2008

EMPIRICAL PROCESSES AND ROC CURVES WITH AN APPLICATION TO LINEAR COMBINATIONS OF DIAGNOSTIC TESTS

Costel Chirila
University of Kentucky

ABSTRACT OF DISSERTATION

Costel Chirila

The Graduate School

University of Kentucky

2008

EMPIRICAL PROCESSES AND ROC CURVES WITH AN APPLICATION TO LINEAR COMBINATIONS OF DIAGNOSTIC TESTS

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Arts and Sciences
at the University of Kentucky

By

Costel Chirila

Lexington, Kentucky

Co-Directors: Dr. Constance L. Wood, Associate Professor of Statistics

and Dr. Arne C. Bathke , Associate Professor of Statistics

Lexington, Kentucky

2008

ABSTRACT OF DISSERTATION

EMPIRICAL PROCESSES AND ROC CURVES WITH AN APPLICATION TO LINEAR COMBINATIONS OF DIAGNOSTIC TESTS

The Receiver Operating Characteristic (ROC) curve is the plot of Sensitivity vs. 1- Specificity of a quantitative diagnostic test, for a wide range of cut-off points c. The empirical ROC curve is probably the most used nonparametric estimator of the ROC curve. The asymptotic properties of this estimator were first developed by Hsieh and Turnbull (1996) based on strong approximations for quantile processes. Jensen *et al.* (2000) provided a general method to obtain regional confidence bands for the empirical ROC curve, based on its asymptotic distribution.

Since most biomarkers do not have high enough sensitivity and specificity to qualify for good diagnostic test, a combination of biomarkers may result in a better diagnostic test than each one taken alone. Su and Liu (1993) proved that, if the panel of biomarkers is multivariate normally distributed for both diseased and non-diseased populations, then the linear combination, using Fisher's linear discriminant coefficients, maximizes the area under the ROC curve of the newly formed diagnostic test, called the generalized ROC curve. In this dissertation, we will derive the asymptotic properties of the generalized empirical ROC curve, the nonparametric estimator of the generalized ROC curve, by using the empirical processes theory as in van der Vaart (1998). The pivotal result used in finding the asymptotic behavior of the proposed nonparametric is the result on random functions which incorporate estimators as developed by van der Vaart (1998). By using this powerful lemma we will be able to decompose an equivalent process into a sum of two other processes, usually called the brownian bridge and the drift term, via Donsker classes of functions. Using a uniform convergence rate result given by Pollard (1984), we derive the limiting process of the drift term. Due to the independence of the random samples, the asymptotic distribution of the generalized empirical ROC process will be the sum of the asymptotic distributions of the decomposed processes. For completeness, we will first re-derive the asymptotic properties of the empirical ROC curve in the univariate case, using the same technique described before. The methodology is used to combine biomarkers in order to discriminate lung cancer patients from normals.

KEYWORDS: Diagnostic test, generalized ROC curve, Nonparametric Estimator,

Empirical Processes, Asymptotic properties

Costel Chirila

November 29, 2008

EMPIRICAL PROCESSES AND ROC CURVES WITH AN APPLICATION TO LINEAR COMBINATIONS OF DIAGNOSTIC TESTS

By

Costel Chirila

Constance L. Wood

(Co-Director of Dissertation)

Arne C. Bathke

(Co-Director of Dissertation)

William S. Griffith

(Director of Graduate Studies)

November 29, 2008

Date

RULES FOR THE USE OF DISSERTATIONS

Name	Date

DISSERTATION

Costel Chirila

The Graduate School

University of Kentucky

2008

EMPIRICAL PROCESSES AND ROC CURVES WITH AN APPLICATION TO LINEAR COMBINATIONS OF DIAGNOSTIC TESTS

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Arts and Sciences
at the University of Kentucky

By

Costel Chirila

Lexington, Kentucky

Co-Directors: Dr. Constance L. Wood, Associate Professor of Statistics

and Dr. Arne C. Bathke, Associate Professor of Statistics

Lexington, Kentucky

2008

To my wife Dana, my sons Andrei and Matei.
To my parents Constantina and Mitica.

ACKNOWLEDGMENTS

I would like to express my profound gratitude to my committee chair, Dr. Constance L. Wood, for her unlimited support and patience, and for her numerous and invaluable suggestions throughout my research. Without her tremendous help and excellent guidance, the completion of this Ph.D dissertation would have not been possible. I am also highly grateful to my committee co-chair, Dr. Arne C. Bathke, for his constant support and encouragement during this challenging work.

I would also like to thank to the other members of my committee, Dr. Arnorld J. Stromberg, Dr. William S. Griffith, and Dr. Thomas V. Getchell. I am very grateful to Dr. Arnold J Stromberg for introducing to me the ROC curve as an interesting research subject and for encouraging me to keep working. I am also very grateful to Dr. William S. Griffith for his excellent teaching of probability and measure theory, that proved to be essential in my research. I really appreciate the kindness of Dr. Edward Hirschowitz, from Department of Internal Medicine at University of Kentucky Medical Center, who provided the lung cancer data. I also want to address special thanks to my former colleagues, Dr. Chris P. Saunders and Dr. Mark J. Lancaster for their useful suggestions and their help with R and Latex.

I am also grateful to my parents for their unconditional faith in me throughout my life. Most especially, I would like to thank to my wife and kids for their understanding and encouragement to finish this work. Without their love and patience, I could have not gotten so far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Overview of the ROC Curve

The ***R****eceiver* ***O****perating* ***C****haracteristic* (ROC) curve has its roots in statistical decision theory and practice of quality control. During the 1950s, the ROC methodology was developed for signal detection experiments in radar. The fundamentals of this methodology, as it was originally applied to signal detection, can be found in Green and Swets (1966). Today, the ROC methodology is applied in a wide variety of scientific areas such as psychology, economics, machine learning, biomedical sciences, and many others (see Swets and Pickett [1982] for other examples). The ROC curve was first introduced in the biomedical area by Lusted (1960) for medical imaging (radiology) applications, but it became a much popular statistical tool after the publication of Swets and Pickett's (1982) text. Nowadays, in the *omics* era, when the discovery of biomarkers is considered the key to personalized medicine, we have seen a huge boom in ROC literature, that ranges from simple applications of the ROC curve to new methodological developments. A search of the PubMed database for "biomarkers and ROC curve" showed that there are slightly more than 2000 publications since the year 2000. Two excellent reviews of ROC methodology applied in the biomedical area are given by Zhou *et al.* (2002) and Pepe (2003).

In the context of biomedical applications, most often the signal event can be replaced by the true status of a disease, diseased or non-diseased, and the "place" of the observer is taken by a diagnostic test or biomarker used as a diagnostic tool (we will use them interchangeably). Let us assume that we know the exact classification of the study subjects in either one of the two categories, a situation in which we say that we have a *gold-standard*. Let D be a dichotomous variable which takes values 0 and 1 for the non-diseased and diseased subjects, respectively. We will assume that the diagnostic test variable is continuous, and that larger values are more likely to appear in the diseased population. Let Z be the random variable of the diagnostic test values. Denote by $X \sim F$ and $Y \sim G$ the continuous random variables and their cumulative distribution functions (cdf) of the test values for the non-diseased and diseased subjects, respectively. By choosing a cut-off value c, a study subject has a positive (negative) test if the values of the diagnostic test is greater than c (less than or equal to c). Since we know exactly whether a subject is either diseased or non-diseased, the result of the test can be classified as *true positive* (TP), *false positive* (FP), *true negative* (TN), or *false negative* (FN). Thus, given N subjects and any cut-off value c, we can construct the following 2x2 table.

Table 1.1: Classification of Diagnostic Test Results

Diagnostic Test/Disease Status	Diseased	Non-diseased	Total
Positive Test	TP	FP	$TP + FP$
Negative Test	FN	TN	$FN + TN$
Total	$TP + FN$	$FP + TN$	N

A test result is TP (FP) when a diseased subject is correctly (erroneously) classified as diseased. Similarly, a test result is TN (FN) when a non-diseased subject is correctly (erroneously) classified as non-diseased. Based on the above classification of a test result let us introduce the following *accuracy measures*

$$TPF(c) = P(Z > c|D = 1) = P(Y > c) = 1 - G(c) = \overline{G}(c) \tag{1.1}$$

$$FPF(c) = P(Z > c|D = 0) = P(X > c) = 1 - F(c) = \overline{F}(c) \tag{1.2}$$

$$TNF(c) = P(Z \leqslant c|D = 0) = P(X \leqslant c) = F(c) \tag{1.3}$$

$$FNF(c) = P(Z \leqslant c|D = 1) = P(Y \leqslant c) = G(c), \tag{1.4}$$

where $\overline{F}$ and $\overline{G}$ are the survival functions. In the medical literature TPF and TNF are also called *Sensitivity* and *Specificity*, respectively. Note that, for any given cut-off value, among the four fractions exist the following relations

$$TPF(c) + FNF(c) = 1$$

$$TNF(c) + FPF(c) = 1.$$

Therefore, only two of the above four fractions, or *"operating characteristics"*, can be really used to gain insights in how well the diagnostic test has done. Let us choose *Sensitivity(c)*and *Specificity(c)*. An ideal diagnostic test would be able to perfectly discriminate between non-diseased and diseased subjects or, in other words, to have sensitivity and specificity equal to 1. This is rarely the case in practice and, as a matter of fact, the sensitivity increases from 0 to 1, while the specificity decreases from 1 to 0 as the cut-off point varies from $+\infty$ to $-\infty$ (it practically only varies on the range of the diagnostic test values). Therefore, by plotting sensitivity versus 1-specificity for all the possible cut-off points c, we obtain a visualization tool, namely the ROC curve, that shows the trade-off, or interdependence, between the sensitivity and specificity of a diagnostic test at each cut-off value. The ROC curve can also be considered as a performance measure of the diagnostic test. An ROC curve close to, but above the first diagonal of the unit square indicates that our diagnostic test has slightly better chances to distinguish between diseased and non-diseased subjects than flipping a coin. The figure below shows an example of a diagnostic test that is better than the flip of a coin. Any point on the ROC curve is determined by its

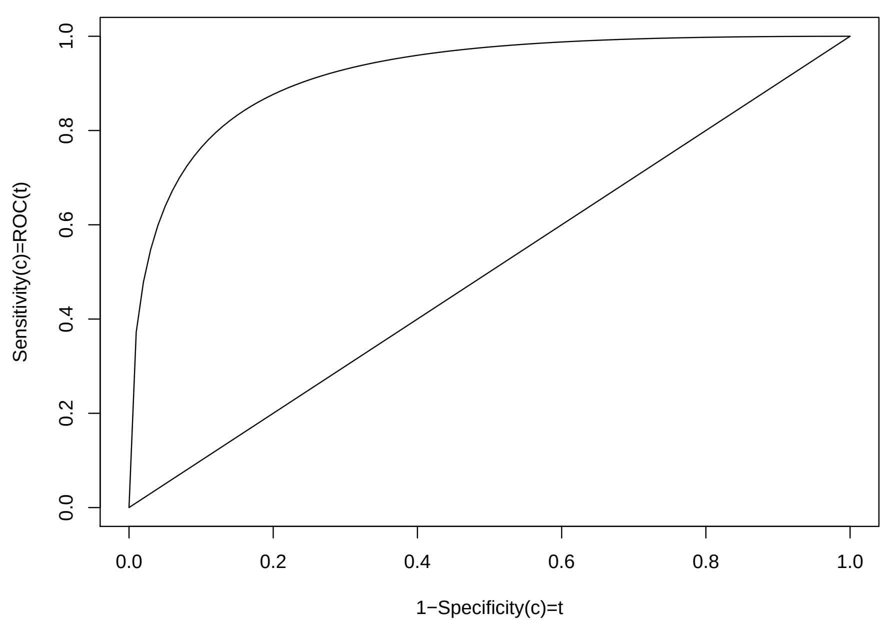

Figure 1.1: Example of an ROC Curve

coordinates

$$ROC(\,\cdot\,) = \{(1 - \text{Specificity}(c), \text{Sensitivity}(c)), \quad c \in \mathbb{R}\}\,.$$

Notice that if we denote 1-specificity by t and the sensitivity by $ROC(t)$, then by using the above formulae for the fractions we obtain

$$ROC(\,\cdot\,) = \left\{(t, ROC(t)) = \left(t, 1 - G\left(F^{-1}(1-t)\right)\right), \quad t \in [0,1]\right\}. \tag{1.5}$$

From (1.5), we see that the ROC curve is completely determined by the quantity $G\left(F^{-1}(p)\right)$ for $p \in [0,1]$.

It is worth noting a few properties of the ROC curve. Firstly, ROC curves are invariant under monotone increasing transformations. If H is such a transformation, then the ROC for X and Y is the same as the ROC for $H(X)$ and $H(Y)$. This property lead to the so called "binormal" assumption, in which the idea is to find the transformation H so that $H(X)$ and $H(Y)$ are both normally distributed (the binormal model). Secondly, the ROC curve lies above the first diagonal of the unit square if X is stochastically smaller than Y, (i.e., $F(c) \geqslant G(c)$, $\forall c$). Thirdly, if the probability density functions (pdf) f and g have monotone likelihood ratio $L(c) = g(c)/f(c)$, then the curve is concave. Next, we introduce the *Area Under the Curve* (AUC), which is one of the most used summary indices of the ROC curve (for other indices see Pepe, section 4.3, [2003]). It was shown by Bamber (1975) that $AUC = P(X \leqslant Y)$, meaning that AUC is the probability that the test can correctly discriminate between a diseased and a non-diseased from a random pair of subjects. However, since clinicians are more interested in specific ranges of the ROC curve, an alternative measure is the *partial Area Under the Curve* (pAUC) proposed by McClish (1989), Thompson and Zucchini (1989), and Dodd and Pepe (2003).

Here, we consider a panel of biomarkers, or multivariate diagnostic tests. Note that most biomarkers do not have high enough sensitivity and specificity to qualify for good diagnostic tests alone. Therefore, by combining the information of each individual biomarker we may obtain a better diagnostic test than each one taken alone. In the past years, there has been an increasing interest in constructing ROC curves based on a combination of biomarkers. The challenge of this problem is given by the fact that the natural ordering of the real numbers, which we used in constructing the ROC curve, is lost when we move up to dimensions higher than two. One solution to this problem, proposed by Baker (2000), was to create a new ordering relationship. Another solution is to use a multivariate model or a transformation that constructs a one-dimensional projection. Among the multivariate models used, we mention here logistic regression and tree-based models. These models estimate the predicted probability of the disease, which, in turn, can be used as a diagnostic test to create an ROC curve. On the other hand, Su and Liu (1993) proposed to create a new diagnostic test as a linear combination of biomarkers such that the AUC under the newly created ROC curve, also called the *generalized ROC curve*, is maximized. Su and Liu actually proved that, if the panel of biomarkers is multivariate normally distributed for both diseased and non-diseased populations, then the linear combination, using Fisher's linear discriminant coefficients, maximizes the AUC. Also, it was pointed out that, if the covariance matrices of the two multivariate normal distributions are assumed proportional, then Fisher's linear discriminant coefficients provide the highest sensitivity uniformly at any given specificity. Pepe and Thompson (2000) were able to drop the normality assumption and obtain estimates of the coefficients, by numerically maximizing the Mann-Whitney U-statistics, a nonparametric estimator of AUC. Pepe *et al.* (2006) reconsidered the problem in the ROC-GLM framework and looked at the AUC maximization as a special case of the maximum rank correlation estimator described by Han (1987). Moreover, it was also shown through simulations, that the AUC maximization approach is comparable with the logistic likelihood maximization (i.e., logistic regression) when the logistic model holds, and it is much better when the model does not hold. Other work on the generalized ROC curve was done by Reiser and Faraggi (1997) who developed confidence intervals for AUC using Wishart distributions, and Schisterman *et al.* (2004) who adjusted the generalized ROC curve for covariates. Using the same argument as in the pAUC case, Liu *et al.* (2005) proposed linear combinations of biomarkers that maximize the sensitivity over a desired range of specificity, instead of AUC as in Su and Liu (1993).

1.2 Estimation of the ROC Curve

Recall that the ROC curve is practically determined by the quantity $G\left(F^{-1}(p)\right)$ where $p \in [0,1]$. Since in practice the cdf's F and G are unknown, we need to estimate them. Therefore, we randomly select a sample of n non-diseased subjects, also called *"controls"*, and m diseased subjects, called *"cases"*. Moreover, based on Table 1.1 we can calculate the fractions of correctly or incorrectly classified subjects, for every given cut-off value c. The methods that are usually used for the estimation of the ROC curve can be roughly classified as parametric, semiparametric, and nonparametric. We will briefly describe them next, and provide some literature references.

The parametric estimation of the ROC curve consists in assuming that the diagnostic test variables X, Y have a known probability distribution which depends on some unknown parameters. The most used model is the "binormal" model, in which the diagnostic test variables are both normally distributed, $X \sim N(\mu_{ND}, \sigma^2_{ND})$ and $Y \sim N(\mu_D, \sigma^2_D)$. Then, it is easy to show that $TPF = ROC = \Phi(a + b\Phi^{-1}(FPF))$ where $a = (\mu_D - \mu_{ND})/\sigma_D$ and $b = \sigma_{ND}/\sigma_D$. The parameters are usually estimated using the maximum likelihood method (see, for example, Dorfman and Alf [1968] and [1969]). Of course, as with any other parametric approach, the estimates can be biased when the data does not follow the Gaussian distribution (see Goddard and Hinberg [1990]).

The semiparametric methods are developed as a compromise solution between parametric and nonparametric approaches. The most known semiparametric method was presented in Section 1.1 as the binormal assumption, although it is also confusingly called, by some authors, the binormal model. After the data transformation, the parameter estimation can be done in several ways, of which we mention here Hsieh and Turnbull (1996), Metz *et al.* (LABROC method) (1998), Zou and Hall (2000), Pepe (ROC-GLM method)(2000, and section 5.5.2, [2003]). Discussions about how realistic this approach is can be found in a series of papers by Hanley (1988) and (1996), Metz *et al.* (1998), among others.

The nonparametric estimation of the ROC curve is appealing because it does not impose any parametric model, with or without transformation, on the cdf's F and G. Therefore, F and G can be estimated either by using kernel (smoothing) methods or empirical methods. Estimation of the ROC curve using kernel based methods was first introduced by Zou *et al.* (1997) and improved by Lloyd (1998) and Zhou and Harezlak (2002). The empirical method consists in estimating F and G by their empirical distribution functions (edf) F_n and G_m. Campbell (1994) presented the empirical ROC curve, $1 - G_m\left(F_n^{-1}(1-t)\right)$ with $t \in [0,1]$, and its associated confidence region, based on Kolmogorov-Smirnov statistics and bootstrapping. Hsieh and Turnbull (1996) obtained the asymptotic properties of the empirical ROC curve on any interval $[a, b] \subset (0,1)$ using strong approximation results from Csörgő and Révész (1981) and Csörgő (1983). Using the asymptotic properties derived by Hsieh and Turnbull, Jensen *et al.* (2000) derived regional confidence bands for the smoothed empirical ROC curve. Li *et al.* (1996) derived the asymptotic properties of the empirical ROC curve under censoring, using empirical processes theory and the functional delta method. By using the same methodology, Li *et al.* (1999) introduced a mixed approach, in which one cdf is modelled parametrically and the other nonparametrically, arguing that this approach will result in smaller asymptotic variance than in the nonparametric case. Claeskens *et al.* (2003) used empirical likelihood to estimate the ROC curve, and based on that they constructed confidence regions. Recently, Gu and Ghoshal (2008a, 2008b, 2008c) proposed new estimation methods of the ROC curve using nonparametric bayesian inference, specifically, bayesian rank-based partial likelihood and bayesian bootstrapping. The asymptotic properties were based on strong approximation theory. Based on this approach, they also constructed credible confidence bounds.

1.3 Proposed Methods and Results

As we said before, developing multivariate diagnostic tests from large datasets, high-throughput screening data from gene expression arrays or mass spectrometry technologies, has become a very interesting and challenging research subject. In this dissertation, we will construct a multivariate diagnostic test as a linear combination of univariate diagnostic tests, using the methodology proposed by Su and Liu (1993). Again, we point out that the coefficients of the linear combination are determined such the AUC under ROC curve of the newly formed diagnostic test is maximized. The unknown coefficients of this transformation are estimated by their maximum likelihood estimators. There seems to be little research about the statistical properties of the generalized empirical ROC curve, the nonparametric estimator of the generalized ROC curve. Therefore, our main goal is to derive the asymptotic distribution of the generalized empirical ROC curve. Note that, given that the asymptotic behavior of the generalized empirical ROC curve is known, one can construct either pointwise or regional confidence bands, as presented in Jensen *et al.* (2000).

Here, we will derive the asymptotic properties by using the empirical processes theory as in van der Vaart (1998). Shortly, the major steps of this technique can be described as follows. Firstly, we rewrite the generalized empirical ROC process in an equivalent form using uniform edf's. Secondly, we decompose this equivalent process into a sum of two other processes, usually called the brownian bridge and the drift term, using the powerful Lemma 19.24 (van der Vaart [1998]), via Donsker classes of functions. Thirdly, we find the asymptotic distribution of each of the decomposed processes. Due to the independence of the random samples, the asymptotic distribution of the generalized empirical ROC process will be the sum of the asymptotic distributions found previously. For completeness, we will first re-derive the asymptotic properties of the empirical ROC curve in the univariate case, using the major steps described before.

1.4 Organization

This dissertation is organized as follows. In Chapter 2, we will introduce the basic concepts from measure and probability theory and provide the main results from empirical processes theory to be used later on. The pivotal results used to derive the asymptotic distribution of the ROC processes are Lemma 19.24 from van der Vaart (1998) and Theorem 37 from Pollard (1984). In Chapter 3, we will re-derive the asymptotic distribution of the empirical ROC process using the empirical processes approach and the functional delta method. The results are presented in Theorem 3.14 and Corollary 3.15. In Chapter 4, we will derive the main result of this dissertation, namely the asymptotic distribution of the generalized empirical ROC process on the interval $[0, 1]$, by using the core technique introduced in the previous chapter. The working assumption is that the biomarker panel is multivariate normally distributed and the covariance matrices for the diseased and non-diseased are the same. The main result is obtained in Theorem 4.55. In Chapter 5, we will apply the methodology to a set of biomarkers used for discrimination between lung cancer and normal subjects and present a simulation study. In Chapter 6, we will discuss the results and future work.

CHAPTER 2: PRELIMINARY TOOLS

2.1 Basic Definitions and Theory

For completeness, we will introduce notations, definitions and main results from empirical process theory that we will be using in the subsequent chapters. In this section, we will start with the basics from measure and probability theory and we will end with some results concerning the generalized inverse function. All results from this chapter will be stated without proof, but, for those interested in their proofs, we will add in parenthesis the source of the result.

Definition 2.1. A *metric* is a map $d : \mathbb{D} \times \mathbb{D} \mapsto [0, \infty)$ with properties

1. $d(x, y) = d(y, x)$;
2. $d(x, z) \leqslant d(x, y) + d(y, z)$ (triangle inequality);
3. $d(x, y) = 0$ if and only if $x = y$.

Definition 2.2. A set $\mathbb{D}$ equipped with a metric d is called a *metric space* and is denoted $(\mathbb{D}, d)$.

Definition 2.3. A subset of a metric space is *dense* if and only if its closure is the whole space. A metric space is *separable* if and only if it has a countable dense subset.

Definition 2.4. A subset K of a metric space is *compact* if and only if it is closed and every sequence in K has a converging subsequence. A subset K is *totally bounded* if and only if for every $\varepsilon > 0$ it can be covered by finitely many balls of radius ε.

Definition 2.5. A *norm* is a map $\|\cdot\| : \mathbb{D} \mapsto [0, \infty)$ such that for every $x, y \in \mathbb{D}$ and $\alpha \in \mathbb{R}$,

1. $\|x + y\| \leqslant \|x\| + \|y\|$ (triangle inequality);
2. $\|\alpha x\| = |\alpha| \, \|x\|$;
3. $\|x\| = 0$ if and only if $x = 0$.

Definition 2.6. A set $\mathbb{D}$ equipped with a norm is called a *normed space*.

Remark 2.7. If $\|\cdot\|$ is norm then $d(x, y) = \|x - y\|$ is a metric.

Remark 2.8. A *semimetric*(*seminorm*) is map that satisfies only conditions 1 and 2 from Definition 2.1(Definition 2.5).

Remark 2.9. Here are some examples of normed spaces that we will work with later on. Let $-\infty \leqslant a < b \leqslant \infty$ and $\mathcal{S} = \{f : [a, b] \mapsto \mathbb{R}\}$. Depending on the type of functions f, the set $\mathcal{S}$ will have different notations. $C[a, b]$ is the set of all continuous functions, $D[a, b]$ is the set of all functions that are right continuous and whose left limits exists everywhere in $[a, b]$ and $l^\infty[a, b]$ is the set of all bounded functions. We will equip these spaces with the *uniform norm* defined as $\|f\|_\infty = sup_{x \in [a,b]} |f(x)|$. When the limits a, b are not included, we will adjust the notation correspondingly.

Definition 2.10. Let Ω be a arbitrary set. A class $\mathcal{U}$ of subsets of Ω is called *σ-field* if:

1. $\emptyset, \Omega \in \mathcal{U}$;
2. if $A \in \mathcal{U}$ then its complement $A^c \in \mathcal{U}$;

3. if $A_1, A_2, \ldots$ is a countable collection of sets in $\mathcal{U}$ then $\bigcup_i A_i \in \mathcal{U}$ and $\bigcap_i A_i \in \mathcal{U}$.

Remark 2.11. A set Ω together with the σ-field $\mathcal{U}$ on it is called a *measurable space.*

Definition 2.12. The smallest σ-field that contains the open sets of a metric space $\mathbb{D}$ is called a *Borel σ-field.*

Remark 2.13. We will denote by $\mathcal{B}(\mathbb{R})$ the Borel σ-field on the real line.

Definition 2.14. Let $\mathcal{U}$ be a σ-field of Ω. A function $\mu : \mathcal{U} \to \mathbb{R}$ is called a *measure* if:

1. $0 \leqslant \mu(A) \leqslant \infty$, $\forall A \in \mathcal{U}$;

2. $\mu(\emptyset) = 0$;

3. if $A_1, A_2, \ldots$ is a countable collection of pairwise disjoint sets in $\mathcal{U}$ then $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$.

Remark 2.15. A measure P for which $P(\Omega) = 1$ is called a *probability measure.* The space $(\Omega, \mathcal{U}, P)$ is called a *probability space.*

Let $(\Omega, \mathcal{U}, P)$ be a probability space and $(\mathbb{D}, d)$ be a metric space with $\mathcal{D}$ a σ-field on it.

Definition 2.16. A map $X : \Omega \to \mathbb{D}$ is called a *$\mathcal{U}/\mathcal{D}$-measurable map* if for any $D \in \mathcal{D}$ the set $\{\omega \in \Omega : X(\omega) \in D\} \in \mathcal{U}$.

Remark 2.17. If $\mathcal{D}$ is the Borel σ-field then X is called *Borel-measurable.*

Definition 2.18. A map $X : \Omega \mapsto \mathbb{D}$ is called a *random element* with values in $\mathbb{D}$ if it is Borel-measurable.

Remark 2.19. When $\mathbb{D} = \mathbb{R}(\mathbb{R}^k)$, X is called a *random variable (vector)*. If $\mathbb{D}$ is a space of functions like $C[a,b]$, $D[a,b]$ or $l^\infty[a,b]$ then, X is called a *random function*.

Definition 2.20. A random element $X : \Omega \to \mathbb{D}$ is called *tight* if for every $\varepsilon > 0$ there exists a compact set K such that $P(X \notin K) < \varepsilon$.

Definition 2.21. Let T be an arbitrary set. A collection $X = \{X_t : t \in T\}$ of random variables indexed by T and defined on the same probability space $(\Omega, \mathcal{U}, P)$ is called a *stochastic process*.

Remark 2.22. For a fixed ω, the map $t \mapsto X_t(\omega)$ is called a *sample path*. If, for example, every sample path is a bounded function, then X can be viewed as a random element with values in $l^\infty(T)$. A classical example of a stochastic process is the empirical distribution function and we will talk more about it in a later section.

Definition 2.23. A stochastic process $X = \{X_t : t \in T\}$ is called *Gaussian* if the random vector $(X_{t_1}(\omega), \ldots, X_{t_k}(\omega))$ is multivariate normal for $\forall k \in \mathbb{N}$ and $\forall t_k \geqslant 0$.

Definition 2.24. Let $X : \Omega \to \mathbb{D}$ be a random element. The induced probability measure $P_X : \mathbb{D} \to \mathbb{R}$ defined by

$$P_X(D) = P\left(X^{-1}(D)\right) = P\left(\omega : X(\omega) \in D\right), \quad \forall D \in \mathcal{D},$$

is called the *probability distribution* or simply *distribution* of X.

Remark 2.25. When there is no confusion, we will drop the subscript and, in order to make a distinction between the two probabilities, we will denote by P the probability measure and by P the induced probability distribution.

Definition 2.26. A random element $X : \Omega \to \mathbb{D}$ is called *separable* if exists a separable, measurable set $D \in \mathcal{D}$ with $P_X(D) = 1$.

Definition 2.27. The *distribution function* of a random variable X, is the right continuous function defined on $\mathbb{R}$ by

$$F(x) = P_X((-\infty, x]) = P(\omega : X(\omega) \leqslant x).$$

Definition 2.28. The *expectation* of a random variable X is the Lebesgue-Stieltjes integral of $X(\omega)$ with respect to probability measure P.

Remark 2.29. Some common notations that we will use are: EX, $\int_\Omega X(\omega)dP(\omega))$ or $\int xdP_X(x)$.

Definition 2.30. The p^{th} *quantile* of a distribution function F is the quantity given by

$$F^{-1}(p) = \inf_{x \in \mathbb{R}} \{x : F(x) \geqslant p\}, \quad 0 < p < 1.$$

Let $F^{-1} : (0, 1) \mapsto \mathbb{R}$ be the *quantile function* or *generalized inverse function.* Next, we will state some very useful properties of the quantile function.

Lemma 2.31. *(Lemma 1.1.4, Serfling, , p. 3), Let F be a distribution function. The quantile function is non-decreasing and left continuous, and satisfies*

1. $F^{-1} \circ F(x) \leqslant x$, $-\infty < x < \infty$ *and*

2. $F \circ F^{-1}(p) \geqslant p$, $0 < p < 1$. *Hence*

3. $F(x) \geqslant p$ *if and only if* $x \geqslant F^{-1}(p)$.

Corollary 2.32. *For every $p \in (0,1)$ and $x \in \mathbb{R}$, $F \circ F^{-1}(p) \equiv p$ iff F is continuous and $F^{-1} \circ F(x) \equiv x$ iff F is strictly increasing.*

Remark 2.33. (Theorem 2.1.3 A, Remark (i), Serfling , p. 59), For any random sample $\{X_i\}_{i=\overline{1,n}}$ from distribution function F one can construct independent uniform $[0,1]$ random variables such that

$$P\left(X_i = F^{-1}(U_i)\right) = 1, \quad i = \overline{1,n}. \tag{2.1}$$

Lemma 2.34. *(Theorem 1, Shorack and Wellner, 1986, p. 3), Let $\xi \sim Unif(0,1)$. Then, for a fixed distribution function F, the random variable, obtained by thequantile transformation, $X \equiv F^{-1}(\xi)$ has distribution function F.*

Lemma 2.35. *Let $X \sim F$. Then, the random variable $U \equiv F(X)$, obtained by the probability integral transformation, is uniformly distributed on $[0,1]$ if and only if F is continuous.*

Lemma 2.36. *(Proposition 6, Shorack and Wellner, 1986, p. 9), If F has a positive continuous density in the neighborhood of $F^{-1}(p)$ where $p \in (0,1)$, then $(d/dp)F^{-1}(p)$ exists and equals $1/f(F^{-1}(p))$.*

2.2 Stochastic Convergence in Metric Spaces

We will introduce now three modes of stochastic convergence in metric spaces and state properties involving these modes of convergence. Also, we will introduce the useful notations $o_p(1)$, $O_p(1)$ and operations with them.

Let $(\Omega, \mathcal{U}, P)$ be an arbitrary probability space and $(\mathbb{D}, d)$ a metric space with $\mathcal{D}$ its Borel σ-field on it. Let $X_n : \Omega_n \mapsto \mathbb{D}$ be a sequence of arbitrary maps defined on probability spaces $(\Omega_n, \mathcal{U}_n, P_n)$ and $X : \Omega \mapsto \mathbb{D}$ be a random element. Note that, in the classical theory of stochastic convergence, X_n are required to be measurable, condition that usually holds when $\mathbb{D}$ is a separable metric space ($\mathbb{R}$ for example). This requirement fails when dealing with empirical processes (See [van der Vaart and Wellner, 1996, p. 3] and [Bilingsley, 1968, pp. 150-152]) for such examples). There were several attempts to solve this problem but none of those was totally satisfactory until Hoffmann-Jørgensen developed a new concept of weak convergence based on outer expectation.

Definition 2.37. Let $X : \Omega \mapsto \mathbb{D}$ an arbitrary map. The *outer expectation* of X with respect to P is given by

$$E^*X = \inf\{EU : U : \Omega \mapsto \mathbb{R},\ \text{measurable}, U \geqslant X, EU \text{ exists}\}.$$

The *outer probability* of an arbitrary subset $B \in \Omega$ is given by

$$P^*(B) = \inf\{P(A) : A \supset B, A \in \mathcal{U}\}.$$

Definition 2.38. The sequence X_n converges *in probability* to X, denoted $X_n \xrightarrow{P} X$, if $P^*(d(X_n, X) > \varepsilon) \to 0$, as $n \to \infty$.

Definition 2.39. The sequence X_n converges *almost surely* to X, denoted $X_n \xrightarrow{a.s.} X$, if there exists a sequence of measurable random variables Δ_n such that $d(X_n, X) \leqslant \Delta_n$

and $\Delta_n \to 0$ almost sure as $n \to \infty$.

Definition 2.40. The sequence X_n converges *weakly* (or in distribution) to X, if $E^* f(X_n) \to Ef(X)$, as $n \to \infty$, for every bounded, continuous function $f : \mathbb{D} \mapsto \mathbb{R}$. We denote this type of convergence by $X_n \rightsquigarrow X$, as $n \to \infty$.

