DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

# A NEW GOODNESS-OF-FIT APPROACH FOR ARCHIMEDEAN COPULA MODELS AND POWER ANALYSIS

by
Selim Orhun SUSAM

March, 2019
İZMİR

# A NEW GOODNESS-OF-FIT APPROACH FOR ARCHIMEDEAN COPULA MODELS AND POWER ANALYSIS

**A Thesis Submitted to the**
**Graduate School of Natural And Applied Sciences of Dokuz Eylül University**
**In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Statistics**

**by**
**Selim Orhun SUSAM**

**March, 2019**
**İZMİR**

## Ph.D. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**A NEW GOODNESS-OF-FIT APPROACH FOR ARCHIMEDEAN COPULA MODELS AND POWER ANALYSIS**" completed by **SELİM ORHUN SUSAM** under supervision of **PROF. DR. BURCU HÜDAVERDİ ÜÇER** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Doctor of Philosophy.

.............................................................

Prof. Dr. Burcu HÜDAVERDİ ÜÇER

Supervisor

.............................................................

Assoc. Prof. Dr. A. Fırat Özdemir

Thesis Committee Member

.............................................................

Asst. Prof. Dr. Celal Cem Sarıoğlu

Thesis Committee Member

.............................................................

Assoc. Prof. Dr. Sevcan DEMİR ATALAY

Examining Committee Member

.............................................................

Assoc. Prof. Dr. Femin YALÇIN

Examining Committee Member

Prof. Dr. Kadriye ERTEKİN

Director

Graduate School of Natural and Applied Sciences

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Prof. Dr. Burcu HÜDAVERDİ ÜÇER for her advice, continual presence, guidance, encouragement and endless patience during the course of this research. Besides my advisor, I would like to thank my thesis committee, Assoc. Prof. Dr. A. Fırat ÖZDEMİR and Dr. C. Cem SARIOĞLU, for their insightful comments. And I would like to thank all staff in Statistics Department of the Faculty of Science of Dokuz Eylül University.

I would also like to thank my parents and brother, whose love and guidance are with me. Most importantly, I wish to thank my loving and supportive wife, Esra.

Selim Orhun SUSAM

**A NEW GOODNESS-OF-FIT APPROACH FOR ARCHIMEDEAN COPULA MODELS AND POWER ANALYSIS**

## ABSTRACT

In this thesis, a new class of bivariate multi-parameter Archimedean copula based on Kendall distribution using Bernstein-Bézier polynomials is introduced. This new class copula has flexible dependence properties depending on the polynomial degree and the control points. Some dependence characteristics such as Kendall's tau, upper tail and lower tail dependence of the this new Archimedean copula class are derived. The simulation procedure based on these desired dependence characteristics is presented.

Also, we propose an estimation method for the Archimedean family of copula in a nonparametric setting. Bernstein polynomials and Bézier curve approaches are used to estimate the Kendall distribution function of the Archimedean copula. Also, a new goodness-of-fit test based on Cramér-Von Mises type statistic is constructed using new estimation methods of Kendall distribution function. A Monte Carlo study is performed to measure the performance of the proposed tests. The simulation results show that both the Bernstein polynomial and the Bézier curve based tests have better performances than the classical one since they have flexible form according to its order $m$.

**Keywords:** Kendall distribution function, Bernstein polynomial, Bézier curve, Cramér-von-Mises statistic, Archimedean copula

**ARCHIMEDEAN KOPULA MODELLERİ İÇİN YENİ BİR UYUM İYİLİĞİ YAKLAŞIMI VE GÜÇ ANALİZİ**

# ÖZ

Bu tezde, Bernstein-Bézier polinomları kullanılarak Kendall dağılımına dayalı yeni bir iki değişkenli Archimedean kopula sınıfı tanıtıldı. Bu yeni sınıf kopula, polinom derecesine ve kontrol noktalarına bağlı olarak esnek bağımlılık özelliklerine sahiptir. Bu yeni Archimedean kopula sınıfının Kendall tau, üst kuyruk ve alt kuyruk bağımlılığı gibi bazı bağımlılık özellikleri elde edilmiştir. İstenen bağımlılık özelliklerine dayanan benzetim prosedürü tanıtılmıştır.

Ayrıca, Archimedean ailesi için parametrik olmayan tahmin yöntemleri önerilmektedir. Archimedean Kopula'nın Kendall dağılım fonksiyonunu tahmin etmek için Bernstein polinomları ve Bézier eğrisi yaklaşımı kullanılmıştır. Ayrıca, Cramér-Von Mises türü istatistiğe dayalı yeni bir uyum iyiliği testi, Kendall dağılım fonksiyonunun yeni tahmin yöntemleri kullanılarak elde edilmiştir. Önerilen testlerin performansını ölçmek için Monte Carlo benzetim çalışması gerçekleştirilmiştir. Benzetim sonuçları, hem Bernstein polinomu hem de Bézier eğrisi temelli testlerin, "$m$" derecesine dayalı olarak esnek bir forma sahip olmasından dolayı klasik olandan daha iyi bir performansa sahip olduğunu göstermektedir.

**Anahtar kelimeler:** Kendall dağılım fonksiyonu, Bernstein polinomu, Bézier eğrisi, Cramér-von-Mises istatistiği, Archimedean kopula

**CONTENTS**

## LIST OF FIGURES

## LIST OF TABLES

**Page**

# CHAPTER ONE
# INTRODUCTION

In order to understand the relationship between the variables, it is necessary to explain the dependence structure of the variables. The dependence structure has been investigated for many years in statistical theory and many methods have been developed to reveal this dependence structure. One of them is the function of Copula. The most important benefit of the copula functions is the distinction between the dependence structure and the univariate distributions of variables. This property allows copula functions to model the dependence structure through the distribution functions. Copula modeling has been also used in many applicational areas, especially finance, insurance, survival analysis, enviroment engineering, medicine, etc. For further information, see Joe (1997) and Nelsen (2006).

First time in Copula literature, in 1940, Hoéffding carried out several studies on multivariate distributions, characteristics of these distributions and dependence measures. In 1959, Sklar demonstrated the presence of Copulas. Nelsen (2006) described the Copula functions as follows: “Copulas are functions that join or “couple” multivariate distribution functions to their one dimensional marginal distribution functions. Alternatively, copulas are multivariate distribution functions whose one-dimensional margins are uniform on the interval $[0, 1]$.” Fisher (1997) answered the questions of “Why are copulas of interest to students of probability and statistics?” in the Encyclopedia of Statistical Sciences. He stated that “Copulas are of interest to statisticians for two main reasons: Firstly, as a way of studying scale-free measures of dependence; and secondly, as a starting point for constructing families of bivariate distributions, sometimes with a view to simulation.”

The flexibility of copula models allows us to capture a complex joint distribution simply by changing the specification of the copula model and the marginal distributions. Because of this flexibility, copulas have gained popularity in many of the applicational areas. Plackett (1965) proposed one of the earliest family of dependence functions. Clayton (1978), Cook & Johnson (1981), Oakes (1982), Oakes

(1989), Tawn (1988) and some others proposed and studied the other type of family of copulas. Genest & MacKay (1986) and Nelsen (2006) studied on Archimedean copula which is wide and mathematically tractable and this class is more attractive for the statistical treatment of data. In this thesis, we firstly propose constructing a multi-parameter Archimedean copula using Kendall distribution function. We use Bernstein-Bézier polynomials to create the new Archimedean class. This new Archimedean copula family is contributed to the expansion of the existed Archimedean copula family which is limited in the literature.

Inference problem about copulas gains a considerable afford recently. Some parametric and nonparametric estimation approaches have been proposed for copula models depending on which the copula and the marginals are assumed to be known. When the true copula does not belong to an assumed parametric family, nonparametric estimation leads to a good fit to data. Also, parametric methods as maximum likelihood leads to a good fit when the true copula belongs to a parametric family. Deheuvels (1978), Genest et al. (1995), Joe & Xu (1996), Joe (1997), Fermanian et al. (2004) studied on the estimation methods and gave some comparisons among them.

Genest & Rivest (1993) presented statistical inference procedure for bivariate Archimedean copulas. Lambert (2007) used a penalized smoothing spline in Archimedean copula parameter estimation via an MCMC algorithm with Bayesian approach. Dimitrova et al. (2008) obtained geometrically designed (GED) spline estimator for Kendall distribution function, $K(t)$ through multivariate Archimedean copula estimation process. Susam & Ucer (2018) defined an estimator of $K(t)$ using Bernstein polynomial approximation. In this thesis, we secondly propose an estimator for Kendall distribution function using Bernstein and Bézier polynomial approximation to estimate smooth curves. For the Archimedean family of copulas, the difficulties in estimating the generator function motivate us for some alternative approaches.

The study of goodness-of-fit tests (GoF) for copulas arise recently as an inferential problem. Several GoF approaches for copulas have been proposed in the literature to see which copula family is convenient. Genest & Rivest (1993) proposed an empirical method to choose the best Archimedean copula. First work is conducted by Wang & Wells (2000) whose foundations are based on the study proposed by Genest & Rivest (1993). Later, Genest et al. (2006) proposed some alternative tests for the Archimedean copulas and these methods are based on probability integral transformation. This transformation is also called as Kendall transformation. Berg (2009) examined several copula goodness-of-fit approaches. Fermanian et al. (2004) defined the distribution-free goodness-of-fit test statistics for copulas and investigated their asymptotic distributions. Diebold et al. (1999), Berkowitz (2001) and Breymann et al. (2003) used the probability integral transform (PIT) of Rosenblatt (1952) in estimation of density models. Susam & Ucer (2018) constructed independence test for Archimedean copula using Bernstein estimation of Kendall distribution function. In this thesis, as a third contribution, we construct a goodness of fit test based on Bernstein and Bernstein-Bézier Empirical Kendall distribution function to choose which Archimedean copula is better.

This thesis is organized as follows. The copula function and its properties are given in Chapter 2. The general concept of Archimedean copula is given in Chapter 3. Kendall distribution function in terms of Archimedean copula is also given. In Chapter 4, Bernstein-Bézier type Archimedean copula is presented and some dependence characteristics are investigated. A simulation procedure of this new class for different polynomial degrees is given in this section. Also, In Chapter 4, we propose an estimator for Kendall distribution function using Bernstein polynomial and Bernstein-Bézier curve approximation to estimate smooth curves. We investigate its estimation performance by a simulation study. In Chapter 5, we construct a Goodness-of-fit test based on Bernstein polynomial and Bézier curve estimation of Kendall distribution for Archimedean copula. Also, the power and the size of the new test statistics are given and compared with the empirical estimation of Kendall distribution function. And the last chapter is devoted to the conclusion.

# CHAPTER TWO
# BIVARIATE COPULA FUNCTIONS

## 2.1 Bivariate Copula Definitions

Copula is used to model the dependence structure between random variables. Sklar (1959) explained the relationship between the marginal distribution functions and the joint distribution function in terms of copula.

**Theorem 2.1** (Sklar (1959))**.** For bivariate random variables $X$ and $Y$, let $H$ be a bivariate joint distribution function with marginals $F$ and $G$, respectively. Then, there exists a Copula $C$ in such a way that

$$H(x, y) = C(F(x), G(y)) \tag{2.1}$$

for all $(x, y) \in R$.

If the margins are continuous, then $C$ is unique; otherwise $C$ is uniquely determined on Range$(F)$ $\times$ Range$(G)$. When the bivariate joint distribution function $H$ and the marginals $F$ and $G$ are known, copula can be constructed as $C(u, v) = H(F^{-1}(u), G^{-1}(v))$ where $U = F(x)$ and $V = G(y)$.

**Definition 2.1.** A bivariate copula $C : [0, 1]^2 \rightarrow [0, 1]$ is a bivariate distribution function with the following properties:

1. $C(u, 0) = C(0, v) = 0$
2. $C(u, 1) = u$ and $C(1, v) = v$
3. $C$ is two increasing. For every rectangle $B : [u_1, u_2] \times [v_1, v_2]$ in $\boldsymbol{I}$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$. The C-Volume of B is non-negative, i.e.

$$V_c(B) = C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

Property 1 is referred to as grounded property of a copula. Property 2 says that if one variable is known with marginal probabilty 1 then joint probabilty of two outcomes is identical with the probability of the remaining uncertain outcome. That is, univariate margins $F(x)$ and $G(y)$ of $C(F(x), G(y)$ are obtained by letting $F(x) \to 1$ and $G(y) \to 1$, respectively. Property 3 is the two-dimensional analogue of a nondecreasing one-dimensional function. Properties 1 and 3 are general properties of bivariate distribution functions.

## 2.2 Properties of Copulas

### *2.2.1 Fréchet-Hoeffding Bounds Inequality*

**Theorem 2.2** (Nelsen (2006))**.** Let $C$ be a copula. Then for every $u, v$ in $[0, 1]^2$ then

$$\max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v)$$

The bounds in Theorem 2.2 are themselves copula and are denoted by $M(u, v) = \max(u + v - 1, 0)$ and $W(u, v) = \min(u, v)$. $M$ is referred as the Fréchet-Hoeffding upper bound which indicates the perfect positive dependence and $W$ is referred as the Fréchet-Hoeffding lower bound which indicates the perfect negative dependence. Third important copula is the product copula $\Pi(u, v) = uv$ which indicates the independent case. From Theorem 2.2, any copula lies between the Fréchet-Hoeffding bounds. In Figure 2.1, copulas $M, W$ and $\Pi$ are presented.

### *2.2.2 Invariance*

Most of the nonparametric statistics for copula are derived from the property of strictly monotone transformations of the random variables. If $\alpha$ is a strictly monotone function then the distribution of random variable $\alpha(X)$ have continuous distribution

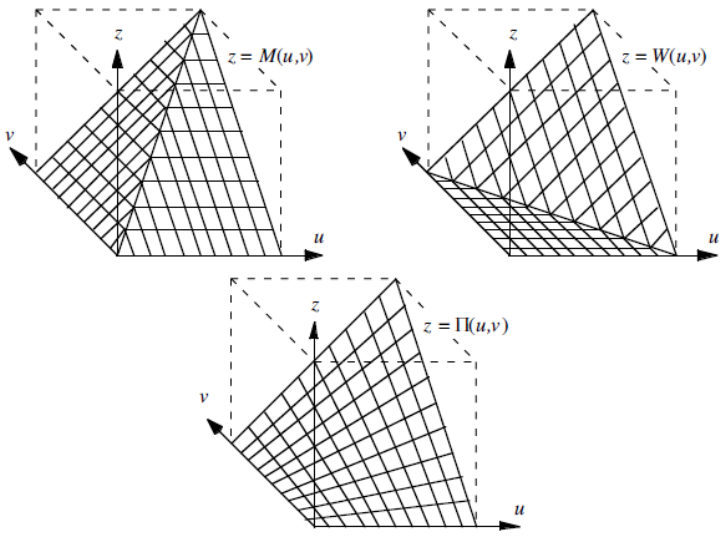



Figure 2.1 $M, W$ and $\Pi$ copulas

function.

**Theorem 2.3** (Nelsen (2006))**.** Let $X$ and $Y$ be a random variable with copula $C_{X,Y}$. If $\alpha$ and $\beta$ are stricty increasing then $C_{\alpha(X),\beta(Y)} = C_{X,Y}$. Thus $C_{X,Y}$ is invariant under strictly increasing transformations of $X$ and $Y$.

### *2.2.3 Symmetry*

Consider continuous random variables $X$ and $Y$ having copula $C$, joint distribution function $H(x, y)$ and marginals $F(x)$ and $G(y)$, respectively. $X$ and $Y$ are exchangeable if and only if for all $(u, v)$ in $[0, 1]^2$, $F = G$ and $C(u, v) = C(v, u)$. When $C(u, v) = C(v, u)$, then $C$ is symmetric.

### *2.2.4 Continuity*

Let $C$ be a copula of random variables $X$ and $Y$. Then for every $u_1, u_2, v_1$ and $v_2$ in domain $C$,

$$|C(u_2, v_2) - C(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1| \ ,$$

hence, $C$ is uniformly continuous on its domain.

### 2.3 Dependence Measure of Copulas

In this section, we restrict the discussion with the bivariate case although generalization to higher dimensions is possible.

Let $X$ and $Y$ be random variables with marginal distribution functions $F$, $G$ and joint distribution function $H$, then $X$ and $Y$ are dependent , if $H(x, y) \neq F(x) \times G(y)$. Let $\delta(X, Y)$ denote the measure of dependence. Embrechts et al. (2002) listed four desirable properties of this measure:

1. $\delta(X, Y) = \delta(Y, X)$ (symmetry)
2. $-1 \leq \delta(X, Y) \leq -1$ (normalization)
3. $\delta(X, Y) = 1 \Longleftrightarrow X$ and $Y$ are comonotonic ; $\delta(X, Y) = -1 \Longleftrightarrow X$ and $Y$ are countermonotonic;
4. For a strictly monotonic transformation $T : R \rightarrow R$ of $X$, if $T$ is increasing then $\delta(T(X), Y) = \delta(Y, X)$ and if $T$ is decreasing then $\delta(T(X), Y) = -\delta(Y, X)$

Dependence can be measured using several alternative concepts: linear correlation, concordance (rank correlation), tail dependence.

#### *2.3.1 Correlation and Dependence*

The most preferred dependence measure is the correlation coefficient between the random variables $X$ and $Y$ given by

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

where $\text{Cov}(X, Y) = E(X, Y) - E(X)E(Y)$ ; $\sigma_X, \sigma_Y > 0$. $\sigma_X$ and $\sigma_X$ denote the standard deviations of $X$ and $Y$, respectively. It is known that: (a) $\rho(X, Y)$ is a measure of linear dependence, (b) $\rho(X, Y)$ is symmetric, (c) that the lower and upper bounds on the inequality $-1 < \rho(X, Y) < 1$ measures perfect negative linear and positive linear dependence (a property referred to as normalization), (d) it is invariant with respect to linear transformations of the variables.

Trivedi & Zimmer (2007) explained the weakness of correlation as: “In general zero correlation does not imply independence. For example, if $X \sim N(0, 1)$ and $Y = X^2$ then $Cov(X, Y) = 0$ but $X$ and $Y$ are clearly dependent. Zero correlation only requires $\text{Cov}(X, Y) = 0$, whereas zero dependence requires $\text{Cov}(\phi_1(X), \phi_2(Y)) = 0$ for any functions $\phi_1$ and $\phi_2$. This represents the weakness of correlation as a measure of dependence. The second limitation of correlation is that it is not defined for some heavy-tailed distributions whose second moments do not exist. The third limitation of the correlation measure is that it is not invariant under strictly increasing nonlinear transformations. Given these limitations, alternative measures of dependence should be considered such as rank correlation (concordance).”

### *2.3.2 Concordance Measure*

Kendall’s tau ($\tau$) and Spearman’s rho ($\rho$) that are defined in terms of copulas are known as measures of associations. Before any discussion about these dependence measures, it is better to define the term concordance.

**Theorem 2.4** (Nelsen (2006))**.** Let $(X, Y)$ be a pair of continuous random variables with joint distribution function $H(X, Y)$, and let $(x_1, y_1)$ and $(x_2, y_2)$ be independent observations from $(X, Y)$. These observations are concordant if $(x_1 - x_2)(y_1 - y_2) > 0$ and discordant if $(x_1 - x_2)(y_1 - y_2) < 0$.

Nelsen (2006) informally defined concordance as: “a pair of random variables are concordant if large values of one tend to be associated with large values of the other and small values of one with small values of the other.”

**Theorem 2.5** (Nelsen (2006))**.** Let $(X_1, Y_1)$ and $(X_2, Y_2)$ be two pairs of independent continuous random variables with a common joint distribution function $H(x, y)$ and copula $C(u,v)$, and let $\tau$ be equal to the probability of concordance minus the probability of discordance:

$$\tau = P((X_1{-}X_2)(Y_1{-}Y_2) > 0) - P((X_1{-}X_2)(Y_1{-}Y_2) < 0)$$

then the Kendall's tau measure in terms of copula is given by

$$\tau = 4\int_0^1\int_0^1 C(u,v)dC(u,v) - 1\ .$$

**Theorem 2.6** (Nelsen (2006))**.** Let $(X_1,Y_1),(X_2,Y_2)$ and $(X_3,Y_3)$ be three pairs of independent continuous random variables with a common joint distribution function $H(x,y)$ and copula $C(u,v)$, and let $\rho$ be equal to the probability of concordance minus the probability of discordance for the pairs $(X_1,Y_1)$ and $(X_2,Y_3)$:

$$\rho = 3P((X_1 - X_2)(Y_1 - Y_3) > 0) - P((X_1 - X_2)(Y_1 - Y_3) < 0)$$

then the Spearman's rho measure in terms of copula is given by

$$\rho = 12\int_0^1\int_0^1 C(u,v)dudv - 3\ .$$

### *2.3.3 Tail Dependence*

Concordance measures are designed to describe how large (or small) values of one random variable appear with large (or small) values of the other. The tail dependence relates to the level of dependence in the upper-right quadrant tail or lower-left quadrant tail of a bivariate distribution. Copula method is used to model the tail dependence structure. Right tail dependence is the probability that random variable $X$ reaches the large values given that random variable $Y$ attains the large values. And, left tail dependence is the probability that random variable $X$ reaches the small values given that random variable $Y$ attains the small values.

