A Dissertation

entitled

Statistical Inference for Binormal ROC Curves
under a Density Ratio Model

by

Wencan He

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the
Doctor of Philosophy Degree in Mathematics

Dr. Biao Zhang, Committee Chair

Dr. Donald White, Committee Chair

Dr. Rong Liu, Committee Chair

Dr. Tian Chen, Committee Chair

Dr. Amanda Bryant-Friedrich, Dean
College of Graduate Studies

The University of Toledo
December 2016
The binormal ROC curve is a classic model for ROC curves. Originally, its functional form is derived from test results that are normally distributed in the diseased and nondiseased populations. However, since the ROC curve pertains to the relationships between two populations rather than to the distributions themselves, the binormal ROC curve applies in more general settings. In this study, we focus on semiparametric inferences of binormal ROC curves under a density ratio model. Density ratio models have a natural connection with generalized linear model, which has been widely used in biostatistics and other areas of applied statistics.

First, by deriving a two-sample density ratio model of the diseased and nondiseased populations from the binormal ROC model, we propose a semiparametric estimator of binormal ROC curve (called pseudo empirical likelihood non-iterative method or pelni method). Our pelni method is proven that the limiting distribution of pelni is normal. It is shown via a simulation study that pelni method surpasses the fully parametric and nonparametric methods in terms of robustness and efficiency. Since the pelni method only uses the empirical distribution function of the nondiseased sample to estimate the unknown transformation, it is less efficient than other semiparametric methods such as the mle and pmle method, especially when sample sizes are small. An analysis of two real examples is presented.
To improve the stability of \textit{pelni} method, we propose a pseudo empirical likelihood (\textit{pel}) estimator. It is proven that the limiting distribution of our \textit{pel} estimator is also normal. Via a simulation study, we show that \textit{pel} is more robust than a fully parametric approach and is more accurate than a fully nonparametric approach. Simulation also shows that, our \textit{pel} method is more efficient than the \textit{pmle} method by Cai & Moskowitz [2004], and is quite comparable to their \textit{mle} method. An analysis of two real examples are presented.

Hazard function of the survival times is among the most popular methodology in risk measurement. Kernel estimator is a nonparametric method widely applied in biostatistics and economics. Finally, using the strong approximation technique, we propose a kernel estimator for the hazard function. We show that the limiting distribution for the proposed kernel estimator of hazard function is normal. A real data application of kernel smoothing on hazard rate is presented.
Acknowledgments

This work has been supervised by Dr. Biao Zhang. It is a great pleasure to be his student working under his outstanding education, suggestion and patience to me. Without his help, I would not be able to finish my Ph.D. degree at this state. Also, I am grateful for Dr. Christopher Cooper, who consistently supports me and gives me opportunities of applying my statistical knowledge into clinical research. I owe a million to Prof. Pam Brewster for her guidance, understanding and encouragement.

In time, I would like to thank all the faculty, staff members in Department of Mathematics and Statistics, Department of Medicine, and Coordinating Center of Clinical Trial, especially, my defense committee members, Department Chair Dr. Donald White, Dr. Rong Liu, Dr. Tian Chen, for their suggestions and comments to my dissertation; my teachers, Dr. Sibylle Weck-Schwarz, Dr. Zeljko Cuckovic, Dr. Qin Shao, Dr. Lihua Chen, Dr. Minghui Paik, Dr. Zhiwei Chen, Dr. Harvey Wolf, Dr. Denis White, Dr. Biao Ou, for their diversified knowledge which enriched me; my teammates, Dr. Jiang Tian, Dr. Steven Haller, Dr. Kaleigh Evans, Dr. David Folt, Dr. Mark Yu, Dr. Christopher Drummond, Dr. Kung Xiang, Dr. Huilin Shi, Emily Cooper, for their great supports. I am also thankful to Mrs. Debra Middleton, Mrs. Michel Shirley and Mrs. Sarah Moomey for their kindly help.

I want to thank my classmates, Kaili Ren, Gang Liu, Jingning Mei, Lin Tang, Hui Shao, Yi Zhang, Yanmei Xie, Hanyang Liu, Haifeng Yu, Ming Zhao, Dong Zhang, for their friendship and encouragement.

I am grateful to Pastor Terry Chen, brothers and sisters in Toledo Chinese Al-
liance Church, Raleigh Chinese Christian Church and Peace Fellowship, for their enlightenment and prayer for me.

Deep love to my parents, my wife and my kids.
Contents

Abstract iii

Acknowledgments v

Contents vii

List of Tables x

List of Figures xii

1 Introduction 1
   1.1 Overview .................................. 1
   1.2 ROC analysis ............................. 5
   1.3 Binormal ROC curve ....................... 7
   1.4 Density ratio model ....................... 12
   1.5 Kernel estimation of hazard rate ............ 14

2 Estimate binormal ROC curve under density ratio model –
   Pseudo Empirical Likelihood Non-Iterative method 16
   2.1 Introduction ............................. 16
   2.2 Methodology ............................. 19
   2.3 Examples ................................. 22
   2.4 Simulations ............................. 23
      2.4.1 Correct specification ................. 25
         2.4.1.1 Normal distribution .............. 25

vii
## List of Tables

1.1 Frequency table of test results by disease status for a binary test . . . . . 5
1.2 Other interpretations for an ROC curve ................................. 6

2.1 Estimation of $(a, b)$ from normal setup, sample sizes $(n_D, n_D) = (50, 50)$, 1000 simulations ......................................................... 27
2.2 Estimation of AUC and ROC from normal setup, sample sizes $(n_D, n_D) = (50, 50)$, 1000 simulations ......................................................... 28
2.3 Estimation of $(a, b)$ from normal distribution, incremental sample sizes, 1000 simulations ................................................................. 29
2.4 Estimation of $(a, b)$ from lognormal distribution, sample sizes $(n_D, n_D) = (50, 50)$, 1000 simulations ......................................................... 32
2.5 Estimation of AUC and ROC from lognormal distribution, sample sizes $(n_D, n_D) = (50, 50)$, 1000 simulations ......................................................... 33
2.6 Estimation of $(a, b)$ from lognormal distribution, incremental sample sizes, 1000 simulations ................................................................. 34
2.7 Estimation of AUC and ROC from exponential distribution, sample sizes $(n_D, n_D) = (50, 50)$, 1000 simulations ......................................................... 36
2.8 Estimation of AUC and ROC from Gamma distribution, sample sizes $(n_D, n_D) = (50, 50)$, 1000 simulations ......................................................... 38

3.1 Example of antigen CA19-9 ......................................................... 63
3.2 Example of UACR data in CORAL study ........................................ 64
3.3 Estimation of \((a,b)\) from normal distribution, sample sizes \((n_D,n_D) = (50,50)\), 1000 times simulation .................................................. 67
3.4 Estimation of AUC and ROC from normal setup, sample sizes \((n_D,n_D) = (50,50)\), 1000 times simulation .................................................. 68
3.5 Estimation of \((a,b)\) from normal setup, incremental sample sizes, 1000 simulation ................................................................. 69
3.6 95% confidence interval for AUC from normal distribution, sample sizes \((n_D,n_D) = (50,50)\), 1000 times simulation .................................................. 70
3.7 Estimation of \((a,b)\) from lognormal setup, sample sizes \((n_D,n_D) = (50,50)\), 1000 times simulation .................................................. 74
3.8 Estimations for AUC and ROC from lognormal setup, sample sizes \((n_D,n_D) = (50,50)\), 1000 times simulation .................................................. 75
3.9 Estimation for \((a,b)\) from lognormal setup, incremental sample sizes, 1000 times simulation ................................................................. 76
3.10 95% confidence interval for AUC from lognormal distributions, sample sizes \((n_D,n_D) = (50,50)\), 1000 simulations .................................................. 77
3.11 AUC and ROC estimation for Exponential distribution, sample sizes \((n_D,n_D) = (50,50)\), 1000 simulations .................................................. 78
3.12 95% confidence interval for AUC from exponential distributions, sample sizes \((n_D,n_D) = (50,50)\), 1000 times simulation .................................................. 79
3.13 AUC and ROC estimation for Gamma distribution, sample sizes \((n_D,n_D) = (50,50)\), 1000 times simulation .................................................. 81
3.14 95% confidence interval for AUC from Gamma distributions, sample sizes \((n_D,n_D) = (50,50)\), 1000 times simulation .................................................. 82
List of Figures

1-1 Comparison of true ROC curve with binormal ROC curves for bilognormal distributions ...................................................... 11
1-2 Comparison of true ROC curve with binormal ROC curves for bigamma distributions ...................................................... 12
3-1 ROC curves for UACR data in CORAL study .............................. 64
3-2 Plot of $\Phi^{-1}(R)$ versus $\Phi^{-1}(s)$ ........................................ 65
4-1 Scatter plot of the point wise estimator of the hazard function, and kernel smooths with different bandwidths from the HMO-HIV+ study .......................... 112
Chapter 1

Introduction

1.1 Overview

Diagnostic tests are an important part of medical care. ROC analysis has become a popular method for evaluating the accuracy of medical diagnostic systems. There are a number of methods for estimating the ROC curve for a continuous test. A fully parametric approach that results in a smooth curve models the constituent distribution functions parametrically in order to arrive at the induced estimator of the ROC curve. A nonparametric method that results in a step function is to use the empirical estimate whose properties have been derived by Hsieh & Turnbull [1996]. An intermediate strategy between these two is a semiparametric approach.

Several different methods have been used to produce a semiparametric ROC curve. For example, Li et al. [1999] propose using a nonparametric approach to estimate the distribution of test results in nondiseased subjects, but then assume a parametric model for the distribution of test results in the diseased subjects. Without assuming a functional relationship between these two distributions, they use maximum likelihood to estimate the unknown parameters in distribution of the diseased population. Qin & Zhang [2003] also assume that the distribution of test results in nondiseased subjects is unknown, but instead assumes a functional form for the density (likelihood) ratio.
function, illustrating a relationship between two distributions. A more commonly taken strategy to semiparametric estimation of the ROC curve is to model the ROC curve parametrically, but avoid making additional assumptions about the distribution of test results. These types of approaches are called parametric distribution-free (Pepe [2000a]; Alonzo & Pepe [2002]). They produce smooth estimated curves while requiring less stringent assumptions than a fully parametric approach. The binormal ROC curve is perhaps the most popular of these intermediate strategies. The binormal ROC curve plays a central role as a classic model in ROC analysis similar to the way that normal distribution is a classic model for distribution functions (Pepe [2003]). Swets [1986] and Hanley [1996] conclude that binormal ROC curve provides a good approximation to a wide range of ROC curves that occur in practice. Further, it has been used extensively in applied research as a simple tool to describe the accuracy of rating data in radiology and psychometric research, to compare tumor markers for various types of cancer, and to compare laboratory blood tests for the screening of prostate cancer.

Metz *et al.* [1998] propose an algorithm called LABROC to estimate the binormal ROC curve. It does not use a transformation of the actual data scale; rather, it fits a latent model to categorized data which was formed from ranks. Pepe [2000a], Alonzo & Pepe [2002] propose that ROC curves can be estimated within the generalized linear model (GLM) binary regression framework. It produces a smooth ROC curve based on the parametric model of the ROC curve. However, these methods are not maximum-likelihood methods. Cai & Moskowitz [2004] propose two methods for estimating the parameters in the binormal ROC curve. The first method uses the profile likelihood and a simple algorithm to produce fully efficient estimates. The second method is based on a pseudo maximum-likelihood that can accommodate adjusting for covariates that could affect the accuracy of the continuous test. Both methods model the binormal ROC curve directly and the estimations are achieved using the
maximum likelihood approach. However, their methods only use the information of
the diseased sample after a monotonic increasing transformation.

Density ratio models have a natural connection with generalized linear models,
which have wide application in many areas, especially in biostatistics and economics
(Agresti [2013]). In the literature, many statistical methods have been developed
under density ratio models. Qin & Zhang [1997] consider a Kolmogorov-Smirnov-type
statistic to test the validity of the logistic regression model under a two-sample density
ratio model. Fokianos et al. [1998] apply a two-sample density ratio model to analyze
a dataset from space borne precipitation radar and space borne radiometer. Fokianos
et al. [2001] test the hypothesis that all population distributions are identical under
a multiple-sample density ratio model. More recently, Wan & Fang [2012] propose
a semiparametric hypothesis testing procedure to test the difference between two
population means under a two-sample density ratio model. Relying on the invariance
property of length-biased failure time data under a two-sample density ratio model,
Shen et al. [2012] present two likelihood approaches for the estimation and assessment
of the difference between two survival distributions. Semiparametric density ratio
models have also been widely used in ROC analysis. Qin & Zhang [2003] propose
a semiparametric approach to estimate ROC curve and AUC by assuming a density
ratio model for the diseased and nondiseased densities, whereas Zhang [2006] develops
a semiparametric hypothesis testing. Wang & Zhang [2014] develop a semiparametric
empirical likelihood confidence interval for AUC.

Our work in Chapters 2 and 3 are motivated by the diverse applications of den-
sity ratio models and the semiparametric ROC curve estimators proposed by Cai &
Moskowitz [2004]. In chapter 2, we present a semiparametric method called pelni
based on the binormal model. The new proposed method is simple to apply, since it
does not require solving \( \hat{h} \) and \( (\hat{a}, \hat{b}) \) iteratively. Simulation results show that pelni
method gives consistent estimations similar to other semiparametric methods. How-
ever, when sample sizes are small, *pelni* method is not as stable as the other methods. The asymptotic distributions of the estimators are derived. Two real examples are presented.

In chapter 3, we propose a semiparametric method (*pel* method) to fit a binormal ROC model. Our *pel* method is still a pseudo likelihood method since we use plug-in method to estimate the parameters. Systematic simulation study shows that the *pel* method is better than the *pmle* method (Cai & Moskowitz [2004]) in terms of estimation for \((a,b)\). Moreover, our *pel* method is close to the *mle* method (Cai & Moskowitz [2004]) in terms of \((a,b)\) estimation. As to the estimation of AUC and ROC, our *pel* method is quite comparable to the *mle* and *pmle* methods. And the *pel* method is similar or better than *labroc* proposed by Metz et al. [1998]. Two examples based on real data are provided. The asymptotic distributions of the estimators are derived. In addition, we use the method to construct Wald-type confidence intervals for AUC.

As we all know, logit regression and ROC analysis targets the binary response variable and whether an event occurs, while survival analysis targets the response variable of time until event. Survival analysis provides insight on not only whether the event will occur, but also how fast the event will occur. The key measurement in survival analysis is the hazard rate. Many literatures are discussing the estimation of the hazard function, among which the kernel estimation is an important nonparametric approach due to its robustness. In chapter 4, we apply the strong approximation technique to study the asymptotic normality for the kernel estimator of hazard rate. An example of kernel estimation of hazard function using real data is provided. Proofs of theoretical results are presented at the end of Chapter 4.
1.2 ROC analysis

An important role of diagnostic medicine research is to estimate and compare the accuracy of diagnostic tests. The intrinsic accuracy of test is measured by comparing test results to the true condition status of the patient. We determine the true condition status by means of a gold standard, such as autopsy reports, surgery findings, pathology results from biopsy specimens, and the results of other diagnostic tests which have perfect or nearly perfect accuracy.

The accuracy of a diagnostic test with binary test results is commonly summarized by sensitivity and specificity. Sensitivity is defined as the probability that test results are positive and the subject is truly diseased. Specificity is the probability that the test result is negative and the subject is truly nondiseased. Table 1.1 is the frequency table of test results by disease status for a binary test. As shown in the table, the sensitivity of the test is given by $\frac{a}{a+b}$, which is a true positive rate (TPR). And, the specificity is given by $\frac{d}{c+d}$, which is a true negative rate (TNR).

Table 1.1: Frequency table of test results by disease status for a binary test

<table>
<thead>
<tr>
<th>Test result</th>
<th>Disease status</th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>True Positive (TP)</td>
<td>a</td>
<td></td>
<td>a+c</td>
</tr>
<tr>
<td>Negative</td>
<td>False Negative (FN)</td>
<td>b</td>
<td></td>
<td>b+d</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>a+b</td>
<td>c+d</td>
<td></td>
</tr>
</tbody>
</table>

For continuous or ordinal test results, a threshold is needed to dichotomize test results, and then calculate the sensitivity and specificity. The receiver operating characteristic (ROC) curve is the most popular approach when test results are not simply positive or negative, but are measured on a continuous or ordinal scale. Each
point on the ROC curve is corresponding to the sensitivity and specificity of the
dichotomized data using a different decision threshold. ROC curves were originally
developed for electronic signal-detection theory. They have been applied in many
medical and non-medical areas, including studies of human perception and decision
making, industrial quality control, and military monitoring (Zhou et al. [2014]).

Let $T$ be a random variable that denotes the outcome of a continuous-scale di-
agnostic test with the convention that higher values of $T$ are more indicative of the
diseased. Let $D$ be the true disease status with $D = 1$ for the diseased and $D = 0$
for the nondiseased. For a given threshold value $c$, sensitivity and specificity are
defined as $se = \Pr(T \geq c \mid D = 1)$ and $sp = \Pr(T < c \mid D = 0)$, respectively. The
ROC curve is a plot of the sensitivity against one minus the specificity across all
possible choices of threshold values. Let $X_1, X_2, \ldots, X_{n_D}$ denote independent and
identically distributed test results from a nondiseased population and, independent
of the $X_i$’s, let $Y_1, Y_2, \ldots, Y_{n_D}$ be independent and identically distributed test results
from a diseased population. Let $G$ and $F$ represent, respectively, the distribution
functions of $X_1$ and $Y_1$, and the ROC curve is a plot of $R(s) = 1 - F\{G^{-1}(1 - s)\}$
against $s \in [0, 1]$. It can be interpreted in multiple ways; for example, TPR versus
FPR, sensitivity versus one minus specificity, power versus the type-I error (see Table
1.2).

<table>
<thead>
<tr>
<th>s</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPR</td>
<td>TPR</td>
</tr>
<tr>
<td>$1 - sp$</td>
<td>$se$</td>
</tr>
<tr>
<td>$Pr(I)$</td>
<td>$1 - Pr(II) = \beta$</td>
</tr>
</tbody>
</table>

The ROC curve has many advantages over isolated measurements of sensitivity
and specificity (Zweig & Campbell [1993]). For example, the ROC curve does not require selection of a particular decision threshold since all possible decision thresholds are included. The ROC curve is a device that allows people to develop different algorithms to decide the optimal threshold. In addition, it does not depend on the scale of test results. That is, it is invariant to monotonic transformations of test results, such as linear, logarithm, and square root. The ROC curve essentially provides a distribution-free description of the separation between the distributions of the diseased and nondiseased group. For reviews of recent developments in ROC curve methodologies and their applications, readers are referred to the works of Begg [1991], Zweig & Campbell [1993], Hsieh & Turnbull [1996], Pepe [2000b], and Alemayehu & Zou [2012].

When it is not feasible to plot the ROC curve itself, it is often useful to summarize the accuracy of a test by a single number. The area under the ROC curve (AUC), defined as $\text{AUC} = \frac{1}{0} \int R(s) ds$, is the most widely used summary index of diagnostic accuracy for a continuous-scale diagnostic test. Bamber [1975] shows that $\text{AUC} = \Pr(X < Y)$, i.e., the probability of a randomly picked diseased patient having a test result greater than that of a randomly picked healthy person. A higher value of AUC is corresponding to more accurate separation between the diseased and nondiseased population by the diagnostic test.

### 1.3 Binormal ROC curve

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 test results. Specifically, the binormal model assumes that $h(X) \sim$
Then the binormal ROC curve is given by

\[ R(s) = \Phi(a + b\Phi^{-1}(s)), \quad s \in [0, 1], \quad (1.1) \]

where

\[ a = \frac{\mu_D - \mu_D}{\sigma_D}, \quad b = \frac{\sigma_D}{\sigma_D}. \quad (1.2) \]

Since \( h(Y) - h(X) \) follows the normal distribution \( N(\mu_D - \mu_D, \sigma_D^2 + \sigma_D^2) \) and AUC has the meaning of \( d = \Pr(Y > X) = \Pr(h(Y) - h(X) > 0) \), we have \( d = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right) \).

Some interesting implications of the binormality assumption are explored by several authors. Green & Swets [1966], Metz et al. [1998] display the binormal model ROC visually in the following way:

\[ \text{TPR}(s) = \Phi(a + b\Phi^{-1}(\text{FPR}(s))), \quad s \in R. \quad (1.3) \]

From which we can see that the binormality also implies a perfect linear relationship between TPR and FPR on “normal-deviate axes”. That is,

\[ \Phi^{-1}(\text{TPR}(s)) = a + b\Phi^{-1}(\text{FPR}(s)), \quad s \in R \quad (1.4) \]

Lloyd [1998] considers a special case to interpret the meaning of the parameters \( a \) and \( b \) in the binormal model. If \( G = N(0, \sigma^2) \), \( F = N(\delta, \sigma^2) \), then \( a = \frac{\delta}{\sigma} \) and \( b = 1 \). Notice that the intercept parameter \( a \) is also equal to the square root of the Mahalanobis distance between the distribution function \( G \) and \( F \). 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.

Cai & Moskowitz [2004] point out that binormality implies that the ratio of the density function for the diseased group over that for the nondiseased group estimated
at $x$ is equal to the ratio of the normal density with mean $\mu_D$ and standard deviation $\sigma_D$ over that of the standard normal density evaluated at $-h(x)$. That is,

$$\frac{f_D(x)}{\bar{f}_D(x)} = \frac{b\phi(a - bh(x))}{\phi(-h(x))}. \quad (1.5)$$

To estimate the binormal parameters $(a, b)$, several methods have been proposed in the literature. 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 is

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

where $k_i$ is the number of observations from the nondiseased group dropped in $i^{th}$ category, $l_j$ is the number of observations from the diseased group dropped in $j^{th}$ category, $P_{i|D}$ is the probability of one observation from the nondiseased 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 $s = (s_0, s_1, \cdots, s_I)$, where $[s_{i-1}, s_i)$ is the range of $i^{th}$ category, $s_0 = -\infty$, $s_I = \infty$, and $s_i$ 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.

Zou & Hall [2000] use the Box-Cox transformation to estimate the strictly increasing transformation function $h(\cdot)$. Pepe [2000a] suggests a GLM method (called BN-G) for estimating the ROC curve, using the relationship

$$\mathbb{E}(I(Y > X)|\bar{F}(X) = s) = \Pr(Y > X|X = \bar{F}^{-1}(s)) = \bar{G}(\bar{F}^{-1}(s)) = R(s).$$

Defining the quantity $U_{js} = I(Y_j > F^{-1}(s))$, Alonzo & Pepe [2002] propose another GLM procedure. The method is computationally simpler than BN-G method.
According to (1.5), Cai & Moskowitz [2004] propose a maximum profile likelihood estimation method (called the mle method). In the same paper, they also propose a pseudo maximum likelihood estimator (called the pmle method) based on the imputation of the transformation function $h(\cdot)$ from the data of two groups. For the pmle method, the parameters $a$ and $b$ are estimated using maximum likelihood method only based on the data transformed from the diseased group. This motivates us to develop the methods in chapter 2 and 3. We propose two methods to estimate the binormal parameters: the pseudo empirical likelihood non-iterative (pelni) method and the pseudo empirical likelihood (pel) method. Instead of estimating the parameters $a$ and $b$ only based on the data from the diseased group, our proposed methods estimate the parameters $a$ and $b$ using the data from both diseased and nondiseased group.

In the literature, the commonly used monotonic increasing transformation function is $h_G(x) = \Phi^{-1}(G(x))$. The binormality implies that $h_G(X) \sim N(0,1)$ and $h_G(Y) \sim N\left(\frac{a}{b}, \frac{1}{b^2}\right)$ (see [Pepe, 2003, p. 85] for proof). Symmetrically we find another monotonic increasing transformation function $h_F(x) = \Phi^{-1}(F(x))$, which transforms the data to $h_F(Y) \sim N(0,1)$ and $h_F(X) \sim N(-a,b^2)$ under the binormal model.

Hanley [1989, 1996] illustrates by many figures that, even the nondiseased $X$ and diseased $Y$ are way from normally distributed, binormal model still works very well. We develop some figures to compare the true ROC curve with those based on binormal model. Figures 1-1 and 1-2 show the correct specification and misspecification of the binormal model, respectively. Figure 1-1 corresponds to a bilognormal case, in which both $X$ and $Y$ have the lognormal distribution. In this case, the binormal model is correct, and the true ROC curve and the binormal ROC curves using $h_G(\cdot)$ or $h_F(\cdot)$ are overlapping. Figure 1-2 corresponds to a bigamma case in which both $X$ and $Y$ have Gamma distribution. The binormal model assumptions are not satisfied in this case. However, the true ROC curve are still very close to the ROC curves using $h_G(\cdot)$
or $h_F(\cdot)$ transformed data based on the binormal model.

Figure 1-1: Comparison of true ROC curve with binormal ROC curves for bilognormal distributions. (a) Density curves of original data where $X \sim \log N(0, 1)$, $Y \sim \log N(1, 0.5)$; (b) Q-Q plot of original data; (c) Density curves of transformed data; (d) Q-Q plot of transformed data; (e) ROC curves
Figure 1-2: Comparison of true ROC curve with binormal ROC curves for bigamma distributions. (a) Density curves of original data where $X \sim \text{Gamma}(0.7, 1)$, $Y \sim \text{Gamma}(2, 1)$; (b) Q-Q plot of original data; (c) Density curves of transformed data; (d) Q-Q plot of transformed data; (e) ROC curves

1.4 Density ratio model

Logistic regression models are commonly used in analyzing binary data which arise in studying relationships between diseases and environment or genetic characteristics. Let $D$ be a binary response variable and $T$ be the associated $p \times 1$ covariate vector. Then the standard logistic regression model is:

$$Pr(D = 1|T = x) = \frac{1}{1 + \exp\{-\alpha^* - \beta^* r(x)\}},$$  \hspace{1cm} (1.6)
where $\alpha^*$ is a scalar parameter, $\beta$ is a $p \times 1$ vector parameter, and $r(x)$ is a $p \times 1$ smooth vector function of $x$. We denote the distribution function and density function by $F(\cdot)$ and $f(\cdot)$, respectively. Let $X_1, X_2, \ldots, X_{n_D}$ be a random sample from $F(x|D = 0)$ and, independent of the $X_i$, let $Y_1, Y_2, \ldots, Y_{n_D}$ be a random sample from $F(x|D = 1)$. Let $f(x|D = i) = dF(x|D = i)$ represents the conditional density of $x$ given $D = i$ for $i = 0, 1$. Qin & Zhang [1997] propose the following two-sample two-sample semiparametric density ratio model in which the two unknown density functions $g(x)$ and $f(x)$ are linked by an “exponential tilt” $\exp\{\alpha + \beta^T r(x)\}$:

\begin{align}
X_1, X_2, \ldots, X_{n_D} & \text{ are independent with density } g(x), \\
Y_1, Y_2, \ldots, Y_{n_D} & \text{ are independent with density } f(x) = \exp\{\alpha + \beta^T r(x)\}g(x).
\end{align}

(1.7)

Kay & Little [1987] discuss various versions for the density ratio model for some conventional distributions. For a choice of $r(x)$, we can use Kolmogorov-Smirnov-type statistic of Qin & Zhang [1997] to test the validity of model (1.7). Voulgaraki et al. [2012] suggest graphic and quantitative diagnostic tools to assess the validity of model (1.7). If more than one form of $r(x)$ is possible for a particular situation, we prefer simple models to complex ones that fit our data almost equally well according to the principle of parsimony. Some choices of $r(x)$ are summarized by

\begin{align*}
&\begin{cases}
  x & \text{for Normal distribution with } \sigma_D = \sigma_D, \text{ Exponential distribution} \\
  x, x^2 & \text{for Normal distribution with } \sigma_D \neq \sigma_D \\
  x, \log(x) & \text{for Gamma distribution} \\
  \log(x), \log(1 - x) & \text{for Beta distribution} \\
  x^2 & \text{for Rayleigh distribution} \\
  \log x & \text{for Lognormal distribution with } \sigma_D = \sigma_D \\
  \log x, \log^2 x & \text{for Lognormal distribution with } \sigma_D \neq \sigma_D
\end{cases}
\end{align*}
Note that all the choices of $r(x)$ are depending on the distribution of the original data. It cab be challenging to select the correct format of $r(x)$ for different distributions. We propose a new two-sample density ratio model based on the binormal model (1.3). That is,

$$\begin{align*}
X_1, X_2, \cdots, X_{nD} & \text{ are independent with density } g(x), \\
Y_1, Y_2, \cdots, Y_{nD} & \text{ are independent with density } f(x) = w(h(x); a, b)g(x),
\end{align*}$$

where

$$w(h(x); a, b) = \exp \left\{ \log b - \frac{a^2}{2} + abh(x) + \frac{1 - b^2}{2}h(x)^2 \right\}.$$

and $h(\cdot)$ is an unknown transformation function which can be estimated using our proposed two methods. As demonstrated by Hanley [1989, 1996], the binormal model applies to broad selection of distributions. Therefore, the density ratio model (1.8) is generic to broad selection of distribution.

Similar to the approach in Qin & Zhang [1997], we employ the Lagrange multiplier method to maximize the empirical likelihood based on the density ratio model (1.7) or (1.8). For more detailed discussions of empirical likelihood and its applications in various areas, the reader is referred to Owen [2001]'s monograph.

### 1.5 Kernel estimation of hazard rate

In the analysis of lifetime data or time-to-event data, a primary interest is to assess the risk of an individual at certain times (or ages). Let $T$ denote a lifetime variable with distribution function $F(t) = \Pr(T \leq t)$ and probability density function $f(t)dF(t)/dt$. The risk of an individual at age $t$ can be measured by the so called
“hazard rate” or “hazard function”, which is defined as:

$$\lambda(t) = \frac{f(t)}{1 - F(t)}, \text{ for } F(t) < 1. \quad (1.9)$$

That is, $\lambda(t)dt$ represents the instantaneous chance that an individual will have event in the interval $(t, t + dt)$ given that this individual is alive at age $t$. The hazard rate provides the trajectory of risk and is widely used also in other fields. Engineers refer to it as “failure rate function” and demographers refer to it as “force of mortality function”. The term “lifetime” simply denotes the time until the occurrence of an event of interest.

While parametric models provide convenient ways to analyze lifetime data, the necessary model assumptions, when violated, can lead to erroneous analyses and thus need to be checked carefully. The nonparametric approach to estimate hazard rates for lifetime data is flexible, model-free and data-driven. No shape assumption is imposed other than that the hazard function is a smooth function. Such an approach typically involves smoothing of an initial hazard estimate, with arbitrary choice of smoother. Tanner & Wong [1983] propose a kernel estimate of the hazard function from censored data by convolution smoothing of the empirical hazards. They give the small and large sample expression for the mean and the variance of the estimator. Conditions for asymptotic normality are investigated using the Hajek projection method. Using the strong approximation technique, we propose a kernel estimator for the hazard function. We show that the limiting distribution for the proposed kernel estimator of hazard function is normal. We apply the kernel estimation to a real data and decide the bandwidth according to Zambom & Dias [2012].
Chapter 2

Estimate binormal ROC curve under density ratio model – Pseudo Empirical Likelihood Non-Iterative method

2.1 Introduction

An important role of medical diagnostic test is to estimate the accuracy of the test. The receiver operating characteristic (ROC) curves are commonly used to test the accuracy of diagnostic tests when the test results are continuous or ordinal. The ROC curve is a plot of sensitivity versus one minus specificity through all possible choices of threshold values. For reviews of recent developments in ROC curve methodologies and their applications, readers are referred to the works of Begg [1991], Zweig & Campbell [1993], Hsieh & Turnbull [1996], Pepe [2000b], Alemayehu & Zou [2012].

Parametric methods, which directly make assumptions on the distributions of the diseased and nondiseased population, are most efficient when the model assumptions are correct. However, when the assumptions are violated, the estimations based on

Semiparametric methods are the intermediate approaches which balance the robustness and efficiency. A commonly taken strategy, called parametric distribution free (Pepe [2000a], Alonzo & Pepe [2002]), is to model the ROC curve parametrically, but avoid making additional assumptions about the distribution. It produces smooth ROC curve based on the parametric model of the ROC curve. The binormal ROC curve is the most popular semiparametric method to estimate ROC curve. It models an ROC curve by

$$\displaystyle R(s) = \Phi(a+b\Phi^{-1}(s)), \, 0 \leq s \leq 1, \, b > 0, \quad (2.1)$$

where $\Phi$ is the CDF of the standard normal distribution and $\theta = (a, b)^T$ are called binormal parameters. This model plays a central role in the development of ROC analysis. Swets [1986] and Hanley [1996] conclude that it appropriately approximates a wide range of ROC curves, even in the case neither the nondiseased nor diseased population looks normally distributed.

Metz et al. [1998] propose an algorithm called LABROC based on the binormal model. It does not use a transformation of the actual data scale; rather it fits a latent model to categorized data which is formed from ranks. Pepe [2000a] gives an interpretation for each point on the ROC curve as being a conditional probability of a test result from a randomly picked diseased subject exceeding that from a randomly picked nondiseased subject. Within the generalized linear model (GLM) binary regression
framework, she highlights a semiparametric estimator of ROC curve. Alonzo & Pepe [2002] consider an ROC model for which the ROC curve is a parametric function of covariates but distributions of the diagnostic test results are not specified. Those two methods are not maximum-likelihood method. Cai & Moskowitz [2004] propose the mle and pmle method for estimating the binormal parameters. The first method uses the profile likelihood and a simple algorithm to produce fully efficient estimates. The second method is based on a pseudo maximum-likelihood that can accommodate adjusting for covariates that could affect the accuracy of the continuous test.