Lemma 2.41. ***Continuous mapping****, (Theorem 1.3.6, van der Vaart and Wellner, 1996, p. 20), Let $g : \mathbb{D} \mapsto \mathbb{E}$ be continuous at every point of a set $\mathbb{D}_0 \subset \mathbb{D}$. If $X_n \rightsquigarrow X$ and X takes its values in $\mathbb{D}_0$, then $g(X_n) \rightsquigarrow g(X)$.*

Lemma 2.42. *Let $X_n, Y_n : \Omega_n \mapsto \mathbb{D}$ be some arbitrary maps and X be a random element with values in $\mathbb{D}$. If $X_n \rightsquigarrow X$ and $d(X_n, Y_n) \xrightarrow{P} 0$, then $Y_n \rightsquigarrow X$.*

Lemma 2.43. ***Slutsky's Lemma****, (Example 1.4.7, van der Vaart and Wellner, 1996, p. 32), Let $X_n : \Omega_n \mapsto \mathbb{D}$, $Y_n : \Omega_n \mapsto \mathbb{E}$ be some arbitrary maps such that $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$ with X separable and c a constant. Then, $(X_n, Y_n) \rightsquigarrow (X, c)$.*

Lemma 2.44. *(Lemma 18.13, van der Vaart, 1998, p. 261), Let $\mathbb{D}_0 \subset \mathbb{D}$ be arbitrary metric spaces equipped with the same metric. If X and every X_n take their values in $\mathbb{D}_0$, then $X_n \rightsquigarrow X$ as maps in $\mathbb{D}_0$ if and only if $X_n \rightsquigarrow X$ as maps in $\mathbb{D}$.*

Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of random variables. Then, the notation $X_n = o_p(1)$ means that $X_n \to 0$ in probability. The notation $X_n = O_p(1)$ means that X_n is bounded in probability or, equivalently, for every $\varepsilon > 0$ there exist $M_\varepsilon < \infty$ and $N_\varepsilon \in \mathbb{N}$ such that $P(|X_n| > M_\varepsilon) < \varepsilon$, $\forall n \geqslant N_\varepsilon$. More generally, let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables. By $X_n = o_p(Y_n)$ we will understand $X_n = Y_n R_n$ where $R_n = o_p(1)$. By $X_n = O_p(Y_n)$ we will understand $X_n = Y_n R_n$ where $R_n = O_p(1)$.

Lemma 2.45. *The following identities are true.*

1. $o_p(1) + o_p(1) = o_p(1)$;

2. $o_p(1) + O_p(1) = O_p(1)$;

3. $O_p(1)o_p(1) = o_p(1)$;

4. $(1 + o_p(1))^{-1} = O_p(1)$;

5. $o_p(R_n) = R_n o_p(1)$;

6. $O_p(R_n) = R_n O_p(1)$.

2.3 Empirical Processes

Now, We are able to talk about some important empirical processes results. First, we will state some classical results regarding empirical distributions. Then, we will introduce the Glivenko-Cantelli, Donsker, and Vapnik-Cervonenskis classes of functions as a main tool in proving weak convergence of empirical processes. The uniform version of Lemma 19.24 from van der Vaart (1998) that will be stated next, will play a pivotal role in finding the asymptotic distribution of the ROC processes. We will continue with the Hadamard differentiability and related results and we will end with Theorem 37 from Pollard (1984).

Let $(\Omega, \mathcal{U}, \mathrm{P})$ be an arbitrary probability space and $(\mathcal{X}, \mathcal{A})$ a measurable space. Let $X_1, \ldots, X_n$ be a *random sample* from probability distribution P with values on $\mathcal{X}$. Notice that since we dropped the subscript X from the induced distribution, we denote the probability measure by P and the distribution of X by P. Let $\mathcal{F}$ be a class of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$.

Definition 2.46. Let A be an arbitrary set. Then the *indicator function* $\mathbf{I}_A(x)$ is defined as

$$\mathbf{I}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \in A^c. \end{cases}$$

Definition 2.47. Let $A \in \mathcal{A}$. Then the *dirac measure*, or (point mass) at the observation, is defined as

$$\delta_x(A) = \begin{cases} 1, & x \in A, \\ 0, & x \in A^c. \end{cases}$$

Definition 2.48. The *empirical distribution* is the discrete uniform measure on the observations, $\mathbb{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}$.

Remark 2.49. The expectations under $\mathbb{P}_n$ and P are, respectively,

$$\mathbb{P}_n f = n^{-1} \sum_{i=1}^{n} f(X_i) \quad \text{and} \quad Pf = \int f dP.$$

For example, if $f = \mathbf{I}_{(-\infty,t]}(x)$ then $\mathbb{P}_n f = F_n$ and $Pf = F$. As we said before F_n is a stochastic process. Since every sample path is cadlag the stochastic process F_n can be viewed as the random function $F_n : \Omega \mapsto D[a,b]$, where $[a,b] \subseteq \mathbb{R}$. Next, we will state a few important results regarding empirical distributions.

Lemma 2.50. ***Bahadur's Theorem****, (Serfling, 1981, pp. 91-92), Let* $p \in (0,1)$.

Suppose F is twice differentiable at $F^{-1}(p)$, with $F'\left(F^{-1}(p)\right) = f\left(F^{-1}(p)\right) > 0$. Then,

$$F_n^{-1}(p) = F^{-1}(p) + \frac{p - F_n\left(F^{-1}(p)\right)}{f\left(F^{-1}(p)\right)} + R_n, \tag{2.2}$$

where with probability one

$$R_n = O\left(n^{-3/4}(\log n)^{3/4}\right), \quad n \to \infty. \tag{2.3}$$

Let $R_n^* = \sup_{p\in(0,1)} f\left(F^{-1}(p)\right)|R_n(p)|$.

Lemma 2.51. ***Kiefer's Theorem****, (Serfling, 1981, p. 101), With probability one*

$$\lim_{n\to\infty} P\left(\frac{n^{3/4}R_n^*}{(\log n)^{1/2}} \leqslant z\right) = 1 - 2\sum_{j=1}^{\infty}(-1)^{j+1}e^{-2j^2z^4}, \quad z > 0. \tag{2.4}$$

Remark 2.52. Notice that (2.4) implies that $R_n^* = O_p\left(n^{-3/4}(\log n)^{1/2}\right)$, as $n \to \infty$.

Lemma 2.53. *(Serfling, 1981, p.283), For $p \in (0,1)$, $\delta \in (0,1/2)$*

$$\sup_{p\in(0,1)}\left|\frac{F_n(p) - F(p)}{[p(1-p)]^{\delta}}\right| = O_p\left(n^{-1/2}\right). \tag{2.5}$$

Lemma 2.54. *(Remark 1(i), Wellner, 1978, p.75), Let U_n be the uniform empirical distribution function. For all $\lambda \geqslant 1$*

$$P\left(\sup_{p\in[0,1]}\frac{U_n(t)}{t} \geqslant \lambda\right) = P\left(\sup_{p\in[1/n,1]}\frac{t}{U_n^{-1}(t)} \geqslant \lambda\right) \leqslant e\lambda^{-1}. \tag{2.6}$$

Proposition 2.55. *(Serfling, 1981, p. 91), Let $X_1, X_2, \ldots, X_n$ be a random sample* *from a standard normal distribution. Let $X_{n:1} \leqslant X_{n:2} \leqslant \ldots \leqslant X_{n:n}$ be the order statistics. Then*

$$P\left(\lim_{n\to\infty} \frac{X_{n:n}}{(2\log n)^{(1/2)}} = 1\right) = 1. \tag{2.7}$$

We will introduce next the *"uniform"* or *"functional"* extensions of the law of large numbers and central limit theorem.

Definition 2.56. The class $\mathcal{F}$ is called *P-Glivenko-Cantelli* if

$$\|\mathbb{P}_n f - Pf\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\mathbb{P}_n f - Pf| \longrightarrow 0, \quad a.s.^*.$$

Theorem 2.57. ***Glivenko-Cantelli**, If $X_1, X_2, \ldots$ are independently and identically distributed random variables with distribution function F then $\|F_n - F\|_\infty \to 0$ a.s.*

Definition 2.58. The *empirical process* evaluated at f is defined as

$$\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf).$$

Definition 2.59. The class $\mathcal{F}$ is called *P-Donsker* if the sequence of processes $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ converges to $\mathbb{G}_P$, a tight limit process in the space $l^\infty(\mathcal{F})$.

Remark 2.60. The limit process $\mathbb{G}_P$, also called a *P-Brownian bridge*, is a Gaussian process with mean zero and covariance structure given by

$$E\mathbb{G}_P f \mathbb{G}_P g = Pfg - PfPg. \tag{2.8}$$

If the functions f are of the form $I_{(-\infty,t]}(x)$ then the limit will be denoted by $\mathbb{G}_F$ and called an *F-Brownian bridge.*

Lemma 2.61. *(Theorem 2.10.1, van der Vaart and Wellner, 1996, p 190), If $\mathcal{F}$ is Donsker and $\mathcal{G} \subset \mathcal{F}$, then $\mathcal{G}$ is Donsker.*

Theorem 2.62. *(Theorem 19.3, van der Vaart 1998, p. 266), If $X_1, X_2, \cdots$ are i.i.d random variables with distribution function F, then the sequence of empirical processes $\sqrt{n}(F_n - F)$ converges in distribution in the space $D[-\infty, \infty]$ to a tight random element $\mathbb{G}_F$ whose marginal distributions are zero-mean normal with covariance function*

$$E\mathbb{G}_F(t_i)\mathbb{G}_F(t_j) = F(t_i \wedge t_j) - F(t_i)F(t_j). \tag{2.9}$$

A class of functions can be Glivenko-Cantelli or Donsker depending on its "size", which can be measured in terms of entropy. The two entropy measures used are the *entropy with bracketing* and the *uniform entropy integral*, of which, the later one will be discussed in more detail . Using the entropy with bracketing the following lemma can be shown.

Lemma 2.63. *(Example 19.12, van der Vaart 1998, p. 273), Let $w : (0, 1) \mapsto \mathbb{R}^+$ be a fixed, continuous function. Let $t \mapsto \mathbb{G}_n^w(t) = \sqrt{n}\,(F_n - F)\,(t)w(F(t))$ be the* weighted empirical process *of a sample of real-values observations. If the weight function w is monotone around 0 and 1 and satisfies $\int_0^1 w^2(s)ds < \infty$, then the weighted empirical process converges weakly in $l^\infty(-\infty, \infty)$ to a tight Gaussian process.*

Definition 2.64. The *covering number* $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ is the minimal number of balls $\{g : \|g - f\| < \varepsilon\}$ of radius ε needed to cover the set $\mathcal{F}$. The centers of the balls need not to belong to $\mathcal{F}$, but they should have finite norms. The *entropy (without bracketing)* is the logarithm of the covering number.

Definition 2.65. The *uniform covering numbers* (relative to L_r) are defined as $\sup_Q N(\varepsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q))$, where, F is a given envelope function, the supremum is over all probability measures Q, with $0 < QF^r < \infty$, and $\|f\|_{Q,r} = (\int |f|^r)^{1/r}$. The *uniform entropy integral* is defined as

$$J(\delta, \mathcal{F}, L_2) = \int_0^\delta \sqrt{log \sup_Q N(\varepsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q))} d\varepsilon.$$

Lemma 2.66. *(Theorem 19.14, van der Vaart 1998, p. 274), Let $\mathcal{F}$ be suitably measurable class of measurable functions with $J(1, \mathcal{F}, L_2) < \infty$. If $P^*F^2 < \infty$, where P^* is the outer probability, then $\mathcal{F}$ is P-Donsker.*

Next, we will introduce the *Vapnik-Cervonenkis* (VC) classes of functions and related results. These classes of functions are very important because it is shown that, under certain conditions, they are Donsker classes.

Let $\mathcal{C}$ be a collection of subsets of a set $\mathcal{X}$. We say that $\mathcal{X}$ *picks out* a certain subset from $\{x_1, \ldots, x_n\}$ if this can be formed as a set of the form $C \cap \{x_1, \ldots, x_n\}$. The collection $\mathcal{C}$ is said to *shatter* $\{x_1, \ldots, x_n\}$ if each of its 2^n subsets can be picked out.

Definition 2.67. The *VC-index* $V(\mathcal{C})$ of the class $\mathcal{C}$ is the smallest n for which no set of size n is shattered by $\mathcal{C}$. The collection $\mathcal{C}$ is called a *VC-class* if its index is finite.

Definition 2.68. The *subgraph* of a function $f : \mathcal{X} \mapsto \mathbb{R}$ is the subset of $\mathcal{X} \times \mathbb{R}$ given by $\{(x,t) : \quad t < f(x)\}$.

Definition 2.69. A collection $\mathcal{F}$ of measurable functions on a sample space is called a *VC-subgraph class* if the collection of all subgraphs of the functions in $\mathcal{F}$ form a VC-class of sets (in $\mathcal{X} \times \mathbb{R}$).

Lemma 2.70. *(Lemma 19.15, van der Vaart 1998, p. 275), There exists a universal constant K such that for any VC-subgraph class $\mathcal{F}$, any $r \geqslant 1$ and $0 < \varepsilon < 1$,*

$$\sup_Q N(\varepsilon \|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \leqslant KV(\mathcal{F})(16e)^{V(\mathcal{F})}(1/\varepsilon)^{r(V(\mathcal{F})-1)}.$$

Remark 2.71. Based on the upper bound obtained in Lemma 2.70, it can be shown that $J(1, \mathcal{F}, L_2) < \infty$. Thus, according to Lemma 2.66, VC-subgraph classes are Q-Donsker classes if they are "suitably measurable" and $P^*F^2 < \infty$, where F is a given envelope of the class of functions.

Lemma 2.72. *(Example 19.17, van der Vaart 1998, p. 276), Let $\mathcal{F}$ be all linear combinations of $\sum \lambda_i f_i$ of a given finite set of functions $f_1, \ldots, f_k$ on $\mathcal{X}$. Then, $\mathcal{F}$ is a VC-subgraph class and hence has a finite uniform entropy integral. Furthermore, the same is true for the class of all sets $\{f > c\}$ if f ranges over $\mathcal{F}$ and c over $\mathbb{R}$.*

Lemma 2.73. *(Lemma 2.6.18, van der Vaart and Wellner, 1996, p. 147), Let $\mathcal{F}$ and $\mathcal{G}$ be VC-subgraph classes on a set $\mathcal{X}$ and $g : \mathcal{X} \mapsto \mathbb{R}$, $\varphi : \mathbb{R} \mapsto \mathbb{R}$ and $\psi : \mathcal{Z} \mapsto \mathcal{X}$ fixed functions. Then,*

1. *$\mathcal{F} \wedge \mathcal{G} = \{f \wedge g : f \in \mathcal{F}, \quad g \in \mathcal{G}\}$ is VC-subgraph;*

2. $\mathcal{F} \vee \mathcal{G}$ *is VC-subgraph;*

3. $\{\mathcal{F} > 0\} = \{\{f > 0\} : f \in \mathcal{F}\}$ *is VC;*

4. $-\mathcal{F}$ *is VC;*

5. $\mathcal{F} + g = \{f + g : f \in \mathcal{F}\}$ *is VC-subgraph;*

6. $\mathcal{F} \cdot g = \{fg : f \in \mathcal{F}\}$ *is VC-subgraph;*

7. $\mathcal{F} \circ \psi = \{f(\psi) : f \in \mathcal{F}\}$ *is VC-subgraph;*

8. $\varphi \circ \mathcal{F}$ *is a VC-subgraph for monotone* φ.

Lemma 2.74. *(Theorem 2.10.6, van der Vaart and Wellner, 1996, p. 192), Let* $\mathcal{F}_1, \ldots, \mathcal{F}_k$ *be Donsker classes with* $\|P\|_{\mathcal{F}_i} < \infty$ *for each* i*. Let* $\varphi : \mathbb{R}^k \mapsto \mathbb{R}$ *satisfy* $|\varphi \circ f(x) - \varphi \circ g(x)|^2 \leqslant \sum_{i=1}^n (f_i(x) - g_i(x))^2$ *for every* $f, g \in \mathcal{F}_1 \times \ldots \mathcal{F}_k$ *and* x*. Then the class* $\varphi \circ (\mathcal{F}_1, \ldots, \mathcal{F}_k)$ *is Donsker, provided* $\varphi \circ (f_1, \ldots, f_k)$ *is integrable for at least one* $(f_1, \ldots, f_k)$.

The following result from van der Vaart (1998) will be essential in finding the asymptotic distribution of the ROC processes.

Lemma 2.75. *(van der Vaart, 1998, p. 281), Let* Θ *be a normed space and* $\mathcal{F}_\delta = \{f_{\theta,t}(x) - f_{\theta_0,t}(x) : \|\theta - \theta_0\| \leqslant \delta, \theta, \theta_0 \in \Theta, t \in \mathbb{R}\}$ *be a P-Donsker class of functions for some* $\delta > 0$*. If*

$$\lim_{\theta \to \theta_0} \sup_{t \in \mathbb{R}} \int (f_{\theta,t}(x) - f_{\theta_0,t}(x))^2 dP(x) \to 0, \quad n \to \infty, \tag{2.10}$$

and

$$\hat{\theta} \xrightarrow{P} \theta_0, \quad as \quad n \to \infty, \tag{2.11}$$

then

$$\sup_{t\in\mathbb{R}} \sqrt{n}(\mathbb{P}_n - P)(f_{\hat{\theta},t}(X) - f_{\theta_0,t}(X)) = o_p(1). \tag{2.12}$$

Remark 2.76. Moreover, one can show that the conclusion of Lemma 2.75 holds with respect to the product probability, when the Donsker class $\mathcal{F}_\delta$ and the estimator $\hat{\theta}$ have different underlying probability spaces. The key result used in this proof is Slutsky's Lemma. Also, we should mention here that the integral $Pf_{\hat{\theta},t}(X)$ uses a notational abuse and it should be understood as follows

$$Pf_{\hat{\theta},t}(X) = \int f_{\theta,t}(x)dP(x)|_{\theta=\hat{\theta}}.$$

In the one dimensional case, the limit process of the ROC process will be obtained by using the functional delta method, via the chain rule. We will actually use the Hadamard differentiability of the operator $G \circ F^{-1}$, as shown in Reeds (1976), Fernholz (1983), Beirlant and Deheuvels (1990), Dudley and Norvaisa (1999), or van der Vaart and Wellner (1996).

Definition 2.77. Let $\mathbb{D}$ and $\mathbb{E}$ be normed spaces. A map $\varphi : \mathbb{D}_\varphi \subset \mathbb{D} \mapsto \mathbb{E}$ is called *Hadamard differentiable* at $\theta \in \mathbb{D}_\varphi$ if there is a continuous linear map $\varphi'_\theta : \mathbb{D} \mapsto \mathbb{E}$ such that

$$\left\| \frac{\varphi(\theta + th_t) - \varphi(\theta)}{t} - \varphi'_\theta(h) \right\|_E \to 0, \tag{2.13}$$

as $t \downarrow 0$, every $h_t \mapsto h$ such that $\theta + th_t \in \mathbb{D}_\varphi$. If $h \in \mathbb{D}_0 \subset \mathbb{D}$ then φ is called *Hadamard differentiable tangentially to* $\mathbb{D}_0$ and φ'_θ is defined on $\mathbb{D}_0$.

Theorem 2.78. ***Chain rule***, *(Lemma 3.9.3, van der Vaart and Wellner, 1996, p. 373), If $\varphi : \mathbb{D}_\varphi \subset \mathbb{D} \mapsto \mathbb{E}_\psi$ is Hadamard differentiable at $\theta \subset \mathbb{D}_\varphi$ tangentially to $\mathbb{D}_0$ and $\psi : \mathbb{E}_\psi \mapsto \mathbb{F}$ is Hadamard differentiable at $\varphi(\theta)$ tangentially to $\varphi'_\theta(\mathbb{D}_0)$ then $\psi \circ \varphi : \mathbb{D}_\varphi \mapsto \mathbb{F}$ is Hadamard differentiable at θ tangentially to $\mathbb{D}_0$ with derivative $\psi'_{\varphi(\theta)} \circ \varphi'_\theta$.*

Theorem 2.79. ***Functional Delta Method***, *(Theorem 3.94, van der Vaart and Wellner, 1996, p. 374), Let $\mathbb{D}$, $\mathbb{E}$ be metrizable topological vector spaces. Let $\varphi : \mathbb{D}_\varphi \subset \mathbb{D} \mapsto \mathbb{E}$ be Hadamard differentiable at θ tangentially to $\mathbb{D}_0$. Let $X_n : \Omega_n \mapsto \mathbb{D}_\varphi$ be maps with $r_n (X_n - \theta) \rightsquigarrow X$ for some sequence of constants $r_n \to \infty$ where X is separable and takes its values in $\mathbb{D}_0$. Then $r_n (\varphi(X_n) - \varphi(\theta)) \rightsquigarrow \varphi'_\theta(X)$. If φ'_θ is defined and continuous on the whole space $\mathbb{D}$ then the sequence $r_n (\varphi(X_n) - \varphi(\theta)) - \varphi'_\theta (r_n (X_n - \theta))$ converges to zero in probability.*

Lemma 2.80. *(Lemma 3.9.23 (ii), van der Vaart and Wellner, 1996, p 386), Let F have compact support $[a, b]$ and be continuously differentiable on its support with strictly positive derivative f. Then the inverse map $A \mapsto A^{-1}$ as a map $\mathbb{D}_2 \subset D[a, b] \mapsto l^\infty(0, 1)$ is Hadamard differentiable at F tangentially to $C[a, b]$. The derivative is given by*

$$\varphi'_F(\alpha) = -(\alpha/f) \circ F^{-1}. \tag{2.14}$$

Lemma 2.81. *(Lemma 3.9.25, van der Vaart and Wellner, 1996, p 388), Let $g : (a, b) \subset \mathbb{R} \mapsto \mathbb{R}$ be differentiable with uniformly continuous and bounded derivative* *and let $\mathbb{D}_\varphi = \{A \in l^\infty(\mathcal{X}) : a < A < b\}$. Then the map $A \mapsto g \circ A$ is Hadamard differentiable as a map $\mathbb{D}_\varphi \subset l^\infty(\mathcal{X}) \mapsto l^\infty(\mathcal{X})$ at every $A \in \mathbb{D}_\varphi$. The derivative is given by*

$$\varphi_A^{'}(\alpha) = g^{'}(A(x))(\alpha(x)). \tag{2.15}$$

Finally, we will state Pollard's Theorem.

Definition 2.82. Let T be a separable metric space and $\mathcal{F} = \{f(\,\cdot\,,t) : t \in T\}$ be a class indexed by T. The class $\mathcal{F}$ is called permissible if it can be indexed by a T in such a way that

1. The function $f(\,\cdot\,,\,\cdot\,)$ is $\mathcal{S} \otimes \mathcal{B}(T)$-measurable as a function from $S \otimes T$ into the real line;

2. T is an analytic subset of a compact metric space $\overline{T}$(from which it inherits its metric and borel σ-field).

Let x_n and y_n be two sequences. By $x_n \gg y_n$ we mean $x_n/y_n \to \infty$.

Theorem 2.83. *(Theorem 37, Pollard, 1984, p. 34), For each n, let $\mathcal{F}_n$ be a permissible class of functions whose covering numbers satisfy*

$$\sup_Q N(\varepsilon, \mathcal{F}_n, L_1(Q)) \leqslant A\varepsilon^{-W} \quad for \quad 0 < \varepsilon < 1, \tag{2.16}$$

with constants A and W not depending on n. Let $\{\alpha_n\}$ be a non-increasing sequence of positive numbers for which $n\delta_n^2\alpha_n^2 \gg \log n$. If $|f| \leqslant 1$ and $(Pf^2)^{1/2} \leqslant \delta_n$ for each f

in $\mathcal{F}_n$ *then*

$$\sup_{\mathcal{F}_n} |\mathbb{P}_n f - Pf| \ll \delta_n^2 \alpha_n \quad almost \quad surely. \tag{2.17}$$

CHAPTER 3: ASYMPTOTIC DISTRIBUTION OF ROC PROCESS

3.1 Notation and Problem Set-up

Let $(\Omega_1, \mathcal{U}_1, \mathrm{P})$ and $(\Omega_2, \mathcal{U}_2, \mathrm{Q})$ be two probability spaces. Let $X : \Omega_1 \mapsto \mathbb{R}$ and $Y : \Omega_2 \mapsto \mathbb{R}$ be two independent random variables that represent the diagnostic tests of healthy and diseased subjects, respectively. Denote by P and F the probability distribution and distribution function induced by X. Similarly, let Q and G denote the probability distribution and distribution function induced by Y. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be two mutually independent random samples from distributions P and Q, respectively. Assume that n and m satisfy condition $m/n \to \lambda \in \mathbb{R}^+$, as $n \to \infty$.

Our goal in this chapter is to find the asymptotic properties of the nonparametric estimator of ROC curve. Recall, from the first chapter, that this estimator is the empirical ROC curve given by $EROC(p) = 1 - G_m(F_n^{-1}(1-t))$, where $t \in (0,1)$ and G_m, F_n are the empirical distribution functions. Therefore, it will be sufficient to focus our attention on the following empirical process

$$\sqrt{m}\left(G_m(F_n^{-1}(p)) - G(F^{-1}(p))\right), \qquad p \in (0,1). \tag{3.1}$$

An equivalent form of the process in (3.1) is given by

$$\sqrt{m}\left(m^{-1}\sum_{j=1}^{m} \mathbf{I}[Y_j \leqslant F_n^{-1}(p)] - G(F^{-1}(p))\right), \qquad p \in (0,1), \tag{3.2}$$

where $\mathbf{I}[A]$ is the indicator function of the event A. The process introduced in (3.1) will be called *empirical ROC process* or, shortly, EROC process. The step function $G_m(F_n^{-1}(p,\omega_1),\omega_2) = m^{-1}\sum_{j=1}^{m}\mathbf{I}[Y_j(\omega_2) \leqslant F_n^{-1}(p,\omega_1)]$ will be called *empirical ROC curve.*

Remark 3.1. Since this is a two sample problem the underlying and induced probability spaces of the random vector (X,Y) are given by $(\Omega_1 \times \Omega_2, \mathcal{U}_1 \otimes \mathcal{U}_2, \mathrm{P} \otimes \mathrm{Q})$ and $(\mathbb{R} \times \mathbb{R}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}), P \otimes Q)$, respectively. However, the notation can easily get complicated if we would like to keep track of the right probability spaces. Hence, whenever is possible, we will work with the marginals and, when deemed necessary, we will provide further clarifications.

Now, we briefly describe how we will proceed to find the asymptotic distribution of the empirical ROC process. First, by Remark 2.33 we will construct independent uniformly $[0,1]$ distributed random variables $\{U_i\}_{i=\overline{1,n}}$ and $\{V_j\}_{j=\overline{1,m}}$ such that (2.1) holds for both random samples. Let U_n and V_m be the empirical distribution functions of the corresponding random samples. Then, we will show that the EROC process defined in (3.1) is equivalent, with probability one, to the process

$$\sqrt{m}\left(\tilde{G}_m(U_n^{-1}(p)) - \tilde{G}(p)\right), \qquad p \in (0,1), \tag{3.3}$$

where $\tilde{G} = G \circ F^{-1}$ and $\tilde{G}_m$ is the empirical distribution function of a random sample $\{Z_j\}_{j=\overline{1,m}}$ with distribution function $\tilde{G}$. This construction will ease our future work by avoiding some technical difficulties that appear in the general case and it will allow us to find the asymptotic distribution on the interval $(0,1)$ by using the functional delta method. Note that Hsieh and Turnbull (1996) obtained the same result on the interval $[a, b] \subset (0, 1)$ using strong approximation theory. Second, starting from the representation process in (3.3), we will construct a Donsker class of functions. Then, by applying Lemma 2.75 to the previous class of functions, we will decompose this process into a sum of two other processes. Third, we will find the asymptotic distribution of each of the decomposed processes. Due to independence of the random samples the asymptotic distribution of this process will be the sum of the asymptotic distributions found before.

Lemma 3.2. *Let F, G be any two distribution functions. Then,*

$$(F \circ G)^{-1} = G^{-1} \circ F^{-1} \tag{3.4}$$

Proof. By definition, $(F \circ G)^{-1}(p) = inf\,\{x : F(G(x)) \geqslant p\}$. But, by Lemma 2.31(3) $F(G(x)) \geqslant p$ iff $G(x) \geqslant F^{-1}(x)$. Hence, $(F \circ G)^{-1}(p) = inf\,\{x : G(x) \geqslant F^{-1}(p)\} = G^{-1} \circ F^{-1}(p)$. □

By Remark 2.33 we can construct the independent and identically Uniform(0,1) distributed random variables U_i such that

$$\mathrm{P}\left(X_i = F^{-1}(U_i)\right) = 1, \quad i = \overline{1, n}. \tag{3.5}$$

Denote by U the uniform distribution on $[0, 1]$ and let $U_n = n^{-1} \sum \mathbf{I}[U_i \leqslant p]$ be the empirical distribution function of random sample $\{U_i\}_{i=\overline{1,n}}$. Notice that since $U(p) = U^{-1}(p) = p$ for any $p \in [0, 1]$, we can conveniently use p instead $U(p)$ or

$U^{-1}(p)$ and vice versa.

Lemma 3.3. *Let F be any distribution function. Then, there exists $\{U_i\}_{i=\overline{1,n}}$, a random sample from Uniform(0,1) distribution, such that (3.5) holds and for any $x \in \mathbb{R}$*

$$F_n(x) = U_n(F(x)), \quad a.s. \tag{3.6}$$

Proof. By (3.5) we have

$$F_n(x) = n^{-1} \sum \mathbf{I}[X_i \leqslant x] = n^{-1} \sum \mathbf{I}[F^{-1}(U_i) \leqslant x], \quad a.s.$$

But, by Lemma 2.31(3) we have

$$n^{-1} \sum \mathbf{I}[F^{-1}(U_i) \leqslant x] = n^{-1} \sum \mathbf{I}[U_i \leqslant F(x)] = U_n(F(x)).$$

□

Lemma 3.4. *Let F be any distribution function. Then, there exists $\{U_i\}_{i=\overline{1,n}}$, a random sample from Uniform(0,1) distribution, such that (3.5) holds and for any $p \in (0,1)$*

$$F_n^{-1}(p) = F^{-1}(U_n^{-1}(p)), \quad a.s. \tag{3.7}$$

Proof. Follows immediately from Lemma 3.3 and Lemma 3.2 □

By analogy, we can construct the independent and identically Uniform(0,1) distributed random variables V_j such that

$$\mathrm{Q}\left(Y_j = G^{-1}(V_j)\right) = 1, \quad j = \overline{1, m}. \tag{3.8}$$

Similarly, if we let $V_m(p) = m^{-1} \sum \mathbf{I}[V_j \leqslant p]$, for $p \in (0, 1)$ be the empirical distribution function of random sample $\{V_j\}_{j=\overline{1,m}}$ we can prove that, for any $y \in \mathbb{R}$,

$$G_m(y) = V_m(G(y)), \quad a.s. \tag{3.9}$$

Lemma 3.5. *Let $\{U_i\}_{i=\overline{1,n}}$ and $\{V_j\}_{j=\overline{1,m}}$ be two mutually independent random samples from a Uniform(0,1) distribution that satisfy (3.5) and (3.8), respectively. If F is strictly increasing and G any distribution function then, for any $p \in (0, 1)$*

$$G_m(F_n^{-1}(p)) = \tilde{G}_m(U_n^{-1}(p)), \quad a.s, \tag{3.10}$$

where $\tilde{G}_m$ is the empirical distribution function of a random sample with distribution function $\tilde{G} = G \circ F^{-1}$.

Proof. By (3.7) and (3.9) we have

$$G_m(F_n^{-1}(p)) = V_m(G(F^{-1}(U_n^{-1}(p)))), \quad a.s. \tag{3.11}$$

Notice that by definition and Lemma 2.31(3) we have

$$V_m(G(F^{-1}(p))) = m^{-1}\sum \mathbf{I}[V_j \leqslant G(F^{-1}(p))] = m^{-1}\sum \mathbf{I}[(G(F^{-1})^{-1}(V_j) \leqslant p]. \tag{3.12}$$

But, if we denote $(G \circ F^{-1})^{-1}(V_j) = Z_j$ for $j = \overline{1,m}$ then, by Lemma 2.34, $\{Z_j\}_{j=\overline{1,m}}$ is a random sample from distribution function $\tilde{G} = G \circ F^{-1}$. The proof is complete by letting $\tilde{G}_m = V_m \circ G \circ F^{-1}$. We should remark though, that the almost sure equality refers to a set $A_1 \times A_2 \in \mathcal{U}_1 \otimes \mathcal{U}_2$ such that $Pr(A_1 \times A_2) = \mathrm{P} \otimes \mathrm{Q}(A_1 \times A_2) = 1$. $\square$

Corollary 3.6. *Let $\{U_i\}_{i=\overline{1,n}}$ and $\{V_j\}_{j=\overline{1,m}}$ be two mutually independent random samples from a Uniform(0,1) distribution that satisfy (3.5) and (3.8), respectively. If F is strictly increasing then, for any $p \in (0,1)$*

$$\sqrt{m}\left(G_m(F_n^{-1}(p)) - G(F^{-1}(p))\right) = \sqrt{m}\left(\tilde{G}_m(U_n^{-1}(p)) - \tilde{G}(p)\right), \quad a.s. \tag{3.13}$$

Proof. Immediate from Lemma 3.5. $\square$

3.2 Decomposition of the Equivalent Empirical ROC Process

As we said before, we will construct a Donsker class of functions and then, by applying Lemma 2.75 to the previous class of functions, we will decompose this process into a sum of two other processes. Let $\tilde{Q}$ be the probability distribution associated with distribution function $\tilde{G}$. Let f_p and $f_{H,p}$ be real functions defined on $[0,1]$ such that $f_p(z) = \mathbf{I}[z \leqslant p]$ and $f_{H,p}(z) = \mathbf{I}[z \leqslant H^{-1}(p)]$, where H is a distribution function on $[0,1]$. Let $\mathcal{H}$ be the class of all distribution functions H defined on $[0,1]$. Construct the following classes of functions:

$$\mathcal{F}_0 = \{f_p : p \in (0,1),\}, \tag{3.14}$$

and

$$\mathcal{F}^{'} = \{f_{H,p} : p \in (0,1), H \in \mathcal{H}\}. \tag{3.15}$$

Lemma 3.7. *The class of measurable functions $\mathcal{F}^{'}$ given in (3.15) is a $\tilde{Q}$-Donsker class.*

Proof. Notice that $H^{-1}(p) \in (0,1)$, for all $H \in \mathcal{H}$ and any $p \in (0,1)$. Thus, $\mathcal{F}^{'} \subset \mathcal{F}_0$ where $\mathcal{F}_0$ is the class defined in (3.14). Since the collection of the segments $(0,t]$, has $V(\mathcal{C}) = 2$ (see Example 19.16, van der Vaart, 1996, p. 276), then $\mathcal{F}_0$ is a VC-subgraph class. Hence, $\mathcal{F}^{'}$ is a VC-subgraph class, too. Let $E \equiv 1$ be an envelope of this class. Since E is a bounded, square integrable and measurable envelope then, according to Remark 2.71, $\mathcal{F}^{'}$ is a $\tilde{Q}$-Donsker class of functions. □

Let $\mathcal{F}$ be the following class of functions

$$\mathcal{F} = \{f_{H,p} - f_p : p \in (0,1), H \in \mathcal{H}\}. \tag{3.16}$$

Lemma 3.8. *The class of measurable functions $\mathcal{F}$ given in (3.16) is a $\tilde{Q}$-Donsker class.*

Proof. Let $\mathcal{F}^{''} = -\mathcal{F}^{'}$. By Lemma 2.73(4), $\mathcal{F}^{''}$ is a VC subgraph class. Notice that $\mathcal{F}^{''}$ has the same envelope E as $\mathcal{F}^{'}$. Thus, $\mathcal{F}^{''}$ is a $\tilde{Q}$-Donsker class. According to Lemma

2.74, $\mathcal{F}' + \mathcal{F}'' = \{\mathbf{I}[z \leqslant H_1^{-1}(p_1)] - \mathbf{I}[z \leqslant H_2^{-1}(p_2)], z \in [0,1], p_1, p_2 \in (0,1), H_1, H_2 \in \mathcal{H}\}$ is a $\tilde{Q}$-Donsker class. Since $\mathcal{F} \subset \mathcal{F}' + \mathcal{F}''$, then, according to Lemma 2.61, $\mathcal{F}$ is a $\tilde{Q}$-Donsker class of functions. □

Lemma 3.9. *Let $\mathcal{F}$ be the class of functions given in (3.16). Then, for any function in $\mathcal{F}$ we have*

$$\sup_{p\in(0,1)} \int (f_{H,p}(z) - f_p(z))^2 d\tilde{Q}(z) = \sup_{p\in(0,1)} \left|\tilde{G}(H^{-1}(p)) - \tilde{G}(p)\right|. \tag{3.17}$$

Proof. Let H be any distribution function in $\mathcal{H}$, p be any fixed value in $(0,1)$. Notice that the integrand on the left-hand side of (3.17) can take a value different from zero only when the inequalities under the indicator functions are in opposite sense. Thus,

$$\begin{aligned}
&\int (\mathbf{I}[z \leqslant H^{-1}(p)] - \mathbf{I}[z \leqslant p])^2 d\tilde{Q}(z) \\
&= \int \mathbf{I}[z \leqslant H^{-1}(p), z > p] d\tilde{Q}(z) + \int \mathbf{I}[z > H^{-1}(p), z \leqslant p] d\tilde{Q}(z) \\
&= \int \mathbf{I}[p < z \leqslant H^{-1}(p)] d\tilde{Q}(z) + \int \mathbf{I}[H^{-1}(p) < z \leqslant p] d\tilde{Q}(z)
\end{aligned} \tag{3.18}$$

Next, consider the following cases.

