal interval length 0.005, $(m, n) = (50, 50)$.

Data/AUC	BB	EL	MW	LT	PB	PTB
Normal:	.9438	.9407	.9379	.9538	.9300	.9690
.8	(.1765)	(.1783)	(.1808)	(.1808)	(.1746)	(.1971)
	.9315	.9352	.9204	.9468	.9150	.9700
.9	(.1234)	(.1281)	(.1271)	(.1326)	(.1228)	(.1591)
	.9066	.8964	.8818	.9289	.8840	.9490
.95	(.0823)	(.0874)	(.0850)	(.0930)	(.0814)	(.1360)
Exponential:	.9465	.9446	.9394	.9551	.9270	.9570
.8	(.1708)	(.1725)	(.1746)	(.1748)	(.1692)	(.1887)
	.9396	.9321	.9200	.9482	.9240	.9740
.9	(.1212)	(.1254)	(.1247)	(.1290)	(.1198)	(.1501)
	.9049	.8977	.8817	NA	.9000	.9460
.95	(.0823)	(.0881)	(.0859)	(NA)	(.0838)	(.1427)

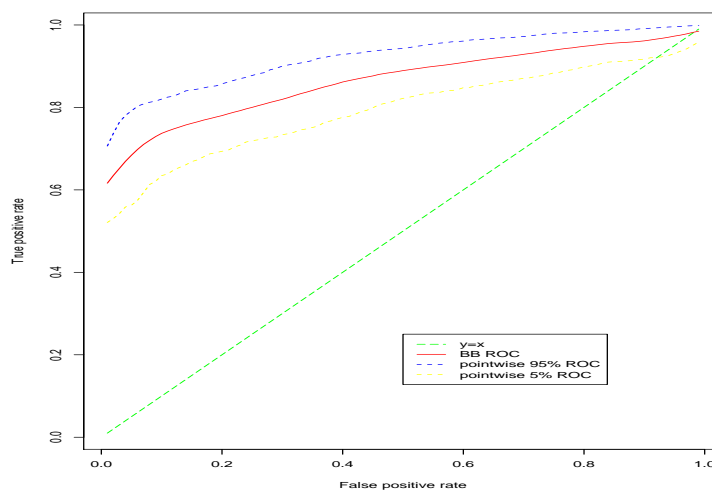


Figure 3.4: Pointwise 90% credible band of ROC curve using biomarker CA 19-9.

Our BB estimate (corresponding 95% credible interval) of AUC is 0.8542 which is similar to those of Qin and Zhang (2003) and Wan and Zhang (2007), but the corresponding confidence interval (0.7834, 0.8995) is slightly shorter than (0.789, 0.911) given by Qin and Zhang (2003), and (0.79128, 0.91053) given by Wan and Zhang (2007).

Using the testing binormality procedure shown in Section 3.2, we fail to reject the binormality assumption of biomarker CA 125 and CA 19-9 for both level 0.05 and 0.1, based on $T = 0.1469$ and $c_{0.05} = 0.2462$, $c_{0.1} = 0.2311$ for CA 125, and $T = 0.3992$ and $c_{0.05} = 0.4603$, $c_{0.1} = 0.4340$ for CA 19-9. These results are based on 1000 resamples and grid points with even interval length 0.05 on $[0,1]$.

Chapter 4

Strong approximations for ROC process

In this chapter, we first obtain strong approximations for quantile processes of resampling processes based on the bootstrap and the the BB resampling distribution by a sequence of appropriate Gaussian processes (more specifically, Kiefer processes). The ROC function is $R(t) = \bar{G}(\bar{F}^{-1}(t))$, where $\bar{F}(x) = 1 - F(x)$ and $\bar{G}(y) = 1 - G(y)$ are the survival functions of independent variables $X \sim F$ and $Y \sim G$. As a consequence, we obtain the strong approximations for the empirical estimate of $R(t)$ and the corresponding bootstrap and Bayesian versions of $R(t)$. Interestingly, it will be seen that frequentist variability of the empirical estimate of ROC can be asymptotically accurately estimated by resampled variability of bootstrap and BB procedures, given that the samples and the posterior mean (i.e, mean of ROC under BB distribution) are asymptotically equivalent to that of the empirical estimate up to the first order ($O(N^{-1/2})$), where N is the total sample size. Also, the result implies that for functionals like AUC, the empirical estimator is asymptotically normal and asymptotically equivalent to the BB estimator, and

the corresponding confidence intervals have asymptotic frequentist validity. Graphical displays of these findings will be shown in the simulation studies in Section 4.4.

4.1 Notation

For $X_1, \dots, X_n \sim \text{i.i.d. } F$, define the domain of X as $[a, b]$: $a = \sup\{x : F(x) = 0\}$, $b = \inf\{x : F(x) = 1\}$; the order statistic $X_{j:n}$, $j = 0, 1, \dots, n+1$, where $X_{0:n} = a$ and $X_{n+1:n} = b$, similarly define $V_{j:n}$ and $U_{j:n}$ for independent uniform samples V_1, \dots, V_n and U_1, \dots, U_n ; the quantile function $F^{-1}(t) = \inf\{x : F(x) \geq t\}$, $t \in [0, 1]$; for a given $0 < y < 1$, $k = \lceil ny \rceil$, where $\lceil \cdot \rceil$ is the ceiling function (that is, the smallest integer greater than or equal to x); $U(0, 1)$ denotes the uniform distribution on $[0, 1]$; abbreviate almost sure convergence by *a.s.*. Condition A [Csörgő and Révész, 1978] and Condition B are assumed, which hold for a large class of distribution functions.

Condition A: The continuous distribution function F is twice differentiable on (a, b) , $F' = f \neq 0$ on (a, b) and for some $\gamma > 0$, $\sup_{a < x < b} F(x)(1 - F(x)) \left| \frac{f'(x)}{f^2(x)} \right| \leq \gamma$.

Condition B: Let F and G satisfy Condition A and let the following conditions hold:

$$\sup_{a < x < b} F(x)(1 - F(x)) \left| \frac{g'(x)}{f^2(x)} \right|, \quad \sup_{a < x < b} F(x)(1 - F(x)) \left| \frac{g(x)}{f(x)} \right| \text{ are bounded.}$$

Define the following orders: $\delta_n = 25n^{-1} \log \log n$, $\delta_n^* = \delta_n + 2n^{-1/2}(\log \log n)^{1/2}$, $\delta_n^\# = \delta_n + n^{-1/2}(\log \log n)^{1/2}$. $l_n = n^{-1/4}(\log \log n)^{1/4}(\log n)^{1/2}$, $\tau_m = m^{-3/4}(\log \log m)^{1/4}(\log m)^{1/2}$, $\gamma_n = n^{-1} \log^2 n$. $\alpha_m^{-1} = o(m^{1/4})$, $\alpha_m^* = \alpha_m + 2m^{-1/2}(\log \log m)^{1/2}$, $\alpha_m^\# = \alpha_m + (m - 1)^{-1/2}(\log \log(m - 1))^{1/2}$.

We define the following empirical and quantile functions and processes:

$$\text{empirical function: } \mathbb{F}_n(x) = j/n, \quad \text{if } X_{j:n} \leq x < X_{j+1:n}, j = 0, 1, \dots, n. \quad (4.1)$$

$$\text{empirical process: } \mathbb{J}_n(x) = \sqrt{n}(\mathbb{F}_n(x) - F(x)) \quad (4.2)$$

$$\text{quantile function: } \mathbb{F}_n^{-1}(y) = X_{k:n} = F^{-1}(U_{k:n}), \text{ where } k = \lceil ny \rceil \quad (4.3)$$

$$\text{quantile process: } \mathbb{Q}_n(y) = \sqrt{n}(\mathbb{F}_n^{-1}(y) - F^{-1}(y)) \quad (4.4)$$

Based on samples X_1, \dots, X_n , bootstrap resample is given by $\{\mathbb{F}_n^{-1}(V_{j:n}), j = 1, \dots, n; V_1, \dots, V_n \sim \text{i.i.d. } U(0, 1) \text{ independent of } X_i\text{'s}\}$. The empirical and quantile functions of bootstrap resampling distribution are denoted as $\mathbb{F}_n^*(x)$ and $\mathbb{F}_n^{*-1}(y)$, respectively, based on the bootstrap resamples. Bootstrap empirical and quantile processes are defined as

$$\mathbb{J}_n^*(x) = \sqrt{n}(\mathbb{F}_n^*(x) - \mathbb{F}_n(x)) \quad (4.5)$$

$$\mathbb{Q}_n^*(y) = \sqrt{n}(\mathbb{F}_n^{*-1}(y) - \mathbb{F}_n^{-1}(y)) = \sqrt{n}(\mathbb{F}_n^{-1}(V_{k^*:n}) - \mathbb{F}_n^{-1}(y)) = \sqrt{n}(F^{-1}(U_{k^*:n}) - F^{-1}(U_{k:n})), \quad (4.6)$$

where $k = \lceil ny \rceil, k^* = \lceil nV_{k:n} \rceil, \mathbb{F}_n^{*-1}(y) = \mathbb{F}_n^{-1}(V_{k^*:n})$.

Based on samples X_1, \dots, X_n , BB's CDF is defined as $\mathbb{F}_n^\#(x) = \sum_{1 \leq j \leq n} \Delta_{j:n} 1(X_{j:n} \leq x), x \in \mathbb{R}$, where $V_1, \dots, V_{n-1} \sim \text{i.i.d. } U(0, 1)$, independent of X_i 's, $\Delta_{j:n} = V_{j:n-1} - V_{j-1:n-1}$. The quantile function, empirical and quantile processes of the BB resampling distributions are:

$$\text{quantile function of BB: } \mathbb{F}_n^{\#-1}(y) = \begin{cases} X_{j:n}, & V_{j-1:n-1} < y \leq V_{j:n-1}, \\ X_{0:n}, & y = 0, j = 1, 2, \dots, n. \end{cases}$$

$$\text{empirical process of BB: } \mathbb{J}_n^\#(x) = \sqrt{n}(\mathbb{F}_n^\#(x) - \mathbb{F}_n(x)), \quad (4.7)$$

$$\text{quantile process of BB: } \mathbb{Q}_n^\#(y) = \sqrt{n}(\mathbb{F}_n^{\#-1}(y) - \mathbb{F}_n^{-1}(y)). \quad (4.8)$$

Table 4.1: Notation for empirical and quantile processes

Definitions	General notation	Uniform case notation
Empirical function	$\mathbb{F}_n(x)$	$\mathbb{U}_n(x)$
Empirical process	$\mathbb{J}_n(x)$	$\mathbb{H}_n(x)$
Quantile function	$\mathbb{F}_n^{-1}(y)$	$\mathbb{U}_n^{-1}(y)$
Quantile process	$\mathbb{Q}_n(y)$	$\mathbb{W}_n(y)$
Bootstrap empirical process	$\mathbb{J}_n^*(x)$	$\mathbb{H}_n^*(x)$
Bootstrap quantile process	$\mathbb{Q}_n^*(y)$	$\mathbb{W}_n^*(y)$
BB empirical process	$\mathbb{J}_n^\#(x)$	$\mathbb{H}_n^\#(x)$
BB quantile process	$\mathbb{Q}_n^\#(y)$	$\mathbb{W}_n^\#(y)$

4.2 Strong approximations for the BB quantile processes

Theorem 4.1 (*Strong approximation for bootstrap quantile process*) Let $X_1, X_2, \dots \sim$ i.i.d. F satisfying Condition A. Then the quantile process of bootstrap $\mathbb{Q}_n^*(y)$ can be strongly approximated by a Kiefer process K , in the sense that

$$\sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} | f(F^{-1}(y))\mathbb{Q}_n^*(y) - n^{-1/2}K(y, n) | =_{a.s.} O(l_n). \quad (4.9)$$

To prove this theorem, we need the following lemma whose proof is given in Section 4.5.

Lemma 4.1 Let $X_1, \dots, X_n \sim$ i.i.d. F satisfying Condition A, and let $V_1, \dots, V_n \sim$ i.i.d. $U(0, 1)$, independent of X_1, \dots, X_n . Then for the quantile process $\mathbb{Q}_n(y)$ of X_i 's, there exists a Kiefer process K such that

$$\sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} | (y - V_{k:n})\mathbb{Q}_n(V_{k:n})f'(F^{-1}(\xi))/f(F^{-1}(\xi)) | =_{a.s.} O(l_n), \quad (4.10)$$

$$\sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} | f(F^{-1}(y))\mathbb{Q}_n(V_{k:n}) - n^{-1/2}K(y, n) | =_{a.s.} O(l_n), \quad (4.11)$$

where $k = \lceil ny \rceil$ and for any ξ lies between y and $V_{k:n}$.

Proof.

When $\delta_n^* \leq y \leq 1 - \delta_n^*$, $[\delta_n^*, 1 - \delta_n^*] \subset [\delta_n, 1 - \delta_n]$. Let $k^* = \lceil nV_{k:n} \rceil$. Then

$$\begin{aligned} \mathbb{Q}_n^*(y) &= \sqrt{n}(F^{-1}(U_{k^*:n}) - F^{-1}(U_{k:n})) \\ &= \sqrt{n}(F^{-1}(U_{k^*:n}) - F^{-1}(V_{k:n})) + \sqrt{n}(F^{-1}(V_{k:n}) - y) - \sqrt{n}(F^{-1}(U_{k:n}) - y) \\ &= \mathbb{Q}_n(V_{k:n}) + \tilde{\mathbb{Q}}_n(y) - \mathbb{Q}_n(y), \end{aligned} \quad (4.12)$$

where $\mathbb{Q}_n(y)$ and $\tilde{\mathbb{Q}}_n(y)$ are independent quantile processes for X_i 's and \tilde{X}_i 's respectively, $X_i = F^{-1}(U_i)$, $\tilde{X}_i = F^{-1}(V_i)$, $U_i \sim \text{i.i.d. } U(0, 1)$, $V_i \sim \text{i.i.d. } U(0, 1)$, and U_i 's and V_i 's are independent, $i = 1, \dots, n$. By Theorem B of Appendix A applied to $\mathbb{Q}_n(y)$ and $\tilde{\mathbb{Q}}_n(y)$, respectively, there exist independent Kiefer processes K and \tilde{K} , satisfying

$$\begin{aligned} \sup_{\delta_n \leq y \leq 1 - \delta_n} |f(F^{-1}(y))\mathbb{Q}_n(y) - n^{-1/2}K(y, n)| &=_{a.s.} O(l_n), \\ \sup_{\delta_n \leq y \leq 1 - \delta_n} |f(F^{-1}(y))\tilde{\mathbb{Q}}_n(y) - n^{-1/2}\tilde{K}(y, n)| &=_{a.s.} O(l_n). \end{aligned} \quad (4.13)$$

By applying Lemma 4.1 and (4.12), (4.13), we obtain $\sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} |f(F^{-1}(y))\mathbb{Q}_n^*(y) - n^{-1/2}\tilde{K}(y, n)| =_{a.s.} O(l_n)$. ■

Theorem 4.2 (Strong approximation for the BB quantile process) *Let $X_1, \dots, X_n \sim \text{i.i.d. } F$ satisfying Condition A. Then the BB quantile process $\mathbb{Q}_n^\#(y)$ can be strongly approximated by a Kiefer process K , in the sense that*

$$\sup_{\delta_n^\# \leq y \leq 1 - \delta_n^\#} |f(F^{-1}(y))\mathbb{Q}_n^\#(y) - n^{-1/2}K(y, n)| =_{a.s.} O(l_n). \quad (4.14)$$

The proof requires the following lemma whose proof is deferred to Section 4.5.