An alternative semiparametric method is to estimate ROC curve under density ratio model. Qin & Zhang [1997] point out that the logistic regression model is equivalent to a two-sample semiparametric model in which the log ratio of two density functions is linear in data. Qin & Zhang [2003] further assume that the distribution of test results in the nondiseased subjects is unknown, but instead a functional form for the density (likelihood) ratio function relating the two distributions to each other. Rather than modeling these distribution functions, their semiparametric approach involves modeling the probability of disease status conditional on the test result. Wan & Zhang [2007] propose a semiparametric kernel distribution function estimator for ROC curve. Wang & Zhang [2014] propose a semiparametric empirical method to estimate the area under ROC curve.

In this Chapter, we propose a semiparametric estimator (called Pseudo Empirical Likelihood Non-Iterative method or pelni). The pelni method is based on the assumption that the data through an unknown transformation satisfy a density ratio model. This is different from the mle or pmle method proposed by Cai & Moskowitz [2004] which assumes a two-sample parametric model. Our pelni method has two advantages. First, the density ratio model in pelni method is more robust than the two-sample parametric model. Second, our pelni method provides a broader choices of distributions than other density ratio methods. The pelni method is an extension
of the density ratio model defined in Qin & Zhang [2003]. The density ratio in the pelni method depends on an unknown transformation \( h(x) \) rather than a known transformation. To simplify the calculation, pelni method uses a non-iterative procedure to estimate the unknown transformation \( h(x) \), which causes larger estimation error than other semiparametric methods, especially when sample sizes are small or the distributions of two samples are very separated. In Chapter 3, we propose another method to improve the estimation.

This Chapter is arranged as follows. The method is described in Section 2.2. Two real data examples are presented in Section 2.3. The simulation results are listed in Section 2.4. The summary is made in Section 2.5. Section 2.6 gives the theoretical proof.

### 2.2 Methodology

Let \( X_1, \cdots, X_{n_D} \) denote independent and identically distributed test results from a nondiseased population and, independently of the \( X_i \)'s, let \( Y_1, \cdots, Y_{n_D} \) be independent and identically distributed test results from the diseased population. Let \( T_1, \cdots, T_n \) denote the pooled test results \( Y_1, \cdots, Y_{n_D}, X_1, \cdots, X_{n_D} \) with \( n = n_D + n_D \).

Let \( x_i \)'s, \( y_i \)'s and \( t_i \)'s be the observations of \( X_i \)'s, \( Y_i \)'s and \( T_i \)'s respectively. Let \( G \) and \( F \) represent respectively the distribution functions of \( X_1 \) and \( Y_1 \). Let \( g \) and \( f \) represent the corresponding density functions.

Let \( h(x) = \Phi^{-1}(G(x)) \) be the unknown transformation function. Under the bi-normal model, the distributions of the transformed data are

\[
h(X_1) \sim N(0, 1), \quad h(Y_1) \sim N\left(\frac{a}{b}, \frac{1}{b^2}\right). \tag{2.2}
\]
On the other hand, since \( h(x) = \Phi^{-1}(G(x)) \) is a monotonically increasing function, we have

\[
h(X_1), \cdots, h(X_{n_D}) \sim G(h(x)) = \Phi(h(x)), \quad h(Y_1), \cdots, h(Y_{n_D}) \sim F(h(x)) = \Phi(bh(x) - a).
\]

Accordingly,

\[
g(x) = \frac{dG(h(x))}{dx} = \phi(h(x)) \frac{dh(x)}{dx},
\]

\[
f(x) = \frac{dF(h(x))}{dx} = b\phi(bh(x) - a) \frac{dh(x)}{dx}.
\]

The ratio of above two density functions is

\[
\frac{f(x)}{g(x)} = \frac{b\phi(bh(x) - a)}{\phi(h(x))} = w(h(x); a, b).
\]

Then we have the following two-sample density ratio model

\[
X_1, \cdots, X_{n_D} \overset{iid}{\sim} g(x), \quad Y_1, \cdots, Y_{n_D} \overset{iid}{\sim} f(x) = g(x)w(h(x); a, b),
\]

where

\[
w(h(x); a, b) = \exp \left\{ \log b - \frac{a^2}{2} + abh(x) + \frac{1 - b^2}{2}h(x)^2 \right\}.
\]

To estimate the unknown transformation \( h(\cdot) \), let \( \hat{G}(x) = \frac{1}{n_D} \sum_{k=1}^{n_D} I(x_k \leq x) \). It follows from (2.2) that as \( n \to \infty \), \( \hat{G}(x) \overset{p}{\to} \Phi(h_0(x)) \) for each fixed \( x \) and \( h_0(\cdot) \) is the true underlying transformation function. This motivates the following estimating equation for \( h(x) \):

\[
\hat{G}(x) = \Phi(h(x)).
\]

Let \( \hat{h}(x) \) be the solution of above equation at \( x \), i.e., \( \hat{h}(x) = \Phi^{-1}(\hat{G}(x)) \). Note that \( \hat{h}(x) \) is not a function of \( (a, b) \), therefore, no iteration is needed to obtain \( \hat{h}(x) \). Since, \(|\Phi^{-1}(x)| \to \infty \) when \( x = 0 \) or \( x = 1 \). We assign \( \hat{G}(\min(T)) = \frac{1}{2n} \) and
\( \hat{G}(\text{max}(T)) = 1 - \frac{1}{4n} \), instead of 0 and 1 respectively.

Next for a fixed \( \hat{h}(T) \), we use the score equations of \( \theta \) derived from the density ratio model to estimate \( \theta_0 \). The estimated empirical log-likelihood function is given by

\[
\hat{l} = \sum_{i=1}^{n} \log p_i + \sum_{j=1}^{n_D} \left\{ \log b - \frac{a^2}{2} + ab\hat{h}(y_j) + \frac{1 - b^2}{2} \hat{h}^2(y_j) \right\}.
\]

where \( p_i (i = 1, \ldots, n) \) is the probability mass on each observation assuming it is nondiseased. Using Lagrange multiplier method, we maximize the estimated empirical log-likelihood subject to constraints: \( \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i(w(\hat{h}(t_i); a, b)-1) = 0 \). As a result,

\[
\hat{p}_i = \frac{1}{n} \frac{1}{1 + \hat{\lambda}[w(\hat{h}(t_i); \hat{a}, \hat{b}) - 1]}.
\]

where \( (\hat{\lambda}, \hat{a}, \hat{b}) \) is the solution of the score equations

\[
\frac{\partial \hat{h}(\lambda, a, b)}{\partial \lambda} = -\sum_{i=1}^{n} \frac{w(\hat{h}(t_i); a, b)-1}{1+\hat{\lambda}[w(\hat{h}(t_i); a, b)-1]} = 0,
\]

\[
\frac{\partial \hat{h}(\lambda, a, b)}{\partial a} = \sum_{j=1}^{n_D} \left[ b\hat{h}(y_j) - a \right] - \sum_{i=1}^{n} \frac{\lambda w(\hat{h}(t_i); a, b) [\hat{h}(t_i) - a]}{1 + \hat{\lambda}[w(\hat{h}(t_i); a, b)-1]} = 0, \tag{2.4}
\]

\[
\frac{\partial \hat{h}(\lambda, a, b)}{\partial b} = \sum_{j=1}^{n_D} \left[ 1 + a\hat{h}(y_j) - \hat{b}^2(y_j) \right] - \sum_{i=1}^{n} \frac{\lambda w(\hat{h}(t_i); a, b) \left[ \frac{1}{2} + a\hat{h}(t_i) - \hat{b}^2(t_i) \right]}{1 + \hat{\lambda}[w(\hat{h}(t_i); a, b)-1]} = 0.
\]

Let

\[
V_{n1} = -\sum_{i=1}^{n} \frac{w(\hat{h}(t_i); a, b)-1}{1+\hat{\lambda}[w(\hat{h}(t_i); a, b)-1]},
\]

\[
V_{n2} = \sum_{j=1}^{n_D} \left[ b\hat{h}(y_j) - a \right] - \sum_{i=1}^{n} \frac{\lambda w(\hat{h}(t_i); a, b) [\hat{h}(t_i) - a]}{1 + \hat{\lambda}[w(\hat{h}(t_i); a, b)-1]},
\]

\[
V_{n3} = \sum_{j=1}^{n_D} \left[ 1 + a\hat{h}(y_j) - \hat{b}^2(y_j) \right] - \sum_{i=1}^{n} \frac{\lambda w(\hat{h}(t_i); a, b) \left[ \frac{1}{2} + a\hat{h}(t_i) - \hat{b}^2(t_i) \right]}{1 + \hat{\lambda}[w(\hat{h}(t_i); a, b)-1]}.
\]
The score equations for $\theta_0$ are

$$V_n(\theta; \hat{h}) = \begin{pmatrix} V_{n1} \\ V_{n2} \\ V_{n3} \end{pmatrix} = 0 \tag{2.5}$$

**Theorem 2.2.1.** Let $\hat{\theta}$ denote the resulting estimation for true value $\theta_0$ which can be solved from (2.5). $\hat{\theta}$ is consistent and $\sqrt{n}(\hat{\theta} - \theta_0)$ converges in distribution to a normal random vector with expectation of 0 and covariance matrix $\Sigma_{pelni}$. That is,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_{pelni})$$

where

$$\Sigma_{pelni} = \frac{1}{1 + \rho} A^{-1} \{ \rho \text{Var}(u_1(Y)) + \text{Var}(u_2(X)) \} (A^{-1})'$$

in which $\rho = \lim_{n \to \infty} \frac{\hat{n}_D}{n_D}$ is assumed a fixed number. $A$, $u_1$, $u_2$ are defined in section 2.6. The estimation of $\Sigma_{pelni}$ is obtained by replacing all theoretical quantities by their empirical counterparts, that is,

$$\hat{\Sigma}_{pelni} = \frac{1}{n} \hat{A}^{-1} \left\{ \sum_{j=1}^{n_D} \left[ \hat{u}_1(y_j) - \frac{1}{n_D} \sum_{j=1}^{n_D} \hat{u}_1(y_j) \right] \left[ \hat{u}_1(y_j) - \frac{1}{n_D} \sum_{j=1}^{n_D} \hat{u}_1(y_j) \right]' \right. \right.$$  
$$\left. + \sum_{k=1}^{n_D} \left[ \hat{u}_2(x_k) - \frac{1}{n_D} \sum_{k=1}^{n_D} \hat{u}_2(x_k) \right] \left[ \hat{u}_2(x_k) - \frac{1}{n_D} \sum_{k=1}^{n_D} \hat{u}_2(x_k) \right]' \right\} (\hat{A}^{-1})'$$

### 2.3 Examples

**Example 2.3.1.** As reported by Wieand et al. [1989], sera from $n_D = 51$ control patients with pancreatitis and $n_D = 90$ case patients with pancreatic cancer were studied at the Mayo Clinic with a carbohydrate antigen CA19-9. Here we apply our
methodology to estimate the ROC curve associated with CA19-9. The pelni estimates of $a$ and $b$ are $\hat{a} = 1.086$ ($\hat{SE} = 0.176$) and $\hat{b} = 0.405$ ($\hat{SE} = 0.073$). The estimate of the AUC is $\hat{d} = 0.843$.

Example 2.3.2. As reported by Cooper et al. [2014], the urine albumin and creatinine ratio are measured at enrollment at core lab for the 931 patients with renal artery stenosis in the CORAL clinical trial. At the end of study, $n_D = 330$ patients have the composite endpoint event. Here we apply our methodology to the treatment outcome associated with urine albumin creatinine ratio. The pelni estimates of $a$ and $b$ are $\hat{a} = 0.496$ ($\hat{SE} = 0.073$) and $\hat{b} = 0.894$ ($\hat{SE} = 0.052$). The estimate of AUC is $\hat{d} = 0.644$.

2.4 Simulations

Two simulation studies are conducted to compare the performance of pelni method with the following six methods:

- Semiparametric maximum likelihood method ($mle$) and profile maximum likelihood method ($pmle$) by Cai & Moskowitz [2004]


$$\hat{d} = \frac{1}{n_D n_D} \sum_{i=1}^{n_D} \sum_{j=1}^{n_D} I(Y_j \geq X_i), \quad \hat{R}(s) = 1 - \hat{F}(\hat{G}^{-1}(1 - s)), \quad s \in [0, 1],$$

According to DeLong et al. [1988] and [Pepe, 2003, p.106],

$$\widehat{\text{Var}}(\hat{d}) = \frac{\widehat{\text{Var}}(\hat{G}(Y))}{n_D} + \frac{\widehat{\text{Var}}(\hat{F}(X))}{n_D},$$
where

\[
\hat{G}(x) = \frac{1}{n_D} \sum_{i=1}^{n_D} I(X_i \leq x),
\]

\[
\hat{F}(x) = \frac{1}{n_D} \sum_{j=1}^{n_D} I(Y_j \leq x),
\]

\[
\hat{\text{Var}}(X) = s^2(X) = \frac{1}{n_D - 1} \sum_{i=1}^{n_D} (X_i - \bar{X})^2,
\]

\[
\hat{\text{Var}}(Y) = s^2(Y) = \frac{1}{n_D - 1} \sum_{i=1}^{n_D} (Y_i - \bar{Y})^2.
\]

Explicitly,

\[
\hat{\text{Var}}(\hat{d}) = \frac{\sum_{j=1}^{n_D} \left( \frac{1}{n_D} \sum_{i=1}^{n_D} I(X_i \leq Y_j) - \bar{G}(Y) \right)^2 + \sum_{i=1}^{n_D} \left( \frac{1}{n_D} \sum_{i=1}^{n_D} I(Y_j \leq X_i) - \bar{F}(X) \right)^2}{n_D(n_D - 1)} + \frac{\sum_{i=1}^{n_D} \left( \frac{1}{n_D} \sum_{i=1}^{n_D} I(Y_j \leq X_i) - \bar{F}(X) \right)^2}{n_D(n_D - 1)}.
\]

where

\[
\bar{G}(Y) = \frac{1}{n_D n_D} \sum_{j=1}^{n_D} \sum_{i=1}^{n_D} I(X_i \leq Y_j),
\]

\[
\bar{F}(X) = \frac{1}{n_D n_D} \sum_{j=1}^{n_D} \sum_{i=1}^{n_D} I(Y_j \leq X_i).
\]

- Parametric maximum likelihood normal method (mlnm), where

\[
\hat{a} = \frac{\bar{Y} - \bar{X}}{s(Y)}, \quad \hat{b} = \frac{s(X)}{s(Y)}, \quad \hat{d} = \Phi \left( \frac{\hat{a}}{\sqrt{1 + \hat{b}^2}} \right), \quad \hat{R}(c) = \Phi(\hat{a} + \hat{b} \Phi^{-1}(c)), \quad c \in [0,1],
\]

are the estimations of \((a, b, d, R)\), respectively. For the variance estimation, according to [Zhou et al., 2009, p.140],

\[
\hat{\text{Var}}(\hat{a}) = \frac{n_D(\hat{a}^2 + 2) + 2n_D \hat{b}^2}{2n_D n_D}, \quad \hat{\text{Var}}(\hat{b}) = \frac{(n_D + n_D)\hat{b}^2}{2n_D n_D}, \quad \hat{\text{Cov}}(\hat{a}, \hat{b}) = \frac{\hat{a} \hat{b}}{2n_D}.
\]
According to Obuchowski & McClish [1997], let \( \hat{c}_1 = \exp\left\{ -\frac{\hat{a}^2}{2(1+b^2)} \right\} \) and \( \hat{c}_2 = \frac{\hat{a}\hat{b}\exp\left\{ -\frac{\hat{a}^2}{2(1+b^2)} \right\}}{\sqrt{2\pi(1+b^2)}} \), then

\[
\hat{\text{Var}}(\hat{d}) = \hat{c}_1^2 \hat{\text{Var}}(\hat{a}) + \hat{c}_2^2 \hat{\text{Var}}(\hat{b}) + 2\hat{c}_1\hat{c}_2 \hat{\text{Cov}}(\hat{a}, \hat{b}).
\]

• *labroc* by Metz et al. [1998], which is a classical benchmark for ROC analysis.

In Chapter 3, let *labroc* denote the AUC estimation by trapezoidal approach.

In the first simulation study, we choose distributions in which the binormal model is correctly specified. Whereas, in the second simulation study, we choose the distributions in which the binormal model is misspecified. 1000 independent random samples are generated from the diseased and nondiseased group. For each sample, the binormal parameters \((a, b)\), area under the ROC curve \(d\), and ROC curve \(R(s)\) at \(s = \{0.1, 0.3, 0.6\}\) are estimated. The performance of each method is measured by bias (BS), standard deviation (SD), root of mean square error (rmse, defined by \(\text{rmse} = \sqrt{\text{BS}^2 + \text{SD}^2}\)), standard error (SE), coverage probability of confidence intervals (CP), average width of confidence intervals (WD) and relative efficiency (e). The relative efficiency between two methods is the ratio of two mean square errors, i.e.,

\[
e(A, B) = \frac{\text{MSE}(B)}{\text{MSE}(A)} = \frac{\text{BS}^2(B) + \text{SD}^2(B)}{\text{BS}^2(A) + \text{SD}^2(A)}.
\]

For example, \(e(A, B) > 1\) indicates that method A is more efficient than method B.

### 2.4.1 Correct specification

#### 2.4.1.1 Normal distribution

The distribution of \(X\) and \(Y\) are chosen to be normally distributed with different means and different variance. That is, \(X \sim N(\mu_D, \sigma_D^2)\) and \(Y \sim N(\mu_D, \sigma_D^2)\). The true parameters are chosen as \(\mu_D = 1\), \(\sigma_D = 1\) for the nondiseased subjects, and \(\mu_D = \{3.311, 2.517, 1.945\}\), \(\sigma_D = 2.25\) for the diseased subjects. By the AUC formula
for a binormal ROC curve \( d = \Phi(\frac{a}{\sqrt{1+b^2}}) \), the AUCs are \( \{0.9, 0.8, 0.7\} \). The parameter \( b = \frac{2}{3} \) is fixed. The simulation results are summarized in Tables 2.1, 2.2 and 2.3.

First we compare the \textit{pelni} method with \textit{mle}, \textit{pmle}, \textit{labroc} and \textit{mlnm} methods in terms of \((a,b)\) estimation. We choose sample sizes \((n_D, n_D) = (50, 50)\). \(e\) denote relative efficiency of each method relative to our proposed \textit{pelni} method. The relative efficiencies for the \textit{pelni} method are all 1, since the mean square error (MSE) from \textit{pelni} is the denominator to calculate the relative efficiency. The results are shown in Table 2.1. Further we compute the AUC and ROC curve at \( s \in \{0.1, 0.3, 0.6\} \) (see Table 2.2). Finally, we compare the large sample performance of the \textit{pelni}, \textit{mle} and \textit{pmle} method (see Table 2.3). We choose incremental sample sizes \((n_D, n_D) = (50, 50), (80, 80), (100, 100)\) and \((120, 120)\) with \(AUC = \{0.9, 0.8\}\).
Table 2.1: Estimation of \((a, b)\) from normal setup, sample sizes \((n_D, n_D) = (50, 50)\), 1000 simulations

<table>
<thead>
<tr>
<th>AUC</th>
<th>mle</th>
<th>pmle</th>
<th>pelni</th>
<th>labroc</th>
<th>mlnm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>0.9</td>
<td>e</td>
<td>1.007</td>
<td>0.942</td>
<td>1.045</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.321</td>
<td>0.191</td>
<td>0.327</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.113</td>
<td>0.063</td>
<td>0.110</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.300</td>
<td>0.180</td>
<td>0.308</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.279</td>
<td>0.176</td>
<td>0.270</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.967</td>
<td>0.945</td>
<td>0.927</td>
<td>0.819</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>1.095</td>
<td>0.719</td>
<td>1.060</td>
<td>0.553</td>
</tr>
<tr>
<td>0.8</td>
<td>e</td>
<td>0.749</td>
<td>0.780</td>
<td>0.743</td>
<td>0.749</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.222</td>
<td>0.137</td>
<td>0.221</td>
<td>0.135</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.069</td>
<td>0.045</td>
<td>0.046</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.211</td>
<td>0.129</td>
<td>0.216</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.212</td>
<td>0.135</td>
<td>0.208</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.962</td>
<td>0.953</td>
<td>0.947</td>
<td>0.884</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.832</td>
<td>0.541</td>
<td>0.817</td>
<td>0.455</td>
</tr>
<tr>
<td>0.7</td>
<td>e</td>
<td>0.588</td>
<td>0.808</td>
<td>0.571</td>
<td>0.789</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.194</td>
<td>0.121</td>
<td>0.191</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.054</td>
<td>0.047</td>
<td>0.025</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.186</td>
<td>0.111</td>
<td>0.190</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.190</td>
<td>0.119</td>
<td>0.187</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.960</td>
<td>0.954</td>
<td>0.954</td>
<td>0.905</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.745</td>
<td>0.477</td>
<td>0.733</td>
<td>0.417</td>
</tr>
</tbody>
</table>
Table 2.2: Estimation of AUC and ROC from normal setup, sample sizes 

\((n_D, n_D) = (50, 50)\), 1000 simulations

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>0.9</th>
<th>0.8</th>
<th>0.7</th>
<th>0.9</th>
<th>0.8</th>
<th>0.7</th>
<th>0.9</th>
<th>0.8</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>e</td>
<td>rmse</td>
<td>BS</td>
<td>SD</td>
<td>e</td>
<td>rmse</td>
<td>BS</td>
<td>SD</td>
<td>e</td>
<td>rmse</td>
</tr>
<tr>
<td>mle</td>
<td>d</td>
<td>0.772</td>
<td>0.030</td>
<td>0.004</td>
<td>0.030</td>
<td>0.588</td>
<td>0.043</td>
<td>0.007</td>
<td>0.042</td>
<td>0.478</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.996</td>
<td>0.069</td>
<td>0.005</td>
<td>0.068</td>
<td>0.979</td>
<td>0.082</td>
<td>0.003</td>
<td>0.082</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.824</td>
<td>0.043</td>
<td>0.008</td>
<td>0.042</td>
<td>0.718</td>
<td>0.061</td>
<td>0.010</td>
<td>0.060</td>
<td>0.634</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.563</td>
<td>0.024</td>
<td>0.004</td>
<td>0.023</td>
<td>0.438</td>
<td>0.041</td>
<td>0.010</td>
<td>0.040</td>
<td>0.398</td>
</tr>
<tr>
<td>pmle</td>
<td>d</td>
<td>0.813</td>
<td>0.031</td>
<td>0.000</td>
<td>0.031</td>
<td>0.602</td>
<td>0.043</td>
<td>0.001</td>
<td>0.043</td>
<td>0.484</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.012</td>
<td>0.069</td>
<td>-0.009</td>
<td>0.069</td>
<td>0.927</td>
<td>0.080</td>
<td>-0.010</td>
<td>0.079</td>
<td>0.867</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.888</td>
<td>0.044</td>
<td>0.004</td>
<td>0.044</td>
<td>0.731</td>
<td>0.061</td>
<td>0.002</td>
<td>0.061</td>
<td>0.624</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.606</td>
<td>0.024</td>
<td>0.004</td>
<td>0.024</td>
<td>0.467</td>
<td>0.042</td>
<td>0.006</td>
<td>0.042</td>
<td>0.414</td>
</tr>
<tr>
<td>pelni</td>
<td>d</td>
<td>1.000</td>
<td>0.034</td>
<td>-0.004</td>
<td>0.034</td>
<td>1.000</td>
<td>0.056</td>
<td>-0.013</td>
<td>0.055</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.000</td>
<td>0.069</td>
<td>-0.003</td>
<td>0.069</td>
<td>1.000</td>
<td>0.083</td>
<td>-0.007</td>
<td>0.082</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.000</td>
<td>0.047</td>
<td>-0.003</td>
<td>0.047</td>
<td>1.000</td>
<td>0.072</td>
<td>-0.013</td>
<td>0.070</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.000</td>
<td>0.031</td>
<td>-0.005</td>
<td>0.031</td>
<td>1.000</td>
<td>0.062</td>
<td>-0.016</td>
<td>0.060</td>
<td>1.000</td>
</tr>
<tr>
<td>labroc</td>
<td>d</td>
<td>0.792</td>
<td>0.030</td>
<td>0.002</td>
<td>0.030</td>
<td>0.609</td>
<td>0.044</td>
<td>0.003</td>
<td>0.044</td>
<td>0.494</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.967</td>
<td>0.068</td>
<td>0.009</td>
<td>0.067</td>
<td>0.968</td>
<td>0.081</td>
<td>0.010</td>
<td>0.081</td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.804</td>
<td>0.042</td>
<td>0.005</td>
<td>0.042</td>
<td>0.708</td>
<td>0.060</td>
<td>0.006</td>
<td>0.060</td>
<td>0.618</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.631</td>
<td>0.025</td>
<td>0.000</td>
<td>0.025</td>
<td>0.474</td>
<td>0.043</td>
<td>0.001</td>
<td>0.043</td>
<td>0.418</td>
</tr>
<tr>
<td>mlnm</td>
<td>d</td>
<td>0.749</td>
<td>0.029</td>
<td>0.001</td>
<td>0.029</td>
<td>0.585</td>
<td>0.043</td>
<td>0.002</td>
<td>0.043</td>
<td>0.478</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.834</td>
<td>0.063</td>
<td>0.004</td>
<td>0.063</td>
<td>0.843</td>
<td>0.076</td>
<td>0.005</td>
<td>0.076</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.710</td>
<td>0.040</td>
<td>0.001</td>
<td>0.040</td>
<td>0.673</td>
<td>0.059</td>
<td>0.003</td>
<td>0.059</td>
<td>0.596</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.479</td>
<td>0.022</td>
<td>-0.003</td>
<td>0.022</td>
<td>0.419</td>
<td>0.040</td>
<td>0.000</td>
<td>0.040</td>
<td>0.391</td>
</tr>
<tr>
<td>np</td>
<td>d</td>
<td>0.831</td>
<td>0.031</td>
<td>0.000</td>
<td>0.031</td>
<td>0.621</td>
<td>0.044</td>
<td>0.000</td>
<td>0.044</td>
<td>0.502</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.209</td>
<td>0.076</td>
<td>0.010</td>
<td>0.075</td>
<td>1.250</td>
<td>0.092</td>
<td>0.016</td>
<td>0.091</td>
<td>1.241</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.189</td>
<td>0.051</td>
<td>0.002</td>
<td>0.051</td>
<td>1.048</td>
<td>0.073</td>
<td>0.005</td>
<td>0.073</td>
<td>0.898</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.973</td>
<td>0.031</td>
<td>0.000</td>
<td>0.031</td>
<td>0.666</td>
<td>0.051</td>
<td>0.002</td>
<td>0.051</td>
<td>0.584</td>
</tr>
</tbody>
</table>
Table 2.3: Estimation of \((a, b)\) from normal distribution, incremental sample sizes, 1000 simulations

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>0.9</th>
<th>0.8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rmse</td>
<td>BS</td>
<td>SD</td>
<td>SE</td>
</tr>
<tr>
<td>(50,50)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mle</td>
<td>a</td>
<td>0.321</td>
<td>0.113</td>
<td>0.300</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.191</td>
<td>0.063</td>
<td>0.180</td>
</tr>
<tr>
<td>pmle</td>
<td>a</td>
<td>0.327</td>
<td>0.110</td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.193</td>
<td>0.095</td>
<td>0.168</td>
</tr>
<tr>
<td>pelni</td>
<td>a</td>
<td>0.320</td>
<td>0.036</td>
<td>0.318</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.197</td>
<td>0.024</td>
<td>0.195</td>
</tr>
<tr>
<td>(80,80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mle</td>
<td>a</td>
<td>0.223</td>
<td>0.057</td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.177</td>
<td>0.036</td>
<td>0.132</td>
</tr>
<tr>
<td>pmle</td>
<td>a</td>
<td>0.227</td>
<td>0.055</td>
<td>0.220</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.139</td>
<td>0.058</td>
<td>0.126</td>
</tr>
<tr>
<td>pelni</td>
<td>a</td>
<td>0.226</td>
<td>0.008</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.142</td>
<td>0.009</td>
<td>0.142</td>
</tr>
<tr>
<td>(100,100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mle</td>
<td>a</td>
<td>0.201</td>
<td>0.062</td>
<td>0.191</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.117</td>
<td>0.032</td>
<td>0.113</td>
</tr>
<tr>
<td>pmle</td>
<td>a</td>
<td>0.205</td>
<td>0.063</td>
<td>0.195</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.121</td>
<td>0.051</td>
<td>0.109</td>
</tr>
<tr>
<td>pelni</td>
<td>a</td>
<td>0.201</td>
<td>0.024</td>
<td>0.199</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.122</td>
<td>0.010</td>
<td>0.122</td>
</tr>
<tr>
<td>(120,120)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mle</td>
<td>a</td>
<td>0.189</td>
<td>0.053</td>
<td>0.181</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.111</td>
<td>0.026</td>
<td>0.108</td>
</tr>
<tr>
<td>pmle</td>
<td>a</td>
<td>0.192</td>
<td>0.052</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.114</td>
<td>0.040</td>
<td>0.106</td>
</tr>
<tr>
<td>pelni</td>
<td>a</td>
<td>0.190</td>
<td>0.019</td>
<td>0.189</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>0.115</td>
<td>0.006</td>
<td>0.115</td>
</tr>
</tbody>
</table>

From Table 2.1, we can see that

1. All biases are close to 0 which indicates all estimators of \((a, b)\) are consistent.

2. As we expected, the relative efficiency of the \textit{mlnm} method is the smallest, since it is correctly specified parametric method.

3. Among all semiparametric methods, the \textit{labroc} method is consistently better than the \textit{mle}, \textit{pmle} and \textit{pelni} methods in terms of relative efficiency.
4. For high value of AUC, the relative efficiency of *pelni* method is close to that of *mle* and *pmle* method. For example, when $AUC = 0.9$, the relative efficiencies of *mle* and *pmle* are $(100.7\%, 94.2\%)$ and $(104.5\%, 96.1\%)$, respectively. However, as AUC decreases, the *pelni* performs worse than *mle* and *pmle* method. When $AUC = 0.8$, the relative efficiencies of the *mle* and *pmle* method decrease to $(72.9\%, 78\%)$ and $(74.3\%, 74.9\%)$. When $AUC = 0.7$, the relative efficiencies of the *mle* and *pmle* method for parameter $a$ decreases to $58.8\%$ and $57.1\%$, respectively.

5. For both *mle* and *labroc* estimator of $(a, b)$, the standard errors are very close to the standard deviation. The standard errors of the *pmle* and *pelni* estimators are always smaller than the corresponding standard deviation.

6. For high or medium values of AUC, the width of the 95% confidence interval for parameter $a$ is the smallest for the *pelni* method. However, for small value of AUC, the *pmle* method gives the narrowest confidence interval for parameter $a$. The average width of the 95% confidence intervals for the parameter $b$ from the *pmle* method is the smallest for all choices of AUC.

7. The *mle* and *pmle* methods are better than *pelni* in terms of coverage probability. The coverage probability of *pelni* method is far from the nominal level 95%, especially for parameter $b$. When $AUC = \{0.9, 0.8, 0.7\}$, the coverage probabilities are $\{0.906, 0.914, 0.908\}$ for parameter $a$, and $\{0.860, 0.855, 0.879\}$ for parameter $b$, respectively.

From Table 2.2, we can see that

1. The *mlnm* method is the correct specification case, therefore, yields the best results in terms of relative efficiency.
2. The *pelni* is more efficient than the *np* method in terms of ROC curve estimation at high and medium specificities \((s = \{0.1, 0.3\})\). When \(AUC = 0.9\), the relative efficiency are \(\{120.9\%, 118.9\%\}\) for ROC curve at \(s = \{0.1, 0.3\}\). When \(AUC = 0.7\), the relative efficiency are \(\{124.1\%, 89.8\%\}\) for ROC curve at \(s = \{0.1, 0.3\}\). However, in terms of estimation for AUC, the *pelni* method is less efficient than the *np* method. The relative efficiencies are ranging between 83.1% at \(AUC = 0.9\) and 50.2% at \(AUC = 0.7\).

3. In terms of ROC curve at \(s = 0.1\), the *pelni* method is close to the *mle*, *pmle* and *labroc* method. The relative efficiency for ROC curve at \(s = 0.1\) is close to 1. However, *pelni* method is less efficient than *mle*, *pmle* and *labroc* method for estimation of \(d\) and ROC curve at \(s = \{0.3, 0.6\}\). The relative efficiency decreases when AUC decreases.

From Table 2.3, we can see that

1. The standard errors of the *pmle* and *pelni* method underestimate the standard deviation. In particular, for small sample sizes \((n_D, n_D) = (50, 50)\) and \(AUC = 0.9\), the percent differences \(\left(\frac{SD - SE}{SD}\right)\) between the standard deviation and the standard error of \((a, b)\), are \((21\%, 27\%)\) for the *pelni* method, and \((12\%, 18\%)\) for the *pmle* method.

2. The percent differences between the standard deviation and the standard error of \((a, b)\) decrease when sample sizes increase. For example, when the sample sizes increase to \((n_D, n_D) = (120, 120)\) at \(AUC = 0.9\), the percent differences decrease to \((15\%, 17\%)\) for the *pelni* method, and \((9\%, 14\%)\) for the *pmle* method.