1. If $p = H^{-1}(p)$ then $\int (f_{H,p}(z) - f_p(z))^2 d\tilde{Q}(z) = 0$.

2. If $p < H^{-1}(p)$ then the second integral from (3.18) is zero and thus,

$$\int (f_{H,p}(z) - f_p(z))^2 d\tilde{Q}(z) = \tilde{Q}(p < Z \leqslant H^{-1}(p)) = \tilde{G}(H^{-1}(p)) - \tilde{G}(p)$$

3. Similarly, if $p > H^{-1}(p)$ then,

$$\int (f_{H,p}(z) - f_p(z))^2 d\tilde{Q}(z) = \tilde{Q}(H^{-1}(p) < Z \leqslant p) = \tilde{G}(p) - \tilde{G}(H^{-1}(p))$$

Therefore, for any $H \in \mathcal{H}$ and any $p \in (0,1)$ we obtain

$$\int (f_{H,p}(z) - f_p(z))^2 d\tilde{Q}(z) = \left|\tilde{G}(H^{-1}(p)) - \tilde{G}(p)\right| \tag{3.19}$$

The conclusion follows by taking supremum in (3.19). □

Lemma 3.10. *Let $\tilde{G}$ be any continuous distribution function. Then, for any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that if $\sup_{t\in(0,1)} |H(t) - t| < \delta_\varepsilon$, where $H \in \mathcal{H}$, we have*

$$\sup_{p\in(0,1)} \left|\tilde{G}(H^{-1}(p)) - \tilde{G}(p)\right| < \varepsilon. \tag{3.20}$$

Proof. Let ε be any given positive value. Let δ be a positive value and $H \in \mathcal{H}$ such that the condition $\sup_{t\in(0,1)} |H(t) - t| < \delta$ is satisfied or, equivalently,

$$-\delta < H(t) - t < \delta, \tag{3.21}$$

for all $t \in (0,1)$. Next, we will show that we can find δ as a function of the given ε such that, from (3.21), we get (3.20).

Let p be any point in $(0,1)$. Then,

$$H^{-1}(p) = \inf\{t : H(t) \geqslant p\} = \inf\{t : U(t) \geqslant p - (H(t) - t)\}. \tag{3.22}$$

By plugging (3.21) into (3.22) we obtain the inequality

$$p - \delta = U^{-1}(p - \delta) \leqslant H^{-1}(p) \leqslant U^{-1}(p + \delta) = p + \delta. \tag{3.23}$$

Since $\tilde{G}$ is continuous, for any given $\varepsilon > 0$, there exists a finite partition $0 = p_0 < p_1 < p_2 < \ldots < p_k < p_{k+1} = 1$ such that

$$\tilde{G}(p_i) - \tilde{G}(p_{i-1}) < \varepsilon/2, \quad i = \overline{1, k+1}. \tag{3.24}$$

Notice, that (3.24) becomes $\tilde{G}(p_1) < \varepsilon/2$ and $1 - \tilde{G}(p_k) < \varepsilon/2$ for $i = 1$ and $i = k+1$, respectively. Let $0 < \delta_\varepsilon < \min_{i=\overline{1,k+1}} \{p_i - p_{i-1}\}$, where δ is a function of ε since the finite partition is determined by ε. Since $p \in (0, 1)$ there exists an i such that $p_{i-1} \leqslant p \leqslant p_i$. Consider the following cases.

1. Let $i = \overline{2, k}$. Then, (3.23) becomes

$$p_{i-2} \leqslant p_{i-1} - \delta \leqslant H^{-1}(p) \leqslant p_i + \delta \leqslant p_{i+1}. \tag{3.25}$$

   Hence, by monotonicity of distribution functions, (3.25) and (3.24) we obtain

$$\begin{aligned} \tilde{G}(H^{-1}(p)) - \tilde{G}(p) &\leqslant \tilde{G}(p_{i+1}) - \tilde{G}(p_{i-1}) < \varepsilon; \\ \tilde{G}(H^{-1}(p)) - \tilde{G}(p) &\geqslant \tilde{G}(p_{i-2})) - \tilde{G}(p_{i-1}) > -\varepsilon. \end{aligned}$$

   Thus, $\sup_{t \in (0,1)} |H(t) - t| < \delta_\varepsilon$ implies that (3.20) is true for all $p \in [p_1, p_k]$.

2. For $i = k+1$ by (3.25) we have $p_{k-1} < H^{-1}(p) \leqslant 1$. Then, by the same arguments as in the first case we obtain

$$\tilde{G}(H^{-1}(p)) - \tilde{G}(p) \leqslant 1 - \tilde{G}(p_k) < \varepsilon/2.$$

   Therefore, (3.20) becomes true for all $p \in [p_k, 1)$.

3. The case $i = 1$ is analogous to the second case.

In conclusion, given $\varepsilon > 0$ we found $\delta_\varepsilon < \min_{i=\overline{1,k+1}} \{p_i - p_{i-1}\}$ such that (3.20) is true for all $p \in (0,1)$ or, equivalently,

$$\sup_{p\in(0,1)} \left| \tilde{G}(H^{-1}(p)) - \tilde{G}(p) \right| < \varepsilon. \tag{3.26}$$

Hence, the lemma is proved. □

Lemma 3.11. *Let $\{U_i\}_{i=\overline{1,n}}$ be a random sample from a Uniform(0,1) distribution such that (3.5) holds and $\{Z_j\}_{j=\overline{1,m}}$ a random sample from distribution function $\tilde{G}$ constructed as in Lemma 3.5. If we assume that F is strictly increasing and G is continuous then,*

$$\begin{aligned}
&\sqrt{m}\left(\tilde{G}_m(U_n^{-1}(p)) - \tilde{G}(p)\right) \\
&= \sqrt{m}\left(\tilde{G}_m(p) - \tilde{G}(p)\right) && (3.27)\\
&\quad + \sqrt{m}\left(\tilde{G}(U_n^{-1}(p)) - \tilde{G}(p)\right) && (3.28)\\
&\quad + o_p(1),
\end{aligned}$$

where $o_p(1)$ *holds uniformly in* $p \in (0,1)$.

Proof. Let $\mathcal{F}$ be the $\tilde{Q}$-Donsker class of functions introduced in Lemma 3.8, U be the $Unif(0,1)$ distribution function and $\{f_{U_n,p} - f_p\}_{n\in\mathbb{N}}$ be a sequence of random functions that takes its values in $\mathcal{F}$. For some $\delta > 0$, define

$$\mathcal{F}_\delta = \{f_{H,p} - f_p \in \mathcal{F} : p \in (0,1), H \in \mathcal{H}, \|H - U\|_\infty \leqslant \delta\},$$

to be a subclass of $\mathcal{F}$. Then, $\mathcal{F}_\delta$ is a $\tilde{Q}$-Donsker class by Lemma 2.61.

First, we will prove

$$\sup_{p\in(0,1)} \sqrt{m}\left|(\tilde{\mathbb{Q}}_m - \tilde{Q})(f_{U_n,p} - f_p)\right| = o_p(1), \quad n, m \to \infty \tag{3.29}$$

or, equivalently, for any $\varepsilon > 0$

$$Pr\left(\sup_{p\in(0,1)} \sqrt{m}\left|(\tilde{\mathbb{Q}}_m - \tilde{Q})(f_{U_n,p} - f_p)\right| > \varepsilon\right) \to 0, \quad n, m \to \infty,$$

where Pr should be understood as $\mathrm{P} \otimes \mathrm{Q}$. Then,

$$\begin{aligned} &Pr\left(\sup_{p\in(0,1)} \sqrt{m}\left|(\tilde{\mathbb{Q}}_m - \tilde{Q})(f_{U_n,p} - f_p)\right| > \varepsilon\right) \\ &\leqslant \; Pr\left(\|U_n - U\|_\infty > \delta\right) &(3.30) \\ &\quad + Pr\left(\sup_{p\in(0,1)} \sqrt{m}\left|(\tilde{\mathbb{Q}}_m - \tilde{Q})(f_{U_n,p} - f_p)\right| > \varepsilon, \|U_n - U\|_\infty \leqslant \delta\right). &(3.31) \end{aligned}$$

Now, we will show that the probability in (3.31) converges to 0 as $n \to \infty$, by applying Lemma 2.75 to $\tilde{Q}$-Donsker class $\mathcal{F}_\delta$. Since $\tilde{G} = G \circ F^{-1}$ is a continuous distribution function, we can set δ to be equal to δ_ε constructed in Lemma 3.10 such that condition (3.20) is satisfied. By (3.17) and the definition of a limit, from (3.20) we have

$$\lim_{H\to U} \sup_{p\in(0,1)} \int (f_{H,p}(z) - f_p(z))^2 d\tilde{Q}(z) \to 0, \quad n \to \infty. \tag{3.32}$$

By Glivenko-Cantelli Theorem we have

$$\|U_n - U\|_\infty \stackrel{a.s.}{\to} 0, \tag{3.33}$$

and, thus, the sequence $\{f_{U_n,p} - f_p\}_{n\in\mathbb{N}}$ takes its values in $\mathcal{F}_\delta$, except for a finite number of n's. Now, from (3.32) and (3.33) we can conclude, after applying Lemma 2.75 to class $\mathcal{F}_\delta$ that,

$$\sup_{p\in(0,1)} \sqrt{m} \left|(\tilde{\mathbb{Q}}_m - \tilde{Q})(f_{U_n,p} - f_p)\right| = o_p(1), \quad n, m \to \infty$$

Noting that

$$\begin{aligned}
&\tilde{\mathbb{Q}}_m f_{U_n,p} = \tilde{\mathbb{Q}}_m \mathbf{I}[Z \leqslant U_n^{-1}(p)] = \tilde{G}_m(U_n^{-1}(p));\\
&\tilde{\mathbb{Q}}_m f_p = \tilde{\mathbb{Q}}_m \mathbf{I}[Z \leqslant p] = \tilde{G}_m(p);\\
&\tilde{Q} f_{U_n,p} = \tilde{Q}\mathbf{I}[Z \leqslant U_n^{-1}(p)] = \tilde{G}(U_n^{-1}(p));\\
&\tilde{Q} f_p = \tilde{Q}\mathbf{I}[Z \leqslant p] = \tilde{G}(p),
\end{aligned}$$

the conclusion of the Lemma follows immediately by regrouping terms in (3.29). □

3.3 Asymptotic Distribution of the Component Processes

In this section we will find the asymptotic distribution of each of the decomposed processes.

Lemma 3.12. *Let* $\{Z_j\}_{j=\overline{1,m}}$ *be a random sample from distribution function* $\tilde{G}$ *constructed as in Lemma 3.5. Then, as* $m \to \infty$,

$$\sqrt{m}\left(\tilde{G}_m - \tilde{G}\right) \rightsquigarrow \mathbb{G}_{\tilde{G}}, \qquad in \quad D[0,1]. \tag{3.34}$$

The tight gaussian process, $\mathbb{G}_{\tilde{G}}$*, has mean zero and covariance structure given by*

$$E\mathbb{G}_{\tilde{G}}(p_i)\mathbb{G}_{\tilde{G}}(p_j) = \tilde{G}(p_i \wedge p_j) - \tilde{G}(p_i)\tilde{G}(p_j), \tag{3.35}$$

where $p_i, p_j \in [0,1]$.

Proof. Let $\mathcal{F}_0$ be the class of functions given in (3.14). Then, $\mathcal{F}_0$ is a $\tilde{Q}$-Donsker class since it is a uniformly bounded VC class. Hence, as $m \to \infty$, we have

$$\sqrt{m}\left(\tilde{\mathbb{Q}}_m - \tilde{Q}\right) \rightsquigarrow \mathbb{G}_{\tilde{Q}}, \quad \text{in} \quad l^\infty(\mathcal{F}_0). \tag{3.36}$$

But, since for all $p \in [0,1]$ we can naturally identify f_p with p and, thus, $l^\infty(\mathcal{F}_0)$ with $l^\infty[0,1]$, then (3.36) is equivalent to

$$\sqrt{m}\left(\tilde{G}_m - \tilde{G}\right) \rightsquigarrow \mathbb{G}_{\tilde{G}}, \quad \text{in} \quad l^\infty[0,1]. \tag{3.37}$$

Since the stochastic processes $\mathbb{G}_{\tilde{G}}$ and $\sqrt{m}\left(\tilde{G}_m - \tilde{G}\right)$ take their values in $D[0,1]$, then, according to Lemma 2.44, the weakly convergence in (3.37) is also true in $D[0,1]$. The covariance structure is obtained immediately by plugging $\tilde{Q}$ and f_{p_i}, f_{p_j} in (2.8). □

Lemma 3.13. *Let $\{U_i\}_{i=\overline{1,n}}$ be a random sample from a Uniform(0,1) distribution such that (3.5) holds and $\tilde{G}$ a distribution function constructed as in Lemma 3.5. If we assume that F and G are differentiable distribution functions with strictly positive derivatives f and g, respectively, such that $\tilde{g} = g(F^{-1})/f(F^{-1})$ is uniformly continuous and bounded on $(0,1)$, and $m/n \to \lambda \in \mathbb{R}^+$ as $n \to \infty$ then, as $n, m \to \infty$,*

$$\sqrt{m}\left(\tilde{G}(U_n^{-1}) - \tilde{G}\right) \rightsquigarrow \sqrt{\lambda}\tilde{g}\mathbb{G}_U, \quad in \quad D(0,1), \tag{3.38}$$

where $\mathbb{G}_U$ is the standard Brownian bridge. The covariance structure of the limit process is given by

$$E\sqrt{\lambda}\tilde{g}(p_i)\mathbb{G}_U(p_i)\sqrt{\lambda}\tilde{g}(p_j)\mathbb{G}_U(p_j) = \lambda\tilde{g}(p_i)\tilde{g}(p_j)(p_i \wedge p_j - p_i p_j), \tag{3.39}$$

where $p_i, p_j \in (0,1)$.

Proof. We will apply the Functional Delta Method. First, by Example 21.6, van der Vaart, 1996, p. 308, we have

$$\sqrt{n}(U_n^{-1} - U^{-1}) \rightsquigarrow \mathbb{G}_U, \quad \text{in} \quad l^\infty(0,1), \tag{3.40}$$

where $\mathbb{G}_U$ is the standard Brownian bridge. Next, let $\varphi : \mathbb{D}_\varphi \subset l^\infty(0,1) \mapsto l^\infty(0,1)$ be a map given by $\varphi(A) = \tilde{G} \circ A$, where $\mathbb{D}_\varphi = \{A \in l^\infty(0,1) : 0 < A < 1\}$. Since $\tilde{G}$ is differentiable with uniformly continuous and bounded derivative $\tilde{g}$, then, by Lemma 2.81, the map φ is Hadamard differentiable at every $A \in \mathbb{D}_\varphi$ with derivative given by $\varphi'_A(\alpha) = \tilde{g}(A)(\alpha)$ and $\alpha \in l^\infty(0,1)$. In particular, φ will be Hadamard differentiable at every $A \in \mathbb{D}_\varphi$ tangentially to $C(0,1)$. Hence, since U_n^{-1} and U^{-1} takes their values in $\mathbb{D}_\varphi$ and $\mathbb{G}_U$ is a tight gaussian process in $C(0,1)$, by applying Functional Delta Method to (3.40), we have, as $n, m \to \infty$,

$$\sqrt{n}\left(\varphi(U_n^{-1}) - \varphi(U^{-1})\right) \rightsquigarrow \varphi'_{U^{-1}}(\mathbb{G}_U), \quad \text{in} \quad l^\infty(0,1). \tag{3.41}$$

By using $m/n \to \lambda \in \mathbb{R}^+$, as $n \to \infty$ and plugging the expressions of φ and φ' in (3.41) we have, as $n, m \to \infty$,

$$\sqrt{m}\left(\tilde{G}(U_n^{-1}) - \tilde{G}\right) \rightsquigarrow \sqrt{\lambda}\tilde{g}\mathbb{G}_U, \quad \text{in} \quad l^\infty(0,1). \tag{3.42}$$

Since the processes in (3.42) take their values in $D(0,1)$, (3.38) follows immediately. The covariance structure in (3.39) is obtained by using the covariance of the standard brownian bridge

$$E\mathbb{G}_U(p_i)\mathbb{G}_U(p_j) = p_i \wedge p_j - p_i p_j, \tag{3.43}$$

where $p_i, p_j \in (0,1)$. □

3.4 The Limit of the Empirical ROC Process

Theorem 3.14. *Let* $\{U_i\}_{i=\overline{1,n}}$ *be a random sample from a Uniform(0,1) distribution such that (3.5) holds and* $\{Z_j\}_{j=\overline{1,m}}$ *a random sample from distribution function* $\tilde{G}$ *constructed as in Lemma 3.5. If we assume that* F *and* G *are differentiable distribution functions with strictly positive derivatives* f *and* g*, respectively, such that* $\tilde{g} = g(F^{-1})/f(F^{-1})$ *is uniformly continuous and bounded on* $(0,1)$*, and* $m/n \to \lambda \in \mathbb{R}^+$ *as* $n \to \infty$ *then, as* $n, m \to \infty$*,*

$$\sqrt{m}\left(\tilde{G}_m(U_n^{-1}) - \tilde{G}\right) \rightsquigarrow \mathbb{G}_{\tilde{G}} + \sqrt{\lambda}\tilde{g}\mathbb{G}_U, \quad in \quad D(0,1). \tag{3.44}$$

The covariance structure of the limit process is given by

$$\begin{aligned}
&\mathbb{E}(\mathbb{G}_{\tilde{G}}(p_i) + \sqrt{\lambda}\tilde{g}(p_i)\mathbb{G}_U(p_i))(\mathbb{G}_{\tilde{G}}(p_j) - \sqrt{\lambda}\tilde{g}(p_j)\mathbb{G}_U(p_j)) \\
&= (\tilde{G}(p_i \wedge p_j) - \tilde{G}(p_i)\tilde{G}(p_j)) + \lambda\tilde{g}(p_i)\tilde{g}(p_j)(p_i \wedge p_j - p_i p_j),
\end{aligned} \tag{3.45}$$

where $p_i, p_j \in (0,1)$*.*

Proof. By Lemmas 3.11, 3.12, 3.13, independence of random samples $\{U_i\}_{i=\overline{1,n}}$ and $\{Z_j\}_{j=\overline{1,m}}$ and by Slutsky's Lemma. □

Corollary 3.15. *Let* $\{X_i\}_{i=\overline{1,n}}$ *and* $\{Y_j\}_{j=\overline{1,m}}$ *be mutually independent random sample from distribution functions* F *and* G*, respectively. If we assume that* F *and* G *are differentiable with strictly positive derivatives derivatives* f *and* g*, respectively, such that* $\tilde{g} = g(F^{-1})/f(F^{-1})$ *is uniformly continuous and bounded on* $(0,1)$*, and* $m/n \to$

$\lambda \in \mathbb{R}^+$ as $n \to \infty$ then, as $n, m \to \infty$,

$$\sqrt{m}\left(G_m(F_n^{-1}) - G(F^{-1})\right) \rightsquigarrow \mathbb{G}_{\tilde{G}} + \sqrt{\lambda}\tilde{g}\mathbb{G}_U, \quad in \quad D(0,1). \tag{3.46}$$

The covariance structure of the limit process is given by (3.45).

Proof. Immediate from definition of weak convergence, Corollary 3.6 and Theorem 3.14. □

CHAPTER 4: ASYMPTOTIC DISTRIBUTION OF GENERALIZED ROC PROCESS

4.1 Notation and Problem Set-up

Let $(\Omega_1, \mathcal{U}_1, \mathrm{P})$ and $(\Omega_2, \mathcal{U}_2, \mathrm{Q})$ be two probability spaces and $\mathbf{X} : \Omega_1 \mapsto \mathbb{R}^k$ and $\mathbf{Y} : \Omega_2 \mapsto \mathbb{R}^k$ be two independent random vectors that represent the multiple diagnostic tests of healthy and diseased subjects, respectively. Denote by P and Q the multivariate probability distributions induced by $\mathbf{X}$ and $\mathbf{Y}$, respectively. Assume $\mathbf{X}$ and $\mathbf{Y}$ are distributed multivariate normal with means $\mu_{\mathbf{x}}$ and $\mu_{\mathbf{y}}$, and covariance matrices $\mathbf{\Sigma_x}$ and $\mathbf{\Sigma_y}$, respectively. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m$ be two mutually independent random samples from distributions P and Q, respectively. The vectors $\mathbf{X_i} = (X_{1i}, X_{2i}, \ldots, X_{ki})'$, $i = \overline{1, n}$ and $\mathbf{Y_j} = (Y_{1j}, Y_{2j}, \ldots, Y_{kj})'$, $j = \overline{1, m}$ are the measurements of the i^{th} and j^{th} healthy and diseased subjects, respectively. Assume that n and m satisfy condition $m/n \to \lambda \in \mathbb{R}^+$, as $n \to \infty$.

Definition 4.1. (Su and Liu, 1993, p.1351) A vector $\mathbf{a}_0 \in \mathbb{R}^k$ is called the *best linear combination* under the ROC criterion, if the Area Under the (ROC) Curve, generated by $\mathbf{a}_0'\mathbf{X}$ and $\mathbf{a}_0'\mathbf{Y}$ is the largest among all linear combinations.

Lemma 4.2. *(Theorem 3.1, Su and Liu, 1993, p. 1352) The coefficients for the best linear combination are*

$$\mathbf{a}_0 \propto (\mathbf{\Sigma_x} + \mathbf{\Sigma_y})^{-1}(\mu_{\mathbf{y}} - \mu_{\mathbf{x}}). \tag{4.1}$$

Without loss of generality assume that $\mu_x = 0$ and denote $\mu_y = \mu$. Also, consider the particular case of equal covariance matrices $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y = \mathbf{\Sigma}$. Let the best linear combination under ROC criterion be

$$\mathbf{a_0} = \mathbf{\Sigma^{-1}}\mu \tag{4.2}$$

Su and Liu (1993) showed how to obtain an unbiased estimator of $\mathbf{a_0}$.

Lemma 4.3. *(Theorem 4.1, Su and Liu, 1993, p. 1352) Let* $\mathbf{\Sigma}_x = \mathbf{\Sigma}_y = \mathbf{\Sigma}$ *and* $\mathbf{S} = \sum_i(\mathbf{X_i} - \overline{\mathbf{X}})(\mathbf{X_i} - \overline{\mathbf{X}})' + \sum_j(\mathbf{Y_j} - \overline{\mathbf{Y}})(\mathbf{Y_j} - \overline{\mathbf{Y}})'$ *be the pooled sum of squares. Then,* $\hat{\mathbf{T}}^{-1} = (n+m-k-3)\mathbf{S}^{-1}$ *is an unbiased estimate of* $\mathbf{\Sigma}^{-1}$ *and* $\hat{\mathbf{a}}_0 = \hat{\mathbf{T}}^{\mathbf{-1}}(\overline{\mathbf{Y}} - \overline{\mathbf{X}})$ *is an unbiased estimate of* $\mathbf{a}_0 = \mathbf{\Sigma^{-1}}(\mu_{\mathbf{y}} - \mu_{\mathbf{x}})$.

Let $\hat{\mathbf{a}}$ be an estimator of $\mathbf{a_0}$ such that the following condition is satisfied

$$\sqrt{n}(\hat{\mathbf{a}} - \mathbf{a_0}) = O_p(1), \quad \text{as} \quad n \to \infty. \tag{4.3}$$

Notice, that an immediate consequence of (4.3) is

$$\hat{\mathbf{a}} - \mathbf{a_0} = o_p(1), \quad \text{as} \quad n \to \infty. \tag{4.4}$$

For any $\mathbf{a} \in \mathbb{R}^k$, by definition of multivariate normal distribution, the random variable $X = \mathbf{a}'\mathbf{X}$ is normally distributed with zero mean and variance $\mathbf{a}'\mathbf{\Sigma}\mathbf{a}$. Similarly, for any $\mathbf{a} \in \mathbb{R}^k$, $Y = \mathbf{a}'\mathbf{Y}$ is normally distributed with mean $\mathbf{a}'\mu$ and variance $\mathbf{a}'\mathbf{\Sigma}\mathbf{a}$. If we denote the distribution functions of the random variables X and Y by

$F(\cdot, \mathbf{a})$ and $G(\cdot, \mathbf{a})$, respectively, then, the following relations can be easily shown

$$F(x, \mathbf{a}) = \Phi\left(\frac{x}{\sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}}\right), \quad x \in \mathbb{R}, \tag{4.5}$$

$$G(y, \mathbf{a}) = \Phi\left(\frac{y - \mathbf{a}'\mu}{\sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}}\right), \quad y \in \mathbb{R}, \tag{4.6}$$

and,

$$G(F^{-1}(p, \mathbf{a}), \mathbf{a}) = \Phi\left(\Phi^{-1}(p) - \frac{\mathbf{a}'\mu}{\sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}}\right), \quad p \in (0, 1). \tag{4.7}$$

Let $F_n(x, \mathbf{a}) = n^{-1} \sum_{i=1}^{n} \mathbf{I}[\mathbf{a}' X_i \leqslant x]$ be the empirical distribution function of the random sample $\{\mathbf{a}'\mathbf{X_i}\}_{i=1}^{n}$ and $F_n^{-1}(p, \mathbf{a})$ be its p^{th} quantile as in Definition 2.30. Similarly, let $G_m(y, \mathbf{a}) = m^{-1} \sum_{j=1}^{m} \mathbf{I}[\mathbf{a}' Y_j \leqslant y]$ be the empirical distribution function of the random sample $\{\mathbf{a}'\mathbf{Y_j}\}_{j=1}^{m}$. Our main goal is to find the asymptotic distribution of the *generalized empirical ROC process* defined as

$$\sqrt{m}\left(G_m\left(F_n^{-1}(p, \hat{\mathbf{a}}), \hat{\mathbf{a}}\right) - G\left(F^{-1}(p, \mathbf{a_0}), \mathbf{a_0}\right)\right), \quad p \in (0, 1). \tag{4.8}$$

Next, we will show that the process in (4.8) is equivalent to another process which is easier to deal with. Notice that the empirical distribution function F_n can be rewritten as

$$\begin{aligned} & F_n(x, \mathbf{a}) \\ = \;& n^{-1} \sum_{i=1}^{n} \mathbf{I}\left[\Phi\left(\frac{\mathbf{a}'\mathbf{X_i}}{\sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}}\right) \leqslant \Phi\left(\frac{x}{\sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}}\right)\right] \\ = \;& U_n\left(\Phi\left(\frac{x}{\sqrt{\mathbf{a}'\mathbf{\Sigma}\mathbf{a}}}\right), \mathbf{b}\right), \end{aligned} \tag{4.9}$$

where

$$\mathbf{b} = \frac{1}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}}\mathbf{a}, \qquad \forall \mathbf{a} \in \mathbb{R}^k, \tag{4.10}$$

and

$$U_n(t, \mathbf{b}) = n^{-1} \sum_{i=1}^{n} \mathbf{I}\left[\Phi\left(\mathbf{b}' X_i\right) \leqslant t\right], \quad t \in (0,1). \tag{4.11}$$

Note that, for $\mathbf{b}$ defined by (4.10), $\Phi\left(\mathbf{b}' X_1\right), \ldots, \Phi\left(\mathbf{b}' X_n\right)$ are i.i.d Uniform[0,1] random variables. For the fixed vector $\mathbf{b_0}$, obtained by using $\mathbf{a_0}$ in (4.10), the process in (4.11) becomes the uniform empirical process and we will simply denote it by U_n. Furthermore, it can be shown that

$$F_n^{-1}(t, \mathbf{a}) = \sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}\,\Phi^{-1}\left(U_n^{-1}(t, \mathbf{b})\right), \quad t \in (0,1). \tag{4.12}$$

By analogy, we can rewrite the empirical distribution function G_m as

$$\begin{aligned} & G_m(y, \mathbf{a}) \\ = \; & m^{-1} \sum_{j=1}^{m} \mathbf{I}\left[\Phi\left(\mathbf{b}' Y_j\right) \leqslant \Phi\left(\frac{y}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}}\right)\right] \\ = \; & \tilde{G}_m\left(\Phi\left(\frac{y}{\sqrt{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}}\right), \mathbf{b}\right), \end{aligned} \tag{4.13}$$

where

$$\tilde{G}_m(t, \mathbf{b}) = m^{-1} \sum_{j=1}^{m} \mathbf{I}\left[\Phi\left(\mathbf{b}' Y_j\right) \leqslant t\right], \quad t \in (0,1). \tag{4.14}$$

Let us denote the distribution function of the random variable $\Phi\left(\mathbf{b}' Y\right)$ by $\tilde{G}(\,\cdot\,, \mathbf{b})$ and notice that it is equal to $\Phi(\Phi^{-1}(\,\cdot\,) - \mathbf{b}'\mu)$. Then, by (4.7), (4.9), (4.12), and

(4.13), the process in (4.8) can equivalently be rewritten as follows

$$\sqrt{m}\left(\tilde{G}_m\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right),\hat{\mathbf{b}}\right)-\tilde{G}\left(p,\mathbf{b_0}\right)\right),\quad p\in(0,1), \tag{4.15}$$

where $\hat{\mathbf{b}}$ is obtained by using $\hat{\mathbf{a}}$ in (4.10).

4.2 Decomposition of the Generalized Empirical ROC Process

In this section we will decompose the equivalent generalized empirical ROC process in (4.15) using the same technique introduced in Chapter 3. But, before checking the conditions of Lemma 2.75, we will derive some useful properties for $\mathbf{b_0}$ and $\hat{\mathbf{b}}$.