**Theorem 2.7** (Nelsen (2006))**.** Let $X$ and $Y$ be continuous random variables with distribution functions $F$ and $G$, respectively. The upper tail dependence parameter $\lambda_U$ is the limit (if it exists) of the conditional probability that:

$$\lambda_U = \lim_{t \to 1^-} P(Y > G^{-1}(t) \mid X > F^{-1}(t))$$

Similarly, the lower tail dependence parameter $\lambda_L$ is the limit (if it exists) of the conditional probability that:

$$\lambda_L = \lim_{t \to 0^+} P(Y \leq G^{-1}(t) \mid X \leq F^{-1}(t))$$

where $F^{-1}$ and $G^{-1}$ denote the generalized inverse distribution functions of $X$ and $Y$, respectively. Then the right tail and the left tail dependence measures in terms of copula are given by

$$\lambda_U = \lim_{t \to 1^-} \frac{1 + 2t - C(t,t)}{1 - t} \;\; , \;\; \lambda_L = \lim_{t \to 0^+} \frac{C(t,t)}{t}$$

respectively. See, Nelsen (2006).

# CHAPTER THREE
# BIVARIATE ARCHIMEDEAN COPULA

In this chapter, we concentrate on Archimedean family of copula. The main feature that separates this class from the others is that it has a generator function $\varphi$ which is used to construct an Archimedean copula. This generator function and its pseudo inverse function are defined in Nelsen (2006).

## 3.1 Generator Function of Archimedean Copulas

**Theorem 3.1** (Nelsen (2006))**.** A generator function $\varphi$ is a continuous, strictly decreasing convex function defined from **I** to $[0, \infty)$ such that $\varphi(1) = 0$. If $\varphi(0) = \infty$, then the generator is called strict. The pseudo inverse of $\varphi$ is the function $\varphi^{[-1]}$, defined on $[0, \infty)$ to **I** is given by

$$\varphi^{[-1]} = \begin{cases} \varphi^{-1}(t) & 0 \leq t \leq \varphi(0) \\ 0 & \varphi(0) \leq t < \infty \end{cases}$$

Then the Copula in terms of generator function is given by the next definition.

**Theorem 3.2** (Nelsen (2006))**.** Let $\varphi$ be a continuous, strictly decreasing function from $I$ to $[0, \infty]$ such that $\varphi(1) = 0$, and let $\varphi^{[-1]}$ be the pseudo-inverse of $\varphi$ defined by Definition 3.1. Let $C$ be the function from $I^2$ to $I$ given by

$$C(u, v) = \varphi^{[-1]}\{\varphi(u) + \varphi(v)\}. \tag{3.1}$$

Then, $C$ satisfies the boundary conditions in definition 2.1

**Example 3.1.** Let $\varphi(t) = -\ln(t)$ for $t \in [0, 1]$. Because $\varphi(0) = \infty$ then $\varphi(t)$ is strict. Thus $\varphi^{[-1]}(t) = \varphi^{-1}(t) = \exp(-t)$. Then

$$C(u, v) = \exp\Big(-(-\ln(u) - \ln(v))\Big) = uv.$$

$C(u, v)$ is known as independent Archimedean copula $\Pi(u, v)$.

**Example 3.2.** Let $\varphi(t) = 1 - t$ . Then $\varphi^{[-1]}(t) = 1 - t$ for $t \in [0, 1]$ and $0$ for $t > 1$. So $\varphi^{[-1]}(t) = \max(1 - t, 0)$. Then

$$C(u, v) = \max(u + v - 1, 0) = W(u, v)$$

Then $C(u, v)$ is known as Fréchet-Hoeffding lower bound.

Some algebric properties of Archimedean copulas are given in next theorem.

**Theorem 3.3** (Nelsen (2006))**.** Let $C$ be an Archimedean copula with generator $\varphi$.

1. $C$ is symmetric; that is $C(u, v) = C(v, u)$
2. $C$ is associative; that is $C(C(u, v), w) = C(u, C(v, w))$
3. If $c > 0$ is any constant, then $c\varphi$ is also a generator of $C$.

## 3.2 Some Archimedean Copulas

Some well-known Archimedean copula functions are proposed by Clayton (1978), Frank (1979), Gumbel (1960). We give some general theory about these Archimedean copulas.

### *3.2.1 Clayton copula*

The Clayton (1978) copula, also referred to as the Cook & Johnson (1981) copula, has been originally investigated by Kimeldorf & Sampson (1975), is defined as takes the form:

$$C(u, v) = (u^{-\theta} + v^{-\theta} + 1)^{-\frac{1}{\theta}} \; ; \; u, v \in [0, 1]^2 \, , \theta > 0$$

with dependence parameter $\theta \in [0, \infty]$. If $\lim_{\theta \to \infty} C(u, v) = \min(u, v)$ then the dependence structure approaches its maximum when $\theta \to \infty$. On the other hand,

$\lim_{\theta \to 0} C(u,v) = uv$; independence is obtained when $\theta \to 0$. The Clayton copula exhibits strong left tail dependence.

### *3.2.2 Frank Copula*

The Frank (1979) copula takes the form:

$$C(u,v) = -\frac{1}{\theta}\log\left(\frac{1+(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}\right) \; ; \; u,v \in [0,1]^2, \theta \in (-\infty,+\infty)/(0)$$

The dependence parameter $\theta \in (-\infty,+\infty)/(0)$. If $\lim_{\theta \to -\infty} C(u,v) = \max(u+v-1,0)$ so Frank copula attains to Fréchet lower bound, $\lim_{\theta \to \infty} C(u,v) = \min(u,v)$ then Frank copula attains to Fréchet upper bound, On the other hand, if $\lim_{\theta \to 0} C(u,v) = uv$; independence is obtained. Frank copula exhibits symmetric and weak tail dependence in both tails.

### *3.2.3 Gumbel Copula*

The Gumbel (1960) copula has the form:

$$C(u,v) = e^{-(-\log(u)^{\theta}-(\log(v))^{\theta})^{\frac{1}{\theta}}} \; ; \; u,v \in [0,1]^2, \theta \in (1,+\infty)$$

The dependence parameter $\theta \in [1,\infty)$. If $\lim_{\theta \to 1} C(u,v) = uv$; then independence is obtained. On the other hand, if $\lim_{\theta \to \infty} C(u,v) = \min(u,v)$ then Frank copula attains to Fréchet upper bound. Gumbel copula cannot model negative dependence, but it contrasts to Clayton, Gumbel has strong right tail dependence.

## 3.3 Kendall Distribution Function and Related Dependence Measure

The most important feature for Archimedean copulas is that its multivariate distribution function can be reduced to univariate distribution function through

Archimedean representation. This feature was investigated by Genest & Rivest (1993). This method is also known as dimension reduction method or Kendall transform method. Genest & Rivest (1993) have given the link between the function $\varphi(t)$ and $K(t)$.

**Theorem 3.4** (Nelsen (2006))**.** Let $C$ be an Archimedean copula wtih generator function $\varphi$. Let $K_C(t)$ denote $C$ measure of set $\Big\{(u,v) \in I^2 | C(u,v) \leq t\Big\}$, or equivalently, of the set $\Big\{(u,v) \in I^2 | \varphi(u) + \varphi(v) \geq \varphi(t)\Big\}$. Then for any $t$ in $[0,1]$

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t)}. \tag{3.2}$$

The Kendall distribution function $K(t)$ has some properties. The following properties were summarized in Nelsen (2006).

**Properties 3.1.** The Kendall distribution function $K(t)$ has properties:

1. $K(0) = 0$
2. $K(1) = 1$
3. $K(t) > t, t \in (0,1)$
4. $K'(t) > 0 \, , t \in (0,1)$

The generator function $\varphi(t)$ can be obtained as

$$\varphi(t) = \varphi(t_0) \exp\left[\int_{t_0}^{t} \frac{1}{\lambda(z)} dz\right], \tag{3.3}$$

where $0 < t_0 < 1$ is a constant. $K(t)$ is an important factor for identifying the function $\varphi(t)$. That means, $K(t)$ determines the dependence structure of Archimedean copula family. For a bivariate Archimedean copula with Kendall distribution function $K(t)$, Genest & MacKay (1986) defined Kendall's Tau $(\tau)$ as given in the next theorem.

**Theorem 3.5** (Genest & MacKay (1986))**.** Let $X$ and $Y$ be random variables with an Archimedean copula $C$ with generator function $\varphi(t)$. Then Kendall's tau for $X$ and $Y$

Table 3.1 Archimedean Copulas with Generator functions $\varphi(t)$

| Copula | $\varphi(t)$ | $K(t)$ | Range of $\theta$ |
|---|---|---|---|
| Clayton | $(t^{-\theta}-1)/\theta$ | $t-\frac{t(1-t^{\theta})}{\theta}$ | $(-1,\infty)-\{0\}$ |
| Frank | $-\log(\frac{\exp(t\theta)-1}{\exp(\theta)-1})$ | $t-\frac{(\exp(t\theta)-1)\log(\frac{\exp(-t\theta)-1}{\exp(-\theta)-1})}{\theta}$ | $(-\infty,\infty)-\{0\}$ |
| Gumbel | $(-\log(t))^{\theta}$ | $t-\frac{t\log(t)}{\theta}$ | $[1,\infty)$ |
| Independence | $-\log(t)$ | $t-t\log(t)$ | - |

Table 3.2 Kendall's Tau $(\tau)$, Lower $\lambda_L$ and Upper $\lambda_U$ tail dependences some Archimedean copulas

| Copula | $\tau(\theta)$ | $\lambda_L$ | $\lambda_U$ |
|---|---|---|---|
| Clayton | $\frac{\theta}{\theta+2}$ | $2^{-\frac{1}{\theta}}$ | 0 |
| Frank | $1+4\theta^{-1}({D_1}^*(\theta)-1)$ | 0 | 0 |
| Gumbel | $\frac{\theta-1}{\theta}$ | 0 | $2-2^{\frac{1}{\theta}}$ |

$^*D_1(x)=x^{-1}\int_0^x t(\exp(t)-1)^{-1}dt$

is given by;

$$\tau=4E(C(u,v))-1=1+4\int_0^1\frac{\varphi(t)}{\varphi'(t)}dt=3-4\int_0^1 K(t)dt. \tag{3.4}$$

And also, lower and upper tail dependence coefficients for Archimedean copulas are given in the next theorem.

**Theorem 3.6** (Michiels et al. (2011))**.** Let $X$ and $Y$ be random variables with an Archimedean copula $C$ with generator function $\varphi(t)$. Then lower $\lambda_L$ and upper $\lambda_U$ tail dependence for $X$ and $Y$ are given by,

$$\lambda_L=2^{\lim_{t\to 0^+}\left(t-K(t)\right)'},\lambda_U=2-2^{\lim_{t\to 1^-}\left(t-K(t)\right)'} \tag{3.5}$$

Some well-known Archimedean copula functions were proposed by Clayton (1978), Frank (1979), Gumbel (1960). These three Archimedean copulas with generator function $\varphi(t)$, Kendall distribution function $K(t)$ are summarized in Table 3.1. And also, Kendall's Tau $(\tau)$, Lower $\lambda_L$ and Upper $\lambda_U$ tail dependence for Archimedean copula are summarized in Table 3.2.

# CHAPTER FOUR
# BERNSTEIN POLYNOMIALS IN ARCHIMEDIAN COPULA

In this chapter, a new class of bivariate multi-parameter Archimedean copula based on Kendall distribution using Bernstein-Bézier polynomials is introduced. The new class copula has flexible dependence properties depending on the polynomial degree and the control points. Some dependence characteristics such as Kendall's tau, upper tail and lower tail dependence of the new Archimedean copula class are derived. The simulation procedure based on these desired dependence characteristics is presented. Also, we propose an estimator for Kendall distribution function using Bernstein and Bézier curve approximation to estimate smooth curves. We investigate the estimation performance by a simulation study. More parameterized generators allow for better flexibility while modelling the copula, however it should be avoided too much of a bias. A possible way to increase the flexibility of the generator $\varphi$, so of the copula $C$ is using the spline functions. For the Archimedean family of copulas, the difficulties in estimating the generator function motivates us for some alternative approaches.

Before any discussion about these topics, we give a brief information about Bernstein polynomials.

## 4.1 Bernstein Polynomials

In many applications of Mathematics and Statistics, there are much more complex functions that are difficult to examine. One of the most important studies of approximation theory is to obtain a representation of these functions in terms of other functions which have better properties. Let $f$ be a continuous function on the support $[a, b]$ and $B$ be any polynomial, Weierstrass (1885) showed that

$$|B(x) - f(x)| < \epsilon$$

for all $\epsilon > 0$ and $x \in [a, b]$. Bernstein (1912) introduced a polynomial $B$ which converges the function $f$ on the support $[0, 1]$. These polynomials are defined as

$$B_m(f, x) = \sum_{k=0}^{m} f(\frac{k}{m}) P_{k,m}(x),$$

where $P_{k,m}(x) = x^k(1-x)^{m-k}\binom{m}{k}$ are Bernstein basis and $m$ is the degree of polynomial. Also we note that $P_{k,m}(x) \geq 0$. Bernstein polynomial has some useful properties. For example, it is clear that $B_m(f, 0) = f(0)$ and $B_m(f, 1) = f(1)$ so that a Bernstein polynomials for $f$ interpolates $f$ at both endpoints of the interval $[0, 1]$. This property is called as end-points rule. Another important property of Bernstein polynomials is shape preserving property. We established non-decreasing property for the Bernstein polynomials. First derivative of the Bernstein polynomials is given as

$$B_m^{'}(f, x) = m \sum_{k=0}^{m-1} \big(f(\frac{k+1}{m}) - f(\frac{k}{m})\big) P_{k,m-1}(x).$$

If $f$ is non-decreasing such that $f(\frac{k+1}{m}) - f(\frac{k}{m}) \geq 0$ then $B_m^{'}(f, x)$ is also non-decreasing because $B_m^{'}(f, x) \geq 0$. This property is also true for increasing, convex or concave function $f$. The third important property of Bernstein polynomial is linear precision property such that $B_m(x, x) = x$.

Bernstein polynomials have been frequently used in various areas of Statistics. First time in the literature, Bernstein polynomials are used by Vitale (1975) for smooth density function estimation. Then, Kim (1996) used Bézier curve for density estimation. Kim et al. (1999) compared performances of Bézier curve density estimation for kernel method. Recently, Babu et al. (2002) used the advantages of Bernstein polynomial approximation to estimate the underlying distribution function. They showed that Bernstein distribution function is uniformly strongly consistent estimator. And also, Leblanc (2009) proved that Bernstein polynomial distribution function has Chung-Simirnov property. Leblanc (2012) introduced a higher order expansion for the asymptotic integrated mean-squared error of Bernstein estimators of distribution functions. They also established the (pointwise) asymptotic normality of Bernstein distribution function. Erdoğan et al. (2019) proposed general method for non-parametric distribution function estimation using the rational Bernstein polynomials. They also investigated theoretical properties of the new estimator. The estimation of the underlying distribution function of population is one of the most interested problem in statistical inference. When $F$ is known to be continuous distribution function, one can consider the estimation of $F$ by using smooth functions rather then the empirical distribution function. Also, we can define a distribution function using properties of the Bernstein polynomials easily.

## 4.2 Bernstein-Bézier type Bivariate Archimedean Copula

In this section, a new Archimedean copula class based on Bernstein-Bézier polynomial is proposed. The proposed class has flexible dependence structure. Because by changing the polynomial degree and the coefficients, different values of Kendall's tau (negative or positive), lower and upper tail dependence coefficients can be obtained. It is possible to create a new distribution function which has desirable dependence characteristics. This is quite useful in power analysis of goodness-of-fit test statistic. Also, an algorithm is proposed to create different distributions with the same dependence level by changing the control points for polynomial degree greater than four. This case is investigated in sub-section 4.2.4.

Michiels et al. (2011) investigated a general method for constructing bivariate Archimedean copula families using $\lambda$ function. They worked with polynomials to construct multi-parameter copula families. Cooray (2018) introduced two-parameter strict Archimedean generator function based on Clayton copula. Najjari et al. (2014) constructed a new generator function $\varphi(t)$ using hyperbolic functions as generators of Archimedean copulas. The majority of the papers proposed some methods based on generator function $\varphi$ for constructing a new Archimedean family of copulas. In this chapter, we propose creating a multi-parameter Archimedean copula using Kendall distribution function $K(t)$. We use Bernstein-Bézier polynomials to create the new

Archimedean class. Kendall's tau, lower and upper tail dependence coefficients are also obtained according to the polynomial degree and the control points. This new Archimedean copula family is contributed to the expansion of the existed Archimedean copula family which is limited in the literature.

A Kendall distribution function, $K(t)$ should satisfy the properties 3.1 (1-4) described in Section 3. The properties of the distribution function $K(t)$ follow the properties of the generator function $\varphi(t)$ for $t \in [0,1]$.

Let $K(m, \alpha; t)$ be a Bernstein-Bézier type Kendall distribution function with polynomial degree $m$ and control points $\alpha$:

$$K(m, \alpha; t) = \sum_{k=0}^{m} \alpha_k B_{k,m}(t) \; , \tag{4.1}$$

where $B_{k,m}(t) = \binom{m}{k} t^k (1-t)^{m-k}$ for $t \in [0,1]$. $K(m, \alpha; t)$ satisfies the following Lemma.

**Lemma 4.1.** A Bernstein-Bézier type Kendall distribution function $K(m, \alpha; t)$ satisfies the properties (1-4) if the following constraints hold:

1. $\alpha_0 = 0 < \alpha_1 < \alpha_2 < \ldots < \alpha_m = 1$
2. $\alpha_k > \frac{k}{m}$, $k = 1, \ldots, m-1$

*Proof:*

$K(m, \alpha, t = 0) = \sum_{k=0}^{m} \alpha_k P_{k,m}(t = 0) = 0$ holds since $\alpha_0 = 0$. Similarly, $K(m, \alpha, t = 1) = \sum_{k=0}^{m} \alpha_k B_{k,m}(t = 1) = 1$ holds since $\alpha_m = 1$ .

Also, $K(m, \alpha, t)' = m \sum_{k=0}^{m-1} (\alpha_{k+1} - \alpha_k) P_{k,m-1}(t) > 0$ holds since $\alpha_0 = 0 < \alpha_1 < \alpha_2 < \ldots < \alpha_m = 1$. See, Duncan (2005).

If the Bézier control points $\alpha_k > \frac{k}{m}, k = 1, \dots, m-1$ where $\alpha_k = k/m + \epsilon_k$, then,

$$
\begin{aligned}
K(m, \alpha, t) &= \sum_{k=0}^{m} \alpha_k \binom{m}{k} t^k (1-t)^{m-k} \\
&= \sum_{k=0}^{m} (\frac{k}{m} + \epsilon_k) \binom{m}{k} t^k (1-t)^{m-k} \\
&= \sum_{k=0}^{m} (\frac{k}{m}) \binom{m}{k} t^k (1-t)^{m-k} + \sum_{k=0}^{m} (\epsilon_k) \binom{m}{k} t^k (1-t)^{m-k} \\
&= t \sum_{k=1}^{m} \binom{m-1}{k-1} t^{k-1} (1-t)^{m-k} + \sum_{k=0}^{m} (\epsilon_k) \binom{m}{k} t^k (1-t)^{m-k} \\
&= t \sum_{p=0}^{m-1} t^p (1-t)^{m-p-1} \binom{m-1}{p} + \sum_{k=0}^{m} (\epsilon_k) \binom{m}{k} t^k (1-t)^{m-k} \\
&= t + \sum_{k=0}^{m} (\epsilon_k) \binom{m}{k} t^k (1-t)^{m-k} > t.
\end{aligned}
$$

We also obtain the Kendall's tau, the lower and the upper tail coefficients of the Bernstein-Bézier type Archimedean copula class using the following lemmas.

**Lemma 4.2.** Kendall's tau for Bernstein-Bézier type Archimedean copula is obtained as

$$\tau = 3 - 4 \sum_{k=0}^{m} \alpha_k \binom{m}{k} \beta(k+1, m-k+1)$$

where $\beta(.,.)$ is the beta function defined as $\beta(v_1, v_2) = \int_0^1 t^{v_1 - 1}(1-t)^{v_2 - 1} dt$ for $v_1, v_2$ positive integers.