3. The percent difference between the standard deviation and the standard error of \((a, b)\) decrease when AUC decreases. For example, the percent differences reduce even further to \((12\%, 15\%)\) for *pelni* method and \((4\%, 9\%)\) for *pmle* method at \(AUC = 0.8\) and sample sizes \((n_D, n_D) = (120, 120)\).
2.4.1.2 Lognormal distribution

The distribution of $X$ and $Y$ are chosen to have lognormal distribution with $\log X \sim N(1, 1)$, $\log Y \sim N(\mu_D, 2.25)$ and $\mu_D = \{3.311, 2.517, 1.945\}$. The true parameter $b = \frac{2}{3}$ and $AUC = \{0.9, 0.8, 0.7\}$. The simulation results are presented in Tables 2.4, 2.5 and 2.6. By comparing the performance measures of all estimators, similar findings are observed.

Table 2.4: Estimation of $(a, b)$ from lognormal distribution, sample sizes $(n_D, n_D) = (50, 50)$, 1000 simulations

<table>
<thead>
<tr>
<th>AUC</th>
<th>mle</th>
<th>pmle</th>
<th>pelni</th>
<th>labroc</th>
<th>mlmm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>0.9</td>
<td>e</td>
<td>1.006</td>
<td>0.948</td>
<td>1.034</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.343</td>
<td>0.194</td>
<td>0.348</td>
<td>0.195</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.109</td>
<td>0.061</td>
<td>0.196</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.325</td>
<td>0.184</td>
<td>0.331</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.279</td>
<td>0.176</td>
<td>0.269</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.959</td>
<td>0.956</td>
<td>0.921</td>
<td>0.819</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>1.095</td>
<td>0.718</td>
<td>1.054</td>
<td>0.548</td>
</tr>
<tr>
<td>0.8</td>
<td>e</td>
<td>0.744</td>
<td>0.788</td>
<td>0.740</td>
<td>0.753</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.233</td>
<td>0.139</td>
<td>0.232</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.060</td>
<td>0.043</td>
<td>0.038</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.225</td>
<td>0.133</td>
<td>0.229</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.212</td>
<td>0.134</td>
<td>0.208</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.951</td>
<td>0.953</td>
<td>0.941</td>
<td>0.889</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.830</td>
<td>0.538</td>
<td>0.815</td>
<td>0.454</td>
</tr>
<tr>
<td>0.7</td>
<td>e</td>
<td>0.572</td>
<td>0.840</td>
<td>0.563</td>
<td>0.770</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.199</td>
<td>0.123</td>
<td>0.198</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.046</td>
<td>0.045</td>
<td>0.016</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.194</td>
<td>0.115</td>
<td>0.197</td>
<td>0.111</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.190</td>
<td>0.119</td>
<td>0.187</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.949</td>
<td>0.954</td>
<td>0.950</td>
<td>0.911</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.743</td>
<td>0.475</td>
<td>0.732</td>
<td>0.415</td>
</tr>
</tbody>
</table>
Table 2.5: Estimation of AUC and ROC from lognormal distribution, sample sizes \((n_D, n_d) = (50, 50)\), 1000 simulations

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>0.9</th>
<th></th>
<th>0.8</th>
<th></th>
<th>0.7</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>e  rms</td>
<td>BS</td>
<td>SD</td>
<td>e  rms</td>
<td>BS</td>
</tr>
<tr>
<td>mle</td>
<td>d</td>
<td>0.780</td>
<td>0.032</td>
<td>0.003</td>
<td>0.032</td>
<td>0.580</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.011</td>
<td>0.071</td>
<td>0.004</td>
<td>0.071</td>
<td>0.965</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.825</td>
<td>0.045</td>
<td>0.007</td>
<td>0.045</td>
<td>0.702</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.571</td>
<td>0.025</td>
<td>0.003</td>
<td>0.025</td>
<td>0.440</td>
<td>0.043</td>
</tr>
<tr>
<td>pmle</td>
<td>d</td>
<td>0.824</td>
<td>0.033</td>
<td>-0.003</td>
<td>0.033</td>
<td>0.601</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.034</td>
<td>0.071</td>
<td>-0.010</td>
<td>0.071</td>
<td>0.927</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.889</td>
<td>0.047</td>
<td>0.003</td>
<td>0.047</td>
<td>0.722</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.616</td>
<td>0.026</td>
<td>0.003</td>
<td>0.026</td>
<td>0.472</td>
<td>0.045</td>
</tr>
<tr>
<td>pelni</td>
<td>d</td>
<td>1.000</td>
<td>0.036</td>
<td>-0.005</td>
<td>0.036</td>
<td>1.000</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.000</td>
<td>0.070</td>
<td>-0.004</td>
<td>0.070</td>
<td>1.000</td>
<td>0.084</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.000</td>
<td>0.050</td>
<td>-0.004</td>
<td>0.050</td>
<td>1.000</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.000</td>
<td>0.033</td>
<td>-0.006</td>
<td>0.032</td>
<td>1.000</td>
<td>0.065</td>
</tr>
<tr>
<td>labroc</td>
<td>d</td>
<td>0.811</td>
<td>0.032</td>
<td>0.001</td>
<td>0.032</td>
<td>0.606</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.978</td>
<td>0.070</td>
<td>0.007</td>
<td>0.069</td>
<td>0.938</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.819</td>
<td>0.045</td>
<td>0.003</td>
<td>0.045</td>
<td>0.697</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.660</td>
<td>0.027</td>
<td>-0.002</td>
<td>0.027</td>
<td>0.490</td>
<td>0.045</td>
</tr>
<tr>
<td>mlnm</td>
<td>d</td>
<td>0.759</td>
<td>0.031</td>
<td>0.009</td>
<td>0.031</td>
<td>0.588</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.867</td>
<td>0.065</td>
<td>0.003</td>
<td>0.065</td>
<td>0.833</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.735</td>
<td>0.043</td>
<td>0.009</td>
<td>0.043</td>
<td>0.673</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.515</td>
<td>0.024</td>
<td>-0.002</td>
<td>0.023</td>
<td>0.440</td>
<td>0.043</td>
</tr>
<tr>
<td>np</td>
<td>d</td>
<td>0.834</td>
<td>0.033</td>
<td>-0.001</td>
<td>0.033</td>
<td>0.619</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.246</td>
<td>0.078</td>
<td>0.011</td>
<td>0.078</td>
<td>1.223</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.141</td>
<td>0.053</td>
<td>0.009</td>
<td>0.053</td>
<td>0.965</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.961</td>
<td>0.032</td>
<td>-0.001</td>
<td>0.032</td>
<td>0.670</td>
<td>0.053</td>
</tr>
</tbody>
</table>
Table 2.6: Estimation of \((a, b)\) from lognormal distribution, incremental sample sizes, 1000 simulations

<table>
<thead>
<tr>
<th>AUC</th>
<th>0.9</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rmse</td>
<td>BS</td>
</tr>
<tr>
<td>(50,50)</td>
<td>mle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pmle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pelni</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td>(80,80)</td>
<td>mle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pmle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pelni</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td>(100,100)</td>
<td>mle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pmle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pelni</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td>(120,120)</td>
<td>mle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pmle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pelni</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
</tbody>
</table>

2.4.2 Misspecification

We consider two cases of misspecification. In the first case, we choose exponential distribution and in the second case, we choose Gamma distribution.
2.4.2.1 Exponential distribution

$X$ and $Y$ are chosen to have exponential distributions with different rates, i.e., $X \sim \text{Exp}(r_D)$ and independently, $Y \sim \text{Exp}(r_D)$. The parameters are $r_D = 1$ and $r_D = \{0.1, 0.25, 0.43\}$. By the AUC formula for exponential setup $d = \frac{r_D}{r_D + r_D}$, we obtain the corresponding $AUC = \{0.91, 0.8, 0.7\}$. The sample sizes for the diseased and nondiseased are both chosen to be 50. Each setup is simulated for 1000 times. The simulation results are listed in Table 2.7. We list the results for the semiparametric, parametric and nonparametric methods. Estimations are made for the area under the curve $d$, ROC curve at $s = \{0.1, 0.3, 0.6\}$. For each estimation, we provide bias, standard deviation, $rmse$ and the relative efficiency $e$ (which is relative to corresponding results from pelni method). The relative efficiencies for the pelni method are all 1, since the mean square error (MSE) from pelni is the denominator to calculate the relative efficiency.
Table 2.7: Estimation of AUC and ROC from exponential distribution, sample sizes \((n_D, n_D) = (50, 50)\), 1000 simulations

<table>
<thead>
<tr>
<th>AUC</th>
<th>0.9</th>
<th>0.8</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>e</td>
<td>rmse</td>
<td>BS</td>
</tr>
<tr>
<td>mle</td>
<td>d</td>
<td>0.604</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.956</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.734</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.449</td>
<td>0.023</td>
</tr>
<tr>
<td>pmle</td>
<td>d</td>
<td>0.634</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.014</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.787</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.502</td>
<td>0.025</td>
</tr>
<tr>
<td>pelni</td>
<td>d</td>
<td>1.000</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.000</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.000</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.000</td>
<td>0.035</td>
</tr>
<tr>
<td>labroc</td>
<td>d</td>
<td>0.679</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.019</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.765</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.515</td>
<td>0.025</td>
</tr>
<tr>
<td>mlnm</td>
<td>d</td>
<td>0.972</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>4.033</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.393</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.983</td>
<td>0.034</td>
</tr>
<tr>
<td>np</td>
<td>d</td>
<td>0.751</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.251</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.172</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.898</td>
<td>0.033</td>
</tr>
</tbody>
</table>

We observe from Table 2.7 that

1. The relative efficiency of \(mlnm\) method is not stable and the value is very large for \(R(0.1)\).

2. The \(pelni\) is more efficient than the \(np\) method in terms of ROC curve estimation at high and medium specificities \((s = \{0.1, 0.3\})\). When \(AUC = 0.9\), the relative efficiency are \(\{125.1\%, 117.2\%\}\) for ROC curve at \(s = \{0.1, 0.3\}\).
$AUC = 0.7$, the relative efficiency are $\{135.4\%, 124.8\%\}$ for ROC curve at $s = \{0.1, 0.3\}$. However, in term of estimation for $d$, the $pelni$ method is less efficient than the $np$ method. The relative efficiency is around $75\%$ at different AUCs.

3. In term of ROC curve at $s = 0.1$, the $pelni$ method is close to $mle$, $pmle$ and $labroc$ method. The relative efficiency for ROC curve at $s = 0.1$ is close to $1$. However, $pelni$ method is less efficient than $mle$, $pmle$ and $labroc$ method for estimation of $d$ and ROC curve at $s = \{0.3, 0.6\}$. Opposite to the previous normal and lognormal setups, the relative efficiency increases when AUC decreases, which means the efficiency of our $pelni$ method increases as the AUC decreases.

2.4.2.2 Gamma distribution

$X$ and $Y$ are chosen to have Gamma distributions with different rates, i.e., $X \sim \Gamma(1, 0.5)$ and independently, $Y \sim \Gamma(\alpha_D, 0.5)$. The parameters are $\alpha_D = \{3.25, 2.5, 1.75\}$. By the definition of AUC, we calculate the corresponding $AUC = \{0.895, 0.823, 0.703\}$. The sample sizes for the diseased and nondiseased are both chosen to be $50$. Each setup is simulated for $1000$ times. The simulation results are listed in Table 2.8. We list the results for the semiparametric, parametric and nonparametric methods. Estimations are made for the area under the curve $d$, ROC curves at $s = \{0.1, 0.3, 0.6\}$. For each estimation, we provide bias, standard deviation, $rmse$ and the relative efficiency $e$ (which is relative to corresponding results from $pelni$ method). The relative efficiencies for the $pelni$ method are all $1$, since the mean square error (MSE) from $pelni$ is the denominator to calculate the relative efficiency.
Table 2.8: Estimation of AUC and ROC from Gamma distribution, sample sizes \((n_D, n_D) = (50, 50)\), 1000 simulations

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>0.895</th>
<th>0.823</th>
<th>0.703</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>e</td>
<td>rmse</td>
<td>BS</td>
<td>SD</td>
</tr>
<tr>
<td>mle</td>
<td>d</td>
<td>0.931</td>
<td>0.030</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.956</td>
<td>0.107</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.938</td>
<td>0.044</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.830</td>
<td>0.011</td>
<td>-0.002</td>
</tr>
<tr>
<td>pmle</td>
<td>d</td>
<td>0.961</td>
<td>0.030</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.912</td>
<td>0.105</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.960</td>
<td>0.045</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.921</td>
<td>0.011</td>
<td>-0.003</td>
</tr>
<tr>
<td>pelni</td>
<td>d</td>
<td>1.000</td>
<td>0.031</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.000</td>
<td>0.110</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.000</td>
<td>0.046</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.000</td>
<td>0.012</td>
<td>-0.003</td>
</tr>
<tr>
<td>labroc</td>
<td>d</td>
<td>0.934</td>
<td>0.030</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.899</td>
<td>0.104</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.939</td>
<td>0.044</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.051</td>
<td>0.012</td>
<td>-0.004</td>
</tr>
<tr>
<td>mlnm</td>
<td>d</td>
<td>2.061</td>
<td>0.045</td>
<td>-0.029</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.790</td>
<td>0.098</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>3.602</td>
<td>0.087</td>
<td>-0.074</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>42.335</td>
<td>0.076</td>
<td>-0.071</td>
</tr>
<tr>
<td>np</td>
<td>d</td>
<td>0.999</td>
<td>0.031</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.160</td>
<td>0.118</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.487</td>
<td>0.056</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.704</td>
<td>0.015</td>
<td>-0.001</td>
</tr>
</tbody>
</table>

From Table 2.8, we can see

1. The parametric mlnm method yields consistent result since the parametric model is misspecified. The relative efficiency are not stable due to large bias.

2. pelni method is more efficient than the np method for \(d\) and ROC estimations at \(s = \{0.1, 0.3, 0.6\}\). For example, when \(AUC = 0.895\), the relative efficiencies are 99.9%, 116%, 148.7% and 170.4% for \(d\) and ROC at \(s = \{0.1, 0.3, 0.6\}\).
When $AUC = 0.703$, the relative efficiencies are 106.7%, 153.2%, 146% and 136% for $d$ and ROC at $s = \{0.1, 0.3, 0.6\}$.

3. The $pelni$ method is closer to the $mle$, $pmle$ and $labroc$ methods in terms of relative efficiency, comparing with previously setups of normal, lognormal and exponential distributions. For example, when $AUC = 0.895$, the relative efficiencies for $d$ and ROC curve at $s = \{0.1, 0.3\}$ are above 90%. When $AUC = 0.703$, for the relative efficiencies of $d$ and ROC at $s = \{0.1, 0.3\}$ are, $\{98.1\%, 107.6\%, 104.3\%\}$ for $mle$ method, $\{100\%, 101\%, 101.6\%\}$ for $pmle$ method, $\{104.1\%, 112.8\%, 107.5\%\}$ for $labroc$ method. Therefore, at low AUC, the $pelni$ method is close to the $mle$, $pmle$ method, and slightly than the $labroc$ method.

2.5 Summary

Based on all the simulation results, we can see that

1. Our proposed $pelni$ method is a consistent estimator for $(a, b)$

2. The $pelni$ method is generally less efficient than $mle$, $pmle$ and $labroc$ method, although it is as efficient as $mle$ and $pmle$ method when AUC is large.

3. The standard error of $(a, b)$ estimator from $pelni$ method is smaller than the standard deviation, especially for high value of AUC and small sample size.

4. The ROC curve estimated by our $pelni$ method is more efficient than that estimated by nonparametric method for large and median value of specificity, when the data is normal, lognormal, exponential and Gamma distributed.

5. Although $pelni$ method is more robust than the parametric methods, it is less efficient than other semiparametric methods when the binormal model data are correctly specified.
6. When the binormal model data are misspecified, \textit{pelni} method yields some estimations close to the other semiparametric method. For example, when the data has exponential distribution, the estimation of ROC curve at high specificity from \textit{pelni} method is very close to \textit{mle}, \textit{pmle} and \textit{labroc} methods. When the data has Gamma distribution and the AUC is median or low, the \textit{pelni} method is more efficient than the \textit{labroc} method in terms of AUC and ROC estimation.

2.6 Proofs

\textbf{Lemma 2.6.1.} Let $V_n(\theta; \hat{h})$ be defined in (2.5). As $n \to \infty$,

$$\frac{1}{n} V_n(\theta_0; \hat{h}) \overset{p}{\to} 0.$$ 

\textit{Proof.} Refer to the proof of Lemma 3.7.2. \hfill \Box

\textbf{Lemma 2.6.2.} Let $\hat{h}(x)$ be the solution of equation

$$\hat{H}(x) = H(h(x))$$

or explicitly

$$\hat{G}(x) = \Phi(h(x))$$

Then, $\hat{h}(x)$ has the closed form

$$\hat{h}(x) = \Phi^{-1}(\hat{G}(x))$$

and is no longer a function of $(a, b)$, as in our \textit{pelni} method. We have

$$\hat{h}(x) \overset{p}{\to} h_0(x), \text{ for each fixed } x.$$
Proof. According to Weak Law of Large Numbers, since $\Phi^{-1}$ is a continuous function, $
abla h(x) \xrightarrow{p} h_0(x)$ for each fixed $x$. \qed

Lemma 2.6.3. Let

$$A = - \lim_{n \to \infty} \frac{1}{n} \frac{\partial V_n(\theta, \hat{h})}{\partial \theta} |_{\theta_0}$$

and assume $\rho = \lim_{n \to \infty} \frac{n_0}{n^\rho}$ and $\lambda_0 = \frac{\rho}{1+\rho}$ are fixed numbers. The matrix of $A$ is denoted by

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

and it is a symmetric matrix.

Proof. Let $\theta = (\lambda, a, b)^T$ be the solution of score equation (2.5). Let $\theta_0 = (\lambda_0, a_0, b_0)^T$ be the true values of $\theta$, $h_0(x)$ be the true transformation function. $\hat{h}(x)$ is defined in Lemma 2.6.2. Define

$$\begin{cases} A_{-k} = \int h_0(x)^k dG(x), & k = 1, \cdots, 4, \\ A_k = \int \frac{h_0(x)^k}{1+\lambda_0[\hat{w}(h_0(x); \theta_0) - 1]} dG(x), & k = 0, \cdots, 4, \end{cases}$$

Clearly,

$$A_{-1} = 0, \quad A_{-2} = 1, \quad A_{-3} = 0, \quad A_{-4} = 3.$$ 

Furthermore, we define

$$U(\hat{h}(x); \theta) = \frac{1 + \lambda_0[\hat{w}(h_0(x); \theta_0) - 1]}{1 + \lambda[\hat{w}(\hat{h}(x); \theta) - 1]}$$

and

$$u(\hat{h}(x); \theta) = \log \hat{w}(\hat{h}(x); \theta) = \log b - \frac{a^2}{2} + ab\hat{h}(x) + \frac{1 - b^2}{2} \hat{h}(x)^2$$

41
Recall

\[
\begin{align*}
\forall & g_1(x; \theta) = \frac{1}{\lambda} \left\{ \frac{1+\lambda_0[w(h_0(x); \theta_0)]}{1+\lambda[w(h(x); \theta)]} - 1 \right\} \\
\forall & g_2(x; \theta) = (1 - \lambda)\left[b\hat{h}(x) - a\right]\left\{ \frac{1+\lambda_0[w(h_0(x); \theta_0)]}{1+\lambda[w(h(x); \theta)]} - \frac{1-\lambda_0}{1-\lambda} \right\} \\
\forall & g_3(x; \theta) = (1 - \lambda)\left[\frac{1}{b} + a\hat{h}(x) - b\hat{h}(x)^2\right]\left\{ \frac{1+\lambda_0[w(h_0(x); \theta_0)]}{1+\lambda[w(h(x); \theta)]} - \frac{1-\lambda_0}{1-\lambda} \right\}
\end{align*}
\]

Using the notation of \( U(\hat{h}(x); \theta) \), we have

\[
\begin{align*}
\forall & g_1(x; \theta) = \frac{1}{\lambda} \left\{ U(\hat{h}(x); \theta) - 1 \right\} \\
\forall & g_2(x; \theta) = (1 - \lambda)\left[b\hat{h}(x) - a\right]\left\{ U(\hat{h}(x); \theta) - \frac{1-\lambda_0}{1-\lambda} \right\} \quad (2.6) \\
\forall & g_3(x; \theta) = (1 - \lambda)\left[\frac{1}{b} + a\hat{h}(x) - b\hat{h}(x)^2\right]\left\{ U(\hat{h}(x); \theta) - \frac{1-\lambda_0}{1-\lambda} \right\}
\end{align*}
\]

Notice that, as \( n \to \infty \)

\[
U(\hat{h}(x); \theta)|_{\theta_0} \xrightarrow{P} 1
\]

Therefore,

\[
g_1(x; \theta_0) \xrightarrow{P} 0, \quad g_2(x; \theta_0) \xrightarrow{P} 0, \quad g_3(x; \theta_0) \xrightarrow{P} 0,
\]

and later on when we calculate \( \lim_{n \to \infty} \frac{\partial g_i(x; \theta)}{\partial \theta} |_{\theta_0} \), \( i = 1, 2, 3 \), the derivative on the part before the big parenthesis in (2.6) vanishes. Thus we only need to calculate the derivative on the parenthesis part. Next, we derive \( \frac{\partial U(\hat{h}(x); \theta)}{\partial \theta} |_{\theta_0} \), which is relatively simpler than the \textit{pel} method in Chapter 3, since \( \hat{h}(x) \) is no longer a function of \((a, b)\).

We have, as \( n \to \infty \),

\[
\frac{\partial u(\hat{h}(x); \theta)}{\partial \lambda} |_{\theta_0} \xrightarrow{P} 0
\]

and

\[
\frac{\partial u(\hat{h}(x); \theta)}{\partial a} |_{\theta_0} = \left\{ -a + b\hat{h}(x) \right\} |_{\theta_0} \xrightarrow{P} -a_0 + b_0 h_0(x)
\]
and
\[
\frac{\partial u(\hat{h}(x); \theta)}{\partial b}|_{\theta_0} = \left\{ \frac{1}{b} + a\hat{h}(x) - b\hat{h}(x)^2 \right\}|_{\theta_0} \overset{p}{\to} \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0(x)^2
\]

Further, we have
\[
\frac{\partial U(\hat{h}(x); \theta)}{\partial \lambda}|_{\theta_0} \overset{p}{\to} -\frac{1}{\lambda_0} \left( 1 - \frac{1}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right)
\]

and
\[
\frac{\partial U(\hat{h}(x); \theta)}{\partial a}|_{\theta_0} \overset{p}{\to} -\left( 1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) \frac{\partial u(\hat{h}(x); \theta)}{\partial a}|_{\theta_0} \left[ -a_0 + b_0 h_0(x) \right]
\]

and
\[
\frac{\partial U(\hat{h}(x); \theta)}{\partial b}|_{\theta_0} \overset{p}{\to} -\left( 1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) \frac{\partial u(\hat{h}(x); \theta)}{\partial b}|_{\theta_0} \left[ \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0(x)^2 \right]
\]

To derive the probability limit of \(-\frac{1}{n} \frac{\partial V_{n1}(\hat{h}; \theta)}{\partial \theta}|_{\theta_0}\), notice that
\[
-\frac{1}{n} \frac{\partial V_{n1}(\hat{h}; \theta)}{\partial \theta}|_{\theta_0} \overset{p}{\to} -\int_{n \to \infty} \frac{\partial g_1(x, \theta)}{\partial \theta}|_{\theta_0} dG(x)
\]
Since
\[
\frac{\partial g_1(x, \theta)}{\partial \theta} \bigg|_{\theta_0} = \frac{1}{\lambda} \left( \begin{array}{c}
\frac{\partial U(h(x); \theta)}{\partial \lambda} \\
\frac{\partial U(h(x); \theta)}{\partial a} \\
\frac{\partial U(h(x); \theta)}{\partial b}
\end{array} \right) \bigg|_{\theta_0}
\]
we have,
\[
a_{11} = - \int \lim_{n \to \infty} \frac{\partial g_1(x, \theta)}{\partial \lambda} \big|_{\theta_0} dG(x) = - \int \lim_{n \to \infty} \frac{1}{\lambda} \frac{\partial U(h(x); \theta)}{\partial \lambda} \big|_{\theta_0} dG(x)
= - \int \frac{1}{\lambda_0} \left\{ - \frac{1}{\lambda_0} \left( 1 - \frac{1}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right) \right\} dG(x)
= \frac{1 - A_0}{\lambda_0^2}
\]
and
\[
a_{12} = - \int \lim_{n \to \infty} \frac{\partial g_1(x, \theta)}{\partial a} \big|_{\theta_0} dG(x) = - \int \lim_{n \to \infty} \frac{1}{\lambda} \frac{\partial U(h(x); \theta)}{\partial a} \big|_{\theta_0} dG(x)
= - \int \frac{1}{\lambda_0} \left\{ - \left( 1 - \frac{1 - \lambda_0}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right) \right\} \left[ -a_0 + b_0 h_0(x) \right] dG(x)
= - a_0 + \frac{1 - \lambda_0}{\lambda_0} (a_0 A_0 - b_0 A_1)
\]
and
\[
a_{13} = - \int \lim_{n \to \infty} \frac{\partial g_1(x, \theta)}{\partial b} \big|_{\theta_0} dG(x) = - \int \lim_{n \to \infty} \frac{1}{\lambda} \frac{\partial U(h(x); \theta)}{\partial b} \big|_{\theta_0} dG(x)
= - \int \frac{1}{\lambda_0} \left\{ - \left( 1 - \frac{1 - \lambda_0}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right) \right\} \left[ \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0(x)^2 \right] dG(x)
= \frac{1}{\lambda_0} \left( \frac{1}{b_0} - b_0 \right) - \frac{1 - \lambda_0}{\lambda_0} \left( \frac{1}{b_0} A_0 + a_0 A_1 - b_0 A_2 \right)
\]
To derive the probability limit of \(-\frac{1}{n} \frac{\partial V_{n2}(\hat{h}; \theta)}{\partial \theta} |_{\theta_0}\), notice that

\[-\frac{1}{n} \frac{\partial V_{n2}(\hat{h}; \theta)}{\partial \theta} |_{\theta_0} \xrightarrow{p} - \int \lim_{n \to \infty} \frac{\partial g_2(x, \theta)}{\partial \theta} |_{\theta_0} dG(x)\]

Since

\[\frac{\partial g_2(x, \theta)}{\partial \theta} |_{\theta_0} = (1 - \lambda)[b\hat{h}(x) - a] \left( \frac{\partial U(\hat{h}(x); \theta)}{\partial \lambda} - \frac{1 - \lambda_0}{(1 - \lambda)^2} \right) |_{\theta_0}\]

we have,

\[a_{21} = -\int \lim_{n \to \infty} \frac{\partial g_2(x, \theta)}{\partial \lambda} |_{\theta_0} dG(x)\]

\[= -\int \lim_{n \to \infty} (1 - \lambda)[b\hat{h}(x) - a] \left( \frac{\partial U(\hat{h}(x); \theta)}{\partial \lambda} - \frac{1 - \lambda_0}{(1 - \lambda)^2} \right) |_{\theta_0} dG(x)\]

\[= -\int (1 - \lambda_0)[b_0h_0(x) - a_0]\]

\[\left[ - \frac{1}{\lambda_0} \left(1 - \frac{1}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) - \frac{1}{1 - \lambda_0} \right] dG(x)\]

\[= \frac{a_0}{\lambda_0} + \frac{1 - \lambda_0}{\lambda_0}(a_0A_0 - b_0A_1) = a_{12}\]

and

\[a_{22} = -\int \lim_{n \to \infty} \frac{\partial g_2(x, \theta)}{\partial a} |_{\theta_0} dG(x)\]

\[= -\int \lim_{n \to \infty} (1 - \lambda)[b\hat{h}(x) - a] \left( \frac{\partial U(\hat{h}(x); \theta)}{\partial a} \right) |_{\theta_0} dG(x)\]

\[= -\int (1 - \lambda_0)[b_0h_0(x) - a_0]\]

\[\left[ - \left(1 - \frac{1}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) [-a_0 + b_0h_0(x)] \right] dG(x)\]

\[= \int (1 - \lambda_0)[b_0h_0(x) - a_0]^2 \left(1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) dG(x)\]

45
\[ \begin{aligned}
&= (1 - \lambda_0)(b_0^2 + a_0^2) - (1 - \lambda_0)^2(b_0^2 A_2 - 2a_0b_0 A_1 + a_0^2 A_0) \\
\end{aligned} \]

and

\[ a_{23} = -\int \lim_{n \to \infty} \frac{\partial g_2(x, \theta)}{\partial b} |_{\theta_0} dG(x) \]
\[ = -\int (1 - \lambda_0)[b_0 h_0(x) - a_0] \left[ \left( 1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) \left( \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0(x)^2 \right) \right] dG(x) \]
\[ = \int (1 - \lambda_0) \left[ \left( 1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) \right] \left( -b_0^2 h_0(x)^3 + 2 a_0 b_0 h_0(x)^2 + (1 - a_0^2) h_0(x) - \frac{a_0}{b_0} \right) dG(x) \]
\[ = (1 - \lambda_0) a_0 \left( \frac{2b_0 - 1}{b_0} \right) \]
\[ -(1 - \lambda_0)^2 \left[ -b_0^2 A_3 + 2 a_0 b_0 A_2 + (1 - a_0^2) A_1 - \frac{a_0}{b_0} A_0 \right] \]

To derive the probability limit of \( -\frac{1}{n} \frac{\partial V_{n3}(\hat{h}; \theta)}{\partial \theta} \bigg|_{\theta_0} \), notice that

\[ -\frac{1}{n} \frac{\partial V_{n3}(\hat{h}; \theta)}{\partial \theta} \bigg|_{\theta_0} \xrightarrow{p} -\int \lim_{n \to \infty} \frac{\partial g_3(x, \theta)}{\partial \theta} |_{\theta_0} dG(x) \]

Since

\[ \frac{\partial g_3(x, \theta)}{\partial \theta} \bigg|_{\theta_0} = (1 - \lambda_0) \left[ \frac{1}{b} + a \hat{h}(x) - b \hat{h}(x)^2 \right] \left( \frac{\partial U(\hat{h}(x); \theta)}{\partial x} - \frac{1 - \lambda_0}{(1 - \lambda)^2} \right) \left( \begin{array}{c} \frac{\partial U(\hat{h}(x); \theta)}{\partial z} \\ \frac{\partial U(\hat{h}(x); \theta)}{\partial b} \end{array} \right) \bigg|_{\theta_0} \]
we have,

\[ a_{31} = - \int \lim_{n \to \infty} \frac{\partial g_3(x, \theta)}{\partial \lambda} |_{\theta_0} dG(x) \]

\[ = - \int \lim_{n \to \infty} (1 - \lambda)[\frac{1}{b} + a \hat{h}(x) - b \hat{h}(x)^2] \left( \frac{\partial U(\hat{h}(x); \theta)}{\partial \lambda} - \frac{1 - \lambda_0}{(1 - \lambda)^2} \right) |_{\theta_0} dG(x) \]

\[ = \int (\lambda_0 - 1)[\frac{1}{b_0} + a_0 h_0(x) - b_0 h_0^2(x)] \]

\[ \left[ - \frac{1}{\lambda_0} \left( 1 - \frac{1}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right) - \frac{1}{1 - \lambda_0} \right] dG(x) \]

\[ = \frac{1}{\lambda_0} \left( \frac{1}{b_0} + a_0 A_{-1} - b_0 A_{-2} \right) - \frac{1 - \lambda_0}{\lambda_0} \left( \frac{A_0}{b_0} + a_0 A_1 - b_0 A_2 \right) \]

\[ = \frac{1}{\lambda_0} \left( \frac{1}{b_0} - b_0 \right) - \frac{1 - \lambda_0}{\lambda_0} \left( \frac{A_0}{b_0} + a_0 A_1 - b_0 A_2 \right) = a_{13} \]

and

\[ a_{32} = - \int \lim_{n \to \infty} \frac{\partial g_3(x, \theta)}{\partial a} |_{\theta_0} dG(x) \]

\[ = - \int \lim_{n \to \infty} (1 - \lambda)[\frac{1}{b} + a \hat{h}(x) - b \hat{h}(x)^2] \left( \frac{\partial U(\hat{h}(x); \theta)}{\partial a} \right) |_{\theta_0} dG(x) \]

\[ = - \int (1 - \lambda_0)[\frac{1}{b_0} + a_0 h_0(x) - b_0 h_0^2(x)] \left[ - \frac{1 - \lambda_0}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right] dG(x) \]

\[ = \int (1 - \lambda_0)[\frac{1}{b_0} + a_0 h_0(x) - b_0 h_0^2(x)] [b_0 h_0(x) - a_0] \]

\[ \left( 1 - \frac{1 - \lambda_0}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right) dG(x) \]

\[ = (1 - \lambda_0) a_0 \left( 2 b_0 - \frac{1}{b_0} \right) \]

\[ -(1 - \lambda_0)^2 [-b_0^2 A_3 + 2 a_0 b_0 A_2 + (1 - a_0^2) A_1 - \frac{a_0}{b_0} A_0] = a_{23} \]

and

\[ a_{33} = - \int \lim_{n \to \infty} \frac{\partial g_3(x, \theta)}{\partial b} |_{\theta_0} dG(x) \]
\[
\int \lim_{n \to \infty} (1 - \lambda) \left[ \frac{1}{b} + a \hat{h}(x) - b \hat{h}(x)^2 \right] \left( \frac{\partial U(\hat{h}(x); \theta)}{\partial b} \right) |_{\theta_0} dG(x)
\]

\[
= - \int (1 - \lambda_0) \left[ \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0^2(x) \right] \left( 1 - \lambda_0 \right) \left[ \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0^2(x) \right] dG(x)
\]

\[
= \int (1 - \lambda_0) \left[ \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0^2(x) \right]^2 \left( 1 - \lambda_0 \right) \left[ \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0^2(x) \right] dG(x)
\]

\[
= (1 - \lambda_0) \left[ 3 b_0^2 + a_0^2 - 2 + \frac{1}{b_0^2} \right] - (1 - \lambda_0)^2 \left[ b_0^2 A_4 - 2 a_0 b_0 A_3 + a_0^2 A_2 - 2 A_2 + \frac{2 a_0}{b_0} A_1 + \frac{1}{b_0} A_0 \right]
\]