Proposition 4.4. *Let* $\mathbf{a_0}$ *be defined by (4.2). Then,*

$$\mathbf{b_0}=\frac{1}{\sqrt{\mu'\Sigma^{-1}\mu}}\Sigma^{-1}\mu, \tag{4.16}$$

and

$$\mathbf{b_0}'\mu=\sqrt{\mu'\Sigma^{-1}\mu}>0. \tag{4.17}$$

Proof. The result follows immediately from (4.2) and (4.10). □

Proposition 4.5. *Let* $\hat{\mathbf{a}}$ *be an estimator of* $\mathbf{a_0}$ *such that (4.3) holds. Then,*

$$\sqrt{n}(\hat{\mathbf{b}}-\mathbf{b_0})=O_p(1),\quad as\quad n\to\infty. \tag{4.18}$$

Proof. By simple algebraic manipulations, we have

$$\begin{aligned}
\sqrt{n}\left(\hat{\mathbf{b}}-\mathbf{b_0}\right) &= \sqrt{n}\left(\frac{1}{\sqrt{\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}}}\hat{\mathbf{a}}-\frac{1}{\sqrt{\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0}}}\mathbf{a_0}\right)\\
&= \sqrt{n}\left(\frac{1}{\sqrt{\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}}}-\frac{1}{\sqrt{\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0}}}\right)(\hat{\mathbf{a}}-\mathbf{a_0})\\
&\quad+\sqrt{n}\frac{1}{\sqrt{\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0}}}(\hat{\mathbf{a}}-\mathbf{a_0})\\
&\quad+\sqrt{n}\left(\frac{1}{\sqrt{\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}}}-\frac{1}{\sqrt{\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0}}}\right)\mathbf{a_0}. \qquad (4.19)
\end{aligned}$$

Notice that

$$\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}-\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0} = (\hat{\mathbf{a}}-\mathbf{a_0})'\,\boldsymbol{\Sigma}\,(\hat{\mathbf{a}}-\mathbf{a_0})+2\,(\hat{\mathbf{a}}-\mathbf{a_0})'\,\boldsymbol{\Sigma}\mathbf{a_0}.$$

Therefore, by (4.4)

$$\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}-\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0} = o_p(1)O(1)o_p(1)+o_p(1)O(1) = o_p(1), \quad \text{as} \quad n\to\infty. \qquad (4.20)$$

Moreover, since $\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}$ is a consistent estimator of $\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0}$ then

$$\frac{1}{\sqrt{\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}}}-\frac{1}{\sqrt{\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0}}} = o_p(1), \quad \text{as} \quad n\to\infty. \qquad (4.21)$$

Notice that the first two terms are $o_p(1)$ and $O_p(1)$, respectively, by using (4.3) and (4.21). Thus, (4.18) is true if the last term is either $o_p(1)$ or $O_p(1)$. Denote $\sqrt{\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}}\sqrt{\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0}}(\sqrt{\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0}}+\sqrt{\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}})$ and $2(\mathbf{a_0}'\boldsymbol{\Sigma}\mathbf{a_0})^{3/2}$ by $\hat{D}$ and D_0, respectively, and notice that $\hat{D}$ is a consistent estimator of D_0. Then, after some further algebraic manipulation, we obtain

$$\begin{aligned}
\sqrt{n}\left(\frac{1}{\sqrt{\hat{\mathbf{a}}'\mathbf{\Sigma}\hat{\mathbf{a}}}} - \frac{1}{\sqrt{\mathbf{a_0}'\mathbf{\Sigma}\mathbf{a_0}}}\right) &= \sqrt{n}\frac{\mathbf{a_0}'\mathbf{\Sigma}\mathbf{a_0} - \hat{\mathbf{a}}'\mathbf{\Sigma}\hat{\mathbf{a}}}{\hat{D}} \\
&= \sqrt{n}(\hat{\mathbf{a}} - \mathbf{a_0})'\Sigma(\hat{\mathbf{a}} - \mathbf{a_0})\left(\frac{1}{D_0} - \frac{1}{\hat{D}}\right) - \sqrt{n}(\hat{\mathbf{a}} - \mathbf{a_0})'\Sigma(\hat{\mathbf{a}} - \mathbf{a_0})\frac{1}{D_0} \\
&\quad +2\sqrt{n}(\hat{\mathbf{a}} - \mathbf{a_0})'\Sigma\mathbf{a_0}\left(\frac{1}{D_0} - \frac{1}{\hat{D}}\right) - 2\sqrt{n}(\hat{\mathbf{a}} - \mathbf{a_0})'\Sigma\mathbf{a_0}\frac{1}{D_0}. \qquad (4.22)
\end{aligned}$$

By using (4.3), (4.4), and the consistency of $\hat{D}$ we obtain

$$\begin{aligned}
&\sqrt{n}\left(\frac{1}{\sqrt{\hat{\mathbf{a}}'\mathbf{\Sigma}\hat{\mathbf{a}}}} - \frac{1}{\sqrt{\mathbf{a_0}'\mathbf{\Sigma}\mathbf{a_0}}}\right) \\
&= O_p(1)o_p(1)o_p(1) + O_p(1)o_p(1)O(1) + O_p(1)O(1)o_p(1) + O_p(1)O(1) \\
&= O_p(1), \quad \text{as} \quad n \to \infty. \qquad (4.23)
\end{aligned}$$

The proof is complete since $o_p(1) + O_p(1) + O_p(1)$ is $O_p(1)$. □

Again, an immediate consequence of (4.18) is

$$\hat{\mathbf{b}} - \mathbf{b_0} = o_p(1), \quad \text{as} \quad n \to \infty. \qquad (4.24)$$

The next proposition will be an important argument in the later proofs.

Proposition 4.6. *Let* $\hat{\mathbf{a}}$ *be an estimator of* $\mathbf{a_0}$ *such that (4.3) holds. Then*

$$\sqrt{n}(\hat{\mathbf{b}} - \mathbf{b_0})'\mu = o_p(1), \quad \textit{as} \quad n \to \infty. \qquad (4.25)$$

Proof. From (4.19), we have

$$
\begin{aligned}
&\sqrt{n}\left(\frac{\hat{\mathbf{a}}'\mu}{\sqrt{\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}}} - \frac{\mathbf{a_0}'\mu}{\sqrt{\mathbf{a_0'}\boldsymbol{\Sigma}\mathbf{a_0}}}\right) \\
&= \sqrt{n}\left(\frac{1}{\sqrt{\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}}} - \frac{1}{\sqrt{\mathbf{a_0}'\Sigma\mathbf{a_0}}}\right)(\hat{\mathbf{a}}-\mathbf{a_0})'\mu \\
&\quad + \sqrt{n}\frac{(\hat{\mathbf{a}}-\mathbf{a_0})'\mu}{\sqrt{\mathbf{a_0}'\Sigma\mathbf{a_0}}} \\
&\quad + \sqrt{n}\left(\frac{1}{\sqrt{\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}}} - \frac{1}{\sqrt{\mathbf{a_0'}\boldsymbol{\Sigma}\mathbf{a_0}}}\right)\mathbf{a_0}'\mu. \qquad (4.26)
\end{aligned}
$$

By using (4.4) and (4.23), the first term on the right hand side of (4.26) is $o_p(1)$. Next, we will show that the sum of the last two terms on the right hand side of (4.26) is $o_p(1)$. By using (4.3), (4.4), the consistency of $\hat{D}$ and the definition of D_0 in (4.22), the last term of (4.26) becomes

$$
\sqrt{n}\left(\frac{1}{\sqrt{\hat{\mathbf{a}}'\Sigma\hat{\mathbf{a}}}} - \frac{1}{\sqrt{\mathbf{a_0'}\boldsymbol{\Sigma}\mathbf{a_0}}}\right)\mathbf{a_0}'\mu = o_p(1) - \sqrt{n}\frac{(\hat{\mathbf{a}}-\mathbf{a_0})'\Sigma\mathbf{a_0}}{(\mathbf{a_0}'\Sigma\mathbf{a_0})^{3/2}}\mathbf{a_0}'\mu. \qquad (4.27)
$$

By using the definition of $\mathbf{a_0}$, the second term in the right hand side of (4.27) can be further simplified as follows

$$
\sqrt{n}\frac{(\hat{\mathbf{a}}-\mathbf{a_0})'\Sigma\mathbf{a_0}}{(\mathbf{a_0'}\boldsymbol{\Sigma}\mathbf{a_0})^{3/2}}\mathbf{a_0}'\mu = \sqrt{n}\frac{(\hat{\mathbf{a}}-\mathbf{a_0})'(\Sigma\Sigma^{-1}\mu\mu'\Sigma^{-1}\mu)}{(\mu'\Sigma^{-1}\Sigma\Sigma^{-1}\mu)^{3/2}} = \sqrt{n}\frac{(\hat{\mathbf{a}}-\mathbf{a_0})'\mu}{\sqrt{\mathbf{a_0}'\Sigma\mathbf{a_0}}}. \qquad (4.28)
$$

Notice that (4.28) cancels out the second term on the right hand side of (4.26). Hence, the result of the lemma will follow immediately. □

In order to apply Lemma 2.75, we need to construct a Donsker class of functions and show that the conditions (2.10) and (2.11) are satisfied. Let g_p and $g_{\mathbf{b},H,p}$ be the following functions

$$g_p : \mathbb{R}^k \longrightarrow \mathbb{R}, \quad g_p(\mathbf{y}) = \mathbf{I}\left[\Phi\left(\mathbf{b_0}'\mathbf{y}\right) \leqslant p\right], \tag{4.29}$$

and

$$g_{\mathbf{b},H,p} : \mathbb{R}^k \longrightarrow \mathbb{R}, \quad g_{\mathbf{b},H,p}(\mathbf{y}) = \mathbf{I}\left[\Phi\left(\mathbf{b}'\mathbf{y}\right) \leqslant H^{-1}(p)\right], \tag{4.30}$$

where $p \in (0,1)$, $\mathbf{b}$, $\mathbf{b_0}$ are defined by (4.10) and $H \in \mathcal{H}$, the class of all distributions functions defined on $[0,1]$. Let $\mathcal{G}'$ and $\mathcal{G}$ be the following classes of functions

$$\mathcal{G}' = \left\{g_{\mathbf{b},H,p} : \mathbf{b} \in \mathbb{R}^k, H \in \mathcal{H}, p \in (0,1)\right\}. \tag{4.31}$$

$$\mathcal{G} = \left\{g_{\mathbf{b},H,p} - g_p : \mathbf{b}, \mathbf{b_0} \in \mathbb{R}^k, H \in \mathcal{H}, p \in (0,1)\right\}. \tag{4.32}$$

Lemma 4.7. *The class of functions $\mathcal{G}'$ defined in (4.31) is a VC subgraph class.*

Proof. Notice that we can write $g_{\mathbf{b},H,p}(\mathbf{y}) = \mathbf{I}[b_1y_1 + b_2y_2 + \ldots + b_ky_k \leqslant \Phi^{-1}(H^{-1}(p))]$. Therefore, $\mathcal{G}' \subseteq \left\{\mathbf{I}[b_1y_1 + b_2y_2 + \ldots + b_ky_k \leqslant v], \mathbf{b} = (b_1, b_2, \ldots, b_k)' \in \mathbb{R}^k, v \in \mathbb{R}\right\}$ since $\Phi^{-1}(H^{-1}(p)) = v \in \mathbb{R}$ for all $p \in (0,1)$. The later class is a VC subgraph class according to Lemma 2.72. Hence, $\mathcal{G}'$ is also a VC subgraph class. □

Lemma 4.8. *The class of functions $\mathcal{G}$ defined in (4.32) is a Q-Donsker class of functions.*

Proof. The proof is similar to that of Lemma 3.8 and is omitted. □

Next, we will prove that condition (2.10) of Lemma 2.75 is satisfied. The steps of the proof are similar to those in Lemma 3.10.

Lemma 4.9. *Let $\mathcal{G}$ be the class of functions defined in (4.32). For any $\varepsilon > 0$, there exists $\delta_1(\varepsilon) > 0$, $\delta_2(\varepsilon) > 0$ such that if*

$$\|\mathbf{b} - \mathbf{b_0}\| < \delta_1(\varepsilon), \tag{4.33}$$

$$\sup_{t \in (0,1)} |H(t) - t| < \delta_2(\varepsilon), \tag{4.34}$$

then, for n sufficiently large,

$$\sup_{p \in (0,1)} \int \left(g_{\mathbf{b},H,p}(\mathbf{y}) - g_p(\mathbf{y})\right)^2 dQ(\mathbf{y}) < \varepsilon. \tag{4.35}$$

Proof. Let $\varepsilon > 0$ be given. For any $p \in (0,1)$, by using properties of the indicator function, the integral from (4.35) can be rewritten as follows

$$\begin{aligned}
&\int \left(g_{\mathbf{b},H,p}(\mathbf{y}) - g_p(\mathbf{y})\right)^2 dQ(\mathbf{y}) \\
&= \int \mathbf{I}\left[\Phi\left(\mathbf{b}'\mathbf{y}\right) \leqslant H^{-1}(p), \Phi\left(\mathbf{b_0}'y\right) > p\right] dQ(\mathbf{y}) \\
&\quad + \int \mathbf{I}\left[\Phi\left(\mathbf{b}'\mathbf{y}\right) > H^{-1}(p), \Phi\left(\mathbf{b_0}'y\right) \leqslant p\right] dQ(\mathbf{y}) \\
&= Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) \leqslant H^{-1}(p), \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) > p\right) \\
&\quad + Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) > H^{-1}(p), \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant p\right).
\end{aligned} \tag{4.36}$$

Moreover, we proved in Lemma 3.10 that if (4.34) is true then, for any $p \in (0,1)$ we have from (3.23)

$$\left|H^{-1}(p) - p\right| \leqslant \delta.$$

Now, we will show that for some properly chosen values p_1 and p_2 in the interval $(0,1)$ there exists $\delta_1(\varepsilon)$ and $\delta_2(\varepsilon)$ such that supremum of the integral in (4.35) can be made arbitrarily small on each of the intervals $(0,p_1)$, $[p_1,p_2]$, and $(p_2,1)$. First, we choose $\max\left\{1/2, 2\Phi\left(\mathbf{b_0'}\mu/2\right)-1\right\} < p_2 < 1$ such that

$$\Phi\left(\Phi^{-1}\left(2(1-p_2)\right)+2\mathbf{b_0}'\mu\right) < \varepsilon/4, \tag{4.37}$$

and $p_1 = 1-p_2$. Let M_ε^* be given by

$$M_\varepsilon^* = \sup_{p\in(0,1-p_1/2)} \frac{\phi\left(\Phi^{-1}(p)-\mathbf{b_0}'\mu\right)}{\phi\left(\Phi^{-1}(p)\right)}, \tag{4.38}$$

and notice that $M_\varepsilon^* > 1$. Also, there exists $M_\varepsilon > 1$ such that

$$Pr\left(\|Y\| > M_\varepsilon\right) = \varepsilon/4. \tag{4.39}$$

Next, choose $\delta_1(\varepsilon)$ and $\delta_2(\varepsilon)$ as follows

$$\delta_1(\varepsilon) < \min\left\{\frac{\mathbf{b_0}'\mu}{\|\mu\|}, \frac{1-p_2}{4M_\varepsilon M_\varepsilon^*}\right\} \tag{4.40}$$

and

$$\delta_2(\varepsilon) < \frac{1-p_2}{4M_\varepsilon^*}. \tag{4.41}$$

Notice that, since $M_\varepsilon^* > 1$, then actually

$$\delta_2(\varepsilon) < \frac{1 - p_2}{4}. \tag{4.42}$$

Next, we will show some inequalities that will be used later in this proof. By using triangle inequality, (4.33) and (4.40) we have

$$\mathbf{b}'\mu = \mathbf{b_0}'\mu + (\mathbf{b} - \mathbf{b_0})'\mu < \mathbf{b_0}'\mu + \delta_1(\varepsilon)\,\|\mu\| < 2\mathbf{b}_\mathbf{0}'\mu, \tag{4.43}$$

and

$$\mathbf{b}'\mu = \mathbf{b_0}'\mu - (\mathbf{b_0} - \mathbf{b})'\mu > \mathbf{b_0}'\mu - \delta_1(\varepsilon)\,\|\mu\| > 0. \tag{4.44}$$

From the definition of p_1, monotonicity of the cumulative and inverse distribution functions, (4.17), and (4.37) we have

$$p_1 = \Phi\left(\Phi^{-1}(1 - p_2)\right) < \Phi\left(\Phi^{-1}(2(1 - p_2)) + 2\mathbf{b_0}'\mu\right) < \varepsilon/4. \tag{4.45}$$

Finally, by simple manipulation of (3.23) and using (4.42), (4.45) we obtain

$$1 - H^{-1}(p_2) < (1 - p_2) + \delta_2(\varepsilon) < 2(1 - p_2) = 2p_1 < 2\varepsilon/4, \tag{4.46}$$

and,

$$H^{-1}(p_1) < p_1 + \delta_2(\varepsilon) < 2p_1 < 2\varepsilon/4. \tag{4.47}$$

Firstly, let $p \in (0, p_1)$. Then, by using the fact that the random variables $\mathbf{b}'\mathbf{Y}$ and

$\mathbf{b_0}'\mathbf{Y}$ are normally distributed with variance one and means $\mathbf{b}'\mu$ and $\mathbf{b}_\mathbf{0}'\mu$, respectively, (4.44), (4.17), monotonicity of the cumulative distribution function, (4.47), and (4.45) we obtain

$$\begin{aligned}
&\sup_{p\in(0,p_1)} Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) \leqslant H^{-1}(p), \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) > p\right) \\
&\qquad + \sup_{p\in(0,p_1)} Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) > H^{-1}(p), \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant p\right) \\
&\leqslant \sup_{p\in(0,p_1)} Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) \leqslant H^{-1}(p)\right) + \sup_{p\in(0,p_1)} Pr\left(\Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant p\right) \\
&\leqslant \Phi\left(\Phi^{-1}(H^{-1}(p_1)) - \mathbf{b}'\mu\right) + \Phi\left(\Phi^{-1}(p_1) - \mathbf{b}_\mathbf{0}'\mu\right) \\
&\leqslant H^{-1}(p_1) + p_1 < 3\varepsilon/4. \qquad (4.48)
\end{aligned}$$

Secondly, let $p \in (p_2, 1)$. Then, by using the same arguments as in the previous case plus the symmetry of the cumulative and inverse standard normal distribution and inequalities (4.43), (4.44), 4.46, and (4.37) we obtain

$$\begin{aligned}
&\sup_{p\in(p_2,1)} Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) \leqslant H^{-1}(p), \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) > p\right) \\
&\qquad + \sup_{p\in(p_2,1)} Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) > H^{-1}(p), \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant p\right) \\
&\leqslant 1 - Pr\left[\Phi\left(\mathbf{b}_\mathbf{0}'\mathbf{Y}\right) \leqslant p_2\right] + 1 - Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) \leqslant H^{-1}(p_2)\right) \\
&= \Phi\left(\Phi^{-1}(1-p_2) + \mathbf{b}_\mathbf{0}'\mu\right) + \Phi\left(\Phi^{-1}(1 - H^{-1}(p_2)) + \mathbf{b}'\mu\right) \\
&< 2\Phi\left(\Phi^{-1}(2(1-p_2)) + 2\mathbf{b}_\mathbf{0}'\mu\right) < 2\varepsilon/4. \qquad (4.49)
\end{aligned}$$

Lastly, let $p \in [p_1, p_2]$ and choose

$$\eta = \delta_1(\varepsilon)M_\varepsilon. \qquad (4.50)$$

Then,

$$
\begin{aligned}
& Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) \leqslant H^{-1}(p), \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) > p\right) \\
= \;& Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) \leqslant H^{-1}(p), \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) > p, \left|\Phi\left(\mathbf{b}'\mathbf{Y}\right) - \Phi\left(\mathbf{b_0}'\mathbf{Y}\right)\right| \leqslant \eta\right) \\
& + Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) \leqslant H^{-1}(p), \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) > p, \left|\Phi\left(\mathbf{b}'\mathbf{Y}\right) - \Phi\left(\mathbf{b_0}'\mathbf{Y}\right)\right| > \eta\right) \\
\leqslant \;& Pr\left(p < \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant H^{-1}(p) + \eta\right) + Pr\left(\left|\Phi\left(\mathbf{b}'\mathbf{Y}\right) - \Phi\left(\mathbf{b_0}'\mathbf{Y}\right)\right| > \eta\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& Pr\left(\Phi\left(\mathbf{b}'\mathbf{Y}\right) > H^{-1}(p), \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant p\right) \\
\leqslant \;& Pr\left(H^{-1}(p) - \eta < \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant p\right) + Pr\left(\left|\Phi\left(\mathbf{b}'\mathbf{Y}\right) - \Phi\left(\mathbf{b_0}'\mathbf{Y}\right)\right| > \eta\right).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& \int \left(g_{\mathbf{b},H,p}(\mathbf{y}) - g_p(\mathbf{y})\right)^2 dQ(\mathbf{y}) \\
\leqslant \;& Pr\left(p < \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant H^{-1}(p) + \eta\right) \\
& + Pr\left(H^{-1}(p) - \eta < \Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant p\right) \\
& + 2Pr\left(\left|\Phi\left(\mathbf{b}'\mathbf{Y}\right) - \Phi\left(\mathbf{b_0}'\mathbf{Y}\right)\right| > \eta\right) \\
= \;& \Phi\left(\Phi^{-1}\left(H^{-1}(p) + \eta\right) - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right) \qquad (4.51) \\
& + \Phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}\left(H^{-1}(p) - \eta\right) - \mathbf{b_0}'\mu\right) \qquad (4.52) \\
& + 2Pr\left(\left|\Phi\left(\mathbf{b}'\mathbf{Y}\right) - \Phi\left(\mathbf{b_0}'\mathbf{Y}\right)\right| > \eta\right). \qquad (4.53)
\end{aligned}
$$

For any $p \in [p_1, p_2]$, by using (3.23), (4.50), (4.40), and (4.42), we obtain the following bound for $H^{-1}(p)+\eta$

$$\begin{aligned} & H^{-1}(p)+\eta \\ < \; & p_2 + \sup_{p\in[p_1,p_2]} \left|H^{-1}(p)-p\right| + \delta_1(\varepsilon)M_\varepsilon \\ < \; & p_2 + \frac{1-p_2}{4} + \frac{1-p_2}{4M_\varepsilon}M_\varepsilon \\ < \; & p_2 + \frac{1-p_2}{2} = 1-\frac{p_1}{2}. \end{aligned} \tag{4.54}$$

Therefore, by applying the first order Taylor expansion to (4.51) we obtain

$$\begin{aligned} & \Phi\left(\Phi^{-1}\left(H^{-1}(p)+\eta\right)-\mathbf{b_0}'\mu\right)-\Phi\left(\Phi^{-1}(p)-\mathbf{b_0}'\mu\right) \\ = \; & \frac{\phi\left(\Phi^{-1}(p^*)-\mathbf{b}_\mathbf{0}'\mu\right)}{\phi\left(\Phi^{-1}(p^*)\right)}\left(H^{-1}(p)+\eta-p\right), \end{aligned} \tag{4.55}$$

where p^* is between p and $H^{-1}(p)+\eta$, or, according to (4.54), p^* is between p and $1-p_1/2$. Hence, by using (4.55), (3.23), (4.40), and (4.41) we obtain

$$\begin{aligned} & \sup_{p\in[p_1,p_2]} \Phi\left(\Phi^{-1}\left(H^{-1}(p)+\eta\right)-\mathbf{b_0}'\mu\right)-\Phi\left(\Phi^{-1}(p)-\mathbf{b_0}'\mu\right) \\ < \; & M_\varepsilon^*\left(\delta_2(\varepsilon)+\eta\right) \\ < \; & M_\varepsilon^*\left(\frac{1-p_2}{4\max\{M_\varepsilon^*,1\}}+\frac{1-p_2}{4M_\varepsilon\max\{M_\varepsilon^*,1\}}\right) \\ < \; & \frac{1-p_2}{2} < \varepsilon/4. \end{aligned} \tag{4.56}$$

Analogously,

$$\begin{aligned}
&\sup_{p\in[p_1,p_2]} \Phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}\left(H^{-1}(p) - \eta\right) - \mathbf{b_0}'\mu\right) \\
&= \sup_{p\in[p_1,p_2]} \frac{\phi\left(\Phi^{-1}(p^{**}) - \mathbf{b_0'}\mu\right)}{\phi\left(\Phi^{-1}(p^{**})\right)} \left(p - H^{-1}(p) + \eta\right) \\
&< M_\varepsilon^* \left(\delta_2(\varepsilon) + \eta\right) \\
&< \frac{1-p_2}{2} < \varepsilon/4, \qquad (4.57)
\end{aligned}$$

where, again it can be shown that p^{**} is bounded above by $1 - p_1/2$. Finally, notice that by applying first order Taylor series expansion to (4.53), we obtain

$$\left|\Phi\left(\mathbf{b}'\mathbf{Y}\right) - \Phi\left(\mathbf{b_0}'\mathbf{Y}\right)\right| < \|\mathbf{b} - \mathbf{b_0}\| \|\mathbf{Y}\| < \delta_1(\varepsilon) \|\mathbf{Y}\|. \qquad (4.58)$$

Therefore, by using (4.58), and (4.50) we obtain

$$2Pr\left(\left|\Phi\left(\mathbf{b}'\mathbf{Y}\right) - \Phi\left(\mathbf{b_0}'\mathbf{Y}\right)\right| > \eta\right) < 2Pr\left(\|\mathbf{Y}\| > M_\varepsilon\right) < 2\varepsilon/4. \qquad (4.59)$$

Thus, from (4.56), (4.57), and (4.59) we have

$$\sup_{p\in[p_1,p_2]} \int \left(g_{\mathbf{b},H,p}(\mathbf{y}) - g_p(\mathbf{y})\right)^2 dQ(\mathbf{y}) < \varepsilon. \qquad (4.60)$$

Conclusion of the Lemma follows immediately from (4.48), (4.49), and (4.60). □

Next, we will prove that condition (2.11) is satisfied by using Theorem 2.83 and we start with the construction of an appropriate class of functions. Let f_p and $f_{\mathbf{b},p}$

be the following functions

$$f_p : \mathbb{R}^k \longrightarrow \mathbb{R}, f_p(\mathbf{x}) = \mathbf{I}\left[\Phi(\mathbf{b_0}'\mathbf{x}) \leqslant p\right], \tag{4.61}$$

and

$$f_{\mathbf{b},p} : \mathbb{R}^k \longrightarrow \mathbb{R}, f_{\mathbf{b},p}(\mathbf{x}) = \mathbf{I}\left[\Phi(\mathbf{b}'\mathbf{x}) \leqslant p\right], \tag{4.62}$$

where $p \in (0,1)$ and $\mathbf{b}$, $\mathbf{b_0}$ are defined by (4.10). Let $\mathcal{F}_n$ be the following class of functions

$$\mathcal{F}_n = \left\{ f_{\mathbf{b},p} - f_p : \|\mathbf{b} - \mathbf{b_0}\| \leqslant \frac{M}{\sqrt{n}}, p \in (0,1) \right\}, \tag{4.63}$$

where $M \in \mathbb{R}^+$.

Lemma 4.10. *Let $\mathcal{F}_n$ be the class defined in (4.63) and $f \in \mathcal{F}_n$ be any function. Then, there exists a constant C such that*

$$Pf^2 \leqslant Cn^{-1/2}(\log n)^{1/2}, \quad n > 2. \tag{4.64}$$

Proof. First we write Pf^2 in a more convenient form, namely

$$\begin{aligned} Pf^2 = & \int \left(f_{\mathbf{b},p}(\mathbf{x}) - f_p(\mathbf{x})\right)^2 dP(\mathbf{x}) \\ = & \int \mathbf{I}\left[\mathbf{b}'\mathbf{x} \leqslant \Phi^{-1}(p), \mathbf{b_0}'\mathbf{x} > \Phi^{-1}(p)\right] dP(\mathbf{x}) \\ & + \int \mathbf{I}\left[\mathbf{b}'\mathbf{x} > \Phi^{-1}(p), \mathbf{b_0}'\mathbf{x} \leqslant \Phi^{-1}(p)\right] dP(\mathbf{x}) \\ = & \; Pr\left(\mathbf{b}'\mathbf{X} \leqslant \Phi^{-1}(p), \mathbf{b_0}'\mathbf{X} > \Phi^{-1}(p)\right) \qquad (4.65) \\ & + Pr\left(\mathbf{b}'\mathbf{X} > \Phi^{-1}(p), \mathbf{b_0}'\mathbf{X} \leqslant \Phi^{-1}(p)\right). \qquad (4.66) \end{aligned}$$

Then, by applying the same technique as in the last case considered in Lemma 4.9, for any $\eta > 0$, we obtain

$$\begin{aligned} & Pr\left(\mathbf{b}'\mathbf{X} \leqslant \Phi^{-1}(p), \mathbf{b_0}'\mathbf{X} > \Phi^{-1}(p)\right) \\ \leqslant\ & \left|\Phi\left(\Phi^{-1}(p) + \eta - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right)\right| \\ & +Pr\left(\left|(\mathbf{b} - \mathbf{b_0})'\,\mathbf{X}\right| > \eta\right), \end{aligned} \tag{4.67}$$

and

$$\begin{aligned} & Pr\left(\mathbf{b}'\mathbf{X} > \Phi^{-1}(p), \mathbf{b_0}'\mathbf{X} \leqslant \Phi^{-1}(p)\right) \\ \leqslant\ & \left|\Phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}(p) - \eta - \mathbf{b_0}'\mu\right)\right| \\ & +Pr\left(\left|(\mathbf{b} - \mathbf{b_0})'\,\mathbf{X}\right| > \eta\right), \end{aligned} \tag{4.68}$$

Note that, by applying Taylor expansion of first order on first right hand side terms from (4.67) and (4.68), we obtain

$$\begin{aligned} & \left|\Phi\left(\Phi^{-1}(p) + \eta - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right)\right| \\ & \quad + \left|\Phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}(p) - \eta - \mathbf{b_0}'\mu\right)\right| \\ \leqslant\ & \frac{2}{\sqrt{2\pi}}\eta. \end{aligned} \tag{4.69}$$

By plugging (4.67), and (4.68) into (4.65) and (4.66), respectively, and by using (4.69) we obtain

$$Pf^2 \leqslant \frac{2}{\sqrt{2\pi}}\eta + 2Pr\left(\left|(\mathbf{b} - \mathbf{b_0})'\,\mathbf{X} > \eta\right|\right) \tag{4.70}$$

Furthermore, the probability in (4.70) can be majored by using the following equality from Serfling, p.81

$$1 - \Phi\left(B(\log n)^{1/2}\right) \leqslant \frac{1}{\sqrt{2\pi}B(\log n)^{1/2}} n^{-1/2B^2}, \quad n > 1, \tag{4.71}$$

where $B \geqslant 1$. Since Σ is positive definite and f is any function in the class $\mathcal{F}_n$ defined by (4.63), then the standard deviation of the random variable $(\mathbf{b}-\mathbf{b_0})'\mathbf{X}$ can be majored as follows:

$$\sqrt{(\mathbf{b}-\mathbf{b_0})'\Sigma(\mathbf{b}-\mathbf{b_0})} \leqslant \frac{M}{\sqrt{n}} \|\Sigma\|^{1/2}. \tag{4.72}$$

Let us choose

$$\eta = C_0 n^{-1/2} (\log n)^{1/2}, \tag{4.73}$$

where the constant C_0 satisfies

$$B = \left(\frac{C_0}{M \|\Sigma\|^{1/2}}\right) \geqslant 1. \tag{4.74}$$

Then, by (4.72),(4.71), and (4.74) we obtain for $n > 2$

$$Pr\left(\left|(\mathbf{b}-\mathbf{b_0})' \mathbf{X} > \eta\right|\right) \leqslant 2\left(1 - \Phi\left(\frac{\eta}{\frac{M}{\sqrt{n}}\|\Sigma\|^{1/2}}\right)\right) \leqslant \frac{2}{\sqrt{2\pi}} n^{-1/2} (\log n)^{1/2}. \tag{4.75}$$

Therefore, inequation (4.64) follows from (4.70) and (4.75) where constant C is equal to $2(C_0+2)/\sqrt{2\pi}$. $\square$

Lemma 4.11. *The class of functions $\mathcal{F}_n$ defined in (4.63)is a permissible class of functions.*

Proof. We will show that $\mathcal{F}_n$ can be indexed by a set T, which satisfies conditions from Definition 2.82. Let $T = \left\{\mathbf{b}, \mathbf{b_0} \in \mathbb{R}^k : \|\mathbf{b} - \mathbf{b_0}\| \leqslant M\right\} \otimes (0,1)$, where $M > 0$, be an index set equipped with the Lebesgue measure. It can be shown that T is a separable metric space by considering the balls with centers belonging to Q. The Borel σ-field is given by $\sigma(S) \otimes \mathcal{B}(0,1)$, where $\sigma(S)$ is σ-field generated by all closed k-dimensional spheres of radius M. Then, any function $f \in \mathcal{F}_n$ is $\mathcal{B}(\mathbb{R}^k) / \left(\sigma(S) \otimes \mathcal{B}(0,1)\right)$ measurable since it is a difference of indicator functions which are measurable. Furthermore, T is an analytic subset of the compact metric space $\overline{T}$ by using the fact that the σ-field generated by all Lebesgue measurable subsets of $\overline{T}$ coincides with its analytic sets. □

Lemma 4.12. *Let $\mathcal{F}_n$ be the class of functions defined in (4.63). Then, for each n and $\varepsilon > 0$, the uniform covering numbers of $\mathcal{F}_n$ satisfy*

$$\sup_Q N\left(\varepsilon, \mathcal{F}_n, L_1(Q)\right) \leqslant A\varepsilon^{-W} \tag{4.76}$$

with constants A, and W, not depending on n.