Lemma 4.2 *Let $X_1, \dots, X_n \sim \text{i.i.d. } F$ satisfying Condition A, $V_1, \dots, V_{n-1} \sim \text{i.i.d. } U(0, 1)$, $\mathbb{F}_n^{-1}(y)$, $\tilde{\mathbb{U}}_{n-1}(x)$ and $\mathbb{F}_n^{\#-1}(y)$ denote the quantile function of X 's, empirical function of V 's and quantile function of the Bayesian bootstrap respectively. Then*

$$\sup_{0 < y < 1} \sqrt{n} |\mathbb{F}_n^{\#-1}(y) - \mathbb{F}_n^{-1}(\tilde{\mathbb{U}}_{n-1}(y))| =_{a.s.} O(n^{-1/2} \log n). \quad (4.15)$$

In addition, there exists a Kiefer process K such that

$$\sup_{\delta_n^\# \leq y \leq 1 - \delta_n^\#} |\sqrt{n}f(F^{-1}(y))(\mathbb{F}_n^{-1}(\tilde{U}_{n-1}(y)) - \mathbb{F}_n^{-1}(y)) - n^{-1/2}K(y, n)| =_{a.s.} O(l_n). \quad (4.16)$$

Proof.

We may represent $X_i = F^{-1}(U_i)$, where U_i 's \sim i.i.d. $U(0, 1)$, $i = 1, \dots, n$, $V_1, \dots, V_{n-1} \sim$ i.i.d. $U(0, 1)$, independent of X 's. Then we have (cf. See Table 4.2 for Values of $\mathbb{F}_n^{\#-1}(y)$ and $\tilde{U}_{n-1}(y)$)

$$\mathbb{F}_n^{\#-1}(y) = \begin{cases} \mathbb{F}_n^{-1}\left(\frac{1+(n-1)\tilde{U}_{n-1}(y)}{n}\right), & V_{k:n-1} < y < V_{k+1:n-1}, k = 0, \dots, n-1, \\ \mathbb{F}_n^{-1}\left(\frac{(n-1)\tilde{U}_{n-1}(y)}{n}\right), & y = V_{k:n-1}, k = 1, \dots, n-1. \end{cases}$$

Table 4.2: Values of $\mathbb{F}_n^{\#-1}(y)$ and $\tilde{U}_{n-1}(y)$

y	$\mathbb{F}_n^{\#-1}(y)$	$\tilde{U}_{n-1}(y)$
$0 \leq y < V_{1:n-1}$	$X_{1:n}$	0
$= V_{1:n-1}$	$X_{1:n}$	$1/(n-1)$
...
$V_{k:n-1} < y < V_{k+1:n-1}$	$X_{k+1:n}$	$k/(n-1)$
$y = V_{k+1:n-1}$	$X_{k+1:n}$	$(k+1)/(n-1)$
...
$V_{n-2:n-1} < y < V_{n-1:n-1}$	$X_{n-1:n}$	$(n-2)/(n-1)$
$y = V_{n-1:n-1}$	$X_{n-1:n}$	1
$V_{n-1:n-1} < y \leq 1$	$X_{n:n}$	1

Also, the BB quantile process $Q_n^\#(y)$ can be split as

$$Q_n^\#(y) = \sqrt{n}(\mathbb{F}_n^{\#-1}(y) - \mathbb{F}_n^{-1}(\tilde{U}_{n-1}(y))) + \sqrt{n}(\mathbb{F}_n^{-1}(\tilde{U}_{n-1}(y)) - \mathbb{F}_n^{-1}(y)). \quad (4.17)$$

Then by applying Lemma 4.2, (4.17) and Condition A, Theorem 4.2 holds for a Kiefer process K satisfying (4.16). ■

4.3 Limit theorem for BB estimator of ROC curve and its functionals

Let $X_1, \dots, X_m \sim \text{i.i.d. } F$, and $Y_1, \dots, Y_n \sim \text{i.i.d. } G$. For example, in medical contexts F may stand for the CDF of a population without disease and G for the CDF of a population with disease. Assume F and G satisfy Condition A and B and X 's and Y 's are independent. Let $N = m + n$, when $N \rightarrow \infty$, $\frac{m}{N} \rightarrow \lambda$, $0 < \lambda < 1$.

Recall Kiefer process defined as follows: $\{K(x, y) : 0 \leq x \leq 1, 0 \leq y < \infty\}$ is determined by $K(x, y) = W(x, y) - xW(1, y)$, where $W(x, y)$ be a two-parameter standard Wiener process. Its covariance function is $E[K(x_1, y_1)K(x_2, y_2)] = (x_1 \wedge x_2 - x_1 x_2)(y_1 \wedge y_2)$.

4.3.1 Functional limit theorems for ROC curve

The ROC curve is defined as $\{(P(X > c_t), P(Y > c_t)) : X \sim F, Y \sim G, c_t \in \mathbb{R}\}$, or alternatively as $R(t) = \bar{G}(\bar{F}^{-1}(t))$; its derivative is $R'(t) = g(\bar{F}^{-1}(t))/f(\bar{F}^{-1}(t))$. An empirical estimate of ROC can be obtained by plugging in their empirical counterparts. Based on the given samples, BB estimate can be obtained by doing the following two steps: (1) get one realization from the posterior distribution of ROC curves by plugging the empirical and quantile function of the BB resampling distributions into the ROC functional form, respectively; (2) get BB estimate by the posterior mean. The BB version of ROC curve described above can be equivalently regarded as putting two independent non-informative Dirichlet priors on F and G . Hence, our BB estimate of ROC is essentially a nonparametric Bayesian estimate. Let $\mathbb{R}(t) = \mathbb{R}_{m,n}(t) = \bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t))$, $\mathbb{R}_{m,n}^*(t) = \bar{\mathbb{G}}_n^*(\bar{\mathbb{F}}_m^{*-1}(t))$, and $\mathbb{R}_{m,n}^\#(t) = \bar{\mathbb{G}}_n^\#(\bar{\mathbb{F}}_m^{\#-1}(t))$.

Theorem 4.3 (*Functional limit theorem for empirical, bootstrap and BB version of ROC*)

curves.) Let $X_1, \dots, X_m \sim i.i.d. F$, $X_i = F^{-1}(U_i)$ and $Y_1, \dots, Y_n \sim i.i.d. G$. Assume F and G satisfy Condition A and B. Then

$$\mathbb{R}_{m,n}(t) = R(t) + R'(t) \frac{K_1(t, m)}{m} + \frac{K_2(R(t), n)}{n} + O(\alpha_m^{-1} \tau_m), \quad t \in (\alpha_m, 1 - \alpha_m) \quad (4.18)$$

$$\mathbb{R}_{m,n}^*(t) = \mathbb{R}_{m,n}(t) + R'(t) \frac{K_1(t, m)}{m} + \frac{K_2(R(t), n)}{n} + O(\alpha_m^{-1} \tau_m), \quad t \in (\alpha_m^*, 1 - \alpha_m^*) \quad (4.19)$$

$$\mathbb{R}_{m,n}^\#(t) = \mathbb{R}_{m,n}(t) + R'(t) \frac{K_1(t, m)}{m} + \frac{K_2(R(t), n)}{n} + O(\alpha_m^{-1} \tau_m), \quad t \in (\alpha_m^\#, 1 - \alpha_m^\#) \quad (4.20)$$

in the sense of a.s. for (4.19) and (4.20), where K_1 and K_2 are independent generic Kiefer processes (not identical in each appearance).

Proof.

1. **Proof of (4.18) in Theorem 4.3:** By Theorem A and B of Appendix A, when $t \in (\alpha_m, 1 - \alpha_m) \subset (\delta_m, 1 - \delta_m)$, the empirical ROC curve estimator $\mathbb{R}_{m,n}(t)$ can be approximated by the following:

$$\begin{aligned} \mathbb{R}_{m,n}(t) &= \bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)) = \bar{G}(\bar{\mathbb{F}}_m^{-1}(t)) + \frac{1}{n} K_2(\bar{G}(\bar{\mathbb{F}}_m^{-1}(t)), n) + O(\gamma_n) \\ &= \bar{G}(\bar{F}^{-1}(t)) + \frac{1}{n} K_2(\bar{G}(\bar{F}^{-1}(t)), n) + I_1 + I_2 + O(\gamma_n), \end{aligned}$$

where

$$I_1 = \bar{G}(\bar{\mathbb{F}}_m^{-1}(t)) - \bar{G}(\bar{F}^{-1}(t)) = m^{-1} R'(t) K_1(t, m) + O(\alpha_m^{-1} \tau_m) \quad (4.21)$$

$$I_2 = n^{-1} K_2(\bar{G}(\bar{\mathbb{F}}_m^{-1}(t)), n) - n^{-1} K_2(\bar{G}(\bar{F}^{-1}(t)), n) = O((m/n)^{1/2} \alpha_m^{-1/2} \tau_m) \quad (4.22)$$

Proof of (4.21) : By Theorem B of Appendix A, we have

$$I_1 = g(\bar{F}^{-1}(t)) \frac{m^{-1} K_1(t, m) - O(\tau_m)}{f(\bar{F}^{-1}(t))} - g'(\bar{F}^{-1}(\xi)) \left(\frac{m^{-1} K_1(t, m) - O(\tau_m)}{f(\bar{F}^{-1}(t))} \right)^2,$$

and ξ lies between t and $U_{\lceil mt \rceil:m}$, which ensures $f(\bar{F}^{-1}(\xi))/f(\bar{F}^{-1}(t)) \leq 10^\gamma$ by Theorem N of Appendix A. By Condition B, we get

$$|R'(t)O(\tau_m)| = \left| t(1-t) \frac{g(\bar{F}^{-1}(t))}{f(\bar{F}^{-1}(t))} \frac{O(\tau_m)}{t(1-t)} \right| \leq O(\alpha_m^{-1}\tau_m) \quad (4.23)$$

$$\begin{aligned} & \left| \xi(1-\xi) \frac{g'(\bar{F}^{-1}(\xi))}{f^2(\bar{F}^{-1}(\xi))} \frac{f^2(\bar{F}^{-1}(\xi))}{f^2(\bar{F}^{-1}(t))\xi(1-\xi)} \right| (m^{-1}K_1(t, m) - O(\tau_m))^2 \\ & \leq O(\alpha_m^{-1}m^{-1} \log \log m) \quad a.s. \quad \square \end{aligned} \quad (4.24)$$

By combining (4.23) and (4.24), we finish the proof of (4.21). \square

Proof of (4.22) : By modulus of continuity for Brownian motion on bounded interval, we get

$$\begin{aligned} |I_2| &= \left| \frac{1}{n} K_2(\bar{G}(\bar{\mathbb{F}}_m^{-1}(t)), n) - \frac{1}{n} K_2(\bar{G}(\bar{F}^{-1}(t)), n) \right| \\ &\leq n^{-1/2} |I_1|^{1/2} (\log(1/|I_1|))^{1/2} \\ &\leq n^{-1/2} \left| \frac{R'(t)}{m} K_1(t, m) + O(\alpha_m^{-1}\tau_m) \right|^{1/2} (\log m)^{1/2} \\ &\leq n^{-1/2} |O(\alpha_m^{-1}m^{-1/2}(\log \log m)^{1/2}) + O(\alpha_m^{-1}\tau_m)|^{1/2} (\log m)^{1/2} \\ &= O((m/n)^{1/2} \alpha_m^{-1/2} \tau_m). \quad \square \end{aligned} \quad (4.25)$$

Therefore, by combining the results (4.21) and (4.22), we complete the proof of (4.18). \square

2. **Proof of (4.19) in Theorem 4.3:** The idea of the proof of (4.19) is similar to (4.18), except for the following: (1) We can expand I_{12}^* in Taylor's series instead of I_1^* ($I_1^* = I_{12}^* + I_{11}^*$); (2) Because the bootstrap version is expanded around the empirical estimator, the Taylor's series expansion is evaluated at the empirical point $\bar{\mathbb{F}}_m^{-1}(t)$ instead of at the true point $\bar{F}^{-1}(t)$. Measuring these difference enlarges the range of domain from $t \in (\alpha_m, 1 - \alpha_m)$ to $t \in (\alpha_m^*, 1 - \alpha_m^*)$, which allows the limiting

distribution based on bootstrap resampling distribution to be the same as that of the empirical one.

By Theorem 4.1 and Theorem C of Appendix A, there exist two independent Kiefer processes K_1 and K_2 , such that when $t \in (\alpha_m^*, 1 - \alpha_m^*) \subset (\delta_m^*, 1 - \delta_m^*)$, the strong approximation for bootstrap version of ROC curve is given by

$$\mathbb{R}_{m,n}^*(t) = \bar{\mathbb{G}}_n^*(\bar{\mathbb{F}}_m^{*-1}(t)) = \bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)) + I_1^* + n^{-1}K_2(\bar{G}(\bar{F}^{-1}(t)), n) + I_2^* + O(\tau_n)$$

where $I_1^* = \bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{*-1}(t)) - \bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)) = I_{11}^* + I_{12}^*$, $I_{12}^* = \bar{G}(\bar{\mathbb{F}}_m^{*-1}(t)) - \bar{G}(\bar{\mathbb{F}}_m^{-1}(t))$, $I_{11}^* = \bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{*-1}(t)) - \bar{G}(\bar{\mathbb{F}}_m^{*-1}(t)) - (\bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)) - \bar{G}(\bar{\mathbb{F}}_m^{-1}(t)))$;

$$I_2^* = I_{21}^* + I_{22}^*, I_{21}^* = n^{-1}K_2(\bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{*-1}(t)), n) - n^{-1}K_2(\bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)), n),$$

$$I_{22}^* = n^{-1}K_2(\bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)), n) - n^{-1}K_2(\bar{G}(\bar{F}^{-1}(t)), n). \text{ We can show}$$

$$I_{12}^* = m^{-1}R'(t)K_1(t, m) + O(\alpha_m^{-1}\tau_m) \quad (4.26)$$

$$I_{11}^* = O((m/n)^{1/2}\alpha_m^{-1/2}\tau_m) \quad (4.27)$$

$$I_{21}^* = O((m/n)^{1/2}\alpha_m^{-1/2}\tau_m) \quad (4.28)$$

$$I_{22}^* = O(\tau_m). \quad (4.29)$$

Proof of (4.26) : By taking Taylor's series expansion of I_{12}^* , we have

$$I_{12}^* = g(\bar{\mathbb{F}}_m^{-1}(t)) \frac{m^{-1}K_1(t, m) - O(\tau_m)}{f(\bar{F}^{-1}(t))} - g'(\bar{F}^{-1}(\xi)) \left(\frac{m^{-1}K_1(t, m) - O(\tau_m)}{f(\bar{F}^{-1}(t))} \right)^2, \quad (4.30)$$

where ξ lies between $U_{k:m}$ and $U_{k^*:m}$, $k = \lceil mt \rceil$, $k^* = \lceil mV_{k:m} \rceil$, $\bar{\mathbb{F}}_m^{*-1}(t) = \bar{F}^{-1}(U_{k^*:m})$, and $\bar{\mathbb{F}}_m^{-1}(t) = \bar{F}^{-1}(U_{k:m})$.

