# SEMI-PARAMETRIC ESTIMATION OF BIVARIATE DEPENDENCE UNDER MIXED MARGINALS

**By**

**Wenmei Huang**

**A DISSERTATION**

**Submitted to**
**Michigan State University**
**in partial fulfillment of the requirements**
**for the degree of**

**DOCTOR OF PHILOSOPHY**

**Department of Statistics and Probability**

**2011**

## ABSTRACT

### SEMI-PARAMETRIC ESTIMATION OF BIVARIATE DEPENDENCE UNDER MIXED MARGINALS

By

Wenmei Huang

Copulas are a flexible way of modeling dependence in a set of variables where the association between variables can be elicited separately from the marginal distributions. A semi-parametric approach for estimating the dependence structure of bivariate distributions derived from copulas is investigated when the associated marginals are mixed, that is, consisting of both discrete and continuous components. The semi-parametric likelihood approach is proposed for obtaining the estimator of the dependence parameter under unknown marginals. The consistency and asymptotic normality of the estimator is established as sample size tends to infinity. For constructing confidence intervals in practice, an estimator of the asymptotic variance is proposed and its properties are investigated via simulation. Extensions to higher dimensions are discussed. Several simulation studies and real data examples are provided for investigating the application of the developed methodology of inference. This work generalizes prior results obtained on the estimation of dependence when the marginals are continuous by Genest et al. [11].

**To my professors and my loving family.**

# ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my dissertation advisor, Prof. Sarat C. Dass, for his patient guidance, encouragement and support during my graduate study at Michigan State University, and his precious advice for my professional career. He helped me to learn not only how to do research in statistics, but also how to write statistical papers using proper English. The education I received under his guidance will be the most valuable in my future career.

I am grateful to Prof. Tapabrata Maiti, Prof. Chae Young Lim of the Department of Statistics and Probability, and Prof. Zhengfang Zhou of the Department of Mathematics, for serving on my thesis committee. I accumulated practical experience from the consulting work at the Center for Statistical Training and Consulting. Prof. Connie Page, Prof. Dennis Gilliland, and Prof. Sandra Herman shared with me their experience and knowledge in dealing with practical problems, and guided me through my projects.

I would like to thank Prof. James Stapleton for his help and encouragement with respect to my teaching and study. He has been very supportive to me, as well as to all his students. I thank him for offering to review my thesis. I would like to thank Prof. Anil K. Jain at Department of Computer Science and Engineering, Michigan State University, for letting me use the biometrics data in the PRIP lab in my research. I would like to thank Dr. Yongfang Zhu for discussions on course work and research. I also would like to thank all the professors of the Department of Statistics and Probability and all my great friends at Michigan State University for making my life there so unique and wonderful.

Finally, I especially thank my loving family for their encouragement and support.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

Copulas are functions that join or "couple" multivariate distribution functions to their one-dimensional marginal distribution functions. Essentially, a copula is a function $C$ from $[0,1]^D$ ($D \geq 2$) to $[0,1]$ with the following properties:

1. For every $(u_1, ..., u_D) \in [0,1]^D$,

$$C(u_1, ..., u_D) = 0 \text{ if at least one of } u_k = 0, \text{ for } k = 1, ..., D \tag{1.1}$$

and

$$C(u_1, ..., u_D) = u_k \text{ if } u_j = 1 \text{ for all } j \neq k, \ k = 1, ..., D \tag{1.2}$$

2. For every $(u_1, ..., u_D)$ and $(v_1, ..., v_D) \in [0,1]^D$ such that $u_k \leq v_k$ for $k = 1, ..., D$, which defines a $D$-box $V = [u_1, v_1] \times [u_2, v_2] \times ... \times [u_D, v_D] \subset [0,1]^D$. Then

$$\sum \text{sgn}(\mathbf{c}) C(\mathbf{c}) \geq 0 \tag{1.3}$$

where the sum is taken over all vertices $\mathbf{c}$ of $V$ and sgn$(\mathbf{c})$ is given by

$$\text{sgn} = \begin{cases} 1, & \text{if } c_k = u_k \text{ for an even number of } k'\text{s}, \\ -1, & \text{if } c_k = u_k \text{ for an odd number of } k'\text{s}. \end{cases}$$

Alternatively, copulas are multivariate distribution functions whose one-dimensional marginals are uniform on the interval $[0,1]$.

Since their introduction by Sklar [46], copulas have proved to be a useful tool for analyzing multivariate dependence structures due to many unique and interesting features. Let $F_k(x_k) = P(X_k \leq x_k)$ for $k = 1, 2, \ldots, D$ denote $D$ continuous distribution functions on the real line, and let $F$ denote a joint distribution function on $R^D$ whose $k$-th marginal distribution corresponds to $F_k$. According to Sklar's Theorem [46], there exists a unique function $C(u_1, ..., u_D)$ from $[0,1]^D$ to $[0,1]$ satisfying

$$F(x_1, ..., x_D) = C(F_1(x_1), \ldots, F_D(x_D)), \tag{1.4}$$

where $x_1, ..., x_D$ are $D$ real numbers. The function $C$ is known as a $D$-copula function that couples the one-dimensional distribution functions $F_k$, $k = 1, 2, \ldots, D$ to obtain $F$. If not all marginals are continuous, $C$ is uniquely determined on $RanF_1 \times ... \times RanF_D$, where $RanF_k$ is the range of $F_k$. Equation (1.4) can also be used to construct $D$-dimensional distribution functions $F$ whose marginals are the pre-specified distributions $F_k, k = 1, \ldots, D$: Choose a copula function $C$ and define the distribution function $F$ as in (1.4). It follows that $F$ is a $D$-dimensional distribution function with marginals $F_k$, $k = 1, \ldots, D$.

Investigating dependence structures of multivariate distributions has always been an important area for researchers; see, for example, [26], [27], and [35]. For example, one of the central problems in statistics concerns testing the hypothesis that random variables are independent. Prior to the very recent explosion of copula theory and application, the only models available in many fields to represent the dependence structure were the classical multivariate models, such as Gaussian multivariate model. These models entailed rigid assumptions on the marginal and joint behaviors of the variables. Therefore, they provide limited usefulness.

Though the phrase "copula" was first used by Sklar [46] in 1959 and traces of copula theory can be found in Hoeffdings work during the 1940s, the study of copulas and their application in statistics is a rather modern phenomenon. Earlier efforts have addressed different statistical aspects of specialized as well as general copula-based joint distributions. Inference procedures for bivariate Archimedean copulas have been developed and discussed by Genest et. al [12]. Demarta et al.,

[8], studied properties related to $t$-copulas, whereas Genest et al., [11] and [13], gave a goodness of fit procedure for general copulas. Estimation techniques and properties of the estimators have been well studied with applications ranging from statistics to mathematical finance and financial risk management; see, for example, Shih et al. [44], Embrechts et al. [9], Cherubini et al. [5], Frey et al. [10] and Chen et al. [4].

Many multivariate models for dependence can be generated by parametric families of copulas, $\{C_\theta : \theta \in \Theta\}$, typically indexed by a real or vector-valued parameter $\theta$. Examples of such systems are given in [23], [30], [22], and others. The recent monograph by Hutchinson and Lai [17], which includes an extensive bibliography, constitutes a handy reference to this expanding literature.

Copula-based models are natural in situations where learning about the association between the variables is important, since the effect of the dependence structure is easily separated from that of the marginals. In such situations, there is typically enough data to obtain nonparametric estimators of the marginal distributions, but insufficient information to afford nonparametric estimation of the structure of the association. In such cases, it is convenient to adopt a parametric form for the dependence function while keeping the marginals unspecified. To estimate the dependence parameter $\theta$, two strategies could be adopted depending on the circumstances: (i) if valid parametric models are already available for the marginals, then it is straightforward in principle to write down a likelihood function for the data, which makes the estimation of $\theta$ margin-dependent, because the estimators of the parameters involved in the marginals would be indirectly affected by the copula. This is the parametric approach. (ii) when nonparametric estimators are contemplated for the marginals, however, inference about the dependence parameter $\theta$ will be margin-free. This is the semi-parametric approach. Clayton [6], Hougaard et al. [15] and Oakes [36] have pointed out that the margin-free requirement is sensible in applications where the focus of the analysis is on the dependence structure.

Given a sample of $n$ observations $\mathbf{X}_j = (X_{1j}, X_{2j}, \ldots, X_{Dj})$, $j = 1, 2, \ldots, n$ from the joint distribution $F(x_1, x_2, \ldots, x_D) = C_\theta(F_1(x_1), \ldots, F_D(x_D))$, the estimation procedure involves selecting $\hat{\theta}_n$ to maximize the semi-parametric likelihood

$$\ell(\theta) = \sum_{j=1}^{n} \log\left[c_\theta(F_{1n}(X_{1j}), \ldots, F_{Dn}(X_{Dj}))\right] \tag{1.5}$$

where $F_{kn}$ denotes $n/(n+1)$ times the empirical distribution function of the $k$-th component observations $X_{kj}$ for $j = 1, 2, \ldots, n$, and $c_\theta$ is the density of $C_\theta$ with respect to Lebesgue measure on $[0,1]^D$. The utilization of $F_{kn}$ instead of the empirical distribution here avoids difficulties arising from the potential unboundedness of log $c_\theta(u_1, ..., u_D)$ as some of the $u_k$'s tend to 1.

A central assumption made in the earlier studies is that the marginals associated with the joint distribution $F$ should be continuous. In reality, there are many situations where this assumption is not satisfied and the marginals can be mixed, that is, they contain both discrete and continuous components (see, for example, Kohn et al. [37]). Based on the continuous assumption and the regularity conditions, Genest et al. [11] have shown that $n^{1/2}\hat{\theta}_n$ is asymptotic normal, where $\hat{\theta}_n$ is the semi-parametric estimator of $\theta$ in (1.5).

Our work extends previous methodology by accounting for mixed marginals for $D = 2$ and $\Theta \subset R$. Uniqueness is an important consideration when estimating the unknown parameters corresponding to a copula. Since our mixed marginals have discrete components, one way to achieve uniqueness is to restrict $C$ to belong to a particular parametric family. Under this assumption, We develop an estimation technique for finding the semi-parametric maximum likelihood estimator, $\hat{\theta}_n$, for the parametric family $\{C_\theta : \theta \in \Theta\}$ in the bivariate situation in Chapter 2. The estimation methodology involves integrals corresponding to the discrete components and is therefore, non-standard. Consistency and asymptotic normality of the semi-parametric maximum likelihood estimator $\hat{\theta}_n$ is developed in Chapter 3. A a variance estimator of $\hat{\theta}_n$, $\varrho$ is developed in Chapter 4. Note that for multivariate observations, copulas allow flexible modelling of the joint distribution via its marginal distributions as well as the correlation between pairwise components of the vector of observations. Therefore, the theory presented in Chapter 2 and Chapter 3 can be extended to higher dimensions, and this is given in Chapter 5. In Chapter 6, $C$ is restricted to a particular parametric family of copula, the $t$-copula, and findings related to the $t$-copula are provided. Numerical simulations and application of the methodology to real biometric data are presented in Chapter 7. We summarize our work and discuss future research directions in Chapter 8.

# CHAPTER 2

# A MODEL FOR BIVARIATE DISTRIBUTIONS

## 2.1 Joint Distributions with Mixed Marginals

Let $F$ and $G$ be two distributions on the real line that are mixed, that is, both $F$ and $G$ consist of a mixture of discrete as well as continuous components. The general form for $F$ is given by

$$F(x) = \sum_{h=1}^{d_F} p_{Fh}\, I_{\{\mathcal{D}_{Fh} \leq x\}} + (1 - \sum_{h=1}^{d_F} p_{Fh}) \int_{-\infty}^{x} f(w)\, dw, \tag{2.1}$$

where $I_{\{A\}}$ is the indicator function of set $A$ (that is, $I_{\{A\}} = 1$ if $A$ is true, and $0$ otherwise); in (2.1), the distribution function $F$ consists of $d_F$ discrete components denoted by $\mathcal{D}_{Fh}$ with $F(\mathcal{D}_{Fh}) - F(\mathcal{D}_{Fh}^{-}) = p_{Fh}$ for $h = 1, 2, \ldots, d_F$ where $x^{-}$ denotes the left limit of $x$, and $f$ is the density (continuous) component of $F$. The discrete components $\mathcal{D}_{Fh}$, $h = 1, 2, \ldots, d_F$ correspond to jump points in the distribution function $F$, and are called the jump points of $F$. All other points on $R$ correspond to points of continuity of $F$, that is, $F(x) - F(x^{-}) = 0$ if $x \neq \mathcal{D}_{Fh}$. The set of jump and continuity points of $F$ are denoted by $\mathcal{J}(F)$ and $\mathcal{C}(F)$, respectively. In the same way, writing

$$G(y) = \sum_{l=1}^{d_G} p_{Gl}\, I_{\{\mathcal{D}_{Gl} \leq y\}} + (1 - \sum_{l=1}^{d_G} p_{Gl}) \int_{-\infty}^{y} g(w)\, dw, \tag{2.2}$$

similar definitions can be given for the quantities $d_G, p_{Gl}, \mathcal{D}_{Gl}$ and $g$. We denote the set of jump and continuity points of $G$ by $\mathcal{J}(G)$ and $\mathcal{C}(G)$, respectively.

Let $H_\theta$ be the bivariate function defined by

$$H_\theta(x, y) = C_\theta(F(x), G(y)) \tag{2.3}$$

for $\theta \in \Theta$ with $F$ and $G$ as in (2.1) and (2.2), respectively. It follows from the properties of a copula function that

**Theorem 2.1.1.** *The function $H_\theta$, as defined in (2.3), is a valid bivariate distribution function on $R^2$ with marginals $F$ and $G$.*

**Proof:** For $H_\theta(x, y)$ to be a valid distribution function on $R^2$, the following four conditions should be satisfied:

(1) $0 \le H_\theta(x, y) \le 1$ for all $(x, y)$,

(2) $H_\theta(x, y) \to 0$ as $\max(x, y) \to -\infty$,

(3) $H_\theta(x, y) \to 1$ as $\min(x, y) \to +\infty$, and

(4) for every $x_1 \le x_2$ and $y_1 \le y_2$,

$$\Delta H \equiv H_\theta(x_2, y_2) - H_\theta(x_1, y_2) - H_\theta(x_2, y_1) + H_\theta(x_1, y_1) \ge 0.$$

Note that since $H_\theta(x, y) = C_\theta(F(x), G(y)) = P(U \le F(x), V \le G(y))$, (1) follows. To prove (2), note that $F(x) \to 0$ and $G(y) \to 0$ when $\max(x, y) \to -\infty$. Hence, $H_\theta(x, y) \to 0$ from the property of a cumulative distribution function. (3) follows similarly. To prove (4), we use the facts that $F(x_1) \le F(x_2)$ for $x_1 \le x_2$ and $G(y_1) \le G(y_2)$ for $y_1 \le y_2$, and

$$\Delta H = \iint_{[F(x_1), F(x_2)] \times [G(y_1), G(y_2)]} c_\theta(u, v)\, du\, dv. \tag{2.4}$$

Since $c_\theta(u, v) \ge 0$, $\Delta H \ge 0$ follows. Proof completed.

We turn now to give a characterization of the density, $h_\theta(x, y)$, of $H_\theta(x, y)$. For a fixed $(x, y) \in R^2$, the two components of $(x, y)$ correspond to either a point of continuity or a jump point of the corresponding marginal distribution function. For $(X, Y) \sim H_\theta$, $h_\theta(x, y)$ has four different expressions, namely,

$$h_\theta(x,y) = \begin{cases} \lim_{a,b \to 0} \dfrac{P\{X \in (x, x+a], Y \in (y, y+b]\}}{ab} & \text{if } x \in \mathcal{C}(F),\, y \in \mathcal{C}(G), \\[2ex] \lim_{b \to 0} \dfrac{P\{X = x, Y \in (y, y+b]\}}{b} & \text{if } x \in \mathcal{J}(F),\, y \in \mathcal{C}(G), \\[2ex] \lim_{a \to 0} \dfrac{P\{X \in (x, x+a], Y = y\}}{a} & \text{if } x \in \mathcal{C}(F),\, y \in \mathcal{J}(G), \text{ and} \\[2ex] P\{X = x, Y = y\} & \text{if } x \in \mathcal{J}(F),\, y \in \mathcal{J}(G), \end{cases} \tag{2.5}$$

The following theorem gives workable expressions for $h_\theta(x, y)$ in terms of the copula:

**Theorem 2.1.2.** *For each* $(x, y) \in R^2$,

$$h_\theta(x,y) = \begin{cases} f(x)\, g(y)\, c_\theta(F(x), G(y)) & \textit{if } x \in \mathcal{C}(F),\, y \in \mathcal{C}(G), \\[2ex] y(y) \displaystyle\int_{[F(x^-), F(x)]} c_\theta(u, G(y))\, du & \textit{if } x \in \mathcal{J}(F),\, y \in \mathcal{C}(G), \\[2ex] f(x) \displaystyle\int_{[G(y^-), G(y)]} c_\theta(F(x), v)\, dv & \textit{if } x \in \mathcal{C}(F),\, y \in \mathcal{J}(G), \\[2ex] \displaystyle\iint_{[F(x^-), F(x)] \times [G(y^-), G(y)]} c_\theta(u, v)\, du\, dv & \textit{if } x \in \mathcal{J}(F),\, y \in \mathcal{J}(G). \end{cases} \tag{2.6}$$

**Proof:** When $x \in \mathcal{C}(F)$ and $y \in \mathcal{C}(G)$,

$$P\{X \in (x, x+a], Y \in (y, y+b]\} = \iint_{[F(x), F(x+a)] \times [G(y), G(y+b)]} c_\theta(u, v)\, du\, dv$$

from (2.4). This is approximately

$$ab\, f(x)\, g(y)\, c_\theta(F(x), G(y))$$

for small $a$ and $b$.

When $x \in \mathcal{J}(F)$ and $y \in \mathcal{C}(G)$, note that

$$P\{X = x, Y \in (y, y+b]\} = \iint_{[F(x^-), F(x)] \times [G(y), G(y+b)]} c_\theta(u, v)\, du\, dv,$$

again from (2.4), which is approximately $b\ g(y) \int_{[F(x^-),F(x)]} c_\theta(u, G(y))\, du$ for small $b$. The third case follows similarly and the fourth expression is straightforward. Proof completed.

The terms in $h_\theta(x, y)$ that involve the densities $f$ and $g$ do not depend on $\theta$ and can, hence, be ignored for the estimation of $\theta$. Subsequently, we focus on the function $c^*_\theta \equiv c^*_\theta(F(x^-), F(x), G(y^-), G(y))$ defined by

$$c^*_\theta = \begin{cases} c_\theta(F(x), G(y)) & \text{if } x \in \mathcal{C}(F),\, y \in \mathcal{C}(G), \\ \displaystyle\int_{[F(x^-),F(x)]} c_\theta(u, G(y))\, du & \text{if } x \in \mathcal{J}(F),\, y \in \mathcal{C}(G), \\ \displaystyle\int_{[G(y^-),G(y)]} c_\theta(F(x), v)\, dv & \text{if } x \in \mathcal{C}(F),\, y \in \mathcal{J}(G), \\ \displaystyle\iint_{[F(x^-),F(x)]\times[G(y^-),G(y)]} c_\theta(u, v)\, du\, dv & \text{if } x \in \mathcal{J}(F),\, y \in \mathcal{J}(G). \end{cases} \tag{2.7}$$

## 2.2 Semi-Parametric Estimation of $\theta$

If not all marginals are continuous, $C$ in (1.4) is no longer unique. Uniqueness is an important consideration when estimating the unknown parameters corresponding to the copula function. Since our marginals have discrete components, one way to achieve uniqueness is to restrict $C$ to belong to a particular parametric family. From now on, let $\{C_\theta :\ \theta \in \Theta\}$ be a restricted parametric family of copulas.

Suppose $\{\, (X_j, Y_j)^T,\ j = 1, 2, \ldots, n \,\}$ is the set of $n$ independent and identically distributed bivariate random vectors arising from the joint distribution $H_\theta(x, y)$ in (2.3). The parameter of interest is the bivariate dependence parameter, $\theta$, and the log-likelihood function corresponding to the $n$ observations is

$$\sum_{j=1}^{n} \log h_\theta(x_j, y_j).$$

The estimation of $\theta$ is complicated by the presence of nuisance parameters consisting of the unknown distributions $F$ and $G$ (only the number $d_F$ and jump points $\mathcal{D}_{Fh}$, $h = 1, 2, \ldots, d_F$

of $F$ (correspondingly, $d_G$ and points $\mathcal{D}_{Gl}$ of $G$) are known). An objective function that has $\theta$ as the only unknown parameter can be obtained by replacing $F$ and $G$ by $F_n$ and $G_n$ in the log-likelihood function, where $F_n$ (and respectively, $G_n$) is $n/(n+1)$ times the empirical cdf of $F$ (respectively, $G$). The rescaling by $n/(n+1)$ avoids difficulties arising from the unboundedness of the log-likelihood function as the empirical cdfs tend to 0 or 1, and has been employed by Genest et al. in [11]. Thus, the unique semi-parametric maximum likelihood estimator of $\theta$, $\hat{\theta}_n$, is the maximizer of

$$L(\theta) = \sum_{j=1}^{n} \log \{c^*_{\theta,j}(F_n(x_j^-), F_n(x_j), G_n(y_j^-), G_n(y_j))\} \tag{2.8}$$

(that is, $\hat{\theta}_n = \arg\max_{\theta \in \Theta} L(\theta)$), where

$$c^*_{\theta,j} \equiv c^*_\theta(F_n(x_j^-), F_n(x_j), G_n(y_j^-), G_n(y_j)),$$

ignoring terms that do not depend on $\theta$ in $\sum_{j=1}^n \log h_\theta(x_j, y_j)$. In the special case when no discrete components are present, Genest et al. [11] developed a semi-parametric approach to estimate the unknown parameters based on the semi-parametric likelihood. Subsequently, it was shown that the resulting estimators were consistent and asymptotically normally distributed; see [11] for details. Expressions (2.7) and (2.8), respectively, are generalizations of the methodology of Genest et al. [11] when $F$ and $G$ contain discrete components. Note that the challenge in maximizing the semiparameteric log-likelihood in (2.8) is that it involves several integrals corresponding to discrete components in $(x_j, y_j)$, $j = 1, 2, \ldots, n$.

# CHAPTER 3

# ASYMPTOTIC PROPERTIES OF SEMI-PARAMETRIC ESTIMATOR IN BIVARIATE DISTRIBUTIONS

## 3.1 Statement of the Main Theorem

In view of its similarity with the semi-parametric maximum likelihood estimator for continuous marginals, we expect that $\hat{\theta}_n$ in the case of mixed marginals to be consistent and asymptotically normal. To prove this, we denote $l(\theta, u_1, u_2, v_1, v_2) = \log\{c^*_\theta(u_1, u_2, v_1, v_2)\}$ and use the notation $l_\theta$ and $l_{\theta,\theta}$ to denote the first and second derivative of $l$ with respect to $\theta$. The estimator $\hat{\theta}_n$ satisfies

$$\frac{1}{n}\frac{\partial L(\theta)}{\partial \theta} = \frac{1}{n}\sum_{j=1}^{n} l_\theta(\theta, F_n(x_j^-), F_n(x_j), G_n(y_j^-), G_n(y_j)) = 0. \tag{3.1}$$

Here is some heuristics of the proof of asymptotic normality. Expanding in a Taylor's series, one obtains,

$$\frac{1}{n}\frac{\partial L(\theta)}{\partial \theta}\bigg|_{\theta=\hat{\theta}_n} = 0 \approx A_n - (\hat{\theta}_n - \theta)B_n \tag{3.2}$$

where

$$A_n = \frac{1}{n}\sum_{j=1}^{n} l_\theta(\theta, F_n(x_j^-), F_n(x_j), G_n(y_j^-), G_n(y_j))$$

and

$$B_n = -\frac{1}{n}\sum_{j=1}^{n} l_{\theta,\theta}(\theta, F_n(x_j^-), F_n(x_j), G_n(y_j^-), G_n(y_j)).$$

It follows from equation (3.2) that $n^{1/2}(\hat{\theta}_n - \theta) \approx n^{1/2}A_n/B_n$. From the theory of multivariate rank statistics, one can get the following limiting behaviors:

$$B_n \to \beta = -E(l_{\theta,\theta}\{\theta, F(X^-), F(X), G(Y^-), G(Y)\}) \tag{3.3}$$

almost surely, and $n^{1/2}A_n$ is asymptotically normal with zero mean and variance $\sigma^2$, of which the explicit form will be provided in Section 3.4.

Thus, we have

**Theorem 3.1.1.** *Under suitable regularity conditions, the semi-parametric estimator $\hat{\theta}_n$ is consistent and $n^{1/2}(\hat{\theta}_n - \theta)$ is asymptotically normal with mean $0$ and variance $\varrho = \sigma^2/\beta^2$.*

We start with the proofs of consistency in Section 3.2, the asymptotic properties of $B_n$ and $A_n$ in Section 3.3 and Section 3.4 respectively, on which Theorem 3.1.1 is based, and complete this Chapter by proving Theorem 3.1.1 in Section 3.5.

Here are some notations which will be used later in this Chapter. Let $\mathcal{S}_{Fh} = [F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh})]$, for $h = 1, 2, ..., d_F$, $\mathcal{S}_{Gl} = [F(\mathcal{D}_{Gl}^-), F(\mathcal{D}_{Gl})]$ for $l = 1, 2, ..., d_G$. Under the situation that both $F$ and $G$ only have one jump point, the previous notations can be simplified to $\mathcal{D}_F$, $\mathcal{D}_G$, $\mathcal{S}_F = [F(\mathcal{D}_F^-), F(\mathcal{D}_F)]$, and $\mathcal{S}_G = [G(\mathcal{D}_G^-), G(\mathcal{D}_G)]$, respectively.

## 3.2 Consistency of $\hat{\theta}_n$

Nothing that $\hat{\theta}_n$ satisfying (3.1), by letting $J = l_\theta$, which is a function on $(0,1)^2$, we only need to show that

$$R_n = \frac{1}{n} \sum_{j=1}^{n} J(F_n(X_j), G_n(Y_j)) \to 0 \qquad a.s. \tag{3.4}$$

Since $H_\theta$ involves one or more jump points, $R_n$ can be written as

$$R_n = \frac{1}{n} \sum_{j=1}^{n} J(F_n(X_j^-), F_n(X_j), G_n(Y_j^-), G_n(Y_j)) \tag{3.5}$$

We introduce some notations for the subsequent presentation. Let $\{cc\}$, $\{cd\}$, $\{dc\}$, and $\{dd\}$ be the events that $\{(X,Y) : X \in \mathcal{C}(F), Y \in \mathcal{C}(G)\}$, $\{(X,Y) : X \in \mathcal{C}(F), Y \in \mathcal{J}(G)\}$, $\{(X,Y) : X \in \mathcal{J}(F), Y \in \mathcal{C}(G)\}$ and $\{(X,Y) : X \in \mathcal{J}(F), Y \in \mathcal{J}(G)\}$, respectively. Further, let $A_{cc} = \{j : (X_j, Y_j) \in \{cc\}\}$. Similarly, one can define $A_{cd}$, $A_{dc}$ and $A_{dd}$. Note that the sets $A_S$ for $S = \{cc\}$, $\{cd\}$, $\{dc\}$, and $\{dd\}$ is a partition of the set of integers $\{1, 2, \ldots, n\}$. We consider the decomposition of $R_n$ based on these four partition sets, namely, $R_n = R_n^{cc} + R_n^{cd} + R_n^{dc} + R_n^{dd}$ where

$$R_n^S = \frac{1}{n} \sum_{j \in A_S} J^S(F_n(X_j^-), F_n(X_j), G_n(Y_j^-), G_n(Y_j)),$$

for $S = \{cc\}$, $\{cd\}$, $\{dc\}$, and $\{dd\}$, and $J^S$ is given by

$$J^S(u_1, u_2, v_1, v_2) = \begin{cases} J^{cc}(u_2, v_2) & \text{if } S = \{cc\} \\ J^{cd}(u_2, v_1, v_2) & \text{if } S = \{cd\} \\ J^{dc}(u_1, u_2, v_2) & \text{if } S = \{dc\} \\ J^{dd}(u_1, u_2, v_1, v_2) & \text{if } S = \{dd\}. \end{cases} \tag{3.6}$$

The functions $J^S$, where $S = \{cc\}$, $\{cd\}$, $\{dc\}$ and $\{dd\}$, are assumed to be continuous on their respective domains. For example, $J^{cc}(u_2, v_2)$ is assumed to be continuous on $[0,1]^2$; this is a reasonable assumption to make since candidates for $J^{cc}$ will be either $\frac{\partial}{\partial\theta} \log c_\theta(u_2, v_2)$ or

$\frac{\partial^2}{\partial\theta^2}\log c_\theta(u_2, v_2)$ in this Chapter, which are continuous functions of $u_2$ and $v_2$. $J^{cd}(u_2, v_1, v_2)$ is assumed to be continuous on $[0,1]^3$. A particular candidate of $J^{cd}$ is

$$J^{cd}(u_2, v_1, v_2) = \frac{\partial}{\partial\theta}\log\int_{[v_1,v_2]} c_\theta(u_2, v)dv$$

which is continuous in $u_2$, $v_1$ and $v_2$. Similarly for $J^{dc}$ and $J^{dd}$.