*Proof:*

$\tau$ is easily derived from equation $\tau = 3 - 4 \int_0^1 K(t) dt$.

$$\tau = 3 - 4 \int_0^1 K(m, \alpha; t) dt$$

$$\tau = 3 - 4 \int_0^1 \sum_{k=0}^{m} \alpha_k \binom{m}{k} t^k (1-t)^{m-k} dt$$

$$\tau = 3 - 4\sum_{k=0}^{m}\binom{m}{k}\alpha_k\int_0^1 t^k(1-t)^{m-k}dt$$

$$\tau = 3 - 4\sum_{k=0}^{m}\binom{m}{k}\alpha_k\beta(k+1, m-k+1)$$

**Lemma 4.3.** The lower tail $\lambda_L$ and the upper tail $\lambda_U$ dependences for Bernstein-Bézier type Archimedean copula are obtained by

$$\lambda_L = 2^{1-m\alpha_1}$$

$$\lambda_U = 2 - 2^{1-m(1-\alpha_{m-1})}$$

*Proof:*

$\lambda_U$ and $\lambda_L$ are easily derived from the equations $\lambda_L = 2^{\lim_{t\to 0^+}\left(t-K(t)\right)'}$ and $\lambda_U = 2 - 2^{\lim_{t\to 1^-}\left(t-K(t)\right)'}$.

$$\lambda_L = 2^{\lim_{t\to 0^+}\left(t-K(m,\alpha;t)\right)'}$$

$$\lambda_U = 2 - 2^{\lim_{t\to 1^-}\left(t-K(m,\alpha;t)\right)'}$$

first order derivative of Bernstein-Bézier polynomial is derived by

$$K'(m,\alpha;t) = m\sum_{k=0}^{m-1}(\alpha_{k+1} - \alpha_k)P_{k,m-1}$$

And also, from the end-point rule of Bersntein-Bézier polynomial, $K'(m,\alpha;t=0)$ and $K'(m,\alpha;t=1)$ are equal to $(\alpha_1 - \alpha_0)$ and $(\alpha_m - \alpha_{m-1})$ respectively, see Duncan (2005). Because of $\alpha_0 = 0$ and $\alpha_1 = 1$ then lower and upper tail dependence are derived by

$$\lambda_L = 2^{1-m\alpha_1},$$

$$\lambda_U = 2 - 2^{1-m(1-\alpha_{m-1})}.$$

It is seen that $\lambda_L$ and $\lambda_U$ are affected by only the control points $\alpha_1$ and $\alpha_{m-1}$, respectively. We can create Bernstein-Bézier type Archimedean copula using $\lambda_L$ and

$\lambda_U$, setting up the control points $\alpha_1$ and $\alpha_{m-1}$.

The following inequalities given in the next lemma provide an information for proper selection of $\lambda_U$ and $\lambda_L$.

**Lemma 4.4.** Let $\lambda_L$ and $\lambda_U$ be lower and upper tail dependence of Bernstein-Bézier type Archimedean copula with polynomial degree $m$. Then,

$$1 > \lambda_L > \frac{2^{2-m}}{2-\lambda_U}$$

is hold for all values of polynomial degree $m$.

*Proof:*

It can ve proved using the equation $\alpha_1 < \alpha_{m-1}$. Also, $0 < \lambda_U, \lambda_L < 1$, see Charpentier & Segers (2009). $\alpha_1$ and $\alpha_{m-1}$ can be written in terms of $\lambda_L$ and $\lambda_U$, respectively.

$$\alpha_1 = \frac{1-\log_2(\lambda_L)}{m}, \quad \alpha_{m-1} = \frac{\log_2(2-\lambda_U)+m-1}{m}$$

Then, using the inequations $\alpha_1 < \alpha_{m-1}$ and $0 < \lambda_L < 1$ , desired result is obtained.

#### *4.2.1 Simulating Data from Bernstein-Bézier type Archimedean Copula*

Data simulation from Bernstein-Bézier type Archimedean copula is given. And also, how it is possible to create a new distribution function which has desirable Kendall's tau and tail dependence are investigated. It is possible that we can create different distributions with the same dependence level.

The following procedure is used to create a distribution with the dependence characteristics represented by Kendall's tau and tail dependence coefficients.

1. The arbitrary value of the upper tail dependence $\lambda_U$ is determined primarily.
2. $\lambda_L$ is determined arbitrarily by using Lemma 4.4.
3. The value of Kendall's tau $\tau$ is determined for the distributions with polynomial degrees 2 and 3. For the distributions having polynomial degree $m \geq 4$, an interval of Kendall's tau is determined. Then, Kendall's tau is selected arbitrarily from this interval.
4. Bivariate data is simulated using following algorithms described in Nelsen (2006).

The algorithm based on Nelsen (2006) allows one to simulate $C(u,v)$ by Kendall distribution function $K(t)$.

1. Simulate two uniformly distributed random variables $S$ and $T$ on $[0,1]$.
2. Set $w = K^{-1}(t)$.
3. Set $u$ such that $\int_w^u \frac{1}{t-K(t)}dt - \ln(s) = 0$.
4. Set $v$ such that $\int_w^v \frac{1}{t-K(t)}dt - \ln(1-s) = 0$.

The range of the parameters and the dependence coefficients depending on the Bernstein-Bézier polynomial degree $m$ are summarized in Table 4.1. The relation among the parameters can also be seen in Table 4.1. It is observed that as the degree of the polynomial increases, the range of the dependence coefficients gets wider.

Kendall's tau, upper and lower tail dependence coefficients obtained by the Bernstein-Bézier type Archimedean copula with control points for degree $(m = 3,4,5)$ are summarized in Table 4.2. Also, different distributions having the same dependence level at the control points $\alpha_2$ and $\alpha_3$ for polynomial degree $5$ are given. All the Bernstein-Bézier control points and dependence coefficients are obtained by applying the simulation procedure (1-4). All cases in Table 4.2 are examined in the subsection (4.1.2-4.1.4).

Table 4.1 Range of parameters and dependence coefficients for different values of degree $m$

| m | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\tau$ | $\lambda_U$ | $\lambda_L$ |
|---|---|---|---|---|---|---|---|
| 3 | $(\frac{1}{3},1)$ | $(\max(\frac{2}{3},\alpha_1),1)$ | - | - | $(0,1)$ | $(0,1)$ | $(\frac{1}{4},1)$ |
| 4 | $(\frac{1}{4},1)$ | $(\max(\frac{2}{4},\alpha_1),1)$ | $(\max(\frac{3}{4},\alpha_2),1)$ | - | $(-0.2,1)$ | $(0,1)$ | $(\frac{1}{8},1)$ |
| 5 | $(\frac{1}{5},1)$ | $(\max(\frac{2}{5},\alpha_1),1)$ | $(\max(\frac{3}{5},\alpha_2),1)$ | $(\max(\frac{4}{5},\alpha_3),1)$ | $(-0.33,1)$ | $(0,1)$ | $(\frac{1}{16},1)$ |

Table 4.2 Parameters and dependence coefficients for different values of degree $m$

| Degree | $\alpha_0$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ | $\tau$ | $\lambda_U$ | $\lambda_L$ |
|---|---|---|---|---|---|---|---|---|---|
| $m=3$ | 0 | 0.7173 | 0.7928 | 1 | - | - | 0.4899 | 0.7 | 0.45 |
| $m=4$ | 0 | 0.3537 | 0.5828 | 0.9815 | 1 | - | 0.68 | 0.1 | 0.75 |
| $m=5$ | 0 | 0.4 | 0.43 | 0.8531 | 0.9169 | 1 | 0.6 | 0.5 | 0.5 |
| | 0 | 0.4 | 0.63 | 0.6531 | 0.9169 | 1 | 0.6 | 0.5 | 0.5 |

#### *4.2.2 Bernstein-Bézier Type Archimedean Copula with Degree Three*

A Bernstein-Bézier type Archimedean copula with degree 3 has the following distribution function,

$$K(3,\alpha;t)=\sum_{k=0}^{m=3}\alpha_k\binom{m}{k}t^k(1-t)^{3-k}, t\in[0,1]$$

from Lemma 4.1, $\alpha_0=0, \alpha_3=1$ , $\alpha_0<\alpha_1<\alpha_2<\alpha_3$ and $\alpha_1>\frac{1}{3}, \alpha_2>\frac{2}{3}$. Kendall's tau of the distribution is given by:

$$\tau=3-4\sum_{k=0}^{3}\alpha_k\binom{3}{k}\beta(k+1,3-k+1)=2-\alpha_1-\alpha_2$$

and lower and upper tail dependence are

$$\lambda_L=2^{1-3\alpha_1}, \lambda_U=2-2^{3\alpha_2-2}$$

$(1-4)$ procedure is applied to determine the Kendall's tau and the tail dependence values of the distribution. The arbitrary value of the upper tail dependence $\lambda_U$ is determined primarily in the range $\lambda_U\in(0,1)$. We select $\lambda_U$ as $0.7$, so $\alpha_2$ is equal to

$0.7928$ . From Lemma 4.4, $1 > \lambda_L > 0.3846$. Then, $\lambda_L$ is determined arbitrarily as $0.45$. So, $\alpha_1$ is equal to $0.7173$. The stage conditions for control points given Lemma 4.1 are satisfied. Finally, Kendall's tau is $0.4899$. $K(3, \alpha; t)$ with control points $\alpha_0 = 0, \alpha_1 = 0.7173, \alpha_2 = 0.7928$ and $\alpha_3 = 1$ has the Kendall's tau value as $\tau = 0.4899$ and the value tail dependence coefficients as $\lambda_L = 0.45$ and $\lambda_U = 0.7$. Simulated data and $K(m = 3, \alpha; t)$ with the sample of size 150 are visualized in Figure 4.1.

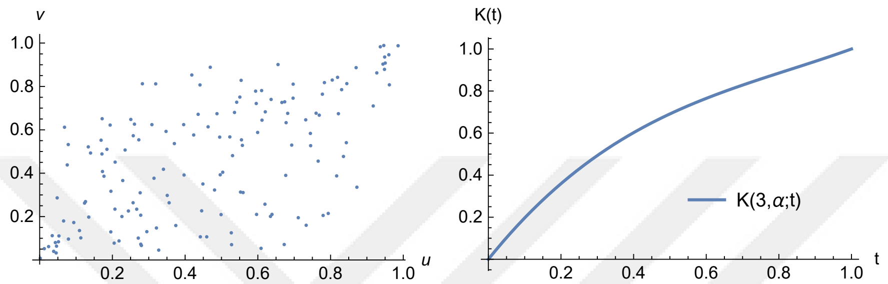



Figure 4.1 Simulated data from $K(3, \alpha; t)$ with $\tau = 0.4899, \lambda_L = 0.45, \lambda_U = 0.7$

### *4.2.3 Bernstein-Bézier Type Archimedean Copula with Degree Four*

Bernstein-Bézier type Archimedean copula with degree 4 has the following distribution function with the dependence characteristics, Kendall's tau, lower and upper tail dependence,

$$K(4, \alpha; t) = \sum_{k=0}^{4} \alpha_k \binom{4}{k} t^k (1-t)^{4-k}, t \in [0, 1],$$

$$\tau = \frac{1}{5}\Big(11 - 4(\alpha_1 + \alpha_2 + \alpha_3)\Big),$$

$$\lambda_L = 2^{1-4\alpha_1}, \lambda_U = 2 - 2^{4\alpha_3 - 3}.$$

$(1 - 4)$ procedure is applied to determine the Kendall's tau and the tail dependence values of the distribution. The arbitrary value of the upper tail dependence $\lambda_U$ is determined primarily in range $\lambda_U \in (0, 1)$. We select $\lambda_U$ as $0.1$ and so $\alpha_3$ is equal to $0.9815$. From Lemma 4.4, $1 > \lambda_L > 0.1315$. Then, $\lambda_L$ is determined arbitrarily as

$0.75$. So, $\alpha_1$ is equal to $0.3537$. Finally from Lemma 4.1, Kendall's tau should be selected in the range $\tau \in (0.3610, 0.7462)$. We determine Kendall's tau arbitrarily as $0.68$. So, $\alpha_2$ is equal to $0.5828$. Kendall distribution function $K(4, \alpha; t)$ with control points $\alpha_0 = 0, \alpha_1 = 0.3537, \alpha_2 = 0.5828, \alpha_3 = 0.9815$ and $\alpha_4 = 1$ has the value of Kendall's tau $\tau = 0.68$ and the values of tail dependences as $\lambda_L = 0.75$ and $\lambda_U = 0.1$. Simulated data and $K(4, \alpha; t)$ with the sample of size 150 is visualized in Figure 4.2.

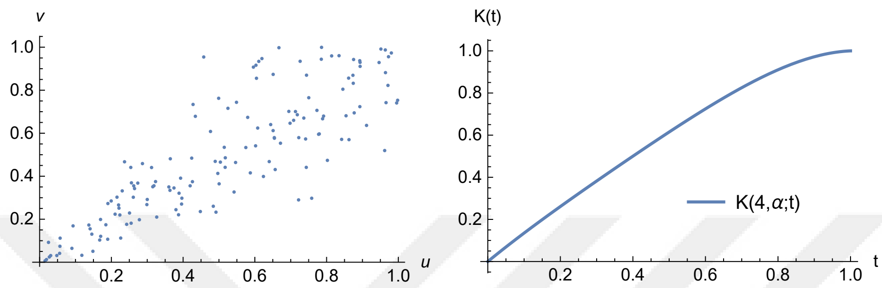



Figure 4.2 Simulated data from $K(4, \alpha; t)$ for $\tau = 0.68, \lambda_L = 0.75, \lambda_U = 0.1$

### *4.2.4 Bernstein-Bézier Type Archimedean Copula with Degree Five*

Bernstein-Bézier type Archimedean copula with degree 5 has the following distribution function with the dependence characteristics Kendall's tau, lower and upper tail dependence,

$$K(5, \alpha; t) = \sum_{k=0}^{5} \alpha_k \binom{5}{k} t^k (1-t)^{5-k}, t \in [0, 1],$$

$$\tau = \frac{1}{3}\Big(7 - 2(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)\Big),$$
$$\lambda_L = 2^{1-5\alpha_1}, \lambda_U = 2 - 2^{5\alpha_4 - 4}.$$

$(1-4)$ procedure is again applied to determine the Kendall's tau and the tail dependence values of the distribution. The arbitrary value of the upper tail dependence $\lambda_U$ is determined primarily in range $\lambda_U \in (0, 1)$. We select $\lambda_U$ as $0.5$ and so $\alpha_4$ is equal to $0.9169$. From Lemma 4.4, $1 > \lambda_L > 0.0833$. Then, $\lambda_L$ is determined arbitrarily as $0.5$. So, $\alpha_1$ is equal to $0.4$. Finally from Lemma 4.1, Kendall's tau should be selected in the range $\tau \in (0.2328, 0.6220)$. We determine Kendall's tau arbitrarily as $0.6$. $\alpha_2$ and $\alpha_3$ can be derived from solving equations $\alpha_2 + \alpha_3 = 1.2831$. From the last equation and Lemma 4.1, $\alpha_2$ and $\alpha_3$ should be selected in the range $\alpha_2 \in (0.4, 0.6415)$ and $\alpha_3 \in (0.6415, 0.8831)$, respectively. Different $\alpha_2$ and $\alpha_3$ values can be selected in order to provide $\alpha_2 + \alpha_3 = 1.2831$ in the range of $\alpha_2$ and $\alpha_3$. This case is important, because we can create different distributions with the same dependence level by selecting different $\alpha_2$ and $\alpha_3$ values. One possible selection is $\alpha_2 = 0.43$ and $\alpha_3 = 0.8531$. Another possible selection is $\alpha_2 = 0.63$ and $\alpha_3 = 0.6531$. Kendall distribution functions $K_1(5, \alpha; t)$ with control points $\alpha_0 = 0$, $\alpha_1 = 0.4$, $\alpha_2 = 0.43$, $\alpha_3 = 0.8531$, $\alpha_4 = 0.9169$, $\alpha_5 = 1$ and $K_2(5, \alpha; t)$ with control points $\alpha_0 = 0$, $\alpha_1 = 0.4$, $\alpha_2 = 0.63$, $\alpha_3 = 0.6531$, $\alpha_4 = 0.9169$, $\alpha_5 = 1$ at the same dependence level are visualized in Figure 4.3.

For the higher order polynomial degree, for example $m = 6$, the range of $\tau$, $\lambda_L$ and $\lambda_U$ are determined as the same as for degree $m < 6$. But the range of $\alpha_2, \alpha_3$ and $\alpha_4$ for the solutions of $\alpha_2 + \alpha_3 + \alpha_4 = a$ cannot be determined easily.

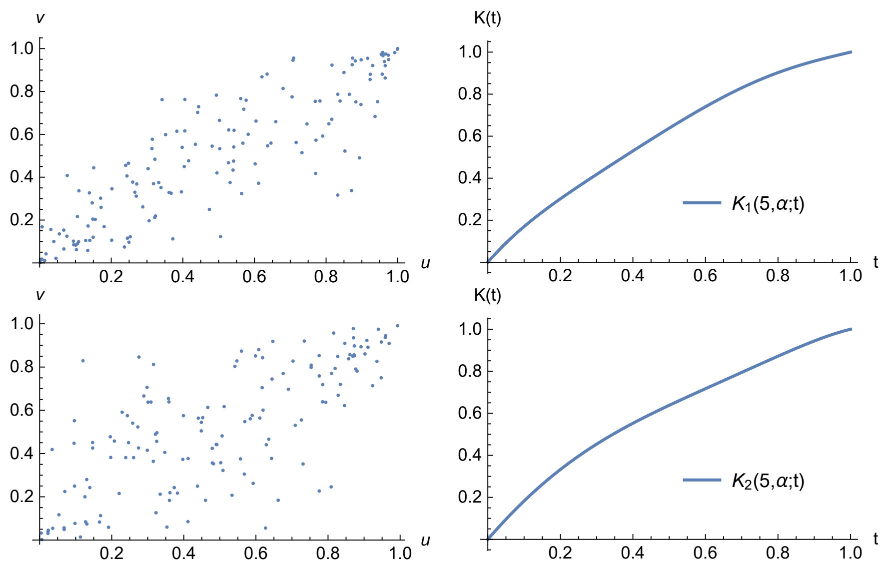



Figure 4.3 Simulated data from $K_1(5, \alpha; t)$ and $K_2(5, \alpha; t)$ for $\tau = 0.6, \lambda_L = 0.5, \lambda_U = 0.5$

## 4.3 Estimation of Archimedean Copula Via Kendall Transform and Bernstein Polynomials

### *4.3.1 Classical Estimate of Kendall Distribution Function*

Genest & Rivest (1993) investigated empirical estimate of Kendall distribution function. From the Theorem 3.2, Archimedean copulas are characterized by random variable $T = C(u, v)$. For the estimation of the random variable $T = H(x, y)$, univariate distribution function $K(T) = P(H(x, y) \leq t) = P(C(u, v) \leq t)$ should be estimated on the interval $[0, 1]$. This estimation process can be accomplished in two steps.

1. Empirical bivariate distribution function $H_n(X, Y)$ should be constructed
2. Then, the psuedo observations of $\hat{T}_i$ are obtained by

$$\hat{T}_i = \sum_{j=1}^{n} \mathrm{I}(X_i < X_j, Y_i < Y_j)/(n-1), i = 1, \ldots, n.$$

By using these psuedo observations, $K(t)$ is estimated by the empirical distribution function as

$$\hat{K}_n(t) = \sum_{i=1}^{n} \mathrm{I}(\hat{T}_i \leq t)/n.$$

Genest & Rivest (1993) stated that the empirical estimation of Kendall distribution function $\sqrt{n}$-consistent estimator. And, Barbe et al. (1996) proved consistency of this estimator.

**Theorem 4.1** (Genest & Rivest (1993))**.** Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be random sample from Copula $C(u, v)$. Under weak regularity conditions, the distribution of $T_i$ converges to Kendall distribution function $K(t) = P(C(u, v) \leq t)$. And also empirical distribution function $\hat{K}_n(t)$ is $\sqrt{n}$-consistent estimator of $K(t)$.

It should also be noted that the pseudo-observations are dependent and hence cannot be viewed as a random sample from $K(t)$. Despite this, the function $\hat{K}_n(t)$ indeed happens to be a consistent estimator of $K(t)$ (see Barbe et al. (1996) and Dimitrova et al. (2008)).

#### *4.3.2 Bernstein Polynomial Based Estimate of Kendall Distribution Function*

Generally, classical empirical distribution function has a good performance as an estimator of distribution function, however since it has jump discontinuities, estimating continuous distribution function may not be appropriate. Babu et al. (2002) proposed a smooth estimator of distribution function by using Bernstein polynomials. Then, Leblanc (2012) studied on estimating distribution function using Bernstein polynomials. The study has shown that Bernstein estimators of distribution functions have very good boundary properties, asymptotically normal and first-order efficient.