- Because the first term of (4.30) can be split as

$$g(\bar{F}^{-1}(t)) \frac{m^{-1}K_1(t, m) - O(\tau_m)}{f(\bar{F}^{-1}(t))} + \frac{g(\bar{\mathbb{F}}_m^{-1}(t)) - g(\bar{F}^{-1}(t))}{f(\bar{F}^{-1}(t))} (m^{-1}K_1(t, m) - O(\tau_m)), \quad (4.31)$$

and then by (4.23), the first term of (4.31) can be majorized by $\frac{R'(t)}{m} K_1(t, m) + O(\alpha_m^{-1} \tau_m)$. The second term of (4.31) can be bounded by

$$\begin{aligned} & \left| \frac{g'(\bar{F}^{-1}(\xi^*))}{f(\bar{F}^{-1}(t))} \right| | \bar{\mathbb{F}}_m^{-1}(t) - \bar{F}^{-1}(t) | | m^{-1}K_1(t, m) - O(\tau_m) | \\ & \leq \left| \xi^*(1 - \xi^*) \frac{g'(\bar{F}^{-1}(\xi^*))}{f^2(\bar{F}^{-1}(\xi^*))} \right| \frac{f^2(\bar{F}^{-1}(\xi^*))}{f^2(\bar{F}^{-1}(t)) \xi^*(1 - \xi^*)} O(m^{-1} \log \log m) \end{aligned}$$

which is $O(\alpha_m^{-1} m^{-1} \log \log m)$, where ξ^* lies between $U_{k:m}$ and t . This condition ensures the boundedness of $\frac{f^2(\bar{F}^{-1}(\xi^*))}{f^2(\bar{F}^{-1}(t))}$ by Theorem N of Appendix A. Also, because $t \in (\alpha_m^*, 1 - \alpha_m^*)$, by Theorem K of Appendix A, $|\frac{1}{\xi^*}| \leq O(\alpha_m^{-1})$. Therefore the first term of (4.30) can be bounded by $m^{-1}R'(t)K_1(t, m) + O(\alpha_m^{-1} \tau_m)$.

- The second term of (4.30) can be majorized as follows:

$$\begin{aligned} & \left| \xi(1 - \xi) \frac{g'(\bar{F}^{-1}(\xi))}{f^2(\bar{F}^{-1}(\xi))} \right| \frac{f^2(\bar{F}^{-1}(\xi))}{f^2(\bar{F}^{-1}(t)) \xi(1 - \xi)} O(m^{-1} \log \log m) \\ & \leq O(\alpha_m^{-1}) O(m^{-1} \log \log m). \end{aligned}$$

By combining the results of these two parts, we prove the statement (4.26). \square

Proof of (4.27) : By Theorem A of Appendix A, there exists Kiefer process K , such that I_{11}^* can be majorized by

$$\begin{aligned} & \left| \frac{K(\bar{G}(\bar{\mathbb{F}}_m^{-1}(t)), n) - K(\bar{G}(\bar{F}^{-1}(t)), n)}{n} \right| + O(n^{-1} \log^2 n) \\ & \leq n^{-1/2} |I_{12}^*|^{1/2} (\log(1/|I_{12}^*|))^{1/2} + O(n^{-1} \log^2 n) \quad (4.32) \\ & = O((m/n)^{1/2} \alpha_m^{-1/2} \tau_m) \text{ (following the similar argument in (4.25))} . \square \end{aligned}$$

By combining the results of I_{11}^* and I_{12}^* , we have

$$I_1^* = m^{-1}R'(t)K_1(t, m) + O(\alpha_m^{-1}\tau_m). \quad (4.33)$$

Proof of (4.28) : By following the same argument of (4.25) using (4.33), the result (4.28) is immediate.

Proof of (4.29) : By applying (4.18), we have

$$\begin{aligned} |I_{22}^*| &\leq n^{-1/2} |\mathbb{R}_{m,n} - R(t)|^{1/2} (\log(|\mathbb{R}_{m,n} - R(t)|))^{-1/2} \\ &\leq n^{-1/2} \left| \frac{R'(t)}{m} K_1(t, m) + \frac{1}{n} K_2(R(t), n) + O(\alpha_m^{-1}\tau_m) \right|^{1/2} O((\log N)^{1/2}) \\ &\leq n^{-1/2} O((m/n)^{1/2} \alpha_m^{-1/2} \tau_m + n^{1/2} \tau_n + \alpha_m^{-1/2} \tau_m^{-1/2} (\log N)^{1/2}) = O(\tau_m). \quad \square \end{aligned} \quad (4.34)$$

Hence, by combining (4.26), (4.27), (4.28) and (4.29), we complete the proof of (4.19). \square

3. **Proof of (4.20) in Theorem 4.3:** The proof of (4.20) follows the same arguments in the proof of (4.19) with the following modifications: (1) In order to use Taylor's series expansion and the boundedness result of the ratio of $\frac{f(\bar{F}^{-1}(\xi))}{f(\bar{F}^{-1}(t))}$, where ξ lies between t and $\tilde{\mathbb{U}}_{m-1}(t)$, we will do Taylor's series expansion on $I_{122}^\#$ instead of $I_{12}^\#$. That is, the decomposition $I_{12}^\# = I_{121}^\# + I_{122}^\#$ is essential; (2) Similarly, one consequence of measuring the remainder term is to shrink the range of domain from $t \in (\alpha_m^*, 1 - \alpha_m^*)$ to $t \in (\alpha_m^\#, 1 - \alpha_m^\#)$.

By Theorem 4.2 and Theorem C of Appendix A, there exist two independent Kiefer processes K_1 and K_2 , such that when $t \in (\alpha_m^\#, 1 - \alpha_m^\#) \subset (\delta_m^\#, 1 - \delta_m^\#)$, the strong approximation for the BB version of ROC curve is given by

$$\mathbb{R}_{m,n}^\#(t) = \bar{\mathbb{G}}_n^\#(\bar{\mathbb{F}}_m^{\#-1}(t)) = \bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)) + I_1^\# + n^{-1}K_2(\bar{G}(\bar{F}^{-1}(t)), n) + I_2^\# + O(\tau_n),$$

where

$$\begin{aligned}
I_1^\# &= \bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{\#-1}(t)) - \bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)) = I_{11}^\# + I_{12}^\# \\
I_{11}^\# &= \bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{\#-1}(t)) - \bar{G}(\bar{\mathbb{F}}_m^{\#-1}(t)) - (\bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)) - \bar{G}(\bar{\mathbb{F}}_m^{-1}(t))) \\
I_{12}^\# &= \bar{G}(\bar{\mathbb{F}}_m^{\#-1}(t)) - \bar{G}(\bar{\mathbb{F}}_m^{-1}(t)) = I_{121}^\# + I_{122}^\# \\
I_{121}^\# &= \bar{G}(\bar{\mathbb{F}}_m^{\#-1}(t)) - \bar{G}(\bar{\mathbb{F}}_m^{-1}(\tilde{\mathbb{U}}_{m-1}(t))) \\
I_{122}^\# &= \bar{G}(\bar{\mathbb{F}}_m^{-1}(\tilde{\mathbb{U}}_{m-1}(t))) - \bar{G}(\bar{\mathbb{F}}_m^{-1}(t)) \\
I_2^\# &= \{n^{-1}K_2(\bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{\#-1}(t)), n) - n^{-1}K_2(\bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)), n)\} \\
&\quad + \{n^{-1}K_2(\bar{\mathbb{G}}_n(\bar{\mathbb{F}}_m^{-1}(t)), n) - n^{-1}K_2(\bar{G}(\bar{F}^{-1}(t)), n)\},
\end{aligned}$$

and $\tilde{\mathbb{U}}_{m-1}(t)$ is the empirical function of $V_1, \dots, V_{m-1} \sim \text{i.i.d. } U(0, 1)$. First, we will show that

$$I_{122}^\# = m^{-1}R'(t)K_1(t, m) + O(\alpha_m^{-1}\tau_m). \quad (4.35)$$

Proof of (4.35) :

$$\begin{aligned}
I_{122}^\# &= \bar{G}(\bar{\mathbb{F}}_m^{-1}(\tilde{\mathbb{U}}_{m-1}(t))) - \bar{G}(\bar{\mathbb{F}}_m^{-1}(t)) = \{-g(\bar{\mathbb{F}}_m^{-1}(t))[\bar{\mathbb{F}}_m^{-1}(\tilde{\mathbb{U}}_{m-1}(t)) - \bar{\mathbb{F}}_m^{-1}(t)]\} \\
&\quad + \{-g'(\bar{\mathbb{F}}_m^{-1}(\xi))[\bar{\mathbb{F}}_m^{-1}(\tilde{\mathbb{U}}_{m-1}(t)) - \bar{\mathbb{F}}_m^{-1}(t)]^2\}, \quad (4.36)
\end{aligned}$$

where ξ lies between t and $\tilde{\mathbb{U}}_{m-1}(t)$. The first term of (4.36) can be strongly approximated by $m^{-1}R'(t)K_1(t, m) + O(\alpha_m^{-1}\tau_m)$, because $-g(\bar{\mathbb{F}}_m^{-1}(t))[\bar{\mathbb{F}}_m^{-1}(\tilde{\mathbb{U}}_{m-1}(t)) - \bar{\mathbb{F}}_m^{-1}(t)]$ can be split as

$$\{-g(\bar{\mathbb{F}}_m^{-1}(t))[\bar{\mathbb{F}}_m^{\#-1}(t) - \bar{\mathbb{F}}_m^{-1}(t)]\} + \{-g(\bar{\mathbb{F}}_m^{-1}(t))[\bar{\mathbb{F}}_m^{-1}(\tilde{\mathbb{U}}_{m-1}(t)) - \bar{\mathbb{F}}_m^{\#-1}(t)]\} \quad (4.37)$$

The first term of (4.37) can be strongly approximated by $m^{-1}R'(t)K_1(t, m) + O(\alpha_m^{-1}\tau_m)$ due to Theorem 4.2, and by an argument similar to the proof of the

bootstrap case; the second term of (4.37) can be bounded *a.s.* by $O(\frac{\log m}{m})$ due to Condition A and Lemma 4.2.

The second term of (4.36) can be bounded *a.s.* by $O(\alpha_m^{-1}\tau_m)$, because

$$\begin{aligned}
& | -g'(\bar{\mathbb{F}}_m^{-1}(\xi)) | [\bar{\mathbb{F}}_m^{-1}(\tilde{\mathbb{U}}_{m-1}(t)) - \bar{\mathbb{F}}_m^{-1}(t)]^2 \\
& \leq 2 | -g'(\bar{\mathbb{F}}_m^{-1}(\xi)) | \{ [\bar{\mathbb{F}}_m^{\#-1}(t) - \bar{\mathbb{F}}_m^{-1}(t)]^2 + [\bar{\mathbb{F}}_m^{-1}(\tilde{\mathbb{U}}_{m-1}(t)) - \bar{\mathbb{F}}_m^{\#-1}(t)]^2 \} \\
& \leq 2 | -g'(\bar{\mathbb{F}}_m^{-1}(\xi)) | [\bar{\mathbb{F}}_m^{\#-1}(t) - \bar{\mathbb{F}}_m^{-1}(t)]^2 + O((\frac{\log m}{m})^2) \text{ (by Lemma 4.2, Condition A)} \\
& \leq 2 | -g'(\bar{\mathbb{F}}_m^{-1}(\xi)) | \left(\frac{m^{-1}K_1(t, m) - O(\tau_m)}{f(\bar{F}^{-1}(t))} \right)^2 + O((m^{-1} \log m)^2) \\
& \leq O\left(\frac{g'(\bar{\mathbb{F}}_m^{-1}(\xi))}{f^2(\bar{F}^{-1}(t))}\right) \cdot O(m^{-1} \log \log m) + O((m^{-1} \log m)^2) \\
& \leq O(\alpha_m^{-1}\tau_m). \tag{4.38}
\end{aligned}$$

The statement (4.38) holds because

$$\begin{aligned}
\left| \frac{g'(\bar{\mathbb{F}}_m^{-1}(\xi))}{f^2(\bar{F}^{-1}(t))} \right| &= \left| \frac{g'(\bar{\mathbb{F}}_m^{-1}(\xi))}{f^2(\bar{\mathbb{F}}_m^{-1}(\xi))} \right| \left| \frac{f^2(\bar{\mathbb{F}}_m^{-1}(\xi))}{f^2(\bar{F}^{-1}(\xi))} \right| \left| \frac{f^2(\bar{F}^{-1}(\xi))}{f^2(\bar{F}^{-1}(t))} \right| \\
&\leq O(\alpha_m^{-1}), \tag{4.39}
\end{aligned}$$

where the second and third term of (4.39) can be bounded by a constant by Theorem N of Appendix A; the first term of (4.39) can be bounded by $O(\alpha_m^{-1})$ due to Condition B and

$$| \xi - t | \leq (m-1)^{-1/2} (\log \log(m-1))^{1/2} \text{ by Theorem J}$$

$$t \in (\alpha_m^{\#}, 1 - \alpha_m^{\#}). \quad \square$$

In addition, $I_{121}^{\#}$ is bounded *a.s.* by $O(\frac{\log m}{m})$ by Condition A and Lemma 4.2. Hence $I_{12}^{\#} = m^{-1}R'(t)K_1(t, m) + O(\alpha_m^{-1}\tau_m)$. $I_{11}^{\#}$ can be bounded *a.s.* by $O((m/n)^{1/2}\alpha_m^{-1/2}\tau_m)$ followed the same argument of (4.32) incorporated with the result of $I_{12}^{\#}$. Therefore

$I_1^\# = m^{-1}R'(t)K_1(t, m) + O(\alpha_m^{-1}\tau_m)$. The same argument used in the proof of the bootstrap case can be used to prove $I_2^\# = O(\tau_m)$. Hence, we complete the proof of (4.20). \square

Hence, we finish the proof of Theorem 4.3. \blacksquare

Remark 4.1 *The similar result of (4.18) can be found in Hsieh and Turnbull (1996). The difference between our strong approximation for $\mathbb{R}_{m,n}(t)$ given by (4.18) and theirs is a different trade-off between interval of validity and the approximation error. They use any fixed subinterval of $[0,1]$ and obtain a better approximation error. From a practical point of view, inclusion of levels near 0 is important. Our results treat the domain of $R(t)$ as the whole of $[0,1]$ in the limiting sense.*

4.3.2 Implication of the functional limit theorems for ROC curves

From (4.18), then as $L^\infty[0,1]$ valued random function,

$$\begin{aligned} & \sqrt{N}(\mathbb{R}_{m,n}(t) - R(t)) \\ &= \frac{\sqrt{N}}{\sqrt{m}}R'(t)\frac{K_1(t, m)}{\sqrt{m}} + \frac{\sqrt{N}}{\sqrt{n}}\frac{K_2(R(t), n)}{\sqrt{n}} + O(\sqrt{N}\alpha_m^{-1}\tau_m) \\ &\rightsquigarrow \frac{1}{\sqrt{\lambda}}R'(t)B_1(t) + \frac{1}{\sqrt{1-\lambda}}B_2(R(t)), \end{aligned} \quad (4.40)$$

in the sense of generalized weak convergence on non-separable spaces [Van der Vaart and Wellner, 1996], where B_1 and B_2 are independent Brownian bridges. Similarly, from (4.20), we can get, *a.s.*, conditionally on samples,

$$\sqrt{N}(\mathbb{R}_{m,n}^\#(t) - \mathbb{R}_{m,n}(t)) \rightsquigarrow \frac{1}{\sqrt{\lambda}}R'(t)B_1(t) + \frac{1}{\sqrt{1-\lambda}}B_2(R(t)). \quad (4.41)$$

Remark 4.2 *Li et al. (1996) obtain the weak convergence limit for the ROC process on a certain subinterval of $[0,1]$ using the product-limit (PL) estimator of the survival*

function instead of the empirical estimator. The approach based on strong approximation lets us work with real valued random variables and ordinary Taylor's series expansion, derive a stronger representation, and treat the whole domain $t \in [0, 1]$ in the limit. In contrast, the weak convergence approach must leave out some neighborhoods of 0 and 1. Further, functional delta method does not apply directly to BB and bootstrap resampling distributions because the asymptotic means are dependent on the sample size.

Result 1 If A is a continuity set for the process on the right hand side of (4.40), then

$$\Pr^\# \{ \sqrt{N}(\mathbb{R}_{m,n}^\#(\cdot) - \mathbb{R}(\cdot)) \in A \} - \Pr \{ \sqrt{N}(\mathbb{R}_{m,n}(\cdot) - R(\cdot)) \in A \} \rightarrow 0. \quad (4.42)$$

where $\Pr^\#$ stands for the BB resampling probability conditional on samples.