Almost sure convergence of $R_n$ is established in the following theorem:

**Theorem 3.2.1.** *Let $J = l_\theta$, $r(u) = u(1-u)$, $\delta > 0$, $p$ and $q$ are positive numbers satisfying $1/p + 1/q = 1$. Let $a$ and $b$ be numbers given by $a = (-1+\delta)/p$ and $b = (-1+\delta)/q$. Consider the conditions*

*(C1) $J^{cc}(u_2, v_2) \leq M_1\, r(u_2)^a r(v_2)^b$,*

*(C2) $J^{cd}(u_2, v_1, v_2) \leq M_2\, r(u_2)^a$ (independent of $v_1$ and $v_2$), and*

*(C3) $J^{dc}(u_1, u_2, v_2) \leq M_3\, r(v_2)^b$ (independent of $u_1$ and $u_2$).*

*(C4) In a small neighborhood of each $(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl})) \in [0,1]^4$, for $h = 1, 2, ..., d_F$ and $l = 1, 2, ..., d_G$, $J^{dd}(u_1, u_2, v_1, v_2)$ is finite.*

*Under conditions (C1-C4), $R_n \to \kappa (= 0)$ almost surely. where*

$$\kappa = E[N^{cc}(X,Y) + N^{cd}(X,Y) + N^{dc}(X,Y) + N^{dd}(X,Y)], \tag{3.7}$$

*with $E[N^{cc}(X,Y)]$, $E[N^{cd}(X,Y)]$, $E[N^{dc}(X,Y)]$ and $E[N^{dd}(X,Y)]$ are given as below, respectively,*

$$\iint_{\{(0,1)\cap(\cup_{h=1}^{d_F}\mathcal{S}_{Fh})^c\}\times\{(0,1)\cap(\cup_{l=1}^{d_G}\mathcal{S}_{Gl})^c\}} J^{cc}(u,v)c_\theta(u,v)\,du\,dv,$$

$$\int_{(0,1)\cap(\cup_{h=1}^{d_F}\mathcal{S}_{Fh})^c}\sum_{l=1}^{d_G}\left\{J^{cd}(u, G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))\int_{\mathcal{S}_{Gl}} c_\theta(u,v)\,dv\right\}du,$$

$$\int_{(0,1)\cap(\cup_{l=1}^{d_G}\mathcal{S}_{Gl})^c}\sum_{h=1}^{d_F}\left\{J^{dc}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), v)\int_{\mathcal{S}_{Fh}} c_\theta(u,v)\,du\right\}dv,$$

$$\sum_{h,l} J^{dd}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))\int_{\mathcal{S}_{Fh}\times\mathcal{S}_{Gl}} c_\theta(u,v)\,du\,dv.$$

**Remark:** In the case of $t$-copulas (as defined in (6.1)), $a$ and $b$ can be chosen such that

$$a \le -\frac{2}{\nu},\ b \le -\frac{2}{\nu},\ a = \frac{-1+\delta}{p},\ b = \frac{-1+\delta}{q} \text{ for some } p,\ q \text{ and } \delta.$$

**Proof:** Without loss of generality, we consider the case that there is only one point of discontinuity in both $F$ and $G$, i. e., $\mathcal{D}_F$ and $\mathcal{D}_G$. Let $C_n(u,v)$ be the empirical copula defined by the sample

$$C_n(u,v) = \frac{1}{n}\sum_{j=1}^{n} I_{\{F_n(X_j)\le u, G_n(Y_j)\le v\}}.$$

We consider the decomposition of the empirical copula measure into 4 sub-measures $C_n^{cc}$, $C_n^{cd}$, $C_n^{dc}$, and $C_n^{dd}$, where

$$C_n^S(u,v) = \frac{1}{n}\sum_{j\in A_S}^{n} I_{\{F_n(X_j)\le u, G_n(Y_j)\le v\}},$$

for $S = \{cc\}$, $\{cd\}$, $\{dc\}$ and $\{dd\}$, and

$$C_n(u,v) = \sum_{S\in\{cc,cd,dc,dd\}} C_n^S(u,v).$$

Then by the Glivenko-Cantelli Theorem, $C_n^{cc}(u,v)$ converges uniformly to $C^1(u,v)$, $C_n^{cd}(u,v)$ converges uniformly to $C^2(u,v)$, $C_n^{dc}(u,v)$ converges uniformly to $C^3(u,v)$, and $C_n^{dd}(u,v)$ converges uniformly to $C^4(u,v)$ respectively, and $C^i(u,v)$ for $i=1,2,3,4$, are given by

$$C^1(u,v) = \begin{cases} P(F(X)\le u \text{ and } G(Y)\le v) & \text{if } (u,v)\in(0,F(\mathcal{D}_F^-))\times(0,G(\mathcal{D}_G^-)) \\ P(F(X)\in(0,u)\cap\mathcal{S}_F^c \text{ and } G(Y)\le v) & \text{if } (u,v)\in[F(\mathcal{D}_F^-),1)\times(0,G(\mathcal{D}_G^-)) \\ P(F(X)\le u \text{ and } G(Y)\in(0,v)\cap\mathcal{S}_G^c) & \text{if } (u,v)\in(0,F(\mathcal{D}_F^-))\times[G(\mathcal{D}_G^-),1) \\ P(F(X)\in(0,u)\cap\mathcal{S}_F^c \text{ and } G(Y)\in(0,v)\cap\mathcal{S}_G^c) & \text{if } (u,v)\in[F(\mathcal{D}_F^-),1)\times[G(\mathcal{D}_G^-),1) \end{cases}$$

$$C^2(u,v) = \begin{cases} 0 & \text{if } v\in(0,G(\mathcal{D}_G)) \\ P(F(X)\le u \text{ and } Y=\mathcal{D}_G) & \text{if } (u,v)\in(0,F(\mathcal{D}_F^-))\times[G(\mathcal{D}_G),1) \\ P(F(X)\in(0,u)\cap\mathcal{S}_F^c \text{ and } Y=\mathcal{D}_G) & \text{if } (u,v)\in[F(\mathcal{D}_F^-),1)\times[G(\mathcal{D}_G),1) \end{cases}$$

$$C^3(u,v) = \begin{cases} 0 & \text{if } u \in (0, F(\mathcal{D}_F)) \\ P(X = \mathcal{D}_F \text{ and } G(Y) \le v) & \text{if } (u,v) \in [F(\mathcal{D}_F), 1) \times (0, G(\mathcal{D}_G^-)) \\ P(X = \mathcal{D}_F \text{ and } G(Y) \in (0,v) \cap \mathcal{S}_G^c) & \text{if } (u,v) \in [F(\mathcal{D}_F), 1) \times [G(\mathcal{D}_G^-), 1) \end{cases}$$

and $C^4(u,v)$ only put mass $P(X = \mathcal{D}_F$ and $Y = \mathcal{D}_G)$ on a single point $\{F(\mathcal{D}_F)\} \times \{G(\mathcal{D}_G)\}$. Note that $C^i(u,v)$, $i = 1, ..., 4$, are not probability measures and

$$C_\theta(u,v) = \sum_{i=1}^{4} C^i(u,v).$$

To obtain the almost sure convergence, it takes four steps.

**Step 1.** Let

$$R_n^{cc} = \iint_{(0,1)^2} J^{cc}(u,v) dC_n^{cc}(u,v).$$

Now we prove that $R_n^{cc} \to \mu^{cc} \equiv E[N^{cc}(X,Y)]$ almost surely. We show that $J^{cc}$ is uniformly integrable with respect to the measures $dC_n^{cc}(u,v)$ by showing $\iint_{(0,1)^2} |J^{cc}(u,v)|^{1+\epsilon} dC_n^{cc}(u,v)$ is bounded for some $\epsilon > 0$. Using the Hölder's inequality and the assumption (C1), one can derive the following chain of inequalities, noting that $dC_n^{cc}(u,v)$ putting mass $\frac{1}{n}$ on each continuous point:

$$\begin{aligned} & \iint_{(0,1)^2} |J^{cc}(u,v)|^{1+\epsilon} dC_n^{cc}(u,v) \\ \le\ & M_1 \iint_{(0,1)^2} r(u)^{a(1+\epsilon)} r(v)^{b(1+\epsilon)} dC_n^{cc}(u,v) \\ \le\ & M_1 \left\{ \int_{(0,1)^2} r(u)^{a(1+\epsilon)p} \, dC_n^{cc}(u,v) \right\}^{1/p} \left\{ \int_{(0,1)^2} r(u)^{b(1+\epsilon)q} \, dC_n^{cc}(u,v) \right\}^{1/q} \\ \le\ & M_1 \left\{ \frac{1}{n} \sum_{j \in A_{cc}} r\left(\frac{j}{n+1}\right)^{a(1+\epsilon)p} \right\}^{1/p} \left\{ \frac{1}{n} \sum_{j \in A_{cc}} r\left(\frac{j}{n+1}\right)^{b(1+\epsilon)q} \right\}^{1/q} \\ =\ & \frac{M_1}{n} \sum_{j \in A_{cc}} r\left(\frac{j}{n+1}\right)^{(-1+\delta)(1+\epsilon)} \\ \le\ & M_1 \int_0^1 \frac{1}{\{u(1-u)^{(1-\delta)(1+\epsilon)}\}} du < \infty, \end{aligned}$$

for $\epsilon < \delta$. Combining the result above with the fact that $C_n^{cc}(u,v)$ uniformly converges to $C^1(u,v)$, we have

$$R_n^{cc} \to \iint_{(0,1)^2} J^{cc}(u,v)dC^1(u,v) = \iint_{((0,1)\cap\mathcal{S}_F^c)\times((0,1)\cap\mathcal{S}_G^c)} J^{cc}(u,v)c_\theta(u,v)\,dudv$$

almost surely, which completes step 1.

**Step 2.** Let

$$R_n^{cd} = \iint_{(0,1)^2} J^{cd}(u,v_1,v_2)dC_n^{cd}(u,v).$$

Now we prove that $R_n^{cd} \to \mu^{cd} \equiv E[N^{cd}(X,Y)]$ almost surely. We show that $J^{cd}(u,v_1,v_2)$ is uniformly integrable with respect to the measures $dC_n^{cd}(u,v)$ by showing $\iint_{(0,1)^2} |J^{cd}(u,v_1,v_2)|^{1+\epsilon}dC_n^{cd}(u,v)$ is bounded for some $\epsilon > 0$. Using the Hölder's inequality and the assumption (C2), one can derive the following chain of inequalities, noting that $dC_n^{cd}(u,v)$ putting mass $\frac{1}{n}$ on each point on $((0,1)\cap\mathcal{S}_F^c)\times G_n(\mathcal{D}_G)$:

$$\begin{aligned}
&\iint_{(0,1)^2} |J^{cd}(u,v_1,v_2)|^{1+\epsilon}dC_n^{cd}(u,v) \\
\leq\ & M_2 \iint_{(0,1)^2} r(u)^{a(1+\epsilon)}dC_n^{cd}(u,v) \\
\leq\ & M_2 \left\{\int_{(0,1)^2} r\,(u)^{a(1+\epsilon)p}\,dC_n^{cd}(u,v)\right\}^{1/p}\left\{\int_{(0,1)^2} 1^{(1+\epsilon)q}dC_n^{cd}(u,v)\right\}^{1/q} \\
\leq\ & M_2 \left\{\frac{1}{n}\sum_{j\in A_{cd}} r\left(\frac{j}{n+1}\right)^{a(1+\epsilon)p}\right\}^{1/p} \cdot 1 \\
=\ & M_2 \left\{\frac{1}{n}\sum_{j\in A_{cd}} r\left(\frac{j}{n+1}\right)^{(-1+\delta)(1+\epsilon)}\right\}^{1/p} \\
\leq\ & M_2 \left\{\int_0^1 \frac{1}{\{u(1-u)^{(1-\delta)(1+\epsilon)}\}}du\right\}^{1/p} \\
<\ & \infty,
\end{aligned}$$

if $\epsilon < \delta$. Combining the result above with the fact that $C_n^{cd}(u,v)$ uniformly converges to $C^2(u,v)$, we have

$$
\begin{aligned}
R_n^{cd} &\to \iint_{(0,1)^2} J^{cd}(u, G(\mathcal{D}_G^-), G(\mathcal{D}_G))dC^2(u,v) \\
&= \int_{((0,1)\cap\mathcal{S}_F^c)} J^{cd}(u, G(\mathcal{D}_G^-), G(\mathcal{D}_G)) \int_{\mathcal{S}_G} c_\theta(u,v)\, dv\, du
\end{aligned}
$$

almost surely, which completes step 2.

**Step 3.** Under the assumption (C3), the proof for $R_n^{dc} \to \mu^{dc} \equiv E[N^{dc}(X,Y)]$ is similar to that in step 2, so it is omitted.

**Step 4.** The convergence of $R_n^{dd}$ can be established using the SLLN. Let

$$
R_n^{dd} = \iint_{(0,1)^2} J^{dd}(u_1,u_2,v_1,v_2)dC_n^{dd}(u,v).
$$

Now we prove that $R_n^{dd} \to \mu^{dd} \equiv E[N^{dd}(X,Y)]$ almost surely. We show that $J^{dd}(u_1,u_2,v_1,v_2)$ is uniformly integrable with respect to the measures $dC_n^{dd}(u,v)$ by showing $\iint_{(0,1)^2} |J^{dd}(u_1,u_2,v_1,v_2)|^{1+\epsilon} dC_n^{dd}(u,v)$ is bounded for some $\epsilon > 0$. Noting that $dC_n^{dd}(u,v)$ putting mass $\dfrac{n_{dd}^*}{n}$ on the single point $(F_n(\mathcal{D}_F), G_n(\mathcal{D}_G))$, where $n_{dd}^*$ is the number of observations of $(X_j, Y_j)$ such that $X_j = \mathcal{D}_F$ and $Y_j = \mathcal{D}_G$:

$$
\begin{aligned}
& \iint_{(0,1)^2} |J^{dd}(u_1,u_2,v_1,v_2)|^{1+\epsilon} dC_n^{dd}(u,v) \\
= \ & \frac{n_{dd}^*}{n} |J^{dd}(u_1,u_2,v_1,v_2)|^{1+\epsilon} \\
< \ & \infty,
\end{aligned}
$$

as long as $J^{dd}(u_1,u_2,v_1,v_2)$ is finite, which is the assumption (C4). Combining the result above with the fact that $C_n^{dd}(u,v)$ uniformly converges to $C^4(u,v)$, we have

$$
\begin{aligned}
R_n^{dd} &\to \iint_{(0,1)^2} J^{dd}(u_1,u_2,v_1,v_2)dC^4(u,v) \\
&= J^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G)) \iint_{\mathcal{S}_F\times\mathcal{S}_G} c_\theta(u,v)\, du\, dv
\end{aligned}
$$

almost surely, which completes step 4.

In summary, note that

$$R_n = R_n^{cc} + R_n^{cd} + R_n^{dc} + R_n^{dd},$$

Theorem 3.2.1 is true.

## 3.3 Asymptotic Behavior of $B_n$

In the case when $H_\theta$ and $J$ are both continuous, in [11], by letting $J = l_{\theta,\theta}$, one can see of interest is the asymptotic behavior of the statistic $R_n$ defined by

$$R_n = \frac{1}{n}\sum_{j=1}^{n} J(F_n(X_j), G_n(Y_j)) \tag{3.8}$$

where $J(\cdot,\cdot)$ is a function on $(0,1)^2$. Genest et al. [11] have shown that

$$B_n \to -E(l_{\theta,\theta}\{\theta, F(X), G(Y)\})$$

as $n \to \infty$ almost surely.

In the present case, the appropriate statistic is similar in form to $R_n$ but with a significant difference, $H_\theta$ involves one or more jump points. Therefore, $R_n$ can be written of the form as in (3.5), with $J = l_{\theta,\theta}$. Almost sure convergence of $R_n$ is established in the following theorem:

**Theorem 3.3.1.** *Let $J = l_{\theta,\theta}$, $r(u) = u(1-u)$, $\delta > 0$, $p$ and $q$ are positive numbers satisfying $1/p + 1/q = 1$. Let $a$ and $b$ be numbers given by $a = (-1+\delta)/p$ and $b = (-1+\delta)/q$. Consider the conditions*

*(C1) $J^{cc}(u_2, v_2) \le M_1\, r(u_2)^a r(v_2)^b$,*

*(C2) $J^{cd}(u_2, v_1, v_2) \le M_2\, r(u_2)^a$ (independent of $v_1$ and $v_2$), and*

*(C3) $J^{dc}(u_1, u_2, v_2) \le M_3\, r(v_2)^b$ (independent of $u_1$ and $u_2$).*

*(C4) In a small neighborhood of each $(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl})) \in [0,1]^4$, for $h = 1, 2, ..., d_F$ and $l = 1, 2, ..., d_G$, $J^{dd}(u_1, u_2, v_1, v_2)$ is finite.*

*Under conditions (C1-C4), $R_n \to \beta$ almost surely where*

$$\beta = E[N^{cc}(X,Y) + N^{cd}(X,Y) + N^{dc}(X,Y) + N^{dd}(X,Y)], \tag{3.9}$$

*with* $E[N^S(X,Y)]$ *defined as in (3.7), and* $S = \{cc\}$, $\{cd\}$, $\{dc\}$.

Since the proof is similar to that in Theorem 3.2.1, it is omitted here.

## 3.4 Asymptotic Behavior of $A_n$

By taking $J = l_\theta$ in (3.5), we have the following Theorem:

**Theorem 3.4.1.** *Let* $r(u) = u(1-u)$, $\delta > 0$, $p$ *and* $q$ *be as in Theorem 3.3.1. Let* $J_u^S$ *and* $J_v^S$ *be the partial derivatives of* $J^S$ *with respect to* $u$ *and* $v$*, respectively, for* $S = \{cc\}$, $\{cd\}$, $\{dc\}$*, and* $\{dd\}$. *Also, let* $a$ *and* $b$ *be numbers given by* $a = (-0.5+\delta)/p$ *and* $b = (-0.5+\delta)/q$. *Consider the conditions*

$(D1)$ $\quad J^{cc}(u_2, v_2) \le M_1\, r(u_2)^a r(v_2)^b$, *with partial derivatives satisfying*

$J^{cc}_{u_2}(u_2, v_2) \le M_2 r(u_2)^{a-1} r(v_2)^b$ *and* $J^{cc}_{v_2}(u_2, v_2) \le M_3 r(u_2)^a r(v_2)^{b-1}$,

$(D2)$ $\quad J^{cd}(u_2, v_1, v_2) \le M_4\, r(u_2)^a$ *with* $J^{cd}_{u_2}(u_2, v_1, v_2) \le M_5\, r(u_2)^{a-1}$,

$$\int_{v_1}^{v_2} c_\theta(u_2, v)dv \le M_6\, r(u_2)^a,$$

$$\int_{(0,1)\cap\{\cup_h \mathcal{S}_{Fh}\}^c} \sum_l \left\{ |J^{cd}_\eta(u_2, G(\mathcal{D}^-_{Gl}), G(\mathcal{D}_{Gl}))| \int_{\mathcal{S}_{Gl}} c_\theta(u_2, v)dv \right\} du_2 < \infty$$

*with* $\eta = v_1$ *or* $v_2$, $J^{cd}_{u_2}(u_2, v_1, v_2)$ *is continuous w.r.t.* $u_2$ *on* $\mathcal{C}(F)$ *almost surely,* $J^{cd}_{v_1}(u_2, v_1, v_2)$ *and* $J^{cd}_{v_2}(u_2, v_1, v_2)$ *are continuous w.r.t.* $v_1$ *and* $v_2$ *respectively in the small neighborhoods of rectangles* $\mathcal{C}(F) \times (G(\mathcal{D}^-_{Gl}), G(\mathcal{D}_{Gl})]$ *almost surely for* $l = 1, ..., D_G$,

$(D3)$ $\quad J^{dc}(u_1,u_2,v_2) \le M_7\, r(v_2)^b$ *with* $J^{dc}_{v_2}(u_1,u_2,v_2) \le M_8\, r(v_2)^{b-1}$,

$\int_{u_1}^{u_2} c_\theta(u,v_2)du \le M_9\, r(v_2)^b$,

$\int_{(0,1)\cap\{\cup_l \mathcal{S}_{Gl}\}^c} \sum_h \left\{ |J^{dc}_\eta(F(\mathcal{D}^-_{Fh}), F(\mathcal{D}_{Fh}), v_2)| \int_{\mathcal{S}_{Fh}} c_\theta(u,v_2)du \right\} dv_2 < \infty$

*with* $\eta = u_1$ *or* $u_2,$ $J^{dc}_{v_2}(u_1,u_2,v_2)$ *is continuous w.r.t.* $v_2$ *on* $\mathcal{C}(G)$ *almost surely,* $J^{dc}_{u_1}(u_1,u_2,v_2)$ *and* $J^{dc}_{u_2}(u_1,u_2,v_2)$ *are continuous w.r.t.* $u_1$ *and* $u_2$ *respectively in the small neighborhoods of rectangles* $(F(\mathcal{D}^-_{Fh}), F(\mathcal{D}_{Fh})] \times \mathcal{C}(G)$ *almost surely for* $h = 1, ..., D_F$,

$(D4)$ $\quad J^{dd}_{u_1}(u_1,\cdot,\cdot,\cdot),$ $J^{dd}_{u_2}(\cdot,u_2,\cdot,\cdot),$ $J^{dd}_{v_1}(\cdot,\cdot,v_1,\cdot),$ *and* $J^{dd}_{v_2}(\cdot,\cdot,\cdot,v_2)$ *are continuous w.r.t.* $u_1, u_2, v_1$ *and* $v_2,$ *respectively, in a small neighborhood of* $(F(\mathcal{D}^-_{Fh}), F(\mathcal{D}_{Fh}),$ $G(\mathcal{D}^-_{Gl}), G(\mathcal{D}_{Gl}))$ *for any* $h = 1, ..., D_F$ *and* $l = 1, ..., D_G$.

*Under conditions (D1-D4),* $n^{1/2}(R_n - \kappa) \to N(0,\sigma^2)$ *in distribution as* $n \to \infty$*, where* $\kappa$ *is defined in (3.7), and*

$$\sigma^2 = \textit{var}\left[M^{cc}(X,Y) + M^{cd}(X,Y) + M^{dc}(X,Y) + M^{dd}(X,Y)\right], \tag{3.10}$$

*with* $M^{cc}(x,y)$*,* $M^{cd}(x,y)$*,* $M^{dc}(x,y)$ *and* $M^{dd}(x,y)$ *are given as below, respectively,*

$$\begin{aligned} &J^{cc}(F(x),G(y))I_{\{x\in\mathcal{C}(F),y\in\mathcal{C}(G)\}} \\ &+ \iint_{\{R\backslash\{\cup_h\mathcal{D}_{Fh}\}\}\times\{R\backslash\{\cup_l\mathcal{D}_{Gl}\}\}} I_{\{x\le x'\}}\, J^{cc}_{u_2}(F(x'),G(y'))dH_\theta(x',y') \\ &+ \iint_{\{R\backslash\{\cup_h\mathcal{D}_{Fh}\}\}\times\{R\backslash\{\cup_l\mathcal{D}_{Gl}\}\}} I_{\{y\le y'\}}\, J^{cc}_{v_2}(F(x'),G(y'))dH_\theta(x',y'), \end{aligned} \tag{3.11}$$

$$
\begin{aligned}
&\sum_{l=1}^{d_G} J^{cd}(F(x), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))I_{\{x\in\mathcal{C}(F),\, y=\mathcal{D}_{Gl}\}}\\
&+\sum_{l=1}^{d_G}\int_{R\setminus\{\cup_h \mathcal{D}_{Fh}\}} I_{\{x\le x'\}}\, J_{u2}^{cd}(F(x'), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))dH_\theta(x', \mathcal{D}_{Gl})\\
&+\sum_{l=1}^{d_G}\int_{R\setminus\{\cup_h \mathcal{D}_{Fh}\}} I_{\{y<\mathcal{D}_{Gl}\}}\, J_{v1}^{cd}(F(x'), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))dH_\theta(x', \mathcal{D}_{Gl})\\
&+\sum_{l=1}^{d_G}\int_{R\setminus\{\cup_h \mathcal{D}_{Fh}\}} I_{\{y\le\mathcal{D}_{Gl}\}}\, J_{v2}^{cd}(F(x'), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))dH_\theta(x', \mathcal{D}_{Gl}),
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{h=1}^{d_F} J^{dc}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(y))I_{\{y\in\mathcal{C}(G)\}}\\
&+\sum_{h=1}^{d_F}\int_{R\setminus\{\cup_l \mathcal{D}_{Gl}\}} I_{\{y\le y'\}} J_{v2}^{dc}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(y'))dH_\theta(\mathcal{D}_{Fh}, y')\\
&+\sum_{h=1}^{d_F}\int_{R\setminus\{\cup_l \mathcal{D}_{Gl}\}} I_{\{x<\mathcal{D}_{Fh}\}}\, J_{u1}^{dc}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(y'))dH_\theta(\mathcal{D}_{Fh}, y')\\
&+\sum_{h=1}^{d_F}\int_{R\setminus\{\cup_l \mathcal{D}_{Gl}\}} I_{\{x\le\mathcal{D}_{Fh}\}}\, J_{u2}^{dc}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(y'))dH_\theta(\mathcal{D}_{Fh}, y'),
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{h,l} J^{dd}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))I_{\{x=\mathcal{D}_{Fh},\, y=\mathcal{D}_{Gl}\}}\\
&+\sum_{h,l}\Big[ I_{\{y<\mathcal{D}_{Gl}\}}\, J_{v1}^{dd}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))\\
&+I_{\{y\le\mathcal{D}_{Gl}\}}\, J_{v2}^{dd}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))\\
&+I_{\{x<\mathcal{D}_{Fh}\}}\, J_{u1}^{dd}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))\\
&+I_{\{x\le\mathcal{D}_{Fh}\}}\, J_{u2}^{dd}(F(\mathcal{D}_{Fh}^-), F(\mathcal{D}_{Fh}), G(\mathcal{D}_{Gl}^-), G(\mathcal{D}_{Gl}))\Big]\\
&\cdot P(X=\mathcal{D}_{Fh},\ Y=\mathcal{D}_{Gl}).
\end{aligned}
$$

In the case when both $H_\theta$ and $J$ are continuous, Genest et al. [11] showed that the statistic $R_n$

was a special case of multivariate rank order statistics whose asymptotic behavior was thoroughly studied by Ruymgaart et al. [41], Rumyaart [42] and Rüschendorf [40]. For continuous $H_\theta$ and $J$ , Genest et al. [11] proposed regularity conditions ensuring almost sure convergence and asymptotic normality. In the present case, $H_\theta$ involves one or more jump points, which causes the $J$ function to be discontinuous on $(0,1)^2$ as well.

In the rest of this Chapter, we will provide a scratch proof of Theorem 3.4.1. Without loss of generality, we will show the theorem is true in the case that there is only one discontinuity point in both $F$ and $G$, which should be completed in the subsequent sections, using arguments similar to those used in [41].

We shall need the following empirical processes

$$\begin{aligned}
&U_n(F(x)) = n^{1/2}(F_n(x) - F(x)),\\
&U_n(F(x^-)) = n^{1/2}(F_n(x^-) - F(x^-)),\\
&V_n(G(y)) = n^{1/2}(G_n(y) - G(y))),\\
&V_n(G(y^-)) = n^{1/2}(G_n(y^-) - G(y^-)).
\end{aligned}$$

Note that

$$P(\Omega_0) = P(\{\omega : F_n(F^{-1}(F)) = F_n, G_n(G^{-1}(G)) = G_n, \text{for all } x, y \text{ and } n\}) = 1. \tag{3.12}$$

The above identities, $F_n(F^{-1}(F)) = F_n$ and $G_n(G^{-1}(G)) = G_n$ are true even for the case when there are jump points. At the jump point of $F$, one can see that $F^{-1}(F(\mathcal{D}_F)) = \mathcal{D}_F$. The similar result holds for $G$ as well.