Proof. Let $\mathcal{F}^{'}$ and $\mathcal{F}^{''}$ be two classes of functions defined by

$$\mathcal{F}^{'} = \left\{ f_{\mathbf{b},p}, \mathbf{b} = \frac{1}{\sqrt{\mathbf{a}^{'}\mathbf{\Sigma}\mathbf{a}}}\mathbf{a} \in \mathbb{R}^k, \quad p \in (0,1) \right\}$$

and

$$\mathcal{F}^{''} = -\mathcal{F}^{'},$$

where $f_{\mathbf{b},p}$ is defined by (4.62). Since $\mathcal{F}_n \subset \mathcal{F}^{'} + \mathcal{F}^{''}$ then,

$$\sup_Q N\left(\varepsilon, \mathcal{F}_n, L_1(Q)\right) \leqslant \sup_Q N\left(\varepsilon, \mathcal{F}^{'} + \mathcal{F}^{''}, L_1(Q)\right).$$

Thus, it will be sufficient to show that

$$\sup_Q N\left(\varepsilon, \mathcal{F}^{'} + \mathcal{F}^{''}, L_1(Q)\right) \leqslant A\varepsilon^{-W}. \tag{4.77}$$

By using Lemma 2.72, $\mathcal{F}^{'}$ is a VC class and thus, $\mathcal{F}^{''}$ is also a VC class by Lemma 2.73. Therefore, by taking $r = 1$ and the envelope function identical to 1 in Lemma 2.70 we obtain

$$\sup_Q N\left(\varepsilon/2, \mathcal{F}^{'}, L_1(Q)\right) \leqslant A^{'}\left(\varepsilon/2\right)^{-W^{'}}, \tag{4.78}$$

and

$$\sup_Q N\left(\varepsilon/2, \mathcal{F}^{''}, L_1(Q)\right) \leqslant A^{''}\left(\varepsilon/2\right)^{-W^{''}}, \tag{4.79}$$

where $A^{'}$, $A^{''}$, $W^{'}$, and $W^{''}$ are independent of n. By Definition 2.65, for any $\varepsilon > 0$ there exist finite sets of functions $\{g_i^{'}\}$ and $\{g_j^{''}\}$, not necessarily in $\mathcal{F}^{'}$, $\mathcal{F}^{''}$, respectively, such that inequalities (4.78) and (4.79) can be rewritten

$$\sup_Q \min\left\{i \in \mathbb{N} : \min_i Q\left|f^{'} - g_i^{'}\right| \leqslant \varepsilon/2, f^{'} \in \mathcal{F}^{'}\right\} \leqslant A^{'}\varepsilon^{-W^{'}} \tag{4.80}$$

$$\sup_Q \min\left\{j \in \mathbb{N} : \min_j Q\left|f^{''} - g_j^{''}\right| \leqslant \varepsilon/2, f^{''} \in \mathcal{F}^{''}\right\} \leqslant A^{''}\varepsilon^{-W^{''}} \tag{4.81}$$

Then, for the following set of functions $g_{(i,j)} = g_i^{'} + g_j^{''}$ and for any $f \in \mathcal{F}^{'} + \mathcal{F}^{''}$, Q, and (i,j) we have

$$\min_{(i,j)} Q\left|f - g_{(i,j)}\right| \leqslant \min_i Q\left|f^{'} - g_i^{'}\right| + \min_j Q\left|f^{''} - g_j^{''}\right| \leqslant \varepsilon.$$

Thus, for any ε and any Q we can find a finite set of functions such that the union of the $L_1(Q)$ balls of radius ε centered at $f^{'} + f^{''}$ covers $\mathcal{F}^{'} + \mathcal{F}^{''}$ and

$$\begin{aligned}\sup_Q N\left(\varepsilon, \mathcal{F}^{'} + \mathcal{F}^{''}, L_1(Q)\right)&\\ \leqslant \sup_Q N\left(\varepsilon, \mathcal{F}^{'}, L_1(Q)\right) \times \sup_Q N\left(\varepsilon, \mathcal{F}^{''}, L_1(Q)\right) &\leqslant (1/2)^{-W^{'}-W^{''}} A^{'}A^{''}\varepsilon^{-W^{'}-W^{''}}.\end{aligned}$$

The proof is now complete. □

Lemma 4.13. *Let $\mathcal{F}_n$ be the class of functions defined by (4.63) and $\hat{\mathbf{b}}$ defined by (4.10) such that (4.18) is satisfied. Then,*

$$\sup_{p\in(0,1)} \left|U_n(p,\hat{\mathbf{b}}) - U_n(p)\right| = o_p(n^{-3/4}\log n) \quad as \quad n \to \infty. \tag{4.82}$$

Proof. We will equivalently show that $\forall\varepsilon > 0$, $\forall\delta > 0$, $\exists N_{\varepsilon,\delta} \in \mathbb{N}^*$ such that for all $n \geqslant N_{\varepsilon,\delta}$ we have

$$Pr\left(\sup_{p\in(0,1)} \left|U_n(p,\hat{\mathbf{b}}) - U_n(p)\right| > \varepsilon n^{-3/4}\log n\right) < \delta. \tag{4.83}$$

Let $\delta > 0$. Then, from (4.18), there exists $M \in (0,\infty)$ and $N_\delta \in \mathbb{N}^*$ such that

$$Pr\left(\sqrt{n}\left\|\hat{\mathbf{b}} - \mathbf{b_0}\right\| > M\right) < \delta/2, \quad \forall n \geqslant N_\delta. \tag{4.84}$$

Next, we will apply Theorem 2.83 to the class of functions $\mathcal{F}_n$. The class $\mathcal{F}_n$ is permissible by Lemma 4.11 and its uniform covering numbers satisfy (2.16) by Lemma 4.12. Let $\alpha_n = n^{-1/4}(\log n)^{1/2}$ be a non-increasing sequence of numbers for $n \geqslant 7$. Let $\delta_n^2 = Cn^{-1/2}(\log n)^{1/2}$ where the constant C is equal to $2(C_0+2)/\sqrt{2\pi}$. Recall, that $x_n \gg y_n$ if $x_n/y_n \to \infty$. Then, it can be easily verified that $n\delta_n^2\alpha_n \gg \log n$. According to Lemma 4.10, for any $f \in \mathcal{F}_n$, which has $|f| \leqslant 1$, we have $(Pf^2)^{1/2} \leqslant \sqrt{Cn^{-1/2}(\log n)^{1/2}}$. Hence, by Theorem 2.83 we obtain

$$\sup_{f\in\mathcal{F}_n} |\mathbb{P}_n f - Pf| \ll Cn^{-3/4}\log n \quad a.s,$$

which implies

$$\sup_{f\in\mathcal{F}_n} |\mathbb{P}_n f - Pf| = o_p\left(n^{-3/4}\log n\right). \tag{4.85}$$

For any $f \in \mathcal{F}_n$, by Remark 2.49 and (4.11), $\mathbb{P}_n f$ can be rewritten as

$$\mathbb{P}_n f = U_n(p, \hat{\mathbf{b}}) - U_n(p). \tag{4.86}$$

Similarly, for any $f \in \mathcal{F}_n$, by Remark 2.49 and the fact that $\Phi(\mathbf{b_0}'\mathbf{X})$ has a Uniform (0,1) distribution, Pf is

$$Pf = U(p) - U(p) = 0. \tag{4.87}$$

Hence, (4.85) can be equivalently rewritten as

$\forall\varepsilon > 0$, $\forall\delta > 0$, $\exists N_{\varepsilon,\delta} \in \mathbb{N}^*$ such that for all $n \geqslant N_{\varepsilon,\delta}$ we have

$$Pr\left(\sup_{p\in(0,1),\|\mathbf{b}-\mathbf{b_0}\|\leqslant M/\sqrt{n}} |U_n(p,\mathbf{b}) - U_n(p)| > \varepsilon n^{-3/4}\log n\right) < \delta/2. \tag{4.88}$$

The conclusion of the lemma follows immediately for $n \geqslant \max\{N_\delta, N_{\varepsilon,\delta}\}$ by using (4.88) and (4.84)

$$\begin{aligned} &Pr\left(\sup_{p\in(0,1)} \left|U_n(p,\hat{\mathbf{b}}) - U_n(p)\right| > \varepsilon n^{-3/4}\log n\right) \\ \leqslant\ & Pr\left(\sup_{p\in(0,1),\left\|\hat{\mathbf{b}}-\mathbf{b_0}\right\|\leqslant M/\sqrt{n}} \left|U_n(p,\hat{\mathbf{b}}) - U_n(p)\right| > \varepsilon n^{-3/4}\log n\right) \\ & + Pr\left(\left\|\hat{\mathbf{b}} - \mathbf{b_0}\right\| > M/\sqrt{n}\right) < \delta. \end{aligned}$$

□

Now, we are able to decompose the process given in (4.15) by using Lemma 2.75.

Lemma 4.14. *Let $\{\mathbf{X_i}\}_{i=1}^n$ and $\{\mathbf{Y_j}\}_{j=1}^m$ be random samples from multivariate normal distributions with mean vectors $\mathbf{0}$ and μ,respectively, and the same covariance matrix Σ. Let $\mathbf{a_0}$ be given by (4.2) and $\hat{\mathbf{a}}$ an estimator of $\mathbf{a_0}$ satisfying (4.3). Let $\mathbf{b_0}$ and $\hat{\mathbf{b}}$ be defined by (4.10). Then, for $m, n \in \mathbb{N}$ such that $m/n \to \lambda \in \mathbb{R}^+$*

$$\sqrt{m}\left(\tilde{G}_m\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right),\hat{\mathbf{b}}\right) - \tilde{G}\left(p,\mathbf{b_0}\right)\right) \tag{4.89}$$

$$= \sqrt{m}\left(\tilde{G}_m\left(p,\mathbf{b_0}\right) - \tilde{G}\left(p,\mathbf{b_0}\right)\right) \tag{4.90}$$

$$+\sqrt{m}\left(\tilde{G}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right),\hat{\mathbf{b}}\right) - \tilde{G}\left(p,\mathbf{b_0}\right)\right) \tag{4.91}$$

$$+o_p(1), \quad as \quad n \to \infty,$$

where $o_p(1)$ holds uniformly in $p \in (0,1)$.

Proof. Let $\mathcal{G}$ be the Q-Donsker class defined in (4.32). Let $\delta_1 > 0$, $\delta_2 > 0$, and $\|\cdot\|_\infty$ be the uniform norm on $(0,1)$. Then, define

$$\mathcal{G}_\delta = \{g_{\mathbf{b},H,p} - g_p \in \mathcal{G} : \|\mathbf{b} - \mathbf{b_0}\| \leqslant \delta_1, \|H - U\|_\infty \leqslant \delta_2\} \tag{4.92}$$

Then, $\mathcal{G}_\delta$ is a Q-Donsker class by Lemma 2.61. Let $\left\{g_{\hat{\mathbf{b}},U_n(\,\cdot\,,\hat{\mathbf{b}}),p}\right\}$ be a sequence of random functions that takes its values in $\mathcal{G}$. We will prove next that for any $\varepsilon > 0$

$$Pr\left(\sup_{p\in(0,1)}\left|(Q_m - Q)\left(g_{\hat{\mathbf{b}},U_n(\,\cdot\,,\hat{\mathbf{b}}),p}(\mathbf{Y}) - g_p(\mathbf{Y})\right)\right| > \varepsilon\right) \to 0, \quad \text{as} \quad n,m\to\infty. \tag{4.93}$$

Let us denote the event $\sup_{p\in(0,1)}\left|(Q_m - Q)\left(g_{\hat{\mathbf{b}},U_n(\,\cdot\,,\hat{\mathbf{b}}),p}(\mathbf{Y}) - g_p(\mathbf{Y})\right)\right| > \varepsilon$ by A. Then, notice that $Pr\,(A)$ can be majored by

$$Pr\left(A, \left\|\hat{\mathbf{b}} - \mathbf{b_0}\right\| \leqslant \delta_1 \cap \left\|U_n\left(\,\cdot\,,\hat{\mathbf{b}}\right) - U\right\|_\infty \leqslant \delta_2\right) \tag{4.94}$$

$$+Pr\left(\left\|\hat{\mathbf{b}} - \mathbf{b_0}\right\| > \delta_1 \cup \left\|U_n\left(\,\cdot\,,\hat{\mathbf{b}}\right) - U\right\|_\infty > \delta_2\right). \tag{4.95}$$

By triangle inequality, Lemma 4.13 and Glivenko-Cantelli Theorem we have

$$\left\|U_n\left(\,\cdot\,,\hat{\mathbf{b}}\right) - U\right\|_\infty = o_p(1), \quad \text{as} \quad n\to\infty. \tag{4.96}$$

Then, from (4.24) and (4.96) we have

$$\left(\hat{\mathbf{b}}, U_n\left(\,\cdot\,,\hat{\mathbf{b}}\right)\right) \xrightarrow{P} (\mathbf{b_0}, U)\,, \quad \text{as} \quad n\to\infty. \tag{4.97}$$

Moreover, for any $\delta_1, \delta_2 > 0$, (4.24) and (4.96) also implies

$$Pr\left(\left\|\hat{\mathbf{b}} - \mathbf{b_0}\right\| > \delta_1 \cup \left\|U_n\left(\,\cdot\,, \hat{\mathbf{b}}\right) - U\right\|_\infty > \delta_2\right) \to 0, \quad \text{as} \quad n \to \infty. \tag{4.98}$$

From Lemma 4.9 we obtain

$$\lim_{\mathbf{b}\to\mathbf{b_0}, H\to U} \sup_{p\in(0,1)} \int \left(g_{\mathbf{b},H,p}(\mathbf{y}) - g_p(\mathbf{y})\right)^2 dQ(\mathbf{y}) \to 0, \quad \text{as} \quad n, m \to \infty. \tag{4.99}$$

Hence, by applying Lemma 2.75 to class of functions $\mathcal{G}_\delta$ we obtain

$$Pr\left(A, \left\|\hat{\mathbf{b}} - \mathbf{b_0}\right\| \leqslant \delta_1, \left\|U_n\left(\,\cdot\,, \hat{\mathbf{b}}\right) - U\right\|_\infty \leqslant \delta_2\right) \to 0, \quad \text{as} \quad n, m \to \infty. \tag{4.100}$$

Then, from (4.100) and (4.98) we obtain (4.93), which implies

$$\sqrt{m}\left(Q_m - Q\right)\left(g_{\hat{\mathbf{b}}, U_n(\,\cdot\,,\hat{\mathbf{b}}), p}(\mathbf{Y}) - g_p(\mathbf{Y})\right) = o_p(1), \quad \text{as} \quad n, m \to \infty, \tag{4.101}$$

where $o_p(1)$ is uniformly in p. The conclusion of the lemma follows immediately by regrouping the following terms

$$\begin{aligned}
&Q_m \mathbf{I}\left[\Phi\left(\hat{\mathbf{b}}'\mathbf{Y}\right) \leqslant U_n^{-1}\left(p, \hat{\mathbf{b}}\right)\right] = \tilde{G}_m\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right), \hat{\mathbf{b}}\right), \\
&Q_m \mathbf{I}\left[\Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant p\right] = \tilde{G}_m\left(p, \mathbf{b_0}\right), \\
&Q \mathbf{I}\left[\Phi\left(\hat{\mathbf{b}}'\mathbf{Y}\right) \leqslant U_n^{-1}\left(p, \hat{\mathbf{b}}\right)\right] = \tilde{G}\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right), \hat{\mathbf{b}}\right), \quad \text{and} \\
&Q \mathbf{I}\left[\Phi\left(\mathbf{b_0}'\mathbf{Y}\right) \leqslant p\right] = \tilde{G}\left(p, \mathbf{b_0}\right).
\end{aligned}$$

□

4.3 Asymptotic Distribution of the Component Processes

Lemma 4.14 decomposed the equivalent generalized empirical process as the sum of two empirical processes. In this section we will find the asymptotic distribution of the empirical processes defined in (4.90) and (4.91). The following lemma gives us the asymptotic distribution of the empirical process defined in (4.90).

Lemma 4.15. *Let $p \in (0,1)$. Then, as $m \to \infty$,*

$$\sqrt{m}(\tilde{G}_m(p, \mathbf{b_0}) - \tilde{G}(p, \mathbf{b_0})) \rightsquigarrow \mathbb{G}_{\tilde{G}}(p), \quad in \quad D[0,1]. \tag{4.102}$$

Proof. The conclusion follows by applying, as in the univariate case, Theorem 2.62 to random variables $W_1, W_2, ..., W_m$, where $W_j = \mathbf{b_0}'\mathbf{Y_j}$. □

Next, we will focus on the process defined in (4.91), also called the drift term. But, before deriving its asymptotic distribution, we will prove a series of propositions and lemmas that will be used later. Note, that for any $p \in (0,1)$ the process in (4.91) can be equivalently written as

$$\sqrt{m}\left(\tilde{G}\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right), \hat{\mathbf{b}}\right) - \tilde{G}\left(p, \mathbf{b_0}\right)\right) =$$

$$\sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right)\right) - \hat{\mathbf{b}}'\mu\right) - \Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right)\right) - \mathbf{b_0}'\mu\right)\right) \tag{4.103}$$

$$+ \sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right)\right) - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p\right)\right) - \mathbf{b_0}'\mu\right)\right) \tag{4.104}$$

$$+ \sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p\right)\right) - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}\left(p\right) - \mathbf{b_0}'\mu\right)\right). \tag{4.105}$$

Next, we will show that the processes (4.103) and (4.104) are $o_p(1)$, as $n \to \infty$, uniformly in $p \in (0,1)$. Finally, we will show that the process (4.105) can be uniformly approximated as

$$\begin{aligned}&\sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right) - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right)\right)\\&= \sqrt{m}\frac{\phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right)}{\phi\left(\Phi^{-1}(p)\right)}\left(p - U_n(p)\right) + o_p(1),\end{aligned} \tag{4.106}$$

where $o_p(1)$ holds uniformly in $p \in (0,1)$. Therefore, the asymptotic distribution of the drift term will be given by the process in (4.106).

Lemma 4.16. *Let* $\hat{\mathbf{b}}$ *be defined by (4.10) such that (4.25) is satisfied. Then,*

$$\sup_{x\in\mathbb{R}} \sqrt{m}\left|\Phi\left(x - \hat{\mathbf{b}}'\mu\right) - \Phi\left(x - \mathbf{b_0}'\mu\right)\right| = o_p(1), \quad \textit{as} \quad n \to \infty. \tag{4.107}$$

Proof. By the first-order Taylor series approximation and the fact that the standard normal density ϕ is bounded by 1, we have

$$\sup_{x\in\mathbb{R}} \sqrt{m}\left|\Phi\left(x - \hat{\mathbf{b}}'\mu\right) - \Phi\left(x - \mathbf{b_0}'\mu\right)\right| < \sqrt{m}\left|\left(\hat{\mathbf{b}} - \mathbf{b_0}\right)'\mu\right| \tag{4.108}$$

The conclusion follows immediately from (4.108) and (4.25). □

Corollary 4.17. *Let* $\hat{\mathbf{b}}$ *be defined by (4.10) such that (4.25) is satisfied. Then, the process* $\sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right) - \hat{\mathbf{b}}'\mu\right) - \Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right) - \mathbf{b_0}'\mu\right)\right)$ *is* $o_p(1)$, *as* $n \to \infty$, *uniformly in* $p \in (0,1)$.

Proof. Let $p \in (0,1)$ and $x = U_n^{-1}\left(p,\hat{\mathbf{b}}\right) \in \mathbb{R}$. Then, conclusion follows immediately from Lemma (4.16). $\square$

The proof for process (4.104) will start with the Taylor series expansion as in the previous case.

Lemma 4.18. *For every $p \in (0,1)$ there exists a point between $U_n^{-1}\left(p, \hat{\mathbf{b}}\right)$ and $U_n^{-1}(p)$, denoted $\theta_n(p)$, such that*

$$\begin{aligned} &\sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right)\right) - \mathbf{b_0}'\mu\right) - \Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right) - \mathbf{b_0}'\mu\right)\right) \\ &= \sqrt{m}\frac{\phi\left(\Phi^{-1}(\theta_n(p)) - \mathbf{b_0}'\mu\right)}{\phi\left(\Phi^{-1}(\theta_n(p))\right)}\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right) - U_n^{-1}(p)\right). \end{aligned} \tag{4.109}$$

Proof. The result follows immediately from the first-order Taylor series expansion of the function $\Phi\left(\Phi^{-1}(\,\cdot\,) - \mathbf{b_0}'\mu\right)$. $\square$

Remark 4.19. From now on, the fact that $\theta_n(p)$ is between $U_n^{-1}\left(p, \hat{\mathbf{b}}\right)$ and $U_n^{-1}(p)$ will be denoted by

$$U_n^{-1}\left(p, \hat{\mathbf{b}}\right) \wedge U_n^{-1}(p) < \theta_n(p) < U_n^{-1}\left(p, \hat{\mathbf{b}}\right) \vee U_n^{-1}(p), \tag{4.110}$$

where, recall, $\wedge$ means minimum and $\vee$ means maximum.

Let $R_\phi : (0,1) \longrightarrow \mathbb{R}$ be defined as

$$R_\phi(p) = \frac{\phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right)}{\phi\left(\Phi^{-1}(p)\right)}. \tag{4.111}$$

Let $\delta \in (0, 1/4)$ and $q_\delta : [0,1] \longrightarrow [0,1]$ be defined as

$$q_\delta(p) = (1-p)^\delta. \tag{4.112}$$

Finally, let $\tilde{q}_\delta : (0,1) \longrightarrow \mathbb{R}$ be defined as

$$\tilde{q}_\delta(p) = q_\delta(p) R_\phi(p). \tag{4.113}$$

Note that for $p \in (0,1)$

$$R_\phi\left(\theta_n(p)\right)\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right) - U_n^{-1}(p)\right) = \tilde{q}_\delta\left(\theta_n(p)\right)\frac{q_\delta(p)}{q_\delta\left(\theta_n(p)\right)}\frac{U_n^{-1}\left(p, \hat{\mathbf{b}}\right) - U_n^{-1}(p)}{q_\delta(p)}. \tag{4.114}$$

We will show next, by using different techniques, that the process (4.109) is $o_p(1)$, as $n \to \infty$, uniformly in $p \in (0, p_n)$, $p \in (p_n, 1-1/n)$, and $p \in (1-1/n, 1)$, where p_n is a properly chosen sequence converging to one. Therefore, by combining these results, process (4.104) will be $o_p(1)$, as $n \to \infty$, uniformly in $p \in (0,1)$. First, we will prove prove some useful properties of the above introduced functions.

Proposition 4.20. *Let $p_0 \in (0,1)$ and R_ϕ be defined by (4.111). Then, R_ϕ is uniformly continuous on $[0, p_0]$.*

Proof. Notice that R_ϕ can be rewritten as

$$R_\phi(p) = e^{\frac{\mathbf{b}_0'\mu}{2}\left(2\Phi^{-1}(p) - \mathbf{b}_0'\mu\right)}, \quad p \in (0,1).$$

Since $\mathbf{b_0'}\mu > 0$, it can be easily shown that R_ϕ is monotonically increasing with $\lim_{p\to 0} R_\phi(p) = 0$ and $\lim_{p\to 1} R_\phi(p) = \infty$. Therefore, for any $p_0 \in (0,1)$, R_ϕ is uniformly continuous on the interval $[0, p_0]$, since it is continuous on the same interval.

□

Proposition 4.21. *Let $\tilde{q}_\delta$ be defined by (4.113). Then, $\tilde{q}_\delta$ is uniformly continuous on* $[0,1]$.

Proof. Notice, that $\tilde{q}_\delta$ is continuous on $(0,1)$. We will show that $\tilde{q}_\delta$ is continuous on the compact interval $[0,1]$, and therefore uniformly continuous on $[0,1]$, by proving that

$$\lim_{p\to 0} \tilde{q}_\delta(p) = \lim_{p\to 1} \tilde{q}_\delta(p) = 0. \tag{4.115}$$

The limit of $\tilde{q}_\delta$ for p converging to zero is immediate from Proposition 4.20 and (4.112). Let $p > 1/2$ and let $x > 0$ be the unique value such that $x = \Phi^{-1}(p)$. Then, by simple algebraic manipulations we have

$$\tilde{q}_\delta(p) = (1-\Phi(x))^\delta \, \frac{\phi\left(x - \mathbf{b_0'}\mu\right)}{\phi(x)} = \left(\frac{1-\Phi(x)}{e^{-\frac{\mathbf{b_0'}\mu}{\delta}x}}\right) e^{-\frac{\left(\mathbf{b_0'}\mu\right)^2}{2}}. \tag{4.116}$$

Since $\mathbf{b_0'}\mu > 0$, then by l'Hopital rule we have

$$\lim_{x\to\infty} \left(\frac{1-\Phi(x)}{e^{-\frac{\mathbf{b_0'}\mu}{\delta}x}}\right) = \lim_{x\to\infty} \frac{\phi(x)}{\frac{\mathbf{b_0'}\mu}{\delta} e^{-\frac{\mathbf{b_0'}\mu}{\delta}x}} = 0. \tag{4.117}$$

The proof is complete since by the change of variable the limit of $\tilde{q}_\delta$ for p converging to one is the same as the limit for x converging to infinity. □

Remark 4.22. By using (4.116), it can be shown that $\tilde{q}_\delta$ is strictly decreasing for $p \geqslant \Phi\left(\frac{\mathbf{b_0}'\mu}{\delta}\right)$.

We can work next on the terms in the right hand side of equality (4.114). We will prove that the first two terms are $O_p(1)$ and the third term is $o_p(1)$, as $n \to \infty$, uniformly in $p \in (0, p_n)$. We will start by proving lemmas that will help us to show that the supremum for $p \in (0,1)$ of the absolute value of term $\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right) - U_n^{-1}(p)\right)$ from (4.109) can be made $o_p\left(n^{-3/4}\log n\right)$. For $i = \overline{1,n}$, let

$$\Sigma^{-1/2}\mathbf{X}_i = \mathbf{Z}_i = \left(Z_{i1}, \ldots, Z_{ik}\right)'. \tag{4.118}$$

Notice that due to the independence of $\mathbf{X}_i$ and (4.118) then $\{Z_{ij}\}_{i=\overline{1,n},j=\overline{1,k}}$ are independent standard normal random variables.

Lemma 4.23. *Let $\mathbf{b} \in \mathbb{R}^k$. Then, the following inequality is true for any $n \in \mathbb{N}^+$*

$$\sup_{p\in(0,1)} \left|U_n^{-1}(p, \mathbf{b}) - U_n^{-1}(p)\right| \leqslant \sqrt{\frac{k}{2\pi}}\, \|\mathbf{b} - \mathbf{b_0}\| \left\|\Sigma^{1/2}\right\| \max_{i=\overline{1,n},j=\overline{1,k}} |Z_{ij}|, \tag{4.119}$$

where $\{Z_{ij}\}$ are defined in (4.118).

Proof. For $i = \overline{1,n}$, denote $\Phi\left(\mathbf{b}'\mathbf{X_i}\right)$ and $\Phi\left(\mathbf{b_0}'\mathbf{X_i}\right)$ by ζ_i and ξ_i, respectively. By definition of an empirical quantile function, let $\zeta_{n:i}$ and $\xi_{n:i}$ be the i^{th} ordered ζ_i and ξ_i value, respectively. Then, for $i = \overline{1,n}$ we have

$$U_n^{-1}(p, \mathbf{b}) = \zeta_{n:i}, \quad \frac{i-1}{n} < p \leqslant \frac{i}{n}, \tag{4.120}$$

and

$$U_n^{-1}(p) = \xi_{n:i}, \quad \frac{i-1}{n} < p \leqslant \frac{i}{n}. \tag{4.121}$$

Notice that it can be shown

$$\max_{i=\overline{1,n}} |\zeta_{n:i} - \xi_{n:i}| \leqslant \max_{i=\overline{1,n}} |\zeta_i - \xi_i| \,. \tag{4.122}$$

From (4.118) we can easily obtain

$$\max_{i=\overline{1,n}} \|\mathbf{Z}_i\| \leqslant \sqrt{k} \max_{i=\overline{1,n}, j=\overline{1,k}} |Z_{ij}| \tag{4.123}$$

Then, for any $n \in \mathbb{N}^+$, by using (4.122), definitions of ζ_i and ξ_i, and first-order Taylor series expansion, we have

$$\begin{aligned} &\sup_{p\in(0,1)} \left|U_n^{-1}(p, \mathbf{b}) - U_n^{-1}(p)\right| \\ &= \max_{i=\overline{1,n}} \sup_{p\in(\frac{i-1}{n}, \leqslant \frac{i}{n}]} \left|U_n^{-1}(p, \mathbf{b}) - U_n^{-1}(p)\right| \\ &= \max_{i=\overline{1,n}} |\zeta_{n:i} - \xi_{n:i}| \\ &\leqslant \frac{1}{\sqrt{2\pi}} \max_{i=\overline{1,n}} \left|(\mathbf{b} - \mathbf{b_0})' \mathbf{X}_i\right|. \end{aligned} \tag{4.124}$$

The conclusion of the lemma follows immediately by noticing that $(\mathbf{b} - \mathbf{b_0})' \mathbf{X}_i$ is equal to $(\mathbf{b} - \mathbf{b_0})' \Sigma^{1/2}\Sigma^{-1/2}\mathbf{X}_i$ and by using triangle inequality and (4.123) in (4.124).

□

For any $p \in (0,1)$ and any $n \in \mathbb{N}$ let $\Delta_n(p)$ be defined as

$$\Delta_n(p) = U_n\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right)\right) - U_n\left(U_n^{-1}(p)\right) - U_n^{-1}\left(p, \hat{\mathbf{b}}\right) + U_n^{-1}(p). \tag{4.125}$$

Lemma 4.24. *Let $\hat{\mathbf{b}}$ be defined by (4.10) such that (4.18) is satisfied. Then,*

$$\sup_{p\in(0,1)} |\Delta_n(p)| = o_p\left(n^{-3/4}(\log n)^{3/4}\beta_n\right), \quad \textit{as} \quad n \to \infty, \tag{4.126}$$

where β_n is any increasing sequence with $n^{-1/4}(\log n)^{1/4}\beta_n$ non-increasing.

Proof. Let $\varepsilon > 0$ be given. We will prove that, for given $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that for all $n \geqslant N_\varepsilon$ we have

$$Pr\left(n^{3/4}(\log n)^{-3/4}\beta_n \sup_{p\in(0,1)} |\Delta_n(p)| > \varepsilon\right) < \varepsilon. \tag{4.127}$$

Notice that $\max_{i=\overline{1,n},j=\overline{1,k}} |Z_{ij}| = O_p\left((\log n)^{1/2}\right)$, as $n \to \infty$, by Proposition 2.55. Therefore, by using this result and (4.18) we have

$$T_n = \frac{\sqrt{n}}{(\log n)^{1/2}}\sqrt{\frac{k}{2\pi}}\left\|\hat{\mathbf{b}} - \mathbf{b_0}\right\|\left\|\Sigma^{1/2}\right\| \max_{i=\overline{1,n},j=\overline{1,k}} |Z_{ij}| = O_p(1), \quad \text{as} \quad n \to \infty,$$

or, equivalently, for given $\varepsilon > 0$ there exists N_ε and C_0 such that

$$Pr\left(T_n > C_0\right) < \varepsilon/2, \quad \forall n \geqslant N_\varepsilon \tag{4.128}$$

Hence, by rewriting (4.119) from Lemma 4.23 as

$$\sup_{p\in(0,1)} \left|U_n^{-1}(p,\mathbf{b}) - U_n^{-1}(p)\right| \leqslant n^{-1/2}(\log n)^{1/2}T_n$$

and, by using the same technique of splitting probabilities, we have

$$\begin{aligned} & Pr\left(n^{3/4}(\log n)^{-3/4}\beta_n^{-1}\sup_{p\in(0,1)}|\Delta_n(p)| > \varepsilon\right) \\ \leqslant\; & Pr\left(n^{3/4}(\log n)^{-3/4}\beta_n^{-1}\sup_{p\in(0,1)}|\Delta_n(p)| > \varepsilon, T_n \leqslant C_0\right) \qquad (4.129) \\ & + Pr\left(T_n > C_0\right). \qquad (4.130) \end{aligned}$$

But,

$$\begin{aligned} & Pr\left(n^{3/4}(\log n)^{-3/4}\beta_n^{-1}\sup_{p\in(0,1)}|\Delta_n(p)| > \varepsilon, T_n \leqslant C_0\right) \\ = \; & Pr\left(n^{3/4}(\log n)^{-3/4}\beta_n^{-1}\sup_{|s-t|\leqslant C_0 n^{-1/2}(\log n)^{1/2}}|U_n(t) - U_n(s) - (t-s)| > \varepsilon\right). \end{aligned}$$

By using Theorem 2.83, we can show that the probability term of the right hand side of the above equality can be made less than $\varepsilon/2$ for n sufficiently large. Let $\mathcal{F}_n$ be the following class of functions

$$\mathcal{F}_n = \left\{\mathbf{I}\left[s < U \leqslant t\right]: \; 0 < s \leqslant t < 1, \; |t-s| \leqslant C_0 n^{-1/2}(\log n)^{1/2}\right\}, \qquad (4.131)$$

where $U \sim Unif(0,1)$. It can be shown that $\mathcal{F}_n$ is a permissible class of functions such that for any n and $\varepsilon > 0$, $\sup_Q N\left(\varepsilon, \mathcal{F}_n, L_1(Q)\right) \leqslant A\varepsilon^W$, where A, W do not depend on n. The proofs are very similar to those of Lemmas 4.11 and 4.12 and they will be omitted. Moreover, it can be easily seen that for any $f \in \mathcal{F}_n$ we have $|f| \leqslant 1$ and $Pf^2 \leqslant C_0 n^{-1/2}(\log n)^{1/2}$. Hence, let δ_n and α_n be two sequences such that $\delta_n^2 = C_0 n^{-1/2}(\log n)^{1/2}$ and $\alpha_n = n^{-1/4}(\log n)^{1/4}\beta_n$ is a non-increasing sequence of numbers with $\beta_n \nearrow \infty$. Notice that $n\delta_n^2\alpha_n^2 \ll \log n$. Therefore, by applying Theorem 2.83 to $\mathcal{F}_n$ we obtain

$$\sup_{f\in\mathcal{F}_n} |\mathbb{P}_n f - Pf| \ll C_0 n^{-3/4}(\log n)^{3/4}\beta_n, \quad \text{a.s,}$$

which implies

$$\sup_{f\in\mathcal{F}_n} |\mathbb{P}_n f - Pf| = o_p\left(n^{-3/4}(\log n)^{3/4}\beta_n\right).$$

But, since $\mathbb{P}_n f = U_n(t) - U_n(s)$ and $Pf = t - s$, then

$$\begin{aligned}
&\sup_{f\in\mathcal{F}_n} |\mathbb{P}_n f - Pf| \\
&\quad = \sup_{|t-s|\leqslant C_0 n^{-1/2}(\log n)^{1/2}} |U_n(t) - U_n(s) - (t-s)| \\
&\quad = o_p\left(n^{-3/4}(\log n)^{3/4}\beta_n\right),
\end{aligned}$$

or, equivalently,

$$Pr\left(n^{3/4}(\log n)^{-3/4}\beta_n^{-1} \sup_{|t-s|\leqslant C_0 n^{-1/2}(\log n)^{1/2}} |U_n(t) - U_n(s) - (t-s)| > \varepsilon\right) < \varepsilon/2. \tag{4.132}$$

The conclusion of the lemma follows immediately from (4.128) and (4.132). □

Corollary 4.25. *Let* $\hat{\mathbf{b}}$ *be defined by (4.10) such that (4.18) is satisfied. Then,*

$$\sup_{p\in(0,1)} \left| U_n^{-1}\left(p,\hat{\mathbf{b}}\right) - U_n^{-1}(p) \right| = o_p\left(n^{-3/4}\log n\right), \quad as \quad n \to \infty. \tag{4.133}$$

Proof. Let $p \in (0,1)$. Notice that

$$\begin{aligned} & U_n\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right),\hat{\mathbf{b}}\right) - U_n\left(U_n^{-1}(p)\right) \qquad (4.134) \\ &= U_n\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right),\hat{\mathbf{b}}\right) - U_n\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right) + U_n\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right) - U_n\left(U_n^{-1}(p)\right). \end{aligned}$$

For ease of presentation will introduce some further notations. For any $p \in (0,1)$ and any $n \in \mathbb{N}$, let $\Delta_n(p)$ be defined in (4.125), $\Delta_n^{'}(p)$, and $\Delta_n^{''}(p)$ defined as follows

$$U_n\left(p,\hat{\mathbf{b}}\right) - U_n(p) = \Delta_n^{'}(p), \tag{4.135}$$

and

$$F_n\left(F_n^{-1}(p)\right) = p + \Delta_n^{''}(p), \tag{4.136}$$

where F_n is any empirical distribution function and $\Delta_n^{''}(p) = O\left(n^{-1}\right)$. By using (4.136) in the left hand side of equality (4.134), and (4.125), (4.135) in the right hand side of the same equality (4.134) we obtain, after some algebraic manipulations,

$$U_n^{-1}\left(p,\hat{\mathbf{b}}\right) - U_n^{-1}(p) = \Delta_n(p) + \Delta_n^{'}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right) + \Delta_n^{''}(p). \tag{4.137}$$

Also, notice that

$$\Delta''_n(p) = o\left(n^{-3/4}\log n\right) \tag{4.138}$$

The conclusion follows by taking supremum after $p \in (0,1)$ in (4.137), by noticing that $U_n^{-1}\left(p, \hat{\mathbf{b}}\right) \in (0,1)$, by using (4.138), Lemma 4.24 with $\beta_n = (\log n)^{1/4}$, Lemma 4.13, and by simple stochastic calculus. □

Lemma 4.26. *Let $\delta \in (0, 1/4)$. Then,*

$$\sup_{(0,1-1/n]} \sqrt{n}\left|\frac{U_n^{-1}\left(p,\hat{\mathbf{b}}\right) - U_n^{-1}(p)}{q_\delta(p)}\right| = o_p\left(n^{\delta-1/4}\log n\right), \quad \textit{as} \quad n \to \infty. \tag{4.139}$$

Proof. For $\delta \in (0, 1/4)$, by monotonicity of q_δ, $\sup_{p\in(0,1-1/n)} q_\delta(p) \geqslant n^{-\delta}$, and Corollary 4.25 we have

$$\begin{aligned}
&\sup_{(0,1-1/n]} \sqrt{n}\left|\frac{U_n^{-1}\left(p,\hat{\mathbf{b}}\right) - U_n^{-1}(p)}{q_\delta(p)}\right| \\
\leqslant\ & n^{\delta-1/4}\log n \sup_{(0,1-1/n]} \frac{\left|U_n^{-1}\left(p,\hat{\mathbf{b}}\right) - U_n^{-1}(p)\right|}{n^{-3/4}\log n} = o_p(1), \quad \text{as} \quad n \to \infty.
\end{aligned}$$

□

We will now introduce two important lemmas that will be a very useful tools for the next proofs.