Susam & Ucer (2018) defined the Bernstein estimator of order $(m > 0)$ of the Kendall distribution function K as,

$$\hat{K}_{m,n}(t) = \sum_{k=0}^{m} K_n(k/m) P_{k,m}(t) \tag{4.2}$$

where $P_{k,m}(t) = \binom{m}{k} t^k (1-t)^{m-k}$ is the Binomial probability and $K_n$ is the empirical distribution function of $K(t)$.

In Figure 4.4, visual comparison between classical empirical and Bernstein empirical estimates of independent Kendall distribution function for sample size $n = 50$ and $m = 7$ is given.

It can be easily seen that Bernstein estimator of Kendall distribution function is genuine distribution function with the following properties for all values of $m$.

**Lemma 4.5.** The Bernstein estimator of Kendall distribution function has the following properties;

1. $\hat{K}_{m,n}(t) = 0$ for all $t \leq 0$,

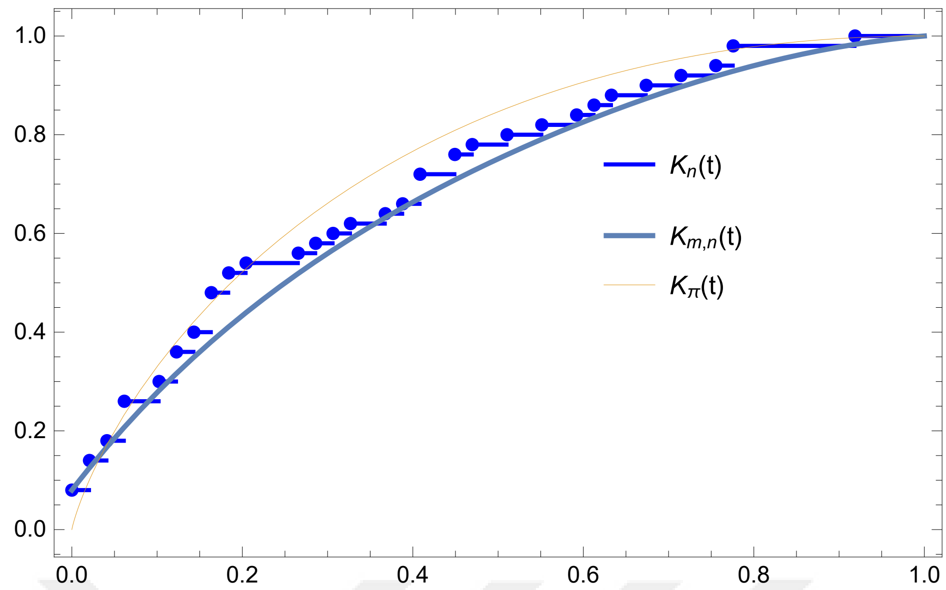



Figure 4.4 Estimated Kendall Distribution Function:$K_{\Pi}(t)$, $K_n(t)$, $K_{m,n}(t)$

2. $\hat{K}_{m,n}(t) = 1$ for all $t \geq 1$,

3. $\hat{K}'_{m,n}(t) > 0$ ,$t \in (0,1)$

4. $\hat{K}_{m,n}(t) > t$, $t \in (0,1)$

*Proof:*

At the interval endpoints 0 and 1, only the first and the last Bernstein polynomials are non-zero;

$$P_{0,m}(0) = 1, P_{k,m}(0) = 0 \;\; k = 1, \ldots, m-1,$$

$$P_{m,m}(1) = 1, P_{k,m}(1) = 0, \;\; k = 1, \ldots, m-1,$$

As a result, Bernstein polynomial based estimator $\sum_{k=0}^{m} K_n(\frac{k}{m})P_{k,m}(t)$ is equal to $K_n(0) = 0$ at $t = 0$ and equal to $K_n(1) = 1$ at $t = 1$. This property is referred to as endpoint interpolation. See Duncan (2005), Lorentz (1986).

For the proof of Property 4, first we note that $\frac{k}{m}\binom{m}{k} = \binom{m-1}{k-1}$ and $\sum_{k=0}^{m}\binom{m}{k}t^k(1-$

$t)^{m-k} = 1$. If $K_n(t) > t$ is satisfied, that is $K_n(\frac{k}{m}) = \frac{k}{m} + \epsilon_k$ then,

$$\begin{aligned}
K_{n,m}(t) &= \sum_{k=0}^{m} K_n(\frac{k}{m})\binom{m}{k} t^k(1-t)^{m-k} \\
&= \sum_{k=0}^{m}(\frac{k}{m} + \epsilon_k)\binom{m}{k} t^k(1-t)^{m-k} \\
&= \sum_{k=0}^{m}(\frac{k}{m})\binom{m}{k} t^k(1-t)^{m-k} + \sum_{k=0}^{m}(\epsilon_k)\binom{m}{k} t^k(1-t)^{m-k} \\
&= t\sum_{k=1}^{m}\binom{m-1}{k-1} t^{k-1}(1-t)^{m-k} + \sum_{k=0}^{m}(\epsilon_k)\binom{m}{k} t^k(1-t)^{m-k} \\
&= t\sum_{p=0}^{m-1} t^p(1-t)^{m-p-1}\binom{m-1}{p} + \sum_{k=0}^{m}(\epsilon_k)\binom{m}{k} t^k(1-t)^{m-k} \\
&= t + \sum_{k=0}^{m}(\epsilon_k)\binom{m}{k} t^k(1-t)^{m-k} > t
\end{aligned}$$

For the proof of Property 4, first derivative of Bernstein estimator of Kendall distribution function is given by

$$\frac{dK_{n,m}(t)}{dt} = k_{n,m}(t) = m\sum_{k=0}^{m-1}(K_n(\frac{k+1}{m}) - K_n(\frac{k}{m}))P_{k,m-1}(t)$$

for all $k$, $\frac{dK_{n,m}(t)}{dt} > 0$ in $t \in (0,1)$. See, Babu et al. (2002).

#### *4.3.2.1 Asymptotic Properties of Bernstein Estimator*

The following theorem which is introduced in Feller (1965) is very useful in demonstrating the asymptotic properties of Bernstein estimator of Kendall distribution function.

**Theorem 4.2** (Feller (1965)). If $u(t)$ is a bounded and continious function on the interval $[0,1]$, then as $m \to \infty$

$$u_m^*(t) = \sum_{k=0}^{m} u(\frac{k}{m})\binom{m}{k} t^k(1-t)^{m-k} \to u(t)$$

.

The following theorem adopted from Babu et al. (2002) shows that $\hat{K}_{n,m}$ is strongly consistent.

**Lemma 4.6.** Let $K$ be Kendall distribution function support on interval $[0, 1]$. If $m, n \to \infty$, then sup$\{\hat{K}_{n,m}(t) - K(t))\} \to 0$ a.s.

*Proof:*

$$\sup\{\hat{K}_{n,m}(t) - K(t))\} \leq \sup\{\hat{K}_{n,m}(t) - K^*_m(t))\} + \sup\{K^*_m(t) - K(t))\},$$

where

$$\hat{K}_{n,m}(t) - K^*_m(t) = \sum_{k=0}^{m} \left(K_n(\frac{k}{m}) - K(\frac{k}{m})\right)\binom{m}{k} t^k(1-t)^{m-k},$$

then we have

$$\sup\{\hat{K}_{n,m}(t) - K^*_m(t))\} \leq \max|K_n(\frac{k}{m}) - K(\frac{k}{m})| \leq \sup\{K_n(t) - K(t)\}$$

Since by Theorem 4.1, sup$\{K_n(t) - K(t)\} \to 0$ as $n \to \infty$.

#### *4.3.3 Bézier Curve Based Estimate of Kendall Distribution Function*

Dimitrova et al. (2008) estimated Kendall distribution function with B-spline estimator $K_\alpha(t, t_{k,m})$ of order $m$ (degree $m - 1$) and the number of internal knots, $k$, defined on the set of $2m + k$ knots,

$$t_{k,m} = (0, \ldots, 0 < t_{m+1} < \ldots < t_{m+k} < 1, \ldots, 1).$$

Then the estimate of Kendall distribution function is

$$\hat{K}_\alpha(t, t_{k,m}) = \sum_{i=0}^{p} \alpha_i N_{i,m}(t),$$

where $\alpha_i$'s are unknown coefficients and $N_{i,m}(t)$'s are $p = k + m$ B-Splines of order order $m$ on the knots $t_{k,m}$. Where

$$N_{i,0}(t) = \begin{cases} 1 & t_i \le t < t_{i+1} \\ 0 & otherwise \end{cases},$$

$$N_{i,m}(t) = \frac{t - t_i}{t_{i+m} - t_i} N_{i,m-1}(t) + \frac{t_{i+m+1} - t}{t_{i+m+1} - t_{i+1}} N_{i+1,m-1}(t).$$

In this approach, $(\alpha_1, \ldots, \alpha_p, k, t_{m+1}, \ldots, t_{m+k})$ should be estimated simultaneously. In our approach, we select the sequence of knots as

$$t_{k=0,m} = (0, \ldots, 0, 1, \ldots, 1),$$

then $N_{i,m}(t)$ are defined as Bernstein basis (a B-spline with no 'interior knots' is a Bézier curve). See, Oruc et al. (2017). So, we obtain the estimator given by

$$\hat{K}_{\alpha,m}(t) = \sum_{i=0}^{m} \alpha_i P_{i,m}(t),$$

where $P_{i,m}(t) = \binom{m}{i} t^i (1-t)^{m-i}$ is a Bernstein basis. The Bézier curve is defined by a set of control points. This estimator is easier to implement than the B-Spline based estimator. In our thesis, we use minimum distance estimator for estimating coefficients $\alpha_i, i = 1, \ldots, m-1$. Weighted Cramér-von-Mises distance between $\hat{K}_{\alpha,m}(t)$ and $\hat{K}_n(t)$ is obtained by the following equation,

$$\int_0^1 n(\hat{K}_n(t) - \hat{K}_{\alpha,m}(t))^2 d\hat{K}_n(t)$$

given the pseudo-observations $\hat{T}_j, j = 1, \ldots, n$, the coefficients can be estimated by minimizing Weighted Cramér-von-Mises distance as

$$\sum_{j=1}^{n} \left( \hat{K}_n(\hat{T}_{n,(j)}) - \sum_{i=0}^{m} \alpha_i P_{i,m}(\hat{T}_{n,(j)}) \right)^2. \tag{4.3}$$

By using the next Lemma, $\alpha_i$ can be estimated. For estimating Bernstein-Bézier coefficients, statistical programming language R is used. The Package "nloptr" is quite handy while optimizing non-linear function.

**Lemma 4.7.** The estimator $\hat{K}_{\alpha,m}(t)$ satisfies properties 3.1 if the following constraints hold;

1. $\alpha_0 = 0 < \alpha_1 < \alpha_2 < \ldots < \alpha_m = 1$,

2. $\alpha_i > \frac{i}{m}$ ,$i = 1, \ldots, m-1$.

*Proof:*

Similarly in the proof of Definition 4.1 $\hat{K}_{\alpha}(t=0,m) = \sum_{i=0}^{m} \alpha_{i+1} P_{i,m}(t=0) = 0$ holds since $\alpha_1 = 0$ similarly $\hat{K}_{\alpha,m}(t=1) = \sum_{i=0}^{m} \alpha_{i+1} P_{i,m}(t=1) = 1$ holds since $\alpha_{m+1} = 1$.

Also, first derivative of function $\hat{K}'_{\alpha,m}(t) = m \sum_{i=0}^{m-1} (\alpha_{i+1} - \alpha_i) P_{i,m-1}(t) > 0$ in $t \in (0,1)$. See, Duncan (2005). So, $\alpha_0 = 0 < \alpha_1 < \alpha_2 < \ldots < \alpha_m = 1$.

If Bézier control points $\alpha_i > \frac{i}{m}$, $i = 1, \ldots, m-1$ that is $\alpha_i = i/m + \epsilon_i$ then,

$$
\begin{aligned}
K_{\alpha,m}(t) &= \sum_{i=0}^{m} \alpha_i \binom{m}{k} t^i (1-t)^{m-i} \\
&= \sum_{i=0}^{m} (\frac{i}{m} + \epsilon_i) \binom{m}{i} t^i (1-t)^{m-i} \\
&= \sum_{i=0}^{m} (\frac{i}{m}) \binom{m}{i} t^i (1-t)^{m-i} + \sum_{i=0}^{m} (\epsilon_i) \binom{m}{i} t^i (1-t)^{m-i} \\
&= t \sum_{i=1}^{m} \binom{m-1}{i-1} t^{i-1} (1-t)^{m-i} + \sum_{i=0}^{m} (\epsilon_i) \binom{m}{i} t^i (1-t)^{m-i} \\
&= t \sum_{p=0}^{m-1} t^p (1-t)^{m-p-1} \binom{m-1}{p} + \sum_{i=0}^{m} (\epsilon_i) \binom{m}{i} t^i (1-t)^{m-i} \\
&= t + \sum_{i=0}^{m} (\epsilon_i) \binom{m}{i} t^i (1-t)^{m-i} > t
\end{aligned}
$$

#### 4.3.3.1 Asymptotic Properties of Bézier Curve Estimator

We know from the Subsection 4.3.3 that the Bézier curve is a special case of the spline estimator. Asymptotic properties of the spline estimator of Kendall distribution have been investigated by Kaishev et al. (2006), Dimitrova et al. (2008). Because of the spline based estimator is consistent, Bernstein-Bézier curve based estimator is also consistent for Kendall distribution function. Also, Wang & Ghosh (2012) introduced Bernstein-Bézier polynomial regression estimation and investigated its asymptotic properties.

We consider equation (4.3) as a Bernstein-Bézier polynomial regression problem. Consistency of Bernstein-Bézier curve based estimation of Kendall distribution function is proved by Bernstein-Bézier polynomial regression. Let $\Big\{x_i, y_i\Big\}_{i=1}^{n}$ be set of observations which is idependently and identically distributed as $(X, Y)$ satisfying regression model $Y = m(X) + \epsilon$, where $m(X) = E(Y|X = x)$ (True regression function). Let $\mathcal{F}$ be a class of continious function subject to a given class of shape restrictions. Then, the regression function $m(.)$ minimizes the distance as

$$m(.) = \underset{f \in \mathcal{F}}{\operatorname{argmin}}\ E\Big(f(X) - Y\Big)^2.$$

We consider the constrained Bernstein-Bézier polynomial family as follows

$$\mathcal{F}_m = \{B_m(x) = \sum_{k=0}^{m} \alpha_k P_{k,m}(x) : \mathbf{A}_m \boldsymbol{\alpha}_m \geq 0, \sum_{k=0}^{m} |\alpha_k| \leq L_m\} \qquad (4.4)$$

for $m = 1, 2, \ldots$, where $P_{k,m}(x) = \binom{m}{k} x^k (1-x)^{m-k}$ are Bernstein basis polynomials. Note that, we can express $B_m(x) = \boldsymbol{P}_m^T \boldsymbol{\alpha}_m$ where $\boldsymbol{P}_m^T = (P_{0,m}(x), \ldots, P_{m,m}(x))'$ and $\boldsymbol{\alpha}_m = (\alpha_0, \ldots, \alpha_m)'$. $\boldsymbol{A}_m$ represents the restriction matrix with dimension $R_m \times (m+1)$ where $R_m$ denotes the rank of $\boldsymbol{A}_m$. The estimator $\hat{m}(.)$, which is selected in $\mathcal{F}_m$, minimizes the function

$$\hat{m}(.) = \boldsymbol{P}_m^T \hat{\boldsymbol{\alpha}}_m; \quad \hat{\boldsymbol{\alpha}}_m = \underset{\hat{\boldsymbol{\alpha}}_m \in \boldsymbol{\xi}_m}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{P}_m^T \boldsymbol{\alpha}_m - Y_i)^2, \qquad (4.5)$$

where

$$\boldsymbol{\xi}_m = \{\boldsymbol{\alpha}_m \in R^{m+1} : \boldsymbol{A}_m\boldsymbol{\alpha}_m \geq 0,\ \sum_{k=0}^{m} |\alpha_k| \leq L_m\}.$$

In following theorem, Wang & Ghosh (2012) stated a bound for $\hat{m}(.)$ and true regression function $m(.)$.

**Theorem 4.3** (Wang & Ghosh (2012))**.** Let $\mathcal{F}_m$ be a class of functions $\mathcal{F}_m : R \to R$ which depends on data $\{(x_i, y_i), i = 1, \ldots, n\}$. If $\hat{m}(.)$ satisfies the equation 4.5 then

$$\begin{aligned}
&\int \Big(\hat{m}(x) - m(x)\Big)^2 \mu(dx) \\
&\leq 2 \sup_{B_m \in \mathcal{F}_m} \left| \frac{1}{n} \sum_{i=1}^{n} (B_m(X_i) - Y_i)^2 - E\Big((B_m(X) - Y)^2\Big) \right| \\
&+ \inf_{B_m \in \mathcal{F}_m} \int \Big(B_m(x) - m(x)\Big)^2 \mu(dx)
\end{aligned}$$

where $\mu$ denotes the distribution of $X$.

**Theorem 4.4** (Wang & Ghosh (2012))**.** Let $\mathcal{F}_m$ be as in equation (4.4) and $\hat{m}(.)$ be estimator defined as in equation (4.5). Let $T_L$ denote the truncation operation as

$$Y_L = T_L(Y) = yI(|y| \leq L) + Lsign(y)I(|y| > L),$$

and $T_L\mathcal{F}_m = \{T_L f : f \in \mathcal{F}_m\}$ is a class of truncated functions. When $L_m \to \infty$ and $L_m > 0$ if

$$\lim_{m\to\infty} \inf_{B_m \in T_L\mathcal{F}_m} \int \Big(B_m(x) - m(x)\Big)^2 \mu(dt) = 0 \ \ a.s.$$

$$\lim_{m\to\infty} \sup_{B_m \in T_L\mathcal{F}_m} \left| \frac{1}{n} \sum_{i=1}^{n} \Big(B_m(X_i - Y_{i,L})\Big)^2 - E\Big((B_m(x) - Y_L)\Big)^2 \right| = 0 \ \ a.s.$$

then

$$\lim_{m\to\infty} \int \Big(\hat{m}(x) - m(x)\Big)^2 \mu(dx) = 0 \ \ a.s.$$

By using Theorem 4.4 and Theorem 4.5 the required conditions proposed estimator $\hat{m}(.)$ defined in equation (4.5), are described in the following theorem.

**Theorem 4.5** (Wang & Ghosh (2012))**.** Let $\mathcal{F}_m$ be a function space defined in equation

(4.4) and $\hat{m}(.)$ be restricted regression function estimator given in equation (4.5). Then $R_m$ denotes the rank of restriction matrix $\boldsymbol{A}_m$ if

$$E(Y^2) < \infty \quad R_m, L_m \to \infty \quad \frac{R_m L_m^4 \log(L_m)}{n} \to 0 \quad \text{and} \quad \frac{L_m^4}{m^{1-\delta}} \to 0 \ \text{ for } \delta > 0$$

then,

$$\int \Big(\hat{m}(x) - m(x)\Big)^2 \mu(dx) \to 0 \ \text{ a.s.}$$

as $n \to \infty$

We consider equation (4.3) as a Bernstein-Bézier polynomial regression problem. For regression estimation of Kendall distribution function, consider data points $\Big\{\hat{T}_i, \hat{K}_n(\hat{T}_i)\Big\}_{i=1}^{n}$ where $\hat{T}_i$ are the psuedo observations defined in Section 4.2. Let $Y_i = \hat{K}_n(\hat{T}_i)$ and $X_i = \hat{T}_i$, then $\hat{m}(x)$ is the Bernstein-Bézier polynomail regression estimation of Kendall distribution function. Note that, $\hat{K}_n$ is the consistent estimator of $K$. The estimation of Kendall distribution function is the solution of the following optimizitation problem:

minimize: $\dfrac{1}{n} \sum_{i=1}^{n} \Big(\boldsymbol{P}_m^T(\hat{T}_i)\boldsymbol{\alpha}_m - \hat{K}_n(\hat{T}_i)\Big)^2$

subject to: $\boldsymbol{A}_m \boldsymbol{\alpha}_m \geq 0, \quad \sum_{k=0}^{m} |\alpha_k| \leq L_m.$

The restriction matrix $\boldsymbol{A}_m$ can be choosen according to Lemma 4.5. In practice $L_m$ can be choosen as reasonably large number and hence we do not need to choose its value to compute $\boldsymbol{\alpha}_m$. They state that $L_m \geq n^d \max_{1 \leq i \leq n} |Y_i|$ for some $d > 0$ can be used. Thus, $L_m$ has larger upper bounds. See, Wang & Ghosh (2012).