### 3.4.1 Asymptotic normality of $R_n^{cc}$

For small positive $\gamma$ define the set

$$\Omega_{\gamma_n} = \{\omega : \sup|F_n - F| < \gamma/2,\ \sup|G_n - G| < \gamma/2\}. \tag{3.13}$$

For $\omega \in \Omega_0 \cap \Omega_{\gamma_n}$, the Mean Value Theorem gives

$$n^{1/2} J^{cc}(F_n, \cdot) = n^{1/2} J^{cc}(F, \cdot) + U_n(F) J^{cc}_{u2}(\Phi_n, \cdot),$$

for all $x \in \mathcal{C}(F)$. In the formula above $\Phi_n$ is defined by $\Phi_n = F + \eta(F_n - F)$, where $\eta = \eta(\omega, x, n)$ is a number between $0$ and $1$. Let

$$\Delta_F = [X_{1n}, \mathcal{D}_F) \bigcup \, (\mathcal{D}_F, X_{nn}],$$
$$\Delta_G = [Y_{1n}, \mathcal{D}_G) \bigcup \, (\mathcal{D}_G, Y_{nn}],$$

where

$$X_{1n}(\text{ or } Y_{1n}) = \min_{1 \le j \le n} X_j \text{ ( or } Y_j),$$
$$X_{nn}(\text{ or } Y_{nn}) = \max_{1 \le j \le n} X_j \text{ ( or } Y_j).$$

Note that

$$n^{1/2}(R_n^{cc} - \mu^{cc}) = \sum_{i=1}^{3} A_{in} + \sum_{i=1}^{4} B_{\gamma_{in}} + B_{5n} + C_n,$$

where

$$\begin{aligned}
\mu^{cc} = \quad & E[N^{cc}(X,Y)] \text{ with } l_{\theta,\theta} \text{ replaced by } l_\theta \text{ in (3.9)},\\
A_{1n} = \quad & n^{1/2} \iint_{\{R \backslash \mathcal{D}_F\} \times \{R \backslash \mathcal{D}_G\}} J^{cc}(F(x), G(y)) d[H_n(x,y) - H_\theta(x,y)],\\
A_{2n} = \quad & \iint_{\{R \backslash \mathcal{D}_F\} \times \{R \backslash \mathcal{D}_G\}} U_n(F(x)) J^{cc}_{u2}(F(x), G(y)) dH_\theta(x,y),\\
A_{3n} = \quad & \iint_{\{R \backslash \mathcal{D}_F\} \times \{R \backslash \mathcal{D}_G\}} V_n(G(y)) J^{cc}_{v2}(F(x), G(y)) dH_\theta(x,y),\\
B_{\gamma\, 1n} = \quad & \chi(\Omega^c_{\gamma_n}) \{ n^{1/2} \iint_{\{R \backslash \mathcal{D}_F\} \times \{R \backslash \mathcal{D}_G\}} [J^{cc}(F_n(x), G(y))\\
& - J^{cc}(F(x), G(y))] dH_n(x,y) - A_{2n} \},
\end{aligned}$$

$$
\begin{aligned}
B_{\gamma\,2n} = {}& \chi(\Omega_{\gamma n}) \iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} U_n(F(x))[J^{cc}_{u2}(\Phi_n(x), G(y)) \\
& -J^{cc}_{u2}(F(x), G(y))\Big]\, dH_n(x,y), \\
B_{\gamma\,3n} = {}& \chi(\Omega_{\gamma n}) \iint_{\{\Delta_F\times\Delta_G\}\cap\{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}\}} U_n(F(x))\cdot \\
& J^{cc}_{u2}(F(x), G(y)) d[H_n(x,y) - H_\theta(x,y)], \\
B_{\gamma\,4n} = {}& \chi(\Omega_{\gamma n}) \Bigg\{ \iint_{\{\Delta_F\times\Delta_G\}\cap\{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}\}} U_n(F(x))\cdot \\
& J^{cc}_{u2}(F(x), G(y)) dH_\theta(x,y) - A_{2n}\Bigg\}, \\
B_{5n} = {}& n^{1/2} \iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} [J^{cc}(F(x), G_n(y)) \\
& -J^{cc}(F(x), G(y))]\, dH_n(x,y) - A_{3n}, \\
C_n = {}& n^{1/2} \iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} [J^{cc}(F_n(x), G_n(y)) - J^{cc}(F_n(x), G(y)) \\
& -J^{cc}(F(x), G_n(y)) + J^{cc}(F(x), G(y))]\, dH_n(x,y),
\end{aligned}
$$

where $\chi(\Omega_{\gamma n})$ denotes the indicator function of $\Omega_{\gamma n}$; $H_n(x,y)$ is the empirical cumulative distribution of $H_\theta(x,y)$ .

Note that with a further decomposition of $C_n$, one has

$$
\begin{aligned}
C_{\gamma\,1n} = {}& \chi(\Omega^c_{\gamma n}) n^{1/2} \iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} [J^{cc}(F_n(x), G_n(y)) - J^{cc}(F_n(x), G(y)) \\
& -J^{cc}(F(x), G_n(y)) + J^{cc}(F(x), G(y))]\, dH_n(x,y), \\
C_{\gamma\,2n} = {}& \chi(\Omega_{\gamma n}) \iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} U_n(F(x))[J^{cc}_{u2}(\Phi_n(x), G_n(y)) \\
& -J^{cc}_{u2}(\Phi_n(x), G(y))\Big]\, dH_n(x,y).
\end{aligned}
$$

We start with a few lemmas to be used in the proof.

**Lemma 3.3.1** For any $\xi \geq 0$ and function $r(u) = u(1-u)$, the function $r(u)^{-\xi}$ is symmetric about $\frac{1}{2}$, decreasing on $\left(0, \frac{1}{2}\right]$ and has the property that for each $\beta$ in $(0,1)$ there exits a constant $M = M_\beta$ such that

$$r(\beta s)^{-\xi} \leq M r(s)^{-\xi} \quad \text{for } 0 < s \leq \frac{1}{2},$$

and

$$r(1-\beta(1-s))^{-\xi} \leq M r(s)^{-\xi} \quad \text{for } \frac{1}{2} < s < 1.$$

Proof of Lemma 3.4.1 can be found in [41].

**Lemma 3.3.2** Let $\Phi_n$ and $\Psi_n$ be functions on $\overline{\Delta}_{n1}$ and $\overline{\Delta}_{n2}$, where

$$\overline{\Delta}_{n1} = \begin{cases} [X_{1n}, X_{nn}], & \text{if } X_{nn} \neq \mathcal{D}_{d_F} \\ \\ [X_{1n}, X_{nn}), & \text{if } X_{nn} = \mathcal{D}_{d_F} \end{cases}$$

and

$$\overline{\Delta}_{n2} = \begin{cases} [Y_{1n}, Y_{nn}], & \text{if } Y_{nn} \neq \mathcal{D}_{d_G} \\ \\ [Y_{1n}, Y_{nn}), & \text{if } Y_{nn} = \mathcal{D}_{d_G} \end{cases}$$

respectively, satisfying

$$\min(F, F_n) \leq \Phi_n \leq \max(F, F_n)$$

and

$$\min(G, G_n) \leq \Psi_n \leq \max(G, G_n)$$

where defined. Then uniformly for $n = 1, 2, \cdots$,

$$\text{(i)} \quad \sup_{\overline{\Delta}_{n1}} r(\Phi_n)^{-\xi} r(F)^{\xi} = O_p(1), \text{ for each } \xi \geq 0,$$

$$\text{(ii)} \quad \sup_{\overline{\Delta}_{n2}} r(\Psi_n)^{-\eta} r(G)^{\eta} = O_p(1), \text{ for each } \eta \geq 0.$$

**Proof:** It suffices to prove (i). From Lemma A.3 in [45] and (3.12), it follows that one needs to show that for each $\epsilon > 0$, there exists a constant $\beta = \beta_\epsilon$ in $(0,1)$ such that

$$P(\Omega_n) = P(\{\omega : \beta F \leq F_n \leq 1 - \beta(1-F) \text{ any } x(\omega) \text{ on } \overline{\Delta}_{n1}\}) > 1 - \epsilon, \tag{3.14}$$

for all $n$ and $F$.

Let $A = \{\omega : \beta F \le F_n \le 1 - \beta(1-F)$ any $x(\omega) \in \overline{\Delta}_{n1} \cap \mathcal{C}(F)\}$, and $B = \{\omega : \beta F \le F_n \le 1 - \beta(1-F)$ any $x(\omega) \in \overline{\Delta}_{n1} \cap \mathcal{J}(F)\}$.

Note that in our case,

$$\begin{aligned}
& P(\{\omega : \beta F \le F_n \le 1 - \beta(1-F) \text{ any } x(\omega) \text{ on } \overline{\Delta}_{n1}\}) \\
= \; & P(A \cup B) \\
= \; & P(A) + P(B) \\
= \; & P(\{\omega : x(\omega) \in \overline{\Delta}_{n1} \cap \mathcal{C}(F)\}) \cdot P(\{\omega : \beta F \le F_n \le 1 - \beta(1-F) \mid x(\omega) \in \overline{\Delta}_{n1} \cap \mathcal{C}(F)\}) \\
& + P(\{\omega : x(\omega) \in \overline{\Delta}_{n1} \cap \mathcal{J}(F)\}) \cdot \\
& P(\{\omega : \beta F \le F_n \le 1 - \beta(1-F) \mid x(\omega) \in \overline{\Delta}_{n1} \cap \mathcal{J}(F)\}).
\end{aligned}$$

In Lemma 6.1 in [41] (3.14) has been verified for all continuous $F$ for a constant $\beta = \beta_\epsilon'$, which gives us $P(\{\omega : \beta F \le F_n \le 1 - \beta(1-F) \mid x(\omega) \in \mathcal{C}(F) \cap \overline{\Delta}_{n1}\}) > 1 - \epsilon$. Noting that

$$P(\{\omega : x(\omega) \in \overline{\Delta}_{n1} \cap \mathcal{C}(F)\}) + P(\{\omega : x(\omega) \in \overline{\Delta}_{n1} \cap \mathcal{J}(F)\}) = 1,$$

to prove (3.14) we only need to show when $x(\omega) \in \overline{\Delta}_{n1} \cap \mathcal{J}(F)$, $P(\{\omega : \beta F \le F_n \le 1 - \beta(1 - F) \mid$ any $x(\omega) \in \overline{\Delta}_{n1} \cap \mathcal{J}(F)\}) > 1 - \epsilon$ is true for a constant $\beta = \beta_\epsilon''$. This is a direct result of the Glivenko-Cantelli Theorem, which completes the proof.

**Lemma 3.3.3** Uniformly in all $F$, we have

(i) $\sup\limits_{\overline{\Delta}_{n1}} |U_n(F) - U_n^*(F)| r(F)^{\rho - 1/2} \to^p 0,$ as $n \to \infty,$ for each $\rho \ge 0,$

(ii) $\sup\limits_{(-\infty,\infty)\backslash\{\cup_h \mathcal{D}_{Fh}\}} |U_n(F)| r(F)^{\rho - 1/2} = O_p(1),$ as $n \to \infty,$ for each $\rho \ge 0,$

where $U_n^*(F(x)) = n^{1/2}(\hat{F}_n(x) - F(x))$, and $\hat{F}_n$ is the empirical distribution of $F$. Note that $F_n$ was defined to be $\dfrac{n}{n+1}\hat{F}_n$.

**Proof:** (i) Note that

$$|U_n(F) - U_n^*(F)| r(F)^{\rho - 1/2} = n^{1/2} \frac{\hat{F}_n}{n+1} r(F)^{\rho - 1/2},$$

and that for any fixed $\beta \in (0,1)$, we have

$$r\left(\frac{\beta}{n}\right)^{\rho-1/2} = r\left(1-\frac{\beta}{n}\right)^{\rho-1/2} = O(n^{-\rho+1/2}).$$

Since $F(X_i)$ are i.i.d. uniform random variables, given any arbitrary $\epsilon > 0$, we can choose a $\beta = \beta_\epsilon$ in $(0,1)$ such that

$$P\left(\frac{\beta}{n} \le F(X_{1n}) \le F(X_{nn}) \le 1-\frac{\beta}{n}\right) > 1-\epsilon$$

for all $n$ and all $F$ with $F(\mathcal{D}_F) \neq 1$, which is obvious since

$$\begin{aligned} & P\left(\frac{\beta}{n} \le F(X_{1n}) \le F(X_{nn}) \le 1-\frac{\beta}{n}\right) \\ = & \left(1-\frac{\beta}{n}-\frac{\beta}{n}\right)^n \\ = & \left(1-\frac{2\beta}{n}\right)^n \to e^{-2\beta} > 1-\epsilon, \end{aligned}$$

for sufficiently small $\beta$.

Combining with all above results, we proved (i). (ii) follows from (i) and (iii) of Lemma 4.2 in [41]. Proof completed.

**Proof of Asymptotic normality of $R_n^{cc}$.**

(1) Note that $A_{1n}$ can be written as

$$A_{1n} = n^{-1/2}\sum_{j=1}^{n} A_{1jn},$$

where $A_{1jn} = J^{cc}(F(X_j), G(Y_j)) \cdot I_{\{j \in A_{cc}\}} - \mu^{cc}$. The $A_{1jn}$ are i.i.d. with mean zero. By the assumption $(D1)$ and the H$\ddot{o}$lder's inequity, we have

$$\begin{aligned}
&\iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} |J^{cc}(F(x),G(y))|^{2+\delta_0} dH_\theta(x,y)\\
\leq\ & \iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} M_1^{2+\delta_0}|r(F(x))^{a(2+\delta_0)}r(G(y))^{b(2+\delta_0)}|dH_\theta(x,y)\\
\leq\ & M_1^{2+\delta_0}\left[\int_{(0,1)} r(u)^{a(2+\delta_0)p_0}du\right]^{1/p_0}\left[\int_{(0,1)} r(v)^{b(2+\delta_0)q_0}dv\right]^{1/q_0}\\
=\ & M_1^{2+\delta_0}\left[\int_{(0,1)} r(u)^{(-\frac{1}{2}+\delta)(2+\delta_0)}du\right]^{1/p_0}\left[\int_{(0,1)} r(v)^{(-\frac{1}{2}+\delta)(2+\delta_0)}dv\right]^{1/q_0}\\
<\ & \infty,
\end{aligned}$$

for some selected $p_0$, $q_0$ and $\delta$ satisfying $(1-\delta)(2+\delta_0)<1$, which means that $A_{1jn}$ has a finite absolute moment of order $2+\delta_0$ for some $\delta_0>0$. Moreover, the term on the left hand is uniformly bounded above of order $2+\delta_0$. Using the CLT, we get the asymptotic normality of $A_{1n}$.

(2) Note that $A_{2n}$ can be written as

$$\begin{aligned}
A_{2n} &= n^{1/2}\iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}}(\hat{F}_n(x)-F(x))J^{cc}_{u_2}(F(x),G(y))dH_\theta(x,y)\\
&\quad +n^{1/2}\iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}}(F_n(x)-\hat{F}_n(x))J^{cc}_{u_2}(F(x),G(y))dH_\theta(x,y)\\
&= A^*_{2n1}+A^*_{2n2},
\end{aligned}$$

where $\hat{F}_n(x)$ is the empirical distribution function of $X$. Let

$$\phi_{X_j}(x)=\begin{cases}0 & \text{if } x<X_j\\ 1 & \text{if } x\geq X_j\end{cases}$$

Then $A^*_{2n1}$ can be written as $n^{-1/2}\sum_{j=1}^{n}A_{2jn}$, where $A_{2jn} = \iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}}(\phi_{X_j}(x)-F(x))J^{cc}_{u_2}(F(x),G(y))dH_\theta(x,y)$ are i.i.d. with mean zero. Note that $|\phi_{X_j}(x)-F(x)|\leq 1$. Under the assumption (D1), we have

$$|A_{2jn}|\leq M_2\iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} r^{a-1}(F(x))r^b(G(y))dH_\theta(x,y).$$

Note that $\iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} r^{a-1}(F(x))r^b(G(y))dH_\theta(x,y) < \infty$ uniformly as long as we can find some $p_1$ and $q_1$ satisfying $1/p_1 + 1/q_1 = 1$ and $(a-1)p_1 > -1$ and $bq_1 > -1$. Thus, we have shown that $A^*_{2n1}$ has an absolute moment of order $2+\delta_1$, which leads to the asymptotic normality of $A^*_{2n1}$. Note that with the same $p_1$ and $q_1$

$$\begin{aligned} A^*_{2n2} &= \frac{n^{1/2}}{n+1}\iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} \hat{F}_n(x)J^{cc}_{u2}(F(x),G(y))dH_\theta(x,y) \\ &\leq \frac{n^{1/2}}{n+1}\sup_x|\hat{F}_n(x)|\iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}} J^{cc}_{u2}(F(x),G(y))dH_\theta(x,y) \\ &= o_p(1). \end{aligned}$$

Therefore, we have the asymptotic normality of $A_{2n}$.

(3) Similarly, note that $A_{3n}$ can be written as

$$\begin{aligned} A_{3n} &= n^{1/2}\iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}}(\hat{G}_n(y)-G(y))J^{cc}_{v2}(F(x),G(y))dH_\theta(x,y) \\ &\quad +n^{1/2}\iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}}(G_n(y)-\hat{G}_n(y))J^{cc}_{v2}(F(x),G(y))dH_\theta(x,y) \\ &= A^*_{3n1}+A^*_{3n2}, \end{aligned}$$

where $\hat{G}_n(y)$ is the empirical distribution function of $Y$. Similarly, the $A^*_{3n1}$ can be written as $n^{-1/2}\sum_{j=1}^n A_{3jn}$, where $A_{3jn} = \iint_{\{R\backslash\mathcal{D}_F\}\times\{R\backslash\mathcal{D}_G\}}(\phi_{Y_j}(y) - G(y))J^{cc}_{v2}(F(x),G(y))dH_\theta(x,y)$ are i.i.d. with mean zero. The same asymptotic conclusion can be drawn for $A_{3n}$.

(4) Note that the $A_{1n}+A_{2n}+A_{3n}$ can be written as $\sum_{j=1}^n n^{-1/2}(A_{1jn}+A_{2jn}+A_{3jn}) + A^*_{2n2}+A^*_{3n2} \equiv \sum_{j=1}^n n^{-1/2}A^{**}_{1jn}+A^*_{2n2}+A^*_{3n2}$, where the $A^{**}_{1jn} = A_{1jn}+A_{2jn}+A_{3jn}$ depends on $(X_j,Y_j)$ only, hence are i.i.d. with mean zero, and the terms $A^*_{2n2} \to^p 0$ and $A^*_{3n2} \to^p 0$ as $n\to\infty$. Using the CLT, we have that the $A_{1n}+A_{2n}+A_{3n}$ is asymptotically normally distributed with mean $0$.

The asymptotic negligibility of the $B-$ and $C-$ terms will be stated as corollaries.

**Corollary 1.** *For fixed* $\gamma$, $B_{\gamma 1n} \to^p 0$ *and* $C_{\gamma 1n} \to^p 0$ *as* $n \to \infty$.

**Proof:** Note that $P(\Omega_{\gamma n}^c) \to 0$ in (3.13) for any $H_\theta$ by the Glivenko-Cantelli Theorem, and because the distribution of $\sup_x |F_n(x) - F(x)|$ does not depend on $H_\theta$. This completes the proof.

**Corollary 2.** *For fixed* $\gamma$, $B_{\gamma 2n} \to^p 0$ *and* $C_{\gamma 2n} \to^p 0$ *as* $n \to \infty$.

**Proof:** According to Lemma 3.3.3 (ii) with $\rho = \frac{1}{2}$, for given $\epsilon > 0$, there exists a constant $M'$ such that

$$P(\Omega_{1n}) = P(\omega : \{\sup_x |U_n(F)| \le M' \text{ and } x \in \mathcal{C}(F)\}) > 1 - \epsilon \tag{3.15}$$

for all $n$ and $H_\theta$, and $M''$ such that

$$P(\Omega_{2n}) = P(\omega : \{\sup_{\overline{\Delta}_{n1}} r^{-\xi}(\Phi_n) r^{\xi}(F) \le M''\}) > 1 - \epsilon,$$

which gives $P(\Omega_{1n} \cap \Omega_{2n}) > 1 - 2\epsilon$. Also,

$$\begin{aligned}
& \chi(\Omega_{1n} \cap \Omega_{2n})|B_{\gamma 2n}| \\
\le\ & M' \ \sup\nolimits_{\Delta_F \times \Delta_G} \left| J_{u2}^{cc}(\Phi_n(x), G(y)) - J_{u2}^{cc}(F(x), G(y)) \right|. \\
& + \chi(\Omega_{1n} \cap \Omega_{2n}) \iint_{\{(-\infty, X_{1n}) \cup (X_{nn}, \infty)\} \times \{(-\infty, Y_{1n}) \cup (Y_{nn}, \infty)\}} U_n(F(x)) \cdot \\
& [J_{u2}^{cc}(\Phi_n(x), G(y)) - J_{u2}^{cc}(F(x), G(y))] dH_n(x, y)
\end{aligned}$$

By the Glivenko-Cantelli Theorem,

$$P(\{(-\infty, X_{1n}) \cup (X_{nn}, \infty)\} \times \{(-\infty, Y_{1n}) \cup (Y_{nn}, \infty)\}) \to 0$$

for any $H_\theta$, so the second term on the right side of the inequality above converges to 0, as $n \to \infty$. The function $J_{u2}^{cc}(u_2, v_2)$ is uniformly continuous on $(0,1)^2$. Since $|\Phi_n - F| \le |F_n - F|$ where $\Phi_n$ is defined, the Glivenko-Cantelli Theorem yields $\sup_{\Delta_F \times \Delta_G} \left| J_{u2}^{cc}(\Phi_n(x), G(y)) - J_{u2}^{cc}(F(x), G(y)) \right| \to^p 0$ uniformly for $H_\theta$. Therefore, we have shown that $B_{\gamma 2n} \to^p 0$ as $n \to \infty$.

A similar argument may be used for $|C_{\gamma 2n}|$. Proof completed.

**Corollary 3.** *For fixed* $\gamma$, $B_{\gamma 3n} \to^p 0$ *as* $n \to \infty$.

**Proof:** Noting that $\{\Delta_F \times \Delta_G\} \cap \{\{R \backslash \mathcal{D}_F\} \times \{R \backslash \mathcal{D}_G\}\} = \Delta_F \times \Delta_G$,

$$\begin{aligned} B_{\gamma 3n} &= \chi(\Omega_{\gamma n}) \iint_{\Delta_F \times \Delta_G} U_n(F(x)) J^{cc}_{u2}(F(x), G(y)) d[H_n(x,y) - H_\theta(x,y)] \\ &= \chi(\Omega_{\gamma n} \cap \Omega_{1n}) \iint_{\Delta_F \times \Delta_G} U_n(F(x)) J^{cc}_{u2}(F(x), G(y)) d[H_n(x,y) - H_\theta(x,y)] \\ &+ \chi(\Omega_{\gamma n} \cap \Omega^c_{1n}) \iint_{\Delta_F \times \Delta_G} U_n(F(x)) J^{cc}_{u2}(F(x), G(y)) d[H_n(x,y) - H_\theta(x,y)] \end{aligned} \tag{3.16}$$

where $\Omega_{1n}$ is as defined in (3.15).

Note that the second term in (3.16) converges to $0$ in probability, and

$$\begin{aligned} & \chi(\Omega_{\gamma n} \cap \Omega_{1n}) \iint_{\Delta_F \times \Delta_G} U_n(F(x)) J^{cc}_{u2}(F(x), G(y)) d[H_n(x,y) - H_\theta(x,y)] \\ \leq\ & \chi(\Omega_{\gamma n} \cap \Omega_{1n}) M' \iint_{\Delta_F \times \Delta_G} J^{cc}_{u2}(F(x), G(y)) d[H_n(x,y) - H_\theta(x,y)]. \\ \leq\ & \chi(\Omega_{\gamma n} \cap \Omega_{1n}) M' \, M''' \iint_{\Delta_F \times \Delta_G} d[H_n(x,y) - H_\theta(x,y)] \end{aligned}$$

where $M''' = \sup_{\Delta_F \times \Delta_G} \left| J^{cc}_{u2}(F(x), G(y)) \right|$. From the Theorem 1.9, (i), from [43], the above integral converges to $0$ in probability as $n \to \infty$. Proof completed.

**Corollary 4.** *For fixed* $\gamma$, $B_{\gamma 4n} \to^p 0$ *as* $n \to \infty$.

**Proof:** Note that

$$\begin{aligned} B_{\gamma 4n} \;=\; & \chi(\Omega_{\gamma n} \cap \Omega_{1n}) \left\{ \iint_{\Delta_F \times \Delta_G} U_n(F(x)) J^{cc}_{u2}(F(x), G(y)) dH_\theta(x,y) - A_{2n} \right\} \\ & + \chi(\Omega_{\gamma n} \cap \Omega^c_{1n}) \left\{ \iint_{\Delta_F \times \Delta_G} U_n(F(x)) J^{cc}_{u2}(F(x), G(y)) dH_\theta(x,y) - A_{2n} \right\} \end{aligned}$$

$$
\begin{aligned}
= \quad & -\chi(\Omega_{\gamma n} \cap \Omega_{1n}) \iint_{\{(-\infty, X_{1n}) \cup (X_{nn}, \infty)\} \times \{(-\infty, Y_{1n}) \cup (Y_{nn}, \infty)\}} U_n(F(x)) \cdot \\
& J^{cc}_{u_2}(F(x), G(y))\, dH_\theta(x, y) \\
& + \chi(\Omega_{\gamma n} \cap \Omega^c_{1n}) \left\{ \iint_{\Delta_F \times \Delta_G} U_n(F(x)) J^{cc}_{u_2}(F(x), G(y)) dH_\theta(x, y) - A_{2n} \right\}
\end{aligned}
$$

where $\Omega_{1n}$ is as defined in (3.15). By the Glivenko-Cantelli Theorem, $P(\{(-\infty, X_{1n}) \cup (X_{nn}, \infty)\} \times \{(-\infty, Y_{1n}) \cup (Y_{nn}, \infty)\}) \rightarrow 0$ for any $H_\theta$. Combining with the fact that $J^{cc}_{u_2}$ is uniformly continuous on $(0, 1)^2$, the righthand side converges to $0$, as $n \rightarrow \infty$.

To see $B_{5n}$ converging to 0 in probability as $n \rightarrow \infty$, let us notice that

$$
\begin{aligned}
\sum_{i=1}^{4} B_{\gamma in} \quad & = n^{1/2} \iint_{\{R \backslash \mathcal{D}_F\} \times \{R \backslash \mathcal{D}_G\}} [J^{cc}(F(x), G_n(y)) - J^{cc}(F(x), G(y))] dH_n(x, y) \\
& - A_{2n}.
\end{aligned}
$$

In summary, we have shown that all these $B$-terms and $C$-term are negligible. Combining with the results of these $A$-terms, we have established the asymptotic normality of $R^{cc}_n$.

### 3.4.2 Asymptotic normality of $R^{cd}_n$ and $R^{dc}_n$

It suffices to show the asymptotic normality of $R^{cd}_n$. Note that

$$
n^{1/2}(R^{cd}_n - \mu^{cd}) = \sum_{i=1}^{4} A_{in} + \sum_{i=1}^{6} B_{in},
$$

where

$$
\begin{aligned}
\mu^{cd} &= E[N^{cd}(X,Y)],\\
A_{1n} &= n^{1/2}\int_{\{R\backslash\mathcal{D}_F\}} J^{cd}(F(x),G(\mathcal{D}_G^-),G(\mathcal{D}_G))d[H_n(x,\mathcal{D}_G)-H_\theta(x,\mathcal{D}_G)];\\
A_{2n} &= \int_{\{R\backslash\mathcal{D}_F\}} U_n(F(x))J_{u2}^{cd}(F(x),G(\mathcal{D}_G^-),G(\mathcal{D}_G))dH_\theta(x,\mathcal{D}_G);\\
A_{3n} &= \int_{\{R\backslash\mathcal{D}_F\}} V_n(G(\mathcal{D}_G^-))J_{v1}^{cd}(F(x),G(\mathcal{D}_G^-),G(\mathcal{D}_G))dH_\theta(x,\mathcal{D}_G);\\
A_{4n} &= \int_{\{R\backslash\mathcal{D}_F\}} V_n(G(\mathcal{D}_G))J_{v2}^{cd}(F(x),G(\mathcal{D}_G^-),G(\mathcal{D}_G))dH_\theta(x,\mathcal{D}_G);\\
B_{1n} &= \int_{\{R\backslash\mathcal{D}_F\}} U_n(F(x))J_{u2}^{cd}(\Phi_n(x),G_n(\mathcal{D}_G^-),G_n(\mathcal{D}_G))d[H_n(x,\mathcal{D}_G)-H_\theta(x,\mathcal{D}_G)]\\
B_{2n} &= \int_{\{R\backslash\mathcal{D}_F\}} V_n(G(\mathcal{D}_G^-))J_{v1}^{cd}(F(x),\Theta_n(\mathcal{D}_G^-),G_n(\mathcal{D}_G))d[H_n(x,\mathcal{D}_G)-H_\theta(x,\mathcal{D}_G)]\\
B_{3n} &= \int_{\{R\backslash\mathcal{D}_F\}} V_n(G(\mathcal{D}_G))J_{v2}^{cd}(F(x),G(\mathcal{D}_G^-),\Psi_n(\mathcal{D}_G))d[H_n(x,\mathcal{D}_G)-H_\theta(x,\mathcal{D}_G)]\\
B_{4n} &= \int_{\{R\backslash\mathcal{D}_F\}} U_n(F(x))[J_{u2}^{cd}(\Phi_n(x),G_n(\mathcal{D}_G^-),G_n(\mathcal{D}_G))\\
&\quad -J_{u2}^{cd}(F(x),G(\mathcal{D}_G^-),G(\mathcal{D}_G))\Big]\,dH_\theta(x,\mathcal{D}_G)\\
B_{5n} &= \int_{\{R\backslash\mathcal{D}_F\}} V_n(G(\mathcal{D}_G^-))[J_{v1}^{cd}(F(x),\Theta_n(\mathcal{D}_G^-),G_n(\mathcal{D}_G))\\
&\quad -J_{v1}^{cd}(F(x),G(\mathcal{D}_G^-),G(\mathcal{D}_G))\Big]\,dH_\theta(x,\mathcal{D}_G)\\
B_{6n} &= \int_{\{R\backslash\mathcal{D}_F\}} V_n(G(\mathcal{D}_G))[J_{v2}^{cd}(F(x),G(\mathcal{D}_G^-),\Psi_n(\mathcal{D}_G))\\
&\quad -J_{v2}^{cd}(F(x),G(\mathcal{D}_G^-),G(\mathcal{D}_G))\Big]\,dH_\theta(x,\mathcal{D}_G)
\end{aligned}
$$

where $\Phi_n(\cdot)$, $\Theta_n(\cdot)$, and $\Psi_n(\cdot)$ are defined by the Mean Value Theorem.

Next, we will show that the $A-$ terms are asymptotic normal and the $B-$ terms converge to $0$ in probability.