Result 2 [Equivalence of empirical and BB estimator of ROC] Conditionally on samples,

$$\sqrt{N} \| \mathbb{E}(\mathbb{R}_{m,n}^\#(\cdot) \mid X_1, \dots, X_m, Y_1, \dots, Y_n) - \mathbb{R}_{m,n}(\cdot) \|_{L^\infty} \rightarrow 0. \quad a.s. \quad (4.43)$$

4.3.3 Limit theorems for AUC

It is not difficult to show the following corollaries.

Corollary 4.1 Let $\psi : \mathbb{R} \rightarrow \mathbb{R}$ be a bounded function with continuous second derivatives.

Then for any $q > 0$ being the continuity point of the limiting random variable

$$\sup_{t \in (0,1)} \{ \psi'(R(t)) [\lambda^{-1/2} R'(t) B_1(t) + (1 - \lambda)^{-1/2} B_2(R(t))] \},$$

we have that almost surely for the given samples,

$$\begin{aligned} & | Pr^* \left\{ \sup_{t \in (\alpha_m^*, 1 - \alpha_m^*)} \sqrt{N} \left| \psi(\mathbb{R}_{m,n}^*(t)) - \psi(\mathbb{R}_{m,n}(t)) \right| \leq q \right\} \\ & - Pr \left\{ \sup_{t \in (\alpha_m^*, 1 - \alpha_m^*)} \sqrt{N} \left| \psi(\mathbb{R}_{m,n}(t)) - \psi(R(t)) \right| \leq q \right\} | \rightarrow 0 \end{aligned} \quad (4.44)$$

$$\begin{aligned} & | Pr^\# \left\{ \sup_{t \in (\alpha_m^\#, 1 - \alpha_m^\#)} \sqrt{N} \left| \psi(\mathbb{R}_{m,n}^\#(t)) - \psi(\mathbb{R}_{m,n}(t)) \right| \leq q \right\} \\ & - Pr \left\{ \sup_{t \in (\alpha_m^\#, 1 - \alpha_m^\#)} \sqrt{N} \left| \psi(\mathbb{R}_{m,n}(t)) - \psi(R(t)) \right| \leq q \right\} | \rightarrow 0, \end{aligned} \quad (4.45)$$

where Pr^* stands for bootstrap resampling probability conditional on samples.

Remark 4.3 Equivalently, a metric d (such as Levy's metric) which can characterize weak convergence can be also used to quantify the nearness of the distributions.

Corollary 4.2 Let $\phi(\cdot) : C[0, 1] \rightarrow \mathbb{R}$ be a bounded linear functional. Then conditionally on the samples, we have for all q

$$Pr^* \{ \sqrt{N} (\phi(\mathbb{R}_{m,n}^*) - \phi(\mathbb{R}_{m,n})) \leq q \} - Pr \{ \sqrt{N} (\phi(\mathbb{R}_{m,n}) - \phi(R)) \leq q \} \rightarrow 0 \quad a.s. \quad (4.46)$$

$$Pr^\# \{ \sqrt{N} (\phi(\mathbb{R}_{m,n}^\#) - \phi(\mathbb{R}_{m,n})) \leq q \} - Pr \{ \sqrt{N} (\phi(\mathbb{R}_{m,n}) - \phi(R)) \leq q \} \rightarrow 0 \quad a.s. \quad (4.47)$$

Corollary 4.3 (For partial AUC (pAUC)) Conditionally on the samples, for any $(e_1, e_2) \subset (0, 1)$, we have $\sqrt{N}(\hat{\mathbb{A}}(e_1, e_2) - A(e_1, e_2)) \rightarrow_d N(0, \sigma^2(e_1, e_2))$, where

$$\begin{aligned} \sigma^2(e_1, e_2) &= \lambda^{-1} \int_{e_1}^{e_2} \int_{e_1}^{e_2} R'(t)R'(s)(s \wedge t - st) dt ds \\ &+ (1 - \lambda)^{-1} \int_{e_1}^{e_2} \int_{e_1}^{e_2} (R(t) \wedge R(s) - R(t)R(s)) dt ds, \\ \hat{\mathbb{A}}(e_1, e_2) &= \int_{e_1}^{e_2} \mathbb{R}_{m,n}(t) dt, \quad A(e_1, e_2) = \int_{e_1}^{e_2} R(t) dt. \end{aligned}$$

Remark 4.4 (1) It can be easily shown that Condition B ensures that $\sigma^2(0, 1) < \infty$.

(2) By redefining the pAUC function by adding an indication function of the range t as

a factor, the asymptotic form for AUC ($e_1 = 0, e_2 = 1$) can be obtained from Corollary 4.2. (3) The same result holds for the BB resampling distributions conditional on the samples.

4.4 Simulation studies

4.4.1 Comparison of the frequentist and the BB resampled variability

To investigate the asymptotic equivalence of frequentist variability of the empirical estimate of AUC and resampled variability of the BB procedures, we simulate 6000 data sets of $X_1, \dots, X_m \sim \text{i.i.d. } F = N(0,1)$, $Y_1, \dots, Y_n \sim \text{i.i.d. } G = N(1.868, 1.5^2)$ to get 6000 frequentist samples of $\sqrt{N}(\hat{\mathbb{A}} - A)$, where $\hat{\mathbb{A}} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \{1(Y_j > X_i) + \frac{1}{2}1(Y_j = X_i)\}$, A is the true AUC value, $N = m + n$, $m = n = 50$ and $m = n = 300$, shown in Figure 4.1 (a) and (b) respectively; based on one simulated data set defined above, BB's one random realization of AUC denoted as $\mathbb{A}^\#$ can be obtained based on that BB's random realization of the ROC curve denoted as $\mathbb{R}^\#(t)$; hence, one BB resample of $\sqrt{N}(\mathbb{A}^\# - \hat{\mathbb{A}})$ can be generated, and then let the resample size be 6000. The density plot (Figure 4.1) of the asymptotic normal distribution $N(0, \sigma^2(0, 1))$, where σ can be calculated by Corollary 4.3, overlays those of $\sqrt{N}(\hat{\mathbb{A}} - A)$ and $(\sqrt{N}(\mathbb{A}^\# - \hat{\mathbb{A}}) \mid \text{data})$. The grid points on $[0,1]$ are chosen with equal interval length 0.001 to calculate AUC.

4.4.2 Comparison of the empirical and BB's estimates of ROC

The asymptotic property of the empirical and BB version of $\psi(R)$ is studied through simulation by examining $P(\psi(R) \leq q)$ and $P^\#(\psi(R^\#) \leq q \mid \text{data})$ where $\psi(\cdot) : C[0, 1] \rightarrow$

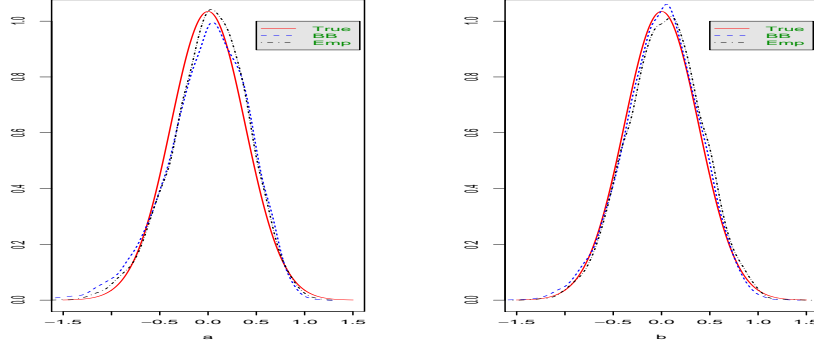


Figure 4.1: Comparison of density plots obtained by empirical $\sqrt{N}(\hat{\mathbb{A}} - A)$ based on 6000 simulated data sets, the BB's ($\sqrt{N}(\mathbb{A}^\# - \hat{\mathbb{A}})|\text{data}$) based on one simulated data set and corresponding 6000 resamples, and the asymptotic normal distribution $N(0, \sigma^2(0, 1))$ where σ can be calculated by Corollary 4.3.

\mathbb{R} is a bounded linear functional and the grid point q on $[0,1]$ is chosen with equal interval length 0.001. By plotting the mid 50% percentile of these quantities on the given grid points, its asymptotic property can be shown.

By simulating the data sets described above, with sample size $m = n = 50$ and $\psi = R(0.5)$, we obtain $P(\hat{\mathbb{R}}(0.5) \leq q)$ based on 1000 simulated data sets, and $P^\#(\mathbb{R}^\#(0.5) \leq q | \text{data})$ based on 1000 resamples. The grid points on $[0,1]$ are chosen with equal interval length 0.001 to calculate ROC. The plot $\hat{P}(\hat{\mathbb{R}}(0.5) \leq q)$ and mid 50% percentile of $P^\#(\mathbb{R}^\#(0.5) \leq q | \text{data})$ are given in Figure 4.2.

4.5 Proof of the Lemmas in the sense of *a.s.*

4.5.1 Proof of the Lemma 4.1

For any $y \in [\delta_n^*, 1 - \delta_n^*]$, let $k = \lceil ny \rceil$, ξ lies between y and $V_{k:n}$. Note that $|y - k/n| \leq 1/n$.

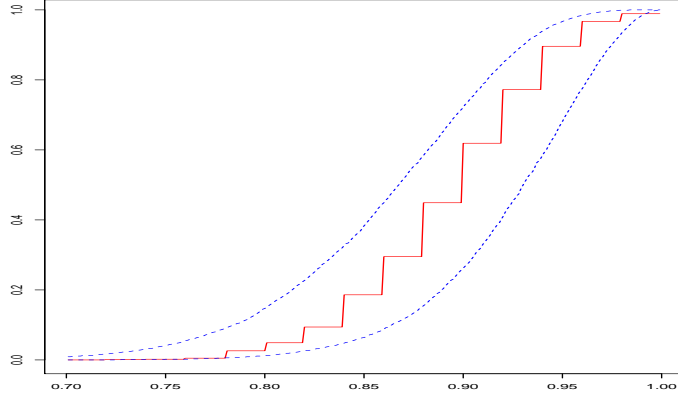


Figure 4.2: Comparison CDF of empirical (solid curve) and the BB's mid 50% percentile (dotted curves) of $R(0.5)$.

Proof of the statement (4.10) in the Lemma 4.1

1. For any $y \in [\delta_n, 1 - \delta_n]$, $V_{k:n} = y + n^{-1/2}\tilde{\mathbb{W}}_n(y)$, where $\tilde{\mathbb{W}}_n(\cdot)$ is the uniform quantile process for V_i 's, $i = 1, \dots, n$, then

$$|y - V_{k:n}| \leq 4n^{-1/2}(y(1-y))^{1/2}(\log \log n)^{1/2} \text{ by Theorem K of Appendix A. (4.48)}$$

2. When n is large, $(y(1-y))^{-1/2} \leq 2(\delta_n^*)^{-1/2}$. Then by (A.1), (A.3) of Theorem N of Appendix A and Condition A,

$$\begin{aligned} & \sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} \frac{\sqrt{y(1-y)} |f'(F^{-1}(\xi))|}{f(F^{-1}(\xi))f(F^{-1}(V_{k:n}))} \\ &= \sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} [(y(1-y))^{-1/2}] \cdot \frac{y(1-y)}{\xi(1-\xi)} \cdot \xi(1-\xi) \frac{|f'(F^{-1}(\xi))|}{f^2(F^{-1}(\xi))} \cdot \frac{f(F^{-1}(\xi))}{f(F^{-1}(V_{k:n}))} \\ &\leq_{a.s.} 2(2n^{-1/2}(\log \log n)^{1/2})^{-1/2} 5\gamma 10^\gamma =_{a.s.} O(n^{1/4}(\log \log n)^{-1/4}). \end{aligned} \tag{4.49}$$

3. δ_n^* ensures $V_{k:n} \in [\delta_n, 1 - \delta_n]$ *a.s.*. By Theorem D of Appendix A, we have

$$\begin{aligned} \sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} |f(F^{-1}(V_{k:n}))\mathbb{Q}_n(V_{k:n}) - \mathbb{W}_n(V_{k:n})| &\leq 40\gamma 10^\gamma n^{-1/2} \log \log n, \\ \sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} |\mathbb{W}_n(V_{k:n})| &\leq 2(\log \log n)^{1/2}. \end{aligned}$$

Hence

$$\sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} |f(F^{-1}(V_{k:n}))\mathbb{Q}_n(V_{k:n})| =_{a.s.} O((\log \log n)^{1/2}). \quad (4.50)$$

Combining results (4.48), (4.49) and (4.50), we now have

$$\begin{aligned} &\sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} \left| (y - V_{k:n})\mathbb{Q}_n(V_{k:n}) \frac{f'(F^{-1}(\xi))}{f(F^{-1}(\xi))} \right| \\ &\leq \sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} \frac{|y - V_{k:n}|}{\sqrt{y(1-y)}} \cdot \left\{ \sqrt{y(1-y)} \frac{|f'(F^{-1}(\xi))|}{f(F^{-1}(\xi))f(F^{-1}(V_{k:n}))} \right\} \\ &\quad \cdot |f(F^{-1}(V_{k:n}))\mathbb{Q}_n(V_{k:n})| =_{a.s.} O(n^{-1/4}(\log \log n)^{3/4}) \leq_{a.s.} O(l_n). \quad \square \end{aligned} \quad (4.51)$$

Proof of the statement (4.11) in the Lemma 4.1

By Condition A,

$$f(F^{-1}(y)) = f(F^{-1}(V_{k:n})) + (y - V_{k:n})f'(F^{-1}(\xi))/f(F^{-1}(\xi)), \quad (4.52)$$

where ξ lies between y and $V_{k:n}$. Note that $y \in [\delta_n^*, 1 - \delta_n^*]$ ensures that $V_{k:n} \in [\delta_n, 1 - \delta_n]$ *a.s.*. Let K be the Kiefer process approximation to \mathbb{Q}_n as in Theorem B of Appendix A.

Then we have

$$\begin{aligned} &\sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} |f(F^{-1}(y))\mathbb{Q}_n(V_{k:n}) - n^{-1/2}K(y, n)| \\ &\leq \sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} |f(F^{-1}(V_{k:n}))\mathbb{Q}_n(V_{k:n}) - n^{-1/2}K(V_{k:n}, n)| \\ &\quad + \sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} |(y - V_{k:n})f'(F^{-1}(\xi))/f(F^{-1}(\xi))\mathbb{Q}_n(V_{k:n})| \\ &\quad + \sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} |n^{-1/2}(K(V_{k:n}, n) - K(y, n))|. \end{aligned} \quad (4.53)$$

The first two terms of (4.53) are $O(l_n)$ *a.s.* respectively by Theorem B of Appendix A and (4.10). The third term of (4.53) can be bounded by $\sup_{\delta_n^* \leq y \leq 1 - \delta_n^*} \{ |n^{-1/2}K(V_{k:n}, n) - n^{-1/2}K(k/n, n)| + |n^{-1/2}K(k/n, n) - n^{-1/2}K(y, n)| \}$, which are $O(l_n)$, respectively, by Theorem E and Theorem F of Appendix A (let $h_n = n^{-1}$). \square

4.5.2 Proof of the Lemma 4.2

Proof of the statement (4.15) in Lemma 4.2

By Theorem M, we have

$$\begin{aligned}
& \sup_{0 < y < 1} \sqrt{n} \left| \mathbb{F}_n^{\#-1}(y) - \mathbb{F}_n^{-1}(\tilde{U}_{n-1}(y)) \right| \\
& \leq \sup_{0 < y < 1} \sqrt{n} \left| \mathbb{F}_n^{-1} \left(\frac{1 + (n-1)\tilde{U}_{n-1}(y)}{n} \right) - \mathbb{F}_n^{-1}(\tilde{U}_{n-1}(y)) \right| \\
& \quad + \sup_{0 < y < 1} \sqrt{n} \left| \mathbb{F}_n^{-1} \left(\frac{(n-1)\tilde{U}_{n-1}(y)}{n} \right) - \mathbb{F}_n^{-1}(\tilde{U}_{n-1}(y)) \right| \\
& \leq \sup_{0 \leq k \leq n} 2\sqrt{n} |X_{k+1:n} - X_{k:n}| =_{a.s.} O(n^{-1/2} \log n). \quad \square
\end{aligned}$$

Proof of the statement (4.16) in Lemma 4.2

Because $y \in [\delta_n^\#, 1 - \delta_n^\#] \subset [\epsilon_n, 1 - \epsilon_n]$, where $\epsilon_n = 0.236n^{-1} \log \log n$, let $k = \lceil ny \rceil$, ξ lies between y and $\tilde{U}_{n-1}(y)$. The following four steps are needed to finish the proof.