#### *4.3.4 Performance of the Proposed Estimators*

In this section, Monte Carlo simulation study is performed to compare the finite sample performances of the classical method, the Bernstein method and the Bézier Curve Method as mentioned in Section 4.2. The data are generated from three Archimedean copulas such as Clayton, Frank and Gumbel with Kendall's Tau 0.05,

Table 4.3 MISE scores of three Archimedean copula estimators for sample size $n = 100$

| Copula | $\tau$ | $\hat{K}_n$ | $\hat{K}_{n,m=5}$ | $\hat{K}_{n,m=10}$ | $\hat{K}_{n,m=15}$ | $\hat{K}_{\alpha,m=5}$ | $\hat{K}_{\alpha,m=10}$ | $\hat{K}_{\alpha,m=15}$ |
|---|---|---|---|---|---|---|---|---|
| Gumbel | 0.05 | 0.00059 | 0.00374 | 0.00111 | 0.00050 | 0.00032 | 0.00037 | 0.00040 |
| | 0.2 | 0.00058 | 0.00275 | 0.00083 | 0.00039 | 0.00034 | 0.00040 | 0.00044 |
| | 0.5 | 0.00052 | 0.00111 | 0.00037 | 0.00021 | 0.00029 | 0.00036 | 0.00040 |
| | 0.8 | 0.00032 | 0.00021 | 0.00008 | 0.00006 | 0.00016 | 0.00020 | 0.00022 |
| Clayton | 0.05 | 0.00061 | 0.00366 | 0.00110 | 0.00052 | 0.00036 | 0.00040 | 0.00044 |
| | 0.2 | 0.00059 | 0.00226 | 0.00071 | 0.00037 | 0.00030 | 0.00041 | 0.00045 |
| | 0.5 | 0.00051 | 0.00097 | 0.00033 | 0.00022 | 0.00027 | 0.00035 | 0.00038 |
| | 0.8 | 0.00031 | 0.00032 | 0.00011 | 0.00009 | 0.00015 | 0.00020 | 0.00021 |
| Frank | 0.05 | 0.00056 | 0.00377 | 0.00109 | 0.00048 | 0.00035 | 0.00036 | 0.00039 |
| | 0.2 | 0.00056 | 0.00257 | 0.00077 | 0.00037 | 0.00034 | 0.00038 | 0.00041 |
| | 0.5 | 0.00054 | 0.00108 | 0.00038 | 0.00023 | 0.00030 | 0.00037 | 0.00040 |
| | 0.8 | 0.00034 | 0.00030 | 0.00013 | 0.00010 | 0.00019 | 0.00021 | 0.00023 |

Table 4.4 MISE scores of three Archimedean copula estimators for sample size $n = 200$

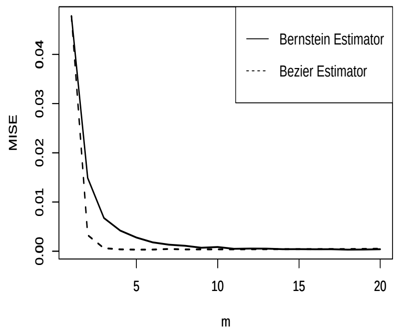
| Copula | $\tau$ | $\hat{K}_n$ | $\hat{K}_{n,m=5}$ | $\hat{K}_{n,m=10}$ | $\hat{K}_{n,m=15}$ | $\hat{K}_{\alpha,m=5}$ | $\hat{K}_{\alpha,m=10}$ | $\hat{K}_{\alpha,m=15}$ |
|---|---|---|---|---|---|---|---|---|
| Gumbel | 0.05 | 0.00026 | 0.00368 | 0.00100 | 0.00044 | 0.00019 | 0.00017 | 0.00018 |
| | 0.2 | 0.00026 | 0.00263 | 0.00072 | 0.00033 | 0.00018 | 0.00018 | 0.00020 |
| | 0.5 | 0.00024 | 0.00105 | 0.00031 | 0.00016 | 0.00013 | 0.00016 | 0.00017 |
| | 0.8 | 0.00015 | 0.00018 | 0.00006 | 0.00003 | 0.00006 | 0.00008 | 0.00009 |
| Clayton | 0.05 | 0.00028 | 0.00349 | 0.00095 | 0.00043 | 0.00018 | 0.00019 | 0.00020 |
| | 0.2 | 0.00027 | 0.00213 | 0.00060 | 0.00029 | 0.00013 | 0.00018 | 0.00020 |
| | 0.5 | 0.00024 | 0.00089 | 0.00027 | 0.00016 | 0.00012 | 0.00015 | 0.00017 |
| | 0.8 | 0.00014 | 0.00029 | 0.00009 | 0.00006 | 0.00005 | 0.00007 | 0.00008 |
| Frank | 0.05 | 0.00025 | 0.00363 | 0.00097 | 0.00042 | 0.00021 | 0.00017 | 0.00018 |
| | 0.2 | 0.00026 | 0.00240 | 0.00067 | 0.00031 | 0.00018 | 0.00017 | 0.00019 |
| | 0.5 | 0.00025 | 0.00102 | 0.00031 | 0.00017 | 0.00015 | 0.00016 | 0.00018 |
| | 0.8 | 0.00016 | 0.00028 | 0.00010 | 0.00006 | 0.00008 | 0.00008 | 0.00009 |

0.2, 0.5 and 0.8. These distributions cover different shapes and characteristics. $10,000$ Monte Carlo samples of size $n = 100, 200$ are generated from each type of Archimedean copulas and investigated the performances of the Bernstein estimator and the Bézier estimator for polynomial degrees $m = 5, 10$. We compare the performances of these estimators using Mean Integrated Square Error (MISE) scores which MISE is defined as

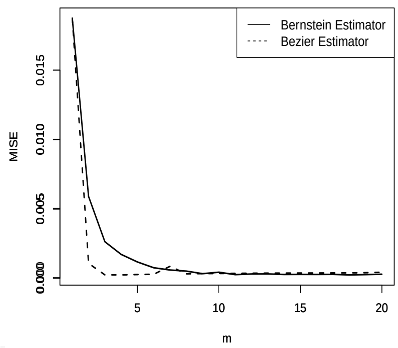
$$\text{MISE}[\hat{K}] = E\left[\int_0^1 (\hat{K}(t) - K(t))^2 dt\right]$$

Simulation results are shown in Table 4.3-4.4. When the results are examined, it is seen that $\text{MISE}[\hat{K}_{\alpha,m}] < \text{MISE}[\hat{K}_{n,m}] < \text{MISE}[\hat{K}_n]$. When the Kendall's tau is increased, MISE scores of all tests decrease. And also, it is a classical fact that when the sample of size $n$ is increased, MISE results decrease.

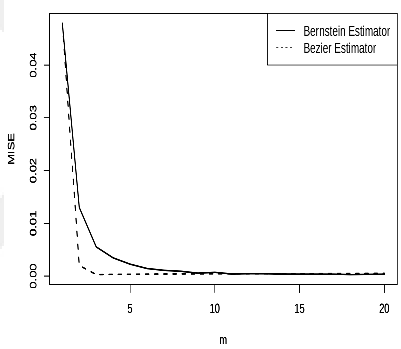
In Figure 4.5, MISE scores of the Bernstein and Bézier based tests are plotted for polynomial degree $m = 1, ..., 20$. It is seen that the Bézier curve based results with polynomial degree $m < 5$ have lower MISE scores than the Bernstein polynomial based results for the same polynomial degree.



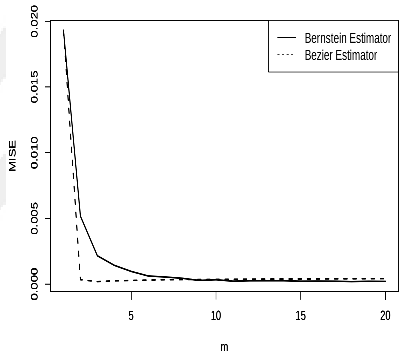

(a) Gumbel Copula with $\tau = 0.2$



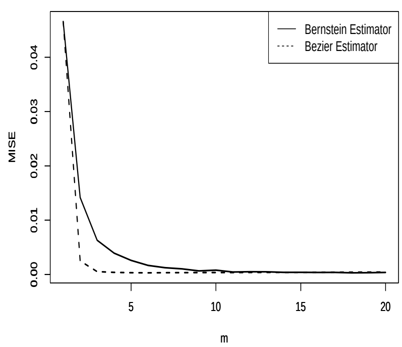

(b) Gumbel Copula with $\tau = 0.5$



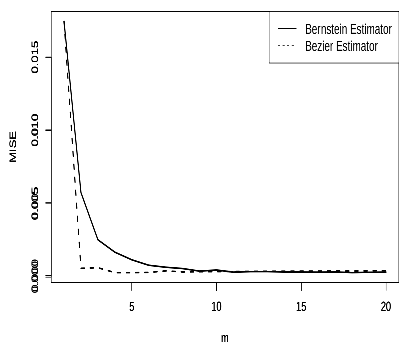

(c) Claytonl Copula with $\tau = 0.2$




(d) Clayton Copula with $\tau = 0.5$




(e) Frank Copula with $\tau = 0.2$




(f) Frank Copula with $\tau = 0.5$

Figure 4.5 MISE values for three Archimedean copulas for $\tau = 0.2, 0.5$

# CHAPTER FIVE
# GOODNESS OF FIT TESTS FOR BIVARIATE ARCHIMEDEAN COPULA

In this chapter, we construct a goodness-of-fit test of the null hypothesis $H_0 : C \in \mathcal{C}$, where the assumed copula belongs to the Archimedean family of copulas based on Cramér-von-Mises type statistics using both Bernstein polynomial and Bézier curve estimation of $K(t)$. We measure the power and the size of the proposed test and compare its performance with the existing ones by Monte Carlo study.

## 5.1 Testing Independence for Archimedean Copula based on Bernstein Estimate of Kendall Distribution Function

Copulas offers a great flexibility in building multivariate models in such a way that various general types of dependence can be represented. Thereupon, testing independence among these random variables gains importance in many disciplines. In the context of linear dependence, Gaussian model based tests have been used, but these tests cannot capture the other types of dependence and these types of tests do not answer how the variables are dependent, just simply answer whether or not they are independent of one another. So, alternative dependence measures and independence tests have been proposed. Blum et al. (1961) developed a nonparametric test based on Cramér-von-Mises measure which uses the joint empirical distribution in comparison. Belalia et al. (2017) proposed also nonparametric tests of independence by using Bernstein Empirical Copula which is also based on Cramér-von-Mises and Kullback-Leibler divergence type functionals.

The concept of estimating copula is related to that of testing independence in a bivariate sample. Copula becomes a product of two uniform distributions when the random variables of the bivariate sample are independent.

In this chapter, we introduce a test of independence of Archimedean Copula family based on Kendall distribution function $K(t)$ using Bernstein polynomial approximation. We obtain a test statistic for Archimedean family and compare its power using Monte Calo simulation as an attempt to choose the best candidate in that class. The new test statistic is based on Cramér-Von-Mises distance that is constructed using Bernstein Empirical Kendall distribution function $\hat{K}_n(t)$ such as,

$$CvM_n = n\int_0^1 (\hat{K}_{n,m}(t) - K(t))^2 dt \tag{5.1}$$

for the null hypothesis,

$$H_0 : K(t) = K_\Pi(t) \tag{5.2}$$

where $K_\Pi = t - t\log(t)$ is the independence form of Kendall distribution function. Genest & Rivest (1993) proposed a test statistic for independence of Archimedean copula given as

$$H_0 : K(t) = K_{\hat{\theta}}(t) \tag{5.3}$$

where $K_{\hat{\theta}}(t)$ is the Kendall distribution function and $\hat{\theta}$ is the inverted Kendall's tau estimator of dependence parameter $\theta$. For Archimedean copulas, Kendall distribution function $K(t)$ is summarized in Table 3.1. The proposed test statistic is given by

$$C_n = \frac{n}{3} + n\sum_{j=1}^{n-1}\hat{K}_n^2(\frac{j}{n})\{K_{\hat{\theta}}(\frac{j+1}{n}) - K_{\hat{\theta}}(\frac{j}{n})\} - n\sum_{j=1}^{n-1}\hat{K}_n(\frac{j}{n})\{K_{\hat{\theta}}^2(\frac{j+1}{n}) - K_{\hat{\theta}}^2(\frac{j}{n})\} \tag{5.4}$$

where $\hat{K}_n$ is the classical empirical Kendall distribution function. Bernstein polynomial based estimator is used to define a new test statistic of independence. For testing independence, the null hyphothesis is given by

$$H_0 : K(t) = K_\Pi(t) \tag{5.5}$$

where $K_\Pi(t) = t - t\log(t)$ is the Kendall distribution function for independent copula.

To test the independence, we use Cramér-von-Mises divergence which measure the distance between Bernstein Empirical Kendall distribution function and the independent copula.

**Theorem 5.1** (Susam & Ucer (2018))**.** Let $\hat{K}_{n,m}(t)$ be the empirical Bernstein Kendall distribution and let $K_{\Pi}(t)$ be the Kendall distribution for independent case, then Cramér-von-Mises test statistic $S_{n,m}$ is given by

$$S_{n,m} = n\int_0^1 (\hat{K}_{n,m}(t) - K_{\Pi}(t))^2 dt, \tag{5.6}$$

then under $H_0$, the test statistic is obtained as,

$$\begin{aligned}
S_{n,m} = {} & n\sum_{k=0}^{m}\binom{m}{k}^2 \hat{K}_n^2(\frac{k}{m})\beta(2k+1, 2m-2k+1) \\
& + 2n\sum_{0\leq k<s\leq m}\binom{m}{k}\binom{m}{s}\hat{K}_n(\frac{k}{m})\hat{K}_n(\frac{s}{m})\beta(k+s+1, 2m-k-s+1) \\
& - 2n\sum_{k=0}^{m}\left(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2, m-k+1)\right. \\
& \left. - \sum_{i=0}^{m-k}(-1)^i\binom{m-k}{i}(-\frac{1}{(k+i+2)^2})\Big)\right) \\
& + 0.6296n
\end{aligned}$$

where $\beta(.,.)$ is the beta function defined as $\beta(v_1, v_2) = \int_0^1 t^{v_1-1}(1-t)^{v_2-1}dt$ for $v_1, v_2$ positive integers.

*Proof:*

Cramér-Von-Mises test statistic measures the distance between Bernstein empirical distribution function and the independent copula. Cramér-von-Mises type test statistic $S_{n,m}$ can be written by

$$S_{n,m} = n\int_0^1 (\hat{K}_{n,m}(t) - K(t))^2 dt \tag{5.7}$$

Under null hyphothesis $H_0 : K(t) = K_{\Pi}(t) = t - t\log(t)$

$$\begin{aligned}
S_{n,m} &= n\int_0^1 (\hat{K}_{n,m}(t) - (t - t\log(t))^2 dt \\
&= n\int_0^1 \hat{K}^2_{n,m}(t)dt + n\int_0^1 (t - t\log(t))^2 dt - 2n\int_0^1 \hat{K}_{n,m}(t)(t - t\log(t))dt \\
&= n\int_0^1 \Big(\sum_{k=0}^{m}\binom{m}{k} t_k(1-t)^{m-k}\hat{K}_n(\frac{k}{m})\Big)^2 dt + n\int_0^1 (t(1-\log(t))^2 dt \\
&- 2n\int_0^1 \sum_{k=0}^{m}\binom{m}{k} t^k(1-t)^{m-k}\hat{K}_n(\frac{k}{m})t(1-\log(t)dt \\
&= I_1 + I_2 - I_3
\end{aligned}$$

Now we calculate part of $I_1$. We know that,

$$(a_1 + a_2 + \ldots + a_n)^2 = \sum_{i=1}^{n} a_i^2 + 2\sum_{1\le i<j\le n} a_i a_j,$$

then we can write

$$\begin{aligned}
I_1 &= n\sum_{k=0}^{m}\binom{m}{k}^2 \hat{K}^2_n(\frac{k}{m})\int_0^1 t^{2k}(1-t)^{2m-2k}dt \\
&+ 2n\sum_{0\le k<s\le m}\binom{m}{k}\hat{K}_n(\frac{k}{m})\binom{m}{s}\hat{K}_n(\frac{s}{m})\int_0^1 t^{k+s}(1-t)^{2m-k-s}dt \\
&= n\sum_{k=0}^{m}\binom{m}{k}^2 \hat{K}^2_n(\frac{k}{m})\beta(2k+1, 2m-2k+1) \\
&+ 2n\sum_{0\le k<s\le m}\binom{m}{k}\hat{K}_n(\frac{k}{m})\binom{m}{s}\hat{K}_n(\frac{s}{m})\beta(k+s+1, 2m-k-s+1).
\end{aligned}$$

Now we calculate the part of $I_3$

$$\begin{aligned}
I_3 &= 2n\sum_{k=0}^{m}\binom{m}{k}\hat{K}_n(\frac{k}{m})\int_0^1 t^k(1-t)^{m-k}t(1-\log(t))dt \\
&= 2n\sum_{k=0}^{m}\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\int_0^1 t^{k+1}(1-t)^{m-k}dt - \int_0^1 t^{k+1}(1-t)^{m-k}\log(t)dt\Big) \\
&= 2n\sum_{k=0}^{m}\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2, m-k+1) - \int_0^1 t^{k+1}(1-t)^{m-k}\log(t)dt\Big).
\end{aligned}$$

If we get binomial expansion term $(1-t)^{m-k}$ in the right side of equation

$$I_3 = 2n\sum_{k=0}^{m}\Bigg(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2,m-k+1)$$
$$-\sum_{i=0}^{m-k}(-1)^i\binom{m-k}{i}\int_0^1 t^{k+1+i}log(t)dt\Big)\Bigg).$$

Because $\log(0)$ does not exist, $0$ is not the domain of the integrand. This is an improper integral, so we use

$$I_3 = 2n\sum_{k=0}^{m}\Bigg(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2,m-k+1)$$
$$-\sum_{i=0}^{m-k}(-1)^i\binom{m-k}{i}\lim_{a\to 0}\int_a^1 t^{k+1+i}\log(t)dt\Big)\Bigg)$$

Recalling that $\log(1) = 0$, we get partial integral of integrand $(u = \log(t), dv = t^{k+1+i}dt)$

$$I_3 = 2n\sum_{k=0}^{m}\Bigg(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2,m-k+1)$$
$$-\sum_{i=0}^{m-k}(-1)^i\binom{m-k}{i}\Big(-\frac{1}{(k+i+2)^2}-\lim_{a\to 0}(\frac{\log(a)}{\frac{k+i+2}{a^{k+i+2}}})\Big)\Big)\Bigg)$$

$\lim_{a\to 0}(\frac{\log(a)}{\frac{k+i+2}{a^{k+i+2}}})$ is indeterminate because it equals to $\frac{-\infty}{+\infty}$ . Applying L'Hopital's rule gives us

$$I_3 = 2n\sum_{k=0}^{m}\Bigg(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2,m-k+1)$$
$$-\sum_{i=0}^{m-k}(-1)^i\binom{m-k}{i}(-\frac{1}{(k+i+2)^2})\Big)\Bigg)$$

Finally, after some calculation $I_2$ is calculated by $0.6296n$.

p-values are obtained by parametric bootstrap method as described in Berg (2009), Genest et al. (2006).

### *5.1.1 Parametric Bootstrap Method for Independence Test*

In order to compute p-values for any test statistic, one requires generating a large number, $N$, of independent samples of size $n$ from $C_{\theta_n}(u, v)$ and computing the corresponding values of the selected statistic, such as $S_{n,m}$. The procedure works as follows:

Step 1: Compute test statistic $S_{n,m}$ for data sets of size n

Step 2: Generate $N$ random samples of size $n$ from independent copula. Then compute test statistcs for all random samples $N$.

Step 3: If $S_{(1:N),n,m} \leq \ldots \leq S_{(N:N),n,m}$ denote the ordered values of the test statistics calculated in step 2, an estimate of the critical value of the test at level $\alpha$ based on $S_{n,m}$ is given by

$$\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{I}\left(S_{(i:N),n,m} \geq S_{n,m}\right)$$

which yields an estimate of the p-value associated with the observed value $S_{n,m}$ of the statistic.

### *5.1.2 Performance of the Independence Test*

Performance of the new test statistic given in Section 5.1 is conducted by Monte Carlo simulation study for different levels of Kendall's tau, sample sizes and the alternative Archimedean copulas. The purpose of this Monte Carlo simulation is to determine the level and the power of the underlying test statistic. In order to evaluate the critical value of the proposed test statistic under the null hypothesis at $\alpha = 0.05$ significance level, data is simulated from independent copula. The results are shown in Table 5.1. Testing procedure consists of rejecting $H_0$ when $S_{n,m}$ is greater than the $100(1-\alpha)th$ percentile of its distribution under the null hypothesis. The following factors are used in simulation;

1. $H_1$ copula (Clayton, Gumbel and Frank)

2. Level of dependence Kendall's tau ($\tau = 0.05, 0.1, 0.2$)

3. Sample sizes ($n = 50, 150$)

In each case, $10.000$ pseudo random samples of size $n = 50, 150$ are generated from the selected model with specified value of Kendall's tau. For each sample, $1.000$ bootstrap sample is generated from identified Archimedean copula for several levels of Kendall's tau, then $S_{n,m}$ statistic is calculated under the null hypothesis $H_0$. In each simulation case, sample of size $n$ is drawn from alternative hypothesis $H_1$ with several levels of Kendall's Tau. The hypotheses are given by

$$H_0 : K(t) = K_\Pi(t)$$
$$H_1 : K(t) \neq K_\Pi(t)$$

where $K_\Pi(t) = t - t\log(t)$.