(1) Note that $A_{1n}$ can be written as

$$A_{1n} = n^{-1/2} \sum_{j=1}^{n} A_{1jn},$$

where $A_{1jn} = J^{cd}(F(X_j), G(\mathcal{D}_G{}^-), G(\mathcal{D}_G)) \cdot I_{\{j \in A_{cd}\}} - \mu^{cd}$. Note that $A_{1jn}$ are i.i.d. with mean zero. Since

$$J^{cd}(u, v_1, v_2) = \frac{\partial}{\partial \theta} \log \int_{v_1}^{v_2} c_\theta(u, v) dv,$$

using assumption (D2), we have

$$\begin{aligned}
& \int_{\{R \setminus \mathcal{D}_F\}} |J^{cd}(F(x), G(\mathcal{D}_G{}^-), G(\mathcal{D}_G))|^{2+\delta_0} dH_\theta(x, \mathcal{D}_G) \\
\leq & \left\{ \int_{\{R \setminus \mathcal{D}_F\}} |J^{cd}(F(x), G(\mathcal{D}_G{}^-), G(\mathcal{D}_G))|^{(2+\delta_0)p_0} dH_\theta(x, \mathcal{D}_G) \right\}^{1/p_0} \\
& \cdot \left\{ \int_{\{R \setminus \mathcal{D}_F\}} 1^{(2+\delta_0)q_0} dH_\theta(x, \mathcal{D}_G) \right\}^{1/q_0} \\
\leq & \left\{ \int_{(0,1) \cap \mathcal{S}_F^c} |J^{cd}(u, G(\mathcal{D}_G{}^-), G(\mathcal{D}_G))|^{(2+\delta_0)p_0} \int_{\mathcal{S}_G} c_\theta(u, v) dv \, du \right\}^{1/p_0}.
\end{aligned}$$

Under the assumption (D2),

$$\begin{aligned}
& \int_{(0,1) \cap \mathcal{S}_F^c} |J^{cd}(u, v_1, v_2)|^{(2+\delta_0)p_0} \int_{\mathcal{S}_G} c_\theta(u, v) dv \, du \\
\leq & \ M_4 \int_{(0,1) \cap \mathcal{S}_F^c} r(u)^{a(2+\delta_0)p_0} \int_{\mathcal{S}_G} c_\theta(u, v) dv \, du \\
\leq & \ M_4 \ M_6 \int_{(0,1) \cap \mathcal{S}_F^c} r(u)^{a(2+\delta_0)p_0} r(u)^a \, du \\
< & \ \infty,
\end{aligned}$$

if $a(2+\delta_0)p_0 + a > -1$. By the CLT, we have the asymptotic normality of $A_{1n}$.

(2) Note that $A_{2n}$ can be written as

$$\begin{aligned} A_{2n} &= n^{1/2}\int_{\{R\backslash\mathcal{D}_F\}}(\hat{F}_n(x)-F(x))J^{cd}_{u_2}(F(x),G(\mathcal{D}_G{}^-),G(\mathcal{D}_G))dH_\theta(x,\mathcal{D}_G) \\ &\quad +n^{1/2}\int_{\{R\backslash\mathcal{D}_F\}}(F_n(x)-\hat{F}_n(x))J^{cd}_{u_2}(F(x),G(\mathcal{D}_G{}^-),G(\mathcal{D}_G))dH_\theta(x,\mathcal{D}_G) \\ &= A^*_{2n1}+A^*_{2n2}. \end{aligned}$$

Furthermore, $A^*_{2n1}$ can be written as $n^{-1/2}\sum_{j=1}^n A_{2jn}$, where $A_{2jn}=\int_{\{R\backslash\mathcal{D}_F\}}(\phi_{X_j}(x)-F(x))J^{cd}_{u_2}(F(x),G(\mathcal{D}_G{}^-),G(\mathcal{D}_G))dH_\theta(x,\mathcal{D}_G)$ are i.i.d. with mean 0. Using the assumption (D2), we have

$$\begin{aligned} |A_{2jn}| &\le M_5\int_{(0,1)\cap\mathcal{S}^c_F} r^{a-1}(u)\int_{\mathcal{S}_G} c_\theta(u,v)dv\,du \\ &\le M_5\,M_6\int_{(0,1)\cap\mathcal{S}^c_F} r^{2a-1}(u)du\,<\infty \end{aligned}$$

as long as $(2a-1)>-1$. Thus we have the asymptotic normality of $A^*_{2n1}$. Using the assumption (D2), we have

$$\begin{aligned} A^*_{2n2} &= \frac{\sqrt{n}}{n+1}\int_{\{R\backslash\mathcal{D}_F\}}\hat{F}_n(x)J^{cd}_{u_2}(F(x),G(\mathcal{D}_G{}^-),G(\mathcal{D}_G))dH_\theta(x,\mathcal{D}_G) \\ &\le \frac{\sqrt{n}}{n+1}\sup_x|\hat{F}_n(x)|\int_{\{R\backslash\mathcal{D}_F\}}J^{cd}_{u_2}(F(x),G(\mathcal{D}_G{}^-),G(\mathcal{D}_G))dH_\theta(x,\mathcal{D}_G) \\ &= o_p(1). \end{aligned}$$

Therefore, the asymptotic normality of $A_{2n}$ has been established.

(3) Note that $A_{3n}$ can be written as

$$\begin{aligned} A_{3n} &= n^{1/2}\int_{\{R\backslash\mathcal{D}_F\}}[\hat{G}_n(\mathcal{D}_G{}^-)-G(\mathcal{D}_G{}^-)]J^{cd}_{v_1}(F(x),G(\mathcal{D}_G{}^-),G(\mathcal{D}_G))dH_\theta(x,\mathcal{D}_G) \\ &\quad +n^{1/2}\int_{\{R\backslash\mathcal{D}_F\}}[G_n(\mathcal{D}_G{}^-)-\hat{G}_n(\mathcal{D}_G{}^-)]J^{cd}_{v_1}(F(x),G(\mathcal{D}_G{}^-),G(\mathcal{D}_G))dH_\theta(x,\mathcal{D}_G) \\ &= A^*_{3n1}+A^*_{3n2}. \end{aligned}$$

Similarly, $A^*_{3n1}$ can be written as $n^{-1/2}\sum_{j=1}^n A_{3jn}$, where $A_{3jn}=\int_{\{R\backslash\mathcal{D}_F\}}(\psi_{Y_j}(\mathcal{D}_G)-$

$G_n(\mathcal{D}_G{}^-))J_{v1}^{cd}(F(x), G(\mathcal{D}_G{}^-), G(\mathcal{D}_G))dH_\theta(x, \mathcal{D}_G)$ are i.i.d. with mean zero. and

$$\psi_{Y_j}(y) = \begin{cases} 1 & \text{if } Y_j < y, \\ 0 & \text{if } Y_j \geq y. \end{cases}$$

Noting that the absolute value of the random part $|\psi_{Y_j}(\mathcal{D}_G) - G_n(\mathcal{D}_G{}^-)|$ in $A_{3jn}$ is bounded above by $1$, we have

$$|A_{3jn}| \leq \int_{(0,1)\cap\mathcal{S}_F^c} |J_{v1}^{cd}(u, G(\mathcal{D}_G{}^-), G(\mathcal{D}_G))| \int_{\mathcal{S}_G} c_\theta(u,v) dv\, du.$$

By the assumption (D2), the integral above is finite. Therefore, $A_{3jn}$ has an absolute moment of order $2+\delta_1$ for some $\delta_1 > 0$, which leads to the asymptotic normality of $A_{3n1}^*$.

Using argument similar to that used for $A_{2n2}^*$ as in (2), one can show that $A_{3n2}^* = o_p(1)$. In summary, we have the asymptotic normality of $A_{3n}$.

(4) Result of $A_{4n}$ can be obtained in a similar way by noting that $A_{4n}$ can be written as

$$\begin{aligned} A_{4n} &= n^{1/2} \int_{\{R\backslash\mathcal{D}_F\}} [\hat{G}_n(\mathcal{D}_G) - G(\mathcal{D}_G)] J_{v2}^{cd}(F(x), G(\mathcal{D}_G{}^-), G(\mathcal{D}_G)) dH_\theta(x, \mathcal{D}_G) \\ &\quad + n^{1/2} \int_{\{R\backslash\mathcal{D}_F\}} [G_n(\mathcal{D}_G) - \hat{G}_n(\mathcal{D}_G)] J_{v2}^{cd}(F(x), G(\mathcal{D}_G{}^-), G(\mathcal{D}_G)) dH_\theta(x, \mathcal{D}_G) \\ &= A_{4n1}^* + A_{4n2}^*, \end{aligned}$$

and $A_{4n1}^*$ can be written as $n^{-1/2}\sum_{j=1}^n A_{4jn}$, where $A_{4jn} = \int_{\{R\backslash\mathcal{D}_F\}} (\phi_{Y_j}(\mathcal{D}_G) - G_n(\mathcal{D}_G))J_{v2}^{cd}(F(x), G(\mathcal{D}_G{}^-), G(\mathcal{D}_G))dH_\theta(x, \mathcal{D}_G)$ are i.i.d. with mean zero.

Using arguments similar to those in (3), we can get that $A_{4n}$ is asymptotically normally distributed with mean $0$.

(5) Finally, we show that the sum of $A_{in}$, $i = 1, ..., 4$, converges to a normal random variable with mean $0$. Note that the sum can be written as $\sum_{j=1}^n n^{-1/2}(A_{1jn} + A_{2jn} + A_{3jn} + A_{4jn}) + A_{2n2}^* + A_{3n2}^* + A_{4n2}^* \equiv \sum_{j=1}^n n^{-1/2} A_{2jn}^{**} + A_{2n2}^* + A_{3n2}^* + A_{4n2}^*$, where $A_{2jn}^{**} = A_{1jn} + A_{2jn} + A_{3jn} + A_{4jn}$ depends on $(X_j, Y_j)$ only, hence are i.i.d. with mean $0$ as shown before

(Similarly, we can define $A^{**}_{3jn}$ for $R^{dc}_n$), and $A^*_{2n2}$, $A^*_{3n2}$ and $A^*_{4n2}$ are negligible. Using the CLT, we have that the sum of $A_{in}$, $i = 1, ..., 4$, is asymptotically normally distributed with mean $0$.

We will finish the proof of $R^{cd}_n$ by a few corollaries.

**Corollary 5.** $B_{1n} \to^p 0$ *as* $n \to \infty$.

**Proof:** Note that

$$\begin{aligned} B_{1n} &= \chi(\Omega_{1n}) \int_{\{R \backslash \mathcal{D}_F\}} U_n(F(x)) J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) \\ &\quad - H_\theta(x, \mathcal{D}_G)] \\ &\quad + \chi(\Omega^c_{1n}) \int_{\{R \backslash \mathcal{D}_F\}} U_n(F(x)) J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) \\ &\quad - H_\theta(x, \mathcal{D}_G)] \end{aligned}$$

where $\Omega_{1n}$ is defined in (3.15), which tells us the second term on the righthand side of the equality above goes to 0 in probability as $n \to \infty$. Also,

$$\begin{aligned} &\chi(\Omega_{1n}) \int_{\{R \backslash \mathcal{D}_F\}} U_n(F(x)) J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) - H_\theta(x, \mathcal{D}_G)] \\ \leq\ &\chi(\Omega_{1n}) \int_{\{R \backslash \mathcal{D}_F\}} M'\, J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) - H_\theta(x, \mathcal{D}_G)] \\ =\ &\chi(\Omega_{1n}) \int_{\Delta_F} M'\, J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) - H_\theta(x, \mathcal{D}_G)] + \\ &\chi(\Omega_{1n}) \int_{\{(-\infty, X_{1n}) \cup (X_{nn}, \infty)\}} M'\, J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) \\ &- H_\theta(x, \mathcal{D}_G)], \end{aligned}$$

the second term on the righthand side of the inequality goes to 0 in probability as $n \to \infty$ by the Glivenko-Cantelli theorem and the assumption (D2). Since $J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G))$ is bounded above almost surely when $x \in \Delta_F$ and $H_n(x, y)$ converges to $H_\theta(x, y)$ in distribution as $n \to \infty$, using Theorem 1.9 (i), [43], once again, we have the first term on the righthand side of the inequality above converges to 0. In summary, we have shown that $B_{1n} \to^p 0$ as $n \to \infty$.

**Corollary 6.** *$B_{2n} \to^p 0$ and $B_{3n} \to^p 0$ as $n \to \infty$.*

**Proof:** It suffices to show the result for $B_{2n}$. By the CLT, $V_n(G(\mathcal{D}_G^-))$ converges to a normal random variable as $n \to \infty$, so we can define a set $\Omega_{2n}$ such that

$$P(\Omega_{2n}) = P(\{\omega : \sup_\omega |V_n(G(\mathcal{D}_G^-))| \leq M'\}) > 1 - \epsilon.$$

Therefore,

$$\begin{aligned} B_{2n} &= \chi(\Omega_{2n}) \int_{\{R \setminus \mathcal{D}_F\}} V_n(G(\mathcal{D}_G^-)) J_{v_1}^{cd}(F(x), \Theta_n(\mathcal{D}_G^-), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) \\ &\quad - H_\theta(x, \mathcal{D}_G)] \\ &\quad + \chi(\Omega_{2n}^c) \int_{\{R \setminus \mathcal{D}_F\}} V_n(G(\mathcal{D}_G^-)) J_{v_1}^{cd}(F(x), \Theta_n(\mathcal{D}_G^-), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) \\ &\quad - H_\theta(x, \mathcal{D}_G)], \end{aligned}$$

and the second term on the righthand side of the equality above goes to 0 in probability as $n \to \infty$. Also,

$$\begin{aligned} & \chi(\Omega_{2n}) \int_{\{R \setminus \mathcal{D}_F\}} V_n(G(\mathcal{D}_G^-)) J_{v_1}^{cd}(F(x), \Theta_n(\mathcal{D}_G^-), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) \\ & - H_\theta(x, \mathcal{D}_G)] \\ \leq\ & \chi(\Omega_{2n}) \int_{\{R \setminus \mathcal{D}_F\}} M' \, J_{v_1}^{cd}(F(x), \Theta_n(\mathcal{D}_G^-), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) - H_\theta(x, \mathcal{D}_G)] \\ =\ & \chi(\Omega_{2n}) \int_{\Delta_F} M' \, J_{v_1}^{cd}(F(x), \Theta_n(\mathcal{D}_G^-), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) - H_\theta(x, \mathcal{D}_G)] + \\ & \chi(\Omega_{2n}) \int_{\{(-\infty, X_{1n}) \cup (X_{nn}, \infty)\}} M' \, J_{v_1}^{cd}(F(x), \Theta_n(\mathcal{D}_G^-), G_n(\mathcal{D}_G)) d[H_n(x, \mathcal{D}_G) \\ & - H_\theta(x, \mathcal{D}_G)], \end{aligned}$$

the second term on the righthand side of the inequality goes to 0 in probability as $n \to \infty$ by the Glivenko-Cantelli theorem and the assumption (D2).

From the assumption (D2), combining with the definition of $\Theta_n(\cdot)$ and the fact that $J_{u_2}^{cd}(u_2, v_1, v_2)$ is continuous with respect to $u_2$, $v_1$, and $v_2$ on $(0, 1)^3$, we know that

$J^{cd}_{u2}(F(x), \Theta_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G))$ is bounded above almost surely for $x \in \Delta_F$. Noting that $H_n(x, y)$ converges to $H_\theta(x, y)$ in distribution as $n \to \infty$, using Theorem 1.9 (i), [43], once again, we have the first term on the righthand side of the inequality above converges to 0. In summary, we have shown that $B_{2n} \to^p 0$ as $n \to \infty$.

**Corollary 7.** $B_{4n} \to^p 0$ *as* $n \to \infty$.

**Proof:** Note that

$$\begin{aligned}
B_{4n} &= \chi(\Omega_{1n}) \int_{\{R \backslash \mathcal{D}_F\}} U_n(F(x)) [J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) \\
&\quad - J^{cd}_{u2}(F(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G))\Big] \, dH_\theta(x, \mathcal{D}_G) \\
&\quad + \chi(\Omega^c_{1n}) \int_{\{R \backslash \mathcal{D}_F\}} U_n(F(x)) [J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) \\
&\quad - J^{cd}_{u2}(F(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G))\Big] \, dH_\theta(x, \mathcal{D}_G)
\end{aligned}$$

where $\Omega_{1n}$ is defined in (3.15). Therefore, the second term on the righthand side of the equality above goes to 0 in probability as $n \to \infty$. Also,

$$\begin{aligned}
& \chi(\Omega_{1n}) \int_{\{R \backslash \mathcal{D}_F\}} U_n(F(x)) [J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) \\
& - J^{cd}_{u2}(F(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G))\Big] \, dH_\theta(x, \mathcal{D}_G) \\
\leq \; & \chi(\Omega_{1n}) \int_{\{R \backslash \mathcal{D}_F\}} M' \, |J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) \\
& - J^{cd}_{u2}(F(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G))\Big| \, dH_\theta(x, \mathcal{D}_G) \\
= \; & \chi(\Omega_{1n}) \int_{\Delta_F} M' \, |J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) \\
& - J^{cd}_{u2}(F(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G))\Big| \, dH_\theta(x, \mathcal{D}_G) + \\
& \chi(\Omega_{1n}) \int_{\{(-\infty, X_{1n}) \cup (X_{nn}, \infty)\}} M' \, |J^{cd}_{u2}(\Phi_n(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G)) \\
& - J^{cd}_{u2}(F(x), G_n(\mathcal{D}^-_G), G_n(\mathcal{D}_G))\Big| \, dH_\theta(x, \mathcal{D}_G),
\end{aligned}$$

the second term on the righthand side of the inequality goes to 0 in probability as $n \to \infty$ by the Glivenko-Cantelli theorem and the assumption (D2). Since $|J_{u2}^{cd}(\Phi_n(x), G_n(\mathcal{D}_G^-), G_n(\mathcal{D}_G)) - J_{u2}^{cd}(F(x), G_n(\mathcal{D}_G^-), G_n(\mathcal{D}_G))|$ is bounded above almost surely when $x \in \Delta_F$ and $H_n(x, y)$ converges to $H_\theta(x, y)$ in distribution as $n \to \infty$, using Theorem 1.9 (i), [43], once again, we have the first term on the righthand side of the inequality above converges to 0. In summary, we have shown that $B_{4n} \to^p 0$ as $n \to \infty$.

**Corollary 8.** $B_{5n} \to^p 0$ *and* $B_{6n} \to^p 0$ *as* $n \to \infty$.

**Proof:** Using similar arguments to those used in the proof of Corollary 7 and Corollary 8, one can see this is true, so proof omitted.

In summary, we have the asymptotic normality of $R_n^{cd}$.

### 3.4.3 Asymptotic normality of $R_n^{dd}$

Note that

$$n^{1/2}(R_n^{dd} - \mu^{dd}) = \sum_{k=1}^{5} A_{kn} + \sum_{k=1}^{4} B_{kn},$$

where

$$\begin{aligned}
\mu^{dd} =& \quad E[N^{dd}(X, Y)], \\
A_{1n} =& \quad n^{1/2} J^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G))[dH_n^{dd} - dH^{dd}], \\
A_{2n} =& \quad V_n(G(\mathcal{D}_G)) J_{v2}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G)) dH^{dd}, \\
A_{3n} =& \quad U_n(F(\mathcal{D}_F)) J_{u2}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G)) dH^{dd}, \\
A_{4n} =& \quad V_n(G(\mathcal{D}_G^-)) J_{v1}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G)) dH^{dd}, \\
A_{5n} =& \quad U_n(F(\mathcal{D}_F^-)) J_{u1}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G)) dH^{dd},
\end{aligned}$$

$$
\begin{aligned}
B_1 = \;& V_n(G(\mathcal{D}_G))[J_{v2}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), \Psi_n)dH_n^{dd} \\
& -J_{v2}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G))dH^{dd}], \\
B_2 = \;& U_n(F(\mathcal{D}_F))[J_{u2}^{dd}(F(\mathcal{D}_F^-), \Phi_n, G(\mathcal{D}_G^-), G(\mathcal{D}_G))dH_n^{dd} \\
& -J_{u2}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G))dH^{dd}], \\
B_3 = \;& V_n(G(\mathcal{D}_G^-))[J_{v1}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), \Theta_n, G(\mathcal{D}_G))dH_n^{dd} \\
& -J_{u1}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G))dH^{dd}], \\
B_4 = \;& U_n(F(\mathcal{D}_F^-))[J_{u1}^{dd}(\Gamma_n, F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G))dH_n^{dd} \\
& -J_{u1}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G))dH^{dd}],
\end{aligned}
$$

with $n_{dd}^* = \{$the number of observations such that $X_j = \mathcal{D}_F$ and $Y_j = \mathcal{D}_G\}$ for $k = 1, 2, ..., n$, and

$$
\begin{aligned}
dH_n^{dd} &= \frac{n_{dd}^*}{n} \\
dH^{dd} &= P(X = \mathcal{D}_F \text{ and } Y = \mathcal{D}_G).
\end{aligned}
$$

We need to show that the $A$-terms are asymptotically normal, and the $B$-terms converge to $0$ in probability as $n \to 0$. By the SLLN, it is obvious that $dH_n^{dd} \to dH^{dd}$ almost surely as $n \to \infty$, and by the CLT, in distribution

$$
\begin{aligned}
& U_n(F(\mathcal{D}_F^-)) \to N(0, [F(\mathcal{D}_F^-)(1 - F(\mathcal{D}_F^-))]) \\
& U_n(F(\mathcal{D}_F)) \to N(0, [F(\mathcal{D}_F)(1 - F(\mathcal{D}_F))]) \\
& V_n(G(\mathcal{D}_G^-)) \to N(0, [G(\mathcal{D}_G^-)(1 - G(\mathcal{D}_G^-))]) \\
& V_n(G(\mathcal{D}_G)) \to N(0, [G(\mathcal{D}_G)(1 - G(\mathcal{D}_G))]) \\
& n^{1/2}[dH_n^{dd} - dH^{dd}] \to N(0, [dH^{dd}(1 - dH^{dd})])
\end{aligned}
$$

as $n \to \infty$. Note that

$$J^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G)),$$
$$J_{v2}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G)),$$
$$J_{u2}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G)),$$
$$J_{v1}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G)), \text{ and}$$
$$J_{u1}^{dd}(F(\mathcal{D}_F^-), F(\mathcal{D}_F), G(\mathcal{D}_G^-), G(\mathcal{D}_G)),$$

in $A_{1n}$, $A_{2n}$, $A_{3n}$, $A_{4n}$ and $A_{5n}$, respectively, are fixed numbers. Therefore, the $A$-terms are asymptotically normally distributed with mean 0. To show the sum of the $A$-terms is still normal, one only need to notice that

$$\begin{aligned}
U_n(F(\mathcal{D}_F^-)) &= n^{1/2}(\hat{F}_n(\mathcal{D}_F^-) - F(\mathcal{D}_F^-)) + n^{1/2}(F_n(\mathcal{D}_F^-) - \hat{F}_n(\mathcal{D}_F^-)) \\
&= n^{-1/2}\sum_{j=1}^{n}(\psi_{X_j}(\mathcal{D}_F^-) - F(\mathcal{D}_F^-)) + op(1) \\
&= n^{-1/2}\sum_{j=1}^{n} H_{1kn}^{*} + op(1)
\end{aligned}$$

$$\begin{aligned}
U_n(F(\mathcal{D}_F)) &= n^{1/2}(\hat{F}_n(\mathcal{D}_F) - F(\mathcal{D}_F)) + n^{1/2}(F_n(\mathcal{D}_F) - \hat{F}_n(\mathcal{D}_F)) \\
&= n^{-1/2}\sum_{j=1}^{n}(\phi_{X_j}(\mathcal{D}_F) - F(\mathcal{D}_F)) + op(1) \\
&= n^{-1/2}\sum_{j=1}^{n} H_{2kn}^{*} + op(1) \\
V_n(G(\mathcal{D}_G^-)) &= n^{1/2}(\hat{G}_n(\mathcal{D}_G^-) - G(\mathcal{D}_G^-)) + n^{1/2}(G_n(\mathcal{D}_G^-) - \hat{G}_n(\mathcal{D}_G^-)) \\
&= n^{-1/2}\sum_{j=1}^{n}(\psi_{Y_j}(\mathcal{D}_G^-) - G(\mathcal{D}_G^-)) + op(1) \\
&= n^{-1/2}\sum_{j=1}^{n} H_{3kn}^{*} + op(1) \\
V_n(G(\mathcal{D}_G)) &= n^{1/2}(\hat{G}_n(\mathcal{D}_G) - G(\mathcal{D}_G)) + n^{1/2}(G_n(\mathcal{D}_G) - \hat{G}_n(\mathcal{D}_G)) \\
&= n^{-1/2}\sum_{j=1}^{n}(\phi_{Y_j}(\mathcal{D}_G) - G(\mathcal{D}_G)) + op(1) \\
&= n^{-1/2}\sum_{j=1}^{n} H_{4kn}^{*} + op(1)
\end{aligned}$$

and

$$n^{1/2}[dH_n^{dd} - dH^{dd}] = n^{-1/2}\sum_{j=1}^{n}\left(I_{\{j\in A_{dd}\}} - dH^{dd}\right) = n^{-1/2}\sum_{j=1}^{n} H_{5kn}^{*},$$

where $H_{ikn}^{*}$ are i.i.d. and depend on $(X_j, Y_j)$ only, for $k = 1, ..., 5$. The $A_{1n} + A_{2n} + A_{3n} + A_{4n} + A_{5n}$ can be written as $\sum_{j=1}^{n} n^{-1/2} A_{4jn}^{**}$, where $A_{4jn}^{**}$ are i.i.d. sum of products of $H_{ikn}^{*}$ and some certain fixed number, and four negligible terms. Using the CLT one more time, we have that the sum of $A$-terms is asymptotically normally distributed with mean $0$.

Under the assumption (D4), $B_i$ converges to $0$ as $n \to \infty$, for $i = 1, ..., 4$, in probability. We have shown that the $R_n^{dd}$ is asymptotically normally distributed with mean $0$.

### 3.4.4 Asymptotic normality of $R_n$

To show that $n^{1/2}R_n = n^{1/2}(R_n^{cc} + R_n^{cd} + R_n^{dc} + R_n^{dd})$ is asymptotic normal, we need the following notations:

$$\begin{aligned}
n^{1/2}(R_n^{cc} - \mu^{cc}) &\Rightarrow R^{cc} \sim N(0, \text{var}(M^{cc}(X,Y))),\\
n^{1/2}(R_n^{cd} - \mu^{cd}) &\Rightarrow R^{cd} \sim N(0, \text{var}(M^{cd}(X,Y))),
\end{aligned}$$

$$\begin{aligned}
n^{1/2}(R_n^{dc} - \mu^{dc}) &\Rightarrow R^{dc} \sim N(0, \text{var}(M^{dc}(X,Y))),\\
n^{1/2}(R_n^{dd} - \mu^{dd}) &\Rightarrow R^{dd} \sim N(0, \text{var}(M^{dd}(X,Y))).
\end{aligned}$$

Noting that

$$\begin{aligned}
n^{1/2}(R_n^{cc} - \mu^{cc}) &= \sum_{j=1}^{n} n^{-1/2} A_{1jn}^{**} + \text{ a few negligible terms },\\
n^{1/2}(R_n^{cd} - \mu^{cd}) &= \sum_{j=1}^{n} n^{-1/2} A_{2jn}^{**} + \text{ a few negligible terms },\\
n^{1/2}(R_n^{dc} - \mu^{dc}) &= \sum_{j=1}^{n} n^{-1/2} A_{3jn}^{**} + \text{ a few negligible terms },\\
n^{1/2}(R_n^{dd} - \mu^{dd}) &= \sum_{j=1}^{n} n^{-1/2} A_{4jn}^{**} + \text{ a few negligible terms },
\end{aligned}$$

by the CLT, we have

$$\begin{aligned} n^{1/2}(R_n - \mu) &= \sum_{j=1}^{n} n^{-1/2}(A_{1jn}^{**} + A_{2jn}^{**} + A_{3jn}^{**} + A_{4jn}^{**}) + \text{ a few negligible terms }, \\ &\Rightarrow N(0, \sigma^2) \end{aligned}$$

where $\mu = (\mu^{cc} + \mu^{cd} + \mu^{dc} + \mu^{dd})$, and

$$\sigma^2 = \text{var}[M^{cc}(X,Y) + M^{cd}(X,Y) + M^{dc}(X,Y) + M^{dd}(X,Y)].$$

## 3.5 Proof of the Main Result

We finish this Chapter by providing the proof of Theorem 3.1.1.

**Proof of Theorem 3.1.1** As shown in Section 3.3, $B_n \to \beta$ almost surely, and in Section 3.4 $A_n$ is asymptotically normal with mean $0$ and variance $\sigma^2$, as $n \to \infty$. Using the Slutsky's Theorem, $n^{1/2}(\hat{\theta}_n - \theta) = n^{1/2} A_n / B_n$ is asymptotically normal with mean $0$ and variance $\varrho = \sigma^2/\beta^2$.

In summary, we have shown the consistency and asymptotic normality of the semi-parametric estimator $\hat{\theta}_n$.

# CHAPTER 4

# A VARIANCE ESTIMATOR OF BIVARIATE DISTRIBUTIONS

In Chapter 3, we provided an explicit formula for $\varrho = \dfrac{\sigma^2}{\beta^2}$. Suppose estimators $\hat{\sigma}^2$ and $\hat{\beta}^2$ could be found for $\sigma^2$ and $\beta^2$ respectively. A rough-and-ready estimator of the asymptotic variance of $\varrho$ would then be given by $\hat{\varrho} = \dfrac{\hat{\sigma}^2}{\hat{\beta}^2}$. If the variables given in (3.9) and (3.10) could be observed, one could simply estimate $\sigma^2$ and $\beta$ by the respective sample variance and sample mean. As this is not possible, the corresponding pseudo-observations can be used instead, which are defined in term of $C_n$, the re-scaled empirical copula function of the bivariate sample, namely

$$C_n(u,v) = \frac{1}{n}\sum_{j=1}^{n} I_{\left\{F_n(X_j)\leq u, G_n(Y_j)\leq v\right\}}.$$

Let $l^{cc}_{\theta,\theta}$, $l^{cd}_{\theta,\theta}$, $l^{dc}_{\theta,\theta}$, and $l^{dd}_{\theta,\theta}$ be the decomposed components of $l_{\theta,\theta}$ with respect to $c^*_\theta$ in (2.7) under different situations. Similarly, one can define $l^{cc}_\theta$, $l^{cd}_\theta$, $l^{dc}_\theta$, and $l^{dd}_\theta$ for $l_\theta$.