1. By similar argument of (4.51), we have

$$\begin{aligned}
& \sup_{\delta_n^\# \leq y \leq 1 - \delta_n^\#} \left| (\tilde{\mathbb{U}}_{n-1}(y) - y) \frac{f'(F^{-1}(\xi))}{f(F^{-1}(\xi))} \mathbb{Q}_n(\tilde{\mathbb{U}}_{n-1}(y)) \right| \\
& \leq \sup_{\delta_n^\# \leq y \leq 1 - \delta_n^\#} \left\{ \frac{|\tilde{\mathbb{U}}_{n-1}(y) - y|}{\sqrt{y(1-y)}} \right\} \left\{ \sqrt{y(1-y)} \frac{|f'(F^{-1}(\xi))|}{f(F^{-1}(\xi))f(F^{-1}(\tilde{\mathbb{U}}_{n-1}(y)))} \right\} \\
& \quad \cdot |f(F^{-1}(\tilde{\mathbb{U}}_{n-1}(y)))\mathbb{Q}_n(\tilde{\mathbb{U}}_{n-1}(y))| \\
& =_{a.s.} O(n^{-1/4}(\log \log n)^{3/4}) \leq_{a.s.} O(l_n).
\end{aligned}$$

2. By Theorem B of Appendix A, $|\sqrt{n}f(F^{-1}(y))(\mathbb{F}_n^{-1}(y) - F^{-1}(y)) - n^{-1/2}K(y, n)| =_{a.s.} O(l_n)$.

3. We have

$$\begin{aligned}
& \sup_{\delta_n^\# \leq y \leq 1 - \delta_n^\#} |\sqrt{n}f(F^{-1}(y))[F^{-1}(\tilde{\mathbb{U}}_{n-1}(y)) - F^{-1}(y)] - n^{-1/2}K(y, n)| \\
& \leq \sup_{\delta_n^\# \leq y \leq 1 - \delta_n^\#} |\sqrt{n}f(F^{-1}(y))(\tilde{\mathbb{U}}_{n-1}(y) - y)/f(F^{-1}(y)) - n^{-1/2}K(y, n)| \\
& \quad + \sup_{\delta_n^\# \leq y \leq 1 - \delta_n^\#} \left| (\tilde{\mathbb{U}}_{n-1}(y) - y)^2 \frac{f'(F^{-1}(\xi))}{f^2(F^{-1}(\xi))} \frac{f(F^{-1}(y))}{f(F^{-1}(\xi))} \right| \\
& \leq \sup_{\delta_n^\# \leq y \leq 1 - \delta_n^\#} |\sqrt{n}(\tilde{\mathbb{U}}_{n-1}(y) - y) - n^{-1/2}K(y, n)| \\
& \quad + \sup_{\delta_n^\# \leq y \leq 1 - \delta_n^\#} \left| (\tilde{\mathbb{U}}_{n-1}(y) - y)^2 \frac{f'(F^{-1}(\xi))}{f^2(F^{-1}(\xi))} \frac{f(F^{-1}(y))}{f(F^{-1}(\xi))} \right| \tag{4.54} \\
& =_{a.s.} O(l_n).
\end{aligned}$$

The reasons for (4.54) are as follows:

(a) By Theorem A of Appendix A, we have

$$\sup_{0 < y < 1} \sqrt{\frac{n}{n-1}} \left| \sqrt{n-1}(\tilde{\mathbb{U}}_{n-1}(y) - y) - \frac{K(y, n-1)}{\sqrt{n-1}} \right| =_{a.s.} O\left(\frac{\sqrt{n} \log^2(n-1)}{n-1}\right).$$

By Theorems G and H of Appendix A, we have

$$\begin{aligned} & \sup_{0 < y < 1} \left| \frac{\sqrt{n}}{n-1} K(y, n-1) - n^{-1/2} K(y, n) \right| \\ & \leq \sup_{0 < y < 1} \left\{ \frac{\sqrt{n}}{n-1} |K(y, n-1) - K(y, n)| + \left| \left(\frac{\sqrt{n}}{n-1} - \frac{1}{\sqrt{n}} \right) K(y, n) \right| \right\} \\ & =_{a.s.} O \left(\sqrt{\frac{n}{n-1}} l_n + \frac{(\log \log n)^{\frac{1}{2}}}{n-1} \right) =_{a.s.} O(l_n) \end{aligned}$$

Therefore, the first term of (4.54) can be majorized by

$$\begin{aligned} & \sup_{0 < y < 1} \sqrt{\frac{n}{n-1}} \left| \sqrt{n-1} (\tilde{U}_{n-1}(y) - y) - \frac{1}{\sqrt{n-1}} K(y, n-1) \right| \\ & + \sup_{0 < y < 1} \left| \frac{\sqrt{n}}{n-1} K(y, n-1) - n^{-1/2} K(y, n) \right| =_{a.s.} O(l_n). \end{aligned}$$

- (b) By Theorem J of Appendix A, we have $(\tilde{U}_{n-1}(y) - y)^2 \leq_{a.s.} 4(n-1)^{-1} y(1-y) \log \log(n-1)$. By following a similar argument to that of Theorem 3 [Csörgő and Révész, 1978], we get $\frac{f(F^{-1}(y))}{f(F^{-1}(\xi))} \leq 10^\gamma$. Also by Condition A, the second term of (4.54) can be majorized by

$$\begin{aligned} & \sup_{\delta_n^\# \leq y \leq 1 - \delta_n^\#} [4(n-1)^{-1} \log \log(n-1)] \left\{ \frac{y(1-y)}{\xi(1-\xi)} \right\} \left\{ \xi(1-\xi) \frac{f'(F^{-1}(\xi))}{f^2(F^{-1}(\xi))} \right\} \\ & \left\{ \frac{f(F^{-1}(y))}{f(F^{-1}(\xi))} \right\} \leq_{a.s.} O(l_n). \end{aligned}$$

4. By Theorems G and I of Appendix A, we have

$$\begin{aligned} & \sup_{0 < y < 1} \frac{1}{\sqrt{n-1}} |K(\tilde{U}_{n-1}(y), n) - K(y, n)| \\ & \leq \sup_{0 < y < 1} \frac{1}{\sqrt{n-1}} |K(\tilde{U}_{n-1}(y), n) - K(\tilde{U}_{n-1}(y), n-1)| \\ & + \sup_{0 < y < 1} \frac{1}{\sqrt{n-1}} |K(y, n) - K(y, n-1)| \\ & + \sup_{0 < y < 1} \frac{1}{\sqrt{n-1}} |K(\tilde{U}_{n-1}(y), n-1) - K(y, n-1)| =_{a.s.} O(l_n). \end{aligned}$$

Notice that $\mathbb{F}_n^{-1}(\tilde{\mathbb{U}}_{n-1}(y)) - \mathbb{F}_n^{-1}(y)$ can be split as

$$[\mathbb{F}_n^{-1}(\tilde{\mathbb{U}}_{n-1}(y)) - F^{-1}(\tilde{\mathbb{U}}_{n-1}(y))] - [\mathbb{F}_n^{-1}(y) - F^{-1}(y)] + [F^{-1}(\tilde{\mathbb{U}}_{n-1}(y)) - F^{-1}(y)]. \quad (4.55)$$

By combining parts (1)- (4) above and (4.55), we finish the proof of (4.16). \square

Chapter 5

ROC analysis without the gold standard

The definition of the ROC curve implies that its underlying assumption is the gold standard for the truth, which means that the true disease status of each patient is known by the most accurate (gold standard) diagnostic test. However, in clinical practice, the gold standard test is not always available for various reasons. When the gold standard exists, due to ethical considerations and because the gold standard test could be invasive and expensive, a patient will be administered the gold standard test only when there is some indication of the presence of the disease through screening tests. For example, the gold standard test for detecting middle ear problems is called visual reinforcement audiometry (VRA) behavioral test, and the commonly used screening test is measured by distortion product otoacoustic emissions (DPOAE). If the DPOAE test of a patient appears abnormal, he (or she) will be scheduled to proceed to a VRA test. A culture of the organism and biopsy of a sample of tissues are usually required to obtain the gold standard for cancer. For some diseases, the gold standard test does not even exist for a

particular time or any time point. For example, the true disease status of a Alzheimer patient can be determined by neuropathological examination after the patient dies; the true disease status of myocardial infarction (MI) is never clearly distinguishable to all patients. Implementation of the ROC curve estimation without the gold standard seems challenging.

Some related statistical issues can be described as follows.

Verification bias occurs when we estimate the sensitivity and specificity of the diagnostic test based on those patients with verified disease status. The commonly used assumption is verification missing at random (MAR), which means that the prevalence within the verified disease status group is equal to the prevalence among the whole group (see Pepe (2003, page 170) for more details). When the screening test, denoted as T , is binary, and under the verification MAR assumption, Begg and Greenes (1983) propose an adjusted estimate of TPF and FPF in terms of predictive values PPV and NPV (see definitions in Glossary). Other more efficient methods are discussed by Reilly and Pepe (1995), Clayton *et al.* (1998). The maximum likelihood estimators of sensitivity and specificity are studied by Zhou (1993, 1994) in a more general setting with and without MAR assumption. If the test variable is ordinal, its verification bias in ROC curves is studied by Gray *et al.* (1984), Hunink *et al.* (1993), Rodenberg and Zhou (2000); see the review by Hui and Zhou (1998) for more details.

When the gold standard is not available, the best available reference test is called the imperfect gold standard test, which is used to compare with other new tests. A commonly used assumption to obtain the error rates of the new test is conditional independence between the tests, which is defined as $P(T_1 \text{ and } T_2 | D) = P(T_1 | D)P(T_2 | D)$, where T_i stands for the i^{th} test, D stands for the true disease status; see the reviews by Walter and Irwig (1988), and Hui and Zhou (1998), and the work by Gart and Buck (1966),

Mantel (1951), and Staquet *et al.* (1981). However, in practice, tests conditional on the true disease status may tend to be positively dependent (see initial discussion by Vacek (1985)). There are several approaches to solve this problem, such as: (i) developing a design to allow more degrees of freedom to estimate the parameters in the conditional dependence model, (ii) assuming that a subset of parameters is known, (iii) adopting a Bayesian approach by putting a strong prior on the prevalence of disease. See Joseph *et al.* (1995) for one imperfect test and Dendukuri and Joseph (2001) for two imperfect tests cases. However, controversy arises due to modeling mechanisms based on latent class analysis (see Pepe, 2003, page 203).

Most of the existing statistical methods deal with the binary diagnostic variables and assume the conditional independence of the diagnostic tests. A few methods study the ordinal or continuous diagnostic tests to obtain the ROC curve estimate. Alonzo (2000) proposes several verification bias correction methods for the continuous tests when covariate data is available to all patients. Hall and Zhou (2003) model continuous diagnostic variables under the conditional independence assumption without the gold standard as a mixture of two distributions. Zhou *et al.* (2005) propose a nonparametric maximum likelihood method to estimate the ROC curve when the diagnostic tests are ordinal and the number of tests is greater than two. Little discussion can be found in the literature dealing with only one continuous test variable and modeling the diagnostic variable up to some strictly increasing transformation with a certain parametric form. We prefer to deal with one available diagnostic test rather than to assume that several tests are available to apply to the same population for identification purpose. Herein, we aim at deriving a ROC curve estimator when there exists only one continuous test variable of one population, and we assume that an accurate proportion of the diseased population among the confounded population is known. This new task can be done by making more specific

reasonable assumptions, such as the binormality assumption without the gold standard. Our motivation is based on observing that the implication of binormality uncovers the underlying distributional relationship between the non-diseased and diseased populations up to some strictly increasing transformation H conditional on the true disease status D . Hence, we may model the diagnostic variable S without the gold standard transformed by H as a mixture of two normal distributions, where one normal component stands exactly for the non-diseased population and the other for the diseased population. The ROC curve can be defined in the usual way and an explicit expression is available because of the normality of the components and the assumption of D . In order to estimate the parameters in the resultant mixture normal model, direct and indirect approaches to estimate H are proposed. Monte Carlo Markov Chain sampling techniques, using WinBUGS software, make the computation feasible.

5.1 Binormal model assumption without the gold standard

Motivated by the binormality assumption, we model the confounded population's diagnostic variable S after transformation H and introduce the disease status as a latent variable D to satisfy the binormality assumption. The proposed binormal model

assumption without the gold standard can be described as follows:

$$\begin{aligned}
 H(S) &\sim (1 - p)N(0, 1) + pN(\mu, \sigma^2), \\
 \text{exist } D &\sim \text{Bernoulli}(p), \text{ where } p \in (0, 1) \text{ is the prevalence of the disease,} \\
 \text{such that } (H(S)|D = 0) &\sim N(0, 1), \quad (H(S)|D = 1) \sim N(\mu, \sigma^2), \\
 (S|D = 0) &\sim F, \quad (S|D = 1) \sim G,
 \end{aligned} \tag{5.1}$$

where F and G are unknown.

5.2 Proposed models

In order to estimate the parameters in model (5.1), two related methods will be proposed by employing a Bayesian approach. The Bernstein polynomial method aims at directly modeling the strictly increasing transformation H . The latent model, which is a Bayesian curve estimation method under binormality using a partial likelihood based on ranks without the gold standard, is also introduced.

5.2.1 Bernstein polynomial model

The advantage of using Bernstein polynomial model is that we can produce a strictly increasing transformation in model (5.1). Due to the invariance property of the ROC curve under common increasing transformations, and for simplicity, we also assume that the diagnostic variable \mathbf{S} has the range of (0,1). The likelihood is given by

$$L(\mu, \sigma | D, H, \mathbf{S}) = \prod_{i: \delta_i=0} \left\{ \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(H(S_i))^2 |J_i|} \right\} \times \prod_{i: \delta_i=1} \left\{ \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(H(S_i)-\mu)^2 |J_i|} \right\}. \tag{5.2}$$

Under the Bernstein polynomial model, the transformation H bringing in normality is given by

$$H(x) = g \left\{ \sum_{j=1}^k h_j \binom{k}{j} x^j (1-x)^{k-j} \right\},$$

where $g : (0, 1) \rightarrow \mathbb{R}$ is some known strictly increasing function. In order to ensure that the transformed random variable is distributed as standard normal conditional on labeling F , we will work with $H(x) = \sum_{j=1}^k h_j \binom{k}{j} x^j (1-x)^{k-j} - 3$, where k is a fixed reasonable integer, h_i 's > 0 , D is an abbreviation of the disease status of each observation, $\mathbf{S} = (S_1, S_2, \dots, S_N)$ are the observations, and the Jacobian evaluated at observation S_i is denoted as $J_i = \frac{dH(x)}{dx} \Big|_{x=S_i}$, $i = 1, \dots, N$, labeling S_i as $\delta_i = 0$ if it is from F , $\delta_i = 1$ from G . This seems to be reasonable strategy since -3 is the effective minimum of the range of a standard normal variable and the other normal has positive mean.