Table 5.1 Percentage rejection of $H_0$ by various tests for data sets arising from independent copula (Level of the tests)

| | | Test based on | | |
|---|---|---|---|---|
| $H_1$ | $n$ | $C_n$ | $S_{n,m=7}$ | $S_{n,m=15}$ |
| Independent | 50 | 0.0483 | 0.0496 | 0.0501 |
| | 150 | 0.0545 | 0.0541 | 0.0546 |

The size and the power of the test are evaluated and the results are summarized in Table 5.1, 5.2 and 5.3. Each line of the tables is the percentage rejection of hypothesis $H_0 : K(t) = K_\Pi(t)$ corresponds to $H_1$ (Clayton, Frank and Gumbel). For example, in Table 6.2, when testing the independent copula from sample size $n = 50$, the test based on Bernstein polynomial with order $m = 7$ has power at $0.1259$ level, when the data are from Clayton copula with Kendall's tau $0.05$. We see that almost all cases, the new test statistic based on Bernstein polynomial has a better power than classical approximation proposed by Genest et al. (2006).

When Table 5.1 is examined, both of the tests based on Bernstein Empirical Kendall

Table 5.2 Percentage rejection of $H_0$ by various tests for data sets of size $n = 50$ arising from three Archimedean Copula models

| | | Test based on | | |
|---|---|---|---|---|
| $H_1$ | $\tau$ | $C_n$ | $S_{n,m=7}$ | $S_{n,m=15}$ |
| Clayton | $0.05$ | $0.0490$ | $0.1259$ | $0.0979$ |
| | $0.1$ | $0.1146$ | $0.2775$ | $0.2249$ |
| | $0.2$ | $0.4492$ | $0.6569$ | $0.6156$ |
| Frank | $0.05$ | $0.0428$ | $0.1272$ | $0.0996$ |
| | $0.1$ | $0.0774$ | $0.2592$ | $0.2132$ |
| | $0.2$ | $0.3028$ | $0.6427$ | $0.5876$ |
| Gumbel | $0.05$ | $0.0463$ | $0.1246$ | $0.0943$ |
| | $0.14$ | $0.0641$ | $0.2492$ | $0.2028$ |
| | $0.2$ | $0.2280$ | $0.6133$ | $0.5506$ |

distribution $S_{n,m}$ and classical empirical distribution $C_n$ hold their nominal level $(0.05)$. The test based on $S_{n,m}$ outperforms the test based on $C_n$ in all cases. It is a classical fact that the power of the test increases, as sample size increases.

Note that when the level of dependence increases for all types of Archimedean copulas, percentage of rejection of the null hypothesis increases in all cases. For the test based on $S_{n,m}$ with sample size $n = 50$, when the order of the Bernstein polynomial $m$ increases, power of the test decreases. However, for the test based on $S_{n,m}$ with sample size $n = 150$, when the order of the Bernstein polynomial $m$ increases, power of the test increases. So, optimum $m$ can be obtained to increase the power of the test for a given sample of size $n$. In short, the test based on $S_{n,m}$ for fitting the data is flexible and performs better when compared with the test based on $C_n$.

Table 5.3 Percentage rejection of $H_0$ by various tests for data sets of size $n = 150$ arising from three Archimedean Copula models

| $H_1$ | $\tau$ | Test based on $C_n$ | $S_{n,m=7}$ | $S_{n,m=15}$ |
|---|---|---|---|---|
| Clayton | 0.05 | 0.1277 | 0.2356 | 0.2384 |
| | 0.1 | 0.4399 | 0.5719 | 0.5905 |
| | 0.2 | 0.9613 | 0.9719 | 0.9792 |
| Frank | 0.05 | 0.0955 | 0.2257 | 0.2366 |
| | 0.1 | 0.3076 | 0.5603 | 0.5705 |
| | 0.2 | 0.8882 | 0.9179 | 0.9972 |
| Gumbel | 0.05 | 0.0767 | 0.2068 | 0.2144 |
| | 0.1 | 0.2234 | 0.5066 | 0.5185 |
| | 0.2 | 0.7717 | 0.9552 | 0.9619 |

## 5.2 Goodness-of-fit Test Based on Bernstein Polynomial and Bézier Curve Estimation of Kendall distribution for Archimedean Copula

We consider methods for assessing the adequacy of Archimedean copula specification and we try to avoid parametric assumptions for the marginals. We provide a way to test the fit of a parametric copula family for bivariate data. Several copula goodness-of-fit approaches have been proposed lately. Genest et al. (2006) introduced a test based on Cramér-von Mises distance of Kendall distribution function $K(t) = Pr\{C(u,v) \leq t\}$ and Empirical Kendall distribution function under the null hypothesis that

$$H_0 : K(t) = K_{\hat{\theta}}(t) \tag{5.8}$$

where $K_{\hat{\theta}}(t)$ is the Kendall distribution function which are listed in Table 3.1 and $\hat{\theta}$ is the estimation of dependence parameter $\theta$. For parameter estimation, inverted Kendall's tau method is used. This method is based on Equation (3.4). Also population value of Kendall's tau $\tau = 4E(T) - 1$ is estimated where $T$'s are pseudo observations defined in Section 4.2. Finally, $\hat{\theta}$ can be expressed by

$$\hat{\theta}_n = \tau^{-1}(\hat{\tau}_n) \tag{5.9}$$

$\tau(\theta)$ function is also listed in Table 3.2. $\sqrt{n}(\hat{\theta}-\theta)$ is asymptotically normal with zero mean. See, Kojadinovic & Yan (2010).

We use Bernstein polynomial and Bézier curve methods to estimate the Kendall distribution function and the Cramér-von Mises test statistics which are given by

$$S_{n,m} = n\int_0^1 (\hat{K}_{n,m}(t) - K(t))^2 dt, \tag{5.10}$$

$$T_{\alpha,m} = n\int_0^1 (\hat{K}_{\alpha,m}(t) - K(t))^2 dt, \tag{5.11}$$

respectively, where $\hat{K}_{n,m}(t)$ is Bernstein based estimator of $K(t)$ with polynomial degree $m$ and $\hat{K}_{\alpha,m}(t)$ is Bézier curve based estimator with control points $\alpha$ and with polynomisl degree $m$.

Proposition 5.1, 5.2 and 5.3 are given for the test statistics based on Bernstein estimator. And also, Proposition 5.4, 5.5 and 5.6 are given for the Bézier curve based polynomial test.

**Proposition 5.1.** Let $\hat{K}_{n,m}(t)$ be the empirical Bernstein Kendall distribution and $K_{G,\hat{\theta}}(t)$ be the Kendall distribution for Gumbel copula and $\hat{\theta}$ is the inverted Kendall's tau estimation of dependence parameter $\theta$, then Cramér-Von Mises test statistic $S^G_{\alpha,m}$ under $H_0 : K(t) = K_{(G,\hat{\theta})}(t)$ is given by

$$\begin{aligned}
S^G_{n,m} &= n\sum_{k=0}^{m}\binom{m}{k}^2 \hat{K}_n^2(\frac{k}{m})\beta(2k+1, 2m-2k+1) \\
&+ 2n\sum_{0\leq k<s\leq m}\binom{m}{k}\binom{m}{s}\hat{K}_n(\frac{k}{m})\hat{K}_n(\frac{s}{m})\beta(k+s+1, 2m-k-s+1) \\
&- 2n\sum_{k=0}^{m}\Bigg(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2, m-k+1) \\
&- \sum_{i=0}^{m-k}\frac{(-1)^i}{\hat{\theta}}\binom{m-k}{i}(-\frac{1}{(k+i+2)^2})\Big)\Bigg) \\
&+ n\frac{2+6\hat{\theta}+9\hat{\theta}^2}{27\hat{\theta}^2}
\end{aligned}$$

where $\beta(.,.)$ is the beta function defined as $\beta(v_1, v_2) = \int_0^1 t^{v_1-1}(1-t)^{v_2-1}dt$ for $v_1, v_2$ positive integers

*Proof:*

Cramér-Von-Mises test statistic measures the distance between Bernstein empirical distribution function and the Gumbel copula. Cramer-Von-Mises type test statistic $S_{n,m}^G$ can be written as

$$S_{n,m}^G = n\int_0^1 (\hat{K}_{n,m}(t) - K_{G,\hat{\theta}}(t))^2 dt \tag{5.12}$$

Under the null hyphothesis of $H_0 : K(t) = K_{G,\hat{\theta}}(t) = t(1 - \frac{t\log(t)}{\hat{\theta}})$

$$\begin{aligned}
S_{n,m}^G &= n\int_0^1 (\hat{K}_{n,m}(t) - t(1 - \frac{tlog(t)}{\hat{\theta}}))^2 dt \\
&= n\int_0^1 \hat{K}_{n,m}^2(t)dt + n\int_0^1 (t(1 - \frac{t\log(t)}{\hat{\theta}}))^2 dt \\
&- 2n\int_0^1 \hat{K}_{n,m}(t)(t(1 - \frac{tlog(t)}{\hat{\theta}}))dt \\
&= n\int_0^1 \Big(\sum_{k=0}^m \binom{m}{k} t_k (1-t)^{m-k}\hat{K}_n(\frac{k}{m})\Big)^2 dt + n\int_0^1 (t(1 - \frac{tlog(t)}{\hat{\theta}}))^2 dt \\
&- 2n\sum_{k=0}^m \hat{K}_n(\frac{k}{m})\binom{m}{k}\int_0^1 t^k(1-t)^{m-k}(t(1 - \frac{t\log(t)}{\hat{\theta}}))dt \\
&= I_1 + I_2 - I_3
\end{aligned}$$

Now, we calculate part of $I_1$. We know that,

$$(a_1 + a_2 + \ldots + a_n)^2 = \sum_{i=1}^n a_i^2 + 2\sum_{0\le i<j\le n} a_i a_j,$$

then we can write

$$\begin{aligned}
I_1 &= n\sum_{k=0}^{m}\binom{m}{k}^2\hat{K}_n^2(\frac{k}{m})\int_0^1 t^{2k}(1-t)^{2m-2k}dt \\
&+ 2n\sum_{0\le k<s\le m}\binom{m}{k}\hat{K}_n(\frac{k}{m})\binom{m}{s}\hat{K}_n(\frac{s}{m})\int_0^1 t^{k+s}(1-t)^{2m-k-s}dt \\
&= n\sum_{k=0}^{m}\binom{m}{k}^2\hat{K}_n^2(\frac{k}{m})\beta(2k+1,2m-2k+1) \\
&+ 2n\sum_{0\le k<s\le m}\binom{m}{k}\hat{K}_n(\frac{k}{m})\binom{m}{s}\hat{K}_n(\frac{s}{m})\beta(k+s+1,2m-k-s+1)
\end{aligned}$$

Now, we calculate the part of $I_3$

$$\begin{aligned}
I_3 &= 2n\sum_{k=0}^{m}\binom{m}{k}\hat{K}_n(\frac{k}{m})\int_0^1 t^k(1-t)^{m-k}(t(1-\frac{t\log(t)}{\hat{\theta}}))dt \\
&= 2n\sum_{k=0}^{m}\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\int_0^1 t^{k+1}(1-t)^{m-k}dt-\int_0^1 t^{k+1}(1-t)^{m-k}\frac{\log(t)}{\hat{\theta}}dt\Big) \\
&= 2n\sum_{k=0}^{m}\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2,m-k+1)-\int_0^1 t^{k+1}(1-t)^{m-k}\frac{\log(t)}{\hat{\theta}}dt\Big)
\end{aligned}$$

If we get binomial expansion term $(1-t)^{m-k}$ in the right side of equation, then we obtain

$$\begin{aligned}
I_3 = 2n\sum_{k=0}^{m}\Bigg(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2,m-k+1) \\
-\sum_{i=0}^{m-k}\frac{(-1)^i}{\hat{\theta}}\binom{m-k}{i}\int_0^1 t^{k+1+i}\log(t)dt\Big)\Bigg)
\end{aligned}$$

Because $\log(0)$ does not exist, $0$ is not the domain of the integrand. This is an improper integral, so we use

$$I_3 = 2n\sum_{k=0}^{m}\Bigg(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2,m-k+1)$$
$$-\sum_{i=0}^{m-k}\frac{(-1)^i}{\hat{\theta}}\binom{m-k}{i}\lim_{a\to 0}\int_a^1 t^{k+1+i}\log(t)dt\Big)\Bigg)$$

Recalling that $\log(1) = 0$, we get partial integral of integrand ($u = \log(t), dv = t^{k+1+i}dt$)

$$I_3 = 2n\sum_{k=0}^{m}\Bigg(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2,m-k+1)$$
$$-\sum_{i=0}^{m-k}\frac{(-1)^i}{\hat{\theta}}\binom{m-k}{i}\Big(-\frac{1}{(k+i+2)^2}$$
$$-\lim_{a\to 0}(\frac{\frac{\log(a)}{k+i+2}}{a^{k+i+2}})\Big)\Big)\Bigg)$$

$\lim_{a\to 0}(\frac{\frac{\log(a)}{k+i+2}}{a^{k+i+2}})$ is indeterminate because it equals to $\frac{-\infty}{+\infty}$ . Applying L'Hopital's rule gives us

$$I_3 = 2n\sum_{k=0}^{m}\Bigg(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2,m-k+1)$$
$$-\sum_{i=0}^{m-k}\frac{(-1)^i}{\hat{\theta}}\binom{m-k}{i}(-\frac{1}{(k+i+2)^2})\Big)\Bigg)$$

Finally, after some calculation $I_2$ is calculated by

$$n\frac{2+6\hat{\theta}+9\hat{\theta}^2}{27\hat{\theta}^2}.$$

**Proposition 5.2.** Let $\hat{K}_{n,m}(t)$ be the empirical Bernstein Kendall distribution function, $K_{C,\hat{\theta}}(t)$ be the Kendall distribution for Clayton copula and and $\hat{\theta}$ be the inverted Kendall's tau estimation of dependence parameter $\theta$, then Cramér-Von Mises test statistic $S^C_{n,m}$ under $H_0 : K(t) = K_{(C,\hat{\theta})}(t)$ is given by

$$
\begin{aligned}
S^C_{n,m} &= n \sum_{k=0}^{m} \binom{m}{k}^2 \hat{K}^2_n(\frac{k}{m})\beta(2k+1, 2m-2k+1) \\
&+ 2n \sum_{0\leq k<s\leq m} \binom{m}{k}\binom{m}{s}\hat{K}_n(\frac{k}{m})\hat{K}_n(\frac{s}{m})\beta(k+s+1, 2m-k-s+1) \\
&- 2n \sum_{k=0}^{m} \left(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2, m-k+1)(\frac{\hat{\theta}+1}{\hat{\theta}}) \right. \\
&\left. - \frac{\beta(k+\hat{\theta}+2, m-k+1)}{\hat{\theta}}\Big)\right) + n\frac{17+13\hat{\theta}+2\hat{\theta}^2}{27(1+\hat{\theta})+6\hat{\theta}^2}
\end{aligned}
$$

where $\beta(.,.)$ is the beta function defined as $\beta(v_1, v_2) = \int_0^1 t^{v_1-1}(1-t)^{v_2-1}dt$ for $v_1, v_2$ positive integers.

*Proof:*

Cramér-Von-Mises type test statistic $S^C_{n,m}$ can be obtained as

$$S^C_{n,m} = n\int_0^1 (\hat{K}_{n,m}(t) - K_{C,\hat{\theta}}(t))^2 dt. \tag{5.13}$$

$K_{C,\alpha}(t)$ can be rewritten as

$$K_{C,\theta}(t) = t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}}. \tag{5.14}$$

Under null hyphothesis of $H_0 : K(t) = K_{C,\hat{\theta}}(t) = t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}}$,

$$
\begin{aligned}
S^C_{n,m} &= n \int_0^1 (\hat{K}_{n,m}(t) - (t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}}))^2 dt \\
&= n \int_0^1 \hat{K}^2_{n,m}(t) dt + n \int_0^1 (t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}})^2 dt - 2n \int_0^1 \hat{K}_{n,m}(t)(t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}}) dt \\
&= n \int_0^1 \Big( \sum_{k=0}^m \binom{m}{k} t_k (1-t)^{m-k} \hat{K}_n(\frac{k}{m}) \Big)^2 dt + n \int_0^1 (t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}})^2 dt \\
&- 2n \sum_{k=0}^m \binom{m}{k} \hat{K}_n(\frac{k}{m}) \int_0^1 t^k (1-t)^{m-k} (t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}}) dt \\
&= I_1 + I_2 - I_3
\end{aligned}
$$

part of the $I_1$ is the same as in proof of the proposition 5.1. Now we calculate the part of $I_3$

$$
\begin{aligned}
I_3 =& 2n \sum_{k=0}^m \binom{m}{k} \hat{K}_n(\frac{k}{m}) \int_0^1 t^k (1-t)^{m-k} (t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}}) dt \\
&= 2n \sum_{k=0}^m \binom{m}{k} \hat{K}_n(\frac{k}{m}) \Big( (1 + \frac{1}{\hat{\theta}}) \int_0^1 t^{k+1} (1-t)^{m-k} dt \\
&- \frac{1}{\hat{\theta}} \int_0^1 t^{k+\theta+1} (1-t)^{m-k} dt \Big) \\
&= 2n \sum_{k=0}^m \binom{m}{k} \hat{K}_n(\frac{k}{m}) \Big( \frac{\hat{\theta}+1}{\hat{\theta}} \beta(k+2, m-k+1) \\
&- \frac{1}{\hat{\theta}} \beta(k + \hat{\theta} + 2, m-k+1) \Big).
\end{aligned}
$$

Finally, after some calculation $I_2$ is calculated as

$$
\frac{17 + 13\hat{\theta} + 2\hat{\theta}^2}{27(1+\hat{\theta}) + 6\hat{\theta}^2}
$$

.

**Proposition 5.3.** Let $\hat{K}_{n,m}(t)$ be the empirical Bernstein Kendall distribution , $K_{F,\hat{\theta}}(t)$ be the Kendall distribution for Frank copula and and $\hat{\theta}$ be the inverted Kendall's tau estimation of dependence parameter $\theta$, then Cramér-Von Mises test statistic $S^F_{n,m}$ under

$H_0 : K(t) = K_{(F,\hat{\theta})}(t)$ is given by

$$
\begin{aligned}
S^F_{n,m} &= n\sum_{k=0}^{m}\binom{m}{k}^2 \hat{K}^2_n(\frac{k}{m})\beta(2k+1, 2m-2k+1) \\
&+ 2n\sum_{0\le k<s\le m}\binom{m}{k}\binom{m}{s}\hat{K}_n(\frac{k}{m})\hat{K}_n(\frac{s}{m})\beta(k+s+1, 2m-k-s+1) \\
&- 2n\sum_{k=0}^{m}\left(\binom{m}{k}\hat{K}_n(\frac{k}{m})\Big(\beta(k+2, m-k+1)\right. \\
&\left.- \sum_{i=0}^{m-k}\frac{(-1)^i}{\hat{\theta}}\binom{m-k}{i}A(t,k,i)\right) + nB(t,\hat{\theta})
\end{aligned}
$$

where $\beta(.,.)$ is the beta function defined as $\beta(v_1, v_2) = \int_0^1 t^{v_1-1}(1-t)^{v_2-1}dt$ for $v_1, v_2$ positive integers and $A(t,k,j)$ and $B(t,\hat{\theta})$ are numerical solution of the $\int_0^1 t^{k+j}(\exp(t*\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})$ and $\int_0^1 \left(t - \frac{(\exp(t\hat{\theta})-1)}{\hat{\theta}}\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})\right)^2 dt$ respectively.