Lemma 4.27. *Let $s \geqslant 1$, $\tau > 0$, $0 \leqslant a < b \leqslant 1$ such that $(1-a)/\tau < 1$, and F be a* *distribution function on* $[0,1]$*. Then,*

$$\sup_{p\in(a,b)} \frac{(1-p)^s}{1-F^{-1}(p)} \leqslant \tau \quad \textit{iff} \quad \sup_{p\in(a,b)} \frac{1-F\left(1-\frac{(1-p)^s}{\tau}\right)}{1-p} \leqslant 1. \tag{4.140}$$

Proof. Notice that by using the following equivalence $\sup f(p) \leqslant t$ iff $f(p) \leqslant t, \forall p$ we have

$$\sup_{p\in(a,b)} \frac{(1-p)^s}{1-F^{-1}(p)} \leqslant \tau \quad \text{iff} \quad \frac{(1-p)^s}{1-F^{-1}(p)} \leqslant \tau,\ \forall p \in (a,b), \tag{4.141}$$

and

$$\sup_{p\in(a,b)} \frac{1-F\left(1-\frac{(1-p)^s}{\tau}\right)}{1-p} \leqslant 1 \quad \text{iff} \quad \frac{1-F\left(1-\frac{(1-p)^s}{\tau}\right)}{1-p} \leqslant 1,\ \forall p \in (a,b). \tag{4.142}$$

For any $p, x \in (0,1)$, by Lemma 2.31 we have

$$1-F(x) \leqslant 1-p \quad \text{iff} \quad 1-x \leqslant 1-F^{-1}(p). \tag{4.143}$$

Notice that $x = 1 - \frac{(1-p)^s}{\tau} \in (0,1)$ for any $p \in (a,b) \subseteq (0,1)$. Therefore, by using (4.143), for any $p \in (a,b)$ we have

$$\frac{(1-p)^s}{\tau} \leqslant 1-F^{-1}(p) \quad \text{iff} \quad \frac{1-F\left(1-\frac{(1-p)^s}{\tau}\right)}{1-p} \leqslant 1. \tag{4.144}$$

The conclusion follows immediately from (4.141), (4.142), and (4.144). □

Remark 4.28. If F is an empirical distribution function than (4.140) becomes

$$Pr\left(\sup_{p\in(a,b)}\frac{(1-p)^s}{1-F^{-1}(p)}>\tau\right)=Pr\left(\sup_{p\in(a,b)}\frac{1-F\left(1-\frac{(1-p)^s}{\tau}\right)}{1-p}>1\right). \tag{4.145}$$

Lemma 4.29. *Let $c>1$. Then,*

$$\lim_{x\to0^+}\frac{x^{c^2}}{1-\Phi\left(c\Phi^{-1}\left(1-x\right)\right)}\longrightarrow 0. \tag{4.146}$$

Proof. If the limit exists, by using l'Hopital's Rule, the fact that $\phi(ct)/\phi(t) = (\sqrt{2\pi}\phi(t))^{c^2-1}$, and Mill's Ratio $t(1-\Phi(t))<\phi(t)$, $\forall t>0$, we have

$$\begin{aligned}
&\lim_{x\to0^+}\frac{x^{c^2}}{1-\Phi\left(c\Phi^{-1}\left(1-x\right)\right)}\\
&= \lim_{x\to0^+}\frac{c^2\left(x^{c^2-1}\right)}{\frac{c\phi(c\Phi^1(1-x))}{\phi(\Phi^1(1-x))}}\\
&= \lim_{x\to0^+}\frac{c}{\sqrt{2\pi}^{c^2}}\left(\frac{1-\Phi\left(\Phi^{-1}\left(1-x\right)\right)}{\phi\left(\Phi^{-1}\left(1-x\right)\right)}\right)^{c^2-1}\\
&< \lim_{x\to0^+}\frac{c}{\sqrt{2\pi}^{c^2}}\left(\frac{1}{\Phi^{-1}\left(1-x\right)}\right)^{c^2-1}=0.
\end{aligned}$$

□

Corollary 4.30. *For $x>0$, sufficiently close to zero, and $c>1$, we have*

$$1-\Phi\left(\frac{\Phi^{-1}\left(1-x^{c^2}\right)}{c}\right)<x,$$

which can be equivalently written as

$$1 - \Phi\left(\frac{\Phi^{-1}(1-x)}{c}\right) < x^{1/c^2}. \tag{4.147}$$

Proof. The conclusion follows immediately from by simple manipulations of (4.146) from Lema 4.29. □

Lemma 4.31. *Let $\delta \in (0, 1/4)$. Then,*

$$\sup_{p\in(0,1)} \frac{q_\delta(p)}{q_\delta\left(U_n^{-1}(p)\right)} = O_p(1), \quad as \quad n \to \infty. \tag{4.148}$$

Proof. Let $\varepsilon > 0$. We will show that for ε given there exists $M_\varepsilon \in (0, \infty)$ such that for n sufficiently large we have

$$Pr\left(\sup_{p\in(0,1)} \frac{1-p}{1-U_n^{-1}(p)} > M_\varepsilon\right) < \varepsilon. \tag{4.149}$$

By setting $s = 1$, $\tau > 1$, $a = 0$, $b = 1$, and the uniform empirical distribution U_n in Remark (4.28), identity (4.145) becomes

$$Pr\left(\sup_{p\in(0,1)} \frac{(1-p)}{1-U_n^{-1}(p)} > \tau\right) = Pr\left(\sup_{p\in(0,1)} \frac{1-U_n\left(1-\frac{1-p}{\tau}\right)}{\frac{1-p}{\tau}} > \tau\right). \tag{4.150}$$

Note that by the symmetry and absolute continuity of the uniform distribution and the definition of indicator function, we have

$$\{U_n(t), t \in [0,1]\} \stackrel{\mathcal{D}}{=} \{1 - U_n(1-t), t \in [0,1]\} \tag{4.151}$$

Hence, from (4.150) and (4.151) and by using change of variable $t = (1-p)/\tau$ we obtain

$$\begin{aligned} &Pr\left(\sup_{p\in(0,1)} \frac{(1-p)}{1-U_n^{-1}(p)} > \tau\right) \\ &= \quad Pr\left(\sup_{p\in(0,1)} \frac{U_n\left(\frac{1-p}{\tau}\right)}{\frac{1-p}{\tau}} > \tau\right) \leqslant Pr\left(\sup_{t\in(0,1/\tau)} \frac{U_n(t)}{t} > \tau\right) \end{aligned} \tag{4.152}$$

By choosing $M_\varepsilon > \max\{1, e/\varepsilon\}$ then, from Lemma (2.54), the right hand side of (4.152) can be made less than ε for n sufficiently large. Thus, the proof is complete.

□

Let p_n be be a sequence converging to one defined by

$$p_n = 1 - n^{-3/4}\log n. \tag{4.153}$$

Lemma 4.32. *Let* $\hat{\mathbf{b}}$ *defined by (4.10) such that (4.18) is satisfied,* $\delta \in (0, 1/4)$, $\theta_n(p)$ *be defined by (4.110), and* p_n *be defined by (4.153). Then,*

$$\sup_{p\in(0,p_n)} \frac{q_\delta(p)}{q_\delta\left(\theta_n(p)\right)} = O_p(1), \quad \text{as} \quad n \to \infty. \tag{4.154}$$

Proof. Let $\varepsilon > 0$. By monotonicity of q_δ and definition of $\theta_n(p)$ we have

$$\sup_{p\in(0,p_n)} \frac{q_\delta(p)}{q_\delta\left(\theta_n(p)\right)} \leqslant \sup_{p\in(0,p_n)} \frac{q_\delta(p)}{q_\delta\left(U_n^{-1}(p)\right)} + \sup_{p\in(0,p_n)} \frac{q_\delta(p)}{q_\delta\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right)\right)}. \tag{4.155}$$

Since the first term on the right hand side of inequality (4.155) is $O_p(1)$ as $n \to \infty$

by Lemma 4.31, it will be sufficient to show that

$$\sup_{p\in(0,p_n)} \frac{q_\delta(p)}{q_\delta\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right)} = O_p(1), \quad \text{as} \quad n\to\infty, \tag{4.156}$$

or, equivalently, for ε given there exists $M_\varepsilon \in (0,\infty)$ such that for n sufficiently large we have

$$Pr\left(\sup_{p\in(0,p_n)} \frac{1-p}{1-U_n^{-1}\left(p,\hat{\mathbf{b}}\right)} > M_\varepsilon\right) < \varepsilon. \tag{4.157}$$

Again, by setting $s=1$, $\tau>1$, $a=0$, $b=p_n$, the uniform empirical distribution U_n in Remark (4.28), and identity (4.151) we have

$$\begin{aligned}
& Pr\left(\sup_{p\in(0,p_n)} \frac{1-p}{1-U_n^{-1}\left(p,\hat{\mathbf{b}}\right)} > \tau\right) \\
\leqslant \;& Pr\left(\sup_{p\in(0,p_n)} \frac{U_n\left(\frac{1-p}{\tau}\right)}{\frac{1-p}{\tau}} > \frac{\tau}{2}\right) \\
& + Pr\left(\sup_{p\in(0,p_n)} \frac{\left|U_n\left(\frac{1-p}{\tau},\hat{\mathbf{b}}\right) - U_n\left(\frac{1-p}{\tau}\right)\right|}{\frac{1-p}{\tau}} > \frac{\tau}{2}\right) \\
\leqslant \;& Pr\left(\sup_{t\in(1/n\tau,1)} \frac{U_n(t)}{t} > \frac{\tau}{2}\right) \qquad (4.158) \\
& + Pr\left(\sup_{t\in(0,1)} \frac{\left|U_n\left(t,\hat{\mathbf{b}}\right) - U_n(t)\right|}{n^{-3/4}\log n} > \frac{1}{2}\right) \qquad (4.159)
\end{aligned}$$

By choosing $M_\varepsilon > \max\{2, 2e/\varepsilon\}$ then, from Lemma (2.54), probability in (4.158) can be made less than $\varepsilon/2$ for n sufficiently large. By Lemma (4.13), probability in (4.159) can also be made less than $\varepsilon/2$ for n sufficiently large. The conclusion of the lemma follows immediately. □

We can now put together the previous results and have the following lemma.

Lemma 4.33. *Let* $\hat{\mathbf{b}}$ *defined by (4.10) such that (4.18) is satisfied,* $\theta_n(p)$ *be defined by (4.110), and* p_n *be defined by (4.153) Then,*

$$\sup_{p\in(0,p_n)} \sqrt{m}\left|R_\phi\left(\theta_n(p)\right)\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)-U_n^{-1}(p)\right)\right| = o_p(1), \quad \text{as} \quad n\to\infty. \tag{4.160}$$

Proof. Let $\delta \in (0, 1/4)$. Then from (4.114) we have

$$\sup_{p\in(0,p_n)} \sqrt{m}\left|R_\phi\left(\theta_n(p)\right)\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)-U_n^{-1}(p)\right)\right|$$

$$\leqslant \sup_{p\in(0,1)} \left|\tilde{q}_\delta\left(\theta_n(p)\right)\right| \tag{4.161}$$

$$\cdot \sup_{p\in(0,p_n)} \left|\frac{q_\delta(p)}{q_\delta\left(\theta_n(p)\right)}\right| \tag{4.162}$$

$$\cdot \sup_{p\in(0,1/n]} \sqrt{m}\left|\frac{U_n^{-1}\left(p,\hat{\mathbf{b}}\right)-U_n^{-1}(p)}{q_\delta(p)}\right|. \tag{4.163}$$

Since $\tilde{q}_\delta$ is uniformly continuous on $(0, 1)$ by Proposition 4.21, then supremum in (4.161) is $O_p(1)$, as $n \to \infty$. Supremum in (4.162) is also $O_p(1)$, as $n \to \infty$, by Lemma 4.32. Finally, supremum in (4.163) is $o_p(1)$, as $n \to \infty$, by Lemma 4.26. The conclusion of the lemma follows from stochastic calculus. □

Note that supremum in (4.162) could be proven to be $O_p(1)$, as $n \to \infty$, only for $p \in (0, p_n)$. In order to show that (4.160) is true when $p \in (p_n, 1 - 1/n)$, we will need to write the process in (4.109) in a slightly different, but important, manner. Let $\delta_1, \delta_2 \in (0, 1/4)$ be such that $\delta_2 = s\delta_1$, where $s \geqslant 1$ is a proportionality factor, whose magnitude will be determined later. Then, note

$$R_\phi\left(\theta_n(p)\right)\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)-U_n^{-1}(p)\right)=\tilde{q}_{\delta_1}\left(\theta_n(p)\right)\frac{q_{\delta_2}(p)}{q_{\delta_1}\left(\theta_n(p)\right)}\frac{U_n^{-1}\left(p,\hat{\mathbf{b}}\right)-U_n^{-1}(p)}{q_{\delta_2}(p)}. \tag{4.164}$$

Hence, by following the same steps and also using some of the results from the case $p\in(0,p_n)$, all we need to show is that

$$\sup_{p\in(p_n,1-1/n)}\frac{q_{\delta_2}(p)}{q_{\delta_1}\left(\theta_n(p)\right)}=O_p(1),\quad \text{as} \quad n\to\infty. \tag{4.165}$$

By using definitions of $\theta_n(p)$ and of function q_δ we have

$$\begin{aligned}
&\sup_{p\in(p_n,1-1/n)}\frac{q_{\delta_2}(p)}{q_{\delta_1}\left(\theta_n(p)\right)}\\
&= \sup_{p\in(p_n,1-1/n)}\left(\frac{(1-p)^s}{1-\theta_n(p)}\right)^{\delta_1}\\
&\leqslant \sup_{p\in(p_n,1-1/n)}\left(\frac{(1-p)^s}{1-U_n^{-1}(p)}\right)^{\delta_1}\\
&\quad + \sup_{p\in(p_n,1-1/n)}\left(\frac{(1-p)^s}{1-U_n^{-1}\left(p,\hat{\mathbf{b}}\right)}\right)^{\delta_1}.
\end{aligned}$$

Notice that (4.165) is true if we prove that the supremums from the right hand side of the above inequality are $O_p(1)$ as $n\to\infty$.

Lemma 4.34. *Let p_n be defined by (4.153) and $s\geqslant 1$. Then,*

$$\sup_{p\in(p_n,1-1/n)}\frac{(1-p)^s}{1-U_n^{-1}(p)}=O_p(1),\quad \textit{as} \quad n\to\infty. \tag{4.166}$$

Proof. Let $\varepsilon > 0$ and $M_\varepsilon = \max\{1, e/\varepsilon\}$ be given. Notice that

$$Pr\left(\sup_{p\in(p_n,1-1/n)} \frac{(1-p)^s}{1-U_n^{-1}(p)} > M_\varepsilon\right) \leqslant Pr\left(\sup_{p\in(p_n,1-1/n)} \frac{(1-p)}{1-U_n^{-1}(p)} > M_\varepsilon\right).$$

Hence, the conclusion of the lemma follows immediately by using the same arguments as in Lemma 4.31, so they will be omitted. □

Proposition 4.35. *Let $s \geqslant 7$. Then,*

$$n\left(1-\Phi\left(\frac{\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)}{2}\right)\right) \longrightarrow 0, \quad as \quad n\to\infty. \tag{4.167}$$

Proof. The conclusion of the lemma follows immediately by using inequality (4.147) with $c = 2$ and $x = n^{-3/4}\log n$ substituted into (4.147). □

Lemma 4.36. *Let $\hat{\mathbf{b}}$ be defined by (4.10), p_n be defined by (4.153), and $s \geqslant 7$. Then,*

$$\sup_{p\in(p_n,1-1/n)} \frac{(1-p)^s}{1-U_n^{-1}\left(p,\hat{\mathbf{b}}\right)} = O_p(1), \quad as \quad n\to\infty. \tag{4.168}$$

Proof. Let $\varepsilon > 0$ and $M_\varepsilon > 1$ be given. Notice that by using Lemma (4.27), monotonicity of $U_n\left(\cdot,\hat{\mathbf{b}}\right)$ given in (4.11), and $\{\mathbf{Z}_i\}$ defined in (4.118) we have the following sequence of inequalities

$$
\begin{aligned}
&Pr\left(\sup_{p\in(p_n,1-1/n)}\frac{(1-p)^s}{1-U_n^{-1}\left(p,\hat{\mathbf{b}}\right)}>M_\varepsilon\right)\\
\leqslant\;& Pr\left(n\sup_{p\in(p_n,1-1/n)}\left(1-U_n\left(1-\frac{1-p}{M_\varepsilon}(1-p)^{s-1},\hat{\mathbf{b}}\right)\right)>1\right)\\
\leqslant\;& Pr\left(n\left(1-U_n\left(1-\left(n^{-3/4}\log n\right)^{s-1},\hat{\mathbf{b}}\right)\right)>1\right)\\
=\;& Pr\left(\sum_{i=1}^n\mathbf{I}\left[\hat{\mathbf{b}}'\mathbf{X_i}>\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)\right]>1\right)\\
\leqslant\;& Pr\left(\sup_{\mathbf{b}\in\mathbb{R}^k,\mathbf{b}=\frac{\mathbf{a}}{\sqrt{\hat{\mathbf{a}}'\boldsymbol{\Sigma}\hat{\mathbf{a}}}}}\sum_{i=1}^n\mathbf{I}\left[\mathbf{b}'\mathbf{X_i}>\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)\right]>1\right)\\
\leqslant\;& Pr\left(\sup_{\mathbf{c}\in\mathbb{R}^k,\|\mathbf{c}\|=1}\sum_{i=1}^n\mathbf{I}\left[\mathbf{c}'\mathbf{Z_i}>\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)\right]>1\right)\\
\leqslant\;& Pr\left(\sum_{i=1}^n\sup_{\mathbf{c}\in\mathbb{R}^k,\|\mathbf{c}\|=1}\mathbf{I}\left[\mathbf{c}'\mathbf{Z_i}>\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)\right]>1\right),
\end{aligned}
\tag{4.169}
$$

where $\mathbf{c}=\boldsymbol{\Sigma}^{\mathbf{1/2}}\mathbf{b}$. We will show next, by mathematical induction, that probability in (4.169) can be made less than ε. Let $k=1$. Then, for $s\geqslant 7$ and n sufficiently large, by (4.167) we have

$$
\begin{aligned}
&\mathbf{E}\left(\sum_{i=1}^n\sup_{|c|=1}\mathbf{I}\left[c'Z_i>\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)\right]\right)\\
\leqslant\;&\sum_{i=1}^n\mathbf{E}\left(\mathbf{I}\left[|Z_i|>\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)\right]\right)\\
\leqslant\;& 2n\left(1-\Phi\left(\frac{\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)}{2}\right)\right)=o(1).
\end{aligned}
$$

Thus, we proved that $\sum_{i=1}^n\sup_{|c|=1}\mathbf{I}\left[c'Z_i>\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)\right]$ converges to zero in L^1, which in turn, implies convergence to zero in probability . Therefore, probability in (4.169) can be made less than ε for n sufficiently large. Now, by using induction, assume that probability in (4.169) is less than ε for $1, 2, \ldots, k-1$ and prove this is also true for k. Notice that for any $c \in \mathbb{R}^k$ such that $\|c\| = 1$ we have $\sqrt{c_1^2 + \ldots + c_{k-1}^2} \leqslant 1$, $\forall k = \overline{1, n}$. Therefore, by using triangle inequality and this fact we have

$$\begin{aligned}
&\sup_{\mathbf{c}\in\mathbb{R}^k, \|\mathbf{c}\|=1} \mathbf{I}\left[\mathbf{c}^{'}\mathbf{Z_i} > \Phi^{-1}\left(1 - \left(n^{-3/4}\log n\right)^{s-1}\right)\right] \\
&\leqslant \sup_{\mathbf{c}\in\mathbb{R}^k, \|\mathbf{c}\|=1} \mathbf{I}\left[c_1 Z_{i1} + \ldots + c_{k-1} Z_{i(k-1)} > \Phi^{-1}\left(1 - \left(n^{-3/4}\log n\right)^{s-1}\right)\right] \\
&\quad + \sup_{\mathbf{c}\in\mathbb{R}^k, \|\mathbf{c}\|=1} \mathbf{I}\left[c_k Z_{ik} > \Phi^{-1}\left(1 - \left(n^{-3/4}\log n\right)^{s-1}\right)\right] \\
&\leqslant \sup_{\mathbf{c}\in\mathbb{R}^k, \|\mathbf{c}\|=1} \mathbf{I}\left[\frac{c_1 Z_{i1} + \ldots + c_{k-1} Z_{i(k-1)}}{\sqrt{c_1^2 + \ldots + c_{k-1}^2}} > \Phi^{-1}\left(1 - \left(n^{-3/4}\log n\right)^{s-1}\right)\right] \\
&\quad + \sup_{\mathbf{c}\in\mathbb{R}^k, \|\mathbf{c}\|=1} \mathbf{I}\left[\frac{c_k Z_{ik}}{\sqrt{c_k^2}} > \Phi^{-1}\left(1 - \left(n^{-3/4}\log n\right)^{s-1}\right)\right] \\
&\leqslant \sup_{\mathbf{d}\in\mathbb{R}^{k-1}, \|\mathbf{d}\|=1} \mathbf{I}\left[\mathbf{d}^{'}\mathbf{Z_i} > \frac{\Phi^{-1}\left(1 - \left(n^{-3/4}\log n\right)^{s-1}\right)}{2}\right] \\
&\quad + \sup_{e\in\mathbb{R}, |e|=1} \mathbf{I}\left[e Z_{ik} > \frac{\Phi^{-1}\left(1 - \left(n^{-3/4}\log n\right)^{s-1}\right)}{2}\right].
\end{aligned}$$

Hence, by using the above inequalities and induction hypothesis, we have

$$
\begin{aligned}
& Pr\left(\sum_{i=1}^{n} \sup_{\mathbf{c}\in\mathbb{R}^k, \|\mathbf{c}\|=1} \mathbf{I}\left[\mathbf{c}^{'}\mathbf{Z_i} > \Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)\right] > 1\right) \\
\leqslant \; & Pr\left(\sum_{i=1}^{n} \sup_{\mathbf{d}\in\mathbb{R}^{k-1}, \|\mathbf{d}\|=1} \mathbf{I}\left[\mathbf{d}^{'}\mathbf{Z_i} > \frac{\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)}{2}\right] > \frac{1}{2}\right) \\
& + Pr\left(\sum_{i=1}^{n} \sup_{e\in\mathbb{R}, |e|=1} \mathbf{I}\left[eZ_{ik} > \frac{\Phi^{-1}\left(1-\left(n^{-3/4}\log n\right)^{s-1}\right)}{2}\right] > \frac{1}{2}\right) \\
\leqslant \; & \varepsilon.
\end{aligned}
$$

Thus, lemma is proved. □

Lemma 4.37. *Let $\hat{\mathbf{b}}$ defined by (4.10) such that (4.18) is satisfied, $\theta_n(p)$ be defined by (4.110), and p_n be defined by (4.153). Then,*

$$
\sup_{p\in(p_n, 1-1/n)} \sqrt{m}\left|R_\phi\left(\theta_n(p)\right)\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right) - U_n^{-1}(p)\right)\right| = o_p(1), \quad as \quad n\to\infty. \tag{4.170}
$$

Proof. Immediate by plugging in (4.164) the following results: uniform continuity of $\tilde{q}_\delta$, Lemmas 4.34, 4.36, and 4.26. □

Finally, we can focus on proving that the process (4.104) is $o_p(1)$ uniformly on $p \in (1-1/n, 1)$. Once again, for the interval $(1-1/n, 1)$ we will have to write the process (4.104) in other equivalent ways. First notice that we can also rewrite (4.104) as follows

$$\sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right)-\mathbf{b}_\mathbf{0}^{'}\mu\right)-\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p\right)\right)-\mathbf{b}_\mathbf{0}^{'}\mu\right)\right)$$
$$=\sqrt{m}\left(\left(1-\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right)-\mathbf{b}_\mathbf{0}^{'}\mu\right)\right)-\left(1-\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p\right)\right)-\mathbf{b}_\mathbf{0}^{'}\mu\right)\right)\right). \tag{4.171}$$

Note that for $p \in (1-1/n, 1)$, $U_n^{-1}(p) = U_{n:n}$, where $U_{n:n}$ is the maximum order statistics from $Unif(0,1)$ random sample.

Lemma 4.38. *Let $U_{n:n}$ be the maximum order statistics of a $Unif(0,1)$ random sample. Then,*

$$\sqrt{n}\left(1-\Phi\left(\Phi^{-1}\left(U_{n:n}\right)-\mathbf{b}_\mathbf{0}^{'}\mu\right)\right)=o(1) \quad as \quad n\to\infty. \tag{4.172}$$

Proof. By Lemma 2.53, for $Z_{n:n} = \Phi^{-1}\left(U_{n:n}\right)$,

$$\lim_{n\to\infty}\frac{Z_{n:n}}{(2\log n)^{1/2}}=1, \quad \text{a.s.}$$

Choose $\varepsilon > 0$ and $c > 0$, such that $\sqrt{2}(1-\varepsilon)-c > 1$ and, for n sufficiently large,

$$\frac{Z_{n:n}}{(2\log n)^{1/2}} > 1-\varepsilon.$$

Note that for n sufficiently large we also have

$$c(\log n)^{1/2} > \mathbf{b}_\mathbf{0}^{'}\mu.$$

Therefore, by using Mill's Ratio we have

$$\begin{aligned}
&\sqrt{n}\left(1-\Phi\left(Z_{n:n}-\mathbf{b_0}'\mu\right)\right)\\
&< \sqrt{n}\left(1-\Phi\left(\sqrt{2}(1-\varepsilon)(\log n)^{1/2}-\mathbf{b_0}'\mu\right)\right)\\
&\leqslant \sqrt{n}\left(1-\Phi\left(\left(\sqrt{2}(1-\varepsilon)-c\right)(\log n)^{1/2}\right)\right)\\
&< \frac{1}{\sqrt{2}(1-\varepsilon)-c}\left(\frac{n}{\log n}e^{-\left(\sqrt{2}(1-\varepsilon)-c\right)^2\log n}\right)^{1/2} = o(1).
\end{aligned}$$

$\square$

Lemma 4.39. *Let* $\hat{\mathbf{b}}$ *be defined by (4.10) such that (4.18) is satisfied. Then, as* $n \to \infty$,

$$\sup_{p\in(1-1/n,1)} \sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right)-\mathbf{b_0}'\mu\right)-\Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right)-\mathbf{b_0}'\mu\right)\right) = o_p(1). \tag{4.173}$$

Proof. Note that process (4.104) can be re-written as in (4.171). First, suppose that $U_n^{-1}\left(p,\hat{\mathbf{b}}\right) > U_n^{-1}(p)$. Then, by monotonicity of functions Φ and Φ^{-1} and the fact that $\mathbf{b_0}'\mu > 0$, we obtain

$$\sqrt{m}\left(1-\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right)-\mathbf{b_0}'\mu\right)\right) < \sqrt{m}\left(1-\Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right)-\mathbf{b_0}'\mu\right)\right). \tag{4.174}$$

Next, suppose $U_n^{-1}\left(p,\hat{\mathbf{b}}\right) \leqslant U_n^{-1}(p)$. Since $\theta_n(p) \leqslant U_n^{-1}(p)$ then, by definition of q_δ, we have $q_\delta(\theta_n(p)) \geqslant q_\delta(U_n^{-1}(p))$. Then, for $\delta < 1/4$, by using the equivalent process

(4.109), we obtain

$$\left|\sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)-\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p\right)\right)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)\right)\right|$$
$$\leqslant \tilde{q}_\delta\left(\theta_n(p)\right)\frac{1}{q_\delta\left(U_n^{-1}(p)\right)}\left|U_n^{-1}(p,\hat{\mathbf{b}})-U_n^{-1}(p)\right|.$$

Note that for $p\in(1-1/n,1)$ we have

$$\frac{1}{q_\delta\left(U_n^{-1}(p)\right)}=\left(\frac{1}{1-U_n^{-1}(p)}\right)^\delta=n^\delta\left(\frac{1}{n\left(1-U_{n:n}\right)}\right)^\delta.$$

Therefore,

$$\sup_{p\in(1-1/n,1)}\sqrt{m}\left|\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)-\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p\right)\right)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)\right)\right|$$
$$\leqslant \max\left\{2\sup_{p\in(1-1/n,1)}\sqrt{m}\left(1-\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p\right)\right)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)\right),\right.$$
$$\left.\sup_{p\in(1-1/n,1)}\sqrt{m}\tilde{q}_\delta\left(\theta_n(p)\right)\left(\frac{1}{n\left(1-U_{n:n}\right)}\right)^\delta n^\delta\left|U_n^{-1}(p,\hat{\mathbf{b}})-U_n^{-1}(p)\right|\right\}.$$

Since the first term in the above inequality is $o_p(1)$ by Lemma 4.39, we only need to show that the second term is also $o_p(1)$. Choose $M>0$. Note, that given a random sample from $Unif(0,1)$, then the j^{th} order statistics has a $Beta(j,n-j+1)$ distribution. Moreover, due to the symmetry of the beta distribution we have

$\{U_{n:1}(t), t \in [0,1]\} \stackrel{\mathcal{D}}{=} \{1 - U_{n:n}(t), t \in [0,1]\}$. Therefore,

$$\begin{aligned}
Pr\left(\frac{1}{n\left(1-U_{n:n}\right)} > M\right) &= Pr\left(\frac{M^{-1}}{n} > U_{n:1}\right) \\
&= \sum_{j=1}^{n} \binom{n}{k} \left(\frac{M^{-1}}{n}\right)^{j} \left(1 - \frac{M^{-1}}{n}\right)^{n-j} \\
&= 1 - \left(1 - \frac{M^{-1}}{n}\right)^{n} \longrightarrow 1 - e^{-M^{-1}}, \quad \text{as} \quad n \to \infty.
\end{aligned}$$

Hence, by choosing M large we can make $Pr\left(\frac{1}{n(1-U_{n:n})} > M\right)$ arbitrarily small. Finally, notice that by using Corollary 4.25 we have

$$\sup_{p\in(1-1/n,1)} n^{1/2+\delta} \left|U_n^{-1}(p,\hat{\mathbf{b}}) - U_n^{-1}(p)\right| = o_p\left(n^{-1/4+\delta}\log n\right) = o_p(1), \quad \text{as} \quad n \to \infty.$$

Therefore, by using Proposition 4.21 and the previous two results we have proved

$$\sup_{p\in(1-1/n,1)} \sqrt{m}\tilde{q}_\delta\left(\theta_n(p)\right)\left(\frac{1}{n\left(1-U_{n:n}\right)}\right)^{\delta} n^{\delta}\left|U_n^{-1}(p,\hat{\mathbf{b}}) - U_n^{-1}(p)\right| = o_p(1),$$

as $n \to \infty$. □

Lemma 4.40. *Let* $\hat{\mathbf{b}}$ *defined by (4.10) such that (4.18) is satisfied. Then, as* $n \to \infty$*,*

$$\sup_{p\in(0,1)} \sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right) - \mathbf{b}_{\mathbf{0}}'\mu\right) - \Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p\right)\right) - \mathbf{b}_{\mathbf{0}}'\mu\right)\right) = o_p(1). \tag{4.175}$$

Proof. Let p_n be defined by (4.153) and $\theta_n(p)$ be defined by (4.110) such that the process (4.104) can be equivalently written as (4.109). Notice that

$$\sup_{p\in(0,1)} \sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right)-\mathbf{b_0}'\mu\right)-\Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right)-\mathbf{b_0}'\mu\right)\right)$$

$$\leqslant \max\left\{\sup_{p\in(0,p_n)} \sqrt{m}\left|R_\phi\left(\theta_n(p)\right)\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)-U_n^{-1}(p)\right)\right|,\right.$$

$$\sup_{p\in(p_n,1-1/n)} \sqrt{m}\left|R_\phi\left(\theta_n(p)\right)\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)-U_n^{-1}(p)\right)\right|,$$

$$\left.\sup_{p\in(1-1/n,1)} \sqrt{m}\left|\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}\left(p,\hat{\mathbf{b}}\right)\right)-\mathbf{b_0}'\mu\right)-\Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right)-\mathbf{b_0}'\mu\right)\right)\right|\right\}.$$

The conclusion follows immediately by using Lemmas 4.33, 4.37, 4.39. □

Next, we will focus on the process given in (4.105) and prove that it uniformly approximated by the process in (4.106).