From (3.9), we have the following:

$$\hat{\beta} = \frac{1}{n}\sum_{j=1}^{n}[\hat{N}^{cc}_j + \hat{N}^{cd}_j + \hat{N}^{dc}_j + \hat{N}^{dd}_j], \tag{4.1}$$

| j | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $(X_j, Y_j)$ | (2.1,1.0) | (2.5, 0.5) | (3.2,1.0) | (3.2,1.5) | (3.2,2.0) | (3.5,2.5) | (3.7,2.5) |
| $R_{X(j)}$ | 1 | 2 | 5 | 5 | 5 | 6 | 7 |
| $R_{Y(j)}$ | 3 | 1 | 3 | 4 | 5 | 7 | 7 |

Table 4.1. Relationship among $(X_j, Y_j)$, $R_{X(j)}$, and $R_{Y(j)}$.

where

$$\begin{aligned}
\hat{N}_j^{cc} &= I_{\{j \in A_{cc}\}} l_{\theta,\theta}^{cc}(\hat{\theta}_n, F_n(X_j), G_n(Y_j)) \\
\hat{N}_j^{cd} &= I_{\{j \in A_{cd}\}} l_{\theta,\theta}^{cd}(\hat{\theta}_n, F_n(X_j), G_n(Y_j^-), G_n(Y_j)) \\
\hat{N}_j^{dc} &= I_{\{j \in A_{dc}\}} l_{\theta,\theta}^{dc}(\hat{\theta}_n, F_n(X_j^-), F_n(X_j), G_n(Y_j)) \\
\hat{N}_j^{dd} &= I_{\{j \in A_{dd}\}} l_{\theta,\theta}^{dd}(\hat{\theta}_n, F_n(X_j^-), F_n(X_j), G_n(Y_j^-), G_n(Y_j)).
\end{aligned}$$

To get $\hat{\sigma}^2$ for $\sigma^2$ in (3.11), we need more notation. Note that rearranging $\{X_j, Y_j\}$, $j = 1, ..., n$, shall not change the value of $\hat{\sigma}^2$. Therefore, we assume the sample is in an order such that $X_j$'s are in non-decreasing order. Let

$$R_{X(j)} = \sum_i I_{\{X_i \leq X_j\}}$$

be the rank of $X_j$ in the sequence of $X$'s (Similarly, we can define $R_{Y(j)}$ for $Y$'s), and let

$$R_{X(j)}^- = \sum_i I_{\{X_i < X_j\}}$$

be the next lower rank to $R_{X(j)}$ within the $X$'s sequence. Similarly, we can define $R_{Y(j)}^-$ for $R_{Y(j)}$ within the $Y$'s sequence.

Under the current setting, for any $i < j$, the following always holds:

$$R_{X(i)} \leq R_{X(j)}.$$

Table 4.1 gives a simple example showing the relationship among $(X_j, Y_j)$, $R_{X(j)}$, and $R_{Y(j)}$. As shown in the table, $R_{X(3)} = R_{X(4)} = R_{X(5)} = 5$, and $R^-_{X(5)} = R^-_{X(4)} = R^-_{X(3)} = R_{X(2)} = 2$ by definition.

Now we can work on the details. Note that in (3.10), $\sigma^2$ is the variance of the sum of $M^S(X,Y)$, where $S = \{cc\}, \{cd\}, \{dc\}, or\{dd\}$. For a given sample $(X_j, Y_j)$, $\sigma^2$ can be estimated by the sample variance of the following items:

$$
\begin{aligned}
\hat{M}_j^{cc} &= l_\theta^{cc}(\hat{\theta}_n, F_n(X_j), G_n(Y_j)) I_{\{j\in A_{cc}\}} + \frac{1}{n}\left[\sum_{k=j}^{n} l_{\theta,u_2}^{cc}\left(\hat{\theta}_n, \frac{R_{X(k)}}{n+1}, \frac{R_{Y(k)}}{n+1}\right)\right. \\
&+ \sum_{R_{Y(k)}\geq R_{Y(j)}} l_{\theta,v_2}^{cc}\left(\hat{\theta}_n, \frac{R_{X(k)}}{n+1}, \frac{R_{Y(k)}}{n+1}\right) I_{\{k\in A_{cc}\}} \\
\hat{M}_j^{cd} &= \sum_{l=1}^{d_G} l_\theta^{cd}(\hat{\theta}_n, F_n(X_j), G_n(Y_j^-), G_n(Y_j)) I_{\{j\in A_{cd}, Y_j=\mathcal{D}_{Gl}\}} \\
&+ \frac{1}{n}\left[\sum_{l=1}^{d_G}\sum_{k=j}^{n} l_{\theta,u_2}^{cd}\left(\hat{\theta}_n, \frac{R_{X(k)}}{n+1}, \frac{R^-_{Y(k)}}{n+1}, \frac{R_{Y(k)}}{n+1}\right) I_{\{k\in A_{cd}, Y_k=\mathcal{D}_{Gl}\}}\right. \\
&+ \sum_{l=1}^{d_G}\sum_{R_{Y(k)}>R_{Y(j)}} l_{\theta,v_1}^{cd}\left(\hat{\theta}_n, \frac{R_{X(k)}}{n+1}, \frac{R^-_{Y(k)}}{n+1}, \frac{R_{Y(k)}}{n+1}\right) I_{\{k\in A_{cd}, Y_k=\mathcal{D}_{Gl}\}} \\
&+ \sum_{l=1}^{d_G}\sum_{R_{Y(k)}\geq R_{Y(j)}} l_{\theta,v_2}^{cd}\left(\hat{\theta}_n, \frac{R_{X(k)}}{n+1}, \frac{R^-_{Y(k)}}{n+1}, \frac{R_{Y(k)}}{n+1}\right) I_{\{k\in A_{cd}, Y_k=\mathcal{D}_{Gl}\}} \\
\hat{M}_j^{dc} &= \sum_{h=1}^{d_F} l_\theta^{dc}(\hat{\theta}_n, F_n(X_j^-), F_n(X_j), G_n(Y_j)) I_{\{j\in A_{dc}, X_j=\mathcal{D}_{Fh}\}} \\
&+ \frac{1}{n}\left[\sum_{h=1}^{d_F}\sum_{R_{Y(k)}\geq R_{Y(j)}} l_{\theta,v_2}^{dc}\left(\hat{\theta}_n, \frac{R^-_{X(k)}}{n+1}, \frac{R_{X(k)}}{n+1}, \frac{R_{Y(k)}}{n+1}\right)\right. . \\
& I_{\{k\in A_{dc}, X_k=\mathcal{D}_{Fh}\}}
\end{aligned}
$$

$$
\begin{aligned}
& + \sum_{h=1}^{d_F} \sum_{R_{X(k)} > R_{X(j)}} l_{\theta,u_1}^{dc} \left( \hat{\theta}_n, \frac{R^-_{X(k)}}{n+1}, \frac{R_{X(k)}}{n+1}, \frac{R_{Y(k)}}{n+1} \right) \cdot \\
& I_{\{k \in A_{dc}, X_k = \mathcal{D}_{Fh}\}} \\
& + \sum_{h=1}^{d_F} \sum_{k=j}^{n} l_{\theta,u_2}^{dc} \left( \hat{\theta}_n, \frac{R^-_{X(k)}}{n+1}, \frac{R_{X(k)}}{n+1}, \frac{R_{Y(k)}}{n+1} \right) I_{\{k \in A_{dc}, X_k = \mathcal{D}_{Fh}\}} \Bigg] \\
\hat{M}_j^{dd} = & \sum_{h=1}^{d_F} \sum_{l=1}^{d_G} l_\theta^{dd} (\hat{\theta}_n, F_n(X_j^-), F_n(X_j), G_n(Y_j^-), G_n(Y_j)) I_{\{X_j = \mathcal{D}_{Fh}, Y_j = \mathcal{D}_{Gl}\}} \\
& + \Bigg[ \sum_{h=1}^{d_F} \sum_{l=1}^{d_G} \sum_{R_{Y(k)} > R_{Y(j)}}^{n} l_{\theta,v_1}^{dd} \left( \hat{\theta}_n, \frac{R^-_{X(k)}}{n+1}, \frac{R_{X(k)}}{n+1}, \frac{R^-_{Y(k)}}{n+1}, \frac{R_{Y(k)}}{n+1} \right) \cdot \\
& I_{\{X_k = \mathcal{D}_{Fh}, Y_k = \mathcal{D}_{Gl}\}} \\
& + \sum_{h=1}^{d_F} \sum_{l=1}^{d_G} \sum_{R_{Y(k)} \geq R_{Y(j)}}^{n} l_{\theta,v_2}^{dd} \left( \hat{\theta}_n, \frac{R^-_{X(k)}}{n+1}, \frac{R_{X(k)}}{n+1}, \frac{R^-_{Y(k)}}{n+1}, \frac{R_{Y(k)}}{n+1} \right) \cdot \\
& I_{\{X_k = \mathcal{D}_{Fh}, Y_k = \mathcal{D}_{Gl}\}} \\
& + \sum_{h=1}^{d_F} \sum_{l=1}^{d_G} \sum_{R_{X(k)} > R_{X(j)}}^{n} l_{\theta,u_1}^{dd} \left( \hat{\theta}_n, \frac{R^-_{X(k)}}{n+1}, \frac{R_{X(k)}}{n+1}, \frac{R^-_{Y(k)}}{n+1}, \frac{R_{Y(k)}}{n+1} \right) \cdot \\
& I_{\{X_k = \mathcal{D}_{Fh}, Y_k = \mathcal{D}_{Gl}\}} \\
& + \sum_{h=1}^{d_F} \sum_{l=1}^{d_G} \sum_{k=j}^{n} l_{\theta,u_2} \left( \hat{\theta}_n, \frac{R^-_{X(k)}}{n+1}, \frac{R_{X(k)}}{n+1}, \frac{R^-_{Y(k)}}{n+1}, \frac{R_{Y(k)}}{n+1} \right) \cdot \\
& I_{\{X_k = \mathcal{D}_{Fh}, Y_k = \mathcal{D}_{Gl}\}} \cdot \frac{n^*_{hl}}{n},
\end{aligned}
$$

where $n^*_{hl} = \{$the number of observations $(X_j, Y_j) : X_j = \mathcal{D}_{Fh}, Y_j = \mathcal{D}_{Gl}\}$.

Next, to establish the consistency of $\hat{\varrho}_n$, it suffices to show that $\hat{\sigma}_n^2$ and $\hat{\beta}_n^2$ are themselves consistent. To prove that $\hat{\beta}_n - \beta$ converges almost surely, for example, express the difference as

$$
\frac{1}{n} \sum_{j=1}^{n} J_{\hat{\theta}_n} (F_n(X_j^-), F_n(X_j), G_n(Y_j^-), G_n(Y_j)) - E[J_\theta(F(X^-), F(X), G(Y^-), G(Y))]
$$

in terms of $J_\theta = l_{\theta,\theta}$. By the triangle inequality, this quantity is no greater than

$$\frac{1}{n}\sum_{j=1}^{n}\Big|J_{\hat{\theta}_n}(F_n(X_j^-), F_n(X_j), G_n(Y_j^-), G_n(Y_j)) \\ - J_\theta(F_n(X_j^-), F_n(X_j), G_n(Y_j^-), G_n(Y_j))\Big|$$

$$+\left|\frac{1}{n}\sum_{j=1}^{n} J_\theta(F_n(X_j^-), F_n(X_j), G_n(Y_j^-), G_n(Y_j)) - E[J_\theta(F(X^-), F(X), G(Y^-), G(Y))]\right|$$

Assuming that $J_\theta$ is bounded by an integrable function in a neighborhood of the true value of $\theta$, by the convergence of maximum likelihood estimators, the first summand then converges to zero by the Dominated Convergence Theorem. Since the second term vanishes asymptotically by a similar argument as used in Chapter 3, it follows that $\hat{\beta}_n \to \beta$ almost surely. The argument for $\hat{\sigma}_n$ is similar, though somewhat more involved. It will not be presented here.

In summary, we provided a variance estimator of $\varrho$, $\hat{\varrho}$, and illustrated its consistency.

# CHAPTER 5

# EXTENSIONS TO HIGHER DIMENSIONS

The previous developments extend more or less automatically to situations where it is desired to estimate a multidimensional dependence semi-parametrically.

Let a boldfaced letter such as $\mathbf{x}$ denotes a $D$-tuple ($D \geq 2$) vector of real numbers, that is, $\mathbf{x} \equiv (x_1, \ldots, x_D)^T$ where $x_k \in R$ for $k = 1, \ldots, D$, and $F_1, ..., F_D$ denote $D$ univariate mixed marginals on the real line given by

$$F_k(x_k) = \sum_{h=1}^{d_k} p_{kh} I_{\{\mathcal{D}_{kh} \leq x_k\}} + (1 - \sum_{h=1}^{d_k} p_{kh}) \int_{w \leq x_k} f_k(w) dw,$$

where $\mathcal{D}_{kh}$ is the $h$-th jump point of $F_k$ with $P(X_k = \mathcal{D}_{kh}) = p_{kh}$, for $h = 1, ..., d_k$, and $f_k(\cdot)$ is a continuous density function with support on the real line.

When restricted to a particular copula family, say $\{C_\theta : \theta = (\theta_1, ..., \theta_q)^T \in \mathcal{A} \subseteq R^q\}$, a multivariate joint distribution function for $(x_1, \ldots, x_D)^T$ can be defined as follows:

$$F^D(x_1, \ldots, x_D) = C_\theta(F_1(x_1), \ldots, F_D(x_D)). \tag{5.1}$$

**Result 1:** $F^D(x_1, \ldots, x_D)$ defined in (5.1) is a valid joint distribution function on $R^D$.

**Proof:** For $F^D(x_1, \ldots, x_D)$ to be a valid distribution function on $R^D$, the following four conditions should be satisfied:

(1) $0 \leq F^D(x_1, \ldots, x_D) \leq 1$ for all $(x_1, \ldots, x_D)$,

(2) $F^D(x_1, \ldots, x_D) \to 0$ as max$(x_1, \ldots, x_D) \to -\infty$,

(3) $F^D(x_1, \ldots, x_D) \to 1$ as min$(x_1, \ldots, x_D) \to +\infty$, and

(4) for every pair of $(x_1, \ldots, x_D)$ and $(y_1, \ldots, y_D) \in R^D$ with $x_k \leq y_k$ for $k = 1, ..., D$, which defines a $D$-box $V = [x_1, y_1] \times [x_2, y_2] \times ... \times [x_d, y_D]$. Then

$$\sum \text{sgn}(\mathbf{c}) F(\mathbf{c}) \geq 0$$

where the sum is taken over all vertices $\mathbf{c}$ of $V$ and sgn$(\cdot)$ is defined as in (1.3).

Note that since $F^D(x_1, \ldots, x_D) = C_\theta(F_1(x_1), \ldots, F_D(x_D)) = P(U_1 \leq F_1(x_1), ..., U_D \leq F_D(x_D))$, (1) follows. To prove (2), note that $F_k(x_k) \to 0$ for all $k = 1, ..., D$ when max$(x_1, \ldots, x_D) \to -\infty$. Hence, $F^D(x_1, \ldots, x_D) \to 0$ from the property of a cumulative distribution function. (3) follows similarly. (4) is the direct result of (1.3) and (5.1). Proof completed.

Define $\mathcal{C}(F_k)$ to be the collection of all points of continuity of $F_k$ and $\mathcal{J}(F_k) = \{\mathcal{D}_{k1}, ..., \mathcal{D}_{k\, d_k}\}$ to be the collection of all jump points of $F_k$.

For a fixed $\mathbf{x} \in R^D$, every component of $\mathbf{x}$ corresponds to either a point of continuity or discontinuity of the corresponding marginal distribution. Suppose the indexes $c_1, c_2, ..., c_{H'} \in \{1, 2, ..., D\}$ corresponds to $x_{c_h} \in \mathcal{C}(F_{c_h})$ (i.e., $x_{c_h}$ belongs to the set of continuity points of $F_{c_h}$) and the remaining indexes, say $d_1, d_2, ..., d_{L'}$, with $H' + L' = D$, are the indexes of marginals such that

$$x_{d_l} = \mathcal{D}_{d_l, i_{d_l}}, \;\; l = 1, 2, ..., L'.$$

When $\{c_1, c_2, ..., c_{H'}\} = \{1, 2, ..., D\}$, the density of $F^D$ is given by

$$dF^D = c_\theta(F_1(x_1), \ldots, F_D(x_D)) \prod_{k=1}^{D} f_k(x_k),$$

and if $\{c_1, c_2, ..., c_{H'}\} \subset \{1, 2, ..., D\}$ the density of $F^D$ is given by

$$dF^D = \int \cdots \int_{\prod_{l=1}^{L'}[F_{d_l}(\mathcal{D}^-_{d_l,i_{d_l}}),F_{d_l}(\mathcal{D}_{d_l,i_{d_l}})]} c_\theta(u_1, \ldots, u_D) du_{d_1} ... du_{d_{L'}} \cdot \prod_{h=1}^{H'} f_{c_h}(x_{c_h}),$$

where $u_{c_h} = F_{c_h}(x_{c_h})$. Note that the above can be written as

$$c^*_\theta(F_1, ...F_D)(\mathbf{x}) \cdot \prod_{h=1}^{H'} f_{c_h}(x_{c_h}).$$

Letting $\mathbf{X}_j = (X_{1j}, X_{2j}, ..., X_{Dj})^T$ where $j = 1, ..., n$ represent a random sample from $F^D$, the semi-parametric estimator $\hat{\theta}_n$ of $\theta$ would then be obtained as a solution of the system

$$\frac{1}{n} \sum_{j=1}^{n} \frac{\partial}{\partial \theta_i} \log[dF^D(X_{1j}, \ldots, X_{Dj})] = 0, \ (1 \le i \le q) \tag{5.2}$$

Since $\theta$ is the only vector of parameters of interest, the components of likelihood that matters is

$$\frac{1}{n} \sum_{j=1}^{n} \frac{\partial}{\partial \theta_i} \log[c^*_\theta(F_1, ...F_D)(X_{1j}, \ldots, X_{Dj})] = 0, \ (1 \le i \le q) \tag{5.3}$$

Using the same techniques as in Chapter 3, Theorem 3.1.1 can be extended to the $D$-dimensional case. Furthermore, the limiting variance-covariance matrix of $n^{1/2}(\hat{\theta}_n - \theta)$ is then $B^{-1}\Sigma B^{-1}$, where $B$ is the information matrix associated with $C_\theta$. To give the explicit form of $\Sigma$, we need some notations:

Let

$$F^p_k(x_k) = \begin{cases} F_k(x_k) \equiv u_{k2}, & \text{if } x_k \in \mathcal{C}(F_k), \\ (F_k(x_k^-), F_k(x_k)) \equiv (u_{k1}, u_{k2}), & \text{if } x_k \in \mathcal{J}(F_k). \end{cases}$$

Note that in the present case, the relevant statistic in (5.3) can be written as, for $1 \le i \le q$,

$$\begin{aligned} R_{i,n} &= \frac{1}{n} \sum_{j=1}^{n} J_i(F_1(X^-_{1j}), F_1(X_{1j}), ..., F_D(X^-_{Dj}), F_D(X_{Dj})) \\ &= \frac{1}{n} \sum_{\mathbf{Q}} \sum_{j=1}^{n} J^{\mathbf{Q}}_i(F^p_1(X_{1j}), ..., F^p_D(X_{Dj})), \end{aligned}$$

where $\mathbf{Q} = \{Q_1, ..., Q_D\}$ is a $D$-sequence consisting of letters of $\{c\}$ or $\{d\}$, with the following properties: (i) when $\mathbf{Q}$ is combined with a point $\mathbf{x}$ and the marginals $F_1, ..., F_D$, say $\mathbf{x} \in \mathbf{Q}(F_1, ..., F_D)$, a '$c$' at the $k$-th component of $\mathbf{Q}$ refers to $x_k$ belonging to $\mathcal{C}(F_k)$, and a '$d$' at the $k$-th component refers to $x_k$ belonging to $\mathcal{J}(F_k)$, for $k = 1, ..., D$; (ii) when $\mathbf{Q}$ is combined with $J_i$ or $J_i$'s derivative with respect to $u_{k1}$ or $u_{k2}$, a '$c$' at the $k$-th component refers to $F_k^p(X_{kj}) = F_k(X_{kj})$, and a '$d$' at the $k$-th component refers to $F_k^p(X_{kj}) = (F_k(X_{kj}^-), F_k(X_{kj}))$, for $k = 1, ..., D$.

Using the same techniques as those used in Chapter 3, one can see that $\Sigma$ is the variance-covariance matrix of the $q$-dimensional random vector whose $i$-th component is given by

$$\text{var}(\sum_{\mathbf{Q}} M_i^{\mathbf{Q}}(X_1, ..., X_D));$$

the following is a general form of $M_i^{\mathbf{Q}}(X_1, ..., X_D))$. Assuming $\mathbf{Q} = \{c_1, c_2, ..., c_{H'}\} \cup \{d_1, d_2, ..., d_{L'}\}$, and $L'$ and $H'$ can be any value among $\{0, ..., D\}$, such that $H' + L' = D$. Then

$$\begin{aligned} M_i^{\mathbf{Q}}(x_1, ..., x_D) = & \sum_{i_{d_1}} \cdots \sum_{i_{d_{L'}}} J_i^{\mathbf{Q}}(F_1^p(x_1), , ..., F_D^p(x_D)) \prod_{h=1}^{H'} I_{\{x_{c_h} \in \mathcal{C}(F_{c_h})\}} \cdot \\ & \textstyle\prod_{l=1}^{L'} I_{\{x_{d_l} = \mathcal{D}_{d_l\, i_{d_l}}\}} + \sum_{k=1}^{D} L_{ik}. \end{aligned}$$

The set of jump points for $F_{d_l}$ is $\{\mathcal{D}_{d_l 1}, ..., \mathcal{D}_{d_l\, d_{d_l}}\}$, and $L_{ik}$ is given as below:
(i) if the $k$-th component in $\mathbf{Q} = c$;

$$\begin{aligned} L_{ik} = & \sum_{i_{d_1}} \cdots \sum_{i_{d_{L'}}} \int_{\{R \backslash \{\cup_{n=1}^{d_{c_1}} \mathcal{D}_{c_1\, n}\}\}} \cdots \int_{\{R \backslash \{\cup_{n=1}^{d_{c_{H'}}} \mathcal{D}_{c_{H'}\, n}\}\}} I(x_k \leq x_k') \cdot \\ & J_{iu_{k2}}^{\mathbf{Q}}(F_1^p(x_1'), ..., F_D^p(x_D')) dF(x_1', ..., x_D') \end{aligned}$$

with $x_{d_l}' = \mathcal{D}_{d_l\, i_{d_l}}$;

(ii) if the $k$-th component in $\mathbf{Q} = d$;

$$
\begin{aligned}
L_{ik} = & \sum_{i_{d_1}} \cdots \sum_{i_{d_{L'}}} \int_{\{R \backslash \{\cup_{n=1}^{d_{c_1}} \mathcal{D}_{c_1\, n}\}\}} \cdots \int_{\{R \backslash \{\cup_{n=1}^{d_{c_{H'}}} \mathcal{D}_{c_{H'}\, n}\}\}} I(x_k < x_k^{'}) \cdot \\
& J_{iu_{k1}}^{\mathbf{Q}} (F_1^p(x_1^{'}), ..., F_D^p(x_D^{'})) dF(x_1^{'}, ..., x_D^{'}) \\
& + \sum_{i_{d_1}} \cdots \sum_{i_{d_{L'}}} \int_{\{R \backslash \{\cup_{n=1}^{d_{c_1}} \mathcal{D}_{c_1\, n}\}\}} \cdots \int_{\{R \backslash \{\cup_{n=1}^{d_{c_{H'}}} \mathcal{D}_{c_{H'}\, n}\}\}} I(x_k \leq x_k^{'}) \cdot \\
& J_{iu_{k2}}^{\mathbf{Q}} (F_1^p(x_1^{'}), ..., F_D^p(x_D^{'})) dF(x_1^{'}, ..., x_D^{'}),
\end{aligned}
$$

with $x_{d_l}^{'} = \mathcal{D}_{d_l\, i_{d_l}}$.

As the parallel with the case $q = 1$ and $D = 2$ described earlier in Chapter 2 and Chapter 3, an estimator of the variance-covariance matrix of $\hat{\theta}_n$ could be found by repeating the procedure described in Chapter 4.

# CHAPTER 6

# JOINT DISTRIBUTIONS USING $t$-COPULA

## 6.1 Joint Distributions Using $t$-copula

In this Chapter, we develop the joint distributions with the family of $t$-copulas, see, [34], and [8]. The reason we choose $t$-copulas is not only it is a generalization of the Gaussian copulas, but also it has the great tail dependence property. This property allows us to study the limiting association between random variables $X$ and $Y$ as both $x$ and $y$ go to their boundaries. Our finding is that the limiting association is governed by the correlation coefficient $\rho$ together with the degrees of freedom $\nu$, which is listed in Lemma 6.1.3. In reality, there are situations where random variables are still associated in a certain level even in the tails. For instance, in biometric recognition, the genuine and imposter distributions generated from the same biometric trait or different biometric traits in a multimodal biometric system, tend to have not exact the same but similar tail behavior, as shown in Chapter 7, Section 7.2.2. The $t$-copula models can fit in this case more appropriately than a copula family which does not have the tail dependency.

Before defining the $t$-copula, we develop some notations for the presentation. Let $\Sigma$ denote a positive definite matrix of dimension $D \times D$. For such a $\Sigma$, the $D$-dimensional $t$ density with $\nu$

degrees of freedom will be denoted by

$$f_{\nu,\Sigma}^{D}(\mathbf{x}) \equiv \frac{\Gamma(\frac{\nu+D}{2})}{(\pi\nu)^{\frac{D}{2}}\Gamma(\frac{\nu}{2})|\Sigma|^{\frac{1}{2}}} \left(1 + \frac{\mathbf{x}^T\Sigma^{-1}\mathbf{x}}{\nu}\right)^{-\frac{\nu+D}{2}},$$

with corresponding cumulative distribution function

$$t_{\nu,\Sigma}^{D}(\mathbf{x}) \equiv \int_{\mathbf{w}\leq\mathbf{x}} f_{\nu,\Sigma}^{D}(\mathbf{w})\, d\mathbf{w}.$$

The matrix $\Sigma$ with unit diagonal entries corresponds to a correlation matrix and will be denoted by $R$.

The $D$-dimensional $t$-copula function is given by

$$C_{\nu,R}(\mathbf{u}) \equiv \int_{\mathbf{w}\leq t_{\nu}^{-1}(\mathbf{u})} f_{\nu,R}^{D}(\mathbf{w})\, d\mathbf{w} \tag{6.1}$$

where $t_{\nu}^{-1}(\mathbf{u}) \equiv (t_{\nu}^{-1}(u_1), \ldots, t_{\nu}^{-1}(u_D))^T$, $t_{\nu}^{-1}$ is the inverse of the cumulative distribution function of univariate $t$ with $\nu$ degrees of freedom, and $R$ is a $D \times D$ correlation matrix. Note that $C_{\nu,R}(\mathbf{u}) = \mathrm{P}\{\mathbf{X} \leq t_{\nu}^{-1}(\mathbf{u})\}$, for $\mathbf{X} = (X_1, \ldots, X_D)^T$ distributed as $t_{\nu,R}^{D}$, demonstrating that $C_{\nu,R}(\mathbf{u})$ is a distribution function on $(0,1)^D$ with uniform marginals. The density corresponding to $C_{\nu,R}(\mathbf{u})$ is given by

$$c_{\nu,R}(\mathbf{u}) = \frac{\partial^D C_{\nu,R}(\mathbf{u})}{\partial u_1 \partial u_2 \ldots \partial u_D} = \frac{f_{\nu,R}^{D}\{t_{\nu}^{-1}(\mathbf{u})\}}{\prod_{k=1}^{D} f_{\nu}\{t_{\nu}^{-1}(u_k)\}}, \tag{6.2}$$

where $f_{\nu}$ in (6.2) is the density of the univariate $t$ distribution with $\nu$ degrees of freedom.

We consider the joint distributions of the form

$$F_{\nu,R}^{D}(\mathbf{x}) = C_{\nu,R}\{F_1(x_1), \ldots, F_D(x_D)\} \tag{6.3}$$

with $C_{\nu,R}$ defined as in (6.1). It follows from the properties of a copula function that $F_{\nu,R}^{D}$ is a valid multivariate distribution function on $R^D$. The identifiability of the marginal distributions $F_k$, $k = 1, \ldots, D$, the correlation matrix $R$, and the degrees of freedom parameter, $\nu$, are established in the following theorem:

**Theorem 6.1.1.** *Let $F^D_{\nu_1,R_1}$ and $G^D_{\nu_2,R_2}$ denote two distribution functions on $R^D$ obtained from equation (6.3) with marginal distributions $F_k, k = 1,\ldots,D$ and $G_k, k = 1,\ldots,D$, respectively. Suppose we have $F^D_{\nu_1,R_1}(\mathbf{x}) = G^D_{\nu_2,R_2}(\mathbf{x})$ for all $\mathbf{x}$. Then, $F_k(x) = G_k(x)$ for all $k$, $\nu_1 = \nu_2$ and $R_1 = R_2$.*

In order to prove identifiability of $(\nu, R)$ in Theorem 6.1.1, we first must state and prove several lemmas.