*Proof:*

Cramér-Von-Mises test statistic measures the distance between Bézier empirical distribution function and the Frank copula. Cramér-von-Mises type test statistic $S^F_{n,m}$ can be written as

$$S^F_{\alpha,m} = n\int_0^1 (\hat{K}_{n,m}(t) - K_{F,\hat{\theta}}(t))^2 dt. \tag{5.15}$$

Under null hyphothesis of $H_0 : K(t) = t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}}$

$$
\begin{aligned}
S^F_{n,m} &= n\int_0^1 (\hat{K}_{n,m}(t) - (t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}}))^2 dt \\
&= n\int_0^1 \hat{K}^2_{n,m}(t)dt + n\int_0^1 (t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})^2 dt \\
&- 2n\int_0^1 \hat{K}_{n,m}(t)(t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})dt \\
&= n\int_0^1 \Big(\sum_{k=0}^{m}\binom{m}{k}t^k(1-t)^{m-k}K_n(\frac{k}{m})\Big)^2 dt \\
&+ n\int_0^1 (t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})^2 dt \\
&- 2n\sum_{k=0}^{m}\binom{m}{k}K_n(\frac{k}{m})\int_0^1 t^k(1-t)^{m-k}(t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})dt \\
&= I_1 + I_2 - I_3
\end{aligned}
$$

Part of the $I_1$ is the same as in proof of the proposition 5.1. Now we calculate the part of $I_3$ as

$$
\begin{aligned}
I_3 &= 2n\sum_{k=0}^{m}\binom{m}{k}K_n(\frac{k}{m})\int_0^1 t^k(1-t)^{m-k}(t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})dt \\
&= 2n\sum_{k=0}^{m}\binom{m}{k}K_n(\frac{k}{m})\Bigg(\int_0^1 t^{k+1}(1-t)^{m-k}dt \\
&- \int_0^1 t^k(1-t)^{m-k}(\frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})dt\Bigg)
\end{aligned}
$$

If we get binomial expansion term $(1-t)^{m-k}$ in second integral of the right side of equation

$$
\begin{aligned}
I_3 = 2n\sum_{k=0}^{m}\binom{m}{k}K_n(\frac{k}{m})\Bigg(&\beta(k+2, m-k+1) - \sum_{j=0}^{m-k}(-1)^j\binom{m-k}{j} \\
&\times \int_0^1 t^{k+j}(\frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})dt\Bigg).
\end{aligned}
$$

There are no explicit solutions of integral

$$A(t,k,j) = \int_0^1 t^{k+j}(\frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{exp(-\hat{\theta})-1})}{\hat{\theta}})dt,$$

We should use numerical solution of the integral.

Finally, $I_2$ is calculated by numerical from the integral $nB(t,\hat{\theta}) = n\int_0^1 (t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})^2 dt$.

**Proposition 5.4.** Let $\hat{K}_{\alpha,m}(t)$ be the empirical Bézier Kendall distribution function, $K_{G,\hat{\theta}}(t)$ be the Kendall distribution for Gumbel copula and $\hat{\theta}$ be the inverted Kendall's tau estimation of dependence parameter $\theta$, then Cramér-Von Mises test statistic $T^G_{\alpha,m}$ under $H_0 : K(t) = K_{(G,\hat{\theta})}(t)$ is given by

$$\begin{aligned} T^G_{\alpha,m} &= n\sum_{i=0}^{m}\binom{m}{i}^2 \alpha_i^2 \beta(2i+1, 2m-2i+1) \\ &+ 2n\sum_{0\le i<s\le m}\binom{m}{i}\binom{m}{s}\alpha_i\alpha_s\beta(i+s+1, 2m-i-s+1) \\ &- 2n\sum_{i=0}^{m}\Bigg(\binom{m}{i}\alpha_i\Big(\beta(i+2, m-i+1) \\ &- \sum_{j=0}^{m-i}\frac{(-1)^j}{\hat{\theta}}\binom{m-i}{j}(-\frac{1}{(i+j+2)^2})\Big)\Bigg) \\ &+ n\frac{2+6\hat{\theta}+9\hat{\theta}^2}{27\hat{\theta}^2}, \end{aligned}$$

where $\beta(.,.)$ is the beta function defined as $\beta(v_1, v_2) = \int_0^1 t^{v_1-1}(1-t)^{v_2-1}dt$ for $v_1, v_2$ positive integers.

*Proof:*

Cramér-von-Mises test statistic measures the distance between Bézier empirical distribution function and the Gumbel copula. Cramér-von-Mises type test statistic

$S^G_{\alpha,m}$ under null hyphothesis of $H_0 : K(t) = K_{(G,\hat{\theta})}(t) = t(1 - \frac{\log(t)}{\hat{\theta}})$ can be written as

$$T^G_{\alpha,m} = n \int_0^1 (\hat{K}_{\alpha,m}(t) - K_{(G,\hat{\theta})}(t))^2 dt \tag{5.16}$$

$$\begin{aligned}
T^G_{\alpha,m} &= n \int_0^1 (\hat{K}_{\alpha,m}(t) - t(1 - \frac{\log(t)}{\hat{\theta}}))^2 dt \\
&= n \int_0^1 \hat{K}^2_{\alpha,m}(t) dt + n \int_0^1 (t(1 - \frac{\log(t)}{\hat{\theta}}))^2 dt - 2n \int_0^1 \hat{K}_{\alpha,m}(t)(t(1 - \frac{\log(t)}{\hat{\theta}}))dt \\
&= n \int_0^1 \Big( \sum_{i=0}^{m} \binom{m}{i} t^i (1-t)^{m-i} \alpha_i \Big)^2 dt + n \int_0^1 (t(1 - \frac{\log(t)}{\hat{\theta}}))^2 dt \\
&- 2n \sum_{i=0}^{m} \alpha_i \binom{m}{i} \int_0^1 t^i (1-t)^{m-i} (t(1 - \frac{\log(t)}{\hat{\theta}}))dt \\
&= I_1 + I_2 - I_3
\end{aligned}$$

Now we calculate part of $I_1$. We know that, $(a_1 + a_2 + \ldots + a_n)^2 = \sum_{i=1}^n a_i^2 + 2\sum_{0 \le i < j \le m} a_i a_j$, then we can write

$$\begin{aligned}
I_1 &= n \sum_{i=0}^{m} \binom{m}{i}^2 \alpha_i^2 \int_0^1 t^{2i} (1-t)^{2m-2i} dt \\
&+ 2n \sum_{0 \le i < s \le m} \binom{m}{i} \alpha_i \binom{m}{s} \alpha_s \int_0^1 t^{i+s} (1-t)^{2m-i-s} dt \\
&= n \sum_{i=0}^{m} \binom{m}{i}^2 \alpha_i^2 \beta(2i+1, 2m-2i+1) \\
&+ 2n \sum_{0 \le i < s \le m} \binom{m}{i} \alpha_i \binom{m}{s} \alpha_s \beta(i+s+1, 2m-i-s+1)
\end{aligned}$$

Now we calculate the part of $I_3$

$$\begin{aligned}
I_3 &= 2n\sum_{i=0}^{m}\binom{m}{i}\alpha_i\int_0^1 t^i(1-t)^{m-i}(t(1-\frac{\log(t)}{\hat{\theta}}))dt \\
&= 2n\sum_{i=0}^{m}\binom{m}{i}\alpha_i\Big(\int_0^1 t^{i+1}(1-t)^{m-i}dt-\int_0^1 t^{i+1}(1-t)^{m-i}\frac{\log(t)}{\hat{\theta}}dt\Big) \\
&= 2n\sum_{i=0}^{m}\binom{m}{i}\alpha_i\Big(\beta(i+2,m-i+1)-\int_0^1 t^{i+1}(1-t)^{m-i}\frac{\log(t)}{\hat{\theta}}dt\Big)
\end{aligned}$$

If we get binomial expansion term $(1-t)^{m-i}$ in the right side of equation

$$\begin{aligned}
I_3 = 2n\sum_{i=0}^{m}\Bigg(\binom{m}{i}\alpha_i\Big(\beta(i+2,m-i+1) \\
-\sum_{j=0}^{m-i}\frac{(-1)^j}{\hat{\theta}}\binom{m-i}{j}\int_0^1 t^{i+1+j}\log(t)dt\Big)\Bigg)
\end{aligned}$$

Because $\log(0)$ does not exist, $0$ is not the domain of the integrand. This is an improper integral, so we use

$$\begin{aligned}
I_3 = 2n\sum_{i=0}^{m}\Bigg(\binom{m}{i}\alpha_i\Big(\beta(i+2,m-i+1) \\
-\sum_{j=0}^{m-k}\frac{(-1)^j}{\hat{\theta}}\binom{m-i}{j}\lim_{a\to 0}\int_a^1 t^{i+1+j}\log(t)dt\Big)\Bigg).
\end{aligned}$$

Recalling that $\log(1)=0$, we get partial integral of integrand $(u=\log(t), dv = t^{k+1+i}dt)$

$$\begin{aligned}
I_3 = 2n\sum_{i=0}^{m}\Bigg(\binom{m}{i}\alpha_i\Bigg(\beta(i+2,m-i+1)-\sum_{j=0}^{m-i}\frac{(-1)^j}{\hat{\theta}}\binom{m-i}{j}\Big(-\frac{1}{(i+j+2)^2} \\
-\lim_{a\to 0}\big(\frac{\frac{\log(a)}{i+j+2}}{a^{i+j+2}}\big)\Big)\Bigg)\Bigg).
\end{aligned}$$

$\lim_{a\to 0}\left(\frac{\frac{\log(a)}{i+j+2}}{a^{i+j+2}}\right)$ is indeterminate because it equals to $\frac{-\infty}{+\infty}$. Applying L'Hopital's rule gives us

$$I_3 = 2n\sum_{i=0}^{m}\left(\binom{m}{i}\alpha_i\Big(\beta(i+2,m-i+1)-\sum_{j=0}^{m-i}\frac{(-1)^j}{\hat{\theta}}\binom{m-i}{j}\big(-\frac{1}{(i+j+2)^2}\big)\Big)\right).$$

Finally, after some algebra $I_2$ is calculated by

$$n\frac{2+6\hat{\theta}+9\hat{\theta}^2}{27\hat{\theta}^2}.$$

**Proposition 5.5.** Let $\hat{K}_{\alpha,m}(t)$ be the empirical Bézier Kendall distribution function, $K_{C,\theta}(t)$ be the Kendall distribution for Clayton copula and and $\hat{\theta}$ be the inverted Kendall's tau estimation of dependence parameter $\theta$, then Cramér-Von Mises test statistic $T^C_{\alpha,m}$ under $H_0 : K(t) = K_{(C,\hat{\theta})}(t)$ is given by

$$\begin{aligned}
T^C_{\alpha,m} &= n\sum_{i=0}^{m}\binom{m}{k}^2\alpha_i^2\beta(2i+1,2m-2i+1)\\
&+2n\sum_{0\le i<s\le m}\binom{m}{i}\binom{m}{s}\alpha_i\alpha_s\beta(i+s+1,2m-i-s+1)\\
&-2n\sum_{i=0}^{m}\left(\binom{m}{k}\alpha_i\Big(\beta(i+2,m-i+1)(\frac{\hat{\theta}+1}{\hat{\theta}})-\frac{\beta(i+\hat{\theta}+2,m-i+1)}{\hat{\theta}}\Big)\right)\\
&+n\frac{17+13\hat{\theta}+2\hat{\theta}^2}{27(1+\hat{\theta})+6\hat{\theta}^2}
\end{aligned}$$

where $\beta(.,.)$ is the beta function defined as $\beta(v_1,v_2)=\int_0^1 t^{v_1-1}(1-t)^{v_2-1}dt$ for $v_1, v_2$ positive integers.

*Proof:*

Cramér-von-Mises type test statistic $T^C_{\alpha,m}$ under null hyphothesis of $H_0 : K(t) = K_{(C,\hat{\theta})}(t)$ can be written as

$$T^C_{\alpha,m} = n\int_0^1(\hat{K}_{\alpha,m}(t)-K_{(C,\hat{\theta})}(t))^2dt \tag{5.17}$$

$K_{C,\hat{\theta}}(t)$ can be rewritten as

$$K_{C,\hat{\theta}}(t) = t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}} \tag{5.18}$$

$$\begin{aligned}
T^C_{\alpha,m} &= n \int_0^1 (\hat{K}_{\alpha,m}(t) - (t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}}))^2 dt \\
&= n \int_0^1 \hat{K}^2_{\alpha,m}(t) dt + n \int_0^1 (t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}})^2 dt - 2n \int_0^1 \hat{K}_{\alpha,m}(t)(t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}}) dt \\
&= n \int_0^1 \Big( \sum_{i=0}^{m} \binom{m}{i} t^i (1-t)^{m-i} \alpha_i \Big)^2 dt + n \int_0^1 (t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}})^2 dt \\
&- 2n \sum_{i=0}^{m} \binom{m}{i} \alpha_i \int_0^1 t^i (1-t)^{m-i} (t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}}) dt \\
&= I_1 + I_2 - I_3.
\end{aligned}$$

Part of the $I_1$ is the same as in proof of the proposition 5.4. Now we calculate the part of $I_3$ as

$$\begin{aligned}
I_3 =& 2n \sum_{i=0}^{m} \binom{m}{i} \alpha_i \int_0^1 t^i (1-t)^{m-i} (t + \frac{t}{\hat{\theta}} - \frac{t^{\hat{\theta}+1}}{\hat{\theta}}) dt \\
&= 2n \sum_{i=0}^{m} \binom{m}{i} \alpha_i \Big( (1 + \frac{1}{\hat{\theta}}) \int_0^1 t^{i+1} (1-t)^{m-i} dt - \frac{1}{\hat{\theta}} \int_0^1 t^{i+\hat{\theta}+1} (1-t)^{m-i} dt \Big) \\
&= 2n \sum_{i=0}^{m} \binom{m}{i} \alpha_i \Big( \frac{\hat{\theta}+1}{\hat{\theta}} \beta(i+2, m-i+1) - \frac{1}{\hat{\theta}} \beta(i+\hat{\theta}+2, m-i+1) \Big)
\end{aligned}$$

Finally, after some calculation $I_2$ is calculated by

$$\frac{17 + 13\hat{\theta} + 2\hat{\theta}^2}{27(1+\hat{\theta}) + 6\hat{\theta}^2}$$

.

**Proposition 5.6.** Let $\hat{K}_{\alpha,m}(t)$ be the empirical Bézier Kendall distribution, $K_{F,\theta}(t)$ be the Kendall distribution for Frank copula and and $\hat{\theta}$ be the inverted Kendall's tau estimation of dependence parameter $\theta$, then Cramér-von Mises test statistic $T^F_{\alpha,m}$ under $H_0 : K(t) = K_{(F,\hat{\theta})}(t)$ is given by

$$
\begin{aligned}
T^F_{\alpha,m} &= n\sum_{i=0}^{m}\binom{m}{i}^2\alpha_i^2\beta(2i+1,2m-2i+1) \\
&+ 2n\sum_{i=0}^{m-1}\sum_{s=i+1}^{m}\binom{m}{i}\binom{m}{s}\alpha_i\alpha_s\beta(i+s+1,2m-i-s+1) \\
&- 2n\sum_{i=0}^{m}\left(\binom{m}{i}\alpha_i\Big(\beta(i+2,m-i+1)-\sum_{j=0}^{m-i}\frac{(-1)^j}{\hat{\theta}}\binom{m-i}{j}A(t,i,j)\Big)\right) \\
&+ nB(t,\hat{\theta})
\end{aligned}
$$

where $\beta(.,.)$ is the beta function defined as $\beta(v_1, v_2) = \int_0^1 t^{v_1-1}(1-t)^{v_2-1}dt$ for $v_1, v_2$ positive integers and $A(t,i,j)$ and $B(t,\hat{\theta})$ are numerical solution of the $\int_0^1 t^{i+j}(\exp(t\hat{\theta}) - 1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})$ and $\int_0^1 \Big(t - \frac{(\exp(t\hat{\theta})-1)}{\hat{\theta}}\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})\Big)^2 dt$ respectively.

*Proof:*

Cramér-von-Mises test statistic measures the distance between Bezier empirical distribution function and the Frank copula. Cramér-von-Mises type test statistic $T^F_{\alpha,m}$ can be written as

$$T^F_{\alpha,m} = n\int_0^1(\hat{K}_{\alpha,m}(t) - K_{F,\hat{\theta}}(t))^2dt \tag{5.19}$$

Under null hyphothesis of $H_0 : K(t) = t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}}$

$$\begin{aligned}
T^F_{\alpha,m} &= n\int_0^1 (\hat{K}_{\alpha,m}(t) - (t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}}))^2 dt \\
&= n\int_0^1 \hat{K}^2_{\alpha,m}(t)dt + n\int_0^1 (t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})^2 dt \\
&- 2n\int_0^1 \hat{K}_{\alpha,m}(t)(t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})dt \\
&= n\int_0^1 \Big(\sum_{i=0}^m \binom{m}{i} t^i(1-t)^{m-i}\alpha_i\Big)^2 dt \\
&+ n\int_0^1 (t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})^2 dt \\
&- 2n\sum_{i=0}^m \binom{m}{i}\alpha_i \int_0^1 t^i(1-t)^{m-i}(t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})dt \\
&= I_1 + I_2 - I_3.
\end{aligned}$$

Part of the $I_1$ is the same as in proof of the Proposition 5.4. Now we calculate the part of $I_3$

$$\begin{aligned}
I_3 &= 2n\sum_{i=0}^m \binom{m}{i}\alpha_i \int_0^1 t^i(1-t)^{m-i}(t - \frac{(exp(t\hat{\theta})-1)log(\frac{exp(-t\hat{\theta})-1}{exp(-\hat{\theta})-1})}{\hat{\theta}})dt \\
&= 2n\sum_{i=0}^m \binom{m}{i}\alpha_i \Bigg(\int_0^1 t^{i+1}(1-t)^{m-i}dt \\
&- \int_0^1 t^i(1-t)^{m-i}(\frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})dt\Bigg).
\end{aligned}$$

If we get binomial expansion term $(1-t)^{m-i}$ in second integral of the right side of equation

$$= 2n \sum_{i=0}^{m} \binom{m}{i} \alpha_i \Bigg( \beta(i+2, m-i+1) - \sum_{j=0}^{m-i} (-1)^j \binom{m-i}{j}$$
$$\times \int_0^1 t^{i+j} \Big( \frac{(\exp(t\hat{\theta}) - 1) \log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}} \Big) dt \Bigg)$$

There is no explicit solutions of integral $A(t,i,j) = \int_0^1 t^{i+j} (\frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}}) dt$, we use numerical solution of the integral.

Finally, $I_2$ is calculated by numerical solutions of the integral $nB(t,\hat{\theta}) = n\int_0^1 (t - \frac{(\exp(t\hat{\theta})-1)\log(\frac{\exp(-t\hat{\theta})-1}{\exp(-\hat{\theta})-1})}{\hat{\theta}})^2 dt.$

### *5.2.1 Parametric Bootstrap Method for Tests Based on Bernstein Polynomial and Bézier Curve Estimation*

In order to compute p-values for any test statistic, one requires generating a large number, $N$, of independent samples of size $n$ from $C_{\hat{\theta}_n}(u,v)$ and computing the corresponding values of the selected statistic, such as $S_{n,m}$ or $T_{\alpha,m}$. The procedure would work as follows:

Step 1: Estimate $\theta$ by a consistent estimator $\hat{\theta}_n$. Then, compute test statistic $S_{n,m}$ for data sets of size n

Step 2: Generate $N$ random samples of size $n$ from $C_{\hat{\theta}_n}(u,v)$. Then compute test statistcs for all random samples $N$.

Step 3: If $S_{(1:N),n,m} \leq \ldots \leq S_{(N:N),n,m}$ denote the ordered values of the test statistics calculated in Step 2, an estimate of the critical value of the test at level $\alpha$ based on $S_{n,m}$ is given by

$$\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{I}\Big(S_{(i:N),n,m} \geq S_{n,m}\Big),$$

which yields an estimate of the p-value associated with the observed value $S_{n,m}$ of the statistic.

### 5.3 Performance of the GoF Tests

Performance of the new test statistics given in Section 5.2 is conducted by Monte Carlo simulation study for different levels of Kendall's tau and the alternative Archimedean copulas. The purpose of this Monte Carlo simulation is to determine the level and the power of the underlying test statistics. In order to evaluate the critical value of the proposed test statistic under the null hypothesis at given significance level, data is simulated from $H_1$ copula. Testing procedure consists of rejecting $H_0$ when $S_{n,m}$ and $T_{\alpha,m}$ are greater than the $100(1-\alpha)th$ percentile of its distribution under the null hypothesis. The following factors are used in the simulation;

1. $H_0$ copula (Clayton, Gumbel and Frank)
2. $H_1$ copula (Clayton, Gumbel and Frank)
3. Level of dependence Kendall's tau ($\tau = 0.2, 0.4, 0.6$)
4. Sample sizes ($n = 100$)

In each case, $10.000$ pseudo random samples of size $n = 100$ are generated from the selected model with specified value of Kendall's tau. For each sample, $1.000$ bootstrap samples are generated from the identified Archimedean copula for several levels of Kendall's tau, then $S_{n,m}$ and $T_{\alpha,m}$ statistics are calculated under the null hypothesis $H_0$. The size and the power of the tests are evaluated. In each simulation case, sample of size $n$ is drawn from alternative hypothesis $H_1$ with several levels of Kendall's Tau. The hypotheses are given by

$$H_0 : K(t) = K_{(*,\hat{\theta})}(t)$$
$$H_1 : K(t) \neq K_{(*,\hat{\theta})}(t)$$

where symbol $(*)$ is denoted for G (Gumbel), C (Clayton) and F (Frank). Also, $K_{(G,\hat{\theta})}(t)$, $K_{(C,\hat{\theta})}(t)$ and $K_{(F,\hat{\theta})}(t)$ are summarized in Table 3.1.