Lemma 4.41. *For every $p \in (0,1)$, there exists $\tilde{\theta}_n(p)$ such that*

$$\begin{aligned}&\sqrt{m}\left(\Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right)-\mathbf{b_0}'\mu\right)-\Phi\left(\Phi^{-1}(p)-\mathbf{b_0}'\mu\right)\right)\\ &= \sqrt{m}R_\phi\left(\tilde{\theta}_n(p)\right)\left(U_n^{-1}(p)-p\right),\end{aligned} \tag{4.176}$$

where

$$U_n^{-1}(p)\wedge p < \tilde{\theta}_n(p) < U_n^{-1}(p)\vee p. \tag{4.177}$$

Proof. Immediate by applying the first-order Taylor series expansion to function $\Phi\left(\Phi^{-1}(\,\cdot\,)-\mathbf{b_0}'\mu\right)$. □

Note that for any $p \in (0,1)$, by using Lemma 2.50 (Bahadur's Theorem) for the

Uniform distribution, we have

$$\begin{aligned}
&\sqrt{m}\left(R_\phi\left(\tilde{\theta}_n(p)\right)\left(U_n^{-1}(p)-p\right)-R_\phi(p)\left(p-U_n(p)\right)\right) \\
&= \sqrt{m}\left(R_\phi\left(\tilde{\theta}_n(p)\right)-R_\phi(p)\right)(p-U_n(p)) \qquad (4.178)\\
&\quad +\sqrt{m}R_\phi\left(\tilde{\theta}_n(p)\right)R_n(p), \qquad (4.179)
\end{aligned}$$

where R_n is the remainder term introduced in (2.2). Therefore, by using Lemma 4.41 we will actually need to show that terms (4.178) and (4.179) are $o_p(1)$ uniformly in $p \in (0, 1-1/n)$, as $n \to \infty$. Similarly, we will show that we can choose $p_0 \in (0,1)$ such that the terms mentioned before are $o_p(1)$ uniformly in $(0, p_0]$ and $(p_0, 1-1/n)$, as $n \to \infty$. Finally, we will combine these results. But first, we will prove other useful lemmas.

Lemma 4.42. *Let $U_1, \ldots, U_n$ be iid Unif(0,1) random variables. Then, almost surely*

$$\sup_{p\in[0,1]}\left|U_n^{-1}(p)-p\right| = o(1), \quad as \quad n \to \infty. \qquad (4.180)$$

Proof. We will prove that $\sup_{p\in[0,1]}|U_n^{-1}(p)-p| = \sup_{p\in[0,1]}|U_n(p)-p|$. This will be sufficient to conclude the lemma since we know that the right hand side of previous equality is $o(1)$ by Glivenko-Cantelli theorem. Recall, that by the definition of an empirical distribution we have

$$U_n^{-1}(p) = U_{n:i}, \quad \frac{i-1}{n} < p \leqslant \frac{i}{n} \qquad (4.181)$$

where $U_{n:i}$ is the i^{th} order statistics and $i = \overline{1,n}$. For $p = 0$ define $U_n^{-1}(p) = U_{n:0} = 0$. Hence,

$$\sup_{p\in[0,1]} \left|U_n^{-1}(p) - p\right| = \max_{i=\overline{1,n}} \sup_{p\in(\frac{i-1}{n},\frac{i}{n}]} |U_{n:i} - p| = \max_{i=\overline{1,n}} \max\left\{\left|U_{n:i} - \frac{i-1}{n}\right|, \left|U_{n:i} - \frac{i}{n}\right|\right\}. \tag{4.182}$$

On the other hand, U_n can be equivalently rewritten as

$$U_n(p) = \frac{i}{n}, \quad U_{n:i} \leqslant p < Un_{n:i+1} \tag{4.183}$$

for $i = \overline{0,n}$. Therefore,

$$\sup_{p\in[0,1]} |U_n(p) - p| = \max_{i=\overline{0,n}} \sup_{p\in[U_{n:i},U_{n:i+1})} \left|\frac{i}{n} - p\right| = \max_{i=\overline{0,n}} \max\left\{\left|\frac{i}{n} - U_{n:i}\right|, \left|\frac{i}{n} - U_{n:i+1}\right|\right\}. \tag{4.184}$$

Since the sets for which we find the maximum in (4.182) and (4.184) are the same, we conclude that the supremum are the same. □

Lemma 4.43. *Let $\tilde{\theta}_n(p)$ be given by (4.177). Then,*

$$\sup_{p\in(0,1)} \left|\tilde{\theta}_n(p) - p\right| = o_p(1), \quad as \quad n \to \infty. \tag{4.185}$$

Proof. By using the definition of $\tilde{\theta}_n(p)$ and triangle inequality we have

$$\sup_{p\in(0,1)} \left|\tilde{\theta}_n(p) - p\right| \leqslant \sup_{p\in(0,1)} \left|U_n^{-1}\left(p, \hat{\mathbf{b}}\right) - U_n^{-1}(p)\right| + 2 \sup_{p\in(0,1)} \left|U_n^{-1}(p) - p\right|.$$

The conclusion is immediate from Corollary 4.25, Lemma 4.43, and stochastic calculus. □

Lemma 4.44. *Let $\tilde{\theta}_n(p)$ be given by (4.177) and $p_0 \in (0,1)$. Then,*

$$\sup_{p\in(0,p_0]} \left| R_\phi\left(\tilde{\theta}_n(p)\right) - R_\phi(p)\right| = o_p(1), \quad as \quad n \to \infty. \tag{4.186}$$

Proof. Let $\varepsilon > 0$ be given. By Proposition 4.20, R_ϕ is uniformly continuous on the interval $[0, p_0]$. Thus, ε given, there exists $\delta > 0$, $p_0 + \delta < 1$, such that

$$\forall p, p' \in [0, p_0] : \sup_{p\in[0,p_0]} |p' - p| < \delta \Rightarrow \sup_{p\in[0,p_0]} |R_\phi(p') - R_\phi(p)| < \varepsilon. \tag{4.187}$$

Notice that

$$\begin{aligned}
&Pr\left(\sup_{p\in[0,p_0]} \left|R_\phi\left(\tilde{\theta}_n(p)\right) - R_\phi(p)\right| > \varepsilon\right) \\
&= Pr\left(\sup_{p\in[0,p_0]} \left|R_\phi\left(\tilde{\theta}_n(p)\right) - R_\phi(p)\right| > \varepsilon, \sup_{p\in[0,p_0]} |\hat{p} - p| < \delta\right) \\
&\quad +Pr\left(\sup_{p\in[0,p_0]} |\hat{p} - p| > \delta\right) \\
&\leqslant Pr\left(\sup_{p,p'\in[0,p_0+\delta]} |R_\phi(p') - R_\phi(p)| > \varepsilon, \sup |p' - p| < \delta\right) \\
&\quad +Pr\left(\sup_{p\in[0,p_0]} \left|\tilde{\theta}_n(p) - p\right| > \delta\right)
\end{aligned}$$

The conclusion follows from (4.187) and Lemma 4.43. □

Lemma 4.45. *Let $\delta \in (0, 1/4)$ and $\tilde{\theta}_n(p)$ be given by (4.177). Then*

$$\sup_{p\in(0,1)} \frac{q_\delta(p)}{q_\delta\left(\tilde{\theta}_n(p)\right)} = O_p(1), \quad as \quad n \to \infty. \tag{4.188}$$

Proof. By using definition of $\tilde{\theta}_n(p)$ and monotonicity of q_δ, we have

$$\sup_{p\in(0,1)} \frac{q_\delta(p)}{q_\delta\left(\tilde{\theta}_n(p)\right)} \leqslant \sup_{p\in(0,1)} \frac{q_\delta(p)}{q_\delta\left(U_n^{-1}(p)\right)} + 1. \tag{4.189}$$

The conclusion follows immediately by applying Lemma 4.31. □

Lemma 4.46. *Let $\delta \in (0, 1/4)$ and R_n be the residual term, as given in (2.2), from Bahadur's theorem applied to Uniform distribution. Then,*

$$\sup_{p\in(0,1-1/n]} \sqrt{n}\left|\frac{R_n(p)}{q_\delta(p)}\right| = o_p(1), \quad as \quad n \to \infty. \tag{4.190}$$

Proof. Let $R_n^* = \sup_{p\in(0,1)} |R_n(p)|$. By using the monotonicity of q_δ we have

$$\sup_{p\in(0,1-1/n]} \sqrt{n}\left|\frac{R_n(p)}{q_\delta(p)}\right| \leqslant n^{\delta-1/4}(\log n)^{1/2}\frac{R_n^*}{n^{-3/4}(\log n)^{1/2}}.$$

The conclusion follows immediately by using Remark 2.52, and stochastic calculus in the above inequality. □

Lemma 4.47. *Let $\tilde{\theta}_n(p)$ be given by (4.177). Then*

$$\sup_{p\in(0,1-1/n)} \sqrt{m}\left|\left(R_\phi\left(\tilde{\theta}_n(p)\right) - R_\phi(p)\right)(p - U_n(p))\right| = o_p(1), \quad as \quad n \to \infty. \tag{4.191}$$

Proof. Let $\varepsilon > 0$ and choose $\delta \in (0, 1/4)$. Notice that for any $p_0 \in (0, 1 - 1/n)$ and n sufficiently large, we have

$$\sup_{p\in(0,1-1/n)} \sqrt{m}\left|\left(R_\phi\left(\tilde{\theta}_n(p)\right) - R_\phi(p)\right)(p - U_n(p))\right|$$

$$\leqslant \max\Big\{\sup_{p\in(0,p_0)} \sqrt{m}\left|\left(R_\phi\left(\tilde{\theta}_n(p)\right) - R_\phi(p)\right)(p - U_n(p))\right|, \tag{4.192}$$

$$\sup_{p\in(p_0,1-1/n)} \sqrt{m}R_\phi\left(\tilde{\theta}_n(p)\right)|p - U_n(p)| \tag{4.193}$$

$$+ \sup_{p\in(p_0,1-1/n)} \sqrt{m}R_\phi(p)\,|p - U_n(p)|\Big\} \tag{4.194}$$

We will prove that all three terms are $o_p(1)$ as $n \to \infty$. Notice that for any choice of $p_0 \in (0, 1 - 1/n)$ supremum in (4.192) is $o_p(1)$, as $n \to \infty$, by Lemma 4.44 and boundness of the uniform empirical process.

Next, consider the process in (4.194). Note that for any any $p_0 > \Phi\left(\frac{\mathbf{b_0'}\mu}{\delta}\right)$ and n sufficiently large, by Remark 4.22 we have

$$\sup_{p\in(p_0,1-1/n)} \sqrt{m}R_\phi(p)\,|p - U_n(p)| \leqslant \tilde{q}_\delta(p_0)\sqrt{m}\sup_{p\in(0,1)}\left|\frac{p - U_n(p)}{q_\delta(p)}\right|. \tag{4.195}$$

Therefore, we will only need to show that there exists p_0 depending on both ε, δ such that

$$Pr\left(\tilde{q}_\delta(p_0)\sqrt{m}\sup_{p\in(0,1)}\left|\frac{p - U_n(p)}{q_\delta(p)}\right| > \varepsilon\right) < \varepsilon, \quad \text{as} \quad n \to \infty. \tag{4.196}$$

By using Lemma 2.53 applied to the Uniform distribution, for ε, δ given, there exists

$M_1 \in (0, \infty)$, depending on both ε, δ, such that for n sufficiently large

$$Pr\left(\sup_{p\in(0,1)} \sqrt{n}\left|\frac{p - U_n(p)}{q_\delta(p)}\right| > M_1\right) < \varepsilon/2. \tag{4.197}$$

Let us choose $p_0 \in (0, 1)$ such that

$$p_0 > \Phi\left(\frac{\mathbf{b_0'}\mu}{\delta}\right) \quad \text{and} \quad \tilde{q}_\delta(p_0) < \frac{\varepsilon}{M_1}. \tag{4.198}$$

Then, (4.196) is true by choosing p_0 that satisfies (4.198), by using (4.197), and by probability manipulations using the splitting probability technique.

For the process in (4.193) consider two cases. First suppose that $U_n^{-1}(p) \geqslant p$, which implies $\tilde{\theta}_n(p) \geqslant p$. For any $p_0 > \Phi\left(\frac{\mathbf{b_0'}\mu}{\delta}\right)$, since $\tilde{\theta}_n(p) \geqslant p \geqslant p_0$, then, by Remark 4.22, we have

$$\sup_{p\in(p_0,1-1/n)} \tilde{q}_\delta(\tilde{\theta}_n(p)) \leqslant \tilde{q}_\delta(p_0). \tag{4.199}$$

Thus, for n sufficiently large,

$$\begin{aligned}&\sup_{p\in(p_0,1-1/n)} \sqrt{m}R_\phi\left(\tilde{\theta}_n(p)\right)|p - U_n(p)| \\ &\quad\leqslant \tilde{q}_\delta(p_0) \sup_{p\in(0,1-1/n)} \frac{q_\delta(p)}{q_\delta(\tilde{\theta}_n(p))}\sqrt{m}\sup_{p\in(0,1)}\left|\frac{p - U_n(p)}{q_\delta(p)}\right|.\end{aligned}$$

Therefore, it will be sufficient to show that there exists p_0 depending on ε, δ such that, as $n \to \infty$,

$$Pr\left(\tilde{q}_\delta(p_0) \sup_{p\in(0,1-1/n)} \frac{q_\delta(p)}{q_\delta(\tilde{\theta}_n(p))}\sqrt{m}\sup_{p\in(0,1)}\left|\frac{p - U_n(p)}{q_\delta(p)}\right| > \varepsilon\right) < \varepsilon. \tag{4.200}$$

By (4.188), for ε, δ given, there exists $M_2 \in (0, \infty)$, depending on both ε, δ, such that for n sufficiently large

$$Pr\left(\sup_{p\in(0,1-1/n)} \frac{q_\delta(p)}{q_\delta\left(\tilde{\theta}_n(p)\right)} > M_2\right) < \varepsilon/2. \tag{4.201}$$

Then, let us choose $p_0 \in (0, 1)$ such that

$$p_0 > \Phi\left(\frac{\mathbf{b_0'}\mu}{\delta}\right) \quad \text{and} \quad \tilde{q}_\delta(p_0) < \frac{\varepsilon}{M_1 M_2}. \tag{4.202}$$

Hence, (4.200) is true by choosing p_0 that satisfies (4.202), by using (4.201), (4.197), and by probability manipulations using the splitting probability technique.

Secondly, suppose $U_n^{-1}(p) \leqslant p$, which implies $\tilde{\theta}_n(p) \leqslant p$. Similarly, for any $p_0 > \Phi\left(\frac{\mathbf{b_0'}\mu}{\delta}\right)$ and n sufficiently large, by monotonicity of R_ϕ and of $\tilde{q}_\delta$ for $p \geqslant p_0$ we obtain

$$\sup_{p\in(p_0,1-1/n)} \sqrt{m}R_\phi\left(\tilde{\theta}_n(p)\right)|p - U_n(p)| \leqslant \tilde{q}_\delta(p_0)\sqrt{m} \sup_{p\in(0,1)} \left|\frac{p - U_n(p)}{q_\delta(p)}\right|.$$

Thus, the proof will be identical to that of the process (4.194).

Therefore, by choosing $p_0 \in (0, 1)$ such that both (4.198) and (4.202) are satisfied, then both terms in (4.193) and (4.194) are $o_p(1)$ as $n \to \infty$. Thus, lemma is proved.

□

Lemma 4.48. *Let $\tilde{\theta}_n(p)$ be given by (4.177). Then*

$$\sup_{p\in(0,1-1/n)} \sqrt{m}R_\phi\left(\tilde{\theta}_n(p)\right)|R_n(p)| = o_p(1), \quad as \quad n\to\infty. \tag{4.203}$$

Proof. The proof is similar to the proof of previous lemma. However, we will be able to use simple stochastic calculus instead of an ε-δ type of proof. Let $\delta\in(0,1/4)$. For any $p_0\in(0,1-1/n)$ and n sufficiently large, we have

$$\sup_{p\in(0,1-1/n)} \sqrt{m}R_\phi\left(\tilde{\theta}_n(p)\right)|R_n(p)|$$
$$\leqslant \max\Big\{\sup_{p\in(0,p_0)} \sqrt{m}R_\phi\left(\tilde{\theta}_n(p)\right)|R_n(p)|, \tag{4.204}$$
$$\sup_{p\in(p_0,1-1/n)} \sqrt{m}R_\phi\left(\tilde{\theta}_n(p)\right)|R_n(p)|\Big\} \tag{4.205}$$

Let $R_n^* = \sup_{p\in(0,1)}|R_n(p)|$. Notice, that

$$\sup_{p\in(0,p_0)} \sqrt{m}R_\phi\left(\tilde{\theta}_n(p)\right)|R_n(p)| \leqslant R_\phi(p_0+\delta_1)\sqrt{m}n^{-3/4}(\log n)^{1/2}\frac{R_n^*}{n^{-3/4}(\log n)^{1/2}},$$

where δ_1 is chosen such that $\sup\left|\tilde{\theta}_n(p)-p\right|<\delta_1$ and $p_0+\delta_1<1$. Therefore, supremum in (4.204) is $o_p(1)$, as $n\to\infty$, by using Proposition 4.20, Remark 2.52, and the fact that the sequence in m,n is converging to zero, as $n\to\infty$. Thus, we only have to prove that supremum in (4.205) is $o_p(1)$ as $n\to\infty$.

We, again, distinguish the following two cases. First, suppose $U_n^{-1}(p)\geqslant p$, which implies $\tilde{\theta}_n(p)\geqslant p$. For any any $p_0>\Phi\left(\frac{\mathbf{b}_0'\mu}{\delta}\right)$ and n sufficiently large, by (4.199) we have

$$\begin{aligned}
&\sup_{p\in(p_0,1-1/n)} \sqrt{m} R_\phi\left(\tilde{\theta}_n(p)\right)|R_n(p)| \\
&\leqslant \tilde{q}_\delta(p_0) \sup_{p\in(0,1)} \frac{q_\delta(p)}{q_\delta(\tilde{\theta}_n(p))} \sup_{p\in(0,1-1/n]} \sqrt{m}\left|\frac{R_n(p)}{q_\delta(p)}\right|.
\end{aligned} \tag{4.206}$$

The right hand side of (4.206) is $o_p(1)$ as $n\to\infty$ by Proposition 4.21, Lemmas 4.45 and 4.46, and stochastic calculus.

Secondly, suppose $U_n^{-1}(p)\leqslant p$, which implies $\tilde{\theta}_n(p)\leqslant p$. For any $p_0>\Phi\left(\frac{\mathbf{b_0'}\mu}{\delta}\right)$ and n sufficiently large, by using again Remark 4.22 we have

$$\begin{aligned}
&\sup_{p\in(p_0,1-1/n)} \sqrt{m} R_\phi\left(\tilde{\theta}_n(p)\right)|R_n(p)| \\
&\leqslant \tilde{q}_\delta(p_0) \sup_{p\in(0,1-1/n]} \sqrt{m}\left|\frac{R_n(p)}{q_\delta(p)}\right|.
\end{aligned} \tag{4.207}$$

Then, the right hand side of (4.207) is $o_p(1)$ as $n\to\infty$ by Proposition 4.21, Lemma 4.46, and stochastic calculus.

Therefore, by choosing $p_0\in(0,1)$ such that $p_0>\Phi\left(\frac{\mathbf{b_0'}\mu}{\delta}\right)$, then both terms in (4.204) and (4.205) are $o_p(1)$ as $n\to\infty$. Hence, lemma is proved. □

Lemma 4.49. *Let $\tilde{\theta}_n(p)$ be given by (4.177). Then*

$$\sup_{p\in(0,1-1/n)} \sqrt{m}\left|\left(R_\phi\left(\tilde{\theta}_n(p)\right)\left(U_n^{-1}(p)-p\right)-R_\phi(p)\left(p-U_n(p)\right)\right)\right| = o_p(1), \quad as \quad n\to\infty.$$

Proof. Immediate from Lemmas 4.47 and 4.48. □

Now, the only proof left is for the interval $(1-1/n,1)$. We will first introduce a lemma that is very similar to Lemma 4.38.

Lemma 4.50. *Let* $p \in (1-1/n, 1)$. *Then, as* $n \to \infty$,

$$\sup_{p\in(1-1/n,1)} \sqrt{n}\left(1-\Phi\left(\Phi^{-1}(p)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)\right) = o(1). \tag{4.208}$$

Proof. Since for $p \in (1-1/n, 1)$

$$1-\Phi\left(\Phi^{-1}(p)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right) \leqslant 1-\Phi\left(\Phi^{-1}\left(1-\frac{1}{n}\right)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right),$$

then, it suffices to show that

$$\sqrt{n}\left(1-\Phi\left(\Phi^{-1}\left(1-\frac{1}{n}\right)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)\right) = o(1), \quad \text{as} \quad n \to \infty.$$

Let $c > \sqrt{2}$ and notice that for n sufficiently large

$$\left(1-\frac{1}{c}\right)\Phi^{-1}\left(1-\frac{1}{n}\right) > \mathbf{b}_{\mathbf{0}}^{'}\mu.$$

Therefore, by Lemma 4.29, as $n \to \infty$, we have

$$\begin{aligned}
&\sqrt{n}\left(1-\Phi\left(\Phi^{-1}\left(1-\frac{1}{n}\right)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)\right) \\
<\ &\sqrt{n}\left(1-\Phi\left(\frac{\Phi^{-1}\left(1-\frac{1}{n}\right)}{c}\right)\right) < n^{\left(\frac{1}{2}-\frac{1}{c^2}\right)} = o(1).
\end{aligned}$$

□

Lemma 4.51. *Let $p \in (1-1/n, 1)$. Then, as $n \to \infty$,*

$$\sup_{p\in(1-1/n,1)} \sqrt{m} \left| \Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right) - \mathbf{b_0'}\mu\right) - \Phi\left(\Phi^{-1}(p) - \mathbf{b_0'}\mu\right) - R_\phi(p)\left(p - U_n(p)\right)\right|$$
$$= o_p(1).$$

Proof. By using triangle inequality and Remark 4.22, for n sufficiently large we have

$$\sup_{p\in(1-1/n,1)} \sqrt{m} \left| \Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right) - \mathbf{b_0'}\mu\right) - \Phi\left(\Phi^{-1}(p) - \mathbf{b_0'}\mu\right) - R_\phi(p)\left(p - U_n(p)\right)\right|$$
$$\leqslant \sup_{p\in(1-1/n,1)} \sqrt{m} \left| \Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right) - \mathbf{b_0'}\mu\right)\right|$$
$$+ \sup_{p\in(1-1/n,1)} \sqrt{m} \left| \Phi\left(\Phi^{-1}(p) - \mathbf{b_0'}\mu\right)\right|$$
$$+ \tilde{q}_\delta\left(1 - \frac{1}{n}\right) \sup_{p\in(0,1)} \sqrt{m} \left| \frac{p - U_n(p)}{q_\delta(p)}\right|$$

The conclusion of the lemma follows immediately by using Lemmas 4.38, 4.50, Proposition 4.21, and Lemma 2.53 applied to the Uniform distribution. □

We are now able to put together the result for the entire $(0, 1)$ interval.

Lemma 4.52. *Let $p \in (0, 1)$. Then, as $n \to \infty$,*

$$\sup_{p\in(0,1)} \sqrt{m} \left| \Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right) - \mathbf{b_0'}\mu\right) - \Phi\left(\Phi^{-1}(p) - \mathbf{b_0'}\mu\right) - R_\phi(p)\left(p - U_n(p)\right)\right|$$
$$= o_p(1).$$

Proof. Let $\tilde{\theta}_n(p)$ be given by (4.177) such that the process (4.105) can be equivalently written as (4.176). Notice that

$$
\begin{aligned}
&\sup_{p\in(0,1)} \sqrt{m}\left|\Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)-\Phi\left(\Phi^{-1}(p)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)-R_\phi(p)\left(p-U_n(p)\right)\right| \\
&\leqslant \max\Bigg\{\sup_{p\in(0,1-1/n)} \sqrt{m}\left|\left(R_\phi\left(\tilde{\theta}_n(p)\right)\left(U_n^{-1}(p)-p\right)-R_\phi(p)\left(p-U_n(p)\right)\right)\right|, \\
&\sup_{p\in(1-1/n,1)} \sqrt{m}\left|\Phi\left(\Phi^{-1}\left(U_n^{-1}(p)\right)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)-\Phi\left(\Phi^{-1}(p)-\mathbf{b}_{\mathbf{0}}^{'}\mu\right)-R_\phi(p)\left(p-U_n(p)\right)\right|\Bigg\}.
\end{aligned}
$$

The conclusion follows immediately by using Lemmas 4.49 and 4.51. □

Lemma 4.53. *Let R_ϕ be defined by (4.111), $m/n \to \lambda \in \mathbb{R}^+$ as $n \to \infty$, and $p \in [0,1]$. Then, as $n \to \infty$, the process $\sqrt{m}R_\phi(p)\left(p - U_n(p)\right)$, defined to be zero if $p = 0$ or $p = 1$, converges weakly in $D[0,1]$ to $\sqrt{\lambda}R_\phi\mathbb{G}_U$, a tight Gaussian process, where $R_\phi(1)\mathbb{G}_U(1)$ is defined to be equal to 0, and with mean zero and covariance function $\lambda R_\phi(s)R_\phi(t)\left(s\wedge t - st\right)$, with $s,t \in [0,1)$, and 0 with $s = 1$ or $t = 1$.*

Proof. Let q_δ be the function defined by (4.112) where $\delta < 1/2$. For any $p \in (0,1)$ we have

$$\sqrt{m}R_\phi(p)\left(p - U_n(p)\right) = \sqrt{m}\tilde{q}_\delta(p)\left(\frac{p - U_n(p)}{q_\delta(p)}\right).$$

By defining the process $\sqrt{m}\frac{p-U_n(p)}{q_\delta(p)}$ to be zero for $p = 0$ and $p = 1$ and by the fact that $\lim_{p\to 0}\tilde{q}_\delta(p) = \lim_{p\to 1}\tilde{q}_\delta(p) = 0$, then the above equality is true for all $p \in [0,1]$. Therefore, it is sufficient to show the weak convergence of the process $\sqrt{m}\tilde{q}_\delta(p)\frac{p-U_n(p)}{q_\delta(p)}$. Notice that for $\delta < 1/2$

$$\int_0^1 \frac{1}{\left(q_\delta(p)\right)^2} = \frac{(1-p)^{1-2\delta}}{1-2\delta} < \infty.$$

Moreover, it can be easily seen that q_δ is monotone around endpoints $p = 0$ and $p = 1$. Thus, by using Lemma 2.63 we have,

$$\sqrt{m}\tilde{q}_\delta(p)\frac{p - U_n(p)}{q_\delta(p)} \rightsquigarrow \sqrt{\lambda}R_\phi(p)\mathbb{G}_U(p) \quad \text{in } D[0,1], \tag{4.209}$$

a tight Gaussian process with mean zero. It can be proved that the limiting process has covariance function $\lambda R_\phi(s)R_\phi(t)\,(s \wedge t - st)$, with $s, t \in (0,1)$. Next, we will show that the covariance function, when we consider the endpoints $p = 0$ and $p = 1$, is equal to zero. Since $\lim_{s\to 0} R_\phi(s) = 0$, then the covariance function for $0 = s \leqslant t < 1$, given by $\lim_{s\to 0} R_\phi(s)R_\phi(t)s(1-t)$, is equal to zero. By using (4.117), the covariance function for $0 < s \leqslant t = 1$, is also equal to zero:

$$\begin{aligned}
&\lim_{t\to 1} R_\phi(s)R_\phi(t)s(1-t) \\
&\quad = \quad sR_\phi(s)e^{-\frac{1}{2}(\mathbf{b_0'}\mu)^2}\lim_{t\to 1} e^{\mathbf{b_0'}\mu\Phi^{-1}(t)}(1-t) \\
&\quad = \quad sR_\phi(s)e^{-\frac{1}{2}(\mathbf{b_0'}\mu)^2}\lim_{x\to\infty} e^{\mathbf{b_0'}\mu x}(1-\Phi(x)) = 0.
\end{aligned}$$

Using similar arguments as above, it can be easily shown that the variances of the limiting process are zero at both endpoints, $p = 0$ and $p = 1$. Thus, the proof is complete. □

Lemma 4.54. *Let* $\hat{\mathbf{b}}$ *defined by (4.10) such that (4.18) and (4.25) are satisfied,* $\tilde{G}(\,\cdot\,, \mathbf{b})$ *be equal to* $\Phi(\Phi^{-1}(\,\cdot\,) - \mathbf{b}'\mu)$, R_ϕ *be defined by (4.111), and* $m/n \to \lambda \in \mathbb{R}^+$ *as* $n \to \infty$. *Define the drift process given by (4.91) to be zero at the endpoints* $p = 0$

and $p = 1$. *Then, for* $R_\phi(1)\mathbb{G}_U(1)$ *defined to be equal to* 0,

$$\sqrt{m}\left(\tilde{G}\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right), \hat{\mathbf{b}}\right) - \tilde{G}\left(p, \mathbf{b_0}\right)\right) \rightsquigarrow \sqrt{\lambda} R_\phi(p)\mathbb{G}_U(p), \quad in \quad D[0,1], \tag{4.210}$$

as $n \to \infty$.

Proof. Recall that the drift process defined by (4.91) was decomposed as sum of three other processes. The first process, defined by (4.103), is $o_p(1)$ on interval $(0,1)$ by Corollary 4.17. The second process, defined by (4.104), is also $o_p(1)$ on interval $(0,1)$ by Lemma 4.40. Finally, by using Lemmas 4.52, 4.53, and Slutsky's Lemma, the third process, defined by (4.105), converges weakly in $D[0,1]$ to the gaussian process $\sqrt{\lambda} R_\phi \mathbb{G}_U$ with mean zero and covariance matrix given by $\lambda R_\phi(s) R_\phi(t)\left(s \wedge t - st\right)$, with $s, t \in [0,1)$, and 0 with $s = 1$ or $t = 1$. The conclusion follows immediately from the previous stated results and Slutsky's Lemma. □

4.4 The Limit of the Generalized Empirical ROC Process

Theorem 4.55. *Let* $\{\mathbf{X_i}\}_{i=1}^{n}$ *and* $\{\mathbf{Y_j}\}_{j=1}^{m}$ *be mutually independent random samples from multivariate normal distributions with mean vectors* $\mathbf{0}$ *and* μ*, respectively, and the same covariance matrix* Σ. *Let* $\mathbf{a_0}$ *be given by (4.2) and* $\hat{\mathbf{a}}$ *an estimator of* $\mathbf{a_0}$ *satisfying (4.3). Let* $\tilde{G}(\,\cdot\,, \mathbf{b})$ *be equal to* $\Phi(\Phi^{-1}(\,\cdot\,) - \mathbf{b}'\mu)$*, where* $\mathbf{b}$ *is given by (4.10), and* R_ϕ *be defined by (4.111). Define the generalized empirical ROC process given by (4.8) to be zero at the endpoints* $p = 0$ *and* $p = 1$. *Then, for* $R_\phi(1)\mathbb{G}_U(1)$ *defined to* *be equal to* 0 *and for* $m, n \in \mathbb{N}$ *such that* $m/n \to \lambda \in \mathbb{R}^+$*, as* $n \to \infty$*,*

$$\sqrt{m}\left(G_m\left(F_n^{-1}\left(p, \hat{\mathbf{a}}\right), \hat{\mathbf{a}}\right) - G\left(F^{-1}\left(p, \mathbf{a_0}\right), \mathbf{a_0}\right)\right) \rightsquigarrow \mathbb{G}_{\tilde{G}}(p) + \sqrt{\lambda} R_\phi(p) \mathbb{G}_U(p), \quad (4.211)$$

in $D[0,1]$*. The covariance structure of the limit process is given by*

$$\tilde{G}(s \wedge t) - \tilde{G}(s)\tilde{G}(t) + \lambda R_\phi(s) R_\phi(t)(s \wedge t - st), \quad (4.212)$$

where $s, t \in [0,1)$*, and* 0 *with* $s = 1$ *or* $t = 1$*.*

Proof. Let $\mathbf{b_0}$ and $\hat{\mathbf{b}}$ be defined by (4.10) with with $\hat{\mathbf{b}}$ satisfying (4.18) and (4.25). Then, from Lemmas 4.14, 4.15, 4.54, independence of random samples $\{\mathbf{X_i}\}_{i=1}^n$ and $\{\mathbf{Y_j}\}_{j=1}^m$, Slutsky's Lemma, Lemma 2.42, we have

$$\sqrt{m}\left(\tilde{G}_m\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right), \hat{\mathbf{b}}\right) - \tilde{G}\left(p, \mathbf{b_0}\right)\right) \rightsquigarrow \mathbb{G}_{\tilde{G}}(p) + \sqrt{\lambda} R_\phi(p) \mathbb{G}_U(p), \quad (4.213)$$

in $D[0,1]$, as $n \to \infty$. The conclusion of the theorem follows immediately since the generalized empirical ROC process was equivalently written as $\sqrt{m}\left(\tilde{G}_m\left(U_n^{-1}\left(p, \hat{\mathbf{b}}\right), \hat{\mathbf{b}}\right) - \tilde{G}\left(p, \mathbf{b_0}\right)\right)$. □

CHAPTER 5: APPLICATION AND SIMULATION STUDY

5.1 Application

In this section we will apply our methodology to a lung cancer data provided by Dr. Edward Hirschowitz, Department of Internal Medicine at University of Kentucky Medical Center. There are 52 normal subjects and 51 subjects with lung cancer. The biomarkers are proteins from cDNAT7 phage library using biopan enrichment technique. Two candidate proteins, T7RL1002 and T7RL1004, were selected to create a new biomarker as a linear combination. The data was log-transformed beforehand.

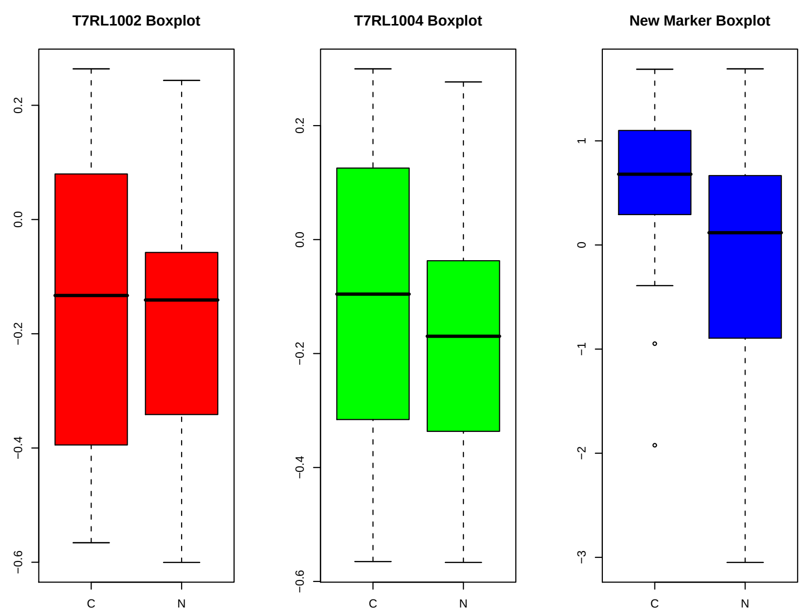

Figure 5.1: Boxplots of T7RL1002, T7RL1004, and the new marker

In Figure 5.1 above, the new marker, constructed as linear combination of T7RL1002 and T7RL1004 using Su and Liu method, seems to better discriminate, between lung cancer and normal subjects, than the individual markers. Under the assumption of equal covariance matrices, the coefficients of the linear combination were estimated by $\hat{\mathbf{a}}_0 = \hat{\mathbf{T}}^{-1}(\overline{\mathbf{Y}} - \overline{\mathbf{X}}) = (-17.77, 18.49)'$, where $\hat{T}^{-1}$ is the inverse of the pooled variance as given in Lemma 4.3. Hence, the linear combination of T7RL1002 and T7RL1004 was given by $\hat{\mathbf{a}}_0'\mathbf{X}$ and $\hat{\mathbf{a}}_0'\mathbf{X}$.