**Lemma 6.1.1** Fix $\nu$. Let $F^D_{\nu,R_1}$ and $G^D_{\nu,R_2}$ denote two cumulative distribution functions on $R^D$ obtained from equation (6.3) with marginal distributions $F_k$, $k = 1,\ldots,D$ and $G_k$, $k = 1,\ldots,D$, respectively. Suppose we have

$$F^D_{\nu,R_1}(\mathbf{x}) = G^D_{\nu,R_2}(\mathbf{x}) \tag{6.4}$$

for all $\mathbf{x}$. Then, $F_k(x) = G_k(x)$ for all $k$ and $R_1 = R_2$.

**Proof:** By taking $x_i$, $i \neq k$ tending to $\infty$, we get that $F^D_{\nu,R_1}(\mathbf{x}) \to F_k(x_k)$ and $G^D_{\nu,R_2}(\mathbf{x}) \to G_k(x_k)$. It follows that $F_k(x) = G_k(x)$ for all $k$.

Next we show that the correlation matrices $R_1$ and $R_2$ are equal. We first prove this result for $D = 2$. Note that when $D = 2$, $R_1$ and $R_2$ can be determined by one correlation parameter, namely, $\rho_1$ and $\rho_2$, respectively, so, we have

$$F^2_{\nu,\rho_1}(\mathbf{x}) = C_{\nu,\rho_1}\{F_1(x_1), F_2(x_2)\} \equiv \int_{\mathbf{w} \leq t_\nu^{-1}(\mathbf{v})} f^2_{\nu,\rho_1}(\mathbf{w})\, d\mathbf{w} \tag{6.5}$$

and

$$G^2_{\nu,\rho_2}(\mathbf{x}) = C_{\nu,\rho_2}\{F_1(x_1), F_2(x_2)\} \equiv \int_{\mathbf{w} \leq t_\nu^{-1}(\mathbf{v})} f^2_{\nu,\rho_2}(\mathbf{w})\, d\mathbf{w} \tag{6.6}$$

where $\mathbf{v} = (F_1(x_1), F_2(x_2))^T$, so, one only needs to prove $\rho_1 = \rho_2$.

Now, for any real numbers $a$ and $b$, the bivariate $t$-cumulative distribution function, $t^2_{\nu,\rho}(a,b)$, with $\nu$ degrees of freedom and correlation parameter $\rho$ is a strictly increasing function of $\rho$. Note that

$$\begin{aligned} t^2_{\nu,\rho}(a,b) &= \int_{-\infty}^{a}\int_{-\infty}^{b} f^2_{\nu,\rho}(m,n)\,dm\,dn \\ &= \int_{-\infty}^{a}\int_{-\infty}^{b}\int_0^\infty \phi(m,n,\rho,w)\, g_\nu(w)\, dw\, dm\, dn, \end{aligned}$$

where

$$\phi(m, n, \rho, w) = \frac{1}{2\pi w(1-\rho^2)^{1/2}} \exp\left\{-\frac{m^2 - 2\rho mn + n^2}{2w(1-\rho^2)}\right\}$$

and $g_\nu(w)$ is the probability density function of the inverse Gamma distribution defined as

$$g_\nu(w^2) \sim IG(\frac{\nu}{2}, \frac{2}{\nu}). \tag{6.7}$$

Differentiating $t^2_{\nu,\rho}(a, b)$ with respect to $\rho$, we get

$$\begin{aligned}
\frac{\partial t^2_{\nu,\rho}(a,b)}{\partial\rho} &= \int_{-\infty}^{a}\int_{-\infty}^{b}\int_{0}^{\infty} \frac{\partial\phi(m,n,\rho,w)}{\partial\rho}\, g_\nu(w)\, dw\, dm\, dn \\
&= \int_{-\infty}^{a}\int_{-\infty}^{b}\int_{0}^{\infty} w\, \frac{d^2\phi(m,n,\rho,w)}{dm\, dn}\, g_\nu(w)\, dw\, dm\, dn \\
&= \int_{0}^{\infty} w\, \phi(m,n,\rho,w)\, g_\nu(w)\, dw > 0,
\end{aligned}$$

using the fact that $\partial\phi(m, n, \rho, w)/d\rho = w\, \partial^2\phi(m, n, \rho, w)/\partial m\partial n$.

Note that (6.5) and (6.6) can be re-written as $F^2_{\nu,\rho_j}(\mathbf{x}) = t^2_{\nu,\rho_j}(a, b)$ for $j = 1, 2$ with $a = t_\nu^{-1}(F_1(x_1))$ and $b = t_\nu^{-1}(F_2(x_2))$. So, $F^2_{\nu,\rho_1}(\mathbf{x}) = F^2_{\nu,\rho_2}(\mathbf{x})$, which gives

$$t^2_{\nu,\rho_1}(a, b) = t^2_{\nu,\rho_2}(a, b).$$

Since when $\nu$ is fixed, for any $a$ and $b$, $t^2_{\nu,\rho_j}(a, b)$ is an strictly increasing function of $\rho$, the equality above implies that $\rho_1 = \rho_2$ must hold. The proof is completed.

Now, we prove Lemma 6.1.1 for general $D$. Let $R_1 = ((\rho_{kk',1}))$ and $R_2 = ((\rho_{kk',2}))$ be the representation of the correlation matrices $R_1$ and $R_2$ in terms of their entries. Note that the condition (6.4) can be reduced to the case $D = 2$ by taking $x_i$, $i \neq k, k'$ tending to infinity. Thus, we have $F^2_{\nu,\rho_{kk',1}}(x_k, x_{k'}) = G^2_{\nu,\rho_{kk',2}}(x_k, x_{k'})$, but this implies $\rho_{kk',1} = \rho_{kk',2}$ from the case $D = 2$.

Next, we state and prove two more relevant lemmas:

**Lemma 6.1.2** For $a > 0$, $t_\nu\{-(a\nu)^{1/2}\}$ is a decreasing function of $\nu$.

**Proof:** Let $Y \sim t_\nu(\cdot)$. Then $Z = Y/\nu^{1/2}$ has pdf $f_\nu(z) = [\Gamma\{(\nu+1)/2\}/\{\Gamma(\nu/2)\}]$ $\left(1 + z^2\right)^{-(\nu+1)/2} \pi^{-1/2}$. One can see that if $\nu_1 \leq \nu_2$, $\{f_{\nu_2}(|z|)/f_{\nu_1}(|z|)\}$ is a strictly decreasing function of $|z|$. Since $I_{\{|Z| \geq a^{1/2}\}}$ is a nondecreasing function of $a$, mimicking the proof of Lemma 2 in [28], one can prove that $E_\nu(I_{\{|Z|\geq a^{1/2}\}})$ is a decreasing function of $\nu$. It follows that

$$t_\nu\{-(a\nu)^{1/2}\} = \mathrm{P}\{Y \leq -(a\nu)^{1/2}\} = \mathrm{P}\{Z \leq -a^{1/2}\} = \frac{E_\nu(I_{\{|Z|\geq a^{1/2}\}})}{2}$$

is decreasing in $\nu$.

**Lemma 6.1.3** For the bivariate $t$-copula $C_{\nu,\rho}(u,v)$, let

$$\lambda^* = \lim_{v\to 0+}\left\{\frac{\partial C_{\nu,\rho}(u,v)}{\partial v}\right\}_{u=v} \quad \text{and} \quad \mu^* = \lim_{v\to 0+}\left\{\frac{\partial C_{\nu,\rho}(u,v)}{\partial v}\right\}_{u=1-v}.$$

Then, it follows that

$$\lambda^* = t_{\nu+1}\left[-\left\{(\nu+1)\frac{1-\rho}{1+\rho}\right\}^{1/2}\right] \text{ and } \mu^* = t_{\nu+1}\left[-\left\{(\nu+1)\frac{1+\rho}{1-\rho}\right\}^{1/2}\right].$$

**Proof:** It is not hard to see that if $(X,Y)^T \sim t^2_{\nu,\rho}$, then given $Y=y$,

$$\left(\frac{\nu+1}{\nu+y^2}\right)^{1/2}\frac{X-\rho y}{(1-\rho^2)^{1/2}} \sim t_{\nu+1}. \tag{6.8}$$

Therefore

$$\begin{aligned} C_{\nu,\rho}(u,v) &= \mathrm{P}(X \leq t_\nu^{-1}(u), Y \leq t_\nu^{-1}(v)) \\ &= \int_{-\infty}^{t_\nu^{-1}(v)} t_{\nu+1}\left[\left(\frac{\nu+1}{\nu+y^2}\right)^{1/2}\left\{\frac{t_\nu^{-1}(u)-\rho y}{(1-\rho^2)^{1/2}}\right\}\right] f_\nu(y)\,dy. \end{aligned}$$

Taking derivative with respect to $v$, we get

$$\begin{aligned} \frac{\partial C_{\nu,\rho}(u,v)}{\partial v} &= t_{\nu+1}\left[\left\{\frac{\nu+1}{\nu+t_\nu^{-1}(v)^2}\right\}^{1/2}\left\{\frac{t_\nu^{-1}(u)-\rho t_\nu^{-1}(v)}{(1-\rho^2)^{1/2}}\right\}\right] \\ &= t_{\nu+1}\left[\left\{\frac{\nu+1}{\nu+t_\nu^{-1}(v)^2}\right\}^{1/2}\left\{\frac{t_\nu^{-1}(v)-\rho t_\nu^{-1}(v)}{(1-\rho^2)^{1/2}}\right\}\right], \text{ putting } u=v \\ &= t_{\nu+1}\left[\left\{\frac{\nu+1}{\nu+t_\nu^{-1}(v)^2}\right\}^{1/2}\left\{\frac{t_\nu^{-1}(v)(1-\rho)}{(1-\rho^2)^{1/2}}\right\}\right] \\ &\to t_{\nu+1}\left[-\left\{(\nu+1)\frac{1-\rho}{1+\rho}\right\}^{1/2}\right] \end{aligned}$$

as $v \to 0+$. The other expression can be derived similarly.

**Remark:** Lemma 6.1.3 tells the property of tail dependence in $t$-copulas.

**Proof of Theorem 6.1.1**

Using similar arguments as before, we only need to prove Theorem 6.1.1 for $D = 2$. It easily follows by taking $x_i \to \infty$ for $i \neq k$, that $F_k(x_k) = G_k(x_k)$ for $k = 1, 2$. It follows that

$$C_{\nu_1,\rho_1}(u, v) = C_{\nu_2,\rho_2}(u, v) \tag{6.9}$$

for all pairs of $(u, v)$ of the form $(F_1(x_1), F_2(x_2))$. Thus, (6.9) holds for all values of $(u, v)$ in $(0, 1)^2$ as in the case when both marginals are continuous. Nevertheless, since both marginals have only a finite number of discontinuities, there are infinite values of $(u, v)$ for which the limits and derivatives in Lemma 6.1.3 can be applied to obtain $\lambda^*$ and $\mu^*$. Since (6.9) holds, we must have $\lambda_1^* = \lambda_2^*$ and $\mu_1^* = \mu_2^*$. Now without loss of generality, assume $\rho_1 \neq \rho_2$ and $\rho_1 \geq \rho_2$, from the equality $\lambda_1^* = \lambda_2^*$ and Lemma 6.1.2, we must have $\nu_1 \geq \nu_2$. On the other hand, the equality $\mu_1^* = \mu_2^*$ gives $\nu_1 \leq \nu_2$. Hence, $\rho_1 \geq \rho_2$ implies that $\nu_1 = \nu_2 = \nu,$ say. Now, using Lemma 6.1.1, we get $\rho_1 = \rho_2$. The proof of theorem is completed.

**Remark:** Note that as the degrees of freedom $\nu \to \infty$, the $t$-distribution converges asymptotically to a Gaussian distribution, so does the $t$-copula. For the Gaussian copulas, the counterpart of Theorem 6.1.1 also holds. Therefore, we have the following lemma:

**Lemma 6.1.4** Let $F^D_{R_1}$ and $G^D_{R_2}$ denote two distribution functions on $R^D$ in the Gaussian copula family with marginal distributions $F_k, k = 1, \ldots, D$ and $G_k, k = 1, \ldots, D$, respectively. Suppose we have $F^D_{R_1}(\mathbf{x}) = G^D_{R_2}(\mathbf{x})$ for all $\mathbf{x}$. Then, $F_k(x) = G_k(x)$ for all $k,$ and $R_1 = R_2$.

**Proof:** Without loss of generality, we prove Lemma for the case that $D = 2$, i.e., $\rho_1 = \rho_2$.

Notice that it suffices to prove that for any fixed $(a, b) \in R^2$,

$$P_\rho(Z_1 \leq a, Z_2 \leq b) = \Phi_{2,\rho}(a, b)$$

is increasing in $\rho$, where $Z_i \sim N(0, 1), i = 1, 2$, with correlation coefficient $\rho$.

Note that

$$\Phi_{2,\rho}(a,b) = \int_{-\infty}^{a}\int_{-\infty}^{b} \phi_\rho(z_1,z_2)dz_1dz_2,$$

which gives us

$$\begin{aligned}
\frac{\partial \Phi_{2,\rho}}{\partial \rho}(a,b) &= \int_{-\infty}^{a}\int_{-\infty}^{b} \phi'_\rho(z_1,z_2)dz_1dz_2 \\
&= \int_{-\infty}^{a}\int_{-\infty}^{b} \frac{1}{2\pi\sqrt{(1-\rho^2)}}\left[\frac{\rho}{1-\rho^2}+\frac{(z_1\rho-z_2)(z_2\rho-z_1)}{(1-\rho^2)^2}\right]\cdot \\
&\quad \exp\left\{-\frac{z_1^2-2\rho z_1z_2+z_2^2}{2(1-\rho^2)}\right\}dz_1dz_2 \\
&= \int_{-\infty}^{a}\int_{-\infty}^{b} \frac{\partial^2\phi_\rho(z_1,z_2)}{\partial z_1\partial z_2}dz_1dz_2 \\
&= \phi_\rho(a,b)
\end{aligned}$$

which is always positive. Therefore, we have proved the uniqueness of $\rho$. The proof of Lemma is completed.

**Remark:** It is not hard to see that for the Gaussian copulas, the tail dependence between random variables no longer exists.

Using notations defined in Chapter 5, the density of $F^D_{\nu,R}$ at a fixed point $\mathbf{x} \in R^D$, $dF^D_{\nu,R}(\mathbf{x})$, is a function on $R^D$ that satisfies

$$F^D_{\nu,R}(\mathbf{x}) = \int_{\mathbf{w}\le\mathbf{x}} dF^D_{\nu,R}(\mathbf{w}), \tag{6.10}$$

where $dF^D_{\nu,R}(\mathbf{x})$ is given by

$$\int\cdots\int_{\prod_{l=1}^{L'}[F_{d_l}(\mathcal{D}^-_{d_l,i_{d_l}}),F_{d_l}(\mathcal{D}_{d_l,i_{d_l}})]} c_{\nu,R}(u_1,\ldots,u_D)du_{d_1},...,du_{d_{L'}}\cdot\prod_{h=1}^{H'} f_{c_h}(x_{c_h}),$$

where $u_{c_h} = F_{c_h}(x_{c_h})$, and $H'$ and $L'$ could be any value among $\{0,1,...,D\}$ such that $H'+L'=D$. Note that the above density can be generally written as

$$c^*(\nu,R,F_1,...,F_D)(\mathbf{x})\cdot\prod_{h=1}^{H'} f_{c_h}(x_{c_h}). \tag{6.11}$$

## 6.2 Estimation of $R$ and $\nu$

### 6.2.1 Estimation of $R$ for fixed $\nu$

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $n$ independent and identically distributed $D$-dimensional random vectors arising from the joint distribution $F^D_{R,\nu}$ in (6.3), $\hat{F}_{kn}$ denote the empirical distribution function of $F_k$ and $F_{kn} = n\hat{F}_{kn}/(n+1)$.

From (6.11), the log-likelihood function corresponding to the $n$ i.i.d. observations $\mathbf{X}_j$, $j = 1, \ldots, n$, is given by

$$\begin{aligned} \tau(\nu, R) &= \sum_{j=1}^{n} \log dF^D_{\nu,R}(\mathbf{X}_j) \\ &= \sum_{j=1}^{n} \log\, c^*(\nu, R, F_1, \ldots, F_D)(\mathbf{X}_j) \end{aligned} \tag{6.12}$$

For fixed $\nu$, our estimator of $R$ is taken to be the maximizer of $\hat{\tau}(\nu, R)$, that is,

$$\hat{R}(\nu) = \text{arg max}_R\, \hat{\tau}(\nu, R). \tag{6.13}$$

Note that there are two main challenges in maximizing the likelihood $\hat{\tau}(\nu, R)$: Firstly, $\hat{\tau}(\nu, R)$ involves several integrals corresponding to discrete components in $\mathbf{X}_j$, $j = 1, \ldots, n$; $c^*$ is not available in a closed form. Secondly, $\hat{\tau}(\nu, R)$ needs to be maximized over all $D \times D$ correlation matrices. The space of all correlation matrices is a compact space of $[-1, 1]^{D^2}$ since all entries of a correlation matrix lie between $-1$ and $1$. However, this space is not easy to maximize over due to the constraint of positive definiteness placed on correlation matrices.

#### The EM Algorithm

The difficulties mentioned can be overcome with the use of the EM algorithm; see, for example, [31]. The EM algorithm is a well-known algorithm used to find the maximum likelihood estimator (MLE) of a parameter $\theta$ based on data $\mathbf{y}$ distributed according to the likelihood $\ell_{obs}(\mathbf{y}; \theta)$. In many situations, obtaining the MLE,

$$\hat{\theta}_{MLE} = \text{arg max}_{\theta}\, \ell_{obs}(\,\mathbf{y};\, \theta),$$

via maximization of $\ell_{obs}(\,\mathbf{y};\, \theta)$ over $\theta$ turns out to be difficult. In such cases, the observed likelihood can usually be expressed in terms of an integral over missing components of a complete likelihood; in other words, if $\mathbf{z}$ and $\ell_{com}(\,\mathbf{y}, \mathbf{z};\, \theta)$, respectively, denote the missing observations and the complete likelihood corresponding to $(\mathbf{y}, \mathbf{z})$, it follows that

$$\ell_{obs}(\,\mathbf{y};\, \theta) = \int_{\mathcal{Z}(\mathbf{y})} \ell_{com}(\,\mathbf{y}, \mathbf{z};\, \theta)\, d\mathbf{z}$$

where $\mathcal{Z}(\mathbf{y})$ is the range of integration of $\mathbf{z}$ subject to the observed data being $\mathbf{y}$. The EM algorithm is an iterative procedure that can be formulated in two steps: First, (i) the E-step, where the quantity

$$Q(\theta^{(N)}, \theta) = E_{\pi(\,\mathbf{z}\,|\,\theta^{(N)}, \mathbf{y})}\, \{\log \ell_{com}(\,\mathbf{y}, \mathbf{z};\, \theta)\} \tag{6.14}$$

is formed; the expectation in (6.14) is taken with respect to the conditional distribution of $\mathbf{Z}$ given $\mathbf{y}$ when evaluated at a particular value, $\theta^{(N)}$, and the conditional distribution of $\mathbf{Z}$ given $\mathbf{y}$ is given by

$$\pi(\,\mathbf{z}\,|\,\theta, \mathbf{y}) = \frac{\ell_{com}(\,\mathbf{y}, \mathbf{z};\, \theta)}{\ell_{obs}(\mathbf{y};\, \theta)}. \tag{6.15}$$

Second, (ii) the M-step, where $Q(\theta^{(N)}, \theta)$ is maximized with respect to $\theta$ to obtain $\theta^{(N+1)}$, that is,

$$\theta^{(N+1)} = \text{arg max}_{\theta}\, Q(\theta^{(N)}, \theta).$$

Starting from an initial value $\theta^{(0)}$, the sequence $\theta^{(N)},\ N \geq 0$ converges to $\hat{\theta}_{MLE}$ under suitable regularity conditions.

The integrals in $\hat{\tau}(\nu, R)$ corresponding to the discrete components can be formulated as missing components of a complete likelihood. Recall that the notations $c_h$ and $d_l$ were used to denote the discrete and continuous components in the vector $\mathbf{x}$. We now extend this notation to represent discrete and continuous components in the $j$-th observation vector $\mathbf{X}_j$, $j = 1, \ldots, n$. For $j =$

$1, \ldots, n$, let $d_{lj}$, $l = 1, \ldots, L'_j$ and $c_{hj}$, $h = 1, \ldots, H'_j$, respectively, denote the discrete and continuous components of $\mathbf{X}_j$ with respect to the corresponding marginal distributions. Next, define the vector $\mathbf{u}_j$ in the following way: The $d_{lj}$-th component of $\mathbf{u}_j$, $u_{d_{lj}}$, is a number that is allowed to vary between $[F_{d_{lj}}(x^-_{d_{lj}}), F_{d_{lj}}(x_{d_{lj}})]$, that is, $u_{lj} \in [F_{d_{lj}}(x^-_{d_{lj}}), F_{d_{lj}}(x_{d_{lj}})] \equiv \mathbf{S}_{lj}$, say, for $l = 1, \ldots, L'_j$. The $c_{hj}$-th component of $\mathbf{u}_j$, $u_{c_{hj}}$, is taken to be $u_{c_{hj}} = F_{c_{hj}}(x_{c_{hj}})$ for $h = 1, \ldots, H'_j$. In that case, we have

$$c^*(\nu, R, F_1, \ldots, F_D)(\mathbf{X}_j) = \int_{S_{1j}} \cdots \int_{S_{L'_j}} c^D_{\nu,R}(\mathbf{u}_j)\, du_{d_{1j}} \ldots du_{d_{L'_j}}$$

Making the transformation $\mathbf{z}_j = t_\nu^{-1}(\mathbf{u}_j)$, the likelihood corresponding to the $j$-th observation in (6.12) can be written as $c^*(\nu, R, F_1, \ldots, F_D)(\mathbf{X}_j) = NUM_j / DENOM_j$, where

$$NUM_j = \int_{t_\nu^{-1}(S_{1j})} \cdots \int_{t_\nu^{-1}(S_{L'_j})} f^D_{\nu,R}(\mathbf{z}_j)\, dz_{d_{1j}} \ldots dz_{d_{L'_j}}$$

and

$$DENOM_j = \prod_{h=1}^{H'_j} f_\nu(z_{c_{hj}}),$$

where $z_{c_{hj}} = t_\nu^{-1}\{F_{c_{hj}}(x_{c_{hj}})\}$ and $t_\nu^{-1}(\mathbf{S}_{lj}) = [t_\nu^{-1}\{F_{d_{lj}}(x^-_{d_{lj}})\}, t_\nu^{-1}\{F_{d_{lj}}(x_{d_{lj}})\}]$. We make several important observations. First, where the maximization of $R$ is concerned for fixed $\nu$, it is enough to consider the $NUM_j$ terms. Thus, in the EM framework, we define $NUM_j$ to be the "observed likelihood" corresponding to the $\mathbf{x}_j$:

$$\ell_{obs}(\mathbf{x}_j;\, \nu, R) = \int_{t_\nu^{-1}(\mathbf{S}_{1j})} \cdots \int_{t_\nu^{-1}(\mathbf{S}_{L'_j})} f^D_{\nu,R}(\mathbf{z}_j)\, dz_{d_{1j}} \ldots dz_{d_{L'_j}}. \tag{6.16}$$

If the variables $z_{d_{lj}}$ $j = 1, \ldots, L'_j$ in (6.16) are treated as missing, the "complete likelihood" corresponding to the $j$-th observation becomes

$$\ell_{com}(\mathbf{x}_j, \mathcal{Z}_j;\, \nu, R) = f^D_{\nu,R}(\mathbf{z}_j) \left( \prod_{l=1}^{L'_j} I_{\{z_{d_{lj}} \in t_\nu^{-1}(\mathbf{S}_{lj})\}} \right),$$

where $\mathcal{Z}_j = \{ z_{d_{lj}}, j = 1, \ldots, L'_j \}$. Next, we note that the $t$ density, $f^D_{\nu,R}$, is an infinite scale mixture of Gaussian densities, namely,

$$f^D_{\nu,R}(\mathbf{z}_j) = \int_{\sigma_j^2=0}^{\infty} \phi^D_{R,\sigma_j}(\mathbf{z}_j)\, g_\nu(\sigma_j^2)\, d\sigma_j^2$$

where

$$\phi^D_{R,\sigma_j}(\mathbf{z}) = \frac{1}{(2\pi)^{\frac{D}{2}}\, \sigma_j^D\, |R|^{\frac{1}{2}}} \exp\left( -\frac{\mathbf{z}^T R^{-1} \mathbf{z}}{2\sigma_j^2} \right)$$

and the mixing distribution on $\sigma_j$ is the inverse Gamma distribution as defined in (6.7). In other words, an extra missing component can be added into the E-step, namely, the mixing parameter, $\sigma_j$. The complete likelihood specification for the $j$-th observation now becomes

$$\ell_{com}(\mathbf{x}_j,\, \mathcal{Z}_j,\, \sigma_j;\, \nu, R) = \phi^D_{R,\sigma_j}(\mathbf{z}_j) \cdot g_\nu(\sigma_j^2) \left( \prod_{l=1}^{L'_j} I_{\{z_{d_{lj}} \in t_\nu^{-1}(\mathbf{S}_{lj})\}} \right). \tag{6.17}$$

The conditional distribution $\pi$ required to obtain $Q$ in (6.14) is defined as

$$\pi(\mathcal{Z}_j,\, \sigma_j,\, j = 1, \ldots, n;\, \nu, R\,) = \prod_{j=1}^{n} \pi_j(\mathcal{Z}_j,\, \sigma_j;\, \nu, R\,)$$

where

$$\pi_j(\mathcal{Z}_j,\, \sigma_j;\, \nu, R\,) = \frac{\ell_{com}(\mathbf{x}_j,\, \mathcal{Z}_j,\, \sigma_j;\, \nu, R\,)}{\ell_{obs}(\mathbf{x}_j;\, \nu, R\,)}; \tag{6.18}$$

see (6.15). Two distributions derived from (6.18) will be used subsequently: (i) the conditional distribution of $\sigma_j$ given $\mathcal{Z}_j$, given by

$$\pi_j(\sigma_j^2 \,|\, \mathcal{Z}_j, \nu, R) = IG\left\{ \frac{\nu + D}{2}, \left( \frac{\nu + \mathbf{z}_j^T R^{-1} \mathbf{z}_j}{2} \right)^{-1} \right\},$$

and (ii) the marginal of $\mathcal{Z}_j$s (after integrating out $\sigma_j$), given by

$$\pi_j(\mathcal{Z}_j\,;\, \nu, R) = \frac{f^D_{\nu,R}(\mathbf{z}_j) \left( \prod_{l=1}^{L'_j} I_{\{z_{d_{lj}} \in t_\nu^{-1}(t_\nu^{-1}(\mathbf{S}_{lj})\}} \right)}{\int_{t_\nu^{-1}(\mathbf{S}_{1j})} \cdots \int_{t_\nu^{-1}(\mathbf{S}_{L'_j})} f^D_{\nu,R}(\mathbf{z}_j)\, dz_{d_{1j}} \cdots dz_{d_{L'_j}}}.$$

The E-step entails that the expected value of the logarithm of the complete likelihood (6.17) is taken with respect to the missing components, $\mathcal{Z}_j$, and the $N$-th iterate of $R$, $R^{(N)}$:

$$E_{\pi(\mathcal{Z}_j,\sigma_j;\nu,R^{(N)})}\{\log\ell_{com}(\mathbf{x}_j,\mathcal{Z}_j,\sigma_j;\,\nu,R\,)\,\} \equiv E_j\,\{\log\ell_{com}(\mathbf{x}_j,\mathcal{Z}_j,\sigma_j;\,\nu,R\,)\},$$

where $\pi_j(\mathcal{Z}_j,\sigma_j;\nu,R\,)$ is as defined in (6.18). The expected value can be simplified to

$$E_j\left(-\frac{\mathbf{z}_j^T R^{-1}\mathbf{z}_j}{2\sigma_j^2}\right) - \frac{1}{2}\log|R| \tag{6.19}$$

plus other terms that do not involve the correlation matrix $R$, and hence are irrelevant for the subsequent M-step. The expectation in (6.19) is taken in two steps, namely, $E_j = E_{j1}\,E_{j2}$ where $E_{j2}$ is the conditional expectation of $\sigma_j$ given $\mathcal{Z}_j$, and $E_{j1}$ is the expectation with respect to the marginal of $\mathcal{Z}_j$. On taking $E_{j2}$, the expression in (6.19) simplifies to

$$-\frac{\nu+D}{2}\,E_{j1}\left\{\frac{z_j^T R^{-1} z_j}{\nu + z_j^T\,(R^{(N)})^{-1} z_j}\right\} = -\frac{\nu+D}{2}\,\mathrm{tr}\left(R^{-1}\,W^{(N)}\right),$$

where $W^{(N)} = ((w_{kk'}^{(N)}))$ is a $D\times D$ matrix with entries

$$w_{kk'}^{(N)} = \int \left\{\frac{z_{kj} z_{k'j}}{\nu + z_j^T\,(R^{(N)})^{-1} z_j}\right\} \pi_j(\mathcal{Z}_j\,;\,\nu, R^{(N)})\,d\mathcal{Z}_j$$

for $\mathbf{z}_j = (z_{1j},\ldots,z_{Dj})^T$.

The last integral is approximated by numerical integration on a grid. We partition each interval $t_\nu^{-1}(\mathbf{S}_{lj})$, $l = 1,\ldots,L_j'$ into a large number of subintervals and evaluate the integrand on the partition points. Finally, a Reimann sum is obtained as an approximation to the integral.