In a summary,

\[
\mathcal{A} = \begin{pmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{12} & a_{22} & a_{23} \\
  a_{13} & a_{23} & a_{33}
\end{pmatrix} = \mathcal{A}^T
\]

\[
\text{Theorem 2.2.1.}
\]

\[\text{Proof.} \] The first order Taylor expansion of \( V_n(\hat{\theta}, \hat{h}) \) around \( \theta_0 \) yields

\[
\sqrt{n}(\hat{\theta} - \theta_0) = n^{-\frac{1}{2}} \mathcal{A}^{-1} V_n(\hat{h}, \theta_0) + o_p(1)
\]  \hspace{1cm} (2.7)

where \( \mathcal{A} = - \lim_{n \to \infty} \frac{1}{n} \frac{\partial V_n(\theta, \hat{h})}{\partial \theta} |_{\theta_0} \) is derived in Lemma 2.6.3. We have

\[
\frac{V_{n1}(\hat{h}; \theta_0)}{n} = \frac{V_{n1}(h_0; \theta_0)}{n} + V_1(\theta_0) + o_p(n^{-\frac{1}{2}})
\]
Next we take $V_1(\theta_0)$ as function of $\hat{h}$ and do Taylor expansion of $V_1(\theta_0)$ around $h_0(x)$,

$$
V_1(\theta_0) = V_1(\theta_0)|_{\hat{h}(x)=h_0(x)} + \int \frac{\partial g_1(x; \theta_0)}{\partial \hat{h}(x)}|_{\hat{h}(x)=h_0(x)}[\hat{h}(x) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}})
$$

$$
= \int \frac{\partial g_1(x; \theta_0)}{\partial \hat{h}(x)}|_{\hat{h}(x)=h_0(x)}[\hat{h}(x) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}})
$$

$$
= -\int \frac{w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]}[\hat{h}(x) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}})
$$

Taylor expansion of $H(\hat{h}(x))$ around $h_0(x)$ yields

$$
\hat{H}(x) = H(\hat{h}(x)) = H(h_0(x)) + \frac{\partial H(h_0(x))}{\partial h(x)}|_{h_0(x)}(\hat{h}(x) - h_0(x)) + o_p(n^{-\frac{1}{2}})
$$

which gives

$$
\hat{h}(x) - h_0(x) = \left(\frac{\partial H(h_0(x))}{\partial h(x)}|_{h_0(x)}\right)^{-1} [\hat{H}(x) - H(h_0(x))] + o_p(n^{-\frac{1}{2}})
$$

$$
= \frac{\hat{G}(x) - \Phi(h_0(x))}{\phi(h_0(x))} + o_p(n^{-\frac{1}{2}}) = \frac{1}{n_D} \sum_{k=1}^{n_D} e_{x_k}(x) + o_p(n^{-\frac{1}{2}})
$$

where

$$
e_{y}(x) = I(y \leq x) - \Phi(h_0(x))
$$

So

$$
V_1(\theta_0) = -\int \frac{w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]}[\hat{h}(x) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}})
$$

$$
= -\int \frac{w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \cdot \frac{1}{n_D} \sum_{k=1}^{n_D} e_{x_k}(x) dG(x) + o_p(n^{-\frac{1}{2}})
$$

$$
= \frac{1}{n_D} \sum_{k=1}^{n_D} g_{1a}(x)e_{x_k}(x)dG(x) + o_p(n^{-\frac{1}{2}})
$$
where
\[
g_{1\alpha}(x) = \frac{-w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)]}{\{1 + \lambda_0[w(h_0(x); \theta_0) - 1]\}\phi(h_0(x))}
\]

Then
\[
\frac{V_{n1}(\hat{h}; \theta_0)}{n} = \frac{V_{n1}(h_0; \theta_0)}{n} + V_{1}(\theta_0) + o_p(n^{-\frac{1}{2}})
\]
\[
= -\frac{1}{n}\sum_{i=1}^{n} \frac{w(h_0(t_i); \theta_0) - 1}{1 + \lambda_0[w(h_0(t_i); \theta_0) - 1]} + \frac{1}{n}\sum_{k=1}^{n_D} \int g_{1\alpha}(x)e_{x_k}(x)dG(x) + o_p(n^{-\frac{1}{2}})
\]
\[
= \frac{1}{n}\sum_{j=1}^{v_{11}(y_j)} + \frac{1}{n}\sum_{k=1}^{v_{12}(x_k)} + o_p(n^{-\frac{1}{2}})
\]

where
\[
v_{11}(y) = -\frac{w(h_0(y); \theta_0) - 1}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]}
\]
\[
v_{12}(y) = \frac{1}{1 - \lambda_0} \int g_{1\alpha}(x)e_{y}(x)dG(x) - \frac{w(h_0(y); \theta_0) - 1}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]}
\]
\[
= \frac{1}{1 - \lambda_0} \int g_{1\alpha}(x)e_{y}(x)dG(x) + v_{11}(y)
\]

Similarly,
\[
V_2(\theta_0) = V_2(\theta_0)|_{\hat{h}=h_0}
\]
\[
+ \int \frac{\partial g_2(x; \theta_0)}{\partial \hat{h}(x)}|_{\hat{h}(x)=h_0(x)} [\hat{h}(x) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}})
\]
\[
= \int \frac{\partial g_2(x; \theta_0)}{\partial \hat{h}(x)}|_{\hat{h}(x)=h_0(x)} [\hat{h}(x) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}})
\]
\[
= -\lambda_0(1 - \lambda_0) \int \frac{w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \cdot
\]
\[
[b_0h_0(x) - a_0][\hat{h}(x) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}})
\]
\[
= -\lambda_0(1 - \lambda_0) \int \frac{w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]}[b_0h_0(x) - a_0]
\]
where

\[ g_{2a}(x) = \frac{-\lambda_0(1 - \lambda_0)w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)][b_0h_0(x) - a_0]}{\{1 + \lambda_0[w(h_0(x); \theta_0) - 1]\} \phi(h_0(x))} \]

Then

\[ \frac{V_{n2}(\hat{h}; \theta_0)}{n} = \frac{V_{n2}(h_0; \theta_0)}{n} + V_2(\theta_0) + o_p(n^{-\frac{1}{2}}) \]

\[ = \frac{1}{n} \sum_{j=1}^{n} \left[ b_0h_0(y_j) - a_0 \right] - \frac{1}{n} \sum_{i=1}^{n} \lambda_0 w(h_0(t_i); \theta_0) \left[ b_0h_0(t_i) - a_0 \right] \frac{1}{1 + \lambda_0[w(h_0(t_i); \theta_0) - 1]} \]

\[ \quad + \frac{1}{n \beta} \sum_{k=1}^{n \beta} \int g_{2a}(x)e_{x_k}(x)dG(x) + o_p(n^{-\frac{1}{2}}) \]

\[ = \frac{1}{n} \sum_{j=1}^{n} v_{21}(y_j) + \frac{1}{n} \sum_{k=1}^{n \beta} v_{22}(x_k) + o_p(n^{-\frac{1}{2}}) \]

where

\[ v_{21}(y) = \frac{(1 - \lambda_0)[b_0h_0(y) - a_0]}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]} \]

\[ v_{22}(y) = \int \frac{g_{2a}(x)e_y(x)}{1 - \lambda_0} dG(x) - \left\{ 1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]} \right\} \left[ b_0h_0(y) - a_0 \right] \]

\[ = \int \frac{g_{2a}(x)e_y(x)}{1 - \lambda_0} dG(x) \left[ b_0h_0(y) - a_0 \right] + v_{21}(y) \]

Third,

\[ V_3(\theta_0) = V_3(\theta_0)\left|_{\hat{h}(x) = h_0(x)} \right| + \int \frac{\partial g_3(x; \theta_0)}{\partial \hat{h}(x)} \left|_{\hat{h}(x) = h_0(x)} \right| [\hat{h}(x) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}}) \]

\[ = \int \frac{\partial g_3(x; \theta_0)}{\partial \hat{h}(x)} \left|_{\hat{h}(x) = h_0(x)} \right| [\hat{h}(x) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}}) \]
\[
\begin{align*}
&= -\lambda_0 (1 - \lambda_0) \int \frac{w(h_0(x); \theta_0)[a_0 b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \\
&\quad + \frac{1}{b_0} \int h(x)[\hat{h}(x) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}}) \\
&= -\lambda_0 (1 - \lambda_0) \int \frac{w(h_0(x); \theta_0)[a_0 b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \left[ b_0 h_0(x) - a_0 \right] \\
&\quad + \frac{1}{n_D} \sum_{k=1}^{n_D} \frac{e_{x_k}(x)}{\phi(h_0(x))} dG(x) + o_p(n^{-\frac{1}{2}}) \\
&= \frac{1}{n_D} \sum_{k=1}^{n_D} g_{3a}(x) e_{x_k}(x) dG(x) + o_p(n^{-\frac{1}{2}})
\end{align*}
\]

where

\[
g_{3a}(x) = \frac{-\lambda_0 (1 - \lambda_0) w(h_0(x); \theta_0)[a_0 b_0 + (1 - b_0^2)h_0(x)]}{\{1 + \lambda_0[w(h_0(x); \theta_0) - 1]\} \phi(h_0(x))}
\]

Then

\[
\frac{V_{n3}(\hat{h}; \theta_0)}{n} = \frac{V_{n3}(h_0; \theta_0)}{n} + V_3(\theta_0) + o_p(n^{-\frac{1}{2}})
\]

\[
= \frac{1}{n} \sum_{j=1}^{n_D} \left\{ \frac{1}{b_0} + a_0 h_0(y_j) - b_0 h_0^2(y_j) \right\} \\
- \frac{1}{n} \sum_{i=1}^{n} \lambda_0 w(h_0(t_i); \theta_0) \left[ \frac{1}{b_0} + a_0 h_0(t_i) - b_0 h_0^2(t_i) \right] \\
\quad + \frac{1}{n_D} \sum_{k=1}^{n_D} \int g_{3a}(x) e_{x_k}(x) dG(x) + o_p(n^{-\frac{1}{2}}) \\
= \frac{1}{n} \sum_{j=1}^{n_D} v_{31}(y_j) + \frac{1}{n} \sum_{k=1}^{n_D} v_{32}(x_k) + o_p(n^{-\frac{1}{2}})
\]

where

\[
v_{31}(y) = \frac{(1 - \lambda_0) \left[ \frac{1}{b_0} + a_0 h_0(y) - b_0 h_0^2(y) \right]}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]}
\]

\[
v_{32}(y) = \int \frac{g_{3a}(x) e_y(x)}{1 - \lambda_0} dG(x) - \frac{\lambda_0 w(h_0(y); \theta_0) \left[ \frac{1}{b_0} + a_0 h_0(y) - b_0 h_0^2(y) \right]}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]} \\
= \int \frac{g_{3a}(x) e_y(x)}{1 - \lambda_0} dG(x) - \left[ \frac{1}{b_0} + a_0 h_0(y) - b_0 h_0^2(y) \right] + v_{31}(y)
\]

52
In a summary,

\[
\frac{1}{n} V_n(\hat{h}, \theta_0) = \frac{1}{n} \begin{pmatrix} V_{n1} \\ V_{n2} \\ V_{n3} \end{pmatrix}
\]

\[
= \frac{1}{n} \sum_{j=1}^{n_D} \begin{pmatrix} v_{11}(y_j) \\ v_{21}(y_j) \\ v_{31}(y_j) \end{pmatrix} + \frac{1}{n} \sum_{k=1}^{n_D} \begin{pmatrix} v_{12}(x_k) \\ v_{22}(x_k) \\ v_{32}(x_k) \end{pmatrix} + o_p(n^{-\frac{1}{2}})
\]

\[
= \frac{1}{n} \sum_{j=1}^{n_D} u_1(y_j) + \frac{1}{n} \sum_{k=1}^{n_D} u_2(x_k) + o_p(n^{-\frac{1}{2}})
\]

where

\[
u_1(y) = \begin{pmatrix} v_{11}(y) \\ v_{21}(y) \\ v_{31}(y) \end{pmatrix} = \begin{pmatrix} \frac{1}{1-\lambda_0} \int g_{1a}(x)e_y(x)dG(x) + v_{11}(y) \\ \int g_{2a}(x)e_y(x)dG(x) - [b_0h_0(y) - a_0] + v_{21}(y) \\ \int g_{3a}(x)e_y(x)dG(x) - \left[\frac{1}{b_0} + a_0h_0(y) - b_0h_0^2(y)\right] + v_{31}(y) \end{pmatrix}
\]

\[
u_2(y) = \begin{pmatrix} v_{12}(y) \\ v_{22}(y) \\ v_{32}(y) \end{pmatrix} = \begin{pmatrix} \frac{1}{1-\lambda_0} \int g_{1a}(x)e_y(x)dG(x) + v_{11}(y) \\ \int g_{2a}(x)e_y(x)dG(x) - [b_0h_0(y) - a_0] + v_{21}(y) \\ \int g_{3a}(x)e_y(x)dG(x) - \left[\frac{1}{b_0} + a_0h_0(y) - b_0h_0^2(y)\right] + v_{31}(y) \end{pmatrix}
\]

Plugging into (2.7) yields

\[
\sqrt{n}(\hat{\theta} - \theta_0) = n^{-\frac{1}{2}} A^{-1} V_n(\hat{h}, \theta_0) + o_p(1)
\]

\[
= n^{-\frac{1}{2}} A^{-1} \left\{ \sum_{j=1}^{n_D} u_1(y_j) + \sum_{k=1}^{n_D} u_2(x_k) + o_p(n^{-\frac{1}{2}}) \right\} + o_p(1)
\]
= n^{-\frac{1}{2}} \mathcal{A}^{-1}\left\{ \sum_{j=1}^{n_{D}} \mathbf{u}_1(y_j) + \sum_{k=1}^{n_{D}} \mathbf{u}_2(x_k) \right\} + o_p(1)

Due to Lemma 2.6.1, \( \hat{\theta} \) is a consistent estimator of \( \theta_0 \), thus

\[ \mathbb{E}\{\sqrt{n}(\hat{\theta} - \theta_0)\} = 0 \]

Due to \( X \) and \( Y \) are independent, the covariance matrix of \( \sqrt{n}(\hat{\theta} - \theta_0) \) is

\[ \Sigma_{pelni} = \lim_{n \to \infty} n^{-1} \mathcal{A}^{-1}\{n_{D}\text{Var}(\mathbf{u}_1(Y)) + n_{D}\text{Var}(\mathbf{u}_2(X))\}\mathcal{A}^{-1}\tau \]

\[ = \mathcal{A}^{-1}\{\lambda_0\text{Var}(\mathbf{u}_1(Y)) + (1 - \lambda_0)\text{Var}(\mathbf{u}_2(X))\}\mathcal{A}^{-1}\tau \]

\[ = \frac{1}{1 + \rho} \mathcal{A}^{-1}\{\rho\text{Var}(\mathbf{u}_1(Y)) + \text{Var}(\mathbf{u}_2(X))\}\mathcal{A}^{-1}\tau \]

where \( \lambda_0 = \lim_{n \to \infty} \frac{n_{D}}{n} = \frac{\rho}{1 + \rho} \) and \( \rho = \lim_{n \to \infty} \frac{n_{D}}{n} \).

By Central Limit Theorem, \( \sqrt{n}(\hat{\theta} - \theta_0) \) converges in distribution to a normal random vector with the expectation of \( \mathbf{0} \) and covariance matrix \( \Sigma_{pelni} \), i.e.,

\[ \sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\mathbf{0}, \Sigma_{pelni}) \]

The estimation of \( \Sigma_{pelni} \) is obtained by replacing all theoretical quantities by their empirical counterparts,

\[ \hat{\Sigma}_{pelni} = \frac{1}{n} \hat{\mathcal{A}}^{-1}\left\{ \sum_{j=1}^{n_{D}} [\hat{\mathbf{u}}_1(y_j) - \frac{1}{n_{D}} \sum_{j=1}^{n_{D}} \hat{\mathbf{u}}_1(y_j)][\hat{\mathbf{u}}_1(y_j) - \frac{1}{n_{D}} \sum_{j=1}^{n_{D}} \hat{\mathbf{u}}_1(y_j)]\mathcal{A}^{-1}\tau + \sum_{k=1}^{n_{D}} [\hat{\mathbf{u}}_2(x_k) - \frac{1}{n_{D}} \sum_{k=1}^{n_{D}} \hat{\mathbf{u}}_2(x_k)][\hat{\mathbf{u}}_2(x_k) - \frac{1}{n_{D}} \sum_{k=1}^{n_{D}} \hat{\mathbf{u}}_2(x_k)]\mathcal{A}^{-1}\tau \right\}(\hat{\mathcal{A}}^{-1})\tau \]

The proof is complete. \( \square \)
Chapter 3

Estimate binormal ROC curve under density ratio model – Pseudo Empirical Likelihood method

3.1 Introduction

In previous chapter, we propose a semiparametric \textit{pelni} method to estimate ROC curve. Although the semiparametric \textit{pelni} method is consistent, it is less efficient than the other semiparametric method (\textit{mle}, \textit{pmle}, \textit{labroc}). In this chapter, we propose another semiparametric estimator - Pseudo Empirical Likelihood (\textit{pel}) for the parameters \((a, b)\) in the model (2.1).

The \textit{pel} method in this chapter has the same assumption as the \textit{pelni} method in Chapter 2, which assume that the data though an unknown transformation satisfy a density ratio model. The score equations of the two methods are the same. Unlike the \textit{pelni} method in Chapter 2, the \textit{pel} method in this chapter uses both the diseased and nondiseased samples to estimate \(h(\cdot)\). We show that the \textit{pel} method is more
efficient than other semiparametric methods, especially when sample sizes are small or AUC is large.

This chapter is organized as follows. In Section 3.2, we propose the new pel method to estimate the parameters in the binormal model. In Section 3.3, we extend the new pel method to inference about AUC and ROC curve. In Section 3.4, we apply the proposed pel method to two real examples. In Section 3.5 we conduct simulation studies to compare the performance of pel method with the mle, pmle, labroc, pelni, mlnm, np methods which are defined in Chapter 2. A summary is given in Section 3.6. Proofs of the main theoretical results are provided in Section 3.7.

3.2 Methodology

Let $X_1, \cdots, X_{n_D}$ denote independent and identically distributed test results from a nondiseased population and, independently of the $X_i$’s, let $Y_1, \cdots, Y_{n_D}$ be independent and identically distributed test results from a diseased population. Let $T_1, \cdots, T_n$ denote the pooled test results $Y_1, \cdots, Y_{n_D}, X_1, \cdots, X_{n_D}$ with $n = n_D + n_D$. Let $x_i$’s, $y_i$’s and $t_i$’s be the observations of $X_i$’s, $Y_i$’s and $T_i$’s respectively. Let $G$ and $F$ represent respectively the distribution functions of $X_1$ and $Y_1$. Let $g$ and $f$ represent the corresponding density functions.

Let $h(x) = \Phi^{-1}(G(x))$ be the unknown transformation function. Under the bi-normal model, the distributions of the transformed data are

$$h(X_1) \sim N(0, 1), \quad h(Y_1) \sim N\left(\frac{a}{b}, \frac{1}{b^2}\right). \quad (3.1)$$

On the other hand, since $h(x) = \Phi^{-1}(G(x))$ is a monotonically increasing function, we have

$$h(X_1), \cdots, h(X_{n_D}) \overset{iid}{\sim} G(h(x)) = \Phi(h(x)),$$

$$h(Y_1), \cdots, h(Y_{n_D}) \overset{iid}{\sim} F(h(x)) = \Phi(bh(x) - a).$$
Accordingly,
\[
g(x) = \frac{dG(h(x))}{dx} = \phi(h(x)) \frac{dh(x)}{dx},
\]
\[
f(x) = \frac{dF(h(x))}{dx} = b\phi(bh(x) - a) \frac{dh(x)}{dx}.
\]

The ratio of above two density functions is
\[
\frac{f(x)}{g(x)} = \frac{b\phi(bh(x) - a)}{\phi(h(x))} = w(h(x); a, b).
\]

Then we have the following two-sample density ratio model
\[
X_1, \cdots, X_{n_D} \overset{iid}{\sim} g(x),
\]
\[
Y_1, \cdots, Y_{n_D} \overset{iid}{\sim} f(x) = g(x)w(h(x); a, b),
\]

where
\[
w(h(x); a, b) = \exp \left\{ \log b - \frac{a^2}{2} + abh(x) + \frac{1 - b^2}{2}h^2(x) \right\}.
\]

Let \( p_i = dG(t_i) \) \((i = 1, \cdots, n)\) be non-negative jump sizes with total mass unity. Based on the density ratio model (3.2), the empirical log-likelihood function is
\[
l = \sum_{i=1}^{n} \log p_i + \sum_{j=1}^{n_D} \log w(h(y_j); a, b)
\]
\[
= \sum_{i=1}^{n} \log p_i + \sum_{j=1}^{n_D} \left\{ \log b - \frac{a^2}{2} + abh(y_j) + \frac{1 - b^2}{2}h^2(y_j) \right\}.
\]

To estimate the unknown transformation function \( h(x) \), let \( \hat{H}(x) = \frac{1}{n} \sum_{i=1}^{n} I(t_i \leq x) \) be the empirical distribution function of the combined sample. It follows from (3.1) and proved in Lemma 3.7.1, that as \( n \to \infty \), \( \hat{H}(x) \overset{P}{\to} H(h_0(x); a_0, b_0) \) for each fixed \( x \); where \( H(h(x); a, b) = \frac{1}{1+\hat{\rho}} \Phi(h(x)) + \frac{\hat{\rho}}{1+\hat{\rho}} \Phi(bh(x) - a) \), \( h_0 \) is the true underling transformation function, \( \hat{\rho} = \frac{na}{n\hat{D}} \). This motivates the following estimating equation
for \( h(x) = \Phi^{-1}(G(x)) \) at any given \((a, b)\):

\[
H(h(x); a, b) = \hat{H}(x). \tag{3.3}
\]

Let \( \hat{h}(x) \) denote the solution to (3.3) for a given \((a, b)\). Plugging \( \hat{h} \) to the above log-likelihood function, we obtain the estimated empirical log-likelihood function:

\[
\hat{l} = \sum_{i=1}^{n} \log p_i + \sum_{j=1}^{n_D} \left\{ \log b - \frac{a^2}{2} + ab\hat{h}(y_j) + \frac{1 - b^2}{2} \hat{h}^2(y_j) \right\}.
\]

Using Lagrange multiplier method, we maximize the estimated empirical log-likelihood subject to constraints \( \sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0, \quad \sum_{i=1}^{n} p_i (w(\hat{h}(t_i); a, b) - 1) = 0. \)

As a result,

\[
\hat{p}_i = \frac{1}{n} \frac{1}{1 + \hat{\lambda} [w(\hat{h}(t_i); \tilde{a}, \tilde{b}) - 1]},
\]

where \((\hat{\lambda}, \tilde{a}, \tilde{b})\) is the solution of the score equations for \((\lambda_0, a_0, b_0)\):

\[
\begin{align*}
\frac{\partial \ell(\lambda, a, b)}{\partial \lambda} &= -\sum_{i=1}^{n} \frac{w(\hat{h}(t_i); a, b) - 1}{1 + \hat{\lambda} [w(\hat{h}(t_i); a, b) - 1]} = 0, \\
\frac{\partial \ell(\lambda, a, b)}{\partial a} &= \sum_{j=1}^{n_D} [b\hat{h}(y_j) - a] - \sum_{i=1}^{n} \frac{\lambda w(\hat{h}(t_i); a, b) [b\hat{h}(t_i) - a]}{1 + \hat{\lambda} [w(\hat{h}(t_i); a, b) - 1]} = 0, \\
\frac{\partial \ell(\lambda, a, b)}{\partial b} &= \sum_{j=1}^{n_D} [\frac{1}{b} + a\hat{h}(y_j) - b\hat{h}^2(y_j)] - \sum_{i=1}^{n} \frac{\lambda w(\hat{h}(t_i); a, b) [\frac{1}{b} + a\hat{h}(t_i) - b\hat{h}^2(t_i)]}{1 + \hat{\lambda} [w(\hat{h}(t_i); a, b) - 1]} = 0.
\end{align*}
\]  

(\(\hat{\lambda}, \tilde{a}, \tilde{b}\)) can be obtained either by solving (3.3) for \( h = (h(t_1), \ldots, h(t_n))^\tau \) and solving (3.4) for \((\lambda, a, b)\) iteratively or by solving them simultaneously using the Newton-Raphson algorithm. We assume as \( n \to \infty, \hat{\rho} \) is approaching a constant \( \rho \), which is the odds in favor of disease in the whole population. Thus \( \rho = \lim_{n \to \infty} \hat{\rho} \). Let \( \tilde{\rho} = \frac{\hat{\lambda}}{1 - \hat{\lambda}} \) which is a number close to \( \rho \). Let \( \theta_0 = (\lambda_0, a_0, b_0)^\tau \) denote the true value of \( \theta = (\lambda, a, b)^\tau \) and \( h_0 \) denote the true transformation function. In particular \( \lambda_0 = \frac{\rho}{1 + \rho}. \)
Write

\[
\mathbf{u}_1(y) = \int \begin{pmatrix} g_{1a}(x) \\ g_{2a}(x) \\ g_{3a}(x) \end{pmatrix} e_y(x) dG(x) + \begin{pmatrix} \frac{(1+\rho)(1-w(h_0(y);\theta_0))}{1+\rho w(h_0(y);\theta_0)} \\ \frac{b_0 h_0(y) - a_0}{1+\rho w(h_0(y);\theta_0)} \\ \frac{1}{\rho} + a_0 h_0(y) - b_0 h_0^2(y) \end{pmatrix},
\]

\[
\mathbf{u}_2(y) = \mathbf{u}_1(y) - \begin{pmatrix} 0 \\ b_0 h_0(y) - a_0 \\ \frac{1}{\rho} + a_0 h_0(y) - b_0 h_0^2(y) \end{pmatrix},
\]

where

\[
g_{1a}(x) = -\frac{(1+\rho)^2 w(h_0(x);\theta_0)[a_0 b_0 + (1-b_0^2) h_0(x)]}{[1+\rho w(h_0(x);\theta_0)][\phi(h_0(x)) + \rho b_0 \phi(b_0 h_0(x) - a_0)]},
\]

\[
g_{2a}(x) = \frac{\rho}{(1+\rho)^2} [b_0 h_0(x) - a_0] g_{1a}(x),
\]

\[
g_{3a}(x) = \frac{\rho}{(1+\rho)^2} \left[ \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0^2(x) \right] g_{1a}(x),
\]

and

\[
e_y(x) = I(y \leq x) - \frac{\Phi(h_0(x)) + \rho \Phi(b_0 h_0(x) - a_0)}{1+\rho}.
\]

Denote \(\hat{h}(x) = \hat{h}(x;\tilde{a},\tilde{b})\). Accordingly, the empirical counterparts of \(\mathbf{u}_1\) and \(\mathbf{u}_2\) are

\[
\hat{\mathbf{u}}_1(y) = \int \begin{pmatrix} \hat{g}_{1a}(x) \\ \hat{g}_{2a}(x) \\ \hat{g}_{3a}(x) \end{pmatrix} \hat{e}_y(x) dG(x) + \begin{pmatrix} \frac{1}{\lambda} \left( \frac{1}{1+\lambda[w(\hat{h}(y);\tilde{a},\tilde{b})-1]} - 1 \right) \\ \frac{(1-\lambda) \left[ \hat{h}(y) - \tilde{a} \right]}{1+\lambda[w(\hat{h}(y);\tilde{a},\tilde{b})-1]} \\ \frac{(1-\lambda) \left[ \frac{1}{2} \hat{h}(y) - \tilde{b} \hat{h}(y) \right]}{1+\lambda[w(\hat{h}(y);\tilde{a},\tilde{b})-1]} \end{pmatrix},
\]

\[
= \sum_{i=1}^{n} \begin{pmatrix} \hat{g}_{1a}(t_i) \\ \hat{g}_{2a}(t_i) \\ \hat{g}_{3a}(t_i) \end{pmatrix} \hat{e}_y(t_i) \hat{p}(t_i) + \begin{pmatrix} \frac{1}{\lambda} \left( \frac{1}{1+\lambda[w(\hat{h}(y);\tilde{a},\tilde{b})-1]} - 1 \right) \\ \frac{(1-\lambda) \left[ \hat{h}(y) - \tilde{a} \right]}{1+\lambda[w(\hat{h}(y);\tilde{a},\tilde{b})-1]} \\ \frac{(1-\lambda) \left[ \frac{1}{2} \hat{h}(y) - \tilde{b} \hat{h}(y) \right]}{1+\lambda[w(\hat{h}(y);\tilde{a},\tilde{b})-1]} \end{pmatrix},
\]

59
\[
\hat{u}_2(y) = \hat{u}_1(y) - \begin{pmatrix}
0 \\
\tilde{b}\tilde{h}(y) - \tilde{a} \\
\frac{1}{\tilde{b}} + \tilde{a}\tilde{h}(y) - \tilde{b}\tilde{h}^2(y)
\end{pmatrix},
\]

where

\[
\hat{g}_{1a}(x) = \frac{-(1+\rho)w(\tilde{h}(x);\tilde{a},\tilde{b})[\tilde{a}\tilde{b} + (1-\tilde{b}^2)\tilde{h}(x)]}{\{1 + \tilde{\lambda}[w(\tilde{h}(x);\tilde{a},\tilde{b}) - 1]\phi(\tilde{h}(x)) + \rho\phi[\tilde{b}\tilde{h}(x) - \tilde{a}]\}},
\]
\[
\hat{g}_{2a}(x) = \tilde{\lambda}(1 - \tilde{\lambda})[\tilde{b}\tilde{h}(x) - \tilde{a}]\hat{g}_{1a}(x),
\]
\[
\hat{g}_{3a}(x) = \tilde{\lambda}(1 - \tilde{\lambda})\left[\frac{1}{\tilde{b}} + \tilde{a}\tilde{h}(x) - \tilde{b}\tilde{h}^2(x)\right]\hat{g}_{1a}(x),
\]

and

\[
\hat{e}_y(x) = I(y \leq x) - \frac{\Phi(\tilde{h}(x)) + \rho\Phi(\tilde{b}\tilde{h}(x) - \tilde{a})}{1 + \hat{\rho}}.
\]

The following Theorem 3.2.1 shows the asymptotic distribution of \((\tilde{\lambda}, \tilde{a}, \tilde{b})\).