We incorporate that Bayesian approach by putting priors on the parameters in model (5.2) as follows:

$$h_1, h_2 - h_1, \dots, h_k - h_{k-1} \sim \text{i.i.d. } \exp(\lambda),$$

$$\mu \sim N(\mu_0, \sigma_0^2), \quad \mu > 0,$$

$$\pi(\sigma) \propto 1/\sigma, \quad \sigma \in (e_1, e_2),$$

$$\delta_i \sim \text{Bernoulli}(p), \quad p \sim \text{Beta}(1 - p_0, p_0),$$

where $k, \lambda, \mu_0, \sigma_0, e_1, e_2, p_0$ are known constants. Hence, Bayes estimators of μ and σ can be obtained through the following Gibbs sampling technique.

1. Initialize the parameters: $\mu, \sigma, h_1, h_2, \dots, h_k, \delta_1, \delta_2, \dots, \delta_N, p$;

2. Start iterations of updating the parameters. One iteration is given by

$(p, \mu, \sigma) | \text{rest} \sim$ Mixture normal problem,

$\delta_i | \text{rest} \sim$ Bernoulli with mean equal to

$$P(\delta_i = 1) = \frac{p\phi_{\mu,\sigma}(H(S_i))}{(1-p)\phi(H(S_i)) + p\phi_{\mu,\sigma}(H(S_i))}, \quad i = 1, \dots, N,$$

$h_1, \dots, h_k | \text{rest}$ can be sampled by Metropolis-Hasting algorithm.

3. After burn-in, Bayes estimates of μ and σ can be calculated by the corresponding posterior means.

5.2.2 Latent model

In Chapter 2, we proposed a Bayesian rank-based, partial likelihood estimator of the ROC curve using the assumption of binormality. The basic idea is to define a set \mathcal{D}_{obs} (see definition (2.10)) invariant under some strictly increasing transformation H . Hence, the posterior distribution of $\pi(\mu, \sigma | \text{ranks and labels of the observations})$ can be sampled by a data augmentation technique. The transformed diagnostic variables are the augmented variables. However, when the gold standard is not available, by imputing the label of each observation through the Bayesian technique, we can convert the problem without the gold standard to the one with the gold standard. This latent model allows us to estimate the parameters in a binormal model μ, σ by treating H as latent rather than directly modeling it. Computation can proceed as follows:

Given observations S_1, \dots, S_N , denote T_1, \dots, T_N as $T_i = H(S_i)$, preserving the same ranks comparing $OS(S_1, \dots, S_N)$ with $OS(T_1, \dots, T_N)$, and the same labels denoted as $(\delta_1, \dots, \delta_N)$ of both (S_1, \dots, S_N) and (T_1, \dots, T_N) , where OS denotes the order statistics, and δ_i 's follow the same definition in model (5.2). In order to sample from the posterior

distribution of

$$(\mu, \sigma | \text{rank of } \mathbf{S}), \quad (5.3)$$

a data augmentation technique is used and the corresponding augmentation variables are labels $(\delta_1, \dots, \delta_N)$, and subsequently (T_1, \dots, T_N) .

The Bayesian approach is implemented to estimate the parameters of interest in the model (5.3). Prior choices are given as follows:

$$\begin{aligned} \delta_i &\sim \text{Bernoulli}(p), \quad p \sim \text{Normal}(p_0, 0.01^2), \\ \mu &\sim \text{N}(\mu_0, \sigma_0^2), \quad \mu > 0, \\ \pi(\sigma) &\propto 1/\sigma, \quad \sigma \in (e_1, e_2), \end{aligned}$$

where $\mu_0, \sigma_0, e_1, e_2, p_0$ are known constants. Therefore, the Bayes estimators of μ and σ can be obtained through the following Gibbs sampling technique:

1. Initialize the parameters: $\mu, \sigma, \delta_1, \dots, \delta_N, p$, and one data set (T_1, \dots, T_N) satisfying $\{(T_1, \dots, T_N) \in \mathcal{D}_{\text{obs}} | \delta_1, \dots, \delta_N\}$.
2. Start iterations of updating the parameters. One iteration is given by

$(T_1, \dots, T_N) | \text{rest}$ —converting the problem to the binormality case;

$(p, \mu, \sigma) | \text{rest}$ —Mixture normal problem;

$\delta_i | \text{rest} \sim \text{Bernoulli}$ with mean equal to

$$P(\delta_i = 1) = \frac{p\phi_{\mu, \sigma}(T_i)}{(1-p)\phi(T_i) + p\phi_{\mu, \sigma}(T_i)}, \quad i = 1, \dots, N.$$

3. After burn-in, the Bayes estimates of μ and σ can be calculated by the corresponding posterior means.

Remark 5.1

- *Though simulation studies are not shown here, we find that the posterior distribution obtained by using the latent model is quite sensitive to the prior choice of the prevalence. A non-informative prior may cause the posterior to be diffuse. In practice, it makes sense to put strong prior when the related information is accessible.*
- *Because sufficiently large numbers of iterations of MCMC are required for convergence, the method is computationally very intensive. Nevertheless, the latent model provides an invariant estimator, which is especially useful when the sample size is small.*

5.3 Simulation studies

We have proposed two related curve estimation methods when the gold standard is unobserved, only one test is available to each patient, and precise prevalence of disease is known. We now compare the performance among our proposed methods and the method when the gold standard with measurement error is available, with several simulation studies.

Because the area under the ROC curve (AUC) is a well known statistic used to compare different modalities in ROC analysis, two paired distributions of F and G based on two different AUC values .85 and .95 with total sample size 50 are considered. The purpose for our simulation studies is to examine the curve fitting using our proposed methods and the BPL method labeling by the gold standard with measurement error, i.e., the BPL method with surrogate labels denoted as BPLS. Bias, mean square errors (MSE) and MCMC errors of the estimates from simulation studies using our proposed methods and the BPLS method are shown in Table 5.1. Coverage probabilities of 90% credible interval for AUC and corresponding average lengths, and L_1 error of the ROC

curve denoted as $IAE = \int_0^1 |\hat{R}(t) - R(t)| dt$ are compared in Table 5.2, where $R(t)$ is the true ROC curve function, and $\hat{R}(t)$ is one of the curve estimates we used in the simulation. Before simulation, the following quantities are prespecified.

1. Bernstein polynomial model: choose constants $k = 4$, $\lambda = 3$, $\mu_0 = 2$ (or $\mu_0 = 3$) for AUC=0.85 (or AUC=0.95), $\sigma_0 = 1$, $(e_1, e_2) = (1, 2.5)$, $p_0 = 0.5$; the initial values of the parameters will be generated by WinBUGS.
2. Latent model: choose constants $\mu_0 = 2$ (or $\mu_0 = 3$) for AUC=0.85 (or AUC=0.95), $\sigma_0 = 1$, $(e_1, e_2) = (1, 2.5)$, $p_0 = 0.5$; initial values of μ and σ will be chosen to be 3 and 2, respectively.
3. BPLS: assume the following error rates $P(D = 1|X) = P(D = 0|Y) = 0.05$, $P(D = 0|X) = P(D = 1|Y) = 0.95$, generate the labels based on the given error rates and regard these surrogate labels as the true labels.

Simulations are based on 100 simulated data sets. The label of each observation is generated by its prior distribution, then based on a different choice of AUC values, a diagnostic test of each observation is generated from either $F = N(0, 1)$ or $G = N(5^{1/2}\Phi^{-1}(AUC), 2^2)$. We select 3000 Gibbs samples with 6000 iterations burn-in at 3000 for the latent model, and 95000 Gibbs samples with 100000 iterations burn-in at 5000 for both the Bernstein polynomial model and the BPLS method. Although we can have a closed form of AUC based on the binormal parameters, we will use Simpson's method to calculate AUC values, where the grid points on $[0,1]$ are chosen to be 0.01, 0.02, \dots , 0.99, because the main interest of the simulation is to examine the curve fitting (using index IAE). See Appendix C for WinBUGS model specification code for the Bernstein polynomial model and the latent model, also Matlab code for the BPL method.

Based on the simulation results shown in Table 5.1 and 5.2, we can conclude that even though the rate of the misspecified labels is only 5%, estimates of intercepts and slopes obtained by the BPLS method have the largest MSE, and from the results of coverage probabilities of AUC for 90% credible intervals, the BPLS method has the lowest coverage. Although when AUC is equal to 0.95, the estimate of intercept obtained by the latent model has the largest bias compared to other methods, it has the smallest MSE. The rest of the simulation results show that the latent model produces the best estimates of the binormal parameters and the ROC curves when precise prevalence of disease is known. Although for both AUC equal to 0.85 and 0.95, the coverage probabilities of AUC for 90% credible intervals obtained by the Bernstein polynomial model are pretty good, the estimates of the corresponding intercept have much larger MCMC errors than the other two methods. In terms of the computing time, the Bernstein polynomial model gives estimates much faster than the latent model without the gold standard.

Table 5.1: Estimates of binormal ROC's parameters a and b without the gold standard. Simulations are based on the Bernstein polynomial and the latent model abbreviated as Poly and Lat in the table, respectively. All the estimates are based on 100 simulated data sets ($X \sim N(0,1)$, $Y \sim N(5^{1/2}\Phi^{-1}(AUC), 2^2)$, total sample size is equal to 50), and corresponding 3000 Gibbs samples (iteration 6000 and burn-in at 3000) for the latent model and 95000 Gibbs samples (iteration 100000 and burn-in at 5000) for both the Bernstein polynomial model and the BPLS method.

AUC		Bias			MSE			MCMC error		
		Poly	Lat	BPLS	Poly	Lat	BPLS	Poly	Lat	BPLS
.85	a	.1031	.0271	-.0296	.1176	.0017	.1413	.6726	.1808	.3300
	b	.1727	.1274	.1757	.0343	.0162	.1179	.1623	.0750	.2029
.95	a	-.3338	-.1495	-.0254	.3942	.0239	1.1570	.7304	.2363	.6089
	b	.1355	.1207	.3338	.0258	.0147	.5624	.1598	.0628	.3735

Remark 5.2 *Due to the limited simulations, our conclusion may have to be interpreted*

Table 5.2: Coverage probabilities of AUC, related mean lengths of the 90% CI and IAE. The same simulation setting in Table 5.1 is used in this table. The grid points on $[0,1]$ are chosen with equal interval length 0.01.

AUC	Coverage			Mean length			IAE		
	Poly	Lat	BPLS	Poly	Lat	BPLS	Poly	Lat	BPLS
.85	.89	1	.84	.3556	.1097	.1998	.0621	.0225	.0648
.95	.94	1	.60	.3060	.0823	.1507	.0804	.0258	.0581

cautiously. However, it cautions practitioners against the use of surrogate labels.

5.4 Real data analysis

We use the same data set used in the previous chapters published by Wieand *et al.* (1989) to illustrate the ROC curve fitting without the gold standard for the truth. The data contain 51 patients in the control group diagnosed with pancreatitis and 90 patients in the case group diagnosed with pancreatic cancer by two biomarkers: a cancer antigen (CA 125) and a carbohydrate antigen (CA 19-9). Our purpose is to examine which method can produce better ROC curve estimate: the Bernstein polynomial model without the gold standard, BPL using surrogate labels denoted as BPLS, or the latent model without the gold standard. As shown in Chapter 2, the performance of BPL estimator is better than its alternatives; here, we will treat the BPL (or BB) estimate based on the true labels as the true ROC curve; the binormal estimates obtained by the BPL method, the Bernstein polynomial model and the BPLS method are based on 95000 Gibbs samples with 100000 iterations burn-in at 5000; that obtained by the latent model is based on 3000 Gibbs samples with 6000 iterations burn-in at 3000; the BB curve estimate is based on 1000 resamples; the grid points on $[0,1]$ are chosen with equal

interval length 0.01.

Recall that BPLS method introduced in the simulation study requires the gold standard and measurement error. In this application, the BPLS curve estimate has different settings from those in the simulation study: (i) surrogate labels are based on the available alternative biomarker, which is not necessarily the gold standard test; (ii) measurement error of the labels can be calculated after we obtain the surrogate labels. In order to get the surrogate labels for one biomarker, information on the alternative biomarker and corresponding threshold value based on p_0 are used. For example, when we evaluate the accuracy of diagnostic test using biomarker CA 125 without the labels, surrogate labels will be determined by biomarker CA 19-9. In this case, the accurate proportion of disease cases within this confounded group is known as $p_0 = 63.83\%$; hence, a patient is determined to have a positive disease status if his CA 19-9 test value is greater than the 36.17% quantile of all the CA 19-9 test values. Overall, we can get the measurement error rates for CA 125 as $\hat{P}(D = 1|X) = 21/51$ and $\hat{P}(D = 0|Y) = 21/90$, and for CA 19-9 as $\hat{P}(D = 1|X) = 17/51$ and $\hat{P}(D = 0|Y) = 17/90$.

From Figure 5.1 and 5.2, we can see that our BB and BPL curve estimates with the gold standard are very close to each other. This is further evidence that both biomarkers satisfy the binormality assumption (see Section 3.4). In general, the curve estimates by the latent model give the most accurate estimation compared to those by the Bernstein polynomial model and the BPLS method. For biomarker CA 19-9, the curve estimate by the latent model performs best compared to those by the Bernstein polynomial and the BPLS methods. For biomarker CA 125, the BPLS curve estimate performs better when t is within $(0,0.2)$ than the latent model, because biomarker CA 19-9 has more accurate diagnostic capability than biomarker CA 125. The estimates by the Bernstein polynomial have the largest bias. From the calculation not shown here, we find that the

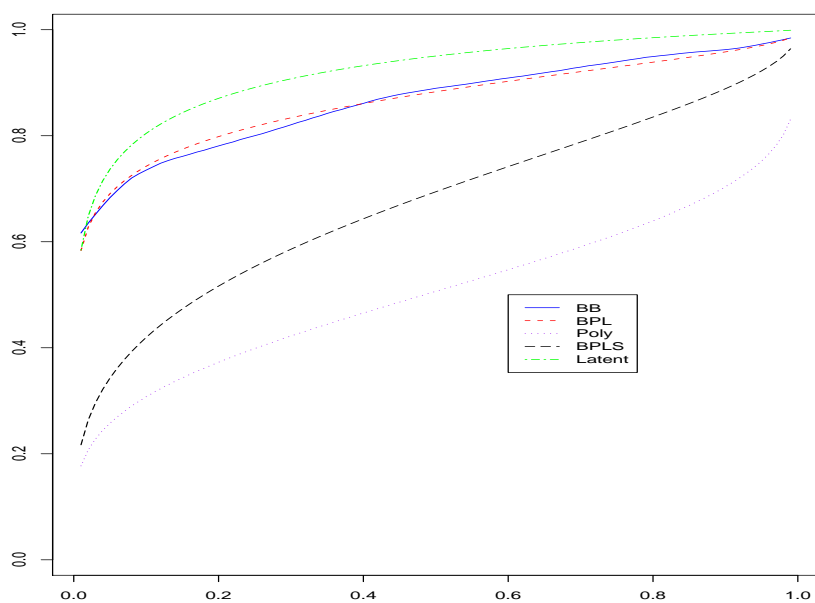


Figure 5.1: CA 19-9: Comparison of ROC curves using the BB and BPL methods (with the gold standard, respectively), the Bernstein polynomial (abbreviated as Poly) and the latent methods (without the gold standard) and BPLS with surrogate labels obtained by biomarker CA 125.