Table 5.4 Percentage rejection of $H_0$ by various tests for data sets of size $n = 100$ arising from $H_0$ Copula (Level of the tests)

| | | Test based on | | | | |
|---|---|---|---|---|---|---|
| $H_0$ | $\tau$ | $C_n$ | $S_{n,m=5}$ | $S_{n,m=10}$ | $T_{\alpha,m=5}$ | $T_{\alpha,m=10}$ |
| Gumbel | 0.2 | 0.0506 | 0.0259 | 0.0524 | 0.051 | 0.0494 |
| | 0.4 | 0.0487 | 0.0276 | 0.0474 | 0.052 | 0.0500 |
| | 0.6 | 0.0478 | 0.0313 | 0.0484 | 0.0458 | 0.0520 |
| Clayton | 0.2 | 0.0411 | 0.0158 | 0.0248 | 0.0495 | 0.0478 |
| | 0.4 | 0.0449 | 0.0398 | 0.0477 | 0.0506 | 0.0532 |
| | 0.6 | 0.0475 | 0.0452 | 0.0572 | 0.0512 | 0.0507 |
| Frank | 0.2 | 0.0490 | 0.0146 | 0.0372 | 0.0523 | 0.0457 |
| | 0.4 | 0.0488 | 0.0234 | 0.0386 | 0.0479 | 0.0487 |
| | 0.6 | 0.0452 | 0.0359 | 0.0419 | 0.0446 | 0.0502 |

The size and the power of the tests are summarized in Table 5.4-5.5. Each line of the tables is the percentage rejection of hypothesis $H_0 : K(t) = K_{(*,\hat{\theta})}(t)$. For example, in Table 5.4, when testing the Gumbel copula from sample size n=100, the test based on Bernstein polynomial with order $m = 10$ has power at $0.0524$ level, when the data is from Gumbel copula with Kendall's tau $0.2$. When Table 5.4 is examined, both of the tests based on Bernstein Empirical Kendall distribution $S_{n,m}$, Bézier Empirical Kendall distribution $T_{\alpha,m}$ and classical empirical distribution $C_n$ hold their nominal level $(0.05)$ except from Bernstein based test with polynomial degree $m = 5$

For Table 5.5, we can conclude that when $H_1$ is Frank copula, the test based on Bernstein polynomial with degree $10$ has a better result than the other tests. When $H_1$ is Clayton copula, the classical test has better power result. And also, when $H_1$ is Gumbel copula, Bézier curve test has better results.

The tests based on Bernstein polynomial and Bézier curve approximations are flexible methods since the power of the test can be increased by changing the value of

Table 5.5 Percentage rejection of $H_0$ by various tests for data sets of size $n = 100$ arising from three Archimedean Copula models

| | | | Test based on | | | | |
|---|---|---|---|---|---|---|---|
| $H_0$ | $H_1$ | $\tau$ | $C_n$ | $S_{n,m=5}$ | $S_{n,m=10}$ | $T_{\alpha,m=5}$ | $T_{\alpha,m=10}$ |
| Gumbel | Clayton | 0.2 | 0.6723 | 0.1028 | 0.5748 | 0.4502 | 0.3548 |
| | | 0.4 | 0.968 | 0.3982 | 0.9468 | 0.8672 | 0.8368 |
| | | 0.6 | 0.9962 | 0.7195 | 0.9957 | 0.981 | 0.9616 |
| | Frank | 0.2 | 0.1266 | 0.0888 | 0.2129 | 0.1236 | 0.0659 |
| | | 0.4 | 0.2417 | 0.1846 | 0.3953 | 0.2255 | 0.1789 |
| | | 0.6 | 0.3581 | 0.2657 | 0.4996 | 0.3151 | 0.2793 |
| Clayton | Gumbel | 0.2 | 0.2007 | 0.0413 | 0.2355 | 0.3894 | 0.4970 |
| | | 0.4 | 0.7188 | 0.4696 | 0.9077 | 0.8542 | 0.9109 |
| | | 0.6 | 0.9436 | 0.8785 | 0.9965 | 0.9785 | 0.9888 |
| | Frank | 0.2 | 0.1206 | 0.0535 | 0.1378 | 0.1550 | 0.2506 |
| | | 0.4 | 0.5273 | 0.3617 | 0.7238 | 0.4787 | 0.6366 |
| | | 0.6 | 0.7941 | 0.6924 | 0.9333 | 0.7239 | 0.8442 |
| Frank | Clayton | 0.2 | 0.4949 | 0.0133 | 0.2437 | 0.2896 | 0.2353 |
| | | 0.4 | 0.8448 | 0.0497 | 0.6494 | 0.6244 | 0.5888 |
| | | 0.6 | 0.9392 | 0.1173 | 0.8672 | 0.8325 | 0.7841 |
| | Gumbel | 0.2 | 0.0487 | 0.0029 | 0.0086 | 0.1231 | 0.1792 |
| | | 0.4 | 0.1212 | 0.0060 | 0.0799 | 0.2955 | 0.3379 |
| | | 0.6 | 0.2281 | 0.0107 | 0.2946 | 0.4473 | 0.4422 |

$m$. This is the main advantage when compared with the other tests. So, optimum $m$ can be obtained by increasing the power of the test for a given sample of size $n$.

## 5.4 Real Data Application: LOSS and ALAE Insurance Data

We illustrate an application of three Archimedean copulas (Gumbel, Frank and Clayton) for modelling the real insurance data. The data comprise $1.500$ general liability claims randomly chosen from late settlement lags and were provided by Insurance Services Office, Inc. Each claim consists of an indemnity payment (the loss) and an allocated loss adjustment expense (ALAE). For more details, see Frees & Valdez (1998). For simplicity, $34$ censored data haven't been used. For uncensored data, general degree of dependence is calculated as $0.30865$.

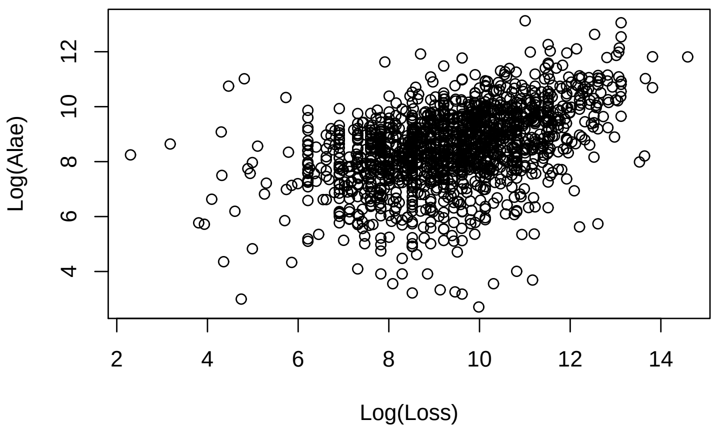



Figure 5.1 Scatterplot of Insurance data for $n = 1500$

Figure 5.1 shows scatter plot of the Insurance data. When Figure 5.2 is examined, right tail dependence is observed. Nonparametric estimate of $\lambda_n(t)$ for Insurance data and parametric estimate of $K_{(*,\hat{\theta})}(t)$ are obtained for three Archimedean copulas (Frank, Clayton and Gumbel) as shown in Figure 5.2. We select $\lambda$-function for the comparison, because $\lambda(t)$ function has better visualization than the function $K(t)$. See for more details, Michiels et al. (2012). In Figure 5.2, Gumbel copula seems to have a good fit for the insurance data, especially at the tails. But there is a deviation center of the graphs between classical estimation of $\lambda_n$ and parametric estimation of Gumbel copula. These visual results verify the goodness-of- fit test results for the classical approximation. Goodness-of-fit results of the Insurance data are shown in Table 5.6. P-values are obtained via $10.000$ bootstrap samples.

Table 5.6 Goodness of fit results for Insurance data

| Copula | $\hat{\theta}_n$ | $C_n, (p - value)$ | | $S_{n,m=10}, (p - value)$ | | $T_{\alpha,m=10}, (p - value)$ | |
|---|---|---|---|---|---|---|---|
| Gumbel | $1.4464$ | $0.081$ | $(0.176)$ | $0.543$ | $(0.980)$ | $0.068$ | $(0.601)$ |
| Frank | $3.0161$ | $0.467$ | $(0.000)$ | $0.572$ | $(0.849)$ | $0.411$ | $(0.021)$ |
| Clayton | $0.8920$ | $3.006$ | $(0.000)$ | $1.557$ | $(0.000)$ | $1.879$ | $(0.000)$ |

When Table 5.6 is examined, Gumbel copula is a good choice for Insurance data sets. P-value for the Gumbel copula increases for the Bernstein and the Bézier based tests which give reliable results for Gumbel copula according to the MISE and the power

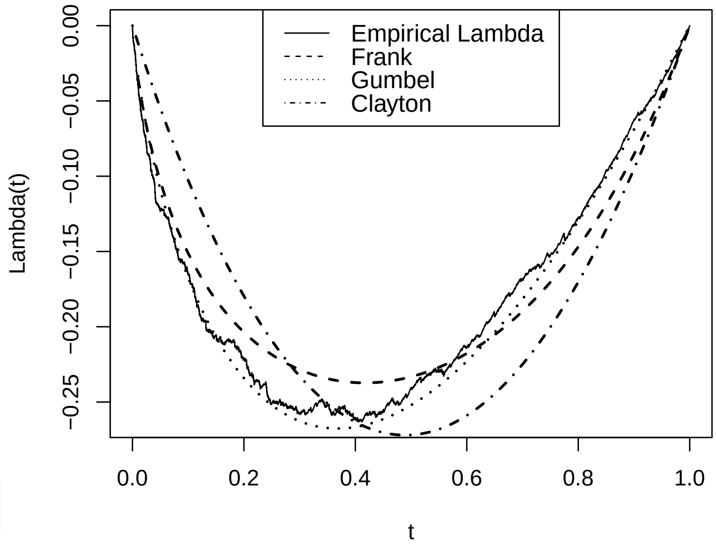



Figure 5.2 Classical Non-parametric and Parametric estimates of $\lambda(t)$ for Insurance Data

results.

For the insurance data, classical test and Bernstein-Bézier polynomial based tests reject the Frank copula with p-values $0.000$, $0.021$, respectively. However, Bernstein polynomial based test does not reject Frank copula with $p-value = 0.849$. Findings obtained in this application confirm the goodness-of-fit results given in Table 5.5.

# CHAPTER SIX
# CONCLUSION

In this thesis, we propose a new family of Archimedean copulas based on Kendall distribution function $K(t)$. We use Bernstein polynomials-Bézier curve to construct this new multi parameter distribution. The method is illustrated for polynomial degree $m = 3, 4, 5$. There are several advantages of this new Archimedean copula class. It is shown that while working with the Bernstein polynomials-Bézier curve structures, a multi-parameter copula family can be constructed in an organized way. It is possible to create a new distribution function which has desirable dependence characteristics using Kendall's tau, lower and upper tail dependence. The parameters of the new model can be interpreted in terms of these dependence characteristics. And also, it is possible that we can create different distributions with the same dependence structures.

Also, we propose two methods of estimating bivariate Archimedean family of copula. We estimate the Kendall distribution function using Bernstein polynomial and Bézier curve approximation. We also establish the MISE values to measure the performance of the estimators. Our simulation suggests that the Bézier estimator has good performance when compared with the other empirical estimators. We show that the Bernstein and the Bézier estimators of Kendall distribution function are flexible by the polynomial degree $m$. The empirical Bernstein polynomial and Bézier curve methods are helpful to construct goodness-of-fit test for Archimedean copulas and also tests of independence.

As a third contrubition of this thesis, we provide a way to test the fit of a parametric copula family for bivariate data. So, we also propose a nonparametric goodness-of-fit test based on Kendall distribution function. We define Cramér-von Mises statistic as a function of the empirical Bernstein and Bézier estimation. The test statistic measures the distance between the estimated Kendall distribution function using Bernstein-Bézier method and the theoretical Kendall distribution function. Finally, a Monte Carlo simulation study is also performed to measure the performance of the proposed test for some Archimedean copula models under given hypothesis and to compare it with the classical empirical Kendall distribution function estimator. The size and the power of the tests are evaluated at a given significance level. We investigate the performance according to the different levels of dependence and sample sizes by comparing the power of classical test which is based on the empirical copula process considered by Genest et al. (2006). The simulation results show that the new Bernstein polynomials and the Bézier curve based estimators have better performances than the classical one since they have flexible forms according to its order $m$. They have satisfactory estimates of the underlying copula when compared with the other empirical methods.

In copula goodness-of-fit test methodology, Bernstein polynomials based estimations not only fix the jump discontinuities problem of classical empirical estimation, but also decrease the estimation error. Thus, the power of the goodness-of-fit test increases with the Bernstein polynomial estimation of Kendall distribution function. And also, the tests based on Bernstein polynomials and Bézier curve approximations outperform the classical one. Also, Bernstein polynomials based tests perform better results in real data applications. On the other hand, real data application results verify the goodness of fit results for the proposed tests.

## REFERENCES


Babu, G. J., Canty, A. J., & Chaubey, Y. P. (2002). Application of bernstein polynomials for smooth estimation of a distribution and density function. *Journal of Statistical Planning and Inference*, *105*, 377-392.

Barbe, P., Genest, C., Ghoudi, K., & R/´emillard, B. (1996). On kendall's process. *Journal of Multivariate Analysis*, *58*, 197-229.

Belalia, M., Bouezmarni, T., Lemyre, F. C., & Taamouti, A. (2017). Testing independence based on bernstein empirical copula and copula density. *Journal of Nonparametric Statistics*, *29*(2), 346-380.

Berg, D. (2009). Copula goodness-of-fit testing: an overview and power comparison. *The European Journal of Finance*, *15*, 675-701.

Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Nonparametric Statistics*, *19*, 465-474.

Bernstein, S. (1912). Démo istration du th'eorème de weierstrass fondée sur le calcul des probabilités. *Сообщенія Харьковскаго математическаго общества*, *13*(1), 1–2.

Blum, J. R., Kiefer, J., & Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics*, *32*(2), 485-498.

Breymann, W., Dias, A., & Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance. *Quantitative Finance*, *1*, 1-14.

Charpentier, A., & Segers, J. (2009). Tails of multivariate archimedean copulas. *Journal of Multivariate Analysis*, *100*, 1521-1537.

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, *65*, 141-151.

Cook, R., & Johnson, M. (1981). A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society B*, *43*, 210-218.

Cooray, K. (2018). Strictly archimedean copulas with complete association for multivariate dependence based on the clayton family. *Dependence Modelling*, *6*, 1-18.

Deheuvels, P. (1978). Caracterisation complete des lois extremes multivariees et de la convergence des types extremes. *Publications de l'Institut de Statistique de l'Université de Paris*, *3*, 1-36.

Diebold, F. X., Hahn, J., & Tay, A. S. (1999). Multivariate density forecast and calibration in financial risk management: High-frequency returns on foreign exchange. *The Review of Economics and Statistics*, *81*(4), 661-673.

Dimitrova, D., Kaishev, K., & Penev, I. (2008). Ged spline estimation of multivariate archimedean copulas. *Computational Statistics and Data Analysis*, *52*, 3570-3582.

Duncan, M. (2005). *Applied Geometry for Computer Graphics and CAD*. London: Springer.

Embrechts, P., McNeil, A., & Straumann, D. (2002). *Correlation and dependence in risk management: Properties and pitfalls*. Cambridge: Cambridge University Press.

Erdoğan, M. S., Dişibüyük, Ç., & Oruç, Ö. E. (2019). An alternative distribution function estimation method using rational bernstein polynomials. *Journal of Computational and Applied Mathematics*, *353*, 232–242.

Feller, W. (1965). *An Introduction to Probability Theory and its Applications*. New York: Wiley.

Fermanian, J., Radulovic, D., & Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, *10*, 847-860.

Fisher, N. (1997). *Copulas*. New York: John Wiley and Sons.

Frank, M. (1979). On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Æquationes Mathematiæ*, *19*, 194-226.

Frees, E., & Valdez, E. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, *2*, 1-25.

Genest, C., & MacKay, R. (1986). Copules archimediennes et families de lois bidimensionelles dont les marges sont donnees. *Canadian Journal of Statistics*, *14*, 145-159.

Genest, C., Molina, J., & Lallena, J. (1995). De l ' impossibilite de construire des lois a marges multidimensionnelles donnees a partir de copules. *Comptes rendus de l'Académie des Sciences*, *320*, 723-726.

Genest, C., Quessy, J., & Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transform. *Scandinavian Journal of Statistics*, *33*, 337-366.

Genest, C., & Rivest, L. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, *88*, 1034-1043.

Gumbel, E. J. (1960). Distributios des valeurs extrémes en plusiers dimensions. *Publications de l'Institut de Statistique de l'Université de Paris*, *9*, 171-173.

Joe, H. (1997). *Multivariate Models and Dependence Concepts, Monographs on Statistics and Applied Probability*. London: Chapmann and Hall.

Joe, H., & Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models.

Kaishev, V., Dimitrova, D., Haberman, S., & Verrall, R. (2006). Geometrically designed, variable knot regression splines: Asymptotics and inference. *Statistical Research Paper*, *28*, 1-37.

Kim, C. (1996). Nonparametric density estimation via the bezier curve. In *Proceedings of the Section on Statistical Graphics of the American Statistical Association* ( 25–28).

Kim, C., Kim, W., Park, B. U., Hong, C., & Jeong, M. (1999). Smoothing techniques via the bezier curve. *Communications in Statistics-Theory and Methods*, *28*(7), 1577–1597.

Kimeldorf, G., & Sampson, A. R. (1975). Uniform representations of bivariate distributions. *Communication in Statistics*, *4*, 617-627.

Kojadinovic, I., & Yan, J. (2010). Comparison of three semiparametric methods for estimating dependence parameters in copula models. *Insurance: Mathematics and Economics*, *47*, 52-63.

Lambert, P. (2007). Archimedean copula estimation using bayesian splines smoothing techniques. *Computational Statistics and Data Analysis*, *51*, 6307-6320.

Leblanc, A. (2009). Chung-smirnov property for bernstein estimator of distribution functions. *Journal of Nonparametric Statistics*, *21*, 133-142.

Leblanc, A. (2012). On estimating distribution functions using bernstein polynomials. *Annals of the Institute of Statistical Mathematics*, *64*, 919-943.

Lorentz, G. G. (1986). *Bernstein Polynomials*. Chelsea: Chelsea Publishing Company.

Michiels, F., Koch, I., & Schepper, A. A. D. (2011). New method for the construction of civariate archimedean copulas based on the $\lambda$ function. *Communications in Statistics - Theory and Methods*, *40*(15), 2670-2679.

Michiels, F., Koch, I., & Schepper, A. D. (2012). How to improve the fit of archimedean copulas by means of transforms. *Statistical Papers*, *53*, 345-355.

Najjari, V., Bacigal, T., & Bal, H. (2014). An archimedean copula family with hyperbolic cotangent generator. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, *20*(5), 761-768.

Nelsen, R. (2006). *An introduction to copulas*. New York: Springer.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of Royal Statistical Society B*, *44*, 414-422.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of American Statistical Association*, *84*, 487-493.

Oruc, E., Erdogan, M., & Oruc, H. (2017). Probabilistic approach to the schoenberg spline operator and unimodal density estimator. *ISTATISTIK, Journal of the Turkish Statistical Association*, *10*, 33-39.

Plackett, R. T. (1965). A class of bivariate distributions. *Journal of American Statistician Associative*, *60*, 516-522.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics*, *23*, 470-472.

Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, *8*, 229–231.

Susam, S., & Ucer, B. (2018). Testing independence for archimedean copula based on bernstein estimate of kendall distribution function. *Journal of Statistical Computation and Simulation*, *88*, 2589-2599.

Tawn, L. A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika*, *75*, 397-415.

Trivedi, P., & Zimmer, D. (2007). Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, *1*(1), 1-111.

Vitale, R. A. (1975). A Bernstein polynomial approach to density function estimation. In *Statistical inference and related topics* ( 87–99). USA:Elsevier.

Wang, J., & Ghosh, S. K. (2012). Shape restricted nonparametric regression with bernstein polynomials. *Computational Statistics and Data Analysis*, *56*, 2729-2741.

Wang, W., & Wells, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association*, *95*, 62-72.

Weierstrass, K. (1885). Über die analytische darstellbarkeit sogenannter willkürlicher functionen einer reellen veränderlichen. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, *2*, 633–639.