**Theorem 3.2.1.** Let \(\tilde{\theta} = (\tilde{\lambda}, \tilde{a}, \tilde{b})^\tau\) denote the resulting estimation for \(\theta_0\) by solving equations (3.3) and (3.4) simultaneously. Then \(\sqrt{n}(\tilde{\theta} - \theta_0)\) converges in distribution to a normal random vector with expectation of \(0\) and covariance matrix \(\Sigma_{pel}\). That is,

\[
\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_{pel}),
\]

where

\[
\Sigma_{pel} = \frac{1}{1 + \rho}A^{-1}\{\rho Var(u_1(Y)) + Var(u_2(X))\}(A^{-1})^\tau,
\]

and \(A\) is given in Lemma 3.7.4. The estimation of \(\Sigma_{pel}\) is obtained by replacing all
theoretical quantities by their empirical counterparts. That is,

$$\hat{\Sigma}_{\text{pel}} = \frac{1}{n} \hat{A}^{-1} \left\{ \sum_{j=1}^{n_D} \left[ \hat{u}_1(y_j) - \frac{1}{n_D} \sum_{j=1}^{n_D} \hat{u}_1(y_j) \right] \left[ \hat{u}_1(y_j) - \frac{1}{n_D} \sum_{j=1}^{n_D} \hat{u}_1(y_j) \right] \right\}^\tau + \sum_{k=1}^{n_D} \left[ \hat{u}_2(x_k) - \frac{1}{n_D} \sum_{k=1}^{n_D} \hat{u}_2(x_k) \right] \left[ \hat{u}_2(x_k) - \frac{1}{n_D} \sum_{k=1}^{n_D} \hat{u}_2(x_k) \right] \right\} \left( \hat{A}^{-1} \right)^\tau.$$

### 3.3 Inference about ROC and AUC

A binormal ROC curve is estimated by

$$\tilde{R}(s) = \Phi(\tilde{a} + \tilde{b}\Phi^{-1}(s)), \ s \in [0, 1] \quad (3.5)$$

where \((\tilde{a}, \tilde{b})\) are the estimation of \((a, b)\) based on semiparametric \(pel\) method. Similarly, the AUC estimation is

$$\tilde{d} = \Phi \left( \frac{\tilde{a}}{\sqrt{1 + \tilde{b}^2}} \right) \quad (3.6)$$

Since \((\tilde{a}, \tilde{b})\) are consistent estimator for \((a_0, b_0)\), as \(n \to \infty\)

$$\tilde{R}(s) = R_0(s) = \Phi(a_0 + b_0\Phi^{-1}(s)), \ s \in [0, 1],$$

$$\tilde{d} = d_0 = \Phi \left( \frac{a_0}{\sqrt{1 + b_0^2}} \right).$$

The variance of \(\tilde{R}(s)\) and \(\tilde{d}\) can be obtained using Delta method. Let \(\tilde{\theta}\) denote the estimator of \(\theta_0\) and \(\Sigma_{\theta}\) denote the asymptotic variance of \(\theta\). According to [Casella & Berger, 2002, p.242], the variance of any continuous function \(f(\tilde{\theta})\) has the property

$$\text{Var}(f(\tilde{\theta})) \approx \left( \frac{\partial f(\theta)}{\partial \theta} \right)^\tau \left|_{\theta_0} \right. \Sigma_{\theta} \left( \frac{\partial f(\theta)}{\partial \theta} \right)^\tau \left|_{\theta_0} \right.$$
Therefore,

\[
\text{Var}(\tilde{d}) \approx n^{-1}(\frac{\partial d}{\partial \theta})^T|_{\theta_0} \Sigma_{\text{pel}} \frac{\partial d}{\partial \theta}|_{\theta_0} = \frac{1}{n(1 + b_0^2)} \phi^2 \left( \frac{a_0}{\sqrt{1 + b_0^2}} \right) \left( 1 - \frac{a_0 b_0}{1 + b_0^2} \right) \Sigma_{\text{pel}} \left( \begin{array}{c} 1 \\ -\frac{a_0 b_0}{1 + b_0^2} \end{array} \right) = \Sigma_d
\]

and

\[
\text{Var}(\tilde{R}(s)) \approx n^{-1}(\frac{\partial R(s)}{\partial \theta})^T|_{\theta_0} \Sigma_{\text{pel}} \frac{\partial R(s)}{\partial \theta}|_{\theta_0} = n^{-1} \phi^2(a_0 + b_0 \Phi^{-1}(s)) \left( 1 - \frac{\phi^{-1}(s)}{\Phi^{-1}(s)} \right) \Sigma_{\text{pel}} \left( \begin{array}{c} 1 \\ \Phi^{-1}(s) \end{array} \right) = \Sigma_{R(s)}
\]

Since \( d \) and \( R(s) \) are between 0 and 1, their Wald confidence intervals are usually constructed with logit transformation. Let \( \tilde{L} = \text{Logit}(\tilde{d}) = \log \frac{\tilde{d}}{1 - \tilde{d}} \), by Delta method

\[
\text{Var}(\tilde{L}) \approx \frac{1}{d_0^2(1 - d_0)^2} \Sigma_d = \Sigma_L
\]

and the \((1 - \alpha)\) margin of error for \( \tilde{L} \) is given by \( M = \Phi^{-1}(1 - \frac{\alpha}{2}) \sqrt{\Sigma_L} \). Replacing \((a_0, b_0)\) by \((\tilde{a}, \tilde{b})\) in \( M \) yields \( \tilde{M} \). The \((1 - \alpha)\) confidence interval for \( \tilde{d} \) is given by

\[
\left[ \frac{1}{1 + \exp(-\tilde{L} - \tilde{M})}, \frac{1}{1 + \exp(-\tilde{L} + \tilde{M})} \right].
\]

The confidence interval for \( \tilde{R}(s) \) can be constructed similarly.

### 3.4 Examples

**Example 3.4.1.** As reported by Wieand et al. [1989], sera from \( n_D = 51 \) control patients with pancreatitis and \( n_D = 90 \) cases patients with pancreatic cancer were studied at the Mayo Clinic with a carbohydrate antigen CA19-9. We apply three methods (mle, pmle and pel) to the data. Table 3.1 lists the resulting estimates
of \( (a, b) \) with standard errors (in parenthesis). All three methods give similar point estimations of \( a \) and \( b \), however, the standard error of pel method is smaller method is smaller than that of mle and pmle. The pel estimator of \( (a, b) \) is \((\hat{a}, \hat{b}) = (1.188, 0.445)\) with the standard error \( SE = (0.185, 0.064)\).

<table>
<thead>
<tr>
<th></th>
<th>mle</th>
<th>pmle</th>
<th>pel</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1.235 (0.161)</td>
<td>1.209 (0.163)</td>
<td>1.188 (0.185)</td>
</tr>
<tr>
<td>b</td>
<td>0.48 (0.089)</td>
<td>0.496 (0.081)</td>
<td>0.445 (0.064)</td>
</tr>
</tbody>
</table>

**Example 3.4.2.** As reported by Cooper et al. [2014], the Cardiovascular Outcomes in Renal Atherosclerotic Lesions (CORAL) study was a multi-center, open-label, randomized, controlled trial that compared medical therapy alone with medical therapy plus renal-artery stenting in patients with atherosclerotic renal-artery stenosis and elevated blood pressure, chronic kidney disease, or both. The urinal albumin and creatinine ratio (UACR) were measured at baseline for 931 patients. The patients were followed up to 5 years after treatment, during which patients’ composite endpoint event including death, stroke, chronic heart failure etc were recorded. To show that baseline UACR is predictive to patients’ composite endpoint event, we conduct an ROC analysis using the composite endpoint event as the binary disease status and baseline UACR as the continuous test results. Table 3.2 lists the resulting estimates of \( (a, b) \) with standard errors. The results of the three methods are very similar. The pel estimator of \( (a, b) \) is \((\hat{a}, \hat{b}) = (0.496, 0.903)\) with standard error \( SE = (0.074, 0.054)\).
Table 3.2: Example of UACR data in CORAL study

<table>
<thead>
<tr>
<th></th>
<th>mle</th>
<th>pmle</th>
<th>pel</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.504 (0.074)</td>
<td>0.496 (0.074)</td>
<td>0.496 (0.074)</td>
</tr>
<tr>
<td>b</td>
<td>0.914 (0.049)</td>
<td>0.905 (0.054)</td>
<td>0.903 (0.054)</td>
</tr>
</tbody>
</table>

Figure 3-1: ROC curves for UACR data in CORAL study: the binormal ROC curve based on pel method (smooth blue curve) and the nonparametric ROC curve (stepped black curve).

We then plot the ROC curve based on the pel estimator of \((a, b)\) (see Figure 3-1). The estimated AUC based on pel method is 0.644, which is the same as the estimated AUC based on nonparametric ROC curve. Further, we plot \(\Phi^{-1}(R(s))\) versus \(\Phi^{-1}(s)\) in Figure 3-2. Clearly it is a linear curve which verifies that the binormal model works well on this data. The added red line is \(\Phi^{-1}(R(s)) = \hat{a} + \hat{b}\Phi^{-1}(s)\) where \((\hat{a}, \hat{b})\) is obtained from pel method.
Figure 3-2: Plot of $\Phi^{-1}(R)$ versus $\Phi^{-1}(s)$ for UACR data in CORAL study

3.5 Simulations

Two simulation studies are conducted to compare the performance of pel method with the other three semiparametric methods (mle and pmle of Cai & Moskowitz [2004], labroc of Metz et al. [1998]), one parametric method (mlnm based on normal assumption) and the nonparametric method np. The details of each method are provided in Chapter 2. In the first part of simulation study, we choose two distributions in which the binormal model is correctly specified. In the second part of simulation study, we choose two distributions in which the binormal model are misspecified. We consider $s = \{0.1, 0.3, 0.6\}$ for ROC curve $R(s)$. 1000 independent random samples are generated from the diseased and nondiseased population. For each setup, the binormal model parameter $(a, b)$ are estimated first. For each estimation, we calculate the relative efficiency (relative to pel method), bias, standard deviation, standard error, coverage probability and average width of 95% confidence interval. Then for
ROC curve $R(s)$ and the area under the curve $d$, we calculate the relative efficiency (relative to pel method), bias, standard deviation. Following that is the simulation with incremental sample size for $(a, b)$ estimation. We also provide simulation for 95% confidence intervals of AUC.

3.5.1 Correct specification

3.5.1.1 Normal distribution

The distribution of $X$ and $Y$ are chosen to be normally distributed. The true parameters are chose as $\mu_D = 1$, $\sigma_D = 1$ for nondiseased subjects and $\mu_D = \{3.311, 2.517, 1.945\}$, $\sigma_D = 1.5$ for diseased subjects. The simulation results are summarized in Tables 3.3, 3.4 and 3.5.

First, we compare the pel method with mle, pmle, labroc and mlnm in terms of $(a, b)$ estimation. The sample size is $(n_D, n_D) = (50, 50)$. The results are shown in Table 3.3. Then based on the estimators of $(a, b)$ in Table 3.3, we compute ROC curve and AUC for the data (see Table 3.4), using the semiparametric methods (mle, pmle, labroc), parametric method (mlnm) and nonparametric method np. In Table 3.5, we summarize the large sample results of our proposed pel method, mle method and pmle method. We consider sample sizes $(n_D, n_D) = (50, 50), (80, 80), (100, 100)$ and $(120, 120)$. Finally, the coverage probability and average width of confidence intervals of AUC are compared among all the parametric (mlnm), semiparametric (mle, pmle, labroc and trapezoidal labroct) and nonparametric (Mann-Whitney np) methods. In addition, we compare the results with and without logit transformation.
Table 3.3: Estimation of \((a,b)\) from normal distribution, sample sizes 
\((n_D,n_D) = (50,50)\), 1000 times simulation

<table>
<thead>
<tr>
<th>AUC</th>
<th>mle</th>
<th>pmle</th>
<th>pel</th>
<th>labroc</th>
<th>mlnm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>0.9</td>
<td>e</td>
<td>1.170</td>
<td>1.287</td>
<td>1.214</td>
<td>1.312</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.321</td>
<td>0.191</td>
<td>0.327</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.113</td>
<td>0.063</td>
<td>0.110</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.390</td>
<td>0.180</td>
<td>0.388</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.279</td>
<td>0.176</td>
<td>0.270</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.967</td>
<td>0.945</td>
<td>0.927</td>
<td>0.819</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>1.095</td>
<td>0.719</td>
<td>1.060</td>
<td>0.553</td>
</tr>
<tr>
<td>0.8</td>
<td>e</td>
<td>1.065</td>
<td>1.129</td>
<td>1.055</td>
<td>1.085</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.222</td>
<td>0.137</td>
<td>0.221</td>
<td>0.135</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.069</td>
<td>0.045</td>
<td>0.046</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.211</td>
<td>0.129</td>
<td>0.216</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.212</td>
<td>0.135</td>
<td>0.208</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.962</td>
<td>0.953</td>
<td>0.947</td>
<td>0.884</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.832</td>
<td>0.541</td>
<td>0.817</td>
<td>0.455</td>
</tr>
<tr>
<td>0.7</td>
<td>e</td>
<td>1.068</td>
<td>1.120</td>
<td>1.037</td>
<td>1.018</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.194</td>
<td>0.121</td>
<td>0.191</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.054</td>
<td>0.047</td>
<td>0.025</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.186</td>
<td>0.111</td>
<td>0.190</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.190</td>
<td>0.119</td>
<td>0.187</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.960</td>
<td>0.954</td>
<td>0.954</td>
<td>0.905</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.745</td>
<td>0.477</td>
<td>0.733</td>
<td>0.417</td>
</tr>
</tbody>
</table>
Table 3.4: Estimation of AUC and ROC from normal setup, sample sizes $(n_D, n_D) = (50, 50)$, 1000 times simulation

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>0.9</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>e</td>
<td>rmse</td>
<td>BS</td>
<td>SD</td>
<td>e</td>
<td>rmse</td>
<td>BS</td>
<td>SD</td>
<td>e</td>
</tr>
<tr>
<td>mle</td>
<td>d</td>
<td>0.933</td>
<td>0.030</td>
<td>0.004</td>
<td>0.030</td>
<td>0.963</td>
<td>0.043</td>
<td>0.007</td>
<td>0.042</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.033</td>
<td>0.069</td>
<td>0.005</td>
<td>0.068</td>
<td>1.056</td>
<td>0.082</td>
<td>0.003</td>
<td>0.082</td>
<td>1.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.974</td>
<td>0.043</td>
<td>0.008</td>
<td>0.042</td>
<td>1.004</td>
<td>0.061</td>
<td>0.010</td>
<td>0.060</td>
<td>1.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.854</td>
<td>0.024</td>
<td>0.004</td>
<td>0.023</td>
<td>0.910</td>
<td>0.041</td>
<td>0.010</td>
<td>0.040</td>
<td>0.962</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.049</td>
<td>0.069</td>
<td>-0.009</td>
<td>0.069</td>
<td>1.001</td>
<td>0.080</td>
<td>-0.010</td>
<td>0.079</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.049</td>
<td>0.044</td>
<td>0.004</td>
<td>0.044</td>
<td>1.022</td>
<td>0.061</td>
<td>0.002</td>
<td>0.061</td>
<td>1.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.919</td>
<td>0.024</td>
<td>0.004</td>
<td>0.024</td>
<td>0.971</td>
<td>0.042</td>
<td>0.006</td>
<td>0.042</td>
<td>1.000</td>
</tr>
<tr>
<td>pmle</td>
<td>d</td>
<td>0.982</td>
<td>0.031</td>
<td>0.000</td>
<td>0.031</td>
<td>0.984</td>
<td>0.043</td>
<td>0.001</td>
<td>0.043</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000</td>
<td>0.068</td>
<td>0.002</td>
<td>0.067</td>
<td>1.000</td>
<td>0.080</td>
<td>0.002</td>
<td>0.080</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000</td>
<td>0.043</td>
<td>0.003</td>
<td>0.043</td>
<td>1.000</td>
<td>0.061</td>
<td>0.002</td>
<td>0.061</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000</td>
<td>0.026</td>
<td>0.000</td>
<td>0.026</td>
<td>1.000</td>
<td>0.043</td>
<td>0.000</td>
<td>0.043</td>
<td>1.000</td>
</tr>
<tr>
<td>pel</td>
<td>d</td>
<td>0.956</td>
<td>0.030</td>
<td>0.002</td>
<td>0.030</td>
<td>0.996</td>
<td>0.044</td>
<td>0.003</td>
<td>0.044</td>
<td>1.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.002</td>
<td>0.068</td>
<td>0.009</td>
<td>0.067</td>
<td>1.045</td>
<td>0.081</td>
<td>0.010</td>
<td>0.081</td>
<td>1.051</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.949</td>
<td>0.042</td>
<td>0.005</td>
<td>0.042</td>
<td>0.991</td>
<td>0.060</td>
<td>0.006</td>
<td>0.060</td>
<td>1.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.958</td>
<td>0.025</td>
<td>0.000</td>
<td>0.025</td>
<td>0.985</td>
<td>0.043</td>
<td>0.001</td>
<td>0.043</td>
<td>1.008</td>
</tr>
<tr>
<td>labroc</td>
<td>d</td>
<td>0.893</td>
<td>0.029</td>
<td>0.001</td>
<td>0.029</td>
<td>0.958</td>
<td>0.043</td>
<td>0.002</td>
<td>0.043</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.865</td>
<td>0.063</td>
<td>0.004</td>
<td>0.063</td>
<td>0.910</td>
<td>0.076</td>
<td>0.005</td>
<td>0.076</td>
<td>0.932</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.839</td>
<td>0.040</td>
<td>0.001</td>
<td>0.040</td>
<td>0.941</td>
<td>0.059</td>
<td>0.003</td>
<td>0.059</td>
<td>0.976</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.727</td>
<td>0.022</td>
<td>-0.001</td>
<td>0.022</td>
<td>0.871</td>
<td>0.040</td>
<td>0.000</td>
<td>0.040</td>
<td>0.944</td>
</tr>
<tr>
<td>mlnm</td>
<td>d</td>
<td>1.003</td>
<td>0.031</td>
<td>0.000</td>
<td>0.031</td>
<td>1.016</td>
<td>0.044</td>
<td>0.000</td>
<td>0.044</td>
<td>1.033</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.254</td>
<td>0.076</td>
<td>0.010</td>
<td>0.075</td>
<td>1.350</td>
<td>0.092</td>
<td>0.016</td>
<td>0.091</td>
<td>1.395</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.404</td>
<td>0.051</td>
<td>0.002</td>
<td>0.051</td>
<td>1.465</td>
<td>0.073</td>
<td>0.005</td>
<td>0.073</td>
<td>1.471</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.477</td>
<td>0.031</td>
<td>0.000</td>
<td>0.031</td>
<td>1.386</td>
<td>0.051</td>
<td>0.002</td>
<td>0.051</td>
<td>1.411</td>
</tr>
</tbody>
</table>

|            | R(0.1) | 1.254 | 0.076 | 0.010 | 0.075 | 1.350 | 0.092 | 0.016 | 0.091 | 1.395 | 0.093 | 0.018 | 0.091 |
|            |     | 1.404 | 0.051 | 0.002 | 0.051 | 1.465 | 0.073 | 0.005 | 0.073 | 1.471 | 0.083 | 0.010 | 0.083 |
|            |     | 1.477 | 0.031 | 0.000 | 0.031 | 1.386 | 0.051 | 0.002 | 0.051 | 1.411 | 0.066 | 0.003 | 0.066 |

68
<table>
<thead>
<tr>
<th>AUC</th>
<th>0.9</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rmse</td>
<td>BS</td>
</tr>
<tr>
<td></td>
<td>mle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pmle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pel</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>mle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pmle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pel</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>mle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pmle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pel</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>mle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pmle</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
<tr>
<td></td>
<td>pel</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
</tr>
</tbody>
</table>

Table 3.5: Estimation of \((a, b)\) from normal setup, incremental sample sizes, 1000 simulation.
Table 3.6: 95% confidence interval for AUC from normal distribution, sample sizes \((n_D, n_D) = (50, 50)\), 1000 times simulation

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>ap</th>
<th>mlnm</th>
<th>mle</th>
<th>pmle</th>
<th>pelni</th>
<th>pel</th>
<th>labroc</th>
<th>labroct</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLT</td>
<td>0.900</td>
<td>CP 0.923 &amp; 0.936 &amp; 0.915 &amp; 0.920 &amp; 0.925 &amp; 0.915 &amp; 0.924 &amp; 0.935</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WD 0.122 &amp; 0.117 &amp; 0.116 &amp; 0.116 &amp; 0.122 &amp; 0.117 &amp; 0.119 &amp; 0.124</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.800</td>
<td>CP 0.945 &amp; 0.944 &amp; 0.929 &amp; 0.933 &amp; 0.929 &amp; 0.933 &amp; 0.936 &amp; 0.941</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WD 0.175 &amp; 0.169 &amp; 0.167 &amp; 0.167 &amp; 0.184 &amp; 0.174 &amp; 0.170 &amp; 0.173</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.700</td>
<td>CP 0.954 &amp; 0.949 &amp; 0.940 &amp; 0.943 &amp; 0.927 &amp; 0.949 &amp; 0.942 &amp; 0.946</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WD 0.208 &amp; 0.198 &amp; 0.199 &amp; 0.198 &amp; 0.232 &amp; 0.209 &amp; 0.201 &amp; 0.203</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LT</td>
<td>0.900</td>
<td>CP 0.951 &amp; 0.955 &amp; 0.957 &amp; 0.942 &amp; 0.939 &amp; 0.939 &amp; 0.958 &amp; 0.969</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WD 0.126 &amp; 0.121 &amp; 0.120 &amp; 0.120 &amp; 0.126 &amp; 0.121 &amp; 0.124 &amp; 0.129</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.800</td>
<td>CP 0.956 &amp; 0.960 &amp; 0.953 &amp; 0.954 &amp; 0.929 &amp; 0.953 &amp; 0.957 &amp; 0.956</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WD 0.176 &amp; 0.169 &amp; 0.168 &amp; 0.168 &amp; 0.183 &amp; 0.174 &amp; 0.171 &amp; 0.173</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.700</td>
<td>CP 0.960 &amp; 0.960 &amp; 0.960 &amp; 0.956 &amp; 0.914 &amp; 0.961 &amp; 0.957 &amp; 0.955</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WD 0.206 &amp; 0.197 &amp; 0.197 &amp; 0.197 &amp; 0.228 &amp; 0.207 &amp; 0.199 &amp; 0.201</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From Table 3.3, we can see that for the estimation of \((a, b)\),

1. All the biases are close to 0 which means the estimations are consistent.

2. The \textit{mlnm} method gives the best results since the model is correctly specified.

3. Our \textit{pel} method gives the biases smaller than the \textit{mle} and \textit{pmle} methods, and quite comparable to the \textit{labroc} method. This indicates that our \textit{pel} method is a consistent estimator of \((a, b)\). We also notice that the difference between the biases of the \textit{pel} method and the \textit{pmle, mle} methods, are increasing when AUC increases.

4. Our proposed \textit{pel} method is comparable to \textit{labroc} method in terms of efficiency, where for \textit{labroc} they are as much as 1.8% and 8.1% greater than 1 for \((a, b)\) respectively.

5. Our \textit{pel} method is uniformly better than the \textit{pmle} method for \((a, b)\) estimation in terms of bias, standard deviation and mean square error. When \(AUC = \ldots\)
0.9, the relative efficiency of the \textit{pmle} method for $(a, b)$ are (121.4\%, 131.2\%) respectively. When $AUC = 0.7$, the relative efficiency of the \textit{pmle} method for $(a, b)$ are (103.7\%, 101.8\%) respectively. Therefore, the relative efficiencies decrease when AUC decreases for the \textit{pmle} method.

6. The \textit{pel} method is quite comparable to \textit{mle} method to estimate the binormal parameters in terms of biases, standard deviation and mean square error.

7. The percent differences of standard deviation and standard error are relative small for \textit{mle} and \textit{labroc} method. For $AUC = 0.9$, the percent differences are 15.4\% and 27.3\% for \textit{pel} method, 12.2\% and 18.1\% for \textit{pmle} method. However the percent differences are only 2.2\% for \textit{mle} method and 0.1\% for \textit{labroc}. As AUC decreases, the percent difference between standard deviation and standard error also decreases. When $AUC = 0.7$, the percent differences reduce to 2.2\% and 10.9\% for \textit{pel} method.

8. The average width of 95\% confidence intervals of $(a, b)$ based on \textit{pel} method are shorter than those based on \textit{mle} and \textit{pmle} method. The coverage probability of 95\% confidence interval of parameter $a$ based on \textit{pel} method is close to the nominal level 95\% for low and medium value of AUC. The coverage probability of confidence interval for parameter $b$ based on \textit{ple} method is far below the nominal level 0.95, especially for high value of AUC. The coverage probabilities of \textit{mle} estimator are close to the nominal value for all values of AUC.

From Table 3.4, we can see that

1. As expected, \textit{mlnm} method is the most efficient method for estimation of ROC curve and AUC.

2. The relative efficiency of the nonparametric method is close to one for estimation of AUC. However, the nonparametric method is not efficient at all when estimate
the ROC curve.

3. The *pel* method gives the estimation for AUC and ROC close to other semi-parametric methods. In particular, when $AUC = 0.7$, the relative efficiencies for *labroc* method are all greater than 1. In general, the estimation of $R(0.1)$ based on *pel* method is more efficient than that of *mle*, *pmle* and *labroc* method. For $d$, $R(0.3)$ and $R(0.6)$, the relative efficiencies are mostly within 5% below 1 for the *mle*, *pmle* and *labroc* method. Therefore, our *pel* method is quite comparable to the *mle*, *pmle* and *labroc* methods, in terms of AUC and ROC estimation.

As shown in Table 3.5,

1. The standard errors of the *pmle* and *pel* method underestimate their standard deviations.

2. The percent difference between the standard deviation and the standard error decreases with increasing sample sizes or decreasing AUC. For example, when $(n_D, n_D) = (50, 50)$ and $AUC = 0.9$, the percent differences between the standard deviation and the standard error of $(a, b)$, are (15.4%, 27.3%) for *pel* method, (12.2%, 18.1%) for *pmle* method. When the sample sizes increase to $(n_D, n_D) = (120, 120)$, the percent differences between standard deviation and standard error drop to (11.6%, 18.9%) for *pel* method, (9.4%, 13.6%) for *pmle* method.

3. The percent difference between the standard deviation and the standard error decreases at lower AUC value. For example, when $AUC = 0.8$ and $(n_D, n_D) = (120, 120)$, the percent differences decrease to (4.1%, 13.5%) for *pel* method, to (4.3%, 8.8%) for *pmle* method.
4. The coverage probabilities for \textit{pmle} and \textit{pel} method are less than nominal level 95%, but they approach the nominal level 95% as the sample sizes increase or AUC decreases.

From Table 3.6, we can see

1. As expected, the \textit{mlnm} method gives the accurate confidence interval estimation, since it is based on correct parametric model.

2. The coverage probabilities after logit transformation (LT) are closer to the nominal level 95% than those without logit transformation (NLT). The average widths after logit transformation are slightly wider than those without logit transformation.

3. For the confidence intervals using logit transformation, the coverage probabilities of \textit{pel} method are less than 95% nominal level when \textit{AUC} = 0.9. The average widths of \textit{pel} method are narrower than \textit{np}, \textit{labroc} and \textit{labroct} method when \textit{AUC} = 0.9. When \textit{AUC} = \{0.8, 0.7\}, the coverage probability of confidence interval based on \textit{pel} method is larger than 95% level, and the average width is larger than the other methods except \textit{pelni} method.

4. \textit{pelni} method has the coverage probabilities always less than the nominal level 95%. The average widths of \textit{pelni} method are larger than the other methods.

\subsection*{3.5.1.2 Lognormal distribution}

The distribution of \(X\) and \(Y\) are chosen to have lognormal distribution with \(\log X \sim N(1, 1), \log Y \sim N(\mu_D, 2.25)\) and \(\mu_D = \{3.311, 2.517, 1.945\}\). The true parameter \(b = \frac{2}{3}\) and \(AUC = \{0.9, 0.8, 0.7\}\). The simulation results are presented in Tables 3.7, 3.8, 3.9 and 3.10. By comparing the performance measures of all estimators, similar findings are observed.
Table 3.7: Estimation of \((a, b)\) from lognormal setup, sample sizes \((n_D, n_D) = (50, 50)\), 1000 times simulation

<table>
<thead>
<tr>
<th>AUC</th>
<th>mle</th>
<th>pmle</th>
<th>pel</th>
<th>labroc</th>
<th>mlnm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
<td>(a)</td>
<td>(b)</td>
<td>(a)</td>
</tr>
<tr>
<td>0.9</td>
<td>e</td>
<td>1.181</td>
<td>1.313</td>
<td>1.215</td>
<td>1.325</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.343</td>
<td>0.194</td>
<td>0.348</td>
<td>0.195</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.199</td>
<td>0.061</td>
<td>0.106</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.325</td>
<td>0.184</td>
<td>0.331</td>
<td>0.171</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.279</td>
<td>0.176</td>
<td>0.269</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.959</td>
<td>0.956</td>
<td>0.921</td>
<td>0.819</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>1.095</td>
<td>0.718</td>
<td>1.054</td>
<td>0.548</td>
</tr>
<tr>
<td>0.8</td>
<td>e</td>
<td>1.054</td>
<td>1.126</td>
<td>1.048</td>
<td>1.075</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.233</td>
<td>0.139</td>
<td>0.232</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.060</td>
<td>0.043</td>
<td>0.038</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.225</td>
<td>0.133</td>
<td>0.229</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.212</td>
<td>0.134</td>
<td>0.208</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.951</td>
<td>0.953</td>
<td>0.941</td>
<td>0.889</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.830</td>
<td>0.538</td>
<td>0.815</td>
<td>0.454</td>
</tr>
<tr>
<td>0.7</td>
<td>e</td>
<td>1.048</td>
<td>1.096</td>
<td>1.031</td>
<td>1.064</td>
</tr>
<tr>
<td></td>
<td>rmse</td>
<td>0.199</td>
<td>0.123</td>
<td>0.198</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>BS</td>
<td>0.046</td>
<td>0.045</td>
<td>0.016</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.194</td>
<td>0.115</td>
<td>0.197</td>
<td>0.111</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.190</td>
<td>0.119</td>
<td>0.187</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.949</td>
<td>0.954</td>
<td>0.950</td>
<td>0.911</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.743</td>
<td>0.475</td>
<td>0.732</td>
<td>0.415</td>
</tr>
</tbody>
</table>
Table 3.8: Estimations for AUC and ROC from lognormal setup, sample sizes

\[(n_D, n_D) = (50, 50), \text{ 1000 times simulation}\]

<table>
<thead>
<tr>
<th>AUC</th>
<th>0.9</th>
<th>0.8</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>e</td>
<td>rmse</td>
<td>BS</td>
</tr>
<tr>
<td>mle</td>
<td>d</td>
<td>0.929</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.031</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.966</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.845</td>
<td>0.025</td>
</tr>
<tr>
<td>pmle</td>
<td>d</td>
<td>0.981</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.054</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.040</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.911</td>
<td>0.026</td>
</tr>
<tr>
<td>pel</td>
<td>d</td>
<td>1.000</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.000</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.000</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.000</td>
<td>0.027</td>
</tr>
<tr>
<td>labroc</td>
<td>d</td>
<td>0.965</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.998</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.958</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.976</td>
<td>0.027</td>
</tr>
<tr>
<td>mlm</td>
<td>d</td>
<td>0.903</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>0.884</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>0.860</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.763</td>
<td>0.024</td>
</tr>
<tr>
<td>np</td>
<td>d</td>
<td>0.993</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.271</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.335</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.421</td>
<td>0.032</td>
</tr>
</tbody>
</table>
Table 3.9: Estimation for \((a, b)\) from lognormal setup, incremental sample sizes, 1000 times simulation

<table>
<thead>
<tr>
<th>AUC</th>
<th>0.9</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rmse</td>
<td>BS</td>
</tr>
<tr>
<td>(50,50) mle a</td>
<td>0.343</td>
<td>0.109</td>
</tr>
<tr>
<td>b</td>
<td>0.194</td>
<td>0.061</td>
</tr>
<tr>
<td>pmle a</td>
<td>0.348</td>
<td>0.106</td>
</tr>
<tr>
<td>b</td>
<td>0.195</td>
<td>0.094</td>
</tr>
<tr>
<td>pel a</td>
<td>0.316</td>
<td>0.059</td>
</tr>
<tr>
<td>b</td>
<td>0.169</td>
<td>0.031</td>
</tr>
<tr>
<td>(80,80) mle a</td>
<td>0.239</td>
<td>0.070</td>
</tr>
<tr>
<td>b</td>
<td>0.139</td>
<td>0.037</td>
</tr>
<tr>
<td>pmle a</td>
<td>0.243</td>
<td>0.068</td>
</tr>
<tr>
<td>b</td>
<td>0.142</td>
<td>0.058</td>
</tr>
<tr>
<td>pel a</td>
<td>0.230</td>
<td>0.041</td>
</tr>
<tr>
<td>b</td>
<td>0.131</td>
<td>0.020</td>
</tr>
<tr>
<td>(100,100) mle a</td>
<td>0.200</td>
<td>0.058</td>
</tr>
<tr>
<td>b</td>
<td>0.120</td>
<td>0.031</td>
</tr>
<tr>
<td>pmle a</td>
<td>0.204</td>
<td>0.057</td>
</tr>
<tr>
<td>b</td>
<td>0.123</td>
<td>0.049</td>
</tr>
<tr>
<td>pel a</td>
<td>0.194</td>
<td>0.036</td>
</tr>
<tr>
<td>b</td>
<td>0.115</td>
<td>0.019</td>
</tr>
<tr>
<td>(120,120) mle a</td>
<td>0.178</td>
<td>0.043</td>
</tr>
<tr>
<td>b</td>
<td>0.105</td>
<td>0.026</td>
</tr>
<tr>
<td>pmle a</td>
<td>0.180</td>
<td>0.042</td>
</tr>
<tr>
<td>b</td>
<td>0.108</td>
<td>0.041</td>
</tr>
<tr>
<td>pel a</td>
<td>0.174</td>
<td>0.025</td>
</tr>
<tr>
<td>b</td>
<td>0.102</td>
<td>0.016</td>
</tr>
</tbody>
</table>
Table 3.10: 95% confidence interval for AUC from lognormal distributions, sample sizes \((n_D, n_D) = (50, 50)\), 1000 simulations

<table>
<thead>
<tr>
<th>AUC</th>
<th>np</th>
<th>plmn</th>
<th>plmle</th>
<th>pelni</th>
<th>pel</th>
<th>labroc</th>
<th>labroct</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>CP</td>
<td>0.928</td>
<td>0.921</td>
<td>0.910</td>
<td>0.915</td>
<td>0.924</td>
<td>0.910</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.122</td>
<td>0.118</td>
<td>0.117</td>
<td>0.117</td>
<td>0.122</td>
<td>0.117</td>
</tr>
<tr>
<td>0.8</td>
<td>CP</td>
<td>0.942</td>
<td>0.936</td>
<td>0.928</td>
<td>0.931</td>
<td>0.922</td>
<td>0.937</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.176</td>
<td>0.169</td>
<td>0.168</td>
<td>0.168</td>
<td>0.184</td>
<td>0.175</td>
</tr>
<tr>
<td>0.7</td>
<td>CP</td>
<td>0.950</td>
<td>0.949</td>
<td>0.937</td>
<td>0.940</td>
<td>0.905</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.208</td>
<td>0.199</td>
<td>0.199</td>
<td>0.199</td>
<td>0.234</td>
<td>0.209</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NLT</td>
<td>0.9</td>
<td>CP</td>
<td>0.941</td>
<td>0.948</td>
<td>0.947</td>
<td>0.936</td>
<td>0.920</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.126</td>
<td>0.121</td>
<td>0.121</td>
<td>0.121</td>
<td>0.126</td>
<td>0.121</td>
</tr>
<tr>
<td>0.8</td>
<td>CP</td>
<td>0.948</td>
<td>0.950</td>
<td>0.952</td>
<td>0.937</td>
<td>0.912</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.176</td>
<td>0.169</td>
<td>0.169</td>
<td>0.168</td>
<td>0.183</td>
<td>0.175</td>
</tr>
<tr>
<td>0.7</td>
<td>CP</td>
<td>0.956</td>
<td>0.949</td>
<td>0.959</td>
<td>0.951</td>
<td>0.897</td>
<td>0.956</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.206</td>
<td>0.197</td>
<td>0.197</td>
<td>0.197</td>
<td>0.229</td>
<td>0.207</td>
</tr>
</tbody>
</table>

3.5.2 Misspecification

We consider two cases of misspecification. In the first case, we choose exponential distribution and in the second case, we choose Gamma distribution.