The comparison between the ROC curves for T7RL1002, T7RL1004 and the new marker is presented in Figure 5.2 below. We can clearly see now, based on the ROC plots below, that the newly constructed marker has a better sensitivity than the individual markers, at all specificity points, except for a very small range of specificity values close to one.

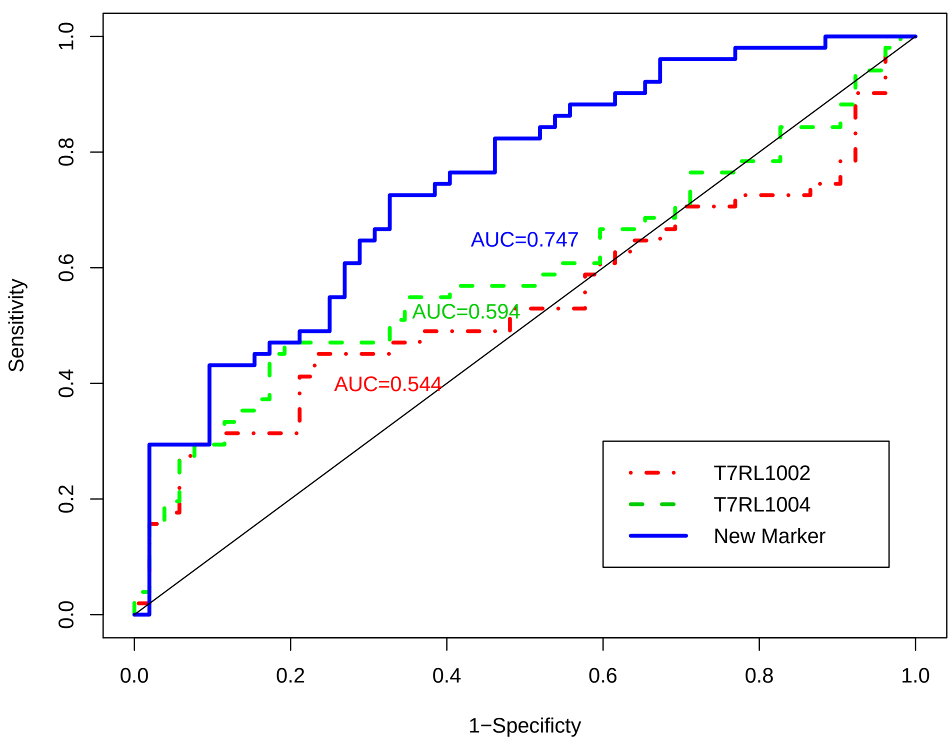

Figure 5.2: ROC curves of T7RL1002, T7RL1004, and the generalized ROC curve

5.2 Simulation Study

A simulation study was performed to estimate the coverage probabilities of the asymptotic pointwise confidence intervals at different specificity values. Since we are mostly interested in large values of specificity, we have chosen to conduct the simulations for the following set of values $\{0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95\}$. The nominal confidence level chosen for all simulations was 95 per cent. The simulations were performed using R software. The multivariate test values for non-diseased and diseased subjects were randomly sampled from bivariate normal distributions $MVN\left((0,0)', \Sigma\right)$ and $MVN\left((\mu_1, \mu_2)', \Sigma\right)$, respectively, using function *mvrnorm* from package *MASS* in R. The diagonal of the covariance matrix Σ was set to 1 and the covariance σ_{12} between biomarkers was chosen from the set $\{0.1, 0.5, 0.9\}$, which can be interpreted as low, medium and high positive correlation levels. The following diseased population mean vectors were used in simulations $\left\{(0.5, 0.5)', (0.5, 1)', (1, 1)'\right\}$. Given that in practice, usually the cases are more difficult to obtain, we considered the situations $m/n \in \{1, 0.5\}$, with $n \in \{20, 40, 100, 250\}$. The variance at each p was determined using the covariance formula (4.212) from Theorem 4.55. Hence, the 95 per cent confidence interval at a specific value of p, where $p \in (0, 1)$, was given by

$$G_m\left(F_n^{-1}(p)\right) \pm 1.96 * \frac{\sqrt{\left(\tilde{G}(p) - \tilde{G}^2(p)\right) + \lambda R_\phi^2(p)(p - p^2)}}{\sqrt{m}},$$

where, recall, $\tilde{G}(p) = \Phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right)$ and $R_\phi(p) = \frac{\phi\left(\Phi^{-1}(p) - \mathbf{b_0}'\mu\right)}{\phi(\Phi^{-1}(p))}$. Since $\mathbf{b_0}'\mu = \sqrt{\mu'\Sigma^{-1}\mu}$ by (4.17), then we can estimate $\mathbf{b_0}'\mu$ by $\sqrt{(\overline{\mathbf{Y}} - \overline{\mathbf{X}})'\hat{\mathbf{T}}^{-1}(\overline{\mathbf{Y}} - \overline{\mathbf{X}})}$ where

$\hat{T}^{-1}$ is calculated as in Lemma 4.3. The results of 10,000 simulations are presented in Tables 5.1 and 5.2, below.

Table 5.1: Estimated Coverage Probabilities of the asymptotic confidence intervals for $m/n = 1$

			Specificity									
(μ_1, μ_2)	n	σ_{12}	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
(0.5,0.5)	20	0.1	91.69	91.46	91.64	91.79	92.22	93.09	93.89	93.85	93.46	93.05
		0.5	91.56	91.79	92.10	92.01	92.64	93.45	94.37	94.68	94.20	92.29
		0.9	91.56	91.75	92.06	92.25	92.71	93.51	94.01	93.02	93.69	91.68
	40	0.1	93.02	93.48	93.39	93.49	93.75	93.79	94.29	95.15	94.71	93.82
		0.5	93.13	93.31	93.29	93.43	93.77	94.26	95.07	94.81	94.74	93.56
		0.9	93.38	93.40	93.21	93.73	94.09	94.15	95.35	94.66	94.82	93.49
	100	0.1	94.27	94.68	94.59	94.32	94.78	94.94	94.85	94.75	94.46	94.48
		0.5	94.29	94.65	94.43	94.39	94.84	94.70	95.06	94.61	94.41	94.19
		0.9	94.36	94.63	94.42	94.13	94.79	94.94	94.53	94.76	94.38	94.32
	250	0.1	94.27	94.46	94.56	95.02	94.76	95.10	94.89	95.39	95.37	95.63
		0.5	94.48	94.44	94.76	95.24	94.98	95.28	94.89	95.29	95.20	95.60
		0.9	94.42	94.36	94.75	95.02	94.99	94.66	94.77	95.35	95.24	95.32
(0.5,1)	20	0.1	91.70	91.36	91.50	91.60	91.78	91.92	92.02	92.62	93.29	94.95
		0.5	92.05	91.63	91.74	91.97	91.91	91.82	92.51	93.00	93.71	93.98
		0.9	92.32	91.69	91.51	91.69	91.58	91.65	91.84	92.40	93.09	94.98
	40	0.1	93.35	93.04	92.98	93.03	93.25	92.75	93.69	93.94	94.36	94.65
		0.5	93.40	93.03	92.84	92.89	92.84	93.03	93.45	94.09	94.30	94.33
		0.9	93.23	93.11	92.94	92.45	92.66	92.56	92.64	93.20	93.81	95.03
	100	0.1	94.25	94.37	94.19	94.37	94.57	94.42	94.42	94.17	94.17	94.38
		0.5	94.56	94.39	94.27	94.28	94.52	94.45	94.34	93.86	94.52	94.45
		0.9	94.32	94.27	94.08	94.04	94.12	93.96	94.01	93.98	94.73	94.41
	250	0.1	94.71	94.66	94.44	94.99	94.97	95.33	94.90	95.46	95.21	95.60
		0.5	94.73	95.19	94.96	95.20	95.06	94.88	94.95	94.98	95.12	95.54
		0.9	94.80	94.60	94.64	94.58	94.71	94.93	94.33	94.28	94.94	95.35
(1,1)	20	0.1	91.93	90.95	91.12	90.87	90.93	91.10	91.55	91.91	92.46	94.96

Continued on next Page...

Table 5.1: Estimated Coverage Probabilities of the asymptotic confidence intervals for $m/n = 1$

			Specificity									
(μ_1, μ_2)	n	σ_{12}	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
		0.5	91.36	91.02	91.09	90.83	91.42	91.69	91.91	92.36	93.31	94.23
		0.9	91.36	91.28	91.33	91.40	91.79	91.86	92.29	92.57	93.67	93.74
	40	0.1	93.27	93.29	92.88	92.66	92.73	92.85	92.86	93.20	94.04	95.09
		0.5	93.33	92.85	92.95	93.08	92.98	93.17	93.31	93.62	94.65	94.84
		0.9	93.24	93.13	93.00	93.15	93.34	93.23	93.55	94.25	94.80	94.45
	100	0.1	94.11	94.09	94.13	94.23	94.45	94.36	94.41	94.48	94.23	94.71
		0.5	94.37	94.14	94.17	94.12	94.43	94.45	94.63	94.46	94.60	95.15
		0.9	94.22	94.03	94.16	94.33	94.68	94.53	94.55	94.38	94.54	94.65
	250	0.1	94.57	94.84	94.53	94.70	94.57	94.49	94.34	94.83	94.76	95.66
		0.5	94.72	94.34	94.52	94.59	94.87	94.72	94.72	95.08	95.02	95.66
		0.9	94.41	94.47	94.46	94.83	94.79	94.85	94.64	95.20	94.75	95.79

Table 5.2: Estimated Coverage Probabilities of the asymptotic confidence intervals for $m/n = 0.5$

			Specificity									
(μ_1, μ_2)	n	σ_{12}	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
(0.5,0.5)	20	0.1	90.41	90.48	91.29	91.22	91.40	92.16	92.50	93.06	93.09	93.80
		0.5	90.20	90.70	91.11	91.32	91.69	92.87	91.26	94.56	93.95	92.80
		0.9	90.13	90.58	90.97	91.19	92.32	92.48	92.95	94.35	94.60	92.14
	40	0.1	92.95	92.80	92.87	93.04	93.37	93.66	94.33	92.25	94.33	93.98
		0.5	92.62	92.35	92.80	93.13	93.38	93.67	94.12	94.79	93.79	93.87
		0.9	92.74	93.03	93.25	93.34	94.12	94.47	94.69	94.01	95.07	93.84
	100	0.1	93.63	93.33	94.23	93.85	94.38	94.71	94.75	94.99	94.95	94.86
		0.5	93.73	93.58	94.30	93.83	95.05	94.30	94.36	95.27	94.99	94.33
		0.9	93.89	93.54	94.46	94.07	95.07	94.78	95.13	95.00	94.93	94.49
	250	0.1	94.56	94.45	94.23	94.64	94.35	95.51	95.17	95.35	95.43	95.39
		0.5	94.02	94.33	94.37	94.78	94.87	94.93	94.96	94.97	95.13	95.37

Continued on next Page. . .

Table 5.2: Estimated Coverage Probabilities of the asymptotic confidence intervals for $m/n = 0.5$

			Specificity									
(μ_1, μ_2)	n	σ_{12}	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
		0.9	94.06	94.43	94.84	94.41	95.05	95.08	95.13	94.99	95.29	95.22
(0.5,1)	20	0.1	91.81	90.95	90.00	90.31	90.26	90.85	90.61	91.05	92.83	93.15
		0.5	91.52	90.50	90.19	90.01	90.31	90.71	90.42	91.44	92.68	94.11
		0.9	93.57	92.52	91.14	90.67	89.45	89.53	89.70	90.24	90.64	93.02
	40	0.1	92.65	92.73	92.54	92.37	92.44	92.92	92.58	92.61	93.75	94.68
		0.5	92.63	92.59	92.01	92.34	92.03	92.15	93.28	93.29	92.52	93.52
		0.9	93.48	92.68	92.26	92.50	92.21	92.22	91.95	92.86	93.31	94.28
	100	0.1	94.14	93.83	93.35	93.72	93.57	93.71	93.97	94.02	94.18	94.69
		0.5	94.00	93.59	93.43	93.65	94.11	94.26	94.02	94.38	94.57	94.53
		0.9	94.15	93.98	93.77	93.58	93.73	93.41	93.42	93.61	93.66	94.83
	250	0.1	94.52	94.56	94.80	94.59	94.92	95.03	94.93	94.75	95.19	95.03
		0.5	94.77	94.87	94.90	94.94	94.73	95.17	94.66	95.28	94.88	95.26
		0.9	95.05	95.05	94.91	94.96	94.63	94.96	94.32	94.94	94.66	95.3
(1,1)	20	0.1	93.48	92.58	91.71	90.73	89.52	89.68	89.96	90.87	91.21	93.17
		0.5	92.11	91.53	90.45	90.03	89.90	90.21	90.85	91.13	92.06	91.53
		0.9	91.61	90.87	90.17	90.21	90.01	91.06	90.99	91.36	92.87	94.01
	40	0.1	93.27	92.60	92.34	92.28	92.46	92.11	92.03	92.73	93.13	94.78
		0.5	92.67	92.31	92.37	92.17	92.14	91.89	92.92	93.15	93.10	94.24
		0.9	92.75	92.40	92.41	92.26	92.27	92.56	93.10	93.42	93.67	94.46
	100	0.1	94.09	93.83	93.60	93.39	93.30	93.53	93.52	93.83	94.28	94.05
		0.5	94.16	93.98	93.47	93.73	93.87	93.72	93.62	93.82	94.72	94.91
		0.9	94.10	93.70	93.66	93.94	93.51	93.71	94.11	94.03	94.52	94.85
	250	0.1	94.49	94.41	94.20	94.42	94.36	94.45	94.25	94.81	94.80	95.20
		0.5	94.18	94.46	94.22	94.37	94.34	94.21	94.75	95.02	95.08	95.43
		0.9	94.38	94.36	94.39	94.34	94.05	94.50	94.74	95.23	95.53	95.52

The estimated coverage probabilities, presented in Tables 5.1 and 5.2, were plotted against the chosen specificity values, for all possible combinations and grouped together by ratio of diseased versus nondiseased samples.

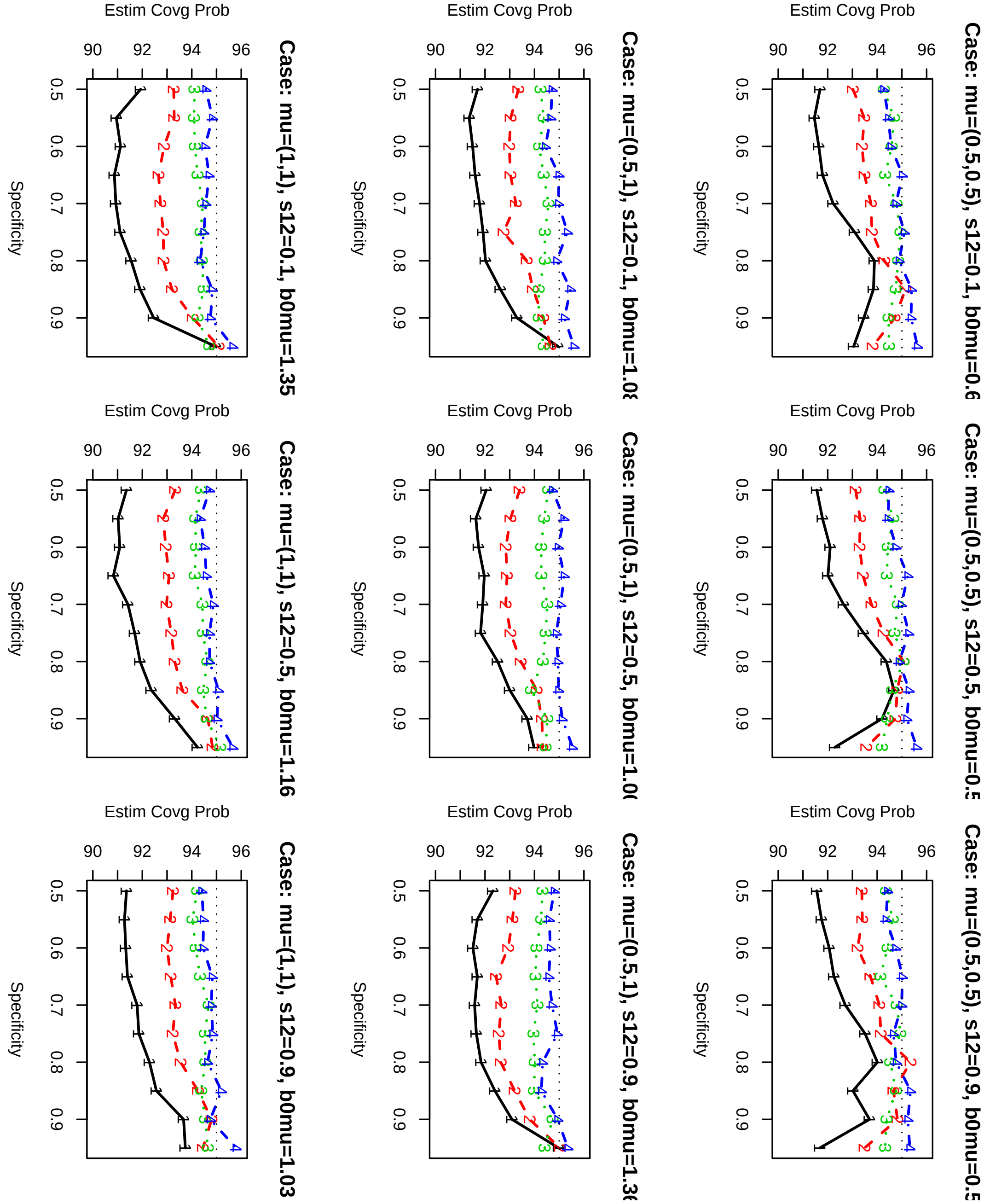

Figure 5.3: Estimated Coverage Probabilities for $m/n = 1$

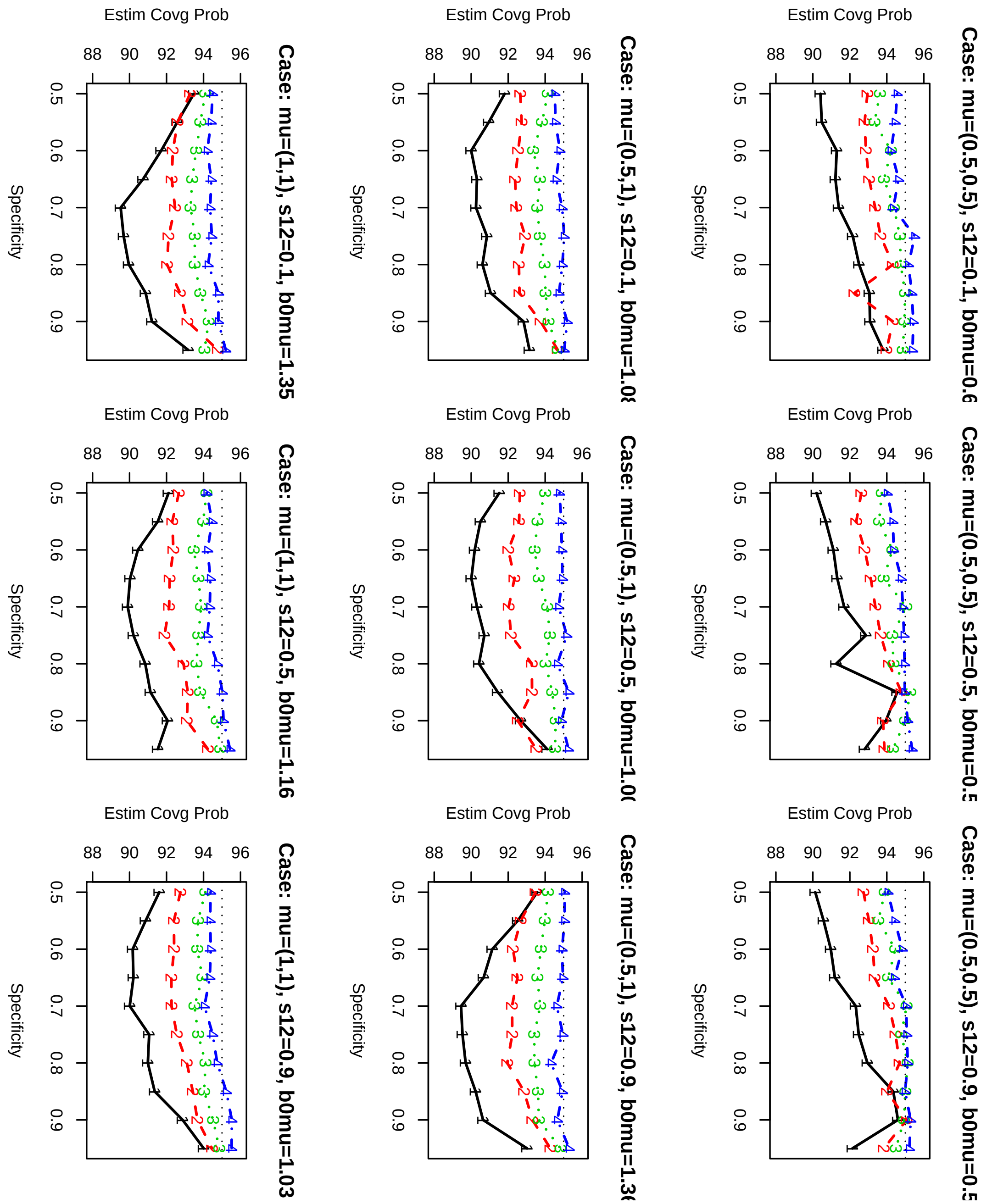

Figure 5.4: Estimated Coverage Probabilities for $m/n = 0.5$

Within each plot, the lines have different types and colors, corresponding to a different sample size. We also used symbols to distinguish the cases, with the lowest sample size numbered 1 and the largest numbered 4. By comparing Figures 5.3 and 5.4, we can observe a slight drop in the coverage only for the lower sample sizes. In other words, if the number of controls is large enough, 100 or more, the estimated coverage varies almost identically around the nominal level, even when the ratio of cases versus controls is 0.5. When the number of controls is either 20 or 40, the coverage probability is underestimated, but it still has a reasonable coverage around 90 per cent. Finally, we notice that when the diseased and nondiseased populations are not well separated, which corresponds to a low value of parameter $\mathbf{b_0'}\mu$, the estimated coverage probability drops for large specificity values.

CHAPTER 6: DISCUSSION AND FUTURE RESEARCH

In this dissertation we considered the ROC curve of a linear combination of diagnostic tests. If both the diseased and non-diseased populations are multivariate normally distributed, then the linear combination, using Fisher's linear discriminant coefficients, maximizes the area under the generalized ROC curve.

In Chapter 4, we derived the asymptotic behavior of the nonparametric estimator, the generalized empirical ROC curve, under the assumption of equal covariance matrices and zero mean for the multivariate normal distribution of the non-diseased population. The coefficients can be estimated by maximum likelihood, however our general requirement was that the estimator is bounded in probability. Future research will be focused on finding the asymptotic distribution of the nonparametric estimator when relaxing one or more conditions. For example, the assumption of equal covariance matrices is not a realistic once, and thus we would be interested in finding the asymptotic distribution for the case of unequal covariance matrices. Also, from a practical standpoint, the normality assumption is not always met. A possible solution would be to consider a situation similar to the binormal assumption, in which data becomes multivariate normal after a monotone transformation is applied. An alternative solution is to consider the multivariate distribution coming from an elliptical family. Finally, another research direction would be to determine the asymptotic distribution of a linear combinations of biomarkers that maximize the sensitivity over a desired range of specificity, as it was proposed by Liu *et al.* (2005).

In Chapter 5, we applied the methodology to a real dataset and created a new marker as a linear combination of two biomarkers that shows a better discrimination between lung cancer patients and normal patients. In the end, we conducted a simulation study for combinations of two biomarkers to determine the estimated coverage probability of the asymptotic pointwise confidence intervals. The results showed a good coverage for sample sizes of at least 100 controls. For lower sample size, the coverage was underestimated with values around 90 per cent. Also, for lower sample sizes we saw a drop in the coverage probability that may be explained by the discreteness nature of the process. As a future work, we will consider constructing confidence intervals using a smoothed empirical distribution G_m. We will also consider more simulations to estimate the coverage probabilities when we have departure from the normal distribution and equal covariance matrices assumption. Finally, we consider developing regional confidence bands for the generalized empirical ROC curve.

BIBLIOGRAPHY

[1] S. G. Baker. Identifying combinations of cancer markers for further study as triggers of early intervention. *Biometrics*, 56(4):1082–1087, 2000.

[2] D. Bamber. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 12:387–415, 1975.

[3] J. Beirlant and P. Deheuvels. On the approximation of P-P and Q-Q plot processes by brownian bridges. *Statistics & Probability Letters*, 9:241–251, 1990.

[4] P. Bilingsley. *Convergence of Probability Measures.* John Wiley & Sons, Inc., New York, 1968.

[5] G. Campbell. Advances in statistical methodology for the evaluation of diagnostic and laboratory tests (Disc: P553-556). *Statistics in Medicine*, 13:499–508, 1994.

[6] G. Claeskens, B. Y. Jing, L. Peng, and W. Zhou. Empirical likelihood confidence regions for comparison distributions and ROC curves. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 31(2):173–190, 2003.

[7] M. Csörgő. *Quantile Processes with Statistical Applications.* SIAM, Philadelphia, 1983.

[8] M. Csörgő and P. Révész. *Strong Approximations in Probability and Statistics.* Academic Press, 1981.

[9] L. E. Dodd and M. S. Pepe. Partial AUC estimation and regression. *Biometrics*, 59(3):614–623, 2003.

[10] D. D. Dorfman and E. Jr. Alf. Maximum likelihod estimation of parameters of signal detection theory - a direct solution. *Psychometrika*, 33, 1968.

[11] D. D. Dorfman and E. Jr. Alf. Maximum likelihod estimation of parameters of signal detection theory - rating method data. *Journal of Mathematical Psychology*, 6:487–496, 1969.

[12] R. M. Dudley and R. Norvaisa. *Differentiability of Six Operators on Nonsmooth Functions and p-Variation*. Lecture Notes in Mathematics, Springer-Verlag, 1999.

[13] M. J. Goddard and I. Hinberg. Receiver operator characteristic (ROC) curves and non-normal data: An empirical study. *Statistics in Medicine*, 9:325–337, 1990.

[14] D. M. Green and J. A. Swets. *Signal Detection Theory and Psychophisics*. John Wiley & Sons, Inc., New York, 1966.

[15] J. Gu and S. Ghoshal. Strong approximations for resample quantile processes and application to ROC methodology. *Journal of Nonparametric Statistics*, 20(3):229–240, 2008a.

[16] J. Gu and S. Ghoshal. Bayesian ROC curve estimation under binormality using a rank likelihood. *Journal of Statistical Planning and Inference*, In Press, Accepted Manuscris, 2008c.

[17] J. Gu, S. Ghoshal, and A. Roy. Bayesian bootstrap estimation of ROC curve. *Statistics in Medicine*, 27(26):5407–5420, 2008b.

[18] A. K. Han. Non-papametric analysis of a generalized regression model. the maximum rank correlation estimator. *Journal of Econometrics*, 35:303–316, 1987.

[19] J. A. Hanley. The robusteness of the 'binormal' assumptions used in fitting ROC curvesd. *Medical Decision Making*, 8:197–203, 1988.

[20] J. A. Hanley. The use of the 'binormal' model for parametric ROC analysis of quantitative diagnostic tests. *Statistics in Medicine*, 15:1575–1585, 1996.

[21] F. Hsieh and B. W. Turnbull. Nonparametric and semiparametric estimation of the receiver opertating characteristic. *The Annals of Statistics*, 24(1):25–40, 1996.

[22] K. Jensen, H. H. Müller, and H. Schäfer. Regional confidence bands for ROC curves. *Statistics in Medicine*, 19:493–509, 2000.

[23] G. Li, R. C. Tiwari, and M. T. Wells. Quantile comparison functions in two-sample problems, with application to comparisons of diagnostic markers. *Journal of the American Statistical Association*, 91:689–698, 1996.

[24] G. Li, R. C. Tiwari, and M. T. Wells. Semiparametric inference for a quantile comparison function with applications to receiver operating characteristic curves. *Biometrika*, 86:487–502, 1999.

[25] A. Liu, E. F. Schisterman, and Y. Zhu. On linear combinations of biomarkers to improve diagnostic accuracy. *Statistics in Medicine*, 24(1):37–47, 2005.

[26] C. J. Lloyd. Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association*, 93:1356–1364, 1998.

[27] L. B. Lusted. Logical analysis in roentgen diagnosis. *Radiology*, 74:178–193, 1960.

[28] D. McClish. Analyzing a portion of the ROC curve. *Medical Decision Making*, 9:190–195, 1989.

[29] C. E. Metz, B. A. Herman, and J. H. Shen. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine*, 17:1033–1053, 1998.

[30] M. S. Pepe. An interpretation for the ROC curve and inference using GLM procedures. *Biometrics*, 56(2):352–359, 2000.

[31] M. S. Pepe. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford Statistical Sciences Series, Oxford University Press, New York, 2003.

[32] M. S. Pepe, T. Cai, and G. Longton. Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics*, 62(1):221–229, 2006.

[33] M. S. Pepe and M. L. Thompson. Combining diagnostic test results to increase accuracy. *Biostatistics*, 1(2), 2000.

[34] J. A. III Reeds. *On the Definition of von Mises Functions*. Ph.D. disertation, Harvard University, 1976.

[35] B. Reiser and D. Faraggi. Confidence intervals for the generalized ROC criterion. *Biometrics*, 53:644–652, 1997.

[36] E. F. Schisterman, D. Faraggi, and B. Reiser. Adjusting the generalized ROC curve for covariates. *Statistics in Medicine*, 23(21):3319–3331, 2004.

[37] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc., New York, 1981.

[38] J. Q. Su and J. S. Liu. Linear combinations of multiple diagnostic markers. *Journal of the American Statistical Association*, 88:1350–1355, 1993.

[39] J. A. Swets and R. M. Pickett. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic Press, 1982.

[40] M. L. Thompson and W. Zucchini. On the statistical analysis of ROC curves. *Statistics in Medicine*, 8:1277–1290, 1989.

[41] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[42] A.W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer-Verlag, New York, 1996.

[43] J. A. Wellner. Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Probability Theory and related Fields*, 45:73–88, 1978.

[44] X. H. Zhou and J. Harezlak. Comparison of bandwidth selection methods for kernel smoothing of ROC curves. *Statistics in Medicine*, 21(14):2045–2055, 2002.

[45] X. H. Zhou, D. K. McClish, and N. A. Obuchowski. *Statistical Methods in Diagnostic Medicine*. John Wiley & Sons, Inc., New York, 2002.

[46] K. H. Zou and W. J. Hall. Two transformation models for estimating an ROC curve derived from continuous data. *Journal of Applied Statistics*, 27(5):621–631, 2000.

[47] K. H. Zou, W. J. Hall, and D. E. Shapiro. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine*, 16:2143–2156, 1997.

VITA: COSTEL CHIRILA

Date and Place of Birth

September 13, 1967
Galati, Romania

Education

UNIVERSITY OF KENTUCKY — Lexington, KY
Ph.D. Candidate in Statistics, November 2008

Master of Science in Statistics, May 2001

ALECSANDRU IOAN CUZA UNIVERSITY — Iasi, Romania
Bachelor of Arts in Economics, July 1997

Bachelor of Science in Mathematics, July 1992

Professional Experience

METABOLON, INC. — Research Triangle Park, NC
Biostatistician, May 2007 - present

PPD, INC. — Research Triangle Park, NC
Biostatistician, December 2005 - May 2007

GILEAD, INC. — Durham, NC
Consultant, August 2005 - December 2005

ELI LILLY AND COMPANY — Indianapolis, IN
Intern, May 2003 - August 2003

UNIVERSITY OF KENTUCKY — Lexington, KY
Department of Statistics:
Graduate Teaching Assistant, August 1999 - May 2000 & August 2003 - May 2005

College of Nursing:
Graduate Research Assistant, August 2000 - May 2003

ALECSANDRU IOAN CUZA UNIVERSITY Iasi, Romania
College of Economics, Department of Statistics and Econometrics:
Instructor, September 1993 - June 1997

Researcher, April 1993 - September 1993

Awards and Honors

R. L. Anderson Teaching Award (University of Kentucky, 2005)
Graduate Fellowship Award (University of Kentucky, 2004)
Graduate Fellowship Award (University of Kentucky, 2001)
Graduate Student Development Award, (University of Kentucky, 2000)

Publications

HAHN, E. J., RAYENS, M. K., WARNICK, T. A., CHIRILA, C., RASNAKE, R. T., PAUL, T. P., & CHRISTIE, D (2005) A Controlled Trial of a Quit and Win Contest. *American Journal of Health Promotion.* **20(2)** 117-126.

HAHN, E. J., RAYENS, M. K., CHIRILA, C., RIKER, C. A., PAUL, T. P., & WARNICK, T. A. (2004). Effectiveness of a quit and win contest with a low-income population. *Preventive Medicine* **39** 543-550.

Contributed Talks

CHIRILA, C (May 2008). Variable selection using Random Forests. Midwest Biopharmaceutical Statistics Workshop.

CHIRILA, C, WOOD, C. L., STROMBERG, A., & BATHKE, A. (May 2005). An application of Empirical Processes to ROC curves. Joint Statistical Meetings. (joint work, Chirila presenting)

Professional Memberships

American Statistical Association
Institute of Mathematical Statistics