The $M$-step consists of maximizing the complete likelihood function (6.19),

$$E_j\left(-\frac{\mathbf{z}_j^T R^{-1}\mathbf{z}_j}{2\sigma_j^2}\right) - \frac{1}{2}\log|R| = -\frac{\nu+D}{2}\mathrm{tr}(R^{-1}W^{(N)}) - \frac{1}{2}\log|R|, \tag{6.20}$$

with respect to the correlation matrix $R$. Since we have to maximize the objective function in the space of all correlation matrices, this is somewhat a difficult task. We adopt the methodology presented by Barnard et al. [1] where an iterative procedure to maximize (6.20) is developed by considering the maximization of one element of $R$, say $\rho$, each time. In order to preserve the positive definiteness of $R$, one can show (see Barnard et al. [1] for details) that $\rho$ should lie in an interval $[\rho_l, \rho_u]$. The lower and upper limits of this interval are derived from the fact that in order for $R$ to be positive definite, it is both necessary and sufficient that all the principal submatrices of $R$ are positive definite. This is equivalent to a non-negativity condition on the determinant of each principal submatrix, which translates to an interval for the range of values of $\rho$. This procedure is repeated for the other elements of $R$ and cycled until convergence before going to the $(N+1)$-st iteration of the EM algorithm.

### 6.2.2 Selection of the Degrees of Freedom, $\nu$

The above procedure is carried out for a collection of degrees of freedom $\nu \in \mathcal{A}$, where $\mathcal{A}$ is finite. For each fixed $\nu$, we obtain the estimator of the correlation matrix $\hat{R}(\nu)$ based on the EM algorithm above. We select the degrees of freedom in the following way: Select $\hat{\nu}$ such that

$$\hat{\nu} = \text{arg max}_{\nu \in \mathcal{A}} \frac{1}{n} \hat{\tau}\{\nu, \hat{R}(\nu)\}. \tag{6.21}$$

## 6.3 Summary

In this Chapter, we introduced the $t$-copula family in $R^D$, and showed that for any joint distribution $F^D$ in $R^D$, there exists a unique pair $(\nu, R)$ such that $F^D_{\nu,R}(\mathbf{x}) = C_{\nu,R}(F_1(x_1), ..., F_D(x_D))$ within this family. Furthermore, we developed a semi-parametric technique to estimate the unknown pair $(\nu, R)$ using EM algorithm. The application of this approach will be presented in the next chapter.

# CHAPTER 7

# SIMULATION AND REAL DATA RESULTS

## 7.1 Simulation Results

The results in this section are based on simulated data of four experiments. The first three are for the bivariate case, whereas the fourth is for a trivariate distribution. The true distributions for the observations $\mathbf{X}_j$, $j = 1, \ldots, n$, in all experiments are of the type as defined in (6.3).

In experiment 1, $\nu_0 = \infty$ in (6.3) corresponding to the Gaussian copula function. We choose $\rho = 0.75$. The two mixed marginals, $F_1$ and $F_2$, have the following cumulative distribution

| Sample size, $n$ | Average Absolute Bias | Relative Average Absolute Bias | Coverage |
|---|---|---|---|
| 500 | 0.0168 | 2.2 % | 92.4 % |
| 1000 | 0.0135 | 1.8 % | 91.6 % |
| 1500 | 0.0116 | 1.5 % | 92.3 % |
| 2000 | 0.0113 | 1.5 % | 93.2 % |
| 2500 | 0.0087 | 1.2 % | 94.7 % |

Table 7.1. Simulation results for Experiment 1 with $\nu_0 = \infty$ and $\rho = 0.75$. The absolute bias and relative absolute bias of the estimator $\hat{\rho}_n$ are provided, together with the empirical coverage of the approximate $95\%$ confidence interval for $\rho$ based on the asymptotic normality of $\hat{\rho}_n$.

functions:

$$F_1(x_1) = 0.3I_{\{x_1 \geq 0.2\}} + 0.7\phi(x_1),$$

and

$$F_2(x_2) = 0.2I_{\{x_2 \geq 0.1\}} + 0.8\phi(x_2),$$

where $\phi(\cdot)$ is the standard normal density function. Thus, we have $d_1 = d_2 = 1$ with $\mathcal{D}_{11} = 0.2$, and $\mathcal{D}_{21} = 0.1$ with probabilities $p_{11} = 0.3$, and $p_{21} = 0.2$, respectively. The set of jump points are $\mathcal{J}(F_1) = \{0.2\}$ and $\mathcal{J}(F_2) = \{0.1\}$ with continuous components of $F_1$ and $F_2$ corresponding to $f_1(x) = f_2(x) = \phi(x)$.

The sample size $n$ is taken from $n = 500$ to $n = 2500$ in increments of $500$. In each trial, we simulate $n$ observations from $F^2_{\infty,0.75}(x_1, x_2)$ and estimate $\rho$ within the Gaussian copulas using the methodology presented in Section 6.2. Table 7.1 contains the simulation result from Experiment 1. For each sample size, the experiment was repeated 1,000 times. The average absolute bias and relative average absolute bias of the estimator $\hat{\rho}_n$ are reported, together with the empirical coverage of the approximate $95\%$ confidence interval for $\rho$ based on the asymptotic normality of $\hat{\rho}_n$.

Experiment 2 consists of the following choices: $\nu_0 = \infty$ in (6.3) corresponding to the Gaussian copula function. We choose $\rho_0 = 0.75$. The two mixed marginals, $F_1$ and $F_2$, have the following cumulative distribution functions:

$$F_1(x_1) = 0.25I_{\{x_1 \geq 0.2\}} + 0.6t_{10}(x_1) + 0.15I_{\{x_1 \geq 0.7\}},$$

and

$$F_2(x_2) = 0.2I_{\{x_2 \geq 0.3\}} + 0.5t_{10}(x_2) + 0.3I_{\{x_2 \geq 0.6\}}.$$

Thus, we have $d_1 = d_2 = 2$ with $\mathcal{D}_{11} = 0.2$, $\mathcal{D}_{12} = 0.7$, $\mathcal{D}_{21} = 0.3$ and $\mathcal{D}_{22} = 0.6$ with probabilities $p_{11} = 0.25$, $p_{12} = 0.15$, $p_{21} = 0.2$ and $p_{22} = 0.3$, respectively. The set of jump points are $\mathcal{J}(F_1) = \{0.2, 0.7\}$ and $\mathcal{J}(F_2) = \{0.3, 0.6\}$ with continuous components of $F_1$ and $F_2$ corresponding to $f_1(x) = f_2(x) = t_{10}(x)$.

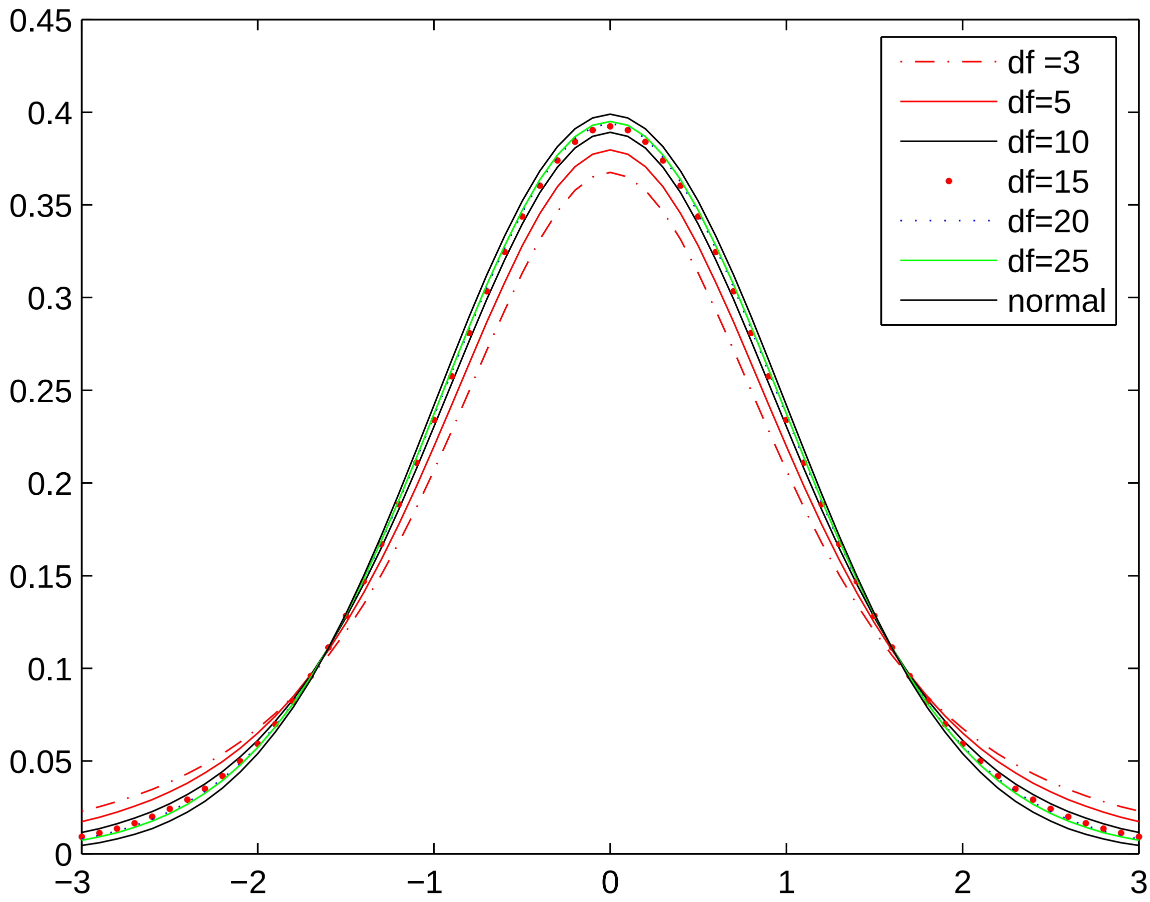


Figure 7.1. Density curves for $t$ distribution with degrees of freedom $\nu$ = 3, 5, 10, 15, 20, 25 and normal distribution. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

For now, we select the set for the degrees of freedom of the $t$-copula to be $\mathcal{A} = \{3, 5, 10, \infty\}$ for illustrative purposes. Note that we choose $\nu_0$ values in $\mathcal{A}$ so that the corresponding $t$-distributions are significantly different from one another. Figure 7.1 gives the $t$-densities for several $\nu_0$ values, including values in $\mathcal{A}$. There exist significant gaps between the $t$-density curves corresponding to $\nu_0 = \{3, 5, 10\}$, but relatively smaller gaps for $\nu_0 = \{15, 20, 25, \infty\}$. Table 7.2 provides the $L_1$-distances between some $t$- distributions in Figure 7.1.

The sample size $n$ is taken from $n = 500$ to $n = 2500$ in increments of $500$. In each trial, we

| $\rho_0$ | $\nu_0$ | 3 | 5 | 10 | $\infty$ |
|---|---|---|---|---|---|
| $\rho_0 = 0.2$ | 3 | 0 | 0.0874 | 0.1705 | 0.2690 |
| | 5 | 0.0874 | 0 | 0.0837 | 0.1835 |
| | 10 | 0.1705 | 0.0837 | 0 | 0.1005 |
| | $\infty$ | 0.2690 | 0.1835 | 0.1005 | 0 |
| $\rho_0 = 0.75$ | 3 | 0 | 0.0896 | 0.1735 | 0.2721 |
| | 5 | 0.0896 | 0 | 0.0845 | 0.1844 |
| | 10 | 0.1735 | 0.0845 | 0 | 0.1006 |
| | $\infty$ | 0.2721 | 0.1844 | 0.1006 | 0 |

Table 7.2. $L_1$ distances for paired values of $\nu_0$ corresponding to two values of $\rho_0$, $0.20$ and $0.75$.

| Sample size, $n$ | Percentage of times $\hat{\nu} = \nu_0$ | Mean $\hat{\rho}(\hat{\nu})$ | $MSE(\hat{\rho}(\hat{\nu}))$ |
|---|---|---|---|
| 500 | 86% | 0.7361 | 0.6312 $\times 10^{-3}$ |
| 1000 | 94% | 0.7392 | 0.3869 $\times 10^{-3}$ |
| 1500 | 100% | 0.7413 | 0.3808 $\times 10^{-3}$ |
| 2000 | 100% | 0.7429 | 0.3035 $\times 10^{-3}$ |
| 2500 | 100% | 0.7450 | 0.2861 $\times 10^{-3}$ |

Table 7.3. Simulation results for Experiment 2 with $\nu_0 = \infty$ and $\rho_0 = 0.75$.

simulate $n$ observations from $F^2_{\infty,0.75}(x_1, x_2)$ and estimate $\rho_0$ and $\nu_0$ based on the methodology presented in Section 6.2. The trial is repeated 50 times. The simulation results, including percentage of times (out of 50) that the true value of $\nu_0$, mean of $\hat{\rho}(\hat{\nu})$, and the $MSE$ of $\hat{\rho}(\hat{\nu})$ is chosen , are presented in Table 7.3.

In Experiment 3, we took $\nu_0 = 10$. The two generalized marginal distributions are the same as in Experiment 2. The correlation parameter $\rho_0$ were selected to be $0.20$ and $0.75$, respectively. The results are presented in Table 7.4. From the entries of Table 7.3 and Table 7.4, we see that the estimation procedure is more effective in selecting the true degrees of freedom when $\nu_0 = \infty$ compared to $\nu_0 = 10$. The reason of being that is the distribution corresponding to $\nu_0 = \infty$ is further away from all the other candidate distributions in $\mathcal{A}$. Also, the estimation procedure is less effected by the value of $\rho_0$ as illustrated by the percentage of times $\hat{\nu} = \nu_0$ column in Table 7.4.

| $\rho_0$ | Sample size, $n$ | Percentage of times $\hat{\nu} = \nu_0$ | Mean $\hat{\rho}(\hat{\nu})$ | $MSE(\hat{\rho}(\hat{\nu}))$ |
|---|---|---|---|---|
| $\rho_0 = 0.20$ | 500 | 84% | 0.1861 | 0.7532$\times 10^{-3}$ |
| | 1000 | 88% | 0.1877 | 0.6871 $\times 10^{-3}$ |
| | 1500 | 92% | 0.1889 | 0.6095 $\times 10^{-3}$ |
| | 2000 | 94% | 0.1923 | 0.4173 $\times 10^{-3}$ |
| | 2500 | 100% | 0.1944 | 0.3664 $\times 10^{-3}$ |
| $\rho_0 = 0.75$ | 500 | 82% | 0.7398 | 0.7235 $\times 10^{-3}$ |
| | 1000 | 88% | 0.7401 | 0.6789 $\times 10^{-3}$ |
| | 1500 | 96% | 0.7429 | 0.6565 $\times 10^{-3}$ |
| | 2000 | 98% | 0.7523 | 0.4546 $\times 10^{-3}$ |
| | 2500 | 100% | 0.7481 | 0.3648 $\times 10^{-3}$ |

Table 7.4. Simulation results for Experiment 3 with $\nu_0 = 10$. The two correlation values considered are $\rho_0 = 0.2$ and $\rho_0 = 0.75$.

| sample size | percentage of getting true $\nu$ | mean $\hat{\rho}_1(\hat{\nu})$ | mean $\hat{\rho}_2(\hat{\nu})$ | mean $\hat{\rho}_3(\hat{\nu})$ | total $MSE$ |
|---|---|---|---|---|---|
| 500 | 85 % | 0.1990 | 0.2760 | 0.1885 | 0.7143 $\times 10^{-3}$ |
| 1000 | 90 % | 0.1830 | 0.2805 | 0.2005 | 0.6732 $\times 10^{-3}$ |
| 1500 | 95 % | 0.1795 | 0.2845 | 0.1970 | 0.6714 $\times 10^{-3}$ |
| 2000 | 100 % | 0.1820 | 0.2880 | 0.1880 | 0.6126 $\times 10^{-3}$ |
| 2500 | 100 % | 0.1955 | 0.2920 | 0.1930 | 0.1335 $\times 10^{-3}$ |

Table 7.5. Simulation results for Experiment 4.

In Experiment 4, we took $D = 3$ and $\nu_0 = 10$. The first two marginal distributions are the same as before. The third marginal distribution is taken to be the $t$-distribution with 10 degrees of freedom (thus, having no points of discontinuity). We took the correlation matrix $R_0$ as

$$R_0 = \begin{pmatrix} 1 & 0.2 & 0.3 \\ 0.2 & 1 & 0.2 \\ 0.3 & 0.2 & 1 \end{pmatrix}_{3\times 3} = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix}_{3\times 3}, \text{say.} \tag{7.1}$$

For different sample sizes, the experiment were repeated 20 times (instead of 50) to reduce computational time. The estimators of $\rho_i$, $\hat{\rho}_i$, $i = 1, 2, 3$, were obtained based on the iterative procedure outlined in Section 6.2. Since the maximization step involves another loop within the M-step, the objective function was maximized over $\rho$-intervals in steps of $0.01$ to reduce computational time. The iterative procedure within the M-step was not required when $D = 2$ which enabled us to maximize the objective function over a finer grid (steps of $0.0001$). The results are given in Table 7.5; note that (i) the estimators converge and (ii) the MSE reduces as $n$ tends to infinity.

## 7.2 Application to Multimodal Fusion in Biometric Recognition

### 7.2.1 Introduction

Biometric recognition refers to the automatic identification of individuals based on their biological or behavioral characteristics [20]. In recent years, recognition of an individual based on his/her biometric trait has become an increasingly important method for testing "you are who you claim you are", see, for example, [18] and [29]. Biometric recognition is more reliable compared to traditional approaches, such as password-based or token-based approaches, as biometric traits cannot be easily stolen or forgotten. Some examples of biometric traits include fingerprint, face, signature, voice and hand geometry (See Figure 7.2). A number of commercial recognition systems based on

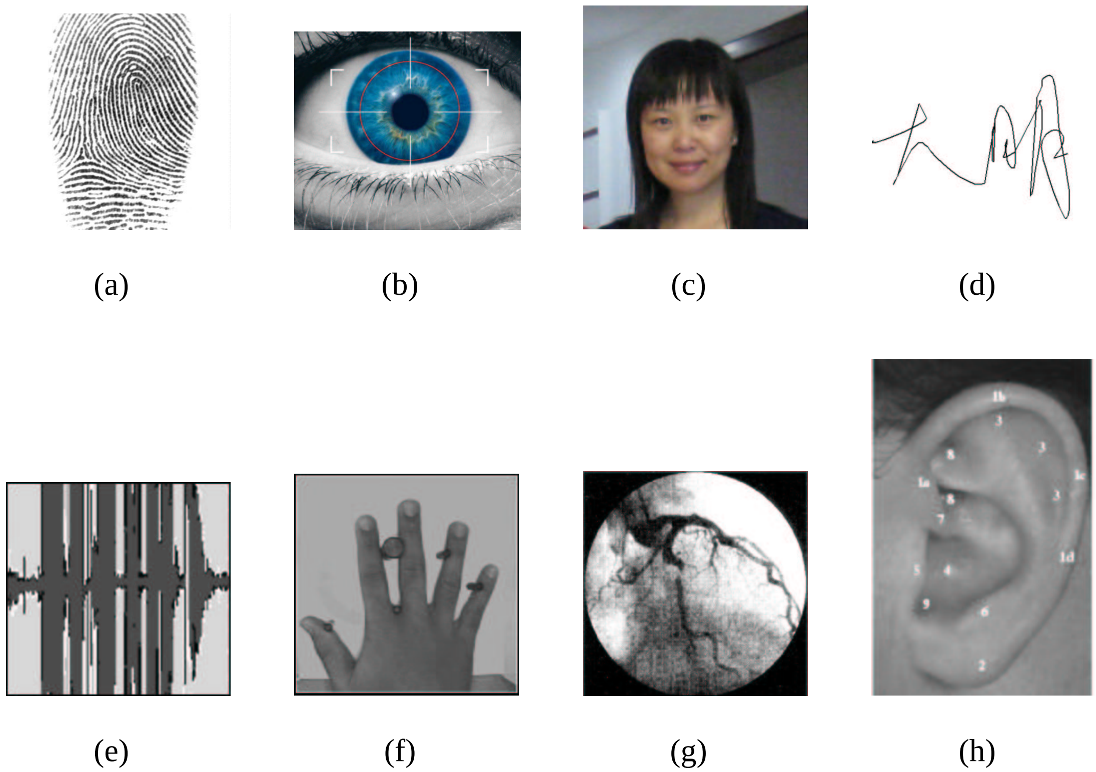


Figure 7.2. Some examples of biometric traits: (a) fingerprint, (b) iris scan, (c) face scan, (d) signature, (e) voice, (f) hand geometry, (g) retina, and (h) ear.

these traits have been deployed and are currently in use. Biometric technology has now become a viable alternative to traditional government applications (e.g., US-VISIT program [48] and the proposed biometric passport which is capable of storing biometric information of the owner in a chip inside the passport). With increasing applications involving human-computer interactions, there is a growing need for recognition techniques that are reliable and secure.

Recognition of an individual can be viewed as a test of statistical hypothesis. Based on the biometric input $Q$ and a claimed identity $I_c$, we would like to test

$$H_0 : I_t = I_c \quad \text{vs.} \quad H_1 : I_t \neq I_c, \tag{7.2}$$

where $I_t$ is the true identity of the user.

The testing in (7.2) is performed by a matcher which computes a similarity measure, $S(Q,T)$, based on $Q$ and $T$; large (respectively, small) values of $S$ indicate that $T$ and $Q$ are close to (far from) each other (A matcher can also compute a distance measure between $Q$ and $T$ in which case similar $Q$ and $T$ will produce distance values that are close to zero and vice versa). The distribution of $S(Q,T)$ is called genuine (respectively, impostor) when $I_t = I_c$ ($I_t \neq I_c$) under $H_0$ ($H_1$).

We denote the genuine (imposter) matching score distribution function by $F_{gen}$ ($F_{imp}$). Assuming that $F_{gen}(x)$ and $F_{imp}(x)$ have densities $f_{gen}(x)$ and $f_{imp}(x)$, respectively. The Neyman-Pearson theorem states that the *optimal* ROC curve is the one corresponding to the likelihood ratio statistic

$$NP(x) = \frac{f_{gen}(x)}{f_{imp}(x)}$$

[14]. The ROC curve corresponding to $NP(x)$ has the highest genuine accept rate (GAR) for every given value of the false acceptance rate (FAR) compared to any other statistic $U(x) \neq NP(x)$.

However, both $f_{gen}(x)$ and $f_{imp}(x)$ are unknown, and are estimated from the observed matching scores. The ROC corresponding to $NP(x)$ may turn out to be suboptimal, which is mainly due to the large errors in the estimation of $f_{gen}(x)$ and $f_{imp}(x)$. Thus, for a set of genuine and imposter matching scores, it is important to be able to estimate $f_{gen}(x)$ and $f_{imp}(x)$ reliably and accurately. The articles, [14] and [38], assume that the distribution function $F$ has a continuous density with no discrete components. In reality most matching algorithms apply thresholds at various stages in the matching process. When the required threshold conditions are not met, specific matching scores are output by the matcher. For example, some fingerprint matchers produce a score of zero if the number of extracted minutiae is less than a threshold. This leads to discrete components in the matching scores distribution that can not be modelled accurately using a continuous density function (see Figure 7.3, Figure 7.4, Figure 7.5 and Figure 7.6) . Thus, discrete components need to be detected and modelled seperately to avoid large errors in estimating $f_{gen}(x)$ and $f_{imp}(x)$.

Another issue is that biometric systems based on a single source of information suffer from limitations like the lack of uniqueness, non-universality and noisy data [21] and hence, may not be able to achieve the desired performance requirements of real-world applications. In contrast,

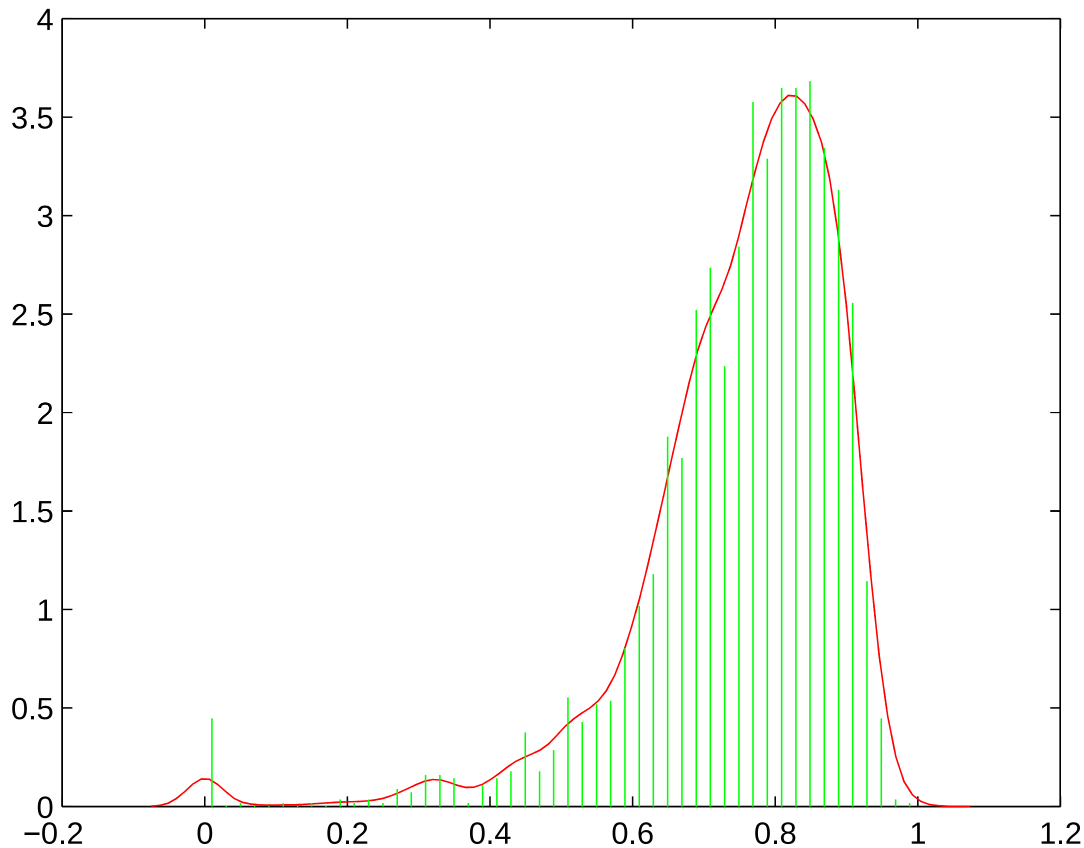


Figure 7.3. Histograms of matching scores, corresponding to genuine scores for Matcher 1. Continuous (respectively, generalized) density estimators is given by the dashed lines (solid lines).

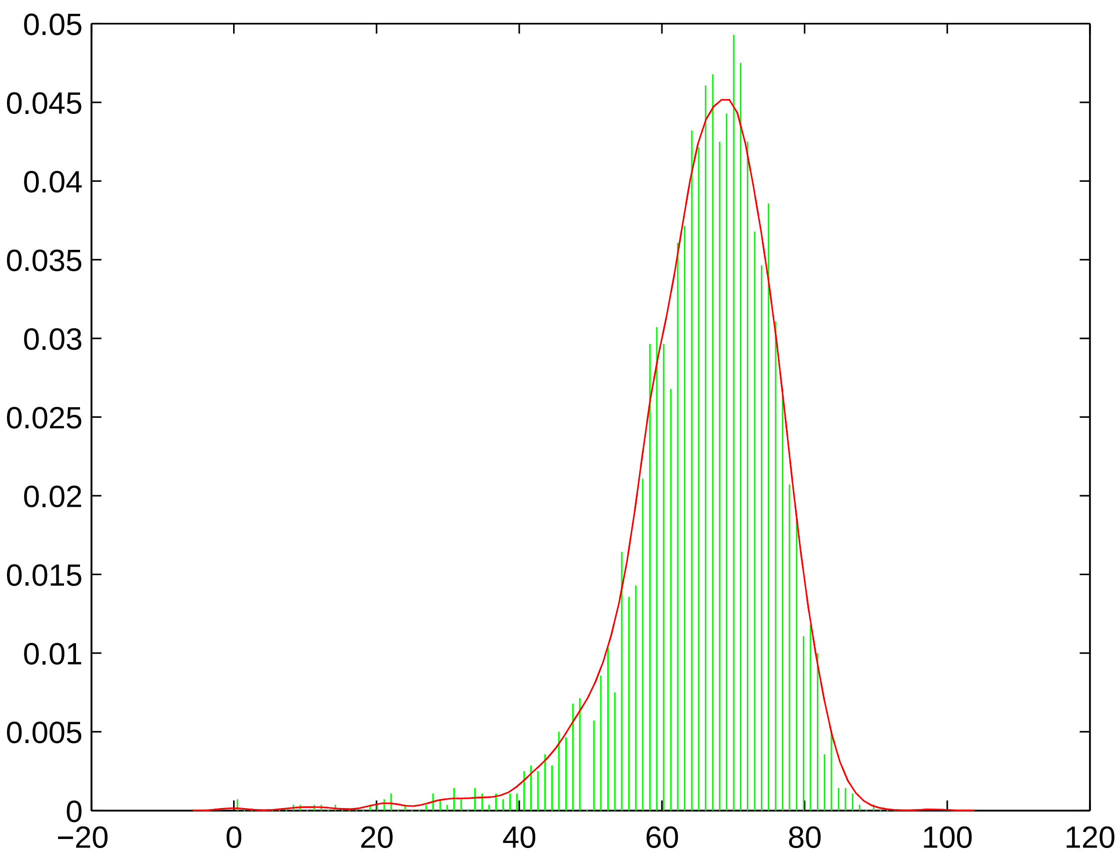


Figure 7.4. Histograms of matching scores, corresponding to genuine scores for Matcher 2. Continuous (respectively, generalized) density estimators is given by the dashed lines (solid lines).