3.5.2.1 Exponential distribution

\(X\) and \(Y\) are chosen to have exponential distributions with different rates, i.e., \(X \sim \text{Exp}(r_D)\) and independently, \(Y \sim \text{Exp}(r_D)\). The parameters are \(r_D = 1\) and \(r_D = \{0.1, 0.25, 0.43\}\). By the AUC formula for exponential setup \(d = \frac{r_D - r_D}{r_D + r_D}\), we obtain the corresponding \(AUC = \{0.91, 0.8, 0.7\}\). The sample sizes for the diseased and nondiseased are both chosen to be 50. Each setup is simulated for 1000 times. The simulation results are listed in Tables 3.11 and 3.12. In Table 3.11, we list the results for the semiparametric, parametric and nonparametric methods. Estimations are made for the area under the curve \(d\), ROC curves at \(s = \{0.1, 0.3, 0.6\}\). For each estimation, we provide bias, standard deviation, \(rmse\) and the relative efficiency \(e\)
(which is relative to the corresponding results from pel method). In Table 3.12, we list the 95% confidence interval estimation for AUC. The parametric mlnm, non-parametric np, semiparametric mle, pmle, pelni, pel, labroc and labroct methods are compared. The coverage probability and average width of the confidence intervals are calculated. NLT denotes the confidence interval without logit transformation, and LT denotes the confidence interval with a logit transformation.

Table 3.11: AUC and ROC estimation for Exponential distribution, sample sizes \((n_D, n_D) = (50, 50)\), 1000 simulations

<table>
<thead>
<tr>
<th>AUC</th>
<th>(0.91)</th>
<th>(0.8)</th>
<th>(0.7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>e</td>
<td>rmse</td>
<td>BS</td>
</tr>
<tr>
<td>mle</td>
<td>d</td>
<td>0.930</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.006</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.000</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.888</td>
<td>0.023</td>
</tr>
<tr>
<td>pmle</td>
<td>d</td>
<td>0.977</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.067</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.072</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>0.994</td>
<td>0.025</td>
</tr>
<tr>
<td>pel</td>
<td>d</td>
<td>1.000</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.000</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.000</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.000</td>
<td>0.025</td>
</tr>
<tr>
<td>labroc</td>
<td>d</td>
<td>1.047</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.073</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.042</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.020</td>
<td>0.025</td>
</tr>
<tr>
<td>mlnm</td>
<td>d</td>
<td>1.497</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>4.243</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.897</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.948</td>
<td>0.034</td>
</tr>
<tr>
<td>np</td>
<td>d</td>
<td>1.158</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>R(0.1)</td>
<td>1.316</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>R(0.3)</td>
<td>1.596</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>R(0.6)</td>
<td>1.779</td>
<td>0.033</td>
</tr>
</tbody>
</table>
Table 3.12: 95\% confidence interval for AUC from exponential distributions, sample sizes \((n_D, n_D) = (50, 50)\), 1000 times simulation

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>np</th>
<th>mlnm</th>
<th>mle</th>
<th>pmle</th>
<th>pelni</th>
<th>pel</th>
<th>labroc</th>
<th>labroct</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLT</td>
<td>0.91</td>
<td>0.903</td>
<td>0.898</td>
<td>0.901</td>
<td>0.906</td>
<td>0.909</td>
<td>0.900</td>
<td>0.902</td>
<td>0.912</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.118</td>
<td>0.116</td>
<td>0.111</td>
<td>0.123</td>
<td>0.114</td>
<td>0.115</td>
<td>0.115</td>
<td>0.118</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.927</td>
<td>0.910</td>
<td>0.937</td>
<td>0.939</td>
<td>0.927</td>
<td>0.937</td>
<td>0.924</td>
<td>0.927</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.174</td>
<td>0.175</td>
<td>0.166</td>
<td>0.178</td>
<td>0.168</td>
<td>0.169</td>
<td>0.169</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.933</td>
<td>0.923</td>
<td>0.942</td>
<td>0.941</td>
<td>0.916</td>
<td>0.942</td>
<td>0.930</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.206</td>
<td>0.202</td>
<td>0.197</td>
<td>0.195</td>
<td>0.205</td>
<td>0.199</td>
<td>0.199</td>
<td>0.203</td>
</tr>
<tr>
<td>LT</td>
<td>0.91</td>
<td>0.939</td>
<td>0.883</td>
<td>0.960</td>
<td>0.949</td>
<td>0.932</td>
<td>0.942</td>
<td>0.956</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.124</td>
<td>0.119</td>
<td>0.116</td>
<td>0.116</td>
<td>0.128</td>
<td>0.119</td>
<td>0.121</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.945</td>
<td>0.898</td>
<td>0.958</td>
<td>0.947</td>
<td>0.922</td>
<td>0.949</td>
<td>0.945</td>
<td>0.946</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.175</td>
<td>0.175</td>
<td>0.167</td>
<td>0.165</td>
<td>0.178</td>
<td>0.169</td>
<td>0.170</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.944</td>
<td>0.916</td>
<td>0.957</td>
<td>0.957</td>
<td>0.916</td>
<td>0.953</td>
<td>0.942</td>
<td>0.941</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.204</td>
<td>0.200</td>
<td>0.195</td>
<td>0.193</td>
<td>0.203</td>
<td>0.197</td>
<td>0.197</td>
<td>0.201</td>
</tr>
</tbody>
</table>

From Table 3.11, we observe

1. The \(mlnm\) method is the misspecification case, therefore, yields inconsistent results.

2. The \(np\) method is less efficient than the semiparametric methods \((mle, pmle, labroc\) and \(pel\)). In particular, the relative efficiencies for \(np\) method are up to \((17.9\%, 42.4\%, 59.6\%, 77.9\%)\) greater than 1 for \(d\), \(R(0.1)\), \(R(0.3)\) and \(R(0.6)\) respectively.

3. The relative efficiencies for \(labroc\) method are all greater than 1, which means the \(pel\) method is more efficient than \(labroc\). For example, when \(AUC = 0.8\), the relative efficiencies of \(labroc\) are \((111.5\%, 108\%, 110.1\%, 114.4\%)\) for \(d\), \(R(0.1)\), \(R(0.3)\) and \(R(0.6)\), respectively.

4. The relative efficiencies for \(mle\) and \(pmle\) are greater than 1 for \(R(0.1)\) and \(R(0.3)\). The relative efficiencies for \(mle\) and \(pmle\) are close to 1 for \(d\) and
Therefore, the pel method is better than the mle and pmle method in terms of ROC estimation at high and median specificities.

From Table 3.12, we can see

1. As the misspecification case, the ml nm method gives the worst results in terms of coverage probability, even data go through a log transformation.

2. The coverage probabilities for 95% confidence intervals with logit transformation are closer to the nominal level than those without logit transformation. The average widths with logit transformation are close to or slightly larger than that without logit transformation.

3. For the confidence intervals built with logit transformation, the pel method is better than the mle and pmle method, since the coverage probability for pel method is closer to the nominal level 95% and the average width is slightly larger. The pel method is better than nonparametric np method in terms of coverage probability and average width of the confidence interval.

4. The pelni method has the coverage probabilities far below the nominal level. Moreover, the pelni method has the largest average width.

3.5.2.2 Gamma distribution

$X$ and $Y$ are chosen to have Gamma distributions with different rates, i.e., $X \sim \Gamma(1, 0.5)$ and independently, $Y \sim \Gamma(\alpha_D, 0.5)$. The parameters are $\alpha_D = \{3.25, 2.5, 1.75\}$. By the definition of AUC, we calculate the corresponding $AUC = \{0.895, 0.823, 0.703\}$. The sample sizes for the diseased and nondiseased are both chosen to be 50. Each setup is simulated for 1000 times. The simulation results are listed in Tables 3.13 and 3.14. In Table 3.13, we list the results for the semi-parametric, parametric and nonparametric methods. Estimations are made for the
area under the curve $d$, ROC curves at $s = \{0.1, 0.3, 0.6\}$. For each estimation, we provide bias, standard deviation, $\text{rmse}$ and the relative efficiency $e$ (which is relative to corresponding results from $\text{pel}$ method). In Table 3.14, we list the 95% confidence interval estimation for AUC.

Table 3.13: AUC and ROC estimation for Gamma distribution, sample sizes $(n_D, n_D) = (50, 50)$, 1000 times simulation

<table>
<thead>
<tr>
<th></th>
<th>0.895</th>
<th>0.823</th>
<th>0.703</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mse</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>0.966</td>
<td>0.030</td>
<td>0.000</td>
</tr>
<tr>
<td>R(0.1)</td>
<td>1.042</td>
<td>0.107</td>
<td>0.107</td>
</tr>
<tr>
<td>R(0.3)</td>
<td>1.005</td>
<td>0.044</td>
<td>0.044</td>
</tr>
<tr>
<td>R(0.6)</td>
<td>0.751</td>
<td>0.011</td>
<td>-0.002</td>
</tr>
<tr>
<td>pmle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>0.997</td>
<td>0.030</td>
<td>-0.002</td>
</tr>
<tr>
<td>R(0.1)</td>
<td>0.994</td>
<td>0.105</td>
<td>-0.002</td>
</tr>
<tr>
<td>R(0.3)</td>
<td>1.028</td>
<td>0.045</td>
<td>0.000</td>
</tr>
<tr>
<td>R(0.6)</td>
<td>0.833</td>
<td>0.011</td>
<td>-0.003</td>
</tr>
<tr>
<td>pel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>1.000</td>
<td>0.031</td>
<td>-0.001</td>
</tr>
<tr>
<td>R(0.1)</td>
<td>1.000</td>
<td>0.105</td>
<td>0.005</td>
</tr>
<tr>
<td>R(0.3)</td>
<td>1.000</td>
<td>0.044</td>
<td>-0.002</td>
</tr>
<tr>
<td>R(0.6)</td>
<td>1.000</td>
<td>0.012</td>
<td>-0.004</td>
</tr>
<tr>
<td>labroc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>0.969</td>
<td>0.030</td>
<td>0.002</td>
</tr>
<tr>
<td>R(0.1)</td>
<td>0.981</td>
<td>0.104</td>
<td>0.015</td>
</tr>
<tr>
<td>R(0.3)</td>
<td>1.006</td>
<td>0.044</td>
<td>0.002</td>
</tr>
<tr>
<td>R(0.6)</td>
<td>0.951</td>
<td>0.012</td>
<td>-0.004</td>
</tr>
<tr>
<td>mlhm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>2.138</td>
<td>0.045</td>
<td>-0.029</td>
</tr>
<tr>
<td>R(0.1)</td>
<td>0.861</td>
<td>0.098</td>
<td>0.057</td>
</tr>
<tr>
<td>R(0.3)</td>
<td>3.860</td>
<td>0.087</td>
<td>-0.074</td>
</tr>
<tr>
<td>R(0.6)</td>
<td>38.285</td>
<td>0.076</td>
<td>-0.071</td>
</tr>
<tr>
<td>np</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>1.036</td>
<td>0.031</td>
<td>-0.001</td>
</tr>
<tr>
<td>R(0.1)</td>
<td>1.265</td>
<td>0.118</td>
<td>0.020</td>
</tr>
<tr>
<td>R(0.3)</td>
<td>1.593</td>
<td>0.056</td>
<td>-0.003</td>
</tr>
<tr>
<td>R(0.6)</td>
<td>1.541</td>
<td>0.015</td>
<td>-0.001</td>
</tr>
</tbody>
</table>
Table 3.14: 95% confidence interval for AUC from Gamma distributions, sample sizes \((n_D, n_D) = (50, 50)\), 1000 times simulation

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>np</th>
<th>mlnm</th>
<th>mle</th>
<th>pmle</th>
<th>pelni</th>
<th>pel</th>
<th>labroc</th>
<th>labroct</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLT</td>
<td>0.895</td>
<td>0.934</td>
<td>0.970</td>
<td>0.936</td>
<td>0.921</td>
<td>0.954</td>
<td>0.920</td>
<td>0.937</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.122</td>
<td>0.139</td>
<td>0.122</td>
<td>0.117</td>
<td>0.132</td>
<td>0.117</td>
<td>0.121</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.932</td>
<td>0.971</td>
<td>0.928</td>
<td>0.922</td>
<td>0.948</td>
<td>0.920</td>
<td>0.931</td>
<td>0.937</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.162</td>
<td>0.167</td>
<td>0.158</td>
<td>0.155</td>
<td>0.169</td>
<td>0.155</td>
<td>0.158</td>
<td>0.164</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.941</td>
<td>0.964</td>
<td>0.933</td>
<td>0.938</td>
<td>0.947</td>
<td>0.937</td>
<td>0.935</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.204</td>
<td>0.198</td>
<td>0.197</td>
<td>0.195</td>
<td>0.207</td>
<td>0.197</td>
<td>0.198</td>
<td>0.203</td>
</tr>
<tr>
<td>LT</td>
<td>0.823</td>
<td>0.961</td>
<td>0.919</td>
<td>0.962</td>
<td>0.952</td>
<td>0.964</td>
<td>0.951</td>
<td>0.963</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.126</td>
<td>0.142</td>
<td>0.125</td>
<td>0.120</td>
<td>0.137</td>
<td>0.120</td>
<td>0.125</td>
<td>0.132</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.949</td>
<td>0.957</td>
<td>0.957</td>
<td>0.953</td>
<td>0.966</td>
<td>0.954</td>
<td>0.948</td>
<td>0.959</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.163</td>
<td>0.168</td>
<td>0.159</td>
<td>0.156</td>
<td>0.170</td>
<td>0.156</td>
<td>0.160</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.943</td>
<td>0.966</td>
<td>0.949</td>
<td>0.944</td>
<td>0.955</td>
<td>0.946</td>
<td>0.943</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>WD</td>
<td>0.202</td>
<td>0.197</td>
<td>0.195</td>
<td>0.193</td>
<td>0.205</td>
<td>0.195</td>
<td>0.196</td>
<td>0.201</td>
</tr>
</tbody>
</table>

From Table 3.13, we can see

1. The \textit{mlnm} method yields inconsistent results, since it is the misspecification case.

2. The \textit{np} method yields the relative efficiencies up to 8.5\%, 41.8\%, 60.6\%, 56.1\% greater than 1, for \(d, R(0.1), R(0.3), R(0.6)\) respectively. Therefore, the \textit{pel} is more efficient than the \textit{np} method.

3. The relative efficiencies for \textit{labroc} are greater than 1 when \(AUC = 0.823\) and 0.703. Thus the \textit{pel} method is more efficient than the \textit{labroc} method at median and low AUCs.

4. The \textit{pel} method gives the estimations for AUC and ROC very close to those from the \textit{mle} and \textit{pmle} method since the relative efficiencies for \textit{mle} and \textit{pmle} method are close to 1.

From Table 3.14, we observe that
1. As expected, the \textit{mlnm} method is the misspecification cases, therefore, gives inconsistent results.

2. The coverage probabilities after logit transformation are closer to the nominal level 95\% than those without logit transformation. The average widths after logit transformation are slightly larger than those without the logit transformation.

3. The \textit{pel} method is better than the \textit{np}, \textit{labroc} and \textit{labroct} method, in terms of both coverage probability and average width. The \textit{pel} method is better than the \textit{mle} and \textit{pmle} method since that the coverage probability is closer to the nominal level 95\% and similar average width.

4. The \textit{pelni} method has coverage probabilities less than the nominal level 95\%. Its average widths are slightly larger than the other methods.

### 3.6 Summary

Based on all the simulation results, we can see that

1. Our proposed \textit{pel} method is more efficient than the nonparametric \textit{np} method. The \textit{pel} method is more efficient than our \textit{pelni} method, especially when sample sizes are small or AUC is large.

2. The \textit{pel} method is more robust than the full parametric \textit{mlnm} method.

3. In the case when binormal model is correctly specified, our \textit{pel} method is uniformly better than the \textit{pmle} method for \((a,b)\) estimation in terms of bias, standard deviation and mean square error. The \textit{pel} method is quite comparable to \textit{mle} method to estimate the binormal parameters. As to the AUC and ROC
estimation, our *pel* method is within 5% difference in mean square error comparing with the *mle* and *pmle* methods.

4. In the case when binormal model is misspecified, the *pel* method still provides the estimations of ROC and AUC similar to other semiparametric *mle* and *pmle* method. The *pel* method is even more efficient than *labroc*.

5. The standard error from the *pel* method is smaller than the standard deviation for the estimation of $b$. The percent difference between standard error and standard deviation decreases when the sample size increases or the AUC decreases.

6. The *pel* method can provide the confidence interval for AUC comparable to other semiparametric methods.

### 3.7 Proofs

It is shown in Lemma 3.7.1 that $\hat{h}(x; a_0, b_0)$ is a consistent estimator of $h_0(x)$. Remark 3.7.1 shows that $\hat{h}(x; a, b)$ is a function of $(a, b)$.

**Lemma 3.7.1.** Let $\hat{h}(x; a, b)$ denote the solution of the equation (3.3). Then as $n \to \infty$,

$$\hat{h}(x; a_0, b_0) \xrightarrow{p} h_0(x), \text{ for each fixed } x.$$

**Proof.** As $n \to \infty$,

$$\hat{H}(x) = \frac{1}{n} \sum_{i=1}^{n} I(t_i \leq x) = \frac{1}{n} \sum_{k=1}^{n_D} I(x_k \leq x) + \frac{1}{n} \sum_{j=1}^{n_D} I(y_j \leq x)$$

$$\xrightarrow{p} \frac{n_D}{n} \mathbb{E}I(X \leq x) + \frac{n_D}{n} \mathbb{E}I(Y \leq x)$$

$$\xrightarrow{p} \frac{1}{1 + \rho} \mathbb{E}I(h_0(X) \leq h_0(x)) + \frac{\rho}{1 + \rho} \mathbb{E}I(h_0(Y) \leq h_0(x))$$
\[
= \frac{1}{1 + \rho} \Phi(h_0(x)) + \frac{\rho}{1 + \rho} \Phi(b_0 h_0(x) - a_0),
\]

and \(H(h_0(x), \theta_0) \to \frac{1}{1+\rho} \Phi(h_0(x)) + \frac{\rho}{1+\rho} \Phi(b_0 h_0(x) - a_0)\) for each fixed \(x\). By the uniform law of large numbers according to [Pollard, 1990, p.41], \(\hat{H}(x) - H(h_0(x); a_0, b_0) \to 0\) almost surely, uniformly in \(x\). It follows from the monotonicity and continuity of \(\Phi(\cdot)\) that there exists a unique \(\hat{h}(x; a, b)\) to equation (3.3) for any given \((a, b)\). Therefore,

\[
\hat{h}(x; a_0, b_0) \xrightarrow{p} h_0(x), \text{ for each fixed } x.
\]

**Remark 3.7.1.** Let \(\hat{h}(x; a, b)\) be the solution of the equation (3.3). \(\hat{h}(x; a, b)\) is a function of \((a, b)\). That is \(\frac{\partial \hat{h}(x; a, b)}{\partial (a, b)} \neq 0\).

**Proof.** We take derivative to (3.3) with respect to \((a, b)\), and solve for \(\frac{\partial \hat{h}(x; a, b)}{\partial (a, b)}\):

\[
\frac{\partial \hat{h}(x; a, b)}{\partial a} = \frac{\rho \rho w(h_0(x); a, b)}{\rho \rho w(h_0(x); a, b) + b} = \frac{\rho w(h_0(x); a, b)}{b + \rho w(h_0(x); a, b)},
\]

\[
\frac{\partial \hat{h}(x; a, b)}{\partial b} = -\hat{h}(x; a, b) \frac{\partial h(x; a, b)}{\partial a}.
\]

Since Lemma 3.7.1 shows that as \(n \to \infty\), \(\hat{h}(x; a_0, b_0) \xrightarrow{p} h_0(x)\) for each fixed \(x\), we have as \(n \to \infty\),

\[
\frac{\partial \hat{h}(x; a, b)}{\partial a} \bigg|_{(a_0, b_0)} \xrightarrow{p} \frac{1}{b_0} \frac{\rho w(h_0(x); a_0, b_0)}{1 + \rho w(h_0(x); a_0, b_0)},
\]

\[
\frac{\partial \hat{h}(x; a, b)}{\partial b} \bigg|_{(a_0, b_0)} \xrightarrow{p} -\frac{h_0(x)}{b_0} \frac{\rho w(h_0(x); a_0, b_0)}{1 + \rho w(h_0(x); a_0, b_0)},
\]

for each fixed \(x\).
Denote \( \hat{h}(x) = \hat{h}(x; a, b) \). Let

\[
V_{n1}(\lambda, a, b) = -\sum_{i=1}^{n} \frac{w(\hat{h}(t_i); a, b) - 1}{1 + \lambda_0 [w(\hat{h}(t_i); \theta_0) - 1]},
\]

\[
V_{n2}(\lambda, a, b) = \sum_{j=1}^{nD} [\hat{h}(y_j) - a] - \sum_{i=1}^{n} \frac{\lambda w(\hat{h}(t_i); a, b) \cdot \left| \hat{h}(t_i) - a \right|}{1 + \lambda [w(\hat{h}(t_i); a, b) - 1]},
\]

\[
V_{n3}(\lambda, a, b) = \sum_{j=1}^{nD} \left[ 1 + a \hat{h}(y_j) - b \hat{h}^2(y_j) \right] - \sum_{i=1}^{n} \frac{\lambda w(\hat{h}(t_i); a, b) \cdot \left[ 1 + a \hat{h}(t_i) - b \hat{h}^2(t_i) \right]}{1 + \lambda [w(\hat{h}(t_i); a, b) - 1]}.
\]

and

\[
V_n(\lambda, a, b) = \left( \begin{array}{c}
V_{n1}(\lambda, a, b) \\
V_{n2}(\lambda, a, b) \\
V_{n3}(\lambda, a, b)
\end{array} \right).
\]

(\( \hat{\lambda}, \hat{a}, \hat{b} \)) is the solution of \( V_n(\lambda, a, b) = 0 \). The following Lemma 3.7.2 shows that \( (\hat{\lambda}, \hat{a}, \hat{b}) \) is consistent estimator of \((\lambda_0, a_0, b_0)\).

**Lemma 3.7.2.** As \( n \to \infty \),

\[
\frac{1}{n} V_n(\lambda_0, a_0, b_0) \overset{p}{\to} 0,
\]

where \( \lambda_0 = \frac{\rho}{1 + \rho} \) and \( \rho = \lim_{n \to \infty} \frac{nD}{nD} \).

**Proof.** Let \( \rho = \lim_{n \to \infty} \frac{nD}{nD} \) and \( \lambda_0 = \frac{\rho}{1 + \rho} \). Let \( \theta_0 = (\lambda_0, a_0, b_0)^\tau \) be the true value of \( \theta = (\lambda, a, b)^\tau \). We write \( w(h(x); \theta) \) and \( \hat{h}(x; \theta) \) for simplicity, while exactly they are functions of \((a, b)\) only. First, as \( n \to \infty \)

\[
\frac{V_{n1}(\theta_0)}{n} = -\frac{1}{n} \sum_{i=1}^{n} \frac{w(\hat{h}(t_i); \theta_0) - 1}{1 + \lambda_0 [w(\hat{h}(t_i); \theta_0) - 1]}
\]

\[
= -\frac{1}{n} \sum_{k=1}^{nD} \frac{w(\hat{h}(x_k); \theta_0) - 1}{1 + \lambda_0 [w(\hat{h}(x_k); \theta_0) - 1]} - \frac{1}{n} \sum_{j=1}^{nD} \frac{w(\hat{h}(y_j); \theta_0) - 1}{1 + \lambda_0 [w(\hat{h}(y_j); \theta_0) - 1]}
\]

\[
\overset{p}{\to} -\frac{1}{1 + \rho} \int \frac{w(h(x); \theta_0) - 1}{1 + \lambda_0 [w(h(x); \theta_0) - 1]} dG(x) - \frac{\rho}{1 + \rho} \int \frac{w(h(x); \theta_0) - 1}{1 + \lambda_0 [w(h(x); \theta_0) - 1]} dF(x)
\]

\[
= -\int \frac{(1 - \lambda_0) [w(h(x); \theta_0) - 1]}{1 + \lambda_0 [w(h(x); \theta_0) - 1]} dG(x) \quad \text{and} \quad -\int \frac{\lambda_0 w(h(x); \theta_0) - 1}{1 + \lambda_0 [w(h(x); \theta_0) - 1]} dF(x)
\]

\[
= -\int \frac{\{1 + \lambda_0 [w(h(x); \theta_0) - 1]\} [w(h(x); \theta_0) - 1]}{1 + \lambda_0 [w(h(x); \theta_0) - 1]} dG(x)
\]

86
\[-\int \left[ w(h_0(x); \theta_0) - 1 \right] dG(x) = 0 \]

Second, as \( n \to \infty \):

\[
\frac{V_{n2}(\theta_0)}{n} = \frac{1}{n} \sum_{j=1}^{n}\left[ b_0 \hat{h}(y_j; \theta_0) - a_0 \right] - \frac{1}{n} \sum_{j=1}^{n} \frac{\lambda_0 w(\hat{h}(t_i; \theta_0) \mid b_0 \hat{h}(t_i; \theta_0) - a_0)}{1 + \lambda_0 [w(\hat{h}(t_i; \theta_0) \mid \theta_0) - 1]}
\]

\[
= \frac{1}{n} \sum_{j=1}^{n}\left[ b_0 \hat{h}(y_j; \theta_0) - a_0 \right] - \frac{1}{n} \sum_{j=1}^{n} \left\{ 1 - \frac{1 - \lambda_0}{1 + \lambda_0 [w(\hat{h}(t_i; \theta_0) \mid \theta_0) - 1]} \right\} \left[ b_0 \hat{h}(t_i; \theta_0) - a_0 \right]
\]

\[
\xrightarrow{p} \lambda_0 \int \frac{(1 - \lambda_0)[b_0 h_0(x) - a_0]w(h_0(x); \theta_0)}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} dG(x)
\]

\[
- (1 - \lambda_0) \int \left( 1 - \frac{1 - \lambda_0}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right) [b_0 h_0(x) - a_0] dG(x)
\]

\[
= (1 - \lambda_0) \left\{ \int \frac{\left[ b_0 h_0(x) - a_0 \right] \left[ 1 + \lambda_0 [w(h_0(x); \theta_0) - 1] \right]}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} dG(x) + a_0 \right\} = 0
\]

Third,

\[
\frac{V_{n3}(\theta_0)}{n} = \frac{1}{n} \sum_{j=1}^{n} \left[ \frac{1}{b_0} + a_0 \hat{h}(y_j; \theta_0) - b_0 \hat{h}^2(y_j; \theta_0) \right]
\]

\[
- \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda_0 w(\hat{h}(t_i; \theta_0) \mid b_0 \hat{h}(t_i; \theta_0) - b_0 \hat{h}^2(t_i; \theta_0))}{1 + \lambda_0 [w(\hat{h}(t_i; \theta_0) \mid \theta_0) - 1]}
\]

\[
= \frac{1}{n} \sum_{j=1}^{n} \left[ \frac{1}{b_0} + a_0 \hat{h}(y_j; \theta_0) - b_0 \hat{h}^2(y_j; \theta_0) \right]
\]

\[
- \frac{1}{n} \sum_{i=1}^{n} \left\{ 1 - \frac{1 - \lambda_0}{1 + \lambda_0 [w(\hat{h}(t_i; \theta_0) \mid \theta_0) - 1]} \right\} \left[ \frac{1}{b_0} + a_0 \hat{h}(t_i; \theta_0) - b_0 \hat{h}^2(t_i; \theta_0) \right]
\]

\[
= \frac{1}{n} \sum_{j=1}^{n} \left( 1 - \lambda_0 \right) \left[ \frac{1}{b_0} + a_0 \hat{h}(y_j; \theta_0) - b_0 \hat{h}^2(y_j; \theta_0) \right]
\]

\[
\left\{ \frac{1}{b_0} + a_0 \hat{h}(y_j; \theta_0) - b_0 \hat{h}^2(y_j; \theta_0) \right\}
\]

\[
= \frac{1}{n} \sum_{j=1}^{n} \left( 1 - \lambda_0 \right) \left[ \frac{1}{b_0} + a_0 \hat{h}(y_j; \theta_0) - b_0 \hat{h}^2(y_j; \theta_0) \right]
\]

\[
\left\{ \frac{1}{b_0} + a_0 \hat{h}(y_j; \theta_0) - b_0 \hat{h}^2(y_j; \theta_0) \right\}
\]

87
Lemma 3.7.3. As \( n \to \infty \),
\[
\frac{1}{n} \sum_{k=1}^{n} \left\{ 1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h(x_k; \theta_0); \theta_0) - 1]} \right\} \left[ \frac{1}{b_0} + a_0 \hat{h}(x_k; \theta_0) - b_0 \hat{h}^2(x_k; \theta_0) \right] \xrightarrow{p} 0.
\]

In a summary, as \( n \to \infty \)
\[
\frac{1}{n} V_n(\theta_0) \xrightarrow{p} 0.
\]

Let \( \hat{\theta} \) denote the unique solution of \( V_n(\theta) = 0 \). Thus \( \hat{\theta} \) is a consistent estimator of \( \theta_0 \). In particular, the true value of \( \lambda \) is \( \frac{\rho}{1+\rho} \).

Define \( \hat{h}_0(x; \theta) = \lim_{n \to \infty} \hat{h}(x; \theta) \). According to Lemma 3.7.1, \( \hat{h}_0(x; \theta_0) = h_0(x) \). We have the limiting form of \( \frac{V_n(\theta)}{n} \) as in Lemma 3.7.3.