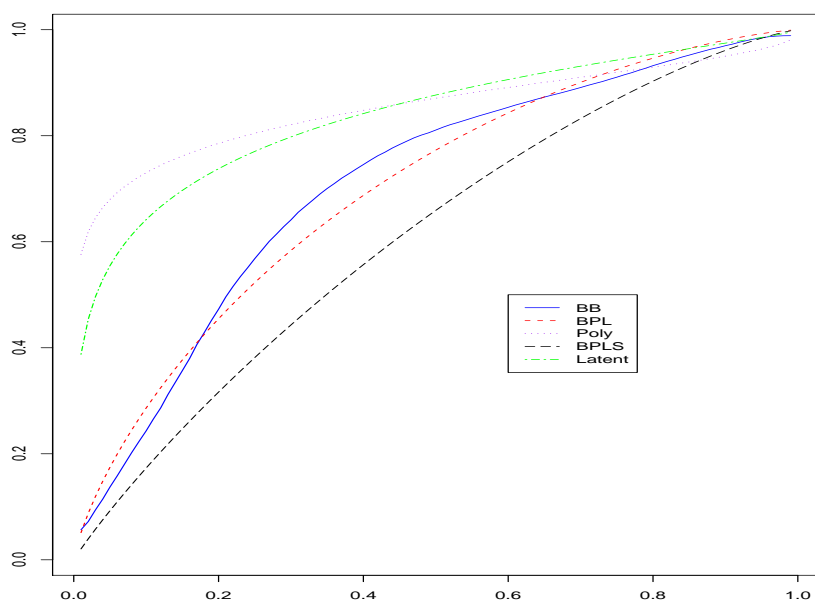


Figure 5.2: CA 125: Comparison of ROC curves using the BB and BPL methods (with the gold standard, respectively), the Bernstein polynomial (abbreviated as Poly) and the latent methods (without the gold standard) and BPLS with surrogate labels obtained by biomarker CA 19-9.

estimates by the Bernstein polynomial for both biomarkers are sensitive to the choice of the constants in the prior. Here, we chose the constants $\mu_0 = 1.5$ and $\mu_0 = 3$ for biomarker CA 125 and CA 19-9, respectively. More accurate hyperparameters in the priors can be used in our future studies.

Chapter 6

Discussion and Future work

We have proposed an estimation method for the ROC curve and its functionals based on the Bayesian bootstrap technique. One of the appealing features of our methodology is that it readily produces standard errors, confidence intervals and confidence bands for the entire ROC curve as well as some associated summary measures. The BB method closely resembles a non-parametric Bayesian analysis using Dirichlet process priors, but substantial simplification is obtained by avoiding the inversion of \bar{F} through the use of the placement variables. This simplification is possible because the BB corresponds to a “non-informative” posterior for which the prior base measure in a Dirichlet process has been chosen to be the null measure, and generating samples from the posterior reduces to finite dimensional random variate generations.

The Bayesian interpretation of our approach means that the method is not based on large sample techniques, and in principle applies to any sample size. However, when the sample size becomes larger, say, more than thousands, in order to get the better results, the resample size should be increased accordingly.

Based on the binormality assumption, choosing the initial parameter values of μ and

σ becomes relatively easier, i.e., the initial value of μ can be chosen among (1,3), because small values of μ indicate that the diagnostic test is non-informative, for X and Y are not well separated, while large values of μ indicate that X and Y are not even overlapped, which is not the case in ROC analysis. The initial value of σ can be chosen based on the density plots of the given diagnostic variables from X and Y . For example, if they have the same shape, then we can choose $b = 1$. Similarly, the density plot of the diagnostic variables without the gold standard for the truth will help us to determine the constants in our proposed models.

In Chapter 5, our first try, not shown here, is to model the transformation H as a smooth function. The B-spline method is one of the choices. Although it does not necessarily satisfy monotonicity, it can provide useful information. But, it seems difficult to implement MCMC algorithm using WinBUGS. One goal of future study is to develop an algorithm to model H using B-splines under the restriction of a strictly increasing function.

If $F = N(0, \sigma^2)$, $G = N(\delta, \sigma^2)$, then $a = \delta/\sigma$ is also equal to the square root of the Mahalanobis distance between F and G (Lloyd, 1998). Motivated by this result, we shall further investigate indices of distance between F and G . Some metrics are invariant under transformation, such as the L_1 distance or the Kullback-Leilber divergence between two density functions f and g . Based on some metric like that, we can develop some indices that will be useful to identify the invariant property of ROC curve.

More future work will be focused on developing ROC curve estimation methods with covariate effects, clustered data sets and correcting the verification bias without the gold standard for the continuous diagnostic tests.

Bibliography

- [1] Alonzo, T. A. (2000). Assessing accuracy of a continuous medical diagnostic or screening test in the presence of verification bias. *University of Washington: Seattle*.
- [2] Alonzo, T. A. and Pepe, M. S. (2002). Distribution-free ROC analysis using binary regression techniques. *Biostatistics* **3**, 421-432.
- [3] Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating graph. *Journal of Mathematical Psychology* . **12**, 387–415.
- [4] Begg, C. B. and Greenes, R. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* **39**, 207–215.
- [5] Begg, C. B. (1991). Advances in statistical methodology for diagnostic medicine in the 1980's. *Statistics in Medicine* **10**, 1887–1895.
- [6] Bozdogan, H. and Ramirez, D. E. (1986). Testing for model fit: Assessing the Box-Cox transformation of multivariate data to near normality. *Computational statistics. Q.* **3**), 127–150.
- [7] Cai, T. and Moskowitz, C. (2004). Semiparametric estimation of the binormal ROC curve. *Biostatistics* **5**(4), 573–586.

- [8] Campbell, G. (1994). Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Statistics in Medicine* **13**, 499–508.
- [9] Clayton, D., Spiegelhalter, D., Dunn, G. and Pickles, A. (1998). Analysis of longitudinal binary data from multiphase sampling. *Journal of the Royal Statistical Society, Series B* **60**, 71–87.
- [10] Csörgő, M. and Révész, P. (1978). Strong approximations of the quantile process. *The Annals of Statistics* **6**, 882–894.
- [11] Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic press, New York.
- [12] De Boor, C. (1978) *A practical guide to splines*. Springer, Berlin.
- [12] DeLong E. R., DeLong D. M. and Clarke-Pearson D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. *Biometrics* **44**, 837–845.
- [13] Dendukuri, N. and Joseph, L. (2001). Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. *Biometrics* **57**, 158–167.
- [14] Dorfman, D. D. and Alf, Jr., E. (1968). Maximum likelihood estimation of parameters of signal-detection theory—A direct solution. *Psychometrika* **33**, 117–124.
- [15] Dorfman, D. D., Berbaum, K. S. and Metz, C. E., Lenth, R. V., Hanley, J. A. and Dagga, H. A. (1997). Proper receiver operating characteristic analysis: The bigamma model. *Academic Radiology* **4**, 138–149.
- [16] Efron, B. (1979). Bootstrap methods: another look at the Jackknife. *The Annals of Statistics* **7**, 1–26.

- [17] Gart, J. J. and Buck, A. A. (1966). Comparison of a screening test and a reference test in epidemiologic studies II: A probabilistic model for the comparison of diagnostic tests. *American Journal of Epidemiology* **83**, 593–602.
- [18] Ghosal, S. and Van der Vaart, A. W. (2008). *Theory of Nonparametric Bayesian Inference*. Cambridge University Press. (To be published)
- [19] Ghosh, J. K. (1969). Only linear transformations preserve normality. *Sankhya Series A* **31**, 309–312.
- [20] Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. Springer-Verlag, New York.
- [21] Goddard, M. J. and Hinberg, I. (1990). Receiver operating characteristic (ROC) curves and non-normal data: An empirical study. *Statistics in Medicine* **9**, 325–337.
- [22] Gray, R., Begg, C. B., Greenes, R. A. (1984). Construction of receiver operating characteristic curves when disease verification is subject to selection bias. *Medical Decision Making* **4**, 151-164.
- [23] Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. John Wiley & Sons, New York.
- [24] Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*. Academic press, New York.
- [25] Hall, P. and Zhou, X. H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *The annals of statistics* **31**, 201–224.
- [26] Hanley, J. A. (1988). The robustness of the “binormal ” assumptions used in fitting ROC curves. *Medical Decision Making* **8**, 197–203.

- [27] Hanley, J. A. (1989). Receiver operating characteristic (ROC) methodology: the state of the art. *Critical Reviews in Diagnostics Imaging* **29**, 307–335.
- [28] Hanley, J. A. (1996). The use of the “binormal” model for parametric ROC analysis of quantitative diagnostic test. *Statistics in Medicine* **15**, 1575–1585.
- [29] Hanley, J. A. and Hajian-Tilaki, K. (1997). Sampling variability of nonparametric estimates of the areas under receiver operating characteristic curves: An update. *Academic Radiology* **4**, 49–58.
- [30] Hoff, P. D. (2006). Extending the rank likelihood for semiparametric copula estimation. (<http://arxiv.org/abs/math/0610413>).
- [31] Hsieh, F. S. and Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *The Annals of Statistics* **24**, 25-40.
- [32] Hsieh, F. S. (1991). Performance of diagnostic tests in a nonparametric setting. Ph.D. dissertation, Cornell University.
- [33] Hui, S. L. and Zhou, X. H. (1998). Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research* **7**, 354–370.
- [34] Hunink, M. G. M., Polak, J. F., Barlan, M. M. and O’Leary, D. H. (1993). Detection and quantification of carotid artery stenosis: Efficacy of various Doppler Velocity parameters. *American Journal of Roentgenology* **160**, 619–625.
- [35] Ishwaran, H. and Gastonis, C. A. (2000). A general class of hierarchical ordinal regression models with applications to correlated ROC analysis. *The Canadian Journal of Statistics* **28**, 731–750.

- [36] Joseph, L. G. Gyorkos, T. and Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology* **141**,263–272.
- [37] Komlós, J. , Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. verw. Gebiete* **32**, 111–131.
- [38] Le, C. T. (1997). Evaluation of confounding effects in ROC studies. *Biometrics* **53**, 998-1007.
- [39] Li, G., Tiwari, R. C. and Wells, M. T. (1996). Quantile comparison functions in two-sample problems: With applications to comparisons of diagnostic markers. *Journal of the American Statistical Association* **91**, 689–698.
- [40] Li, G., Tiwari, R. C. and Wells, M. T. (1999). Semiparametric inference for a quantile comparison function with applications to receiver operating characteristic curves. *Biometrika* **86**, 487-502.
- [41] Lin, C. and Mudholkar, G. (1980). A simple test for normality against asymmetric alternatives. *Biometrika* **67**, 455–461.
- [42] Lloyd, C. J. (1998). Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association* **93**, 1356–1364.
- [43] Lo, A. Y. (1987). A large sample study of the Bayesian bootstrap. *The Annals of Statistics* **15**, 360–375.
- [44] Lusted, L. B. (1960). Logical analysis in roentgen diagnosis. *Radiology* **74**, 178-193.

- [45] Mantel, N. (1951). Evaluation of a class of diagnostic tests. *Biometrics* **7**, 240–246.
- [46] Metz, C. E., Herman, B. A. and Shen, J. (1998). Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine* **17**, 1033–1053.
- [47] Moise, A., Clement, B., Raissis, M. and Nanopoulos, P. (1988). A test for crossing receiver operating characteristic (ROC) curves. *Communications in Statistics—Theory and methods* **17**:6, 1985–2003.
- [48] Murphy, S. A. and Van Der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association* **95**, 449–465.
- [49] Obuchowski, N. A. (1998). Sample size calculations in studies of test accuracy. *Statistical Methods in Medical Research* **7**, 371–392.
- [50] Pepe, M. S. (1998). Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. *Biometrics* **54**, 124–135.
- [51] Pepe, M. S. (2000). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics* **56**, 352–359.
- [52] Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford Statistical Science Series, Oxford University Press.
- [53] 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*, 171–212.
- [54] Qin, G. and Zhou, X. H. (2006). Empirical likelihood inference for the area under the ROC curve. *Biometrics* **62**, 613–622.

- [55] Qin, J. and Zhang, B. (2003). Using logistic regression procedures for estimating receiver operating characteristic curves. *Biometrika* **90**, 585–596.
- [56] Reilly, M. and Pepe, M. S. (1995). A mean-score method for missing and auxiliary covariate data in regression models. *Biometrika* **82**, 299–314.
- [57] Rodenberg, C. A. and Zhou X. H. (2000). ROC curve estimation when covariates affect the verification process. *Biometrics* **56**, 1256–1262.
- [58] Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of Statistics* **9**, 130–134.
- [59] Sethuraman, J. A. (1994). Constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.
- [60] Shapiro, D. E. (1999). The interpretation of diagnostic tests. *Statistical Methods in Medical Research* **8**, 113–134.
- [61] Slud, E. (1978). Entropy and maximal spacings for random partitions. *Z. Wahrsch. Verw. Gebiete* **41**, 341–352.
- [62] Staquet, M., Rozencweig, M., Lee, Y. J. (1981). Methodology for the assessment of new dichotomous diagnostic tests. *Journal of Chronic Diseases*, **34**, 599–610.
- [63] Swets, J. A. (1973). The relative operating characteristic in psychology. *Science* **182**, 990–1000.
- [64] Swets, J. A. and Pickett, R. M. (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, New York.
- [65] Swets, J. A. (1986a). Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychological Bulletin* **99**, 100–117.

- [66] Swets, J. A. (1986b). Form of empirical ROCs in discrimination and diagnostic tasks: implications for theory and measurement of performance. *Psychological Bulletin* **99**, 181–198.
- [67] Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics* **34**, 1265–1269.
- [68] Toledano, A. and Gatsonis, C. A. (1995). Regression analysis of correlated receiver operating characteristic data. *Academic Radiology* **2**, **Supplement 1**, S30-36.
- [69] Tosteson, A. N. A. and Begg, C. B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making* **3**, 204–215.
- [70] Vacek, P. M. (1985). The effect of conditional dependence on the evaluation of diagnostic test. *biometrics* **41**, 959–968.
- [71] Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.
- [72] Wald, A. (1950). *Statistical Decision Functions*. John Wiley & Sons, New York.
- [73] Walter, S. D. and Irwig, L. M. (1988). Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *Clinical Epidemiology* **41**, 923–937.
- [74] Wan, S. and Zhang B. (2007). Smooth semiparametric receiver operating characteristic curves for continuous diagnostic tests. *Statistics in Medicine* **26** , 2565–2586.
- [75] Wieand, S., Gail, M. H., James, B. R. and James, K. L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika* **76**, 585–592.

- [76] Zhou, X. H. (1993). Maximum likelihood estimators of sensitivity and specificity corrected for verification bias. *Communication in Statistics–Theory and Methods* **22**, 3177–3198.
- [77] Zhou, X. H. (1994). Effect of verification bias on positive and negative predictive values. *statistics in Medicine* **13**, 1737–1745.
- [78] Zhou, X. H. (1998). Correcting for verification bias in studies of a diagnostic test's accuracy. *Statistical Methods in Medical Research* **7**, 337–353.
- [79] Zhou, X. H. and Harezlak, J. (2002). Comparison of bandwidth selection methods for kernel smoothing of ROC curves. *Statistics in Medicine* **21**, 2045–2055.
- [80] Zhou, X. H., Castelluccio, P. and Zhou, C. (2005). Nonparametric estimation of ROC curves in the absence of a gold standard. *Biometrics* **61**, 600–609.
- [81] Zhou, X. H., McClish, D. K. and Obuchowski, N. A. (2002). *Statistical Methods in Diagnostic Medicine*. John Wiley & Sons, Inc., New York.
- [82] Zou, K. H. and Hall, W. J. (2000). Two transformation models for estimating an ROC curve derived from continuous data. *Journal of Applied Statistics* **27**:5, 621–631.
- [83] Zou, K. H., Hall, W. J. and Shapiro, D. E. (1997). Smooth nonparametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine* **16**, 2143–2156.
- [84] Zou K. H., Warfield S. K., Bharatha A., Tempany C. M., Kaus M. R., Haker S.J., Wells W. M. 3rd, Jolesz F. A., Kikinis R. (2004). Statistical validation of image

segmentation quality based on a spatial overlap index. *Academic Radiology* **11**, 178–189.

Appendix A

Theorems used for strong approximation theories

The same notations are used in the following Theorems; see Section 4.1 for definitions.