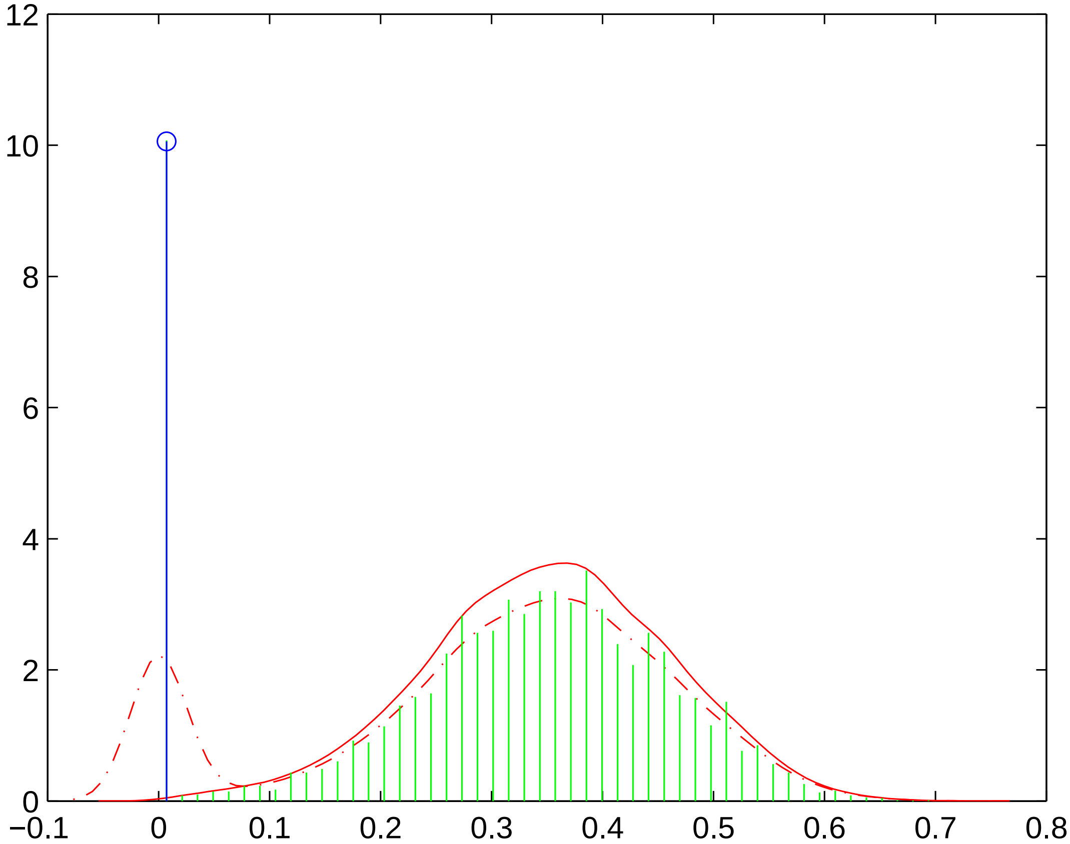


Figure 7.5. Histograms of matching scores, corresponding to impostor scores for Matcher 1. Continuous (respectively, generalized) density estimators is given by the dashed lines (solid lines). The spike corresponds to discrete components. Note how the generalized density estimator performs better compared to the continuous estimator (assuming no discrete components).

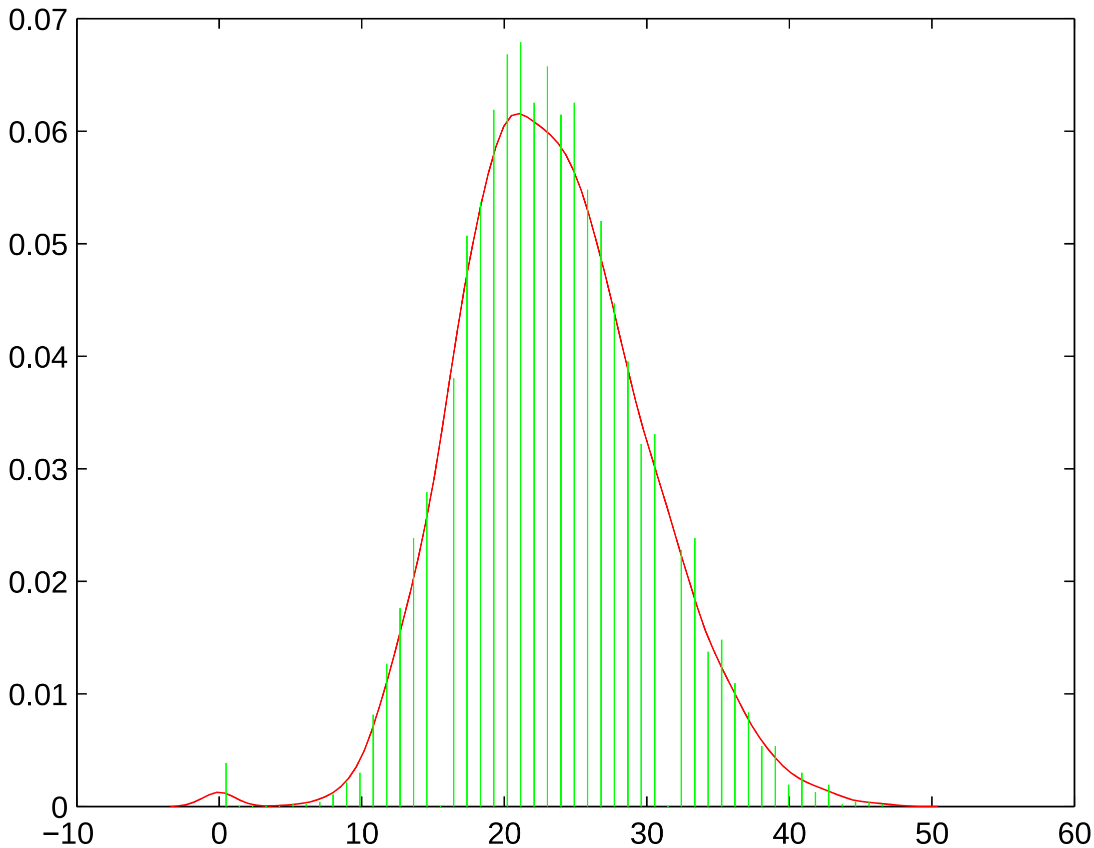


Figure 7.6. Histograms of matching scores, corresponding to impostor scores for Matcher 2. Continuous (respectively, generalized) density estimators is given by the dashed lines (solid lines).

some of the limitations imposed by unimodal biometric systems (that is, biometric systems that rely on the evidence of a single biometric trait) can be overcomed by using multiple biometric modalities[2], [24], [3], [25], [49] and [47]. Such systems, known as multibiometric systems, are expected to be more reliable due to the presence of multiple pieces of evidence. In a Multimodal biometric system, fusion can be done at (i) feature level, (ii) matching scores level, or (iii) decision level. Matching score level fusion is commonly preferred because matching scores are easily available and contain sufficient information to distinguish between a genuine and an imposter case.

Dass et al. [7] proposed a biometric fusion using generalized densities. In [7], a Gaussian copula model is chosen to estimate the correlation structure. In reality, sometimes the joint distribution can be fitted better by using a $t$-copula model instead of a Gaussian copula model, which is due to the nature of the data set, so we consider the Gaussian copula and $t$-copula models together, and choose the more appropriate model by model selection method based on $BIC$ criteria (Publication for this research is [16]).

### 7.2.2 Application in Biometric Fusion

When people deal with biometric fusion, a natural question people need to answer first is how to get the joint distribution of multiple modalities. Previously, people simply assume independence between the individual modalities, but this assumption is not always true, especially, when the fusion is done on the same biometric trait with different matchers. Here we deal with the correlation structure via semi-parametric copula models.

Based on fingerprint images in the MSU-Multimodal database, see [19], corresponding to 100 users, genuine and impostor similarity scores were obtained for two matchers: a correlation matcher, $S_1$ (see [32]), and a minutiae-based matcher, $S_2$ (see [39]). A total of 2,800 and 4,950 vectors of similarity scores, $\mathbf{X}_j \equiv (S_1(Q,T), S_2(Q,T))^T$, were obtained for the genuine and impostor cases, respectively. The histogram plot of each marginal (both genuine and impostor) gives strong indication of non-Gaussianity, thus, justifying the need for the methodology developed in

| Matching score type | $\hat{\nu}$ | ( $\hat{\rho}_1$, $\hat{\rho}_2$, $\hat{\rho}_3$ ) |
|---|---|---|
| Genuine | 14 | ( 0.76, -0.11, -0.14 ) |
| Impostor | 25 | ( 0.3, 0.04, 0.02 ) |

Table 7.6. Results of the estimation procedure for $R$ and $\nu$ based on the NIST database.

this thesis. The match scores are highly correlated since $S_1$ and $S_2$ are applied to the same fingerprint images. Further, both $S_1$ and $S_2$ output the discrete score '0' if certain "initial conditions" are not met, resulting in a spike at $0$ in the corresponding marginal distributions. For both the genuine and impostor distributions, the set of degrees of freedom, $\nu$, considered is $\mathcal{A} = \{1, 2, 3, \ldots 25, \infty\}$. We obtained $\hat{\nu} = 3$ and $\hat{\rho}(\hat{\nu}) = 0.4178$ for the genuine scores, $\hat{\nu} = 3$ and $\hat{\rho}(\hat{\nu}) = 0.1563$ for the impostor scores.

For the reasons mentioned above, joint distribution functions of the form (6.3) with $D = 2$ are appropriate for strongly correlated biometric data as we have here.

The second experiment was carried out on the first partition of the Biometric Scores Set - Release I (BSSR1) released by NIST, [33], consisting of face and fingerprint images from 517 users. Like the previous case, the marginal distributions have discontinuities: The first matcher, $S_1$, for the face biometric outputs the value -1 if certain alignment conditions do not hold for the query and template face pair. The second face matcher, $S_2$, outputs continuous values and therefore, does not have any discrete components. The second fingerprint matcher of the MSU-Multimodal database, renamed $S_3$ here, is used for the fingerprint images resulting in a discrete score of '0'. We obtained 517 and 266,772 vectors of similarity scores, $\mathbf{X}_j \equiv (S_1(Q,T), S_2(Q,T), S_3(Q,T))^T$, corresponding to the genuine and impostor scores, respectively. In this case, $D = 3$ and the correlation matrix $R_0$ can be written as in (7.1). $\mathcal{A}$ is taken as before. Table 7.6 gives the results of the estimation procedure.

We investigate the performance of fusion obtained by combining the $D$ similarity scores obtained from the $D$ different modalities. Since we assume that the genuine and impostor distributions are of the form (6.3), the test of hypotheses (7.2) can be re-stated in terms of the score

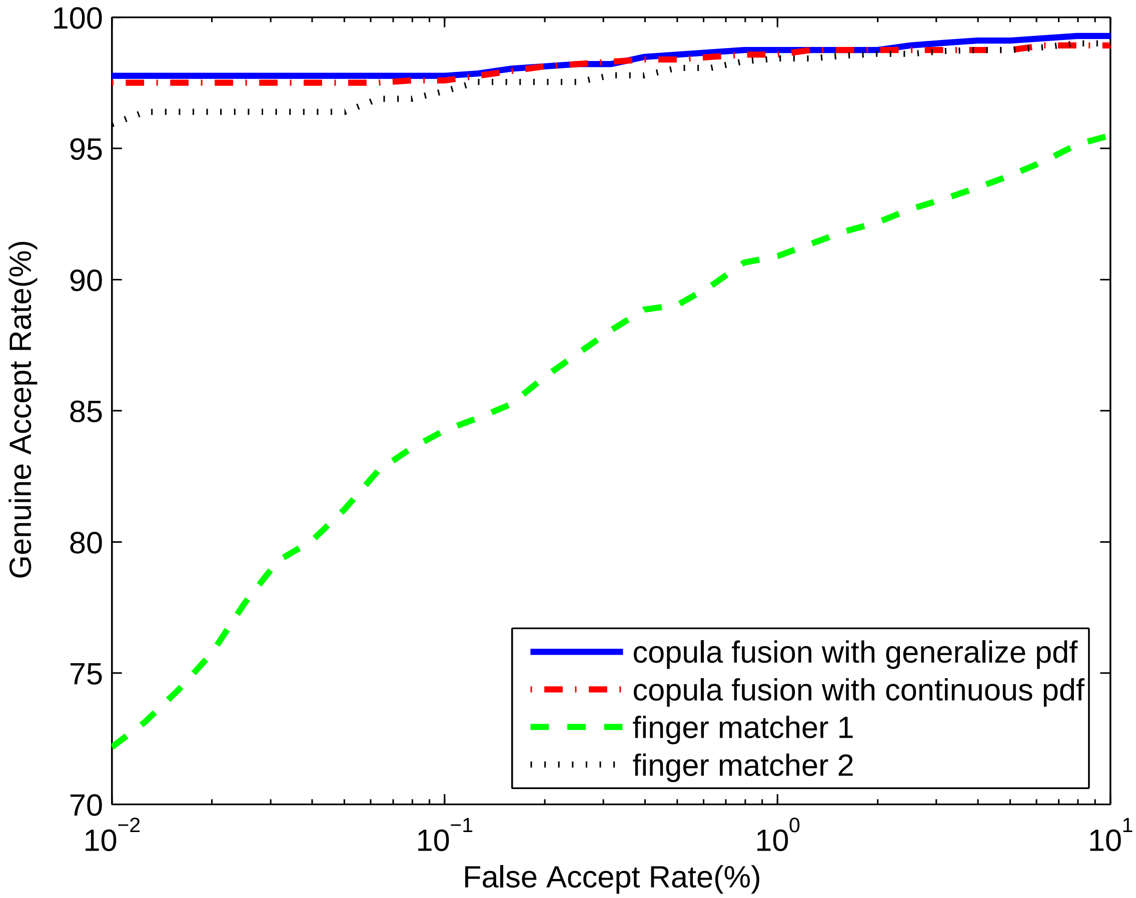


Figure 7.7. Performance of copula fusion on the MSU-Multimodal database

distributions under $H_0$ and $H_1$, namely,

$$H_0 : F_Q(\mathbf{x}) = F^D_{\nu_0,R_0}(\mathbf{x}) \qquad \text{vs.} \quad H_1 : F_Q(\mathbf{x}) = F^D_{\nu_1,R_1}(\mathbf{x})$$

for some $(\nu_0, R_0)$ and $(\nu_1, R_1)$. The optimal decision rule (fusion rule) then turns out to be the likelihood ratio $LR(\mathbf{x}) = dF^D_{\nu_0,R_0}(\mathbf{x})/dF^D_{\nu_1,R_1}(\mathbf{x})$from the Neyman-Pearson Lemma, following in a similar fashion for the case in the single modality explained earlier. However, the $LR$ rule cannot be used in the current form since the parameters $\nu_0, R_0, \nu_1$ and $R_1$ are unknown. The methodology developed here can be used to obtain estimators of all of these parameters, thus obtaining the estimated likelihood ratio statistic $\hat{LR}(\mathbf{x}) = dF^D_{\hat{\nu}_0,\hat{R}_0}(\mathbf{x})/dF^D_{\hat{\nu}_1,\hat{R}_1}(\mathbf{x})$.

The effectiveness of the (estimated) $LR$ fusion rule can be evaluated based on a $K$-fold cross

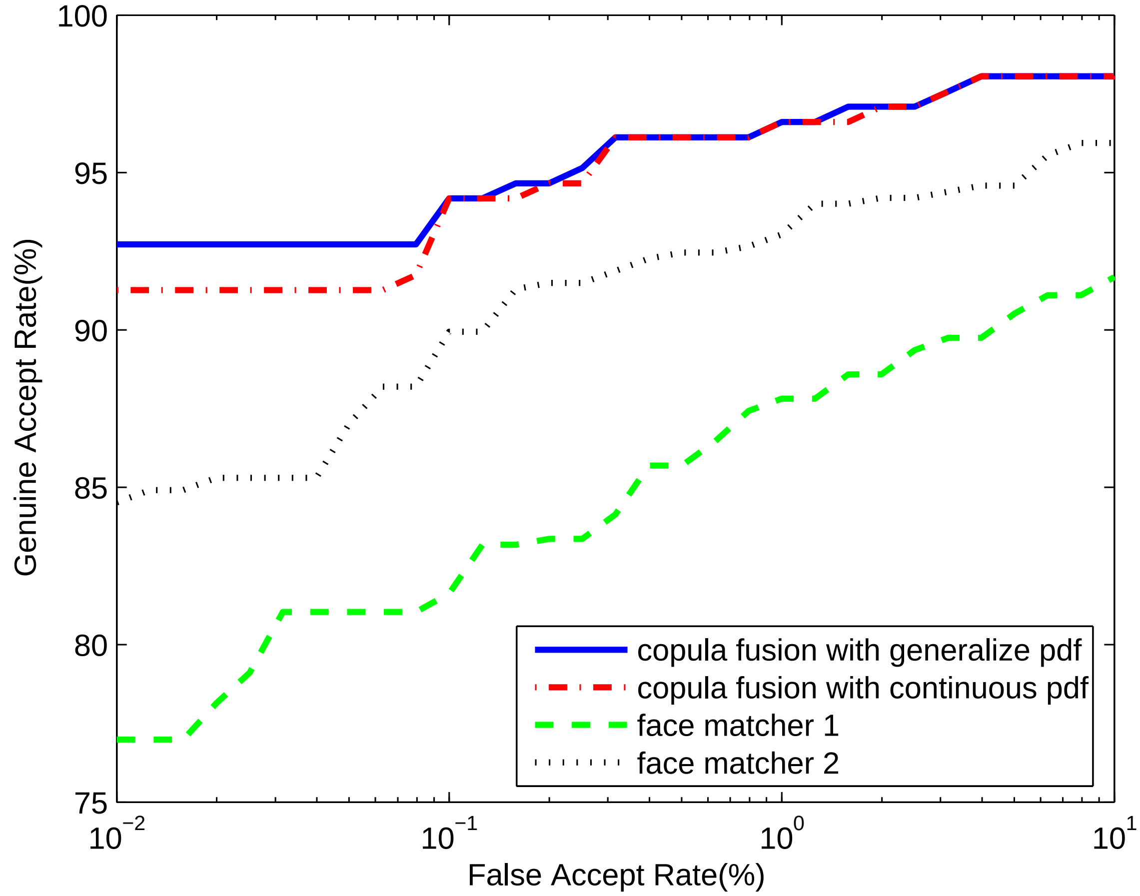


Figure 7.8. Performance of copula fusion on the NIST database.

validation procedure. In the $k$-th iteration, $k = 1, \ldots, K$, a random subset, say $\mathcal{S}_0$, of $n_0$ genuine scores from the total of $n_{gen}$ scores is selected for estimating the parameters $\nu_0$ and $R_0$. Similarly, for estimating $\nu_1$ and $R_1$, a (random) subset $n_1$ impostor scores, say $\mathcal{S}_1$, is selected from a total of $n_{imp}$ scores. The remaining genuine and impostor scores are used to obtain an estimator of the false accept and genuine accept rates (FAR and GAR, respectively) for each threshold $\lambda$. The relevant formulas are

$$F\hat{A}R(\lambda) = \frac{\sum_{j \in \mathcal{S}_1^c} I_{\{L\hat{R}(\mathbf{x}_j) > \lambda\}}}{n_{imp} - n_1} \text{ and } G\hat{A}R(\lambda) = \frac{\sum_{j \in \mathcal{S}_0^c} I_{\{L\hat{R}(\mathbf{x}_j) > \lambda\}}}{n_{gen} - n_0},$$

where $\mathcal{S}_0^c$ and $\mathcal{S}_1^c$ are the complements of $\mathcal{S}_0$ and $\mathcal{S}_1$, respectively. The ROC (Receiver Operating Characteristics) curve is the plot of $F\hat{A}R(\lambda)$ versus $G\hat{A}R(\lambda)$ with higher ROC values indicating better recognition performance. Our experiments on the MSU-Multimodal and NIST databases were based on the following choices: $K = 10$ and $n_0/n_{gen} = n_1/n_{imp} \approx 0.8$. The fusion results are presented in Figure 7.7 and Figure 7.8. Note that there is an dramatic overall improvement of the performance indicating that the elicited joint distributions are good models for the distribution of similarity scores.

# CHAPTER 8

# SUMMARY AND CONCLUSION

Investigating dependence structures of multivariate distributions has always been an interesting area for researchers. Copulas have proved to be a useful tool for analyzing multivariate dependence structures by providing more flexibility than the classic parametric approach as they can easily separate the effect of dependence structure from that of the marginals, especially in situations that the marginals contain discrete points.

This thesis developed a semi-parametric approach to estimate the dependence structure for the bivariate distributions with mixed marginals. The semi-parametric estimator established in this thesis has been shown to be asymptotically consistent. A variance estimator has been provided as well. The estimation methodology involves integrals corresponding to the discrete componets and is therefore, non-standard. Furthermore, our approach was generalized to the higher dimensional case under similar assumptions and using the same arguments. The higher dimensional case requires more computing time.

The semi-parametric approach developed has been utilized in the $t$-copula family and the Gaussian family, which is the limiting distribution of the $t$-copula family. Estimation of the correlation matrix as well as the degrees of freedom corresponding to the $t$-copula were carried out based on the EM algorithm. This estimation was also done for the estimation of the correlation matrix corresponding to the Gaussian copula.

We have shown large sample consistency of our estimates and demonstrated this based on several simulation examples. Finally, the methodology was applied to real data consisting of matching scores from various biometric sources. Fusion based on the generalized distributions gave improved performance compared to the individual systems.

As future work, we will consider extensions to copula functions derived from general elliptical distributions.

# BIBLIOGRAPHY

# BIBLIOGRAPHY

[1] Barnard, J. and McCulloch, R. and Meng, X. L., Matas, J.G., Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage, *Statistica Sinica*, **10** (2000), 1281–1311.

[2] Bigun, E. S., Bigun, J., Duc, B., Fischer, S., Expert Conciliation for Multimodal Person Authentication Systems using Bayesian Statistics, *In: Proceedings of First International Conference on AVBPA, Crans–Montana, Switzerland* (1997), 291–300.

[3] Brunelli, R., and Falavigna, D., Person identification using multiple cues, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20** (1998), 226–239.

[4] Chen, X. and Fan, Y., Estimation of Copula-Based Semiparametric Time Series Models, *Journal of Econometrics*, **130** (2006), 307–335.

[5] Cherubini, U., Luciano, E., and Vecchiato, W., Copula Methods in Finance, *Wiley*, 2004.

[6] Clayton, D. G., A Model for association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence, *Biometrika*, **65** (1978) 141–151.

[7] Dass, S., Nandakumar, K., Jain. A. K., A principled approach to score level fusion in multimodal biometric systems, *To appear in Proceedings of AVBPA*, (2005).

[8] Demarta, S., McNeil, A. J., The $t$–copula and Related Copulas, *International Statistical Review*, **73** (2005), no. 1, 111–129.

[9] Embrechts P., Lindskog F., and McNeil A., Modelling Dependence with Copula and Applications to Risk Management, *Handbook of Heavy Tailed Distributions in Finance*, Elsevier, (2003), 329–384.

[10] Frey, R. and McNeil, A. J., Copulas and Credit Models, *RISK*, (2001), 111–114.

[11] Genest, C., Ghoudi, K., and Rivest, L. –P., A semiparametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika*, **82** (1995), no. 3, 543–552.

[12] Genest, C. and Rivest, L. –P., Statistical Inference Procedures for Bivariate Archimedean Copulas, *Journal of the American Statistical Association*, **88** (1993), 1034–1043.

[13] Genest, C., and Quessy, J. F., and Remillard, B., Goodness–of–fit procedures for copula models based on the probability integral transform, *Scandinavian Journal of Statistics*, (2006), 337–366.

[14] Griffin, P., Optimal Biometric Fusion for Identity Verification, *Identix Corporate Research Center Preprint RDNJ–03–0064*, (2004).

[15] Hougaard, P., Harvald, B., and Holm, N. V., Measuring the Similarities Between the Lifetimes of Adult Danish Twins Born Between 1881 and 1930, *J. Am. Statist. Assoc.*, **87** (1992) 17–24.

[16] Huang, W. and Dass, C. S., Generalized t-Copula and Its Application on Biometric, *JSM Proceedings*, (2007).

[17] Hutchinson, T. P. and Lai, C. D., Continuou Bivariate Distributions, *Emphasising Applications*. Adelaide: Rumsby Scientific.

[18] Jain, A. K. and Bolle, R. and Pankanti, S., BIOMETRICS: Personal Identification in Networked Society. Kluwer Academic Publishers, Boston, 1999.

[19] Jain, A. K. and Prabhakar, S. and Ross, A., Fingerprint Mathcing: Data Acquisition and Performance Evaluation. *MSU Technical Report TR99-14*, (1999).

[20] Jain, A. K., Ross, A., Prabhakar, S., An Introduction to Biometric Recognition, *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image– and Video–Based Biometrics*, **14** (2004), 4–20.

[21] Jain, A. K., Ross, A., Multibiometric Systems, *Communications of the ACM, Special Issue on Multimodal Interfaces*, **47** (2004), 34–40.

[22] Joe, H., Parametric families of multivariate distributions with given marginals, *J. Mult. Anal.*, **46**, 262–282.

[23] Kimeldorf, G., Sampson, A. R., One parameter families of bivariate distributions with fixed marginals, *Commun. Statist.*, **4**, 293–301.

[24] Kittler, J., Hatef, M., Duin, R. P., Matas, J.G., On Combining Classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(10) (1998), 955–966.

[25] Lam, L., Suen, C. Y., Optimal Combination of Pattern Classifers, *Pattern Recognition Letters*, **16** (1995), 945–954.

[26] Lancaster, H., The Structure of Bivariate Distributions, *Annals of Mathematical Statistics*, **29** (1958), 719–736.

[27] Lehmann, E., Some Concepts of Dependence, *Annals of Mathematical Statistics,* **37** (1966), 1137–1153.

[28] Lehmann, E. L., Testing Statistical Hypotheses, *Chapman & Hall*, *NY*, 1994.

[29] Maltoni, D., Maio, D., Jain, A. K. and Prabhakar, S., Handbook of Fingerprint Recognition. Springer-Verlag, 2003.

[30] Marshall, A. W. and Olkin, I., Families of Mulitvariate Distributions, *J. Am. Statist. Assoc.*, **83** (1988) 834–841.

[31] McLachlan, G., J. and Krishnan, T., The EM Algorithm and Extensions, *John Wiley & Sons*, 1997.

[32] Nandakumar, K. and Jain, A. K., Local Correlation-based Fingerprint Matching, *Proc. of Indian Conference on Computer Vision, Graphics and Image Processing, (Kolkata)*, (2004), 503–508.

[33] National Institute of Standards and Technology: NIST Biometric Scores Set. Available at http://www.itl.nist.gov/iad/894.03/biometricscores (2004).

[34] Nelsen. R., An introduction to Copulas, *Springer*, 1998.

[35] Oakes, D., Semiparametric inference in a model for association in bivariate survival data, *Biometrika*, **73** (1986), 353–361.

[36] Oakes, D., Multivariate Survival Distributions, *J. Nonparam. Statist.*, **3** (1994), 343–354.

[37] Pitt, M., and Chan, D., and Kohn, R., Efficient Bayesian inference for Gaussian copula regression models, *Biometrika*, **93** (2006), 537–554.

[38] Prabhakar, S., Jain, A. K., Decision–level Fusion in Fingerprint Verification, *Pattern Recognition*, **35** (2002), 861–874.

[39] Ratha, N. K., Chen, S., Karu, K., and Jain, A. K., A Real-time Matching System for Large Fingerprint Databases, *IEEE Transactions on PAMI*, **18** (1996), 799–813.

[40] Rüschendorf, L., Asymptotic Distribution of Multivarite Rank Order Statistics, *The Annals of Statistics*, **4** (1976), 912–923.

[41] Ruymgaart, F. H., Shorack G. R., and Zwet, W. R.,Asymptotic Normality of Nonparametric Tests for Independence, *Annals of Mathematical Statistics*, **43** (1972), 1122–1135.

[42] Ruymgaart, F. H., Asymptotic Normality of Nonparametric Tests for Independence, *Annals of Statistics*, **2** (1974), 892–910.

[43] Shao, J., Mathematical Statistics, *Springer*, 2003.

[44] Shih, J. and Louis, T., Inferences on the Association Parameter in Copula Models for Bivariate Survival Data, *Biometrics*, **51** (1995), 1384–1399.

[45] Shorack, G. R., Functions of Order Statistics, *Annals of Mathematical Statistics*, **43** (1974), 412–427.

[46] Sklar, A., Fonctions de répartition à $n$ dimensions et leurs marges, *Publ. Inst. Statist. Univ. Paris*, **8**, 229–231.

[47] Toh, K. A., Jiang, X., Yau, W. Y., Exploiting Global and Local Decisions for Multi– modal Biometrics Verification, *IEEE Transactions on Signal Processing*, **52** (2004), 3059–3072.

[48] USVISIT, U. S. Department of Homeland Security. U. S. Department of Homeland Security. Online: http://www.dhs.gov/dhspublic/display?theme=91.

[49] Wang, Y., Tan, T., Jain, A. K., Combining Face and Iris Biometrics for Identity Verification, *In: Proceedings of Fourth International Conference on AVBPA, Guildford, U.K.*, (2003) 805–813.