Lemma 3.7.3. As \( n \to \infty \),
\[
\frac{V_n(\theta)}{n} \xrightarrow{p} \int \begin{pmatrix} g_1(x; \theta) \\ g_2(x; \theta) \\ g_3(x; \theta) \end{pmatrix} \, dG(x) = \begin{pmatrix} V_1(\theta) \\ V_2(\theta) \\ V_3(\theta) \end{pmatrix}
\]

where
\[
g_1(x; \theta) = \frac{1}{\lambda} \left\{ \frac{1 + \lambda_0[w(h_0(x); \theta_0) - 1]}{1 + \lambda[w(h_0(x); \theta_0) - 1]} - 1 \right\},
\]
\[
g_2(x; \theta) = (1 - \lambda) \left\{ \frac{1 + \lambda_0[w(h_0(x); \theta_0) - 1]}{1 + \lambda[w(h_0(x); \theta_0) - 1]} - \frac{1 - \lambda_0}{1 - \lambda} \right\} \left[ \hat{h}_0(x; \theta) - a \right],
\]
\[
g_3(x; \theta) = (1 - \lambda) \left\{ \frac{1 + \lambda_0[w(h_0(x); \theta_0) - 1]}{1 + \lambda[w(h_0(x); \theta_0) - 1]} - \frac{1 - \lambda_0}{1 - \lambda} \right\} \left[ \frac{1}{b} + \hat{h}_0(x; \theta) - b \hat{h}_0(x; \theta)^2 \right].
\]

88
Proof. As \( n \to \infty \)

\[
\frac{V_{n1}(\theta)}{n} = -\frac{1}{n} \sum_{i=1}^{n} \frac{w(\hat{h}(t_i; \theta); \theta) - 1}{1 + \lambda [w(\hat{h}(t_i; \theta); \theta) - 1]}
\]

\[
= -\frac{1}{n} \sum_{k=1}^{n^D} \frac{w(\hat{h}(x_k; \theta); \theta) - 1}{1 + \lambda [w(\hat{h}(x_k; \theta); \theta) - 1]}
- \frac{1}{n} \sum_{j=1}^{n^D} \frac{w(\hat{h}(y_j; \theta); \theta) - 1}{1 + \lambda [w(\hat{h}(y_j; \theta); \theta) - 1]}
\]

\[
\rightarrow -\frac{1}{1 + \rho} \int \frac{w(\hat{h}_0(x; \theta); \theta) - 1}{1 + \lambda [w(\hat{h}_0(x; \theta); \theta) - 1]} dG(x)
- \frac{\rho}{1 + \rho} \int \frac{w(\hat{h}_0(x; \theta); \theta) - 1}{1 + \lambda [w(\hat{h}_0(x; \theta); \theta) - 1]} dF(x)
\]

\[
= -(1 - \lambda_0) \int \frac{w(\hat{h}_0(x; \theta); \theta) - 1}{1 + \lambda [w(\hat{h}_0(x; \theta); \theta) - 1]} dG(x)
- \lambda_0 \int \frac{w(\hat{h}_0(x; \theta); \theta) - 1}{1 + \lambda [w(\hat{h}_0(x; \theta); \theta) - 1]} w(h_0(x; \theta_0)) dG(x)
\]

\[
= -\int 1 + \lambda_0 [w(h_0(x; \theta_0) - 1] \frac{w(\hat{h}_0(x; \theta); \theta) - 1]}{1 + \lambda [w(\hat{h}_0(x; \theta); \theta) - 1]} dG(x)
\]

\[
= -\frac{1}{\lambda} \int \left\{ 1 + \lambda_0 [w(h_0(x; \theta_0) - 1] \right\} \left\{ 1 - \frac{1}{1 + \lambda [w(h_0(x; \theta); \theta) - 1]} \right\} dG(x)
\]

\[
= -\frac{1}{\lambda} \left\{ 1 - \int \frac{1 + \lambda_0 [w(h_0(x; \theta_0) - 1]}{1 + \lambda [w(h_0(x; \theta); \theta) - 1]} dG(x) \right\}
\]

\[
= \int \frac{1}{\lambda} \left\{ 1 + \lambda_0 [w(h_0(x; \theta_0) - 1] \right\} dG(x) = \int g_1(x; \theta) dG(x) = V_1(\theta),
\]

where

\[
g_1(x; \theta) = \frac{1}{\lambda} \left\{ 1 + \lambda_0 [w(h_0(x; \theta_0) - 1] \right\} \frac{1}{1 + \lambda [w(h_0(x; \theta); \theta) - 1] - 1}
\]

Second, as \( n \to \infty \)

\[
\frac{V_{n2}(\theta)}{n} = \frac{1}{n} \sum_{j=1}^{n^D} [b\hat{h}(y_j; \theta) - a] - \frac{1}{n} \sum_{i=1}^{n} \lambda w(\hat{h}(t_i; \theta); \theta)[b\hat{h}(t_i; \theta) - a]
\]

\[
= \frac{1}{n} \sum_{j=1}^{n^D} \frac{(1 - \lambda)[b\hat{h}(y_j; \theta) - a]}{1 + \lambda [w(\hat{h}(y_j; \theta); \theta) - 1]}
- \frac{1}{n} \sum_{k=1}^{n^D} \left\{ 1 - \frac{1 - \lambda}{1 + \lambda [w(\hat{h}(x_k; \theta); \theta) - 1]} \right\} [b\hat{h}(x_k; \theta) - a]
\]

89
\[ \rightarrow \begin{align*}
\lambda_0 & \int \frac{(1 - \lambda)[\hat{h}_0(x; \theta) - a]}{1 + \lambda[w(\hat{h}_0(x; \theta); \theta) - 1]} \ dF(x) \\
& - (1 - \lambda_0) \int \left\{ 1 - \frac{1 - \lambda}{1 + \lambda[w(\hat{h}_0(x; \theta); \theta) - 1]} \right\} [\hat{h}_0(x; \theta) - a] \ dG(x) \\
& = \int \frac{(1 - \lambda)[\hat{h}_0(x; \theta) - a]}{1 + \lambda[w(\hat{h}_0(x; \theta); \theta) - 1]} \left\{ 1 + \lambda_0[w(\hat{h}_0(x; \theta); \theta_0) - 1] - (1 - \lambda_0) \right\} \ dG(x) \\
& - (1 - \lambda_0) \int \left\{ 1 - \frac{1 - \lambda}{1 + \lambda[w(\hat{h}_0(x; \theta); \theta) - 1]} \right\} [\hat{h}_0(x; \theta) - a] \ dG(x) \\
& = \int (1 - \lambda) \left\{ \frac{1 + \lambda_0[w(\hat{h}_0(x; \theta); \theta_0) - 1]}{1 + \lambda[w(\hat{h}_0(x; \theta); \theta) - 1]} - \frac{1 - \lambda_0}{1 - \lambda} \right\} [\hat{h}_0(x; \theta) - a] \ dG(x) \\
& = \int g_2(x; \theta) \ dG(x) = V_2(\theta),
\end{align*} \]

where

\[ g_2(x; \theta) = (1 - \lambda) \left\{ \frac{1 + \lambda_0[w(\hat{h}_0(x; \theta); \theta_0) - 1]}{1 + \lambda[w(\hat{h}_0(x; \theta); \theta) - 1]} - \frac{1 - \lambda_0}{1 - \lambda} \right\} [\hat{h}_0(x; \theta) - a]. \]

And third, as \( n \to \infty \)

\[ \frac{V_{n3}(\theta)}{n} = \frac{1}{n} \sum_{j=1}^{n} \left[ \frac{1}{b} + a\hat{h}(y_j; \theta) - b\hat{h}(y_j; \theta)^2 \right] \\
- \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda w(\hat{h}(t_i; \theta); \theta)[\frac{1}{b} + a\hat{h}(t_i; \theta) - b\hat{h}(t_i; \theta)^2]}{1 + \lambda[w(\hat{h}(t_i; \theta); \theta) - 1]} \\
= \frac{1}{n} \sum_{j=1}^{n} \frac{(1 - \lambda)[\frac{1}{b} + a\hat{h}(y_j; \theta) - b\hat{h}(y_j; \theta)^2]}{1 + \lambda[w(\hat{h}(y_j; \theta); \theta) - 1]} \\
- \frac{1}{n} \sum_{k=1}^{n} \left\{ 1 - \frac{1 - \lambda}{1 + \lambda[w(\hat{h}(x_k; \theta); \theta) - 1]} \right\} \left[ \frac{1}{b} + a\hat{h}(x_k; \theta) - b\hat{h}(x_k; \theta)^2 \right] \\
\rightarrow \begin{align*}
\lambda_0 & \int \frac{(1 - \lambda)[\frac{1}{b} + a\hat{h}_0(x; \theta) - b\hat{h}_0(x; \theta)^2]}{1 + \lambda[w(\hat{h}_0(x; \theta); \theta) - 1]} \ dF(x) \\
& - (1 - \lambda_0) \int \left\{ 1 - \frac{1 - \lambda}{1 + \lambda[w(\hat{h}_0(x; \theta); \theta) - 1]} \right\} \left[ \frac{1}{b} + a\hat{h}_0(x; \theta) - b\hat{h}_0(x; \theta)^2 \right] \ dG(x) \\
& = \int \frac{(1 - \lambda)[\frac{1}{b} + a\hat{h}_0(x; \theta) - b\hat{h}_0(x; \theta)^2]}{1 + \lambda[w(\hat{h}_0(x; \theta); \theta) - 1]} \left\{ 1 + \lambda_0[w(\hat{h}_0(x; \theta); \theta_0) - 1] - (1 - \lambda_0) \right\} \ dG(x)
\end{align*} \]
\[-(1 - \lambda_0) \int \left\{ 1 - \frac{1 - \lambda}{1 + \lambda[w(h_0(x); \theta)]} \right\} \frac{1}{b} \left[ b + a\hat{h_0}(x; \theta) - b\hat{h_0}(x; \theta)^2 \right] dG(x) \]

\[= \int (1 - \lambda) \left\{ \frac{1 + \lambda_0[w(h_0(x); \theta_0)] - 1}{1 + \lambda[w(h_0(x); \theta)] - 1} \right\} \frac{1}{b} \left[ b + a\hat{h_0}(x; \theta) - b\hat{h_0}(x; \theta)^2 \right] dG(x) \]

\[= \int g_3(x; \theta) dG(x) = V_3(\theta), \]

where

\[g_3(x; \theta) = (1 - \lambda) \left\{ \frac{1 + \lambda_0[w(h_0(x); \theta_0)] - 1}{1 + \lambda[w(h_0(x); \theta)] - 1} \right\} \frac{1}{b} \left[ b + a\hat{h_0}(x; \theta) - b\hat{h_0}(x; \theta)^2 \right]. \]

Note that, by Lemma 3.7.1, \( \hat{h_0}(x; \theta_0) = h_0(x) \), we have

\[g_1(x; \theta_0) = g_2(x; \theta_0) = g_3(x; \theta_0) = 0, \text{ and } V_1(\theta_0) = V_2(\theta_0) = V_3(\theta_0) = 0. \]

Define

\[A_k = \int \frac{h_0(x)^k}{1 + \lambda_0[w(h_0(x); a_0, b_0)]} dG(x), \quad k = 0, \ldots, 4, \]

\[C_k = \int \frac{h_0(x)^k}{(1 + \lambda_0[w(h_0(x); a_0, b_0)])^2} dG(x), \quad k = 0, \ldots, 4. \]

**Lemma 3.7.4.** Write

\[A = -\lim_{n \to \infty} \frac{1}{n} \frac{\partial V_n(\lambda, a, b)}{\partial (\lambda, a, b)} \big|_{(\lambda_0, a_0, b_0)}. \]

Then

\[A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \]
Proof. Let $A_{-k} = \int h_0(x)^k dG(x)$, $k = 1, \ldots, 4$, denote the first four moments of $h_0(X)$ which has standard normal distribution. Thus,

$$A_{-1} = 0, \quad A_{-2} = 1, \quad A_{-3} = 0, \quad A_{-4} = 3.$$
Write

\[ U(\theta) = \frac{1 + \lambda_0[w(h_0(x); \theta_0) - 1]}{1 + \lambda[w(\hat{h}_0(x); \theta) - 1]}, \]

and

\[ u(\theta) = \log w(\hat{h}(x; \theta); \theta) = \log b - \frac{a^2}{2} + ab\hat{h}(x; \theta) + \frac{1 - b^2}{2} \hat{h}(x; \theta)^2. \]

According to Lemma 3.7.3 and the definition of \( U(\theta) \), write

\[ g_1(x; \theta) = \frac{1}{\lambda} \{ U(\theta) - 1 \}, \]
\[ g_2(x; \theta) = (1 - \lambda)[b\hat{h}_0(x; \theta) - a] \{ U(\theta) - \frac{1 - \lambda_0}{1 - \lambda} \}, \] (3.9)
\[ g_3(x; \theta) = (1 - \lambda)[\frac{1}{b} + a\hat{h}_0(x; \theta) - b\hat{h}_0(x; \theta)] \{ U(\theta) - \frac{1 - \lambda_0}{1 - \lambda} \}. \]

Since \( \hat{h}_0(x; \theta_0) = h_0(x) \), we have \( U(\theta_0) = 1 \), which gives \( g_1(x; \theta_0) = g_2(x; \theta_0) = g_3(x; \theta_0) = 0 \), the same result as in Lemma 3.7.2. When we calculate \( \frac{\partial u(x; \theta)}{\partial \theta} |_{\theta_0} \), \( i = 1, 2, 3 \), only the derivatives onto the part within the parenthesis in (3.9) remain. Next, we derive \( \frac{\partial u(\theta)}{\partial \lambda} |_{\theta_0} \) where \( \frac{\partial \hat{h}(x,a,b)}{\partial (a,b)} \neq 0 \) (Refer to Remark 3.7.1). First we have, as \( n \to \infty \)

\[ \frac{\partial u(\theta)}{\partial \lambda} |_{\theta_0} \xrightarrow{p} 0, \]

and

\[ \frac{\partial u(\theta)}{\partial a} |_{\theta_0} = \left\{ -a + b\hat{h}(x; \theta) + ab\frac{\partial \hat{h}(x; \theta)}{\partial a} + (1 - b^2)\hat{h}(x; \theta)\frac{\partial \hat{h}(x; \theta)}{\partial a} \right\} |_{\theta_0} \]
\[ = \left\{ -a + b\hat{h}(x; \theta) + [ab + (1 - b^2)\hat{h}(x; \theta)]\frac{\partial \hat{h}(x; \theta)}{\partial a} \right\} |_{\theta_0} \]
\[ \xrightarrow{p} -a_0 + b_0 \cdot h_0(x) + [a_0 b_0 + (1 - b_0^2) h_0(x)] \]
\[ \frac{1 - \lambda_0}{b_0} \left( \frac{1}{1 - \lambda_0} - \frac{1}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right). \]

93
\[ \begin{align*}
&= -a_0 + b_0 h_0(x) + a_0 + \frac{1 - b_0^2}{b_0} h_0(x) - \frac{(1 - \lambda_0)[a_0 + \frac{1 - b_0^2}{b_0} h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \\
&= \frac{h_0(x)}{b_0} - \frac{(1 - \lambda_0)[a_0 + \frac{1 - b_0^2}{b_0} h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} = \frac{h_0(x)}{b_0} - \frac{[a_0 + \frac{1 - b_0^2}{b_0} h_0(x)]}{1 + \rho w(h_0(x); \theta_0)},
\end{align*} \]

and

\[ \frac{\partial u(\hat{h}(x; \theta); \theta)}{\partial b} \bigg|_{\theta_0} = \left\{ \frac{1}{b} + a \hat{h}(x; \theta) - b \hat{h}(x; \theta)^2 + [ab + (1 - b^2)\hat{h}(x; \theta)] \frac{\partial \hat{h}(x; \theta)}{\partial b} \right\} \bigg|_{\theta_0} \]

\[ \xrightarrow{b} \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0(x)^2 - \frac{h_0(x)}{b_0} \left( 1 - \frac{1}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) b_0 a_0 + (1 - b_0^2) h_0(x) \]

\[ = 1 - \frac{h_0(x)^2}{b_0} + \frac{(1 - \lambda_0) h_0(x) \left[ a_0 + \frac{1 - b_0^2}{b_0} h_0(x) \right]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \]

\[ = \frac{1 - h_0(x)^2}{b_0} + \frac{h_0(x) \left[ a_0 + \frac{1 - b_0^2}{b_0} h_0(x) \right]}{1 + \rho w(h_0(x); \theta_0)}, \]

for each fixed \( x \). Second we have

\[ \frac{\partial U(\theta)}{\partial \lambda} \bigg|_{\theta_0} = -\frac{1}{\lambda_0} \left( 1 - \frac{1}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right), \]

and

\[ \frac{\partial U(\theta)}{\partial a} \bigg|_{\theta_0} = -\left( 1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) \lim_{n \to \infty} \frac{\partial u(\theta)}{\partial a} \bigg|_{\theta_0} \]

\[ = -\left( 1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) \left( \frac{h_0(x)}{b_0} - \frac{(1 - \lambda_0) [a_0 + \frac{1 - b_0^2}{b_0} h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \right) \]

\[ = -\frac{h_0(x)}{b_0} + \frac{(1 - \lambda_0) \left[ a_0 + \frac{1 - b_0^2}{b_0} h_0(x) \right]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \left( 1 - \lambda_0 \right)^2 \left( a_0 + \frac{1 - b_0^2}{b_0} h_0(x) \right) \]

\[ - \frac{(1 - \lambda_0)^2 \left( a_0 + \frac{1 - b_0^2}{b_0} h_0(x) \right)}{(1 + \lambda_0[w(h_0(x); \theta_0) - 1])^2}. \]
and

\[
\frac{\partial U(\theta)}{\partial b}|_{\theta_0} = - \left(1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]}\right) \lim_{n \to \infty} \frac{\partial u(\theta)}{\partial b}|_{\theta_0} \\
= - \left(1 - \frac{1 - \lambda_0}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]}\right) \\
\cdot \left(\frac{1 - h_0(x)^2}{b_0} + \frac{(1 - \lambda_0)h_0(x)[a_0 + \frac{1 - b_0^2}{b_0} h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]}\right) \\
= h_0(x)^2 - 1 \frac{(1 - \lambda_0)\left(-h_0(x)[a_0 + \frac{1 - b_0^2}{b_0} h_0(x)] + \frac{1 - h_0(x)^2}{b_0}\right)} {b_0} \\
+ \frac{(1 - \lambda_0)^2 h_0(x)[a_0 + \frac{1 - b_0^2}{b_0} h_0(x)]}{\{1 + \lambda_0[w(h_0(x); \theta_0) - 1]\}^2} \\
= h_0(x)^2 - 1 \frac{(1 - \lambda_0)\left(-h_0(x)[a_0 + \frac{2 - b_0^2}{b_0} h_0(x)] + \frac{1}{b_0}\right)} {b_0} \\
+ \frac{(1 - \lambda_0)^2 h_0(x)[a_0 + \frac{1 - b_0^2}{b_0} h_0(x)]}{\{1 + \lambda_0[w(h_0(x); \theta_0) - 1]\}^2}.
\]

For the calculation of \(- \lim_{n \to \infty} \frac{1}{n} \frac{\partial V_n(\theta)}{\partial \theta}|_{\theta_0}\), notice that:

\[- \lim_{n \to \infty} \frac{1}{n} \frac{\partial V_n(\theta)}{\partial \theta}|_{\theta_0} = - \frac{\partial}{\partial \theta} \left( \lim_{n \to \infty} \frac{V_n(\theta)}{n}\right)|_{\theta_0} = - \frac{\partial V_1(\theta)}{\partial \theta}|_{\theta_0} = - \int \frac{\partial g_1(x; \theta)}{\partial \theta}|_{\theta_0} dG(x).
\]

Since

\[
\frac{\partial g_1(x; \theta)}{\partial \theta}|_{\theta_0} = \frac{1}{\lambda} \left( \begin{array}{c} \frac{\partial U(\theta)}{\partial \lambda} \\ \frac{\partial U(\theta)}{\partial a} \\ \frac{\partial U(\theta)}{\partial b} \end{array} \right)^\top |_{\theta_0},
\]

we have,

\[
a_{11} = - \int \frac{\partial g_1(x; \theta)}{\partial \lambda}|_{\theta_0} dG(x) = - \int \frac{1}{\lambda} \frac{\partial U(\theta)}{\partial \lambda}|_{\theta_0} dG(x).
\]
\[ a_{12} = -\int \frac{\partial g_1(x; \theta)}{\partial a} |_{\theta_0} dG(x) = -\int \frac{\partial U(\theta)}{\lambda} \frac{\partial a}{\partial a} |_{\theta_0} dG(x) \]
\[ = -\int \frac{1}{\lambda_0} \left\{ -\frac{h_0(x)}{b_0} + \frac{(1 - \lambda_0) \left( a_0 + \frac{2 - b_0^2}{b_0} h_0(x) \right)}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right\} dG(x) \]
\[ = -\frac{1 - \lambda_0}{\lambda_0} \left( a_0 A_0 + \frac{2 - b_0^2}{b_0} A_1 \right) + \frac{(1 - \lambda_0)^2}{\lambda_0} \left( a_0 C_0 + \frac{1 - b_0^2}{b_0} C_1 \right), \]

and

\[ a_{13} = -\int \frac{\partial g_1(x; \theta)}{\partial b} |_{\theta_0} dG(x) = -\int \frac{\partial U(\theta)}{\lambda} \frac{\partial b}{\partial b} |_{\theta_0} dG(x) \]
\[ = -\int \frac{1}{\lambda_0} \left\{ \frac{h_0(x)^2 - 1}{b_0} + \frac{(1 - \lambda_0) \left( h_0(x) [a_0 + \frac{2 - b_0^2}{b_0} h_0(x)] + \frac{1}{b_0} \right)}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right\} dG(x) \]
\[ + \frac{(1 - \lambda_0)^2 h_0(x) [a_0 + \frac{1 - b_0^2}{b_0} h_0(x)]}{(1 + \lambda_0 [w(h_0(x); \theta_0) - 1])^2} \right\} dG(x) \]
\[ = \frac{1 - A_{-2}}{b_0 \lambda_0} - \frac{1 - \lambda_0}{\lambda_0} \left( A_0 - a_0 A_1 - \frac{2 - b_0^2}{b_0} A_2 \right) - \frac{(1 - \lambda_0)^2}{\lambda_0} \left( a_0 C_1 + \frac{1 - b_0^2}{b_0} C_2 \right) \]
\[ = -\frac{1 - \lambda_0}{\lambda_0} \left( A_0 - a_0 A_1 - \frac{2 - b_0^2}{b_0} A_2 \right) - \frac{(1 - \lambda_0)^2}{\lambda_0} \left( a_0 C_1 + \frac{1 - b_0^2}{b_0} C_2 \right). \]

Similarly,

\[ \frac{\partial g_2(x; \theta)}{\partial \theta} |_{\theta_0} = (1 - \lambda) \left[ b h_0(x; \theta) - a \right] \begin{pmatrix} \frac{\partial U(\theta)}{\partial \lambda} - \frac{1 - \lambda_0}{(1 - \lambda)^2} \\ \frac{\partial U(\theta)}{\partial a} \\ \frac{\partial U(\theta)}{\partial b} \end{pmatrix} |_{\theta_0}, \]

96
we have

\[
a_{21} = - \int \frac{\partial g_2(x; \theta)}{\partial \lambda} \bigg|_{\theta_0} dG(x) = \int (\lambda - 1) \left[ b h_0(x; \theta) - a \right] \left( \frac{\partial U(\theta)}{\partial \lambda} - \frac{1 - \lambda_0}{(1 - \lambda)^2} \right) \bigg|_{\theta_0} dG(x)
\]

\[
= \int (1 - \lambda_0) \left[ b h_0(x) - a \right] \left[ \frac{1}{\lambda_0} \left( \frac{1}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right) + \frac{1}{1 - \lambda_0} \right] dG(x)
\]

\[
= - \frac{a_0}{\lambda_0} - \frac{1 - \lambda_0}{\lambda_0} (b_0 A_1 - a_0 A_0),
\]

and

\[
a_{22} = - \int \frac{\partial g_2(x; \theta)}{\partial a} \bigg|_{\theta_0} dG(x) = - \int (1 - \lambda) \left[ b h_0(x; \theta) - a \right] \left( \frac{\partial U(\theta)}{\partial a} \right) \bigg|_{\theta_0} dG(x)
\]

\[
= - \int (1 - \lambda_0) \left[ b h_0(x) - a \right] \left[ - \frac{h_0(x)}{b_0} + \frac{(1 - \lambda_0) \left( a_0 + \frac{2b_0^2 - b_0^2 h_0(x)}{b_0} \right)}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right]
\]

\[
\times \left[ \frac{1}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right] dG(x)
\]

\[
= (1 - \lambda_0) A_{-2} - (1 - \lambda_0)^2 \left[ -a_0^2 A_0 + a_0 (b_0 - \frac{2b_0^2 - b_0^2}{b_0}) A_1 + (2 - b_0^2) A_2 \right]
\]

\[
+ (1 - \lambda_0)^3 \left[ -a_0^2 C_0 + a_0 (b_0 - \frac{1 - b_0^2}{b_0}) C_1 + (1 - b_0^2) C_2 \right],
\]

and

\[
a_{23} = - \int \frac{\partial g_2(x; \theta)}{\partial b} \bigg|_{\theta_0} dG(x) = - \int (1 - \lambda) \left[ b h_0(x; \theta) - a \right] \left( \frac{\partial U(\theta)}{\partial b} \right) \bigg|_{\theta_0} dG(x)
\]

\[
= - \int (1 - \lambda_0) \left[ b h_0(x) - a \right] \left[ \frac{h_0(x)^2 - 1}{b_0} \right]
\]

\[
+ \frac{(1 - \lambda_0) \left[ -h_0(x) \left( a_0 + \frac{2b_0^2 - b_0^2 h_0(x)}{b_0} \right) + \frac{1}{b_0} \right]}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} dG(x)
\]

97
\[\frac{(1 - \lambda_0)^2 h_0(x) [a_0 + \frac{1 - b_0^2}{b_0} h_0(x)]}{\{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]\}^2} dG(x) = -(1 - \lambda_0) [A_{-3} - \frac{a_0}{b_0} A_{-2} - A_{-1} + \frac{a_0}{b_0}] - (1 - \lambda_0)^2 \left[ -\frac{a_0}{b_0} A_0 + (1 + a_0^2) A_1 + 2a_0\left( \frac{1}{b_0} - b_0 \right) A_2 - (2 - b_0^2) A_3 \right] - (1 - \lambda_0)^3 \left[ -a_0^2 C_1 + a_0 (2b_0 - \frac{1}{b_0}) C_2 + (1 - b_0^2) C_3 \right] = -(1 - \lambda_0)^2 \left[ -\frac{a_0}{b_0} A_0 + (1 + a_0^2) A_1 + 2a_0\left( \frac{1}{b_0} - b_0 \right) A_2 - (2 - b_0^2) A_3 \right] - (1 - \lambda_0)^3 \left[ -a_0^2 C_1 + a_0 (2b_0 - \frac{1}{b_0}) C_2 + (1 - b_0^2) C_3 \right].\]

Similarly,

\[
\frac{\partial g_3(x; \theta)}{\partial \theta} \big|_{\theta_0} = (1 - \lambda) \left[ \frac{1}{b} + a \tilde{h}_0(x; \theta) - b \tilde{h}_0(x; \theta)^2 \right] \left( \frac{\partial U(\theta)}{\partial \lambda} - \frac{1 - \lambda_0}{(1 - \lambda)^2} \right) \big|_{\theta_0},
\]

we have,

\[
a_{31} = - \int \frac{\partial g_3(x; \theta)}{\partial \lambda} \big|_{\theta_0} dG(x) = - \int (1 - \lambda) \left[ \frac{1}{b} + a \tilde{h}_0(x; \theta) - b \tilde{h}_0(x; \theta)^2 \right] \left( \frac{\partial U(\theta)}{\partial \lambda} - \frac{1 - \lambda_0}{(1 - \lambda)^2} \right) \big|_{\theta_0} dG(x) = - \int (1 - \lambda_0) \left[ \frac{1}{b_0} + a_0 \tilde{h}_0(x) - b_0 \tilde{h}_0^2(x) \right] \left[ -\frac{1}{\lambda_0} \left( 1 - \frac{1}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right) - \frac{1}{1 - \lambda_0} \right] dG(x) = \frac{1}{\lambda_0} \left( \frac{1}{b_0} + a_0 A_{-1} - b_0 A_{-2} \right) - \frac{1 - \lambda_0}{\lambda_0} \left( \frac{A_0}{b_0} + a_0 A_1 - b_0 A_2 \right) = \frac{1}{\lambda_0} \left( \frac{1}{b_0} - b_0 \right) - \frac{1 - \lambda_0}{\lambda_0} \left( \frac{A_0}{b_0} + a_0 A_1 - b_0 A_2 \right),
\]

98
and

\[ a_{32} = - \int \frac{\partial g_3(x; \theta)}{\partial a} |_{\theta_0} dG(x) \]
\[ = - \int (1 - \lambda) \left[ \frac{1}{b} + a h_0(x; \theta) - b \hat{h}_0(x; \theta)^2 \right] \left( \frac{\partial U(\theta)}{\partial a} \right) |_{\theta_0} dG(x) \]
\[ = - \int (1 - \lambda_0) \left[ \frac{1}{b_0} + a_0 h_0(x) - b_0 \hat{h}_0^2(x) \right] \left[ \frac{h_0(x)^2 - 1}{b_0} + \frac{(1 - \lambda_0) \left( -h_0(x) [a_0 + \frac{2 - b_0^2}{b_0} h_0(x)] + \frac{1}{b_0} \right)}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right] dG(x) \]
\[ = - (1 - \lambda_0) \left[ -A_{-4} + \frac{a_0}{b_0} A_{-3} + \frac{1}{b_0^2} + 1 \right] A_{-2} - \frac{a_0}{b_0} A_{-1} - \frac{1}{b_0^2} \]

and

\[ a_{33} = - \int \frac{\partial g_3(x; \theta)}{\partial b} |_{\theta_0} dG(x) \]
\[ = - \int (1 - \lambda) \left[ \frac{1}{b} + a h_0(x; \theta) - b \hat{h}_0(x; \theta)^2 \right] \left( \frac{\partial U(\theta)}{\partial b} \right) |_{\theta_0} dG(x) \]
\[ = - \int (1 - \lambda_0) \left[ \frac{1}{b_0} + a_0 h_0(x) - b_0 \hat{h}_0^2(x) \right] \left[ \frac{h_0(x)^2 - 1}{b_0} + \frac{(1 - \lambda_0) \left( -h_0(x) [a_0 + \frac{2 - b_0^2}{b_0} h_0(x)] + \frac{1}{b_0} \right)}{1 + \lambda_0 [w(h_0(x); \theta_0) - 1]} \right] dG(x) \]
\[ = - (1 - \lambda_0) \left[ -A_{-4} + \frac{a_0}{b_0} A_{-3} + \frac{1}{b_0^2} + 1 \right] A_{-2} - \frac{a_0}{b_0} A_{-1} - \frac{1}{b_0^2} \]
The first order Taylor expansion of $V$ yields

\[
(1 - \lambda_0)^2 \left[ \frac{A_0}{b_0^2} - \frac{2}{b_0^2} + a_0^2 \right] A_2 + 2a_0(b_0 - \frac{1}{b_0}) A_3 + (2 - b_0^2) A_4 \\
-(1 - \lambda_0)^3 \left[ \frac{a_0}{b_0} C_1 + (a_0^2 + 1) C_2 + a_0(\frac{1}{b_0} - 2b_0) C_3 - (1 - b_0^2) C_4 \right] \\
= 2(1 - \lambda_0)
\]

Since $a_{12} \neq a_{21}, a_{13} \neq a_{31}$ and $a_{23} \neq a_{32}$, $A$ is not a symmetric matrix.

\[\square\]

**Theorem 3.2.1.**

**Proof.** The first order Taylor expansion of $V_n(\hat{\theta})$ around $\theta_0$ yields

\[
\sqrt{n}(\hat{\theta} - \theta_0) = n^{-\frac{1}{2}} A^{-1} V_n(\theta_0) + o_p(1), \tag{3.10}
\]

where $A = - \lim_{n \to \infty} \frac{1}{n} \frac{\partial^2 V_n(\theta)}{\partial \theta \partial \theta} |_{\theta_0}$ as derived in Lemma 3.7.4. We have

\[
\frac{V_n(\theta_0)}{n} = \frac{V_n(\theta_0)|_{h(x; \theta_0) = h_0(x)}}{n} + \lim_{n \to \infty} \frac{V_n(\theta_0) - n V_1(\theta_0)}{n} + o_p(n^{-\frac{1}{2}})
\]

\[
= \frac{V_n(\theta_0)|_{h(x; \theta_0) = h_0(x)}}{n} + V_1(\theta_0) + o_p(n^{-\frac{1}{2}}).
\]

Here we take $V_1(\theta_0)$ as function of $h(x; \theta_0)$ and do Taylor expansion around $h_0(x)$ yields

\[
V_1(\theta_0) = V_1(\theta_0)|_{h(x; \theta_0) = h_0(x)} + \int \frac{\partial g_1(x; \theta_0)}{\partial h(x; \theta_0)}|_{h(x; \theta_0) = h_0(x)} [\hat{h}(x; \theta_0) - h_0(x)] dG(x) + o_p(n^{-\frac{1}{2}})
\]

\[
= \int \frac{\partial g_1(x; \theta_0)}{\partial h(x; \theta_0)}|_{h(x; \theta_0) = h_0(x)} [\hat{h}(x; \theta_0) - h_0(x)] dG(x) + o_p(n^{-\frac{1}{2}})
\]
\[
\begin{align*}
\hat{H}(x) = H(h(x; \theta_0); \theta_0) & = H(h_0(x); \theta_0) + \frac{\partial H(h(x); \theta_0)}{\partial h(x)}|_{h_0(x)}(\hat{h}(x; \theta_0) - h_0(x)) + o_p(n^{-\frac{1}{2}}), \\
\hat{h}(x; \theta_0) - h_0(x) & = \left( \frac{\partial H(h(x); \theta_0)}{\partial h(x)}|_{h_0(x)} \right)^{-1} [\hat{H}(x) - H(h_0(x); \theta_0)] + o_p(n^{-\frac{1}{2}}) \\
& = \frac{1}{n} \sum_{i=1}^{n} I(t_i < x) - H(h_0(x); \theta_0) + o_p(n^{-\frac{1}{2}}) \\
& = \frac{1}{n} \sum_{i=1}^{n} e_{t_i}(x) + o_p(n^{-\frac{1}{2}}),
\end{align*}
\]

where \( e_y(x) = I(y \leq x) - H(h_0(x); \theta_0) \). Plugging \( \hat{h}(x; \theta_0) - h_0(x) \) into \( V_1(\theta_0) \) yields

\[
\begin{align*}
V_1(\theta_0) & = - \int \frac{w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} [\hat{h}(x; \theta_0) - h_0(x)]dG(x) \\
& \quad + o_p(n^{-\frac{1}{2}}) \\
& = - \int \frac{w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \frac{1}{n} \sum_{i=1}^{n} e_{t_i}(x) dG(x) + o_p(n^{-\frac{1}{2}}) \\
& = \frac{1}{n} \sum_{i=1}^{n} \int g_{1a}(x)e_{t_i}(x)dG(x) + o_p(n^{-\frac{1}{2}}),
\end{align*}
\]

where

\[
g_{1a}(x) = \frac{-w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)]}{\{1 + \lambda_0[w(h_0(x); \theta_0) - 1]\}\{p_D\phi(h_0(x)) + p_Db_0\phi[b_0h_0(x) - a_0]\}}.
\]
Plugging $V_1(\theta_0)$ into $\frac{V_n(\theta_0)}{n}$, we have

$$\frac{V_n(\theta_0)}{n} = \frac{V_n(\theta_0)}{n} \bigg|_{\hat{g}(x;\theta_0) = h_0(x)} + V_1(\theta_0) + o_p(n^{-\frac{1}{2}})$$

$$= -\frac{1}{n} \sum_{i=1}^{n} \frac{w(h_0(t_i); \theta_0) - 1}{1 + \lambda_0[w(h_0(t_i); \theta_0) - 1]} + \frac{1}{n} \sum_{i=1}^{n} \int g_{1a}(x) e_{t_i}(x) dG(x) + o_p(n^{-\frac{1}{2}})$$

$$= \frac{1}{n} \sum_{j=1}^{n} v_1(y_j) + \frac{1}{n} \sum_{k=1}^{n} v_1(x_k) + o_p(n^{-\frac{1}{2}}),$$

where

$$v_1(y) = \int g_{1a}(x) e_y(x) dG(x) - \frac{w(h_0(y); \theta_0) - 1}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]}$$

$$= \int g_{1a}(x) e_y(x) dG(x) - \frac{1}{\lambda_0} \left( 1 - \frac{1}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]} \right).$$