Theorem A: [Implied by Komlós, Major and Tusnády, 1975, Theorem 4]: Let $X_1, X_2, \dots \sim$ i.i.d. F , where F is any arbitrary continuous distribution function and $\mathbb{G}_n(x)$ be the empirical process. Then there exists a Kiefer process $\{K(y, t); 0 \leq y \leq 1, t \geq 0\}$, such that $\sup_x |\mathbb{J}_n(x) - n^{-1/2}K(F(x), n)| =_{a.s.} O(n^{-1/2} \log^2 n)$.

Theorem B: [Csörgő and Révész, 1978, Theorem 6]: Let $X_1, X_2, \dots \sim$ i.i.d. F , satisfying Condition A. Then the quantile process $\mathbb{Q}_n(x)$ of X can be approximated by a Kiefer process $\{K(y, t); 0 \leq y \leq 1, t \geq 0\}$. That is, $\sup_{\delta_n \leq y \leq 1 - \delta_n} |f(F^{-1}(y))\mathbb{Q}_n(y) - n^{-1/2}K(y, n)| =_{a.s.} O(l_n)$.

Theorem C: [Lo, 1987, Lemma 6.3]: There exists a Kiefer process $\{K(s, t); 0 \leq s \leq 1, t \geq 0\}$ independent of X such that $\sup_x |\mathbb{J}_n^\#(x) - n^{-1/2}K(\mathbb{F}_n(x), n)| =_{a.s.} O(l_n)$.

Theorem D: [Csörgő and Révész, 1978, Theorem 3]: Let X_1, X_2, \dots be i.i.d. F satisfying Condition A. Then $\limsup_{n \rightarrow \infty} \frac{\sqrt{n}}{\log \log n} \sup_{\delta_n \leq y \leq 1 - \delta_n} |f(F^{-1}(y))\mathbb{Q}_n(y) - \mathbb{W}_n(y)| \leq 40\gamma 10^\gamma$ a.s.

(γ is defined in Condition A).

Theorem E: [Csörgő and Révész, 1981, Lemma 4.5.1]: Let U_1, U_2, \dots be i.i.d. $U(0, 1)$ random variables, and let Kiefer process $\{K(y, t); 0 \leq y \leq 1, 0 \leq t\}$ defined on the same probability space. Then $\sup_{1 \leq k \leq n} n^{-1/2} |K(U_{k:n}, n) - K(k/n, n)| =_{a.s.} O(l_n)$.

Theorem F: [Csörgő and Révész, 1981, Theorem 1.15.2]: Let h_n be a sequence of positive numbers for which $\lim_{n \rightarrow \infty} \frac{\log h_n^{-1}}{\log \log n} = \infty$, $\gamma_n = (2nh_n \log h_n^{-1})^{-1/2}$. Then

$$\lim_{n \rightarrow \infty} \sup_{0 \leq t \leq 1 - h_n} \sup_{0 \leq s \leq h_n} \gamma_n |K(t + s, n) - K(t, n)| =_{a.s.} 1.$$

Theorem G: Let $K(s, n)$ be a Kiefer process, $\sup_{0 < s < 1} \left| \frac{K(s, n-1) - K(s, n)}{\sqrt{n-1}} \right| =_{a.s.} O(l_n)$.

The result is contained in the proof of Lemma 6.3 of Lo (1987).

Theorem H: [Law of Iterated Logarithm (LIL)]: Let $K(y, n)$ be Kiefer process, then

$$\limsup_{n \rightarrow \infty} \sup_{0 \leq y \leq 1} \frac{|K(y, n)|}{(2n \log \log n)^{1/2}} =_{a.s.} \frac{1}{2}.$$

Theorem I: [Special case of Lemma 6.2, Lo (1987) when $F = U(0, 1)$]: Let $U_1, \dots, U_m \sim$ i.i.d. $U(0, 1)$, $\mathbb{U}_m(s)$ is the empirical function of U_i 's. For any Kiefer process K independent of U_1, \dots, U_m , $\sup_{0 < s < 1} \left| \frac{K(\mathbb{U}_m(s), m) - K(s, m)}{\sqrt{m}} \right| =_{a.s.} O(l_m)$.

Theorem J: [Csörgő and Révész, 1978, Theorem D]: For $\mathbb{H}_n(x)$ the uniform empirical process and $\epsilon_n = 0.236 \frac{\log \log n}{n}$, we have $\limsup_{n \rightarrow \infty} \sup_{\epsilon_n \leq x \leq 1 - \epsilon_n} (x(1-x) \log \log n)^{-1/2} |\mathbb{H}_n(x)| =_{a.s.} 2$.

Theorem K: [Csörgő and Révész, 1978, Theorem 2]: Let $\mathbb{W}_n(y)$ be the uniform quantile process, then $\limsup_{n \rightarrow \infty} \sup_{\delta_n \leq y \leq 1 - \delta_n} (y(1-y) \log \log n)^{-1/2} |\mathbb{W}_n(y)| =_{a.s.} 4$.

Theorem L: [Csörgő and Révész, 1978, Lemma 1]: Under Condition A,

$$\frac{f(F^{-1}(y_1))}{f(F^{-1}(y_2))} \leq \left\{ \frac{y_1 \vee y_2}{y_1 \wedge y_2} \cdot \frac{1 - y_1 \wedge y_2}{1 - y_1 \vee y_2} \right\}^\gamma \text{ for any pair } y_1, y_2 \in (0, 1).$$

Theorem M: [Slud, 1978]: Let $U_1, \dots, U_n \sim$ i.i.d. $U(0, 1)$ and $M_n = \max_{0 \leq k \leq n} (U_{k+1:n} - U_{k:n})$. Then $nM_n / \log n \rightarrow_{a.s.} 1$.

Theorem N: Under condition A, for any $y \in [\delta_n^*, 1 - \delta_n^*] \subset [\delta_n, 1 - \delta_n]$, let $k = \lceil ny \rceil$, ξ lies between y and $V_{k:n}$. Then

$$\frac{y(1-y)}{\xi(1-\xi)} \leq 5 \quad (\text{A.1})$$

$$\frac{\xi}{V_{k:n}} \leq 5, \quad \frac{1-\xi}{1-V_{k:n}} \leq 5 \quad (\text{A.2})$$

$$\sup_{\delta_n^* \leq y \leq 1-\delta_n^*} f(F^{-1}(\xi))/f(F^{-1}(V_{k:n})) \leq 10^\gamma. \quad (\text{A.3})$$

Remark A.1 When $y \in [\delta_n^\#, 1 - \delta_n^\#]$, ξ lies between y and $\tilde{U}_{n-1}(y)$. By Theorem J of Appendix, the inequalities (A.1), (A.2) and (A.3) hold for y , ξ and $\tilde{U}_{n-1}(y)$ by the same arguments, yielding

$$\sup_{\delta_n^\# \leq y \leq 1-\delta_n^\#} \frac{y(1-y)}{\xi(1-\xi)} \leq 5 \quad (\text{A.4})$$

$$\sup_{\delta_n^\# \leq y \leq 1-\delta_n^\#} \frac{\xi}{\tilde{U}_{n-1}(y)} \leq 5, \quad \sup_{\delta_n^\# \leq y \leq 1-\delta_n^\#} \frac{1-\xi}{1-\tilde{U}_{n-1}(y)} \leq 5 \quad (\text{A.5})$$

$$\sup_{\delta_n^\# \leq y \leq 1-\delta_n^\#} f(F^{-1}(\xi))/f(F^{-1}(\tilde{U}_{n-1}(y))) \leq 10^\gamma. \quad (\text{A.6})$$

The results are contained in the proof of Theorem 3 of Csörgő and Révész (1978).

Appendix B

Theorems used for the proofs

B.1 Theorems used for proof the consistency of BPL estimator

Doob's Theorem [cf. Ghosal, S. and Van der Vaart, A. W., 2008]: Assume that θ is equivalent to an $\mathcal{X}^{(\infty)}$ -measurable random variable, i.e., there exists a $\mathcal{X}^{(\infty)}$ measurable function f on $\mathfrak{X}^{(\infty)}$ such that $\theta = f(\omega^\infty)$ a.e. $[\Pi \times P_\theta^{(\infty)}]$. Then the posterior $\Pi(\cdot|X^{(n)})$ is strongly consistent at θ for almost every θ $[\Pi]$.

Theorem a [Hájek and Šidák, 1967, page 157]: If $\int_0^1 \varphi^2(u)du < \infty$ holds, then

$$\lim_{N \rightarrow \infty} E\{E[\varphi(U_1)|R_{1N} = i] - \varphi(U_1)\}^2 = 0.$$

Appendix C

Programs used for simulation studies

C.1 Matlab code for the BPL method with the gold standard

```

% Given information x, m observations from nondiseased group
%                   y, n observations from disease group
%                   grid, the length of even interval of FPF
%                   NSCAN, an integer of MCMC iteration
%                   burnin, an integer of burn-in
%                   gap, an integer of lag
%                   a, initial value of normal mean
%                   b, initial value of normal standard deviation
%                   gib, vector of MCMC samples
% Help function--generate size=nsample random variables from
% truncated normal with non-truncated normal mean a, standard
% deviation b and corresponding truncation region (kc1,kc2)
function [TN2] =TN2(nsample,a,b,kc1,kc2)
xq=ones(nsample,1);
din= normcdf(kc2,a,b)- normcdf(kc1,a,b);% CDF
for i= 1:nsample
u=rand(1,1);
xq(i)=norminv(din*u+normcdf(kc1,a,b),a,b);
end;
TN2=xq;

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%start the main codes
z=[x;y];%(m+n)*1 matrix
idd=[zeros(m,1);ones(n,1)];
[temp, index]=sortrows(z);
id=idd(index);%labels of combined ordered statistics denoted as id
orid=1:(m+n);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%start to get the iniatial values R=order(Z,W)
Rtemp=zeros(m+n,1);
Rtemp(id==1)=sort(normrnd(a0,b0,n,1));%fill in W initial values
indexy=orid(id==1);%get the indices of W's initial denoted as indexy
if (indexy(1)>1) %fill in Z initial values
Rtemp(1:(indexy(1)-1))=sort(TN2((indexy(1)-1),0,1,-100,Rtemp(indexy(1))));
end;
gum=ones(size(indexy,2),2);
gum(:,2)= [diff(indexy) -1]-1;
gum(:,1)=indexy;
for i=1:n
if (gum(i,2)>0)
Rtemp((gum(i,1)+1):(gum(i+1,1)-1))
=sort(TN2(gum(i,2),0,1, Rtemp(gum(i,1)), Rtemp(gum(i+1,1))));
end;
end;
R=Rtemp;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%start Gibbs sampling
for j=1:NSCAN
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%update R based on current values a(j) and b(j)
i=1;
if (id(i)==0)
R(i)=TN2(1,0,1,-100,R(2));
else
R(i)=TN2(1,a(j),b(j),-100,R(2));
end;
for i=2:(m+n-1)
if (id(i)==0)
R(i)=TN2(1,0,1,R(i-1),R(i+1));
else
R(i)=TN2(1,a(j),b(j),R(i-1),R(i+1));
end;
end;
end;

```

```

i=(m+n);
if (id(i)==0)
R(i)=TN2(1,0,1,R(m+n-1),100);
else
R(i)=TN2(1,a(j),b(j),R(m+n-1),100);
end;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%update Z, W values based on R
Z=R(id==0);
W=R(id==1);
vrate=var(W)*(n-1)/2;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%update a and b by (Z,W)
b(j+1)=sqrt(1/gamrnd((n-1)/2, 1/vrate));
a(j+1)=normrnd(mean(W), b(j+1)/sqrt(n),1,1);
end;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%end of Gibbs sampling

%%%BPL estimate of intercept and slope in the binormal model for ROC
BPLintercept=mean(a(gib)./b(gib));
BPLslope=mean(1./b(gib));

```

Other sampling error information and BPL estimate of ROC can be obtained easily.

C.2 Matlab code for the BB method

The Matlab code to implement our BB estimate of ROC curve is given as follows:

```
%Given data: x,      m observations from nondisease group
%              y,      n observations from disease group
%              grid,  the length of even interval of FPF
%              rep,   resample size

%Define FPF and helper vectors, based on the information given before.;
t= [grid:grid:1-grid] % FPF vector
ot=ones(length(t),1) % vector of 1 with the same length as vector t
onx=ones(m,1);ony=ones(n,1)% vectors of 1 with the same length as x and y

%AUC function (using Simpson's method);
function [auc] =auc(roctrue,grid) %input ROC curve vector as roctrue.
auc=1/3*grid*(roctrue(1)+roctrue(length(roctrue))
+2*sum(roctrue(2:(length(roctrue)-1)))
+2*sum(roctrue(2:2:(length(roctrue)-1))))

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%BB estimate of ROC, AUC;
for r=1:rep
% note: to generate Dirichlet weight vectors p and q
p=exprnd(1,1,m);p=p/sum(p);q=exprnd(1,1,n);q=q/sum(q);
z=p*(x*ony'> onx*y');
roc(r,:) = q*(z'*ot'<ony*t);
aucbb(r) = auc(roc(r,:),grid)
end;
rocbb=mean(roc)% rocbb--BB estimate of ROC
aucbb=auc(rocbb, grid)% BB estimate of AUC
```

Other sampling error information can be obtained easily.

C.3 WinBUGS code for the Bernstein polynomial model

The software WinBUGS is available from the website

<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>. In order to use the truncated normal distribution function, please download the function from shared component from the WinBUGS development site at <http://www.winbugs-development.org.uk/>.

WinBUGS model specification code for the Bernstein polynomial model:

```

model {#trick using zeros
for (i in 1:N) {
zeros[i] <- 0
zeros[i] ~ dpois(phi[i])
phi[i] <- -log( 0.3989423*pow(exp(-0.5*pow(HS[i],2)),2-L[i])
*pow(1/sigma0*exp(-0.5*1/(sigma0*sigma0)*pow(HS[i]-mu0,2)),(L[i]-1))
*abs(J[i]))
HS[i]<-dh1*4*S[i]*pow((1-S[i]),3)+(dh1+dh2)*6*S[i]*S[i]*pow((1-S[i]),2)
+(dh1+dh2+dh3)*4*pow(S[i],3)*(1-S[i])+(dh1+dh2+dh3+dh4)*pow(S[i],4)-3
J[i]<-abs(dh1*4*pow((1-S[i]),2)*(1-4*S[i])+(dh1+dh2)*12*S[i]*(1-S[i])
*(1-2*S[i])+(dh1+dh2+dh3)*4*pow(S[i],2)*(3-4*S[i])
+(dh1+dh2+dh3+dh4)*pow(S[i],3)*4)
}
dh1~dexp(lamb)
dh2~dexp(lamb)
dh3~dexp(lamb)
dh4~dexp(lamb)
mu0~djl.dnorm.trunc(2,1,0,4)#
logsigma0~dunif(0,0.916)
sigma0<-exp(logsigma0)
for (j in 1:N){ L[j] ~ dcat(P[])}
P[1:2] ~ ddirch(alpha[])
}

```

C.4 WinBUGS code for the Latent model

WinBUGS model specification code for the Latent model

```

model
{
R[1] ~ djl.dnorm.trunc(lambda[T[1]],tau[T[1]],0,R[2])
for( i in 2 :(N-1) ) {
R[i] ~ djl.dnorm.trunc(lambda[T[i]],tau[T[i]],R[i-1],R[i+1])
}
R[N] ~ djl.dnorm.trunc(lambda[T[N]],tau[T[N]],R[N-1],1000)
for( i in 1 :N ) {
T[i] ~ dcat(P[])
lambda[1]<-0
lambda[2]<- lambda[1] + c
c ~ djl.dnorm.trunc(2,1,0,4)
tau[1]~dbern(1)
logsigma0~dunif(0,0.916)
sigma<-exp(logsigma0)
tau[2]<-1/(sigma*sigma)
P[1]~ djl.dnorm.trunc(alpha,10000,0,1)
P[2]<-1-P[1]
}

```