Through similar procedure,

$$V_2(\theta_0) = V_2(\theta_0) \bigg|_{\hat{g}(x;\theta_0) = h_0(x)} + \int \frac{\partial g_2(x; \theta_0)}{\partial h(x; \theta_0)} \bigg|_{\hat{g}(x;\theta_0) = h_0(x)} [\hat{h}(x; \theta_0) - h_0(x)] dG(x) + o_p(n^{-\frac{1}{2}})$$

$$= \int \frac{\partial g_2(x; \theta_0)}{\partial h(x; \theta_0)} \bigg|_{\hat{g}(x;\theta_0) = h_0(x)} [\hat{h}(x; \theta_0) - h_0(x)] dG(x) + o_p(n^{-\frac{1}{2}})$$

$$= -\lambda_0(1 - \lambda_0) \int \frac{w(h_0(x); \theta_0)[a_0 b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} \cdot$$

$$[b_0 h_0(x) - a_0][\hat{h}(x; \theta_0) - h_0(x)] dG(x) + o_p(n^{-\frac{1}{2}})$$

$$= -\lambda_0(1 - \lambda_0) \int \frac{w(h_0(x); \theta_0)[a_0 b_0 + (1 - b_0^2)h_0(x)]}{1 + \lambda_0[w(h_0(x); \theta_0) - 1]} [b_0 h_0(x) - a_0] \cdot$$

$$\frac{1}{n} \sum_{i=1}^{n} e_{t_i}(x) dG(x) + o_p(n^{-\frac{1}{2}})$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int g_{2a}(x) e_{t_i}(x) dG(x) + o_p(n^{-\frac{1}{2}}),$$

102
where

\[ g_{2a}(x) = -\lambda_0(1 - \lambda_0)[b_0h_0(x) - a_0]w(h_0(x); \theta_0)[a_0b_0 + (1 - b_0^2)h_0(x)] \{1 + \lambda_0[w(h_0(x); \theta_0) - 1]\} \{p_D\phi(h_0(x)) + p_Db_0\phi[b_0h_0(x) - a_0]\} = \lambda_0(1 - \lambda_0)[b_0h_0(x) - a_0]g_{1a}(x). \]

We have

\[
\frac{V_{n2}(\theta_0)}{n} = \frac{V_{n2}(\theta_0)}{n} |_{h(x; \theta_0) = h_0(x)} + V_2(\theta_0) + o_p(n^{-\frac{1}{2}})
= \frac{1}{n} \sum_{j=1}^{n} [b_0h_0(y_j) - a_0] - \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda_0w(h_0(t_i); \theta_0)}{1 + \lambda_0[w(h_0(t_i); \theta_0) - 1]} \lambda_0w(h_0(t_i); \theta_0) [b_0h_0(t_i) - a_0]
+ \frac{1}{n} \sum_{i=1}^{n} \int g_{2a}(x)e_{t_i}(x)dG(x) + o_p(n^{-\frac{1}{2}})
= \frac{1}{n} \sum_{j=1}^{n} v_{21}(y_j) + \frac{1}{n} \sum_{i=1}^{n} v_{22}(t_i) + o_p(n^{-\frac{1}{2}})
= \frac{1}{n} \sum_{j=1}^{n} [v_{21}(y_j) + v_{22}(y_j)] + \frac{1}{n} \sum_{k=1}^{n} v_{22}(x_k) + o_p(n^{-\frac{1}{2}}),
\]

where

\[
v_{21}(y) = b_0h_0(y) - a_0, \quad v_{22}(y) = \int g_{2a}(x)e_y(x)dG(x) - \frac{\lambda_0w(h_0(y); \theta_0)}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]} [b_0h_0(y) - a_0]
= \int g_{2a}(x)e_y(x)dG(x) + \frac{(1 - \lambda_0) [b_0h_0(y) - a_0]}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]} - v_{21}(y).\]

Similarly,

\[
V_3(\theta_0) = V_3(\theta_0) |_{h(x; \theta_0) = h_0(x)} + \int \frac{\partial g_3(x; \theta_0)}{\partial h(x; \theta_0)} |_{h(x; \theta_0) = h_0(x)} [\hat{h}(x; \theta_0) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}})
= \int \frac{\partial g_3(x; \theta_0)}{\partial h(x; \theta_0)} |_{h(x; \theta_0) = h_0(x)} [\hat{h}(x; \theta_0) - h_0(x)]dG(x) + o_p(n^{-\frac{1}{2}})
\]

103
\[ g_{3a}(x) = \frac{-\lambda_0(1 - \lambda_0) \left[ \frac{1}{b_0} + a_0 h_0(x) - b_0 h_0^2(x) \right] w(h_0(x); \theta_0) w(h_0(x); \theta_0) - a_0^2 h_0(x)}{(1 + \lambda_0 w(h_0(x); \theta_0) - 1) \{ p_D \phi(h_0(x)) + p_D \phi b_0 h_0(x) - a_0 \}} \]

where

\[ V_{n3}(\theta_0) = \frac{V_{n3}(\theta_0)}{n} \bigg|_{h(x; \theta_0) = h_0(x)} + V_3(\theta_0) + o_p(n^{-\frac{1}{2}}) \]

\[ = \frac{1}{n} \sum_{j=1}^{n_{p}} \left[ \frac{1}{b_0} + a_0 h_0(y_j) - b_0 h_0^2(y_j) \right] \]

\[ - \frac{1}{n} \sum_{i=1}^{n} \frac{\lambda_0 w(h_0(t_i); \theta_0) \left[ \frac{1}{b_0} + a_0 h_0(t_i) - b_0 h_0^2(t_i) \right]}{1 + \lambda_0 w(h_0(t_i); \theta_0) - 1} \]

\[ + \frac{1}{n} \sum_{i=1}^{n} g_{3a}(x) e_{t_i}(x) dG(x) + o_p(n^{-\frac{1}{2}}) \]

\[ = \frac{1}{n} \sum_{j=1}^{n_{p}} v_{31}(y_j) + \frac{1}{n} \sum_{i=1}^{n} v_{32}(t_i) + o_p(n^{-\frac{1}{2}}) \]

\[ = \frac{1}{n} \sum_{j=1}^{n_{p}} \left[ v_{31}(y_j) + v_{32}(y_j) \right] + \frac{1}{n} \sum_{k=1}^{n_{p}} v_{32}(x_k) + o_p(n^{-\frac{1}{2}}), \]

where

\[ v_{31}(y) = \frac{1}{b_0} + a_0 h_0(y) - b_0 h_0^2(y), \]
\[ v_{32}(y) = \int g_{3a}(x)e_y(x)dG(x) - \frac{\lambda_0 w(h_0(y); \theta_0) \left[ \frac{1}{\lambda_0} + a_0 h_0(y) - b_0 h_0^2(y) \right]}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]} \]

\[ + \int g_{3a}(x)e_y(x)dG(x) + \frac{(1 - \lambda_0) \left[ \frac{1}{\lambda_0} + a_0 h_0(y) - b_0 h_0^2(y) \right]}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]} - v_{31}(y). \]

In a summary,

\[
\frac{1}{n} V_n(\theta_0) = \frac{1}{n} \begin{pmatrix}
V_{n1}(\theta_0) \\
V_{n2}(\theta_0) \\
V_{n3}(\theta_0)
\end{pmatrix}
\]

\[
= \frac{1}{n} \sum_{j=1}^{n_D} \begin{pmatrix}
v_1(y_j) \\
v_21(y_j) + v_22(y_j) \\
v_31(y_j) + v_32(y_j)
\end{pmatrix} + \frac{1}{n} \sum_{k=1}^{n_D} \begin{pmatrix}
v_1(x_k) \\
v_22(x_k) \\
v_32(x_k)
\end{pmatrix} + o_p(n^{-\frac{1}{2}})
\]

\[
= \frac{1}{n} \sum_{j=1}^{n_D} u_1(y_j) + \frac{1}{n} \sum_{k=1}^{n_D} u_2(x_k) + o_p(n^{-\frac{1}{2}}),
\]

where

\[
\begin{pmatrix}
v_1(y) \\
v_21(y) + v_22(y) \\
v_31(y) + v_32(y)
\end{pmatrix} = \begin{pmatrix}
\int g_{1a}(x)e_y(x)dG(x) - \frac{1}{\lambda_0} \left( 1 - \frac{1}{1 + \lambda_0 w(h_0(y); \theta_0) - 1} \right) \\
\int g_{2a}(x)e_y(x)dG(x) + \frac{(1 - \lambda_0) [b_0 h_0(y) - a_0]}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]} \\
\int g_{3a}(x)e_y(x)dG(x) + \frac{(1 - \lambda_0) \left[ \frac{1}{\lambda_0} + a_0 h_0(y) - b_0 h_0^2(y) \right]}{1 + \lambda_0[w(h_0(y); \theta_0) - 1]}
\end{pmatrix}.
\]

\[
\begin{pmatrix}
v_1(y) \\
v_22(y) \\
v_32(y)
\end{pmatrix} = \begin{pmatrix}
v_1(y) \\
v_21(y) \\
v_31(y)
\end{pmatrix} - \begin{pmatrix}
0 \\
v_21 \\
v_31
\end{pmatrix} = \begin{pmatrix}
0 \\
b_0 h_0(y) - a_0 \\
\frac{1}{b_0} + a_0 h_0(y) - b_0 h_0^2(y)
\end{pmatrix}.
\]
Plugging into (3.10) yields

\[ \sqrt{n}(\hat{\theta} - \theta_0) = n^{-\frac{1}{2}}A^{-1}V_n(\theta_0) + o_p(1) \]

\[ = n^{-\frac{1}{2}}A^{-1}\left\{ \sum_{j=1}^{n_D} u_1(y_j) + \sum_{k=1}^{n_D} u_2(x_k) + o_p(n^\frac{1}{2}) \right\} + o_p(1) \]

\[ = n^{-\frac{1}{2}}A^{-1}\left\{ \sum_{j=1}^{n_D} u_1(y_j) + \sum_{k=1}^{n_D} u_2(x_k) \right\} + o_p(1). \]

Due to that \( X \) and \( Y \) are independent, the covariance matrix of \( \sqrt{n}(\hat{\theta} - \theta_0) \) is

\[ \Sigma_{pel} = \lim_{n \to \infty} n^{-1}A^{-1}\{n_D \text{Var}(u_1(Y)) + n_D \text{Var}(u_2(X))\}(A^{-1})^\top \]

\[ = A^{-1}\{\lambda_0 \text{Var}[u_1(Y)] + (1 - \lambda_0) \text{Var}[u_2(X)]\}(A^{-1})^\top \]

\[ = \frac{1}{1 + \rho} A^{-1}\{\rho \text{Var}[u_1(Y)] + \text{Var}[u_2(X)]\}(A^{-1})^\top, \]

where \( \lambda_0 = \lim_{n \to \infty} \frac{n_D}{n} = \frac{\rho}{1 + \rho} \) and \( \rho = \lim_{n \to \infty} \frac{n_D}{n_D} \). According to Lemma 3.7.2, \( \hat{\theta} \) is a consistent estimator of \( \theta_0 \). Thus, \( \mathbb{E}\{\sqrt{n}(\hat{\theta} - \theta_0)\} = 0 \). By the Central Limit Theorem, \( \sqrt{n}(\hat{\theta} - \theta_0) \) converges in distribution to a normal random vector with expectation \( 0 \) and covariance matrix \( \Sigma_{pel} \). That is

\[ \sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\to} N(0, \Sigma_{pel}). \]

The estimation of \( \Sigma_{pel} \) is obtained by replacing all theoretical quantities by their empirical counterparts. That is

\[ \hat{\Sigma}_{pel} = \frac{1}{n^2} \hat{A}^{-1}\left\{ \sum_{j=1}^{n_D} \left[ \hat{u}_1(y_j) - \frac{1}{n_D} \sum_{j=1}^{n_D} \hat{u}_1(y_j) \right]^\top \left[ \hat{u}_1(y_j) - \frac{1}{n_D} \sum_{j=1}^{n_D} \hat{u}_1(y_j) \right] \right. \]

\[ + \sum_{k=1}^{n_D} \left[ \hat{u}_2(x_k) - \frac{1}{n_D} \sum_{k=1}^{n_D} \hat{u}_2(x_k) \right]^\top \left[ \hat{u}_2(x_k) - \frac{1}{n_D} \sum_{k=1}^{n_D} \hat{u}_2(x_k) \right] \left( \hat{A}^{-1} \right)^\top \right\} \left( \hat{A}^{-1} \right)^\top, \]
where

\[
\begin{align*}
\mathbf{u}(y) &= \int \begin{pmatrix} \widehat{g}_{1a}(x) \\ \widehat{g}_{2a}(x) \\ \widehat{g}_{3a}(x) \end{pmatrix} \mathbf{e}_y(x) dG(x) + \begin{pmatrix} \frac{1}{\lambda} (1 + \lambda w(h(y); \bar{a}, \bar{b}) - 1) \\ \frac{1}{1 + \lambda w(h(y); \bar{a}, \bar{b}) - 1} \end{pmatrix}, \\
&= \sum_{i=1}^{n} \begin{pmatrix} \widehat{g}_{1a}(t_i) \\ \widehat{g}_{2a}(t_i) \\ \widehat{g}_{3a}(t_i) \end{pmatrix} \mathbf{e}_y(t_i) \bar{p}(t_i) + \begin{pmatrix} \frac{1}{\lambda} (1 + \lambda w(h(y); \bar{a}, \bar{b}) - 1) \\ \frac{1}{1 + \lambda w(h(y); \bar{a}, \bar{b}) - 1} \end{pmatrix}, \\
\mathbf{u}(y) &= \mathbf{u}(y) - \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},
\end{align*}
\]

with \( \hat{h}(x) = h(x; \bar{a}, \bar{b}) \) and

\[
\begin{align*}
\widehat{g}_{1a}(x) &= \frac{-(1 + \rho) w(h(x); \bar{a}, \bar{b}) [\bar{a} \bar{b} + (1 - \bar{b}^2) \hat{h}(x)]}{\{1 + \lambda w(h(x); \bar{a}, \bar{b}) - 1\}\{\phi(h(x)) + \rho b \phi[h(x) - \bar{a}]\}}, \\
\widehat{g}_{2a}(x) &= \lambda (1 - \lambda) [\bar{b} h(x) - \bar{a}] \widehat{g}_{1a}(x), \\
\widehat{g}_{3a}(x) &= \lambda (1 - \lambda) \left[ \frac{1}{\lambda} + \bar{a} h(x) - \bar{b} \hat{h}^2(x) \right] \widehat{g}_{1a}(x),
\end{align*}
\]

and

\[
\mathbf{e}_y(x) = I(y \leq x) - \frac{\Phi(\hat{h}(x)) + \hat{\rho} \Phi(\bar{b} h(x) - \bar{a})}{1 + \hat{\rho}}.
\]

where \( \hat{\rho} = \frac{nD}{nD} \). The proof is complete.
Chapter 4

Asymptotic normality of kernel estimators of hazard functions under random censorship

4.1 Introduction

Let $T_1, \cdots, T_n$ be a sequence of independent, nonnegative random variables with common continuous distribution function $F$ and density function $f$. Independent of the $T_i$’s, let $U_1, \cdots, U_n$ be another sequence of independent, nonnegative random variables with common continuous distribution function $G$. We will refer to the $T_i$’s as survival times and to the $U_i$’s as censoring times. Under the random censorship model, we are only able to observe the smaller of $T_i$ and $U_i$ and an indicator of which variable was smaller:

$$X_i = \min(T_i, U_i), \quad \delta = I_{[T_i \leq U_i]}, \quad \text{for } i = 1, \cdots, n,$$

(4.1)

where $I_A$ for any event $A$ denotes the indicator function of $A$. Let $\lambda(t) = f(t)/(1 - F(t))$ be the hazard function of the survival times and $\Lambda = \int_0^t \lambda(s)ds$ be the corre-
sponding cumulative hazard function. Based on randomly censored data in (4.1), the Nelson [1969] cumulative hazard estimator for \( \Lambda(t) \) is defined by

\[
\Lambda_n(t) = \sum_{k: X(k) \leq t} \frac{\delta(k)}{n-k+1},
\]

where \( X(1) < X(2) < \cdots < X(n) \) is the order statistics of \( X_1, \cdots, X_n \), and \( \delta(i) \) is the value of \( \delta \) associated with \( X(i) \), that is, \( \delta(i) = \delta_j \) when \( X(i) = X_j \). By smoothing the increments in the Nelson estimator \( \Lambda_n \), the kernel estimator for the hazard function \( \lambda(t) \) is defined by

\[
\lambda_n(t) = \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) d\Lambda_n(s) \tag{4.3}
\]

where \( K \) is a kernel function having finite support on \((-1, 1)\) and \( h_n \) is a sequence of positive bandwidths tending to 0 as \( n \to \infty \). The properties of the kernel estimator \( \lambda_n \) have been examined by Ramlau-Hansen [1983], Tanner & Wong [1983], and Yandell [1983], among others. It is the purpose of this chapter to study the asymptotic normality for \( \lambda_n - \lambda \) using the strong approximation technique developed by Burke et al. [1981, 1988] in the censored case. Our approach is first to apply the strong approximation technique to establish the asymptotic normality of \( \lambda_n - \tilde{\lambda}_n \) (see (4.7) for the definition of \( \tilde{\lambda}_n \)). Then we show that \( \tilde{\lambda}_n - \mathbb{E}\lambda_n \) converges to zero at an exponential rate using the theory of martingales for counting processes. The martingale approach to the statistical analysis of counting processes was introduced by Aalen [1976, 1977, 1978] and has proved remarkably successful in yielding results about statistical methods for many problems arising in randomly censored data from biomedical studies. Fleming & Harrington [1991] provided an excellent exposition on the counting process and the martingale methods used with censored survival data.

In order to formulate our results, we first introduce some notations. Let \( S(t) = 1 - F(t) \), \( C(t) = 1 - G(t) \), \( \pi(t) = \Pr(X_1 \geq t) = S(t)C(t) \). In this paper we assume
that the kernel function $K$ is of bounded variation on $(-1, 1)$ with total variation denoted by $V_K$ and satisfies the following conditions

$$
\int_{-1}^{1} K(t) dt = 1, \quad \int_{-1}^{1} tK(t) dt = 0, \quad \text{and} \quad \int_{-1}^{1} t^2K(t) dt = k_2 \neq 0, \quad (4.4)
$$

Furthermore, for the sake of notational simplicity, we assume that $K(t) = 0$ if $t \notin (-1, 1)$.

Our main result of this chapter is the following theorem, whose proof is given in Section 4.3.

**Theorem 4.1.1.** Let $\tau$ be such that $\pi(\tau) > 0$. Suppose that $\lambda$ is continuous at $t \in [0, \tau)$ and that

$$
\lim_{n \to \infty} \min_{n} \left( \log n \right) \frac{\sigma}{n} = 0.
$$

Then for $t \in [0, \tau)$, as $n \to \infty$ and $h \to 0$,

$$
\sqrt{nh} (\lambda_n(t) - \mathbb{E}\lambda_n(t)) \overset{d}{\to} N(0, \sigma^2(t)),
$$

where

$$
\sigma^2(t) = \frac{\lambda(t)}{\pi(t)} \int_{-1}^{1} K^2(u) du. \quad (4.6)
$$

**Corollary 4.1.1.1.** In addition to the conditions in Theorem 4.1.1, we assume that $\lambda$ has a bounded derivative in a neighborhood of $t$ and that $nh^3 \to 0$ as $n \to \infty$, then

$$
\sqrt{nh} (\lambda_n(t) - \lambda(t)) \overset{d}{\to} N(0, \sigma^2(t)).
$$

Note that the assumption that $nh^3 \to 0$, as $n \to \infty$, requires that the bandwidth
tends to zero faster than $n^{-1/3}$. The following corollary gives the asymptotic normality of $\lambda_n$ for bandwidths of the order $n^{-1/5}$.

**Corollary 4.1.1.2.** In addition to the conditions in Theorem 4.1.1, we assume that $\lambda$ is twice continuously differentiable in a neighborhood of $t$ and that the bandwidth $h_n$ satisfies $h_n = O(n^{-1/5})$ as $n \to \infty$. Then, we have

$$\sqrt{n}h_n(\lambda_n(t) - \lambda(t)) - \frac{1}{2} h_n^2 \lambda''(t) k_2 \xrightarrow{d} N(0, \sigma^2(t))$$

as $n \to \infty$, where $k_2$ is given in (4.4).

## 4.2 Example

As described in Hosmer & Lemeshow [1999], a large HMO wishes to evaluate the survival time of its HIV+ members using a follow-up study. Subjects were enrolled in the study from January 1, 1989 to December 31, 1991. The study ended on December 31, 1995. After a confirmed diagnosis of HIV, members were followed until death due to AIDS or AIDS-related complications, until the end of the study or until the subject was lost to follow-up. The primary outcome variable of interest is survival time after a confirmed diagnosis of HIV. Data for 100 subjects are: Time: the follow-up time is the number of months between the entry date (ENT DATE) and the end date (END DATE), and CENSOR: vital status at the end of the study (1 = Death due to AIDS, 0 = Lost to follow-up or alive). We apply the kernel method to estimate the hazard rate and compare to KM estimates. Gaussian kernel and triangle kernel are selected. For illustration, the bandwidths are chosen to be $kn^{-\frac{1}{5}}$ where $k = 3, 6, 9, 12$. According to Zambom & Dias [2012], assuming that the reference density is Gaussian,
and a Gaussian kernel is used, we have

\[ b_{MISE} = \left( \frac{R(K)}{\mu_2(K)R(f''')} \right)^{1/5} n^{-1/5} = \left[ \frac{(2\sqrt{\pi})^{-1}}{3\pi^{-1/2}\sigma^{-5}} \right]^{1/5} n^{-1/5} = 1.06\sigma n^{-1/5} \]

By using an estimate of \( \sigma \), one has a data-based estimate of the optimal bandwidth. Applying this method to our example data, we obtain the optimal band width of 9.87. Figure 4-1 shows the K-M hazard rate estimation together with Gaussian and triangle kernel different bandwidths. Clearly, the increasing of bandwidth will result a smoother kernel estimation of hazard rate. With the optimal bandwidth 9.87 and a Gaussian kernel, we obtain a relatively flat kernel estimation (on the left of Figure 4-1). The same bandwidth of 9.87, but a triangle kernel function, yields less flat curve (on the right of Figure 4-1).

Figure 4-1: Scatter plot of the point wise estimator of the hazard function, and 5 kernel smooths from the HMO-HIV+ study

### 4.3 Proofs

The proof of Theorem 4.1.1 is based on the following two lemmas. We begin with introducing some further notations. Let \( \{W(t), t \geq 0\} \) be a standard Wiener process.
and \(\{W_n(t), t \geq 0\}\) be a sequence of standard Wiener processes. Without loss of generality, we assume throughout that all the random variables and the processes of this paper are defined on the same probability space. Furthermore, we define

\[
\tilde{\lambda}_n(t) = \frac{1}{h_n} \int_0^\infty K\left(\frac{t-s}{h_n}\right) \lambda(s) ds,
\]

\[
v(t) = \int_0^t \frac{1}{S^2(s)C(s)} dF(s),
\]

\[
B_n(t) = -\frac{1}{\sqrt{h_n}} \int_0^\infty K\left(\frac{t-s}{h_n}\right) dW_n(v(s)),
\]

\[
C_n(t) = -\frac{1}{\sqrt{h_n}} \int_0^\infty K\left(\frac{t-s}{h_n}\right) \sqrt{v'(s)} dW(s). \tag{4.7}
\]

Note that \(W_n(t) \overset{d}{=} W(t)\) for each \(n\).

We first establish the asymptotic normality for \(\lambda_n - \tilde{\lambda}_n\).

**Lemma 4.3.1.** Under the same conditions as in Theorem 4.1.1, as \(n \to \infty\) and \(h_n \to 0\), we have

\[
\sqrt{nh_n}(\lambda_n(t) - \tilde{\lambda}_n(t)) \overset{d}{\to} N(0, \sigma^2(t)).
\]

**Proof.** According to Theorem 2 of Burke et al. [1988], we can define a sequence of standard Wiener process \(\{W_n(t), t \geq 0\}\) such that

\[
\sup_{0 \leq t \leq \tau} |\sqrt{n}(\Lambda_n(t) - \Lambda(t)) - W_n(v(t))| = O(n^{-1/2} \log n) \quad \text{a.s.,}
\]

where \(\tau\) is such that \(\pi(\tau) > 0\). For \(t \in [0, \tau)\) and large \(n\), we have, with probability
\[
\sqrt{n h_n} (\lambda_n(t) - \tilde{\lambda}_n(t)) = -\frac{1}{\sqrt{h_n}} \int_0^\infty \sqrt{n} [\Lambda_n(s) - \Lambda(s)] dK \left( \frac{t - s}{h_n} \right)
\]
\[
= \frac{1}{\sqrt{h_n}} \int_{-1}^1 W_n(v(t - h_n u)) dK(u) + O \left( \frac{\log n}{\sqrt{n h_n}} \right)
\]
\[
= B_n(t) + O \left( \frac{\log n}{\sqrt{n h_n}} \right). \tag{4.8}
\]

Let \( \sigma_n^2(t) = \text{Var}(C_n(t)) \), then \( \sigma_n^2(t) = \int_{-1}^1 K^2(u) v'(t - h_n u) du \). It is easy to show by the continuity of \( \lambda \) at \( t \) that
\[
\lim_{n \to \infty} \sigma_n^2(t) = \sigma^2(t), \tag{4.9}
\]
where \( \sigma^2(t) \) is given in (4.6). Since \( C_n(t) \) is a normal random variable with mean O and variance \( \sigma_n^2(t) \), it follows from (4.9) and Slutsky’s Theorem that
\[
C_n(t) \xrightarrow{d} N(0, \sigma^2(t)). \tag{4.10}
\]
Furthermore, since \( B_n(t) \equiv C_n(t) \) for each \( n \), (4.10) implies that
\[
B_n(t) \xrightarrow{d} N(0, \sigma^2(t)). \tag{4.11}
\]
Combining (4.5), (4.8), (4.9) and ((4.11)) completes the proof of Lemma 4.3.1.

Next, we prove the following lemma, which indicates that \( \tilde{\lambda}_n - \mathbb{E} \lambda_n \) converges to zero at an exponential rate.

**Lemma 4.3.2.** Suppose that \( \lambda \) is continuous at \( t \) and \( \pi \) is positive in a neighborhood of \( t \), then for large \( n \),
\[
|\mathbb{E} \lambda_n(t) - \tilde{\lambda}_n(t)| \leq e^{-n \pi(t + h_n)} \int_{-1}^1 |K(u)| \lambda(t - h_n u) du.
\]
\[
114
\]
As a result, $\mathbb{E}\lambda_n(t) - \tilde{\lambda}_n(t)$ converges to zero at an exponential rate as $n \to \infty$ and $h_n \to 0$.

**Proof.** For $t \in [0, \infty)$, define

\[
N(t) = \sum_{i=1}^{n} I_{[X_i \leq t, \delta_i = 1]},
\]

\[
Y(t) = \sum_{i=1}^{n} I_{[X_i \geq t]},
\]

\[
A(t) = \int_{0}^{t} Y(u)d\Lambda(u),
\]

\[
M(t) = N(t) - A(t) = N(t) - \int_{0}^{t} Y(u)\lambda(u)du,
\]

\[
\Lambda_n^* = \int_{0}^{t} I_{[Y(u) > 0]}d\Lambda(u),
\]

\[
\lambda_n^*(t) = \frac{1}{h_n} \int_{0}^{\infty} K \left( \frac{t - s}{h_n} d\Lambda_n^*(s) \right). \quad (4.12)
\]

According to Theorem 1.3.1 of Fleming & Harrington [1991], the process $M$ given in (4.12) is an $\mathcal{F}_t$-martingale, where $\mathcal{F}_t = \sigma\{I_{[X_i \leq t, \delta_i = 0]} : 0 \leq s \leq t, i = 1, \ldots, n\}$. In fact, $M$ is a local square integrable martingale. Furthermore, the Nelson estimator in (4.2) can be expressed in terms of stochastic processes $N(t)$ and $Y(t)$ as follows:

\[
\Lambda_n(t) = \int_{0}^{t} \frac{dN(s)}{Y(s)} \quad (4.13)
\]

Simple algebra shows from (4.3), (4.12) and (4.13) that

\[
\Lambda_n(t) - \Lambda_n^*(t) = \int_{0}^{t} \frac{I_{[Y(u) > 0]}}{Y(u)} dM(u)
\]
\[
\lambda_n(t) - \lambda_n^*(t) = \frac{1}{h_n} \int_0^\infty K \left( \frac{t - s}{h_n} \right) \frac{I_{[Y(s) > 0]}}{Y(s)} dM(s).
\]

Thus, \( \lambda_n(t) - \lambda_n^*(t) \) is a stochastic integral with respect to the local square integrable martingale \( M(t) = N(t) - \int_0^t \lambda(s) Y(s) ds \). Using Theorem 2.4.5 of Fleming & Harrington [1991] and noting that \( \langle M, M \rangle = \int_0^t Y(s) \lambda(s) ds \), we have

\[
\mathbb{E}[\lambda_n(t) - \lambda_n^*(t)]^2 = \frac{1}{h_n^2} \mathbb{E} \left[ \int_0^\infty K \left( \frac{t - s}{h_n} \right) \frac{I_{[Y(s) > 0]}}{Y(s)} dM(s) \right]^2
\]

\[
= \frac{1}{h_n^2} \int_0^\infty K^2 \left( \frac{t - s}{h_n} \right) \frac{I_{[Y(s) > 0]}}{Y^2(s)} d\langle M, M \rangle(s)
\]

\[
\leq \frac{1}{h_n} \int_{-1}^1 K^2(u) \lambda(t - h_n u) du < \infty,
\]

and hence

\[
\mathbb{E} \lambda_n(t) = \mathbb{E} \lambda_n^*(t) = \mathbb{E} \left[ \frac{1}{h_n} \int_0^\infty K \left( \frac{t - s}{h_n} \right) d\Lambda_n^*(s) \right] = \frac{1}{h_n} \int_0^\infty K \left( \frac{t - s}{h_n} \right) \mathbb{E}[I_{[Y(s) > 0]}] d\Lambda(s).
\]

As a result, for large \( n \),

\[
|\mathbb{E} \lambda_n(t) - \tilde{\lambda}_n(t)| = |\mathbb{E} \lambda_n^*(t) - \tilde{\lambda}_n(t)|
\]

\[
\leq \frac{1}{h_n} \int_0^\infty \left| K \left( \frac{t - s}{h_n} \right) \right| \mathbb{E}[I_{[Y(s) = 0]}] \lambda(s) ds
\]

\[
= \int_{-1}^1 \left| K(u) \right| \left[ 1 - \pi(t - h_n u) \right] \lambda(t - h_n u) du
\]

116
\begin{align*}
&\leq [1 - \pi(t + h_n)]^n \int_{-1}^{1} |K(u)| \lambda(t - h_n u) du \\
&\leq e^{-n\pi(t + h_n)} \int_{-1}^{1} |K(u)| \lambda(t - h_n u) du.
\end{align*}

The proof is complete. \hfill \Box

\textbf{Proof of Theorem 4.1.1}: Theorem 4.1.1 is a straightforward consequence of Lemmas 4.3.1 and 4.3.2 and the equality $\lambda_n - \mathbb{E}\lambda_n = \lambda_n - \bar{\lambda} + \bar{\lambda} - \mathbb{E}\lambda_n$.

\textbf{Proof of Corollary 4.1.1.1}: Suppose that $|\lambda'(s)| \leq M$ for $s$ in a neighborhood of $t$, where $M$ is a constant depending only on $t$. Applying the mean value theorem with $\xi_n \in (\min(t, t - h_n u), \max(t, t - h_n u))$ gives

\begin{equation}
\sqrt{nh_n} |\tilde{\lambda}_n(t) - \lambda(t)| = \sqrt{nh_n} \left| \int_{-1}^{1} K(u)[\lambda(t - h_n u) - \lambda(t)] du \right| \\
= \sqrt{nh_n} \left| \int_{-1}^{1} K(u)[\lambda'(\xi_n)(-h_n u)] du \right| \\
\leq M \sqrt{nh_n^{3/2}} \int_{-1}^{1} |uK(u)| du \longrightarrow 0, \quad (4.14)
\end{equation}

as $n \to \infty$. Combining (4.14) and Lemma 4.3.1 completes the proof.

\textbf{Proof of Corollary 4.1.1.2}: Applying a two-term Taylor expansion gives, as $n \to \infty$,

\begin{align*}
\sqrt{nh_n} \left| \tilde{\lambda}_n(t) - \lambda(t) - \frac{1}{2} h_n^2 \lambda''(t)k_2 \right| &= \frac{1}{2} \sqrt{nh_n^{5/2}} \left| \int_{-1}^{1} u^2 K(u)\lambda''(\xi_n) du - \lambda''(t)k_2 \right| \\
&= \frac{1}{2} \sqrt{nh_n^{5/2}} \left| \int_{-1}^{1} u^2 K(u)[\lambda''(\xi_n) - \lambda''(t)] du \right|
\end{align*}
\[
\leq \frac{1}{2} \sqrt{\text{m}h_n^{5/2}} \int_{-1}^{1} u^2 |K(u)||\lambda''(\xi_n) - \lambda''(t)| du \rightarrow 0, \tag{4.15}
\]

where \( \xi_n \in (\min(t, t - h_n u), \max(t, t - h_n u)) \).

The proof is complete by combining (4.15) and Lemma 4.3.1.
References


Bamber, Donald. 1975. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12(4), 387–415.


